Block-Based Feature Adaptive Compressive Sensing for Video

Xin Ding*,Wei Chen†*,Ian Wassell*
* *Computer Laboratory, University of Cambridge, UK, Email: (xd225, wc253, ijw24)@cam.ac.uk*
†*State Key Laboratory of Rail Traffic Control and Safety, Beijing Jiaotong University, China.*

***Abstract*—This paper focuses on the problem of feature adaptive reconstruction of Compressive Sensing (CS) captured video. In CS, sparse signals can be recovered with high probability of success from very few random samples. Utilizing the temporal correlations between video frames, it is possible to exploit improved CS reconstruction algorithms. Features that relate to the changes between frames are one of the options to benefit reconstruction. However, to choose the optimal feature for every particular region in each frame is difficult, as the true images are unknown in a CS framework. In this paper, we propose two systems for block-based feature adaptive CS video reconstruction, i.e., a Cross Validation (CV) based system and a classification based system. The CV based system achieves the selection of the optimal feature by applying the techniques of CV to the results of extra reconstructions and the classification based system reduces complexity by classifying the CS samples directly, where the optimal feature for the particular class is employed for the reconstruction. Simulations demonstrate that both of our systems work appropriately and their performance is better than uniformly using any single feature for the whole video reconstruction.**

I. Introduction

The Nyquist-Shannon sampling theorem requires a signal sampling rate that is too high or too costly in applications such as imaging, video, medical imaging, remote surveillance, spectroscopy, etc [1]. Conventionally, to address the challenges involved in dealing with these high-dimensional data, we depend on signal compression after acquisition, e.g., using JPEG and JPEG2000 [2]. Thus a lot of data is thrown away after sensing, making this sense-then-compress framework wasteful of resources, which is particularly problematic for battery powered devices.

Compressive Sensing (CS) [3][4] has emerged as a new framework for signal acquisition and is attracting increasing attention. It builds upon the fact that many signals can be represented with only a few non-zero coefficients in a suitable basis or dictionary. CS achieves sensing and compression at the same time and enables recovery of such signals from samples which are far fewer than that required by the Nyquist-Shannon sampling theorem. In particular, CS is a very attractive technique for applications with high data acquisition cost. For example, it has had notable impacts on medical imaging [5], sensor networks [6], compressive radar [7].

CS is also making significant contributions in the field of image processing. For practical imaging systems, a conventional camera would require one sensor (e.g., focal plane array element) per pixel to capture an image; however, a CS-based camera alleviates this constraint. Some well-known examples are the Rice single-pixel camera [8], coded aperture imagers [9], CMOS CS imagers [10][11] and spectral imagers [12][13].

Video sequences, that can essentially be considered as sets of images, can be sampled and reconstructed using CS theory. Naively, we can recover videos frame by frame in an independent manner, but unfortunately this approach is inefficient. The correlation present between video frames make it possible for us to exploit more efficient sampling and reconstruction methods. In [14], differences between frames are employed to provide better reconstruction results on videos that have only small spatial changes. In [15] and [16], frames are initially assumed as unchanged, i.e., the current frame is same as the previous frame, permitting simplified reconstructions and further steps are proposed to refine the initial results. In [17], in order to involve temporal correlation, CS video is reconstructed by utilizing a dictionary that is formed using the blocks in neighbor frames. In [18] and [19], motion estimation is incorporated into the CS reconstruction procedure, thereby improving the reconstruction performance of CS captured video.

In contrast to these methods, in this paper, we consider feature-specific video reconstruction [20]. More precisely, in a video sequence, many kinds of changes may occur between frames, for example, a whole scale change in intensity, small movements or a new object appearing. These various changes results in different features that can benefit the reconstruction. Besides, regions in a frame and all frames in a sequence may move independently, which will generate different inter-frame correlation in different regions [21]. Utilizing the optimal feature for each local region according to its inter-frame changes can provide an optimal reconstruction for the whole video, which motivates the investigation of block-based feature adaptive reconstruction for CS captured video.

