Abstract--- Object detection is a fundamental step for automated video analysis in many vision applications. Object detection in a video is usually performed by object detectors or background subtraction techniques. Often, an object detector requires manually labeled examples to train a binary classifier, while background subtraction needs a training sequence that contains no objects to build a background model. To automate the analysis, object detection without a separate training phase becomes a critical task. Several experiments have been done using motion information. But existing motion-based methods are usually limited when coping with complex scenarios such as nonrigid motion and dynamic background. The project Motion Detection in Low Rank Representation addresses the above challenges in a unified framework named DETecting Contiguous Outliers in the LOw-rank Representation (DECOLOR). This formulation integrates object detection and background learning into a single process of optimization, which can be solved by an alternating algorithm efficiently. Experiments on both simulated data and real sequences demonstrate that DECOLOR outperforms the state-of-the-art approaches and it can work effectively on a wide range of complex scenarios.

Keywords: - Background subtraction, Low Rank Modeling, Motion segmentation Moving Object detection

I. INTRODUCTION

Video surveillance systems have long been in use to monitor security sensitive areas. The making of video surveillance systems “smart” requires fast, reliable and robust algorithms for moving object detection, classification, tracking and activity analysis[1][2]. There are three key steps for automated video analysis: object detection, object tracking, and behavior recognition. As the first step, object detection aims to locate and segment interesting objects in a video. Then, such objects can be tracked from frame to frame, and the tracks can be analyzed to recognize object behavior. Thus, object detection plays a critical role in practical applications. Moving object detection is the basic step for further analysis of video. It handles segmentation of moving objects from stationary background objects.

Object classification step categorizes detected objects into predefined classes such as human, vehicle, animal, clutter, etc. It is necessary to distinguish objects from each other in order to track and analyse their actions reliably. In previous system background subtraction [3] is performed by using Canny Edge Detection [2]. In Canny Edge Detection process taking two images for comparison those are background image and foreground image. Backend image are images which is already stored. Foreground images are images which are captured by the webcam and these are compared with the background image to get the status.

The most natural way for motion-based object detection is to classify pixels according to motion patterns, which is usually named motion segmentation [4]. These approaches achieve both segmentation and optical flow computation accurately and they can work in the presence of large camera motion. However, they assume rigid motion or smooth motion in respective regions, which is not generally true in practice. In practice, the foreground motion can be very complicated with nonrigid shape changes. Also, the background may be complex, including illumination changes and varying textures such as waving trees and sea waves. It is difficult to handle the scenarios with complex background or moving cameras.

Previous strategies for object detection are immense, including object detectors, image segmentation, Background subtraction, etc. This methodology aims to segment objects supported motion info and it comprises an element of background modeling. In the previous strategies background subtraction is performed only for images. For this paper a pixel wise background modeling [5] and subtraction technique using multiple features is proposed. Hence, in this color, gradient and Haar-Like features are integrated to handle the variations in each pixel. Thus, motion segmentation and background subtraction are the most connected topics to the current paper.

In this paper, we propose a novel algorithm for moving object detection which falls into the category of motion-based methods. It solves the challenges mentioned above in a unified framework named DETecting Contiguous Outliers in the LOw-rank Representation (DECOLOR) [6]. We assume that the underlying background images are linearly correlated. Thus, the matrix composed of vectorized video frames can be approximated by a low-rank matrix, and the moving objects can be detected as outliers in this low-rank representation. Formulating the problem as outlier detection allows us to get rid of many assumptions on the behavior of foreground. The low-rank representation of background makes it flexible to accommodate the global variations in the background. Moreover, DECOLOR performs object detection and background estimation simultaneously without training sequences. The main contributions can be summarized as follows:

1. We propose a new formulation of outlier detection in the low-rank representation in which the outlier support and the low-rank matrix are estimated simultaneously.

We establish the link between our model and other relevant models in the framework of Robust Principal Component Analysis (RPCA) [7] differently from other formulations of RPCA; we model the outlier support explicitly. DECOLOR can be interpreted as ‘$\ell_0$-penalty’ regularized RPCA, which is a more faithful model for the problem of moving object segmentation. DECOLOR achieves better accuracy in terms of both object detection and background estimation compared

Motion Detection in Low Rank Representation
Midhun Omanakuttan$^1$ Mrs.Porkodi Prabhakaran$^2$

$^1$Final Year M.E CSE $^2$Assistant Professor, Department of CSE, Al-ameen college of Engineering, Erode
against the state-of-the-art algorithm of RPCA.

