ADAPTIVE SALIENCY-BASED COMPRESSIVE SENSING IMAGE RECONSTRUCTION

A. Akbari, D. Mandache, and M. Trocan

Institut Suprieur d'Electronique de Paris (ISEP)
Paris, France
(ali.akbari, diana.mandache, maria.trocan)@isep.fr

B. Granado

Pierre et Marie Curie University
Paris, France
bertrand.granado@lip6.fr

ABSTRACT

This paper proposes an adaptive compressive sensing reconstruction method which provides a higher recovered image quality. Based on an initial compressive sampling reconstruction at a given sampling rate, the visually salient regions of the image that are more conspicuous to the human visual system are extracted using a classical graph-based method. The target acquisition subrate is further adaptively allocated among these regions, such that the new acquisition will favor the interest areas. The measurements produced by this adaptive method are fully compatible with the existing sparse reconstruction algorithms, which favors the utilization of the proposed scheme. Simulation results show that the saliency-based compressive sensing recovery method outperforms the conventional sparse reconstruction algorithms in terms of image quality at the same target sampling ratio with a smaller increment in the complexity.

Index Terms— Compressive sensing, adaptive sampling rate, sparse reconstruction, saliency map, HVS

1. INTRODUCTION

The theory of compressed sensing (CS) provides a new approach for signal acquisition wherein signal can be exactly reconstructed using a small number of random linear measurements, under certain sparsity conditions [1]. Since most signals are indeed compressible in some transform domains, CS has attracted a lot of attention in many applications, including medical imaging, camera design, and multimedia sensor networks due to its potential of reduction of sampling rates, power consumption and computation complexity in the image acquisition.

In case of image and video, low complexity of the sensing scheme plays an important role in designing an imaging sensor. Block-based CS (BCS) scheme, where non-overlapped blocks are sensed separately, comes to solve this issue due to its advantage of low complexity sampling and high reconstruction quality [2]. However, the conventional BCS samples all blocks with the same number of measurements, neglecting the subjective importance of each block to Human Visual System (HVS), that can lead to difficulty of recovering fine grained details. On the other hand, from the view of biological vision and scientific analysis, the visual significance of each block varies with its spatial position [3]. Some regions can be more sensitive to the HVS, while others have a lower level of visual interest. Therefore, it is necessary to design an adaptive block based compressive sensing scheme by joining the block based sampling and the HVS characteristics in order to achieve a smart-human recovery performance.

In order to improve the limitations of the traditional fixed-rate BCS, several studies have been conducted in the adaptive BCS from various perspectives. In [4], the authors give a theoretical analysis of the adaptive compressive sampling. In this paper, authors confirm that adaptive measurements significantly outperform non-adaptive systems in practice. Some acquisition techniques take benefit from local features extraction in the measurement domain such as standard deviation [5]- [6], edge counting [7] or estimation of reconstruction error [8]. Inspired by relationship between the compressibility and the redundancy of images, an adaptive scheme is proposed in [9] by estimation of the compressibility based on the local redundancy which is measured by statistics of the pre-sensed measurements. These measurement-domain based recovery algorithms can benefit from on-the-fly adaptive sampling (they do not require feedback channel from receiver side) but are often not so accurate in the reconstruction. Only few works perform the adaptive BCS by combining the sampling and reconstruction together. In [10] this procedure is used to improve the reconstruction quality for pedestrian tracking in the video surveillance applications.

Visual saliency is a cognitive mechanism of the HVS in order to accurately identify significant visual information (salient or foreground regions) and filter out other redundant visual information (non-salient or background regions) for the natural scenes. HVS focuses on the salient regions while ignoring non-salient areas when exploring a natural scene.

To the best of our knowledge, there is no application of the visual saliency in the image compressive sensing domain. Different from previous approaches in the adaptive CS, we adopt the adaptive compressive sensing framework by access to side information, such as a saliency map, which leads to an efficient CS image reconstruction. In this sense, we try to

obtain a saliency model of the original image.

Since the complex human visual system performs numerous functions when viewing, it is not possible to combine it into the direct sensing process. Beside of the hardware problems, the process of integrating the salient region detection into the cameras is not straightforward due to its time and energy consuming process. However, by considering a feedback channel between the sensor and the receiver, the CS can gain from the visual saliency for improving the image reconstruction quality. In addition, the feedback information should be as few as possible, then a low bandwidth channel is required.

