

A surveillance video compression algorithm based on regional dictionary

Suyu Wang^{1,2}, Yujie Chen^{1,2} and Yuheng Bai^{1,2}

¹Beijing Engineering Research Center for IoT Software and Systems, 100124 Beijing, China

²School of Software Engineering, Beijing University of Technology, 100124 Beijing, China

Abstract. Surveillance video system as an effective means of public safety widely appeared in people's life in recent years. Traditional video coding technology, the encoding process is complex, computation, and conflicts with hardware limitations of the monitoring system. This paper presents a kind method used to design and train redundant dictionary based on image region. Each frame is divided into same-sized area with equal blocks, then the image areas of the same position are put together as the training sample library of regional redundant dictionary. In this way, redundant dictionary compression results in the occupied volume can be reduce. On top of this dictionary, surveillance video compression coding algorithm based on the regional dictionary is designed and implemented. Experimental results show that the proposed algorithm can effectively improve the compression ratio of the algorithm.

With the need for Social and public security and big data analysis, video surveillance system is increasingly applied to people's life^[1]. Surveillance systems made of various cameras are lying over all positions of the society and provide services for people's safe life. According to IDC (Data Corporation International) research report estimates, the total amount of global data will reach 40ZB in 2020, among which the data of surveillance video is 5.8ZB. And China will account for 21%, which means that in 2020, China will have 1.2ZB (One billion and two hundred million TB) surveillance video data needed to be stored, transmitted and analyzed. Therefore, facing with massive surveillance video, considering the transmission and storage costs, we need to carry out new research and achieve a major breakthrough in surveillance video encoding and analysis technology, it is very necessary to research higher compression efficiency of surveillance video encoding technologies. This paper proposes a method for studying of balancing the redundant dictionary size and the overall algorithm of encoding effect. The way to reduce the volume of redundant dictionary can be found out by analysing the compression result. The redundancy of each image block in the way of training the dictionary itself can be reduced by improving the original training method. A redundant dictionary trained by a region which is as a simple formed by multiple unit image blocks is applied to all of the image block in the area. This method use of redundant information can reduce the number of redundant dictionary, improve the training efficiency further, and do not cause significant losses for the compression algorithm of encoding efficiency. This paper also proposes a algorithm that combined with surveillance video

compression and coding algorithm based on key frame to build the public background of surveillance and use the redundant dictionary to sparse decomposition the background image block or the image block with the approximate background, then quantize and compress correspondingly, while the foreground image block is reserved separately. Decoding the foreground background combined decoding, so as to improve the quality of encoding video in the premise of a small amount of loss compression ratio.

1 Algorithms' outline

The surveillance video compression coding algorithm based on the regional dictionary which is similar to that based on sub block dictionary, and it is divided into two stages: the training of the regional dictionary and the video compression and coding. The compression results of the surveillance video compression coding algorithm are generally made up of encoding's results of surveillance video frames and redundant dictionaries. The training process of the regional dictionary is the same as that of the sub block mostly, but there are also differences in details. Firstly, each frame of the video captured from the whole surveillance video was divided into a number of image regions with the same size in the same way, each of the image regions has the same shape and the same number of image blocks. After that, the same region in the picture are collected together, as the training sample library of the region's redundant dictionary. The image block took out from the training sample library randomly is trained redundant dictionary using K-SVD algorithm to get redundant dictionary of the

corresponding area. Then, uniform quantization, subtraction Background, JPEG-LS encoding, after those get the encoding results.

2 Sparse representation theory based on redundant dictionary

The sparse representation theory of signal points out that [2], when the signal is projected to a certain transformation base, the obtained transform coefficients are mostly zero or very small absolute value, that is, the transform coefficient vector is sparse or approximate sparse, then we can transform the transformation vector as a simple expression of the original signal in the transform domain. To obtain a sparse representation of the signal, it is usually necessary to select the appropriate transformation basis for the characteristics of the signal. At this point, the transformation base is also called sparse group, the coefficient of sparse decomposition and sparse matrix can be combined to reconstruct the original signal [3].

In the model of signals sparse representation, a linear weighted combination of a few atoms in a redundant dictionary can be used to represent the signal [4]. Sparse decomposition is the process of decomposing signals based on the redundant dictionary and getting the sparse representation coefficients. It means that redundant dictionary is very important in sparse representation, and it will affect the effect of sparse representation and the efficiency of sparse decomposition, whether the redundant dictionary is appropriate.