2. In other models of RPCA, no prior knowledge on the spatial distribution of outliers has been considered. In real videos, the foreground objects usually are small clusters. Thus, contiguous regions should be preferred to be detected. Since the outlier support is modeled explicitly in our formulation, we can naturally incorporate such contiguity prior using Markov Random Fields (MRFs) [8].

II. RELATED WORK

Previous methods for object detection are vast, including object detectors (supervised learning), image segmentation, background subtraction, etc., our method aims to segment objects based on motion information and it comprises a component of background modeling. Thus, motion segmentation and background subtraction are the most related topics to this paper.

A. Motion Segmentation

In motion segmentation, the moving objects are continuously present in the scene, and the background may also move due to camera motion. The target is to separate different motions.

A common approach for motion segmentation is to partition the dense optical-flow field [9] this is usually achieved by decomposing the image into different motion layers. The assumption is that the optical-flow field should be smooth in each motion layer, and sharp motion changes only occur at layer boundaries. Dense optical flow and motion boundaries are computed in an alternating manner named motion competition, which is usually implemented in a level set framework. A similar scheme is later applied to dynamic texture segmentation. While high accuracy can be achieved in these methods, accurate motion analysis itself is a challenging task due to the difficulties raised by aperture problem, occlusion, video noises, etc... Moreover, most of the motion segmentation methods require object contours to be initialized and the number of foreground objects to be specified.

An alternative approach for motion segmentation tries to segment the objects by analyzing point trajectories. Some sparse feature points are first detected and tracked throughout the video and then separated into several clusters via subspace clustering or spectral clustering. The formulation is mathematically elegant and it can handle large camera motion. However, these methods require point trajectories as input and only output a segmentation of sparse points. The performance relies on the quality of point tracking and post processing is needed to obtain the dense segmentation. Also, they are limited when dealing with noisy data and nonrigid motion

B. Background Subtraction

In background subtraction, the general assumption is that a background model can be obtained from a training sequence that does not contain foreground objects. Moreover, it usually assumes that the video is captured by a static camera. Thus, foreground objects can be detected by checking the difference between the testing frame and the background model built previously.

A considerable number of works have been done on background modeling, i.e., building a proper representation of the background scene. Typical methods include single Gaussian distribution[9], Mixture of Gaussian (MoG)[10], kernel density estimation[11], block correlation, codebook model, Hidden Markov model[12], and linear autoregressive models.

Learning with sparsity has drawn a lot of attention in recent machine learning and computer vision research, and several methods based on the sparse representation for background modeling have been developed. One pioneering work is the Eigen backgrounds model, where the principal component analysis (PCA) is performed on a training sequence. When a new frame arrives, it is projected onto the subspace spanned by the principal components, and the residues indicate the presence of new objects. An alternative approach that can operate sequentially is sparse signal recovery[12]. Background subtraction is formulated as a regression problem with the assumption that a new-coming frame should be sparsely represented by a linear combination of preceding frames except for foreground parts. These models capture the correlation between video frames. Thus, they can naturally handle global variations in the background such as illumination change and dynamic textures.

Background subtraction methods mentioned above rarely consider the scenario where the objects appear at the start and are continuously present in the scene (i.e., the training sequence is not available). Very little literature considers the problem of background initialization. Most of them seek a stable interval, inside which the intensity is relatively smooth for each pixel independently. Pixels during such intervals are regarded as background, and the background scene is estimated from these intervals. The validity of this approach relies on the assumption of static background. Thus, it is limited when processing dynamic background or videos captured by a moving camera.

III. CONTIGUOUS OUTLIER DETECTION IN THE LOW-RANK REPRESENTATION

This section, focus on the problem of detecting contiguous outliers in the low-rank representation[13]. First consider the case without camera motion.

A. Notations

This paper, use following notations. $I_j \in \mathbb{R}^{m \times n}$ denotes the jth frame of a video sequence, which is written as a column vector consisting of m pixels. The ith pixel in the jth frame is denoted as $i_j$. $D=[I_1; \ldots; I_L] \in \mathbb{R}^{m \times n \times L}$ is a matrix representing all n frames of a sequence. $B \in \mathbb{R}^{m \times n}$ is a matrix with the same size of $D$, which denotes the background image. $S \in \{0, 1\}^{n \times n}$ is a binary matrix denoting the foreground support:

\[ S_{ij} = \begin{cases} 0, & \text{if } i \text{ is background} \\ 1, & \text{if } i \text{ is foreground} \end{cases} \]

We use $P_s(X)$ to represent the orthogonal projection of a matrix $X$ onto the linear space of matrices supported by $S$,

\[ P_s(X) = \begin{cases} X_{ij} = 0, & \text{if } S_{ij} = 0 \\ X_{ij}, & \text{if } S_{ij} = 1 \end{cases} \]

B. Formulation

Given a sequence $D$, our objective is to estimate the foreground support $S$ as well as the underlying background images $B$. To make the problem well posed, we have the following models to describe the foreground, the back-
ground, and the formation of observed signal.

