Proceedings Paper:

Kuzin, D., Isupova, O. and Mihaylova, L. (2015) Compressive Sensing Approaches for Autonomous Object Detection in Video Sequences. In: Sensor Data Fusion: Trends, Solutions, Applications (SDF), 2015. Sensor Data Fusion: Trends, Solutions, Applications (SDF), 2015, 06-08 Oct 2015, Bonn, Germany. IEEE .

https://doi.org/10.1109/SDF.2015.7347706

Compressive Sensing Approaches for Autonomous Object Detection in Video Sequences

Danil Kuzin
The University of Sheffield
Sheffield, UK
Email: dkuzin1@sheffield.ac.uk

Olga Isupova
The University of Sheffield
Sheffield, UK
Email: o.isupova@sheffield.ac.uk

Lyudmila Mihaylova
The University of Sheffield
Sheffield, UK
Email: l.s.mihaylova@sheffield.ac.uk

***Abstract*—Video analytics requires operating with large amounts of data. Compressive sensing allows to reduce the number of measurements required to represent the video using the prior knowledge of sparsity of the original signal, but it imposes certain conditions on the design matrix. The Bayesian compressive sensing approach relaxes the limitations of the conventional approach using the probabilistic reasoning and allows to include different prior knowledge about the signal structure. This paper presents two Bayesian compressive sensing methods for autonomous object detection in a video sequence from a static camera. Their performance is compared on real datasets with the non-Bayesian greedy algorithm. It is shown that the Bayesian methods can provide more effective results than the greedy algorithm in terms of both accuracy and computational time.**

I. Introduction

The significant developments in the field of sparse methods during the last decades lead to the new research and application fields. One of the first applications of sparse modelling is the linear regression problem where l_0 and l_1-norm regularisation is considered. The latter has the advantage that a norm term is convex, while it has not so obvious sparse interpretation [1].

Sparse modelling is further developed in the field of signal processing in compressive sensing [2], where the main idea is to extract only the meaningful information from the measurements. Two main problems are considered in compressive sensing: selecting the optimal design matrix and solving ill-posed regression, that arises in the original signal decoding from the measurements [3].

In Bayesian modelling sparseness of the data can be achieved by imposing special sparse prior distributions [4]. For example, in [5] the authors propose the Laplace prior on the data. The full inference to this model is provided in [6], using the Expectation Propagation (EP) technique. Another work is [7], where the prior is modified to the hierarchical Gaussian-Gamma distribution. These models are used as a basis for Bayesian compressive sensing in [8] and [9].

The overview of recently developed sparse algorithms and models for image and video processing is presented in [10]. One of the essential problems in video processing is autonomous object detection which is mostly solved by background subtraction. Background subtraction aims to distinguish the foreground (moving objects) from the background (static ones). Sparseness is natural for the background subtraction problem as the foreground objects occupy the small regions on a frame. Background subtraction hence represents a natural application area for sparse modelling.

The idea to apply compressive sensing for background subtraction is proposed, for example, in [11] and developed in [12]. In contrast to these works in this paper we focus on the sparse Bayesian methods for background subtraction and the comprehensive comparison of these methods with the conventional compressive sensing one.

The contribution of this paper is the development of a Bayesian compressive sensing algorithm for the background subtraction problem. In Bayesian compressive sensing the estimated sparse coefficients are random variables with some prior distribution. Also this paper presents the comparison of several algorithms to evaluate their applicability in different situations.

This paper is organised as follows. In Section II the proposed approach is described. The experimental results are presented in Section III. Section IV concludes the paper and discusses the future work.

II. Framework

Assume that we have a static camera. The frame $\mathbf{B} \in \mathbb{R}^{n_1 \times n_2}$ at a time instant 0 is considered as a background frame. The video from the camera consists of the sequential frames $\mathbf{V}_k \in \mathbb{R}^{n_1 \times n_2}$, $k \in \{1, \dots, K\}$. The aim is to detect the moving objects in these frames.

A. Video preprocessing

The camera provides video frames in the Red-Green-Blue (RGB) format and they are next converted to the greyscale format. For the purposes of the compressive sensing approach the frames are converted to vectors. Thus, the background frame $\mathbf{B}$ is converted to a vector $\mathbf{b} \in \mathbb{R}^n$, the video frames $\mathbf{V}_k$ are converted to vectors $\mathbf{v}_k \in \mathbb{R}^n$, where $n = n_1 n_2$.