The procedure begins with a fixed-rate compressive sampling for all blocks and a primary recovery at the receiver. Then a classical graph based method [11] is used to compute the saliency map, so that the HVS interesting regions are extracted. Then, a binary map, where the ones represent the blocks belonging to the salient region, is organised. We achieve the adaptivity in sampling by establishing a feedback channel between the sensor and receiver, making subsequent measurements more directly into the salient regions of the image.

Experimental results manifest superior improvement of the proposed method compared with the conventional fixed-rate CS reconstruction scheme at the expense of a low complexity increase. Also, subjective results demonstrate that the proposed method reduces reconstruction artifacts and preserves more details.

The remainder of the paper is organized as follows. In Section 2, an overview of the compressive sensing is given. Section 3 focuses on the proposed saliency-based compressive sensing framework. The results analysis is shown in Section 4 and some conclusions are drawn in Section 5.

2. COMPRESSIVE SENSING

In the compressed sensing, given an original signal $x \in \Re^N$, the observed measurements $y \in \Re^M$ are sampled by $y = \Phi x$, where Φ is a $M \times N$ random orthonormal matrix ($M \ll N$). The sampling rate (*subrate*) is defined as M/N. The fundamental concept of the CS theory states that the signal x can still be exactly recovered from the measurements y, if it is sparse enough [12], although the number of unknowns is larger than the number of measurements. Quite often, real world signals, such as image and video, are sparse (or compressible) with respect to some transform basis Ψ. Then, the recovery problem turns into the generation of a sparse set of transform coefficients. It should be noted that Φ must satisfy the Restricted Isometry Property (RIP) and be incoherent with Ψ [12].

How to efficiently reconstruct the image with a good recovery quality and low computational complexity is a hot topic in the CS applications fields. To address these issues, there have been some great efforts to develop the CS reconstruction algorithms which differ based on their computational complexity and reconstruction quality. The algorithm proposed in [13] searches for the x with smallest l_1-norm consistent with the observed y,

$$x = \underset{x}{\operatorname{argmin}} \|x\|_1 \qquad \text{s.t.} \qquad y = \Phi x. \tag{1}$$

Because of the high computational complexity of this algorithm, a fast iterative method based on successive projections and thresholding in the transform domain, called Projected Landweber (PL), has been proposed which provides a good trade-off between computational complexity and reconstruction quality [14]. Given an initial approximation of the transform coefficients X^0, the algorithm updates X^i at iteration i as following:

$$X^i = X^{i-1} + \Psi\Phi^T(y - \Phi\Psi^{-1}X^{i-1}), \tag{2}$$

where Ψ^{-1} denotes the inverse transform. In addition, a thresholding process controls the sparsity of the generated transform coefficients at each iteration.

In the case of images and videos acquisition, saving the measurement matrix Φ requires a large memory; moreover, it increases the computational burden during the reconstruction. Block compressed sensing (BCS) comes to solve these issues by splitting the image into $B \times B$ distinct blocks and acquiring all the blocks using an appropriately-sized measurement matrix Φ_B. Assuming x_i is the raster-scan representation of block i of the input image, the corresponding measurements are $y_i = \Phi_B x_i$ [2].

In the followings, we propose to use the Block Compressed Sensing with Smoothed Projected Landweber (BCS-SPL) algorithm introduced in [14], in order to illustrate the proposed adaptive CS reconstruction method, due to its good performance in terms of complexity and image recovery quality. Note that any other block-based CS reconstruction method can replace the BCS-SPL for the image recovery.

3. PROPOSED METHOD

Given the good performance of the Graph Based Visual Saliency (GBVS) model in [11] that aims to predict the scene locations focused by a human observer, we propose to use it in the sequel in order to build up a saliency-based compressive sensing framework. This method is adaptive to the human visual perception by applying a high sampling rate to the salient regions and a low one to the non-salient regions.

Fig. 1 presents the block diagram of the proposed saliency-based compressive sensing reconstruction framework. Firstly, it follows the basic BCS acquisition . The image is partitioned into non-overlapping blocks of $B \times B$ size; further, the blocks are sampled at the same subrate. At the receiver, an initial reconstruction of the image is obtained using the BCS-SPL algorithm and a saliency map for the whole reconstructed image is computed using the GBVS model [11].