The construction of redundant dictionary based on sparse representation has two ways [5]: (1) the method based on generating functions, such as Contour lets transformation; (2) the method based on sample training, such as the Optimum-Direction method [6], GPCA [7], K-SVD algorithm (K-singular value decomposition) [8]. K-SVD is the algorithm for redundant dictionary training in this paper.

The basic idea of the algorithm is based on the given training sample, and the dictionary is updated by searching the best representation of the training sample under the strict sparsity degree. To sum up, the K-SVD algorithm can be said to be the iteration of two steps:

- Sparse decomposition procedure. Samples are sparsely decomposed with any tracking algorithms.
- The step of dictionary updating. If the coefficient matrix of the sparse decomposition is given, the singular value decomposition method is used to decompose the residual matrix of the dictionary, and update the corresponding atom in the dictionary and the corresponding sparse decomposition coefficient simultaneously. Then back to the sparse decomposition step, multiple iterations are completed by the loop.

The best dictionary of the sparse representation of data samples $\{y_i\}_{i=1}^N$ will be found through solving the formula. The formula is as follows:

$$\min_{D,X} \{ \|Y - DX\|_F^2 \} \quad \text{subject to} \quad \forall i, \|x_i\|_0 \leq T_0 \quad (1)$$

3 The surveillance video compression coding algorithm based on the regional dictionary

3.1 Description of training process of regional dictionary

In this paper, the compression result of the surveillance video compression coding algorithm is made up of encoding's results and the redundant dictionary. The redundant dictionary is trained for the surveillance video encoded specially. It is not adaptive to other surveillance video or signal, and the algorithm must use the redundant dictionary to reconstruct the original surveillance video, so redundant dictionary also needs to be counted in the compression results. The atoms of redundant dictionary are dense and the redundant dictionary, the key to the decomposition and reconstruction of the video in the process of sparse representation, should not be compressed. So in the compression results, it has a large proportion that reduces the overall compression ratio. If the volume of redundant dictionary can be controlled, better compression effect should be got, but at the same time, we should pay attention to the reconstructed quality of the surveillance video.

Although the volume of redundant dictionary is fixed. With the increase of the length of encoded surveillance video, the proportion of redundant dictionary in the compression results will fall, because the volume of each frame encoded residual is too small, the reduction is slow and slow. Therefore, it is very important to reduce the overall size of the redundant dictionary, which is the key to improve the compression ratio of the algorithm. In this paper, each frame of the video intercepted from the whole surveillance video is segmented in the same way. Each frame was divided into a number of regions with same size, and each of the image regions has the same shape and the same number of image blocks. After that, the same region in the picture are collected together as the training sample library of the region's redundant dictionary. The image block took out from the training sample library randomly is trained redundant dictionary using K-SVD algorithm to get redundant dictionary of the corresponding area. Since each region's dictionary training is independent of each other without no overlap in the image block, it can be executed in parallel to speed up the training efficiency.

3.2 Compressed encoding

In the video compression algorithm based on the regional dictionary, during the process of sparse decomposition, every frame of the surveillance video is read in order, and the image block is processed one by one. Firstly, each image block is decided which region it belongs to, and then according to corresponding region dictionary the coefficients of sparse decomposition of the image block are obtained, finally the coefficients of sparse decomposition of the video are obtained. After the uniform quantization, differential background [9], JPEG-LS encoding, encoding results can be got.

In the process of sparse decomposition, each image block of each frame is decomposed by the corresponding position regional dictionary trained by the K-SVD algorithm, the coefficients of all the image blocks are the coefficients of this frame. The coefficients are between 1 and 0, and there are a large number of decimal numbers. Therefore, the coefficients are quantized to 0-255 with details remained utmost, which is convenient for the following processing and preservation. As the background of each frame is similar, the sparse decomposition coefficients of the background are also similar. In order to remove the redundant information, the first step is establishing a public background of the surveillance video, the public background is also sparsely decomposed by the same set of redundant dictionary, the coefficients obtained are quantified uniformly, while the other frames are different from the coefficients of public background, and retain the residual error for the final entropy encoding only. In front, the sparse decomposition and uniform quantization have caused loss to the video quality. In order to avoid losing more video information, the last of the encoding uses JPEG_LS which is Near-lossless Coding. Complete compression encoding process is as follows:

- Sparse decomposition: Each frame of the public background and the surveillance video is decomposed, and the image block of the video frame is decomposed by the regional dictionary of the corresponding region and then the sparse decomposition coefficients of each frame are obtained;
- The final result S_i of each frame is obtained through uniform quantization, differential background and JPEG-LS encoding.