B. Compressive sensing

The non-zero elements of the difference $\mathbf{f}_k = \mathbf{v}_k - \mathbf{b}$ are supposed to correspond to the moving objects. As the foreground objects take only a part of the image the vector $\mathbf{f}_k$ has many values that are close to zero:

$$\|\mathbf{f}_k\|_{l_0} \leq s \ll n, \tag{1}$$

l_0-pseudonorm is the number of non-zero elements of a vector.

Within the compressive sensing approach the number of measurements that are need to be taken can be reduced [2] and the image quality may be improved [10]. The values of the vector $\mathbf{f}_k$ are calculated based on the set of the compressed measurements $\mathbf{g}_k \in \mathbb{R}^s$:

$$\mathbf{g}_k = \mathbf{\Phi} \mathbf{f}_k, \quad (2)$$

where the design matrix $\mathbf{\Phi} \in \mathbb{R}^{s \times n}$ consists of i.i.d Gaussian variables. It is selected according to the method proposed in [13].

Since $\mathbf{f}_k = \mathbf{v}_k - \mathbf{b}$, the computation of the coefficients $\mathbf{g}_k$ can be done on the acquisition step as

$$\mathbf{g}_k = \mathbf{\Phi} \mathbf{f}_k = \mathbf{\Phi} \mathbf{v}_k - \mathbf{\Phi} \mathbf{b}. \quad (3)$$

The vectors $\mathbf{\Phi b}$ and $\mathbf{\Phi v}_k$ are linear combinations of the pixels of the video frames. Therefore, a single pixel camera may be used. It means that the acquisition and background subtraction steps use the signal with the length s rather than n. The reconstruction step is used to restore the signal with the length n.

The linear system (3) is underdetermined when $n > s$ and therefore an infinite number of solutions exists. The problem can be determined by introducing a sparse structure in the $\mathbf{f}_k$ signal. This can be done by l_p norm minimisation, where $p < 2$ typically.

The sparse methods for this kind of problems are [14, Chapter 13]:

- l_0 - *minimisation.* The greedy algorithms based on least squares estimates, stochastic search, variational inference;
- l_1 - *minimisation.* The coordinate descent, LARS, the proximal and gradient projection methods;
- *Non-convex minimisation.* The bridge regression, the hierarchical adaptive lasso.

In this paper we focus on the Bayesian compressive sensing methods [8], [15] and compare them with orthogonal matching pursuit (OMP) [16], that is a greedy algorithm for l_0-minimisation.

1) Bayesian compressive sensing (BCS): The system (3) is reformulated as a linear regression model in [8]:

$$\mathbf{g}_k = \mathbf{\Phi} \mathbf{f}_k + \boldsymbol{\xi}, \quad (4)$$

where $\boldsymbol{\xi}$ is a vector which elements are the independent noise from the Gaussian distribution: $\xi_i \sim \mathcal{N}(\xi_i; 0, \beta^{-1})$. Therefore, the likelihood can be expressed as

$$p(\mathbf{g}_k | \mathbf{f}_k, \beta) = \prod_{i=1}^{n} \mathcal{N}(g_{i,k}; \mathbf{\Phi}_i \mathbf{f}_k, \beta^{-1}), \quad (5)$$

where $g_{i,k}$ is the i-th element of the vector $\mathbf{g}_k$, $\mathbf{\Phi}_i$ – the i-th row of the matrix $\mathbf{\Phi}$.

To implement the full Bayesian approach, the prior distributions are imposed on all the parameters:

$$p(\mathbf{f}_k | \boldsymbol{\alpha}) = \prod_{i=1}^{n} \mathcal{N}(f_{i,k}; 0, \alpha_i^{-1}), \quad (6)$$

where $f_{i,k}$ is the i-th element of the vector $\mathbf{f}_k$, $\boldsymbol{\alpha}$ is a prior parameter vector, α_i is the i-th element of the vector $\boldsymbol{\alpha}$;

$$p(\boldsymbol{\alpha}) = \prod_{i=1}^{n} \Gamma(\alpha_i; a, b), \quad (7)$$

$$p(\beta) = \Gamma(\beta; c, d), \quad (8)$$

where $\Gamma(\cdot)$ denotes the Gamma distribution. The values of the hyperparameters a, b, c, d are set uniform and close to zero.

