An Improved Background Subtraction Method for Adaptive Rate Compressive Sensing

Jianming Wang, Jianhua Chen
School of Information Science and Engineering, Yunnan University, Kunming, 650500, China
Email: chenjh@ynu.edu.cn

Abstract. In Compressive Video Sensing application, the using of Adaptive Rate Compressive Sensing (ARCS) method can predict and adjust the sampling rate for each video frame, reduce the total sampling rate and improve the quality of reconstructed image. In order to use the inter-frame correlation of the video signal efficiently and reduce the sampling rate of the whole signal, an Improved Background Subtraction (IBS) method is proposed in this paper. By alternately using the frame without any foreground objects and the previous frame as the background of the signal, the sparsity of the foreground signal is improved and the total sampling rate is reduced. Experimental results show that, compare with the traditional background subtraction method, the IBS method can significantly reduce the total video sampling rate without apparent degradation of reconstructed image quality under the same ARCS method.

1. Introduction
In the application of Compressive Sensing (CS), the implementation of Adaptive Rate Compressive Sensing (ARCS) [1,2] is one of the hot issues. In the related researches, some fast and practical ARCS methods have been proposed, which makes it possible to apply adaptive rate sampling method in the process of Compressive Video Sensing (CVS). For example, in the methods proposed in reference [3,4,5], the sampling rate can be allocated by observing the sampling results of CS without digital conversion and storage of the original signal, and these methods are all independent of signal reconstruction.

By using an ARCS method, the sampling device can estimate the signal sparsity and allocate the appropriate sampling rate before sampling. Therefore, if the complexity of the target signal can be reduced and the sparsity can be improved, the total sampling rate can be reduced.

In surveillance video, the background subtraction [6] method is often used to generate sparse target signal by subtracting the background frame from the current frame. The background subtraction method has the characteristics of fast calculation speed in practical use, and can be used in many different ARCS methods. However, generally speaking, compared with the background frame, a frame at time \( t \) is always more similar to the image at time \( t-1 \). If the similarity can be used, the sparsity of the target signal can be improved obviously.

In order to make better use of inter-frame correlation and further improve the signal sparsity, a new Improved Background Subtraction (IBS) method is proposed, which uses the frame without any foreground objects and the previous frame alternately as the background of the signal to improve the sparsity of the sampled signal and reduce the overall sampling rate. At the same time, according to the characteristics of ARCS, we design a two-step sensing method, and optimize the implementation of the IBS method according to the characteristics of the two-step method. Finally, by comparing the
performance of Traditional Background Subtraction (TBS) with that of IBS in the same ARCS-FSE [3] method, we show the performance characteristics of IBS.

2. Proposed Method

2.1. Improved background subtraction method

The surveillance video is not a sparse signal itself, so it is necessary to obtain a sparse frame through background subtraction method before CS measurement. At time $t$, for a frame of the video signal $v_t \in \mathbb{R}^{h \times c}$, the size of the frame is $h$ rows and $c$ columns, and $v_t$ can be represented by a vector $x_t \in \mathbb{R}^N$ where $N = h \times c$. The $x_t$ can be decomposed into a time-varying foreground signal $f_t \in \mathbb{R}^N$ and a known background signal $b \in \mathbb{R}^N$. Considering the noise $n_t \in \mathbb{R}^N$ in the sampling process, the video signal at time $t$ can be expressed as:

$$x_t = f_t + b + n_t,$$

(1)

where it is assumed that the noise $n_t$ obeys i.i.d. Gaussian distribution, $n_t \sim N(0, \sigma^2)$.

It can be obtained from the characteristics of the monitoring video that $f_t$ is a description of the object that appears in the monitoring range at time $t$. By subtracting $b$ from $x_t$, we can get the signal $f_N$ which is the sum of $f_t$ and $n_t$:

$$f_N = x_t - b = f_t + n_t.$$

(2)

In the TBS method, a frame that does not contain any foreground signal is generally used as a background frame. In practice, due to the influence of noise, we use $J$ frames of video images that do not contain any foreground objects to approximate the background signal $b$. In these images, we denote the video signal at time $j$ as $x_b$, take the average value:

$$\overline{x_b} = \frac{1}{J} \sum_{j=1}^{J} x_b$$

(3)

and consider $\overline{x_b}$ as the background signal:

$$b = \overline{x_b}.$$

(4)

In monitoring videos, the current frame is usually more similar to the previous frame than the background frame, if let

$$b = x_{t-1},$$

(5)

then it will be able to make better use of the inter-frame correlation of the video signal, and the sparsity of the signal after subtraction is higher, which is beneficial to further reduce the sampling rate.

However, the CS reconstruction method is a lossy method. At the reconstruction end, the original signal $x_{t-1}$ is unknown, only $x_{t-1}$ can be obtained. Due to the error accumulation effect, if we set $b = x_{t-1}$ for all the frames, the reconstruction quality of the frames is going to be worse and worse, which is not acceptable in practice.