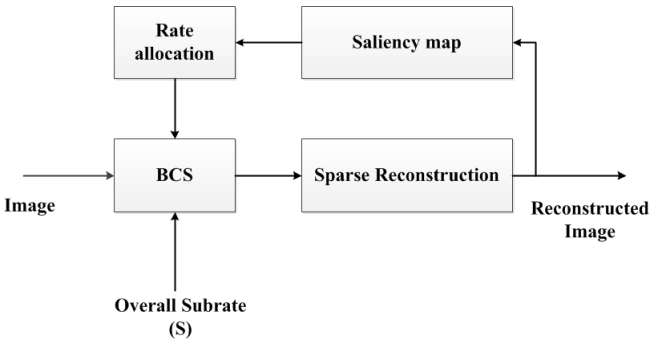


Fig. 1. Block diagram of the proposed algorithm.

On the saliency map, a binary label is assigned to each block, indicating if it belongs or not to a salient region. The new sampling subrates are computed on this binary block labeling, such that the total acquisition subrate equals the original one used in the first acquisition. The block labelings and the new sampling subrates are further sent to the sensor in order to perform a new acquisition which will serve for the final image reconstruction.

In order to obtain block labelings, first, we normalize the saliency map of the input image into $[0, 1]$. The output is a map where the intensity of each pixel represents the probability of that pixel belonging to the salient regions. After calculating the average saliency value for each block, we label each block as a salient block (labelled with one) or non-salient one (labelled with zero) to get a binary mask by using a threshold T in the range $[0, 1]$. Obviously, the percentage of blocks falling into the salient regions will be increased by decreasing the threshold. In the next section, a simple procedure for finding T is proposed. However, finding an appropriate T can be considered as a future work.

It should be noted that each block has different saliency value and definitely, assigning the sampling rate to each block according to its saliency value will improve the reconstructed quality. But, it will increase the overhead information for sending back to sensor. We just use a binary label that ensures the system does not need a high bandwidth feedback channel.

Our goal is thus to allocate different subrates for the salient and non-salient blocks, while satisfying a subrate constraint. This constraint states that the overall subrate should be equal (or slightly inferior) to the target subrate S. Assume the number of salient blocks and non-salient blocks are N_s and N_{ns}, respectively. The subrates of salient and non-salient blocks are adjusted so that the overall subrate equals with the target subrate such that:

$$N_s S_s + N_{ns} S_{ns} = (N_s + N_{ns})S, \tag{3}$$

where S_s and S_{ns} are adaptively allocated subrates for the salient and non-salient blocks, respectively. For obtaining S_s and S_{ns}, we let the salient and non-salient subrates be a fraction of a constant value U:

$$S_s = K_s U \;\text{ and }\; S_{ns} = K_{ns} U, \tag{4}$$

subject to

$$0 < K_{ns} < K_s < 1 \quad \text{and} \quad K_s + K_{ns} = 1, \tag{5}$$

where K_s and K_{ns} are predefined parameters to determine the proportion of the salient blocks and non-salient blocks in the number of measurements. From (4) and (5) , one can easily solve (3) for U:

$$U = \frac{(N_s + N_{ns})S}{K_s N_s + K_{ns} N_{ns}}. \tag{6}$$

A small K_{ns} can lead to a serious distortion of the non-salient blocks, while a selection of large K_{ns} cannot reflect the superiority of the proposed algorithm because of inconspicuous contribution for the adaptive sampling. In this paper, we consider $K_{ns} = 0.2$ and $K_s = 0.8$.

This process maybe produce $S_s > 1$ (especially in a high target subrate S.) Thus, we modify the solution for this case and set $S_s = 1$ and S_{ns} can be easily computed from (3):

$$S_{ns} = \frac{(N_s + N_{ns})S - N_s}{N_{ns}}. \tag{7}$$

For a low target subrate S, S_{ns} is very small and can result into inaccurate reconstruction of the non-salient blocks. In order to ensure that the assigned subrate of the non-salient blocks is not too small, we assign a minimum subrate S_m for S_{ns}. We set up $S_m = K \times S$, where $0 < K < 1$ is a predefined parameter to adjust the minimum subrate for the non-salient blocks. In this paper, we consider $K = 0.2$. After obtaining S_s and S_{ns} by (4), if $S_{ns} \leq S_m$, we set $S_{ns} = S_m$ and the subrate of the salient blocks is re-calculated using (3):

