Available online at www.sciencedirect.com

SciVerse ScienceDirect

AASRI Procedia 1 (2012) 492 – 497

AASRI Procedia

www.elsevier.com/locate/procedia

2012 AASRI Conference on Computational Intelligence and Bioinformatics

Moving-object Detection Based on Sparse Representation and Dictionary Learning

Xiaosheng Huang[a],Feng Wu[a,*],Ping Huang[a]

[a]*School of Information Engineering,East China Jiaotong University,Nanchang,330013,China*

Abstract

In this paper, we propose an algorithm of moving-object detection via the sparse representation and learned dictionary. First, compress image in order to reduce data redundancy and bandwidth. Then initialize data dictionary with CS measurement values and sparse basis, train and update it through the K-SVD which can get the sparsest representations. At the same time, we consider the correlation between the dictionaries which can effectively reduce the dictionary redundancy. Carry out the selective reconstruction depending on the sparse coefficients to determine whether the target invades, which can decrease the amount of data to calculate and is better to achieve real-time. We segment the moving-object through the robust principal component pursuit(PCP) for that the image is consisted with low-rank of the background regions and the sparsity of the foreground regions. Analysis, simulation, and experimental results show that our scheme has a good detection result, which can significantly decrease data redundancy and the demand for bandwidth at the same time.

Keywords: dictionary learning; sparse sepresentation; moving-object detection ; principal component pursuit

1. Introduction

Moving-object detection is one of the core problems in computer vision, which is widely applied in the fields of artificial intelligence, intelligent monitoring, video analysis and so on.Common methods for moving-object detection are optical flow method[1], inter-frame difference method[2] and background subtraction

* Corresponding author. Tel.: +8615270896920.
E-mail address: hawf1988@163.com.

doi:10.1016/j.aasri.2012.06.077

method[3].Those algorithms are all based on pixel values, which the common drawback is data redundancy occupied lots of bandwidth and complex calculation. To solve the problem of large amount of data, scholars have introduced compressive sensing (CS) technology[4-5], which can reduce the transmission bandwidth and storage memory effectively. Unfortunately, signal decomposition of the above algorithms uses nonlinear orthogonal transformation which can't consider the correlation between images, making signal less sparse.

However, dictionary learning methods determine the proper representation of data via decreased dimensionality subspaces, which are appropriate for the characteristics of the signals and image processing. These representations are based on the principle that our observations can be described by a sparse subset of atoms taken from a redundant dictionary, which represents the causes of observations in the real world.

In this paper, we propose an algorithm of moving-object detection via the sparse representation and learned dictionary. Make the sparse coefficient sparser according to the CS measurements and sparse basis which are used to train dictionary initially. Update the dictionary continually through K-SVD to strengthen the sparse representation of the dictionary capabilities. Reconstruct the image depended on whether object occurs, which can reduce the amount of data to calculate and is better to achieve real-time. Finally, image consisted of low-rank of the background component and the sparsity of the foreground component can be recovered through the robust principal component pursuit(PCP) individually.

2. Sparse Representation of image

At the phase of video acquisition, write a two-dimensional image of each frame one-dimensional column vector in $x_i \in R^N$, where $i = 1, 2, \cdots, \mathrm{I}$, I is the total frames of this video, N is the total pixels of an image. $X = \left[x_1, x_2, \cdots, x_{\mathrm{I}}\right] \in R^{N \times \mathrm{I}}$, each column vector is sub-modules of streaming video, and independent of each other.

Reduce dimensionality of a video sequence X via compressive sensing[6-7]:

$$Y = \phi X = \phi\varphi\theta \tag{1}$$

where φ is the sparse basis, ϕ is the measurement matrix, $\phi \in R^{M \times N}\,(M < N)$, θ is sparse coefficient matrix, Y is measurements, $Y = \left[y_1, y_2, \cdots, y_{\mathrm{I}}\right]$, each column values correspond to X. When (1) satisfy RIP[8], we can solve X by measurements Y to achieve accurate reconstruction:

$$\hat{\theta} = \arg\min \left\|\theta\right\|_2 \quad s.t. \;\; \phi\varphi\theta = Y \tag{2}$$

Where $\left\|\bullet\right\|_2$ is the l^2 norm. Then we can get X through (1).