According to the Bayes rule the posterior distribution can be written as follows:

$$p(\mathbf{f}_k, \boldsymbol{\alpha}, \beta | \mathbf{g}_k) = \frac{p(\mathbf{g}_k | \mathbf{f}_k, \boldsymbol{\alpha}, \beta) p(\mathbf{f}_k, \boldsymbol{\alpha}, \beta)}{p(\mathbf{g}_k)}, \quad (9)$$

where $p(\mathbf{g}_k | \mathbf{f}_k, \boldsymbol{\alpha}, \beta)$ is the likelihood term, $p(\mathbf{f}_k, \boldsymbol{\alpha}, \beta)$ is the prior term, $p(\mathbf{g}_k)$ is the evidence term. The latter can be expressed as:

$$p(\mathbf{g}_k) = \int\limits_{\mathbf{f}_k, \boldsymbol{\alpha}, \beta} p(\mathbf{g}_k | \mathbf{f}_k, \boldsymbol{\alpha}, \beta)\, p(\mathbf{f}_k, \boldsymbol{\alpha}, \beta)\, d\mathbf{f}_k\, d\boldsymbol{\alpha}\, d\beta. \quad (10)$$

This integral is intractable, therefore some kind of approximation should be used.

In Bayesian compressive sensing [8] the decomposition of the posterior probability into the product of the tractable and intractable probabilities is used and the intractable one is approximated with the delta-function in its mode:

$$p(\mathbf{f}_k, \boldsymbol{\alpha}, \beta | \mathbf{g}_k) = p(\mathbf{f}_k | \mathbf{g}_k, \boldsymbol{\alpha}, \beta) p(\boldsymbol{\alpha}, \beta | \mathbf{g}_k). \quad (11)$$

The Bayes rule for the first term of (11) is as follows:

$$p(\mathbf{f}_k | \mathbf{g}_k, \boldsymbol{\alpha}, \beta) = \frac{p(\mathbf{g}_k | \mathbf{f}_k, \beta) p(\mathbf{f}_k | \boldsymbol{\alpha})}{p(\mathbf{g}_k | \boldsymbol{\alpha}, \beta)}. \quad (12)$$

These are all the Gaussians, so the probability $p(\mathbf{f}_k | \boldsymbol{\alpha}, \beta, \mathbf{g}_k)$ can be calculated straightforwardly. It is the Gaussian distribution with the parameters

$$\mathbf{\Sigma} = (\beta \mathbf{\Phi}^\top \mathbf{\Phi} + \mathbf{A})^{-1}, \quad (13)$$

$$\boldsymbol{\mu} = \beta \mathbf{\Sigma} \mathbf{\Phi}^\top \mathbf{g}_k, \quad (14)$$

where $\mathbf{A} = \mathrm{diag}(\alpha_1, \ldots, \alpha_n)$.

The second term of the posterior probability (11) can be expressed as:

$$p(\boldsymbol{\alpha}, \beta | \mathbf{g}_k) = \frac{p(\mathbf{g}_k | \boldsymbol{\alpha}, \beta) p(\boldsymbol{\alpha}) p(\beta)}{p(\mathbf{g}_k)}. \quad (15)$$

The denominator in (15) is not tractable. The values of $\boldsymbol{\alpha}, \beta$ which maximise (15) are used. The hyperpriors are uniform, therefore only the term $p(\mathbf{g}_k | \boldsymbol{\alpha}, \beta)$ needs to be maximised:

$$p(\mathbf{g}_k | \boldsymbol{\alpha}, \beta) = \int p(\mathbf{g}_k | \mathbf{f}_k, \beta) p(\mathbf{f}_k | \boldsymbol{\alpha}) d\mathbf{f}_k. \quad (16)$$

Maximisation of (16) w.r.t. $\boldsymbol{\alpha}, \beta$ gives the following iterative process:

$$\alpha_i^{new} = \frac{\gamma_i}{\mu_i^2}, \quad (17)$$

$$(\beta^{-1})^{new} = \frac{\|\mathbf{g}_k - \mathbf{\Phi} \mu\|_{l_2}^2}{s - \Sigma_{ii} \gamma_i}, \quad (18)$$