II. The Theory of Compressive Sensing

CS relies on the fact that many natural signals have concise representations when expressed in some certain basis. Mathematically, a signal $\mathbf{x} \in \mathbb{R}^N$ is said to be K sparse when it can be expressed as:

$$\mathbf{x} = \mathbf{\Psi s}, \tag{1}$$

where $\mathbf{\Psi} \in \mathbb{R}^{N \times N}$ is an orthonormal basis and $\mathbf{s} \in \mathbb{R}^N$ has only K ($K \ll N$) non-zero coefficients. Then, such a signal is sampled via CS as:

$$\mathbf{y} = \mathbf{\Phi x} + \mathbf{n} = \mathbf{\Phi \Psi s} + \mathbf{n} = \mathbf{As} + \mathbf{n}, \tag{2}$$

This work is supported by EPSRC Research Grant (EP/K033700/1); the Natural Science Foundation of China (61401018); Beijing Jiaotong University; the Fundamental Research Funds for the Central Universities (2014JBM149).

where $\mathbf{\Phi} \in \mathbb{R}^{M \times N} (M \ll N)$ represents the sensing matrix, $\mathbf{n} \in \mathbb{R}^M$ denotes the measurement noise and $\mathbf{A} = \mathbf{\Phi\Psi}$ is the equivalent sensing matrix.

Conventionally, it is impossible to recover $\mathbf{s}$ from the measurements taken in (2) as it is an under-determined problem. However, CS asserts that a solution to the problem is guaranteed provided specific conditions are imposed on the sensing matrix. A criterion called the Restricted Isometry Property (RIP) is commonly used to evaluate whether the matrix $\mathbf{A} = \mathbf{\Phi\Psi}$ is qualified [3][22].

Definition 1 *For integers K=1, 2, ..., define the Restricted Isometry Constant (RIC) δ_K of a matrix $\mathbf{A}$ as the smallest number such that*

$$(1 - \delta_K)||\mathbf{s}||_{l_2}^2 \leq ||\mathbf{As}||_{l_2}^2 \leq (1 + \delta_K)||\mathbf{s}||_{l_2}^2 \tag{3}$$

holds for all K-sparse vectors $\mathbf{s}$.

A Gaussian or Bernoulli distributed matrix is often used as the sensing matrix $\mathbf{\Phi}$ because of their universal incoherence, in which case RIP holds for the matrix $\mathbf{A}$ with high probability regardless of the choice of $\mathbf{\Psi}$[23]. However, in the sampling process when using such matrices, although the number of measurements is less than the total number of pixels, all the pixels are in fact obtained in the form of linear combinations. To achieve simpler hardware implementation, another technique, namely random sampling is proposed [24], [25]. In this scheme, only a small, uniformly distributed, randomly chosen fraction of the coefficients is captured. The entries of such a sensing matrix $\mathbf{\Phi}_t \in \mathbb{R}^{M \times N}$ are all zeros except for M unity valued entries in M different columns and rows.

The CS reconstruction procedure is then modeled as the following optimization problem:

$$\min ||\mathbf{s}||_{l_0} \quad s.t. ||\mathbf{y} - \mathbf{As}||_{l_2} \leq \varepsilon, \tag{4}$$

in which $||\cdot||_{l_0}$ denotes the l_0 norm which is given by the number of nonzero entries and ε is a tolerance parameter. Many reconstruction methods have been proposed to solve this problem [26], [27], [28], [29], among which Basis Pursuit (BP) is considered to have the best performance in the sense that it can recover accurate signals from the fewest number of measurements since it has the property of being the global optimal [30]. In BP, the NP hard problem in (4) is relaxed as:

$$\min ||\mathbf{s}||_{l_1} \quad s.t. ||\mathbf{y} - \mathbf{As}||_{l_2} \leq \varepsilon. \tag{5}$$

The reconstruction accuracy of the l_1 minimization problem is guaranteed by the following theorem.

Theorem 2 *Assume that $\delta_{2K} < \sqrt{2} - 1$ and $||\mathbf{n}||_{l_2} \leq \varepsilon$. Then the solution $\mathbf{s}^*$ to (5) obeys*

$$||\mathbf{s}^* - \mathbf{s}||_{l_2} \leq C_0 K^{-1/2} ||\mathbf{s} - \mathbf{s}_K||_{l_1} + C_1 \varepsilon \tag{6}$$

where $C_0 = \frac{2+(2\sqrt{2}-2)\delta_{2K}}{1-(\sqrt{2}+1)\delta_{2K}}$, $C_1 = \frac{4\sqrt{1+\delta_{2K}}}{1-(\sqrt{2}+1)\delta_{2K}}$, δ_{2K} is the RIC of matrix $\mathbf{A}$, $\mathbf{s}_K$ is an approximation of $\mathbf{s}$ with all but the K largest entries set to zero.