C. Background model. The background intensity should be unchanged over the sequence except for variations arising from illumination change or periodical motion of dynamic textures. Thus, background images are linearly correlated with each other, forming a low-rank matrix B. Besides the low-rank property, we don’t make any additional assumption on the background scene. Thus, we only impose the following constraint on B:

\[
\text{Rank}(B) \leq K;
\]

Where K is a constant to be predefined. Intrinsically, K constrains the complexity of the background model.

D. Foreground model. The foreground is defined as any object that moves differently from the background. Foreground motion gives intensity changes that cannot be fitted into the low-rank model of background. Thus, they can be detected as outliers in the low-rank representation. Generally, we have a prior that foreground objects should be contiguous pieces with relatively small size. The binary states of entries in foreground support S can be naturally modeled by a Markov Random Field. To prefer \( S_{ij} = 0 \) that indicates sparse foreground, we define the unary potential \( u_{ij} \) as

\[
U_{ij}(S_{ij}) = \begin{cases} 0, & \text{if } S_{ij} = 0 \\ |A_{ij}|, & \text{if } S_{ij} = 1 \end{cases}
\]

E. Signal model. The signal model describes the formation of D, given B and S. In the background region where \( S_{ij} = 0 \), we assume that \( D_{ij} = B_{ij} + \varepsilon_{ij} \) where \( \varepsilon_{ij} \) denotes i.i.d. Gaussian noise. That is, \( D_{ij} \sim \mathcal{N}(B_{ij}; \sigma^2) \) with \( \sigma^2 \) being the variance of Gaussian noise. Then, \( B_{ij} \) should be the best fitting to \( D_{ij} \) in the least squares sense when \( S_{ij} = 0 \). In the foreground regions where \( S_{ij} = 1 \), the background scene is occluded by the foreground. Thus, \( D_{ij} \) equals the foreground intensity. Since we don’t make any assumption about the foreground appearance, \( D_{ij} \) is not constrained when \( S_{ij} = 1 \).

Combining above three models, we propose to minimize the following energy to estimate B and S:

\[
\text{ie, rank}(B) \leq K
\]

This formulation says that the background images should form a low-rank matrix and fit the observed sequence in the least squares sense except for foreground regions that are sparse and contiguous.

To make the energy minimization tractable, we relax the rank operator on B with the nuclear norm. The nuclear norm has proven to be an effective convex surrogate of the rank operator. Moreover, it can help to avoid over fitting.

F. Low Rank Modelling: The matrix completion problem is to recover a low-rank matrix from a subset of its entries. The main solution strategy for this problem has been based on nuclear-norm minimization which requires computing singular value decompositions – a task that is increasingly costly as matrix sizes and ranks increase. To improve the capacity of solving large-scale problems, we propose a low-rank factorization model and construct nonlinear successive over-relaxation (SOR) [11] algorithm that only requires solving a linear least squares problem per iteration. Convergence of this nonlinear SOR algorithm is analyzed. Numerical results show that the algorithm can reliably solve a wide range of problems at a speed at least Several times faster than many nuclear-norm minimization algorithms.

The problem of minimizing the rank of a matrix arises in many applications, for example, control and systems theory, model reduction and minimum order control synthesis recovering shape and motion from image streams data mining and pattern recognitions and machine learning such as latent semantic indexing, collaborative prediction and low-dimensional embedding. In this paper, we consider the Matrix Completions(MC) problem of finding a lowest-rank matrix given a subset of its entries, that is, \( W \in \mathbb{R}^{m \times n} \) where \( \text{rank}(W) \) denotes the rank of \( W \), and \( M_{ij} \in \mathbb{R} \) are given for \( (i, j) \in \Omega \), where \( \Omega \) is the subset of \( \mathbb{R}^{m \times n} \).

\[
\text{Convex optimization problem}
\]

\[
\begin{align*}
\min & W \in \mathbb{R}^{m \times n} \\
\text{s.t.} & W_{ij} = M_{ij}, \forall (i, j) \in \Omega,
\end{align*}
\]

Where the nuclear or trace norm \( kWk^* \) is the summation of the singular values of \( W \). In particular, Candès and ’Resht proved in [12] and [13] that a given rank- \( r \) matrix \( M \) satisfying certain incoherence conditions can be recovered exactly by with high probability from a subset \( \Omega \) of uniformly sampled entries whose cardinality \( |\Omega| \) is of the order \( O(r(m + n)\text{polylog}(m + n)) \).