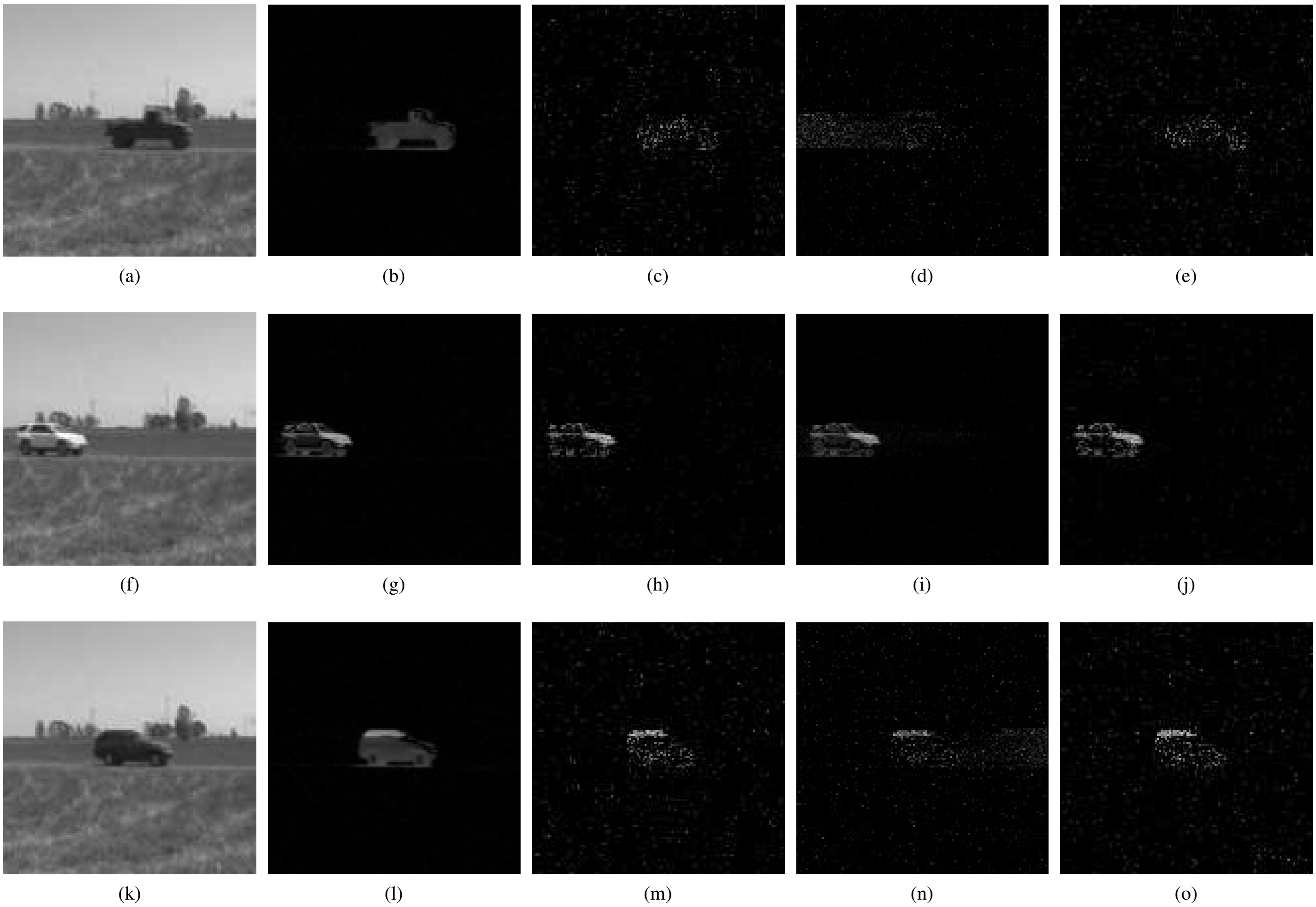

Figure 1. Comparison of the foreground restoration based on 2000 measurements by the algorithms. The three rows correspond to the three sample frames. From left to right columns: the input uncompressed frame, uncompressed background subtraction, compressed background subtraction with Bayesian compressive sensing, compressed background subtraction with multi-task Bayesian compressive sensing and compressed background subtraction with orthogonal matching pursuit

where $\gamma_i = 1 - \alpha_i \Sigma_{ii}$.

This process together with (13) – (14) converges to the optimal estimates.

Note that

$$p(f_{i,k}) = \frac{b^a \Gamma\left(a + \frac{1}{2}\right)}{(2\pi)^{\frac{1}{2}} \Gamma(a)} \left(b + \frac{f_{i,k}^2}{2}\right)^{-\left(a+\frac{1}{2}\right)}. \qquad (19)$$

This is the Student-t distribution that has the most probable area concentrated around zero. Thereby it leads to the sparse vector $\mathbf{f}_k$.

2) Multitask Bayesian compressive sensing (Multitask BCS): In [15] the Bayesian method to process several signals that have a similar sparse structure is proposed. The multitask setting reduces the number of measurements that should be taken compared to processing all the signals independently. The hyperparameter $\boldsymbol{\alpha}$ is considered to be shared by all the tasks.

3) Matching Pursuit: The greedy algorithms are proposed for the l_0 minimisation in [16]. These methods start with a null vector and iteratively add non-zeros values to it until a convergence to a threshold.