III. Joint Compressive Sensing for Video

A. Joint Sensing Formulation

Temporal information can also be exploited when it comes to the CS video reconstruction. Using a joint sensing formula [20], every two neighbor frames in a video sequence are sensed by:

$$\mathbf{y} = \mathbf{\Phi x} + \mathbf{n} = \mathbf{\Phi\Psi s} + \mathbf{n}, \tag{7}$$

where the matrices and vectors are all in joint forms as:

$$\mathbf{y} = \begin{bmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \end{bmatrix}, \mathbf{\Phi} = \begin{bmatrix} \mathbf{\Phi_1} & \mathbf{0} \\ \mathbf{0} & \mathbf{\Phi_2} \end{bmatrix}, \mathbf{x} = \begin{bmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \end{bmatrix}, \mathbf{n} = \begin{bmatrix} \mathbf{n}_1 \\ \mathbf{n}_2 \end{bmatrix}. \tag{8}$$

Note that in (7), even though $\mathbf{x}$ can be expressed by $\mathbf{x}_1$ and $\mathbf{x}_2$, $\mathbf{s}$ may not be separable. Estimating $\mathbf{s}$ jointly allows us to take advantage of correlation between the frames at two time instants. To enable causal reconstruction (i.e., only the current measurements and the previously reconstructed frame are employed), in the rest of this paper, we consider $\mathbf{\Phi}_1$ as the identity matrix $\mathbf{I}$, which means the first frame in the joint scheme is completely sampled as a reference or that it has been recovered in the previous step. We choose $\mathbf{\Phi}_2$ as a random sampling matrix [24], [25] as described in Section II.

B. Feature Specific CS Video Reconstruction

To exploit the correlation between frames, the reconstruction scheme for jointly sampled images could be different to that for conventional CS. Here, we consider four reconstruction schemes that employ different features. Note that in the following schemes, $\mathbf{x}$, $\mathbf{y}$ and $\mathbf{\Phi}$ are all in the joint forms, as indicated in (8).

- **Scheme 1:** $\min ||\mathbf{s}||_{l_1} \quad s.t. ||\mathbf{y} - \mathbf{\Phi\Psi s}||_{l_2} \leq \varepsilon$, where $\mathbf{\Psi}$ is the sparsifying dictionary, in this case the DWT. Note that $\mathbf{s}$ is a joint estimation of the two frames, therefore it is not separable.
- **Scheme 2:** $\min ||\triangle \mathbf{x}||_{l_1} \quad s.t. ||\mathbf{y} - \mathbf{\Phi x}||_{l_2} \leq \varepsilon$, where $||\triangle \mathbf{x}||_{l_1} = ||\mathbf{x}_2 - \mathbf{x}_1||_{l_1}$. This scheme takes into account the difference between frames by favoring values closer to 0.
- **Scheme 3**: $\min (\mathbf{x}_2)_{TV} \quad s.t. ||\mathbf{y} - \mathbf{\Phi x}||_{l_2} \leq \varepsilon$, where $(\mathbf{x})_{TV}$ denotes the Total Variation (TV) of $\mathbf{x}$. It is defined as: $(\mathbf{x})_{TV} = \sum_i |\triangle_i^h \mathbf{x}| + |\triangle_i^v \mathbf{x}|$, where $\triangle_i^h$ and $\triangle_i^v$ are the first order horizontal and vertical difference operators, respectively.
- **Scheme 4:** $\min (\triangle \mathbf{x})_{TV} \quad s.t. ||\mathbf{y} - \mathbf{\Phi x}||_{l_2} \leq \varepsilon$, in which we impose the TV condition on the difference image between the frames, instead of the frame itself which needs to be recovered.

Empirically, the performances of the schemes are different for various scenes. In other words, each of the schemes has a specific scenario for which it is optimal. Specifically, Scheme 2 and 4 outperform Scheme 1 and 3 when the changes between frames are quite small, while Scheme 3 is the best when the scene totally changes and this change or the difference image is not sufficiently smooth (otherwise Scheme 4 may be the best instead). The reason for this is that for a specific scenario, the scheme that has the most sparse objective will result in the best performance. This observation is in accordance with the conclusion in [14]. Therefore, in the following section, we will design systems to optimally choose the features that are