An improved background subtraction (IBS) method is proposed to solve this problem. First, we divide all video frames into odd frames and even frames. For odd frames, where $\text{MOD}(t, 2) = 1$, we set $b = \overline{x_b}$. For even frames, where $\text{MOD}(t, 2) = 0$, we set $b = x_{t-1}$. In this way, we can make full use of the inter-frame correlation to reduce the sampling rate while ensuring that the reconstructed image quality does not have apparent deterioration.

2.2. Adaptive Rate Compressive Sensing method

For ARCS, the size of its measurement matrix $\Phi_t \in \mathbb{R}^{M_t \times N}$ is variable depending on the sparsity of the signal. When $\Phi_t$ is determined, $x_t$ can be measured by:

$$y_t = \Phi_t x_t.$$

(6)

Let
\[ \beta = \Phi_t b \]

be the CS measurement result for \( b \). Then \( \beta \) can be subtracted from \( y_t \), we have:
\[
\xi_t = y_t - \beta = \Phi_t(x_t - b) = \Phi_t(f_t + n_t) = \Phi_t f N_t.
\]

Here, the dimension parameter \( M_t \) of \( \Phi_t \) is closely related the sparsity of \( f N_t \).

The ARCS-FSE method [3] can be used to estimate the sparsity of \( f N_t \). Let the index of the pixel in \( f N_t \) be \( i \), and the index set of the large value pixels be \( F_t = \{ i : |f N_t(i)| \geq a \} \). Note that the value of \( f_t \) is 0 when its pixels does not contain any foreground object information, we have:
\[
f N_t(i) = \begin{cases} 
    f_t(i) + n_t, & i \in F_t \\
    n_t, & i \notin F_t
\end{cases}
\]

Assume that pixels are uniformly distributed when \( i \in F_t \), and consider that \( \sigma \) is much smaller than \( \alpha \), we can assume that:
\[
f N_t(i) \sim \begin{cases} 
    U([-b, -a] \cup [a, b]), & i \in F_t \\
    N(0, \sigma^2), & i \notin F_t
\end{cases}
\]

From the work of [7] and [8], for a measurement matrix \( \Phi' \in \mathbb{R}^{P \times N} \), accuracy parameter \( \varepsilon \) and confidence parameter \( \rho \), when
\[
r \geq 8 \varepsilon^{-2} \log(1/2\rho),
\]

it can be sure
\[
(1 - \varepsilon^2) \leq \frac{\|f N_t\|_2^2}{\|\xi_t\|_2^2} \leq (1 + \varepsilon^2)
\]

with probability exceeding \( 1 - \rho \). In practice, by appropriately setting the value of \( r \), it can be roughly considered that:
\[
\|f N_t\|_2^2 \approx \|\xi_t\|_2^2.
\]

When \( |F_t| = k \) and \( |\bar{F}_t| = N - k \), from the work of [3], the energy expectation of \( f N_t \) is:
\[
E(\|f N_t\|_2^2) = k \left( \frac{b^2 + ab + a^2}{3} \right) + (N - k) \sigma^2.
\]

And the estimation of \( k \) can be
\[
k^* = \arg \min_{k=0,1,\ldots,N} E(\|f N_t\|_2^2) - \|\xi_t\|_2^2.
\]

For a threshold \( \tau \), all the points with values larger than it are considered as effective large value points, which are the targets for sensing and reconstruction, while the points smaller than \( \tau \) are not considered as effective large value points, and the value is regarded as 0. Then the estimation of sparsity can be
\[
s_t^* = k^* \cdot Pu + (N - k^*) \cdot Pn,
\]

where
\[
Pu = \begin{cases} 
    \int_{|r| > |a|} U([-b, -a] \cup [a, b]) \, dx, & |r| > |a| \\
    1, & |r| < |a|
\end{cases}
\]
\[
Pn = 2 \int_{|r| < |a|} N(0, \sigma^2) \, dx.
\]

2.3. Two-step CS measurement and Reconstruction

Since the size parameter \( r \) of \( \Phi' \) is often much smaller than \( M_t \), we can design a method of two-step CS measurement. When \( \beta \) is known, we can use a lower fixed-rate CS matrix \( \Phi' \) to measure \( x_t \), then the energy of \( f N_t \) and the pre-measurement result \( y_t' \) can be obtained. Using the method mentioned above, \( s_t^* \) and \( M_t \) can be solved. Supplementary measurement can be applied and the result \( y_t'' \) can be obtained.
by using a supplementary measurement matrix $\Phi'' \in \mathbb{R}^{(M_t-r) \times N}$. The $\mathbf{y}_t'$ and $\mathbf{y}_t''$ are sent to the reconstruction side. In particular, when $s_t^r$ is small enough, $M_t$ may be smaller than $r$, in this case, the supplementary measurement process should be abandoned, the measurement rate is greater than what is actually needed. However, due to the existence of the noise signal, the value of $s_t^r$ is often not that small, this situation rarely happens if we set up a reasonable value for $r$.