There is Gaussian noise in the image, can use the following formula to denoising:

$$\hat{X} = \arg\min \left\|X\right\|_2 \quad s.t. \;\; \left\|\phi\varphi\theta\text{-Y}\right\|_2 \le \varepsilon, X = \varphi\theta \tag{3}$$

During transmission, the transmission of just the measurements can restore the original high-dimensional image ($M < N$), which reduces the bandwidth of the signal transmission. Encoding and decoding of compressive sensing are used for reducing the data redundancy, but failed to consider the relation between adjacent frames. So we mention the dictionary in the next section.

3. Learning and Updating Dictionary Based on K-SVD

There are lots of studies on signal sparse representation. Using the data dictionary[9-10] to sparse represent the original images which is consisted of sparse linear combination of dictionary atoms and can effectively resolve the perceived deficiencies in the image processing. In this paper, adaptive dictionary to get the most

sparse dictionary to represent the CS measurements based on the K-SVD algorithm. In addition, the update of the dictionary can exclude the interference of the external environment (light, etc.). The K-SVD algorithm is promoted by the K-means clustering algorithm, which is the iterative approach. What's more, allowing a change in the coefficient values while updating the dictionary column vectors will accelerate convergence, since the subsequent column vectors update which is based on more relevant coefficients.

3.1. Dictionary Learning

Given the data dictionary $D=\left[d_1,d_2,\cdots,d_J\right]\in R^{M\times J}$, where $J<I$, in this way, the minimization corresponding to (3) is that of searching the sparse representation and the best dictionary:

$$(\hat{D},\hat{\theta}) = \arg\min\left\|Y-D\theta\right\|_2 \quad \text{s.t.}\ \forall i \ \ \left\|\theta_i\right\|_0 \le T_0 \tag{4}$$

where $\left\|\bullet\right\|_0$ is l^0 norm simply counts the number of nonzero elements. T_0 is the upper limitation of the number of non-zero elements in the coefficient, which is also the greatest difference in the degree of the coefficient vector.

We solve the (4) iteratively, specific algorithm is as follows:

① Fix D the with $\phi\varphi$ in (1) to initialize the dictionary;

② Find the sparsest sparse coefficient θ using an approximation pursuit method(BP,MP,OMP). Once the θ is solved, we search for a better D.

3.2. Dictionary Updating

In reality, the background will change because of the light, angle of the camera, which can lead to fatal mistake at the time of detecting the moving target if we don't automatically update the background models. It is assumed that both D and θ are fixed, and we only care the jth column of D which corresponds to the kth row of θ. (4) can be rewritten as:

$$(\hat{D},\hat{\theta}) = \left\|Y-D\theta\right\|_2 = \left\|Y-\sum_{p=1}^{J} d_p\theta^k\right\|_2 = \left\|\left(Y-\sum_{p\neq j} d_p\theta^k\right)-d_j\theta^k\right\|_2 = \left\|E_j-d_j\theta^k\right\|_2 \tag{5}$$

Where E_j stands for the error for all the J examples when the kth element is removed. Here, It would be tempting to use SVD to find alternative d_j and θ^k directly, however, which step will cause a mistake because of the new vector θ^k is likely to be filled. We only save the nonzero elements in θ^k, then it will be solved by SVD. $E_j^{'}=U\Delta V^T$,where $E_j^{'}$ is the error for all the J examples when the nonzero atoms of kth element is removed. We define first column of U as d_j,and the first column of V multiplied $\Delta(1,1)$ as θ^k. The two kinds of the end of the loop iterations, one is to set a number of iterations, this article is set to 20; another is to set the error value of the reconstructed image and original image.

In K-SVD algorithm, the video stream can be approximately represented by linear combination of small enough (less than T_0) atoms. Each sparse coding step (BP, MP, OMP), reduce the total error $\left\|Y-D\theta\right\|_2$ posed in (4). Moreover, in the iterative process to ensure the error is decreased or unchanged which can guarantee that the total error decreases monotonically, which is to ensure the convergence of the K-SVD algorithm.

4. Moving-object Detection and Segment

Read into a video stream X , the measurement matrix and wavelet sparse basis are ϕ and φ respectively. We will get measurement Y via CS that can complete the image compression to reduce data redundancy and bandwidth requirements. Assign values $\phi\varphi$ to the data dictionary to complete the initialization of the dictionary. Through the OMP algorithm for continuous learning dictionary, get the best dictionary $\hat{D}$, and the most sparse coefficient $\hat{\theta}$, strengthening the sparse representation of the dictionary capabilities. Read into the test image, using CS K-SVD algorithm to update the dictionary after CS. The measured values compared with the data dictionary refactoring, to determine whether the target invade, thus selective reconstruction. The final problem can be transformed into a classification problem by setting a threshold of the two types of data individually. The whole process is shown in the figure 1.