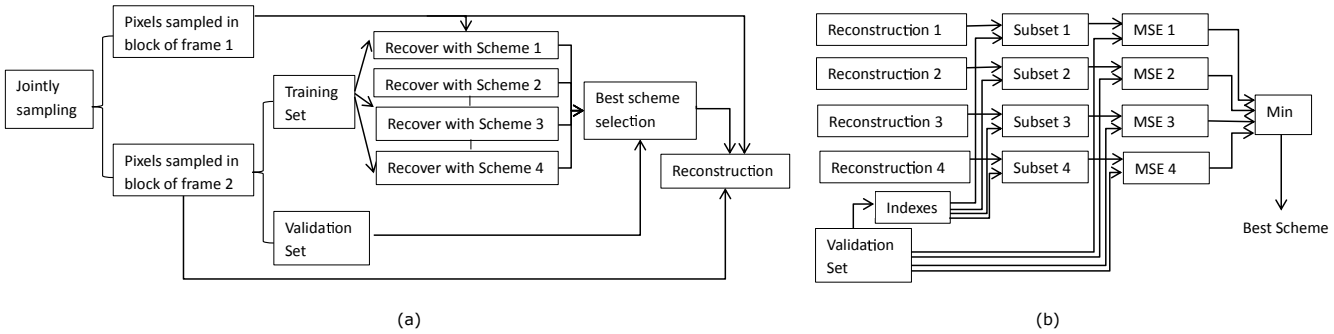

Fig. 1. Illustration of CV based system. (a) Framework; (b) Best Scheme selection procedure.

utilized in CS reconstruction for every local region according to its temporal changes in video.

IV. Feature Adaptive CS Video Reconstruction

In this section, we propose two systems for CS video reconstruction with adaptive feature selection. Considering the changes in each region of video can move independently, we conduct block-based sensing and reconstruction so that a locally optimal choice can be made. We employ the joint sensing schemes and the reconstruction schemes utilizing different features as described in Section III-A.

A. Cross Validation Based System

As discussed, the performance of the reconstruction schemes using various features is scene-related. However, in CS video system, we have no knowledge about the scene, and indeed all we have are some random CS samples. In order to tackle this difficulty, we propose to employ Cross Validation (CV) to adaptively choose the most appropriate reconstruction scheme.

Specifically, we first obtain the joint CS samples, in which the samples of the current frame are partitioned into a training set and a validation set. Then, using the samples in the training set, and also the samples in the reference frame, joint reconstruction using all four candidate schemes are conducted. The reconstruction results are then compared with the validation set to determine the optimal scheme for the particular block. The framework of this system is shown in Fig. 1 (a).

The procedure for selecting the best scheme is shown in detail in Fig. 1 (b), where the four reconstructions (1-4) arise owing to each of the four recovery schemes and the validation set is from the previous step in Fig. 1 (a). As shown, in the best scheme selection procedure, we first pick the subsets of pixels from the reconstruction results that correspond with those in the validation set and calculate the Mean Square Error (MSE) between them. The scheme with smallest MSE is chosen as the best scheme, which is then employed for the final reconstruction in Figure 1. Note that all the processes operate in a block-based manner. The whole frame can be recovered by repeating the process block by block and the whole video sequence can be reconstructed frame by frame.

It is noticeable that the CV based system has very high computational complexity, as it involves repeatedly solving the under-determined optimization problem. We will thus design a classification based system in next section, where the optimal feature is selected directly according to the CS samples.

B. Classification Based System

In this section, we propose a system for adaptively reconstructing video images using its local optimal features based on classification. The framework of the system is demonstrated in Fig. 2 (a). As shown, we first classify the type of the block according to the samples in both the current and the previous frame, after which the reconstruction is performed using the scheme selected as being the most appropriate for the particular block under consideration. Fig. 2 (b) shows the detailed procedure for classification, where $\mathrm{T}_1, \mathrm{T}_2, \mathrm{T}_3$ are particular threshold values used in the classification procedure. Note that Scheme 1 is no longer included in the candidates in this section, since by conducting simulations as in Section V-A, we found that scheme 1 has the generally worst performance.

In the process of classification, two indicators as follows are employed:

- **Indicator A**: $\mathrm{A} = \frac{||\mathbf{y}_2 - \mathbf{y}_1||_{l_p}}{||\mathbf{y}_1||_{l_p}}$, where $\mathbf{y}_2 = \mathbf{\Phi}_2\mathbf{x}_2$ is the measurement vector of the current frame, $\mathbf{y}_1$ is the corresponding vector of the previous frame with the pixels in the same positions as $\mathbf{y}_2$. $||\cdot||_{l_p}$ denotes the l_p norm and $0 < p < 1$. Therefore, the numerator in indicator A indicates the degree of change between frames and the denominator is included for normalization. For the selection of the value of p, consider the following examples: in a 32×32 image, compared to its previous frame, (i) 510 pixels change their values from 10 to 11 (the value range of pixels is [0, 255]); (ii) 510 pixels change their values from 10 to 200; (iii) 2 pixels change their values from 0 to 255. If $p = 0$, the indicator cannot distinguish between (i) and (ii); while if $p = 1$, the indicator cannot distinguish between (i) and (iii). Therefore we use $0 < p < 1$ to balance between the effects coming from the number of changed pixels and the degree of change of pixel intensity.
- **Indicator B**: $\mathrm{B} = std[(\mathbf{y}_2 - \mathbf{y}_1)_v]$, where std means Standard Deviation (SD) which is defined as $\sigma = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(x_i - \mu)^2}$, where $\mu = \frac{1}{N}\sum_{i=1}^{N} x_i$. Note that SD indicates how much variation exists around the mean value. A low SD means that the data points tend to be very close to the mean while high SD shows that the