III. Experiments

We use the Convoy dataset [12], which consists of 260 greyscale frames and the background frame. The frames are scaled to the less resolution of 128×128 to avoid memory problems. For the multitask algorithm the batches of 40 frames are run together, while for the Bayesian compressive sensing and OMP algorithms all the frames are processed independently. There are two sets of the experiments: one with $s = 2000$ measurements and the other with $s = 5000$ measurements. For both sets of the experiments all three methods are run for 10 times with 10 different design matrices $\mathbf{\Phi}$ shared among the methods. For the quantitative comparison the median values of quality measures among these runs are presented.

The qualitative comparison of the methods with the same design matrix $\mathbf{\Phi}$ is displayed in Figures 1 – 2. The three

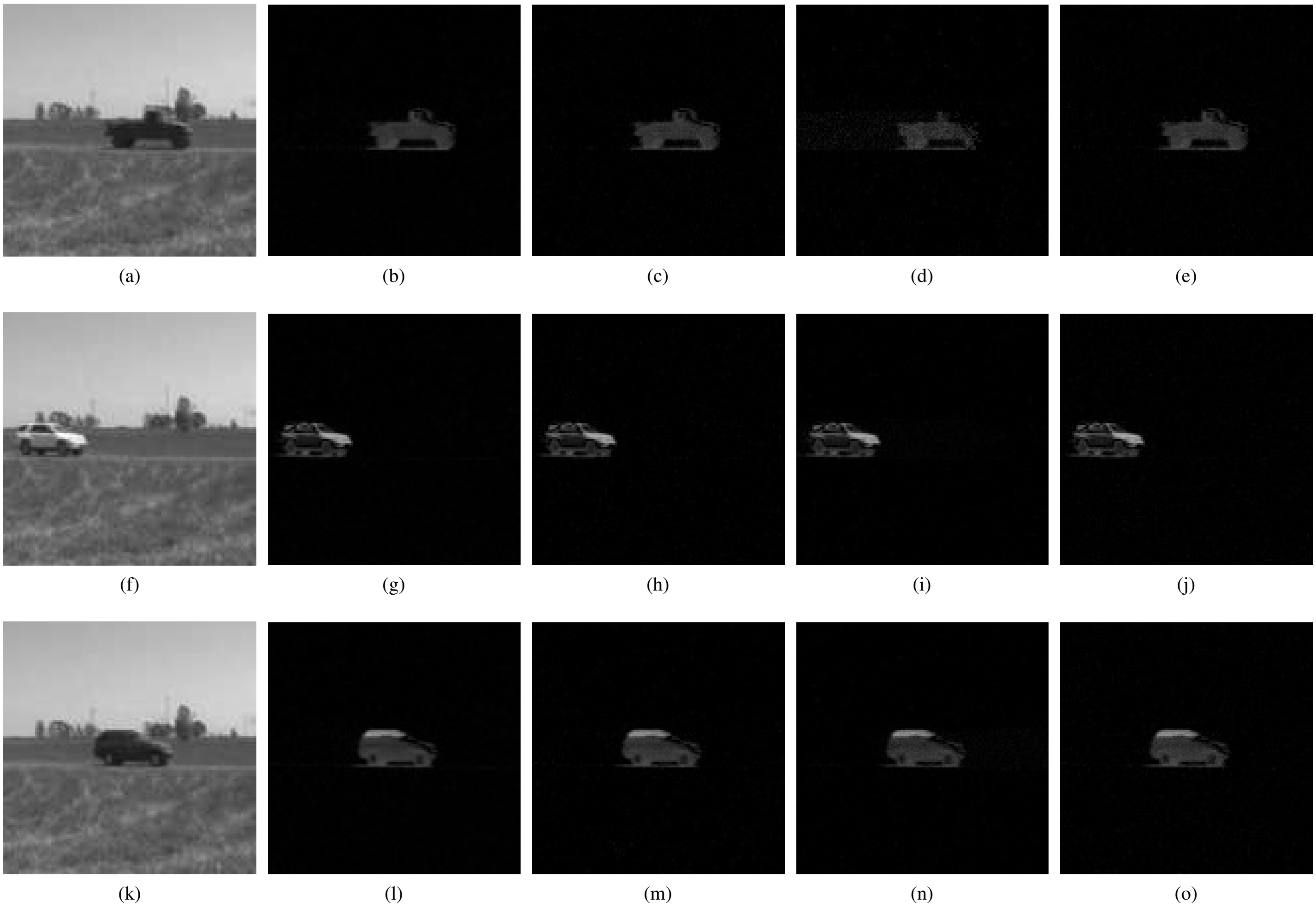

Figure 2. Comparison of the foreground restoration based on 5000 measurements by the algorithms. The three rows correspond to the three sample frames. From left to right columns: the input uncompressed frame, uncompressed background subtraction, compressed background subtraction with Bayesian compressive sensing, compressed background subtraction with multi-task Bayesian compressive sensing and compressed background subtraction with orthogonal matching pursuit

demonstrative frames are presented. One can notice that with the same design matrix the models demonstrate similar results. The figures show that 2000 measurements can be used for object region detection, while 5000 measurements which is only about 30% of the input resolution are enough even to distinguish parts of the objects like doors and windows of the cars.

