ONLINE ADAPTIVE DICTIONARY LEARNING AND WEIGHTED SPARSE CODING FOR ABNORMALITY DETECTION

Sheng Han[1], *Ruiqing Fu*[2],*Suzhen Wang*[3], *Xinyu Wu*[2,3]

[1]Shenzhen Key Lab for Computer Vision and Pattern Recognition,
Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, China
[2]Guangdong Provincial Key Lab of Robotics and Intelligent System,
Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, China
[3]The Chinese University of Hong Kong, Hong Kong

ABSTRACT

This paper focuses mainly on adaptive dictionary updating and abnormality detection via weighted space coding in video surveillance. Generally, abnormality analysis conducted on a large amount of video data is very complicated, time-consuming and time-variant. However, our dictionary is very efficient at following up on shifted contents in video and abandoning old inactive information in time. The adaptability characteristic also helps reduce the dictionary's size to a small scale, since it only needs to keep recent or active information. We also introduce a simple, but effective, judgement criterion for abnormal detection based on sparse coding over weighted bases. Because of the condensed dictionary and the simplified judgment criterion, our algorithm performs online learning and online detection with a high speed and a high accuracy in various scenes.

Index Terms— Dictionary Learning, Adaptive Learning, Sparse Coding, Abnormality Detection

1. INTRODUCTION

Recently, very large amounts of video data has generated great interests in automatic abnormality analysis of the video contents in surveillance systems. In fact, abnormal events usually appear with suddenly changed features in a few successive frames or with significantly changed patterns in a short time. Following the above clues, we proposed an effective framework to detect such abnormal events in this paper.

In recent years, sparse coding has be widely used in machine learning and pattern recognition[1, 2, 3, 4], because of its efficiency in finding succinct representations of stimuli; dictionary is actually a succinct representation for a specific data set. In this paper, we propose an online and adaptive dictionary learning algorithm based on sparse coding. The advantage of our online adaptive learning is: online learning can efficiently handle a tremendous amount of video data; adaptive learning can always maintain pace with moderately-shifted contents from new observations and, therefore, makes off-line training unnecessary. Adaptability also contributes to a much smaller dictionary and, hence, a much faster learning rate; this is because our dictionary only needs to record active information in recent time, which is enough information to discriminate abnormal events from normal ones. Inspired by adaptive GMM[5], we realize adaptability by placing more emphasis on the new observations. Actually, our dictionary consists of normalized bases with assigned weights indicating the current active levels. In Fig.1, we try to show how the dictionary changes when new events continue to appear randomly inside the rectangular area. From Fig.1, we can see that the dictionary is gradually attracted to the rectangular area from initial state (a) to state (c).

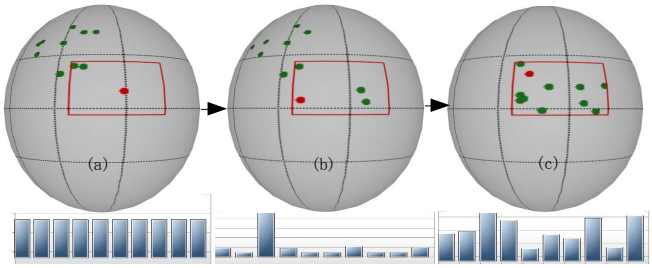

Fig. 1. Over the unit ball, the blue dots indicate normalized bases in the dictionary, the red dots indicate new samples randomly inside the rectangular. Below is the relative weight histogram of the bases at each state

In Fig.1 (b), we can see that only one weight has an overwhelming value after the red point appears several times. In our algorithm, however, the appearance of several dominated weights is not occasional when abnormal events occur;

The work described in this paper is partially supported by the National Natural Science Foundation of China (61005012), Shenzhen Fundamental Research Program (JC201105190948A), Guangdong Innovative Research Team Program (201001D0104648280).

based on this phenomenon, we obtained our effective judgment criterion shown in section 3.3. In section4, we apply our algorithm into various videos from the UMN dataset to the subway dataset, and the experiment results fully validate the efficiency, accuracy and flexibility of our approach.