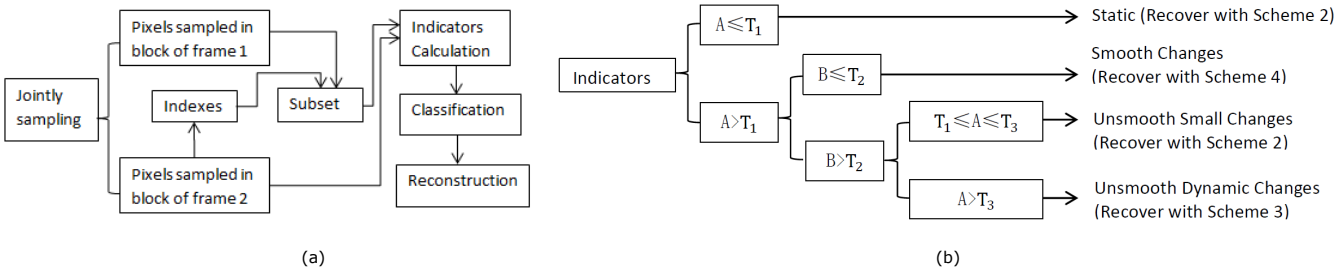

Fig. 2. Illustration of Classification Based System. (a) Framework; (b) classification procedure.

Fig. 3. Examples of reconstructed frames by CV based system.

samples are spread out over a large range of values. We use $(\mathbf{y}_2 - \mathbf{y}_1)_v$ to denote the pixel set corresponding to changed areas of the difference image. Thus indicator B reveals the smoothness presented in the changed areas.

We classify the blocks into four types: 'static', 'smooth changes', 'unsmooth small changes' and 'unsmooth dynamic changes'. Each type of block is allocated one specific reconstruction scheme. As investigated in Section III-B and V-A, when the blocks barely change, both Scheme 2 and Scheme 4 can provide quite high quality reconstruction. Considering that the complexity of Scheme 4 is higher, we allocate Scheme 2 for the 'static' blocks. Also, we have found that Scheme 4 performs best when the changes are sufficiently smooth, no matter whether the change is small or large. Therefore, when some change occurs, we first evaluate the smoothness of the change. If it is sufficiently smooth, the block is classified as 'smooth changes' and will be reconstructed by Scheme 4. Otherwise, we need to evaluate the degree of change in a further step. If the change is small enough, the block is viewed as 'unsmooth small changes', which is suitable to be recovered by Scheme 2. If the change is large, Scheme 3 outperforms Scheme 2, as discussed in Section III-B. Therefore Scheme 3 is chosen for the blocks identified as having 'unsmooth dynamic changes'.

Based on this classification scheme, we could also include an optional function in the system to further improve the reconstruction quality. That is, if the block is classified as 'unsmooth dynamic changes', we could take some extra pixels to increase the sampling ratio. It is noticeable from our experimental results in Section V-A that even though Scheme 3 is the best option for this type of block, its reconstruction quality is still much lower than that for blocks of other types. This is because Scheme 3 does not utilize any correlation between frames. More samples can compensate for this and thus improve the performance.

V. Numerical Experiments

A. Performance of Cross Validation Based System

In this section, we evaluate the framework proposed in Section IV-A. The video that we employ in the experiments has dimensions of 641×509 and it consists of various types of blocks in the frames. The block size is set as 32×32. The sampling strategy as described in Section III-A is utilized and for the random sampling, we use a sampling ratio of 20% (i.e., $M/N = 0.2$), of which 30% are kept as the validation set. The previously recovered frame is used as the reference for the reconstruction of the next frame. We employ the Structural Similarity Index (SSI) [31] to evaluate the reconstruction quality. The value of SSI ranges from 0 to 1 and the higher the value indicates better reconstruction.