For the quantitative comparison of the results the following measures are used:

- *Reconstruction error*: $\frac{\|\mathbf{f}-\hat{\mathbf{f}}\|_{l_2}}{\|\mathbf{f}\|_{l_2}}$, where $\mathbf{f}$ is the signal ground truth, $\hat{\mathbf{f}}$ is the signal, reconstructed by the algorithm;
- *Background subtraction quality measure (BS quality)*: $\frac{|S(\mathbf{f}) \cap S(\hat{\mathbf{f}})|}{|S(\mathbf{f}) \cup S(\hat{\mathbf{f}})|}$, where $S(\mathbf{f})$ is the ground truth foreground area, $S(\hat{\mathbf{f}})$ is the algorithm detected foreground area, $|\cdot|$ is the cardinality of the set;
- *Peak signal-to-noise ratio (PSNR)*: $10\log_{10}\left(\frac{\text{peakval}^2}{\text{MSE}}\right)$, where peakval is the maximum possible pixel value, that is 255 in our case. MSE is the mean square error between $\mathbf{f}$ and $\hat{\mathbf{f}}$;
- *Structural similarity index (SSIM)* [17]: $\frac{(2\mu_{\mathbf{f}}\mu_{\hat{\mathbf{f}}}+C_1)(2\sigma_{\mathbf{f}\hat{\mathbf{f}}}+C_2)}{(\mu_{\mathbf{f}}^2+\mu_{\hat{\mathbf{f}}}^2+C_1)(\sigma_{\mathbf{f}}^2+\sigma_{\hat{\mathbf{f}}}^2+C_2)}$, where $\mu_{\mathbf{f}}$, $\mu_{\hat{\mathbf{f}}}$, $\sigma_{\mathbf{f}}$, $\sigma_{\hat{\mathbf{f}}}$, $\sigma_{\mathbf{f}\hat{\mathbf{f}}}$ are the local means, standard deviations, and cross-covariance for the images $\mathbf{f}$, $\hat{\mathbf{f}}$ respectively, and C_1, C_2 are the regularisation constants.

The difference between the uncompressed current frame $\mathbf{v}_k$ and the uncompressed background frame $\mathbf{b}$ is used as the ground truth signal $\mathbf{f}$ for every frame (the second columns in Figures 1 – 2), since this is the signal which is compressed by (3).

The results are presented in Figure 3. All the quality measures – reconstruction error, BS quality, PSNR and SSIM – are calculated for every frame. The mean values among the frames for each measure can be found in Tables I – II.