2. RELATED WORK

Previous approaches to abnormal detection roughly fall into two classes: object-tracking based [6], and spatio-temporal analysis based [7, 8, 9, 9]. Recently, methods based on dictionary learning and sparse coding have appeared for effective automatic video analysis and can be easily applied into various scenarios. In [2, 10], J. Mairal et al managed to work out an online dictionary learning algorithm. Based on J. Mairal's work, Bin Zhao et al [11] proposed an effective, unusual event detection algorithm. However, J. Mairal's dictionary is essentially non-adaptive since it will eventually converge to a stable state; that is also the case with B. Zhao's algorithm. In [12], propose a new criteria(SRC) for abnormal detection is proposed, but their training is completely off-line. Compared to the above methods, our dictionary completely and efficiently realized online learning, adaptive learning and online detection without any off-line training. Additionally, our algorithm performs much faster than theirs while still maintaining a high level of accuracy.

3. ALGORITHM

3.1. OVERVIEW

To transform the raw data into feature vectors, we simply adopts a feature descriptor called Multi-scale Histogram of Optical Field (MHOF) [12], which not only describes motion direction but also preserves more motion-energy information. In the following two subsections, we present our learning algorithm and the entire framework for abnormality detection.

3.2. DICTIONARY LEARNING AND UPDATING

Our dictionary is indeed a matrix with n columns and m rows (m is the dimension of feature vector). Each column can be considered as a base assigned with a weight $w_{ii}, i = 1, ..., n$, the weights should satisfy $\sum_{i=1}^{n} w_{ii} = n$. Dictionary updating only relies on the current dictionary A_t, current weight matrix $W_t = diag(w_{11}, w_{22}, ..., w_{nn})$, the new sample y_{t+1} ($\|y_{t+1}\|_2 = 1$). To emphasize the new samples, we will assign each with a fixed initial weight ω^* ($\omega^* \geq 1$) and, meanwhile, lower the weights in matrix W_t by multiplying each weight in W_t with λ where $\lambda = 1 - \frac{\omega^*}{n}$. Then theoretically we can get the updated dictionary A_{t+1} through solving the following optimization problem:

$$\begin{aligned}(A_{t+1}, X) = \underset{\substack{A_{t+1}\in R^{m\times n}\\ X\in R^{n\times(n+1)}}}{\arg\min} \|A' * W'_t - A_{t+1} * X\|_F \\ s.t.\ \|A_{t+1,i}\|_2 = 1, \quad \|X_j\|_1 \leq 1, \\ i = 1, 2, ..., n; \quad j = 1, 2, ..., n+1\end{aligned} \tag{1}$$

Where the matrix $A' = [A_t, y_{t+1}]$, $W'_t = \begin{bmatrix} \lambda W_t & 0 \\ 0 & \omega^* \end{bmatrix}$ with $trace(W'_t) = n$, $A_{t+1,i}$ indicates the i-th column of A_{t+1}; X_j, the j-th column of X.

The main idea of the above optimization problems is: for each new sample y_{t+1}, we will re-seek a new set of bases A_{t+1} so that the new set is still the most succinct representation for the new sample y_{t+1} and all old bases in A_t taking their weights W'_t into consideration.

In fact, problem (1) is a non-convex optimization problem and its solution is not unique; however, if we fix either A or X, the problem then becomes convex. Therefore, we can obtain a local optimal solution by alternately minimizing one parameter(A or X) while fixing the other(X or A). Unfortunately, this kind of alternating iteration may be very time-consuming. Luckily, through the investigation and analysis conducted in this paper[2], it is possible for us to find a warm start that will obtain an empirically adequate solution after only a single iteration for problem (1).

To obtain the warm start of A, X, we observe that the original optimization problem (1) can be simplified into two parts: problem (2) and problem (3), both are much easier to solve; $X_{[1:n]}$ indicates the first n columns in X and x is the $(n+1)$-th vector in X. In problem (2), the optimization is obtained when $A_{t+1} = A_t, X_{[1:n]} = diag(min(\lambda W_{t,11}, 1), .., min(\lambda W_{t,nn}, 1))$. In problem (3), given $A_{t+1} = A_t$, we can easily obtain the optimal solution $\hat{x}_{t+1}$ by LARS[3]. So taking $A_{t+1} = A_t, X_{[1:n]} = diag(min(\lambda W_{t,11}, 1), ..., min(\lambda W_{t,nn}, 1)), X_{n+1} = \hat{x}_{t+1}$ as the warm start, then we apply the single iteration in **Algortithm1** to obtain an approximate solution.