Some reconstructed examples are shown in Fig. 3. It can be seen that the recovery has a high quality, especially in the parts of the images with no or small changes. To investigate the performance improvements made possible by the CV based system, we also perform reconstructions employing the same scheme for every block. For the frames shown in Fig. 3, we summarize the SSI results of our CV based system and for each of the four schemes in Table I. The first frame is not included in the table since it is completely sampled as a reference. In the table, we also include the SSI of the 'Oracle', which means that the recovery result for each block is chosen based on the true MSE, i.e., the MSE between recovered block and the original block. This is infeasible in practice because doing so requires the original image to be known.

From the SSI summary, we can observe that the CV based system can achieve a similar performance to that of the 'Oracle', which means that the system works appropriately. Also, it can be seen that the system can reconstruct the images having a quality, either as good as the best scheme, or better than that of any of the schemes alone. This is determined by the changes between frames. For example, frame 2 has barely changed compared to frame 1, therefore both Scheme

TABLE I
SSI SUMMARY

SSI	Frame 2	Frame 3	Frame 4	Frame 5
Oracle	0.9838	0.9077	0.8391	0.8896
CV	0.9833	0.9050	0.8320	0.8866
Scheme 1	0.3953	0.4303	0.4616	0.4110
Scheme 2	0.9804	0.7847	0.5599	0.7781
Scheme 3	0.6327	0.6730	0.7176	0.6263
Scheme 4	0.9799	0.8733	0.7141	0.8416

2 and Scheme 4 can provide very high quality recovery in such a case. Accordingly, in the CV based system, all of the blocks are selected to be reconstructed with Scheme 2 or 4, which results in the similar SSI for 'CV', 'Scheme 2' and 'Scheme 4' in the reconstruction of frame 2. However, for frame 4, the changes compared to the previous frame happen in the two areas having cars, which leads to a diverse selection of schemes. Scheme 2 or 4 is chosen for the non-changed area and Scheme 3 or 4 is selected for the regions with cars. Thus the CV system outperforms any of the constant-scheme reconstructions. Clearly, the advantage of the CV based system is more distinct when the changes between frames can lead to more diverse scheme selections.

To illustrate the difference in performance between the schemes, we also show the four recovered versions of frame 4 as an example in Figure 4. By comparing these results with frame 4 in Figure 3, we can conclude that all the reconstruction schemes give worse results than that for the CV based system. For Scheme 1, obvious block artifacts are visible and the recovery quality is quite poor. Scheme 2 can achieve perfect reconstruction for the blocks with no changes, however, for the dynamically changing blocks, the recovery is poor, for example, we can only observe a ghost-like shape of the car in the frame. For Scheme 3, it can reconstruct the car with an acceptable quality, but for the other regions, the reconstruction quality is not as good as that in Scheme 2 and 4. For Scheme 4, its recovery of the car region is worse than that by Scheme 3, because there are more discontinuities in the difference image than in the frame itself. Our CV based system performs best since that it picks the best reconstruction for every region and combines them to form the whole frame.

B. Performance of Classification Based System

In this section, we evaluate the performance of the classification based system using the same video sequence as in Section V-A. The block size is again 32×32 and the sampling ratio is 20%. In the indicators calculation and classification processes, we employ the following parameters: $p = 0.5$, $\mathrm{T}_1 = 0.03$, $\mathrm{T}_2 = 20$, $\mathrm{T}_3 = 0.07$. These parameters were established empirically by conducting experiments and it turns out that T_1 and T_3 have similar values to those obtained in [14]. In the first instance, we have not included the optional function of increasing the sampling ratio and the recovered sequence is shown in Figure 5. The recovery has a similar quality to that of the previous CV based system.

We also summarize the SSI results in Table II. The oracle data has previously been presented in Table I. The results for

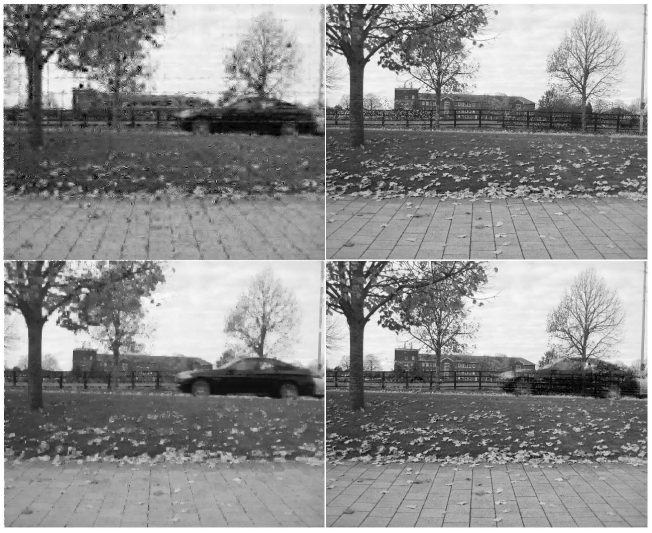

Fig. 4. Reconstructed Frame 4 for Each Scheme

TABLE II
SSI FOR CLASSIFICATION BASED SYSTEM

SSI	Frame 2	Frame 3	Frame 4	Frame 5
Oracle	0.9838	0.9077	0.8391	0.8896
Classify (exc. optional)	0.9830	0.9070	0.8324	0.8832
Classify (inc. optional)	0.9871	0.9411	0.8906	0.9018