In the reconstruction side, we can get $\mathbf{y}_t'$ by taking data for the first $r$ measurements. Using the same operation as the sampling side, $s_t^r$ and $M_t$ can be solved. The data for the next $M_t-r$ measurements can be taken and we have $\mathbf{y}_t''$, using $\mathbf{y}_t'$, $\mathbf{y}_t''$ we get the measurement result. We can reconstruct the $\mathbf{f}_N^r$ using the SPGL1 [9] method and then the reconstructed $\mathbf{x}_t^r$ can be obtained.

2.4. Implementation considerations

In the process of ARCS using IBS, a special case needs to be considered. For even frames, where $\text{MOD}(t, 2) = 0$, we set $\mathbf{b} = \mathbf{x}_{t-1}$. Here, if the sampling rate of the current frame is higher than that of the previous frame, $M_t > M_{t-1}$, the reconstruction cannot be realized by using the measurement result of the previous frame as $\mathbf{b}$, at this time, the background should be reset as $\mathbf{b} = \bar{x}\mathbf{b}$ and recalculate $\mathbf{y}_t'$, $s_t^r$ and $\mathbf{y}_t''$.

3. Experiments

The video Hall are used as the test sequence, only gray channel frames are used to test the performance of the proposed method. The video is a typical surveillance video, with a fixed background and a significantly changing foreground. The video was shot indoors, using indoor light as the light source, and the background of each frame was significantly affected by noise. Example images from the video sequence are shown in Figure 1.

![Example frame of video sequences](image)

**Figure 1.** Example frame of video sequences

For the video sequence, the ARCS-FSE is used to realize adaptive rate sensing. TBS and IBS are used to acquire foreground signal respectively. Through adaptive compressed sensing and reconstruction of the signals processed by two different background subtraction methods, we compare the sampling rates and the reconstructed image quality of the two methods.

The relevant parameters are set and shown in Table 1.

| Parameters | $\Sigma$ | $a$ | $b$ | $\tau$ | $r$ |
|------------|---------|-----|-----|--------|-----|
| Values     | 2.65    | 16  | 127 | 8      | 600 |

In Table 2, the average CS sampling rate and average PSNR of different methods are shown, where the CS sampling rate is equal to the ratio of $M_t$ to $N$.

| Method | Average Sampling Rate | Average PSNR (dB) |
|--------|-----------------------|-------------------|
| TBS    | 0.2232                | 37.25             |
| IBS    | 0.1812                | 36.73             |

It can be seen that the sampling rate reduces by about 20% when the IBS method is used. At the same time, the average PSNR of the reconstructed image decreases by only about 0.5 dB.

The sampling speed and PSNR of each frame are shown in Figure 2 and Figure 3 respectively. It can be seen that for odd frames, the sampling rate and reconstruction quality of the two methods are
the same, for even frames, the IBS method can significantly reduce the sampling rates, and will cause a slight decrease in the quality of the reconstructed image.

![Figure 2. Measurements number of each frame](image)

4. Conclusion
An improved background subtraction method suitable for two-step ARCS is proposed in this paper. Relevant experiment results show that it can greatly reduce the CS sampling rate under the condition of slightly reducing the image reconstruction quality, which shows better practicability under the condition of limited transmission bandwidth.

5. References
[1] Testa M and Magli E 2016 Compressive estimation and imaging based on autoregressive models *IEEE Transactions on Image Processing* **25**(11) 5077
[2] Tom A J and George S N 2020 Simultaneous Reconstruction and moving object detection from compressive sampled surveillance videos *IEEE Transactions on Image Processing* **29** 7590
[3] Wang J and Chen J 2019 Adaptive-rate compressive sensing for monitoring video based on fast sparsity estimation *ICITEE-2019: 2nd International Conference on Information Technologies and Electrical Engineering* (New York: ACM Int. Conf. Proc. Ser.) **1** 68
[4] Wang Y, Tang C, Chen C, Feng H, Xu Z and Li Q 2018 Adaptive temporal compressive sensing for video with motion estimation *Opt. Rev.* **25** 215
[5] Li H 2019 Compressive domain spatial–temporal difference saliency-based realtime adaptive measurement method for video recovery *IET Image Processing* **13**(11) 2008
[6] Cevher V, Sankaranarayanan A, Duarte M F, Reddy D, Baraniuk R G and Chellappa R 2013 Compressive sensing for background subtraction *10th European Conference on Computer Vision* (New York: Springer Publishing Company) **5303** 155
[7] Ward R 2009 Compressed sensing with cross validation *IEEE Transactions on Information Theory* **55**(12) 5773
[8] Warnell G, Bhattacharya S, Chellappa R and Basar T 2015 Adaptive-rate compressive sensing using side information *IEEE Transactions on Image Processing* **24**(11) 3846
[9] Ewout V D B and Friedlander M P 2009 Probing the Pareto Frontier for basis pursuit solutions *SIAM Journal on Scientific Computing* **31**(2) 890