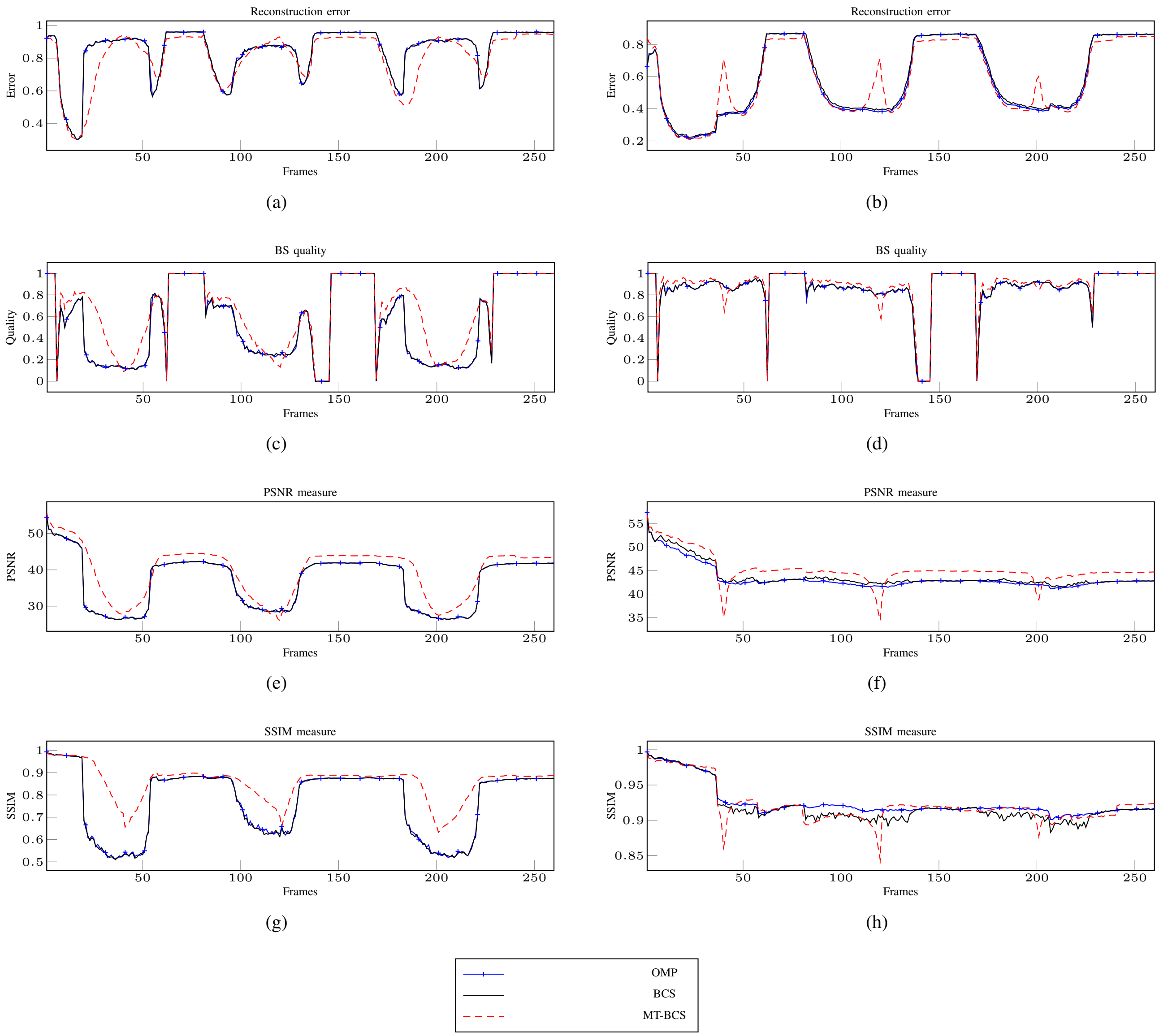

Figure 3. Quantitative method comparison on the frame level. The left column corresponds to the set of the experiments with 2000 measurements, the right column corresponds to the set of the experiments with 5000 measurements. From top to bottom rows: the reconstruction error measure (values close to 1 refer to the frames without any foreground objects), the background subtraction quality measure, the PSNR measure and the SSIM measure.

Table I
METHOD COMPARISON BASED ON 2000 MEASUREMENTS.

Algorithm	Mean frame recon-struction error	Mean frame BS quality	Mean frame PSNR	Mean frame SSIM	Mean compu-tational time (hours)[1]
BCS	0.8037	0.3518	34.2007	0.7198	0.23
Multitask BCS	0.7608	0.4820	37.542	0.8384	0.67
OMP	0.8028	0.3510	34.1705	0.7204	0.51

Table II
METHOD COMPARISON BASED ON 5000 MEASUREMENTS.

Algorithm	Mean frame recon-struction error	Mean frame BS quality	Mean frame PSNR	Mean frame SSIM	Mean compu-tational time (hours)[1]
BCS	0.4713	0.8119	43.8251	0.9186	0.9
Multitask BCS	0.4702	0.8421	45.0028	0.9212	8.5
OMP	0.4578	0.8109	43.2720	0.9266	4.8

[1]The computational time is provided for a batch of 40 frames (BCS and OMP process each frame independently with 4 parallel workers, multitask BCS processes all 40 frames together). Implementation is made on the laptop with i7-4702HQ CPU with 2.20GHz, 16 GB RAM using MATLAB 2015a.

The multitask Bayesian compressive sensing algorithm demonstrates the best results according to almost each measure. The Bayesian compressive sensing and OMP algorithms show the competitive results but the Bayesian compressive sensing algorithm works faster. It is worth while to note that the multitask Bayesian compressive sensing algorithm has the biggest variance amongst the runs with the different design matrices, while the variances of the Bayesian compressive sensing and OMP runs for the same matrices are quite small.

IV. Conclusions and future work

This work presents two Bayesian compressive sensing algorithms for autonomous object detection in video sequences. These are the conventional Bayesian compressive sensing and the multitask Bayesian compressive sensing algorithms. The results presented in Figures 1 – 2 demonstrate the appropriate reconstruction quality of the original image based on only 5000 measurements (that is $\approx 30\%$ of the original image size).