H. Basic Low Ranking Approximation

Minimize over \( D_{ij} \) subject to \( \text{rank}(D_{ij}) \leq r \) has analytic solution in terms of the singular value decomposition of the data matrix. The result is referred to as the matrix approximation lemma or Eckart–Young–Mirskey theorem

\[
D = U \Sigma V^*
\]

where U is a \( m \times m \) real or complex unitary matrix, \( \Sigma \) is an \( m \times n \) rectangular diagonal matrix with nonnegative real numbers on the diagonal, and \( V^* \) (the conjugate transpose of \( V \), or simply the transpose of \( V \) if \( \mathbb{R} \) is real) is an \( m \times n \) real or complex unitary matrix. The diagonal entries \( \Sigma_{ij} \) of \( \Sigma \) are known as the singular values of \( M \). The m columns of \( U \) and the n columns of \( V \) are called the left-singular vectors and right-singular vectors of \( M \), respectively. The Frobenius norm weights uniformly all elements of the approximation error \( DD^* \). Prior knowledge about distribution of the errors can be taken into account by considering the weighted low-rank approximation problem. The general weighted low-rank approximation problem does not admit an analytic solution in terms of the singular value decomposition and is solved by local optimization methods. The image representation of the rank constraint suggests a parameter optimization methods, in which the cost function is minimized alternatively over one of the variables (P or L) with the other one fixed. Although simultaneous minimization over both P and L is a difficult no convex optimization problem, minimization over one of the variables alone is a linear least squares problem and can be solved globally and efficiently.

The resulting optimization algorithm (called alternating projections) is globally convergent with a linear convergence rate to a locally optimal solution of the weighted low-rank approximation problem. Starting value for the P (or L) parameter should be given. The iteration is
stopped when a user defined convergence condition is satisfied.

I. Parameter Tuning

The parameter α controls the complexity of the background model. In our algorithm, we first give a rough estimate to the rank of the background model, i.e., K in . Then, we start from a large α . After each run of SOFT-IMPUTE[10], if rank(B)≤K, we reduce and repeat SOFT-IMPUTE until rank(B)≥K . In conclusion, DECOLOR performs stably if K and γ are in proper ranges.

J. Motion segmentation

Background subtraction is the first step in the process of segmenting and tracking people. Distinguishing between foreground and background in a very dynamic and unconstrained outdoor environment over several hours is a challenging task. The background model is kept in the data storage and four individual modules do training of the model, updating of the model, foreground/background classification and post processing. The first k video frames are used to train the background model to achieve a model that represents the variation in the background during this period. The following frames (from k + 1 and onwards) are each processed by the background subtraction module to produce a mask that describes the foreground regions identified by comparing the incoming frame with the background model. Information from frames k + 1 and onwards are used to update the background model either by the continuous update mechanism, the layered Updating, or both. The mask obtained from the background subtraction is processed further in the post processing module, which minimizes the effect of noise in the mask.

IV. Relation To Other Methods

A. Robust Principal Component Analysis

RPCA has drawn a lot of attention in computer vision . Recently, the seminal work showed that, under some mild conditions, the low-rank model can be recovered from unknown corruption patterns via a convex program named Principal Component Pursuit (PCP)[8]. The examples demonstrate the superior performance of PCP compared with previous methods of RPCA and its promising potential for background subtraction.

DECOLOR can be regarded as a special form of RPCA where the γ-penalty on E is not relaxed and the problem in is converted to the optimization over Sin. One recent work has shown that the γ-penalty works effectively for outlier detection in regression, while the γ1-penalty does not. The theoretical reason for the unsatisfactory performance of the γ-penalty is that the irrepresentable condition is often not satisfied in the outlier detection problem. In order to go beyond the γ1-penalty, nonconvex penalties have been explored in recent literature . Compared with the γ1-norm, nonconvex penalties give an estimation with less bias but higher variance. Thus, these nonconvex penalties are superior to the γ1-penalty when the signal-noise-ratio (SNR)[5] is relatively high . For natural video analysis, it is the case.

In summary, both PCP and DECOLOR aim to recover a low-rank model from corrupted data. PCP uses the convex relaxation by replacing rank(B) with ||B||γ and ||E||0 with ||E||. DECOLOR only relaxes the rank penalty and keeps the γ-penalty on E to preserve the robustness to outliers. Moreover, DECOLOR estimates the outlier support S explicitly by formulating the problem as the energy minimization over S, and models the continuity prior on S using MRFs to improve the accuracy of detecting contiguous outliers.