the system including the optional function is also included in the table, where the sampling ratio for the blocks categorized as 'unsmooth dynamic changes' is increased to 30%. From the table, it can be seen that the classification based system can achieve similar recovery quality to that of 'oracle', which means that our proposed framework operates as expected. Furthermore, when the optional function is employed, the performance improves. The more blocks that utilize Scheme 3, the greater the improvement that can be obtained, but at the cost of a higher sampling ratio.

However, there are some aspects that can impact on the performance of the classification based system. The block size should be smaller than the size of the changing objects. Otherwise, some sensitive cases may exist in the difference image, for example, as shown in Figure 6, where the black area means no changes. In this example, the SD of changed area is quite small because all changes have the same values. Obviously it will be classified as 'smooth changes' and recovered by Scheme 4. However, the changed area is not consecutive. The TV of the difference image is not that sparse, so Scheme 3 may be more suitable. Setting the block size appropriately can generally help to avoid this kind of sensitive case. This is not a requirement that is difficult to implement and the smaller block size can also dramatically speed up the reconstruction. Other sensitive cases may occur as well, for example the calculation of the SD is affected by the extreme values. More work is needed to explore robust solutions to the sensitive cases.

VI. CONCLUSION

In this paper, we proposed two systems for block-based feature adaptive CS video reconstruction. Our systems are capable of adaptively selecting optimal features that are uti-

Fig. 5. Examples of reconstructed frames by Classification Based System

Fig. 6. Sensitive Case Example

lized in the reconstruction for different regions of the frames. Experiments show that the CV based system can successfully choose the best scheme for the blocks and provide high quality reconstruction. Its reconstruction performance is better than employing a single scheme for the entire image sequence particularly when the images have high diversity. The classification based system reduces complexity compared with the CV based system by classifying blocks according to indicators directly calculated from the samples. It is shown to perform as expected and can be further improved by utilizing the optional function of increasing the sampling ratio for a specific block type.

REFERENCES

[1] M. Davenport and M. Duarte, *Introduction to Compressed Sensing*, Y. C. Eldar and G. Kutyniok, Eds. Cambridge University Press, 2012, vol. Chapter in Compressive Sensing: Theory and Applications.

[2] C. Christopoulos, A. Skodras, and T. Ebrahimi, "The JPEG2000 still image coding system: an overview," *IEEE Trans. Consumer Electron.*, vol. 46, no. 4, pp. 1103–1127, Nov 2000.

[3] E. J. Candes, J. Romberg, and T. Tao, "Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information," *IEEE Trans. Information Theory*, Feb. 2006.

[4] D. L. Donoho, "Compressed sensing," *IEEE Trans. Information Theory*, April 2006.

[5] M. Lustig, D. L. Donoho, J. M. Santos, and J. M. Pauly, "Compressed sensing MRI," *IEEE Signal Process. Magazine*, March 2008.

[6] J. Haupt, W. Bajwa, M. Rabbat, and R. Nowak, "Compressed sensing for networked data," *IEEE Signal Processing Magazine*, vol. 25, no. 2, pp. 92–101, March 2008.

[7] J. H. G. Ender, "On compressive sensing applied to radar," *Signal Processing*, vol. 90, no. 5, pp. 1402–1414, 2010.

[8] M. Duarte, M. Davenport, D. Takhar, J. Laska, T. Sun, K. Kelly, and R. Baraniuk, "Single-pixel imaging via compressive sampling," *IEEE Signal Process. Magazine*, vol. 25, no. 2, pp. 83 –91, March 2008.

[9] R. F. Marcia, Z. T. Harmany, and R. M. Willett, "Compressive coded aperture imaging," *SPIE 7246, Comput. Imag. VII*, p. 72460G, 2009.

[10] R. Robucci, J. Gray, L. K. Chiu, J. Romberg, and P. Hasler, "Compressive sensing on a CMOS separable-transform image sensor," *in Proc. IEEE ICASSP*, vol. 98, no. 6, pp. 1089 –1101, June 2010.