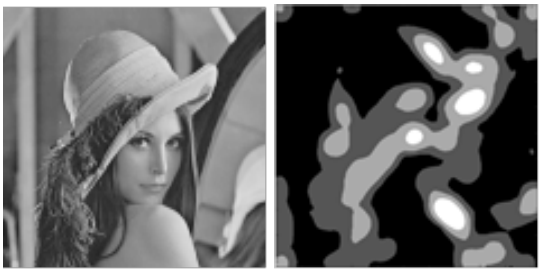
$$S_s = \frac{(N_s + (1 - K)N_{ns})S}{N_s}. \tag{8}$$

As an overall result, the sampling of block i will be carried out by

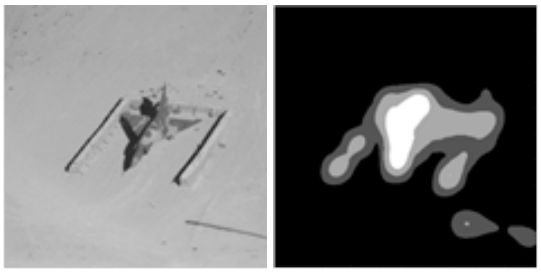
$$y_i = \begin{cases} \Phi_s x_i & x_i \in \text{Saliency} \\ \Phi_{ns} x_i & x_i \in \text{General} \end{cases} \tag{9}$$

with a smaller subrate than S for the non-salient regions, while the visually significant blocks are assigned a higher subrate than S, so that the overall subrate of image is unchanged.

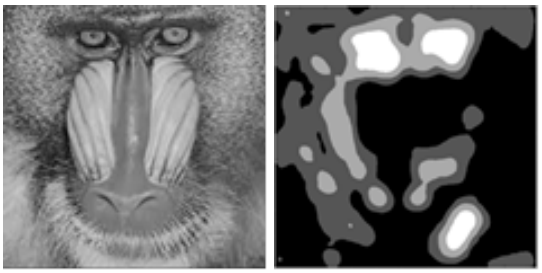
As previously mentioned, any CS reconstruction can be straightforwardly used to recover the image from these adaptive measurements. In this paper, the BCS-SPL algorithm [14] is used in order to exemplify the proposed acquisition method.



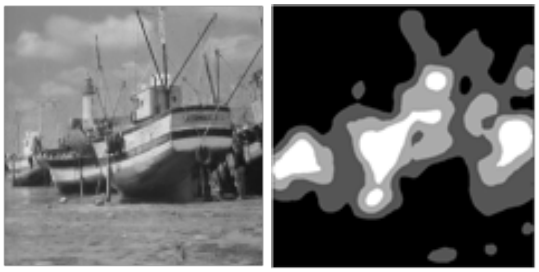
Lena



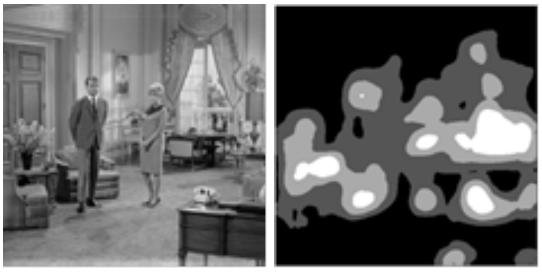
Plane



Mandrill



Boat



Living room

Initial image Saliency map

Fig. 2. Reconstructed images and saliency map.

4. EXPERIMENTAL RESULTS

In this section, we evaluate the performance of the proposed adaptive saliency-based compressive sensing image reconstruction scheme presented in Section 3 and compare it with the original BCS-SPL recovery algorithm. The performance of the proposed scheme is evaluated on 5 popular 8 bits per pixel, grayscale test images (*Lena*, *Plane*, *Mandrill*, *Boat* and *Living room*) of 512×512 pixels resolution. All results are averaged over 5 independent trials, since the reconstruction performance varies due to the randomness of the sampling matrix. We make the assessment in terms of quality of the reconstructed image, i.e. PSNR (Peak Signal-to-Noise Ratio), and computational burden.