4 Experimental result

4.1 The description of experimental material

The material used in this section is the 432nd to 562nd frames of S5-T1-G, a group of PETS2006 video sequences. In the 130 frames, the foreground has a large change, pure background without the foreground can be obtained when public background is established. In order to have a reference, the algorithm will be compared to the MJPEG Motion JPEG (Motion, Joint, Photographic, Experts, Group). MJPEG is a video compression format without considering the difference of the different frames of the video basically, and it uses JPEG encoding to compress each frame image independently. MJPEG is the main product of video surveillance equipment manufacturers, which has a large market, and its characteristics are similar to the algorithm in this paper.

On the quality evaluation of the recovered after compression, the paper uses two methods. one is comparing the objective value with the test sample, the other one is the subjective evaluation. The common objective evaluation criteria are MSE, SNR, PSNR and so on currently, which can be transformed with the formula. This paper chooses PSNR (the peak signal to noise ratio) as the standard. Supposing the original image is

$f(x, y) \in R^{m \times n}$, the restored image is $\hat{f}(x, y)$, then the PSNR is as follows:

$$PSNR = 10 \log_{10} \left\{ \frac{255^2 mn}{\sum_{x=0}^{m-1} \sum_{y=0}^{n-1} (f(x, y) - \hat{f}(x, y))^2} \right\} \quad (2)$$

4.2 Analysis of experimental results

This paper selects 432nd to 562nd frames. In the 130 frames, the foreground has a large change, pure background without the foreground can be obtained when public background is established. While the unit image block is selected as $8 * 8$, then the region composed by $11 * 4$ image blocks is trained redundancy dictionary based on the region. As a result, the number of redundant dictionaries is greatly reduced, from 6336 to 144. During the sparse decomposition, first the image block is judge which region it belongs to, and then it is sparsely decomposed by the corresponding region dictionary, as well as the reconstruction.

Table 1. Comparison between surveillance video compression coding algorithm based on the regional dictionary and the algorithm in the video compression algorithm MJPEG.

algorithm	compression ratio	PSNR(dB)
Motion JPEG algorithm	15.0207	31.8202
	9.7771	34.9453
	5.9689	36.6654
Algorithm based on the regional dictionary	20.2682	31.6745
	10.4001	34.6090
	5.4762	36.7784

The compression ratio and PSNR of several algorithms is compared showed in Table 1. The algorithm based on regional dictionary is compared with MJPEG algorithm, when the reconstruction quality is low, namely the PSNR value is low, the compression ratio is better than MJPEG algorithm; but when PSNR is higher, the compression ratio is similar with MJPEG algorithm.

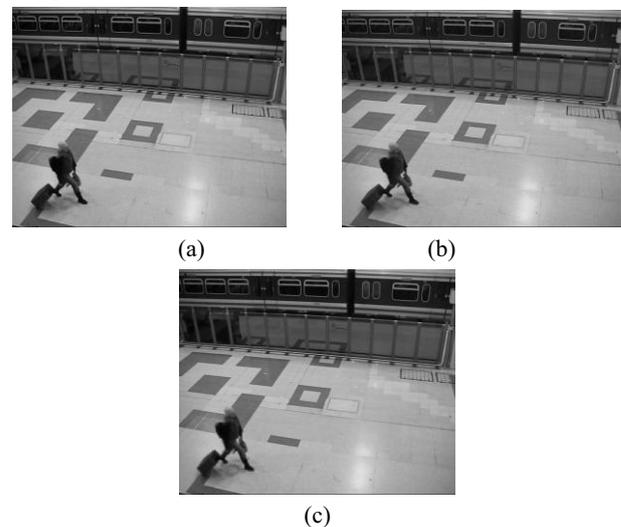


Figure 1. Overall subjective contrast.

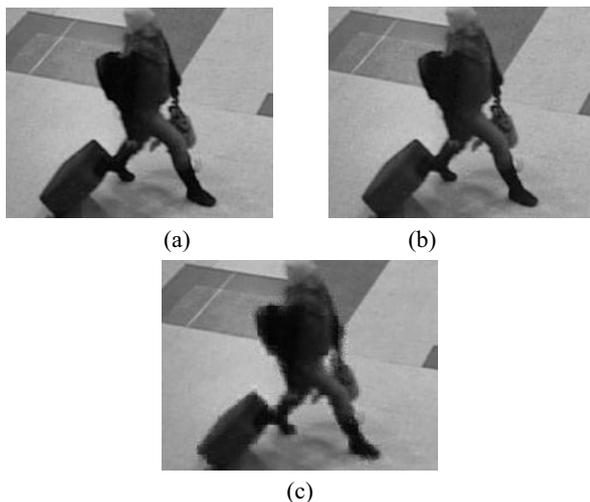


Figure 2. Local subjective contrast(1).