[11] V. Majidzadeh, L. Jacques, A. Schmid, P. Vandergheynst, and Y. Leblebici, "A (256x256) pixel 76.7mW CMOS imager/ compressor based on real-time in-pixel compressive sensing," in *in Proc. IEEE ISCAS*, June 2010, pp. 2956 –2959.

[12] M. Gehm, R. John, D. Brady, R. Willett, and T. Schulz, "Single-shot compressive spectral imaging with a dual-disperser architecture," *Opt. Express*, vol. 15 (21), pp. 14 013–14 027, 2007.

[13] A. Wagadarikar, R. John, R. Willett, and D. Brady, "Single disperser design for coded aperture snapshot spectral imaging," *Appl. Opt.*, vol. 47 (10), pp. B44–B51, 2008.

[14] J. Zheng and E. L. Jacobs, "Video compressive sensing using spatial domain sparsity," *Optical Engineering*, vol. 48, no. 8, pp. 087 006–10, 2009.

[15] N. Vaswani, "KF-CS: Compressive sensing on Kalman filtered residual," *Proc. CoRR*, vol. abs/0912.1628, 2009.

[16] N. Vaswani and W. Lu, "Modified-CS: Modifying compressive sensing for problems with partially known support," *IEEE Trans. Signal Process.*, vol. 58, no. 9, pp. 4595–4607, 2010.

[17] T. T. Do, Y. Chen, D. Nguyen, N. Nguyen, L. Gan, and T. Tran, "Distributed compressed video sensing," in *Proc. IEEE ICIP*, 2009, pp. 1393–1396.

[18] H. Jung and J. C. Ye, "Motion estimated and compensated compressed sensing dynamic magnetic resonance imaging: What we can learn from video compression techniques," *Inter. J. IST*, vol. 20, pp. 81–98, 2010.

[19] X. Ding, W. Chen, and I. J. Wassell, "Generalized-KFCS: Motion estimation enhanced Kalman filtered compressive sensing for video," in *Proc. IEEE ICIP*, Oct. 2014, pp. 1297–1301.

[20] S. Uttam, N. Goodman, and M. Neifeld, "Feature-specific difference imaging," *IEEE Trans. Image Process.*, vol. 21, no. 2, pp. 638 –652, Feb. 2012.

[21] Z. Liu, H. Zhao, and A. Elezzabi, "Block-based adaptive compressed sensing for video," in *in Proc. IEEE ICIP*, Sep 2010, pp. 1649 –1652.

[22] E. Candès and T. Tao, "Decoding by linear programming," *IEEE Trans. Information Theory*, vol. 51, no. 12, pp. 4203 – 4215, Dec. 2005.

[23] E. J. Candès and M. B. Wakin, "An introduction to compressive sampling," *IEEE Signal Process. Magazine*, vol. 25, no. 2, pp. 21 –30, March 2008.

[24] W. Chen and I. Wassell, "Energy-efficient signal acquisition in wireless sensor networks: a compressive sensing framework," *IET Wireless Sensor Systems*, vol. 2, no. 1, pp. 1–8, March 2012.

[25] F. A. Boyle, J. Haupt, G. L. Fudge, and C.-C. A. Yeh, "Detecting signal structure from randomly-sampled data," in *in Proc. IEEE workshop on SSP*, Aug. 2007, pp. 326 –330.

[26] J. Tropp and A. Gilbert, "Signal recovery from random measurements via orthogonal matching pursuit," *IEEE Trans. Information Theory*, vol. 53, no. 12, pp. 4655 –4666, Dec. 2007.

[27] D. Needell and J. Tropp, "CoSaMP: Iterative signal recovery from incomplete and inaccurate samples," *Applied and Computational Harmonic Analysis*, vol. 26, no. 3, pp. 301 – 321, 2009.

[28] D. L. Donoho, A. Maleki, and A. Montanari, "Message-passing algorithms for compressed sensing," *in Proc. National Academy of Sciences*, vol. 106, no. 45, pp. 18 914–18 919, 2009.

[29] M. Duarte and Y. Eldar, "Structured compressed sensing: From theory to applications," *IEEE Trans. Signal Process.*, vol. 59, no. 9, pp. 4053 –4085, Sep 2011.

[30] S. Chen, D. Donoho, and M. Saunders, "Atomic decomposition by basis pursuit," *SIAM review*, vol. 43(1), pp. 129–159, 2001.

[31] Z. Wang, A. Bovik, H. Sheikh, and E. Simoncelli, "Image quality assessment: from error visibility to structural similarity," *IEEE Trans. Image Process.*, vol. 13, no. 4, pp. 600 –612, april 2004.