For the BCS sampling, we use block dimension of size $B = 32$, as proposed in [14]. The target subrate S is fixed for two algorithms (BCS-SPL and proposed scheme) and the corresponding subrates for the salient and non-salient blocks (i.e. S_s and S_{ns}) are determined by (4). For the CS reconstruction, we adopt the well-known SPL algorithm due to its high performance and low computational burden [14]. The CS recovery is coupled with the Double Discrete Wavelet Transform (DDWT) as the sparsity basis.

After a fixed-rate block compressive sampling, an initial image is reconstructed using the BCS-SPL algorithm; based on this image, the salient and non-salient regions are computed using the GBVS [11]. The first and second columns in Fig. 2 show the initial reconstructed images at the target subrate $S = 0.3$ and their corresponding saliency maps, respectively (the saliency maps were quantized for better display.) In the saliency maps, brighter regions represent the salient locations on which a human observer pays more attention to, while the darker areas represent the less-saliency regions.

To find an appropriate threshold T to identify the salient blocks and compare the quality of the reconstructed image, we proceed empirically and vary T from 0 to 1 over several test images and compute the corresponding PSNR value at each threshold. The resulting PSNR values versus T for the *Lena* image at the target subrate $S = 0.3$ is shown in Fig. 3. This curve provides a reliable comparison of how proportion of the salient regions has an effect on the reconstructed image. It is important to note that the percentage of the salient regions decreases when the threshold value increases, however, the PSNR does not change proportionally with. Over several tests, we found out that the appropriate value for the threshold T with a good recovery precision can be determined by:

$$T = \frac{1}{2(R \times C)} \sum_{x=0}^{R-1} \sum_{y=0}^{C-1} H(x, y), \tag{10}$$

where R and C are the width and height of the normalized saliency map H, respectively, and $H(x, y)$ is the saliency value of the pixel at the position (x, y). As can be observed in Fig. 3, for the *Lena* image, the obtained threshold is $T = 0.13$

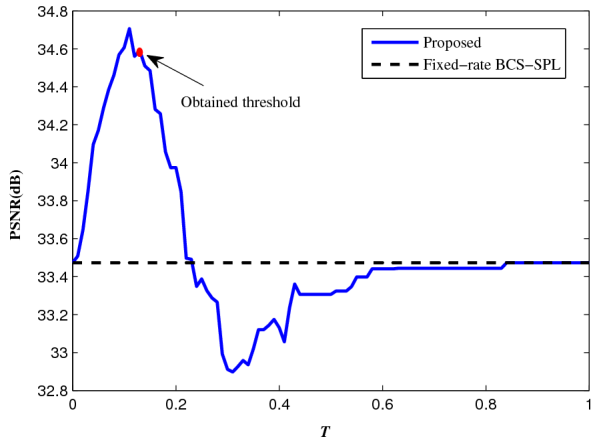


Fig. 3. Comparison of the effect of threshold value on PSNR value of the reconstructed image for the Lena image at the overall subrate $S = 0.3$.

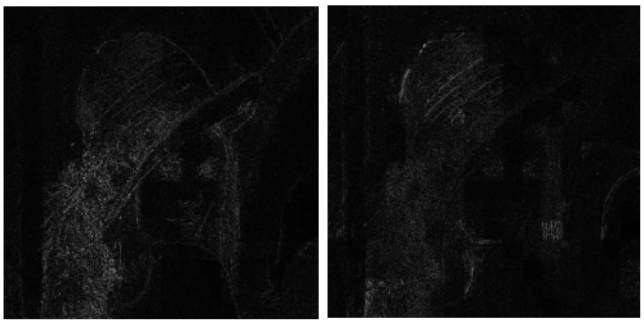
Fixed-rate BCS-SPL Adaptive Proposed method

Fig. 4. Zoomed portions extracted from the reconstructed image of Lena at the overall subrate $S = 0.3$.

which is near to the real maximum of the curve. At this case, the percentage of salient regions is %53.9.

According to the saliency map in Fig. 2, relatively high sampling rates are assigned to the blocks in the salient regions and low sampling rates are assigned to the rest of blocks. For instance, for the *Lena* image at the target subrate S = 0.3, the found subrates for the salient and non-salient areas are $S_s = 0.46$ and $S_{ns} = 0.11$, respectively.