The conventional Bayesian compressive sensing method demonstrates similar results to the greedy OMP algorithm but the former is more effective in terms of the computational time. If the computational time is not critical the extension of the Bayesian method designed for a multitask problem can improve the performance in terms of the different measures. Therefore, other extensions of the Bayesian method to include the prior information need further research.

Future work will be focused on different sparse Bayesian methods, such as the EP-based framework with the Laplace prior proposed in [6] and the Markov Chain Monte Carlo (MCMC) framework proposed in [18]. Future work will also consider a correlated sparse structure of the data.

V. Acknowledgements

The authors Olga Isupova and Lyudmila Mihaylova are grateful for the support provided by the EC Seventh Framework Programme [FP7 2013-2017] TRAcking in compleX sensor systems (TRAX) Grant agreement no.: 607400. Lyudmila Mihaylova acknowledges also the support from the UK Engineering and Physical Sciences Research Council (EPSRC) via the Bayesian Tracking and Reasoning over Time (BTaRoT) grant EP/K021516/1.

References

[1] F. Bach, R. Jenatton, J. Mairal, and G. Obozinski, "Optimization with sparsity-inducing penalties," *Found. Trends Mach. Learn.*, vol. 4, no. 1, pp. 1–106, Jan. 2012.

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

[3] A. Y. Carmi, L. S. Mihaylova, and S. J. Godsill, "Introduction to compressed sensing and sparse filtering," in *Compressed Sensing and Sparse Filtering*, ser. Signals and Communication Technology, A. Y. Carmi, L. Mihaylova, and S. J. Godsill, Eds. Springer Berlin Heidelberg, 2014, pp. 1–23.

[4] N. G. Polson and J. G. Scott, "Shrink globally, act locally: Sparse Bayesian regularization and prediction," *Bayesian Statistics*, vol. 9, pp. 501–538, 2010.

[5] R. Tibshirani, "Regression shrinkage and selection via the lasso," *Journal of the Royal Statistical Society. Series B (Methodological)*, vol. 58, no. 1, pp. 267–288, 1996.

[6] M. Seeger, "Bayesian inference and optimal design in the sparse linear model," *Journal of Machine Learning Research*, vol. 9, pp. 759–813, 2008.

[7] M. E. Tipping, "Sparse Bayesian learning and the relevance vector machine," *The journal of machine learning research*, vol. 1, pp. 211–244, 2001.

[8] S. Ji, Y. Xue, and L. Carin, "Bayesian compressive sensing," *IEEE Transactions on Signal Processing*, vol. 56, no. 6, pp. 2346–2356, June 2008.

[9] M. W. Seeger and H. Nickisch, "Compressed sensing and Bayesian experimental design," in *Proceedings of the 25th International Conference on Machine Learning*, ser. ICML '08. New York, NY, USA: ACM, 2008, pp. 912–919.

[10] J. Mairal, F. R. Bach, and J. Ponce, "Sparse modeling for image and vision processing," *CoRR*, vol. abs/1411.3230, 2014.

[11] V. Cevher, A. Sankaranarayanan, M. F. Duarte, D. Reddy, and R. G. Baraniuk, "Compressive sensing for background subtraction," in *European Conf. Comp. Vision (ECCV)*, 2008, pp. 155–168.

[12] G. Warnell, S. Bhattacharya, R. Chellappa, and T. Basar, "Adaptive-rate compressive sensing using side information," *ArXiv e-prints*, Jan. 2014.

[13] R. Baraniuk, M. Davenport, R. DeVore, and M. Wakin, "A simple proof of the restricted isometry property for random matrices," *Constructive Approximation*, vol. 28, no. 3, pp. 253–263, 2008.

[14] K. P. Murphy, *Machine Learning: A Probabilistic Perspective*. The MIT Press, 2012.

[15] S. Ji, D. Dunson, and L. Carin, "Multitask compressive sensing," *IEEE Transactions on Signal Processing*, vol. 57, no. 1, pp. 92–106, Jan 2009.

[16] S. Mallat and Z. Zhang, "Matching pursuits with time-frequency dictionaries," *IEEE Transactions on Signal Processing*, vol. 41, no. 12, pp. 3397–3415, Dec 1993.

[17] Z. Wang, A. Bovik, H. Sheikh, and E. Simoncelli, "Image quality assessment: from error visibility to structural similarity," *IEEE Transactions on Image Processing*, vol. 13, no. 4, pp. 600–612, April 2004.

[18] T. Park and G. Casella, "The Bayesian lasso," *Journal of the American Statistical Association*, vol. 103, no. 482, pp. 681–686, 2008.