$$\begin{aligned}(A_{t+1}, X_{[1:n]}) = \underset{\substack{A_{t+1}\in R^{m\times n}\\ X\in R^{n\times(n+1)}}}{\arg\min} \|A_t * (\lambda W_t) - A_{t+1} * X_{[1:n]}\|_F \\ s.t.\ \|A_{t+1,i}\|_2 = 1; \quad \|X_j\|_1 \leq 1\end{aligned} \tag{2}$$

$$\begin{aligned}\hat{x}_{t+1} = \underset{x}{\arg\min} \quad \|\omega^* y_{t+1} - A_{t+1} x\|_2 \\ s.t.\ \|x\|_1 \leq 1; \quad A_{t+1} = A_t\end{aligned} \tag{3}$$

If some weights in weight matrix W_t become very small and below a pre-determined threshold, meaning that the corresponding base becomes inactive, then we will directly replace the least-weight base (i.e., i-th column) with the new sample by (6), and skip **Algortithm1**.

$$\begin{aligned}A_{t+1,i} &= y_{t+1} \\ W_{t+1,jj} &= \frac{W_{t,jj} * (n - \omega^*)}{n - W_{t,ii}} \quad j \neq i \\ W_{t+1,ii} &= \omega^*\end{aligned} \tag{6}$$

Algorithm 1 Online Adaptive Dictionary Updating with Weights

Require: $A_t, W_t, y_{t+1}, \hat{x}_{t+1}, \lambda, \omega^*$, where $\lambda = 1 - \frac{\omega^*}{n}$
Ensure: A_{t+1}, W_{t+1}
Step1:

$$\begin{aligned} B &= diag(B'_{11}, ..., B'_{nn}) + \hat{x}_{t+1}\hat{x}_{t+1}^T \\ L &= (L'_1, .., L'_n) + y_{t+1}\hat{x}_{t+1}^T \end{aligned} \tag{4}$$

where $B'_{ii} = min(\lambda^2 W^2_{t,ii}, 1)$ and $L'_i = min(\lambda W_{t,ii}, 1)$
Step2:
for $i = 1$ to n **do**

$$\begin{aligned} A'_{t+1,i} &= \frac{L_i - A_t * B_i}{B_{ii}} + A_{t,i} \\ A_{t+1,i} &= \frac{A'_{t+1,i}}{\|A'_{t+1,i}\|_2} \\ W_{t+1} &= diag(W_{t+1,11}, .., W_{t+1,nn}) \end{aligned} \tag{5}$$

where $W_{t+1,ii} = \lambda W_{t,ii} + \frac{|\hat{x}_{t+1,i}| * \omega^*}{\|\hat{x}_{t+1,i}\|_1}$
end for
return A_{t+1}, W_{t+1}

Generally, feature vectors, extracted from continuously appearing normal events, are highly related to each other and are very likely to fall into a low rank subspace; this phenomena also guarantees the feasibility of a succinct dictionary with a small scale. Bases in current dictionary and their corresponding weights essentially form a density distribution of the normalized feature vector, and each new sample helps update the density distribution to adapt to dynamic scenario.

3.3. JUDGEMENT CRITERIA AND FRAMEWORK OF ABNORMALITY DETECTION

In this section, we come to discuss about the judgement criteria shown in (7) for abnormal event detection in formula and then present the entire framework of the detection algorithm.

$$J(y_{t+1}) = \|W_t * \hat{x}_{t+1}\|_1 \tag{7}$$

What we aim to detect are events with significantly changed features which may lie far away from the feature space spanned by current dictionary (refer to Fig.1). At each dictionary updating, the abnormal feature will asymptotically affect the closest base and simultaneously increase its weight, which in turn, makes this base much closer to the abnormal feature. So during the subsequent frames, abnormal features will mainly act on the same bases, resulting in abruptly increasing weights of these bases. Based on this, we obtained our abnormality judgement criteria $J(y)$ as shown in (7), which will select all these dominant weights and sum them up to show its rapidly-increasing trend. It is important to note that when some weight falls below a threshold, the corresponding base will be replaced by the new sample and such an operation will help recovery the weight distribution back to balanced state later.