The superiority of the adaptive algorithm is highlighted in Fig. 4 and Fig. 5 for the *Lena* image at target subrate $S = 0.3$. Fig. 4 shows the zoomed portions that are extracted from the images generated by the fixed-rate BCS-SPL algorithm and adaptive proposed method, respectively. Fig. 4 validates the performance of the adaptive compressive sensing algorithm based on the visual attention. As can be seen, the fine details in the salient regions become much clearer compared with the BCS-SPL, especially in some complex regions which contain high frequencies (strong hairs). Fig. 5 shows the reconstruction error images, obtained for the BCS-SPL and proposed algorithms with respect to the original image at the target subrate $S = 0.3$. We can generally infer that the error in the reconstructed image by the proposed method in the salient regions is lower in comparison with the BCS-SPL algorithm.



Fixed-rate BCS-SPL Adaptive Proposed method

Fig. 5. Error map between the reconstructed and original image of Lena at the overall subrate $S = 0.3$.

Table 1. Image reconstruction PSNR in dBs

		Target Subrate (S)				
		0.1	0.2	0.3	0.4	0.5
Lena	BCS-SPL	28.02	31.41	33.48	35.15	36.73
	Proposed	28.63	32.26	34.56	36.42	37.84
	S_s	0.15	0.30	0.45	0.61	0.77
	S_{ns}	0.04	0.07	0.11	0.15	0.17
Plane	BCS-SPL	31.68	35.14	37.19	38.86	40.41
	Proposed	34.76	37.79	39.62	40.94	42.38
	S_s	0.23	0.47	0.71	0.93	1
	S_{ns}	0.05	011	0.17	0.23	0.34
Mandrill	BCS-SPL	20.65	21.85	22.91	23.97	25.09
	Proposed	20.73	22.17	23.23	24.17	25.44
	S_s	0.18	0.34	0.52	0.67	0.78
	S_{ns}	0.04	0.08	0.13	0.17	0.20
Boat	BCS-SPL	25.04	27.72	29.49	31.08	32.57
	Proposed	26.05	28.72	30.46	31.96	33.30
	S_s	0.16	0.30	0.45	0.60	0.76
	S_{ns}	0.04	0.07	0.11	0.15	0.19
Living room	BCS-SPL	24.78	27.03	28.66	30.13	31.61
	Proposed	25.24	27.74	29.36	30.88	32.38
	S_s	0.14	0.28	0.42	0.58	0.70
	S_{ns}	0.03	0.07	0.10	0.14	0.18

To better illustrate the improvement, an objective comparison is provided in Table 1 at different target subrates for the fixed-rate BCS-SPL and the proposed algorithm. Additionally, Table 1 shows the allocated subrates for the salient region and non-salient regions (S_s and S_{ns}). Clearly, the proposed method gains remarkably compared with the case of fixed subrate BCS. It can be observed that for the highly textured images, like *Mandrill*, the PSNR improvement is between 0.1 dBs to 0.35 dBs. For smoother images, like Plane, the improvement is up to 3 dBs.It can be concluded that if the salient parts are extracted accurately, better performance of the CS reconstruction can be realized both in the subjec-

Table 2. Comparison results for Lena in term of PSNR (dBs)

	Target Subrate (S)				
Algorithm	0.1	0.2	0.3	0.4	0.5
BCS-SPL	28.02	31.41	33.48	35.15	36.73
[9]	28.11	31.61	33.94	35.96	37.35
Proposed	28.63	32.26	34.56	36.42	37.84

Table 3. Execution time (Sec.) For Lena at $S = 0.3$

Algorithm	Time
BCS-SPL	21.63
Proposed	62.34

tive/visual and objective tests.

In Table 2, the performance of he the proposed method is also compared with the adaptive compressive sensing technique introduced in [9]. Both algorithms use the same reconstruction algorithm (SPL). As can be seen, the proposed algorithm outperforms the algorithm proposed in [9] in term of PSNR for the *Lena* image.