B. Sparse Signal Recovery

With the success of compressive sensing , sparse signal recovery has become a popular framework to deal with various problems in machine learning and signal processing . To make use of structural information about nonzero patterns of variables, the structured sparsity is defined in recent works and several algorithms have been developed and applied successfully on back-ground subtraction, such as Lattice Matching Pursuit (LaMP)[10] Dynamic Group Sparsity (DGS) recovery[11], and Proximal Operator using Network Flow (ProxFlow)[12].

LaMP, DGS, and ProxFlow aim to detect new objects in a new testing image given a training sequence not containing such objects. The problem is formulated as linear regression with fixed bases. DECOLOR aims to segment moving objects from a short sequence during which the objects continuously appear, which is a more challenging problem. To this end, DECOLOR estimates the foreground and background jointly by outlier detection during matrix learning.

A key difference between DECOLOR and sparse signal recovery is the assumption on availability of training sequences. Background subtraction via sparse signal recovery requires a set of background images without foreground, which is not always available, especially for surveillance of crowded scenes.

For sparse signal recovery, we apply the ProxFlow algorithm to solve the model in . ProxFlow cannot recover the background and gives inaccurate segmentation. Instead, DECOLOR can estimate a clean background from occluded data. In practice, DECOLOR can be used for background initialization. DECOLOR maintains better performance than PCP if SNR is relatively high, but drops dramatically after SNR < 2. This can be interpreted by the property of nonconvex penalties. Compared with γ1-norm, nonconvex penalties are more robust to gross errors but more sensitive to entrywise perturbations. In general cases of natural video analysis, SNR is much larger than 1. Thus, DECOLOR can work stably.

Since we aim to evaluate the ability of algorithms in detecting moving objects at the start of videos, we focus on short clips composed of beginning frames of videos. We compare DECOLOR with three methods that are simple in implementation but effective in practice. The first one is PCP, which is the state-of-the-art algorithm for RPCA.

The second method is median filtration, a baseline method

![Fig 1: An example of smoke detection. (a) Sample frame. (b) Estimated background. (c) Segmentation.](image)
for unimodal background modeling. The median intensity value around each pixel is computed, forming a background image. Then, each frame is subtracted by the background image and the difference is threshold to generate a foreground mask. The advantage of using median rather than mean is that it is a more robust estimator to avoid blending pixel values, which is more proper for background estimation. The third method is mixture of Gaussians. It is popularly used for multimodal background modeling and has proven to be very competitive compared with other more sophisticated techniques for background subtraction.

The low-rank modeling of background gives better results with less false detection on the water surface and DECOLOR obtains a cleaner background compared against PCP.

Dynamic texture segmentation has drawn some attention in recent computer vision research. While we have shown that DECOLOR can model periodically varying textures like escalators or water surfaces as background, it is also able to detect fast changing textures whose motion has little periodicity and cannot be modeled as low rank. The memory costs of DECOLOR and PCP are almost the same since both of them need to compute SVD[9].

V. CONCLUSION AND FUTURE WORK

In this paper, we propose a novel framework named DECOLOR to segment moving objects from image sequences. It avoids complicated motion computation by formulating the problem as outlier detection and makes use of the low-rank modeling to deal with complex background.

We established the link between DECOLOR and PCP. Compared with PCP, DECOLOR uses the nonconvex penalty and MRFs for outlier detection, which is more greedy to detect outlier regions that are relatively dense and contiguous. Despite its satisfactory performance in our experiments, DECOLOR also has some disadvantages. Since DECOLOR minimizes a nonconvex energy via alternating optimization, it converges to a local optimum with results depending on initialization of $S^o$, while PCP always minimizes its energy globally. In all our experiments we simply start from $S^o=0$. Also we have tested other random initialization of $S^o$ and it generally converges to a satisfactory result. This is because the SOFT-IMPUTE step will output similar results for each randomly generated as long as is $S^o$ not too dense.

DECOLOR may misclassify unmoved objects or large texture less regions as back-ground since they are prone to entering the low-rank model. To address these problems, incorporating additional models such as object appearance or shape prior to improve the power of DECOLOR can be further explored in future.

Currently, DECOLOR works in a batch mode. Thus, it is not suitable for real-time object detection. In the future, we plan to develop the online version of DECOLOR that can work incrementally, e.g., the low-rank model extracted from beginning frames may be updated online when new frames arrive.

REFERENCES


[6] Xiaowei ,Can Yang, And Wei chuan Yu “Moving Object Detection By Detecting Contiguous Outliers In The Low Rank Representation”