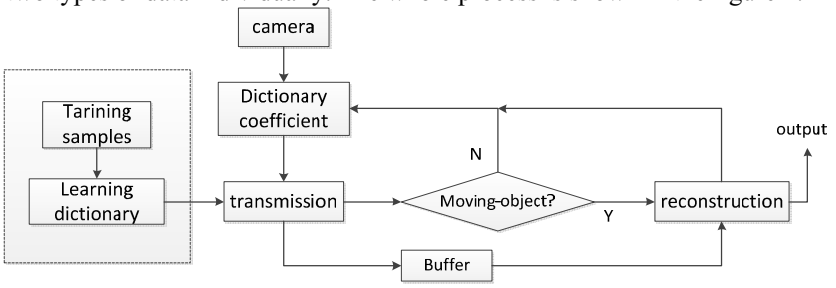


Fig.1. block diagram of the proposed method

Compare the measurements of the test image and data dictionary to determine whether the target invasive, leading to selective reconstruction. This problem can be regarded as a classification problem, by setting a threshold of these two types of data separately.

Segment the moving-object regions while reconstruct the measurements in this paper. Video streams are formed with background region matrix, foreground region matrix and error matrix. As in paper[11] mentioned the principal component pursuit analysis, background is low rank, foreground is the sparse component. Given the measurement matrix $\hat{Y}$ reconstructed from dictionary, the video volume X can be reconstructed by following minimization problem:

$$\hat{X} = X_{\mathrm{b}} + X_{\mathrm{f}} + X_{\mathrm{e}} \tag{6}$$

$$(X_{\mathrm{b}}, X_{\mathrm{f}}) = \arg\min \left\|X_{\mathrm{b}}\right\|_{*} + \lambda \left\|X_{\mathrm{f}}\right\|_{2} \quad \text{s.t.} \quad \left\|\hat{X} - X_{\mathrm{b}} - X_{\mathrm{f}}\right\|_{2} \le \varepsilon, \hat{X} = \varphi\hat{\theta} \tag{7}$$

In (7), $\left\|X_{\mathrm{b}}\right\|_{*}$ is the nuclear norm of the matrix X_{b}, which is defined as:

$$\left\|X_{\mathrm{b}}\right\|_{*} = \sum_{\mathrm{i}=1}^{\min(\mathrm{N},\mathrm{I})} \left|\sigma_{\mathrm{i}}\right| \tag{8}$$

Where σ_i are the singular values of matrix X_{b}. The nuclear norm of the matrix X_{b} is the l^1 norm of its singular values.

In (6), X_{b}, X_{f} and X_{e} represent three different components of the reconstructed video volume. The low rank component is a relatively stationary component, which represents the background region of the video

volume. The foreground region is enough smaller compared to the background region, which can be treated as sparse matrix. What we propose above satisfy the condition in paper [11], which can be solved by (7).

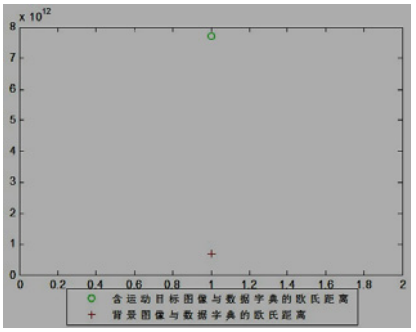
5. Experiment and analysis

We form 100 frames into the matrix to construct the background model, and each frame is written as a column vector(streaming video at 25 frames per second, size of image is 512*512). So training sample size is 262144*100 and CS sampling rate is 20 %. Then initially compressed measurement matrix is 52428*100. The initial data dictionary is represented by the measurement, training iterations is 30, sampling column number 25. The data dictionary matrix decompose as 52428*25and25*50 matrix which is smaller than the original video stream. Figure 2 shows the image constructed via the data dictionary.