While adaptive sensing results in the improved performance, we mention a few limitations. In Table 3, we compare the execution time for each algorithm on the *Lena* image at the target subrate $S = 0.3$. The experiments are conducted using MATLAB (R2014a) on a PC with an Intel(R) Xeon(R) processor at 3.20GHz and 8-GB of RAM at 64-bits operating system of WINDOWS 7. Our proposed method is slower than the BCS-SPL, however, because of the block-based sensing, it is plausible to the extent that a parallel implementation of the proposed algorithm can be trivially accomplished in hardware, which is suitable for the real-time applications. In addition, Although the proposed method has been employed in single-image acquisition, one can imagine the application of the proposed method for pedestrian tracking in the video surveillance applications and the sampling of low-to-normal dynamics video sequences, where only the first frame within a group of pictures (GOP) is used for the saliency computation and rate allocation. Given the high resemblance among the frames within the same GOP, adaptive sensing uses this saliency information to guide the sampling of the next frames in order to optimize the gain of new information.

5. CONCLUSION

Inspired by the saliency based model of visual attention, a novel adaptive compressive sampling scheme has been proposed in order to enhance the reconstruction performance of the block-based compressive sensing. At the first step, an initial recovery is obtained using fixed subrate sampling. Further, based on a binary saliency map obtained with a graph-based algorithm, adaptive compressive sampling is performed, such that more visual-significant details can be captured in the salient regions. This way, the proposed method enhances the recovery quality by finding an optimal subrate trade-off between the salient and non-salient areas. The experimental results show a remarkable reconstruction quality improvement of the proposed method compared with the existing algorithms which adopt a fixed-rate sampling.

6. REFERENCES

[1] E. J. Candes and M. B. Wakin, "An introduction to compressive sampling," *IEEE Signal Processing Mag.*, vol. 25, pp. 21–30, March 2008.

[2] L. Gan, "Block compressed sensing of natural images," in *Proc. Int. Conf. Digital Signal Processing.*, Cardiff, UK, July 2007, pp. 403–406.

[3] A. Borji, D. N. Sihite, and L. Itti, "Salient object detection: A benchmark," in *Proc. 12th Eur. Conf. Computer Vision*, Italy, Oct. 2012, pp. 414–429.

[4] M. L. Malloy and R. D. Nowak, "Near-optimal adaptive compressed sensing," *IEEE Trans. Information Theory*, vol. 60, no. 7, pp. 4001–4012, 2014.

[5] I. Noor and E. L. Jacobs, "Adaptive compressive sensing algorithm for video acquisition using single pixel camera," *SPIE J. Electronic Imaging*, vol. 22, no. 2, pp. 021013–021013, July 2013.

[6] W. Guicquero, A. Verdant, A. Dupret, and P. Vandergheynst, "Nonuniform sampling with adaptive expectancy based on local variance," in *Proc. Int. Conf. Sampling Theory and Applications (SampTA)*, Washington, DC, May 2015, pp. 254–258.

[7] W. Guicquero, A. Dupret, and P. Vandergheynst, "An adaptive compressive sensing with side information," in *Asilomar Conf. Signals, Systems and Computers*, Pacific Grove, CA, Nov 2013, pp. 138–142.

[8] D.M. Malioutov, S.R. Sanghavi, and A.S. Willsky, "Sequential compressed sensing," *IEEE J. Sel. Topics in Signal Processing*, vol. 4, no. 2, pp. 435–444, April 2010.

[9] J. Chen, X. Zhang, and H. Meng, "Self-adaptive sampling rate assignment and image reconstruction via combination of structured sparsity and non-local total variation priors," *J. Digital Signal Processing*, vol. 29, pp. 54–66, June 2014.

[10] G. Warnell, S. Bhattacharya, R. Chellappa, and T. Basar, "Adaptive-rate compressive sensing using side information," *IEEE Trans. Image Processing*, vol. 24, no. 11, pp. 3846–3857, Nov 2015.

[11] J. Harel, C. Koch, and P. Perona, "Graph-based visual saliency," in *Proc. Neural Information Processing Systems*, 2006, pp. 545–552.

[12] D.L. Donoho, "Compressed sensing," *IEEE Trans. Information Theory*, vol. 52, no. 4, pp. 1289–1306, April 2006.

[13] S. S. Chen, D. L. Donoho, and M. A. Saunders, "Atomic decomposition by basis pursuit," *SIAM J. Scientific Computing*, vol. 20, no. 1, pp. 33–61, Dec. 1998.

[14] S. Mun and J.E. Fowler, "Block compressed sensing of images using directional transforms," in *Proc. IEEE Int. Conf. Image Processing (ICIP)*, Nov 2009, pp. 3021–3024.