Algorithm 2 The Framework for Abnormal Detection

Require: The first n video series$\{a_1, ..., a_n\}$, $A_t, W_t, y_{t+1}, \omega^*$, threshold δ.
Ensure: A_{t+1}, W_{t+1}
Step0: Initialization $t = 0$, $A_t = [a_1, ..., a_n]$,
for $i = 1$ to n
$W_{0,ii} = 1$
endfor
Step1: use LARS to compute:

$$\begin{aligned} \hat{x}_{t+1} = & argmin \quad \|\omega^* y_{t+1} - A_t x\|_2 \\ & s.t. \quad \|x\|_1 \leq 1 \end{aligned} \tag{8}$$

Step2: **if** $J(y_{t+1}) > \delta_1$ (**ALARM**)
Go to **Algorithm1** or formula (6) to update A_{t+1}, W_{t+1}
Step3: Let $t = t + 1$, go to Step1

4. EXPERIMENTS

To demonstrate the performance of our algorithm, we apply it into both global abnormal events(GAE) detection and local abnormal events(LAE) detection with videos from two typical data sets: the UMN dataset and the subway dataset [13] respectively. We will adopt type A-MHOF ([12]) feature descriptor for GAE and type B-MHOF for LAE.

4.1. GAE FOR UMN DATASET

We test the algorithm's GAE performance in three different scenes from the UMN dataset. Each scene (Scene1-Scene3) consists of 1075, 725 and 2207 frames respectively with resolution of 320×240. We first transform the RGB frames into gray style and split each frame into 3×4 blocks, and then we extract the feature from each block and concatenate them into a final feature vector with a dimension $m = 192$. We only take the first 30, 60 and 90 frames from each scene to initialize each dictionary with a size of 30, 60 and 90 columns respectively. Then we begin learning and detection immediately without off-line training. The results are shown in Fig.2, from the charts below, we can read easily about the occurrence of escape events from $J(y)$ curve, and the algorithm's processing rate for each scene is 0.05, 0.08, 0.11 seconds/frame.

4.2. LAE FOR SUBWAY DATASET

We test the algorithm's LAE performance in two subway videos: "entrance" with 144249 frames and "exit" with 64900 frames from the subway dataset, which was first provided by Adam[13]. In our experiments, we choose four ROIs in both videos, and then we use our dictionary updating algorithm twice to construct two dictionaries. The first dictionary is

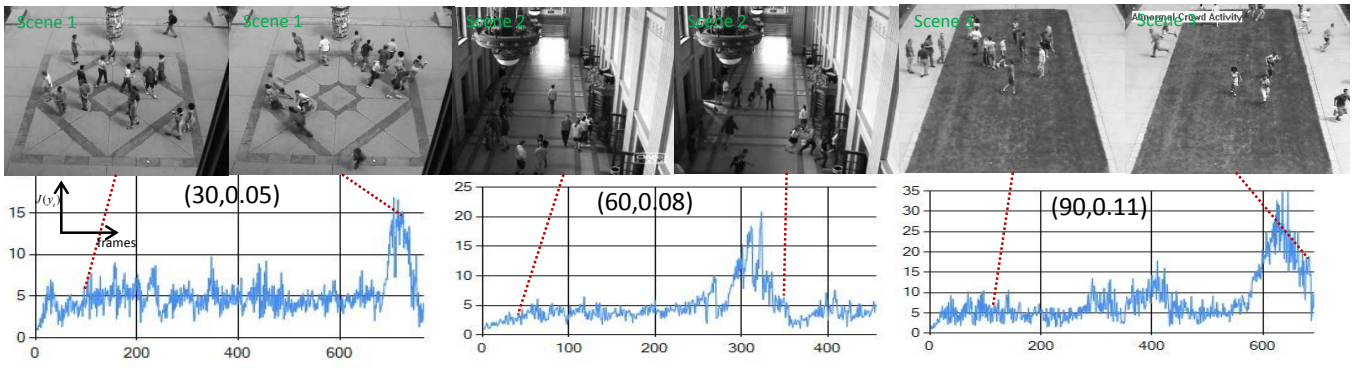

Fig. 2. The results of escape events in three scenes from the UMN dataset. The first pic of each scene indicates its normal pattern; the second pic indicates the abnormal pattern. The curve of $J(y)$ will increase abruptly when escape event appears, and drop back to low value when the abnormal event disappears. the numbers in the bracket indicate the size of the dictionary and the processing speed (sec/frame).

Fig. 3. The results of the LAE detection for Subway dataset. The top row are from exit video; the bottom row entrance video. The the first three columns are frames with irregular behaviour being correctly detected; the last column is the frame with false alarm

used to detect all the people passing by in each ROI; the second dictionary is used to detect people with irregular behaviour among all the detected people.