Fig.2. (a)original image; (b)CS reconstruction; (c)dictionary reconstruction

After establishing background model, the video starts to monitor. Read into the images of moving objects, complete the CS's compression and K-SVD's background reconstruction. By comparing the Euclidean distance of data dictionary, determine whether there is a moving target, and then conduct selective reconstruction. As shown in Figure 3(a), column vector is the Euclidean distance between measurements for the reconstruction of the input image and background models. The circle stands for images include the moving target, and plus sign for the non-moving target images. Thus, we can see that by setting a threshold between two values then determine whether the goal invasion or not. Here we set the threshold value as $4*10^{12}$.

By using the proposed K-SVD algorithm, we can obtain the background model as Figure 2(c), which ruled out suspicious foreground and the movement goal to the background's disturbance. As this method is based on the brightness of the color property, the three RGB components, if use traditional background difference method for processing, it is very sensitive to brightness change. Here, for the time difference between moving target image and background image is two hours, the light intensity of car surface metal reflector changes sharply, reflecting the vehicle location's pixel gray values are mutated on the image. After the traditional method of background difference, stationary cars will be detected, mistaken for moving targets, as is shown in Figure 3(c). Shown in Figure 3(a), the proposed segmentation method is an effective solution to a brightness change, rule out the disturbance of suspicious foreground, and improve the accuracy of target detection.



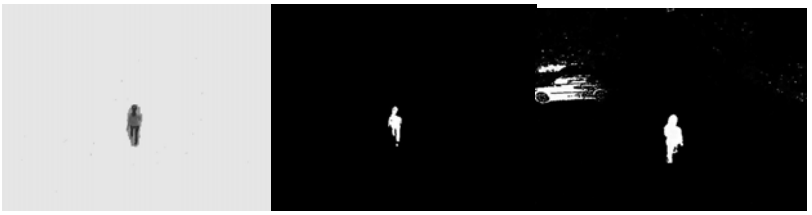


Fig.3. (a)threshold of object detection; (b) object segment based on PCP; (c)binary image; (d) original background subtraction

6. Conclusions

The study proposes an algorithm based on the sparse representation and learned dictionary to detect and segment the moving-object, which can effectively decrease the data transmission through the CS and K-SVD algorithm. Update the dictionary continually through K-SVD to strengthen the sparse representation of the dictionary capabilities. The selectivity of the reconstructed image reduces the amount of calculation achieving practical application of real-time. Reconstruct the moving-object and background individually through PCP method which the image is consisted of low-rank component and the sparse component.

Acknowledgements

This work is supported by the Foundation of postgraduate innovation of Jiangxi Province under Grant No.YC2011- X013.

Reference

[1] ALI S, SHAH M. Human action recognition in videos using kinematic features and multiple instance learning[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2010, 32(2):288-303.
[2] HUI K C, SIU W C. Extended analysis of motion-compensated frame difference for block-based motion prediction error[J]. IEEE Transactions on Imaging Processing, 2007, 16(5):1232-1245.
[3] TSAI D M, LAI S C. Independent component analysis-based background subtraction for indoor surveillance[J]. IEEE Transactions on Imaging Processing, 2009, 18(1):158-167.
[4] Li jie, Li wang zong. Background Subtraction Based on Compressive Sensing[J]. Computer Knowledge and Technology,2010,6(2):410-412.
[5] Volkan Cevher, Aswin Sankaranarayanan,Marco Duarte, et al. Compressive Sensing for Background Subtraction[J]. Computer Vision. 2008, 5303(2008):155-168.
[6] Emmanuel Candes, Michael Wakin. An introduction to compressive sampling[J]. IEEE Signal Processing Magazine,2008,25(2):21-30.
[7] Deng C, Lin W, Lee B, et al. Robust image coding based upon compressive sensing. IEEE Transactions on Multimedia. 2012, 14(2):278-290.
[8] Richard Baraniuk, Compressive sensing[J]. IEEE Signal Processing Magazine,2007, 24(4):118-121.
[9] Tosic I, Frossard P. Dictionary Learning. Signal Processing Magazine[J],IEEE.2011, 28(2):27-38.
[10] Michal Aharon, Michael Elad, and Alfred Bruckstein. K-SVD: An Algorithm for Designing Overcomplete Dictionaries for Sparse Representation[J]. IEEE Transactions On Signal Processing . 2006,54(11):4311-4321.
[11] E. J. Candes, X. Li, Y. Ma ,et al. Robust principal component analysis[J]? Journal of ACM, 2009,58(1):1–37.