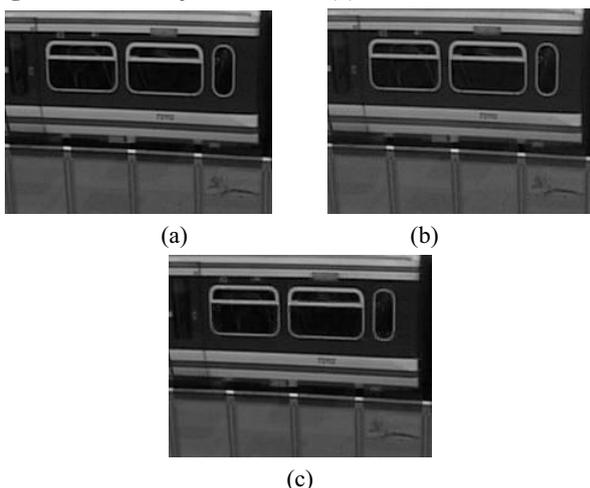


Figure 3. Local subjective contrast(2).

Figure 1 is the whole subjective comparison of a frame, figure a is the original label of the 477th frame material, figure b is the MJPEG encoding map of the frame, figure c is the reconstruction based on the regional dictionary algorithm of the frame. Figure 2 and figure 3 are a subjective comparison of the details. a, b, and c are the same meanings. The regional dictionary's performance is very good in the two algorithm through subjective comparison, the quality of reconstructed image and original is similar, and also no less than MJPEG algorithm. Although each image block is not corresponding to a redundant dictionary, the background area can still be clearly reconstructed. Moreover, with the regional dictionary displacing the block dictionary, reconstruction ability of foreground based on regional dictionary algorithm becomes stronger without separate encoding. Because samples of the regional dictionary samples are more diverse, the atomic structure is also more, the algorithm can find a result which is more similar to the original image.

Seen from the contrast of foreground details of Figure 2, the outline and details of reconstructed foreground based on the algorithm of the regional dictionary are still relatively vague.

5 Summary

The paper proposes a new algorithm for surveillance video, which is based on the regional dictionary. The key of the algorithm is to train a regional dictionary which is more conforms to surveillance video, and it can bring a good sparse decomposition and restoration effect. The experimental results show that the image compression coding algorithm based on the regional dictionary has a good effect on the background reconstruction, and has a general effect on the foreground reconstruction. There are still some shortcomings in this paper, the most obvious one is that, the process of the foreground is relatively simple. In the algorithm, the foreground is only judged by the average error. If it can be judged by the corresponding algorithm, it should be able to improve the quality of the picture further.

Acknowledgments

This work is supported by the National Natural Science Foundation of China (No. 61201361), Science Foundation of the Beijing Education Commission (KM20131 0005028), Training Programme Foundation for the Talents in Beijing City (2013D005015000008), and Training Programme Foundation for the Youth Talents by Beijing Education Commission.

References

1. Zhong Bin. Video surveillance system and intelligent video surveillance technology analysis [J]. electronic technology and software engineering, **23**,187-188(2014)
2. Deng Chengzhi. Image sparse representation theory and its application [D]. Huazhong University of Science and Technology (2008)
3. JIANG Ping. Research on construction method of based on compressed video sensing dictionary [D]. Xidian University (2008)
4. Lian Qiusheng, Shi Baoshun, Chen Anzhen. Research Advances on Dictionary Learning Models ,Algorithms and Applications[J]. Acta Automatica Sinica, **41(2)**:240-260 (2015)
5. Rubinstein, Brucksteina, Eladm. Dictionaries for Sparse Representation Modeling[J]. Proceedings of the IEEE, **98(6)**:1045-1057(2010)
6. KreutzDelgado K, Murray J F, Rao B D, et al. Dictionary learning algorithms for sparse representation[J]. Neural computation, **15(2)**:349-396(2003)
7. Aharon M, Elad M, Bruckstein A. K-SVD: An Algorithm for Designing Overcomplete Dictionaries for Sparse Representation[J]. Signal Processing, IEEE Transactions on, **54(11)**:4311-4322 (2006)
8. Sun Yubao. Image sparse representation model and its application in image processing inverse problem [D]. Nanjing University of Science and Technology (2010)
9. Zhang Xianguo. Research on surveillance video based on background model[D]. Peking University (2013)