The dimension of feature vector for each ROI is $m = 16 \times 5 = 80$ and both dictionaries have a size of 100 columns. We can see the results from Fig.3 and the comparison data between our approach and the state-of-art methods from Table1.

From Table1, we can see our algorithm have a high processing rate and can detect wrong directions with a high accuracy. However, as to No-Pay detection, its performance becomes worse since there are no radical feature changes during the No-Pay events.

All the experiments are run on a computer with 2GB RAM and 2.6GHz CPU dual core, and the software we use is C#. The average online computation time is 0.08s/frame for GAE and 0.25s/frame for LAE. Such speeds are much faster than the state-of-the-art methods based on sparse coding.

	WD	NP	FA	Speed(s/f)
Ground Truth	26/9	13/-	0/0	–
Yang Cong[12]	21/9	6/-	4/0	4.6
Bin Han[11]	25/9	9/-	5/2	2
Ours	22/9	2/-	9/5	0.25

Table 1. Comparisons between different algorithms based on sparse coding for subway detection. The number before (/)denotes the data from "entrance"; the number behind is the data from "exit". WD means Wrong Direction; NP No-Pay; FA False Alarm; (s/f) seconds/frame

5. CONCLUSION

In this paper we propose an effective abnormal detection framework which aims to follow up on shifted scenario over time and simultaneously maintain competence in discriminating abnormal events based on recently learned knowledge. Our algorithm has very good flexibility: it can be applied directly into both global and local event detection, and into both crowded and uncrowded scenarios. The experiment results fully validate the efficiency, accuracy and flexibility of our algorithm in various scenes even though its learning and detection is completely online.

6. REFERENCES

[1] John Wright, Allen Y.Yang, Arivind Ganesh, S.Shankar Sastry, and Yi Ma, "Robust face recognition via sparse representation," *TPAMI*, vol. 31, no. 2, pp. 210–227, 2009.

[2] Julien Mairal, Francis Bach, Jean Ponce, and Guillermo Sapiro, "Online dictionary learning for sparse cod-

ing," *International Conference on Machine Learning*, pp. 689–696, 2009.

[3] Bradley Efron, Trevor Hastie, Iain Johnstone, and Robert Tibshirani, "Least angle regression," *Annals of Statistics*, vol. 32, no. 2, pp. 407–499, 2004.

[4] Robert Tibshirani, "Regression shrinkage and selection via the lasso," *Journal of the Royal Statistical Society*, vol. 58, no. 1, pp. 267–288, 1996.

[5] Chris Stauffer and W.E.L Grimson, "Adaptive background mixture models for real-time tracking," *CVPR*, 1999.

[6] Weiming Hu; Tieniu Tan; Liang Wang; S. Maybank, "A survey on visual surveillance of object motion and behaviors," *IEEE Systems, Man, and Cybernetics Society*, pp. 334–352, 2004.

[7] vijay Mahadevan, Weixin Li, Viral Bhalodia, and Nuno Vasconcelos, "Anomaly detection in crowded scenes," *CVPR*, pp. 1975–1981, 2010.

[8] Louis Kratz and Ko Nishino, "Learning object motion patterns for anomaly detection and improved object detection," *CVPR*, pp. 1–8, 2008.

[9] Hannah M.Dee and Alice Caplier, "Anomalous video event detection using spatio-temporal context," *ICIP*, pp. 1545–1548, 2010.

[10] Julien Mairal, Francis Bach, Jean Ponce, and Guillermo Sapiro, "Online learning for matrix factorization and sparse coding," *Journal of Machine Learning Research*, vol. 11, no. 3, pp. 19–60, 2010.

[11] Bin Zhao, Li Fei-Fei, and Eric P. Xing, "Online detection of unusual events in videos via dynamic sparse coding," *CVPR*, pp. 3313–3320, 2011.

[12] Yang Cong, Junsong Yuan, and Ji Liu, "Sparse reconstruction cost for abnormal event detection," *CVPR*, pp. 3449–3456, 2011.

[13] Amit Adam, Ehud Rivlin, Ilan Shimshoni, and David Reinitz, "Robust real-time unusual event detection using multiple fixed location monitors.," *TPAMI*, pp. 555–560, 2008.