Sparse Representation for Paddy Plants Nutrient Deficiency Tracking System

Zar Zar Tun, Khin Htar Nwe
University of Information Technology (UIT), Myanmar
zarzarhtun@uit.edu.mm, khinhtarnwe@uit.edu.mm

Abstract

Moving object detection and tracking from consecutive frames of sensing devices (Unmanned Aerial Vehicles-UAV) needs efficient sampling from mass data with sufficient memory saving. Objects with super pixels are tracked by Compressive Sensing (CS) and the generative structural part model is designed to be adaptive to variation of deformable objects. CS can precisely reconstruct sparse signal with a small amount of sampling data. This system creates the sparse representation (SR) dictionary representing the nutrient deficiency tracking system for paddy plants to support the healthily growth of the whole field. This system uses compressed domain features that can be exploited to map the semantic features of consecutive frames. As the CS is a developing signal processing technique, a sparse signal is reconstructed with efficient sampling rate and creates the sparse dictionary. The SR for paddy plant health system can build rich information about paddy plants from signaling devices and can alert the deficiency conditions accurately in real time.

Keywords- Compressed Domain Features, Compressive Sensing (CS), Sparse Representation (SR), Dictionary Learning (DL), Paddy Plants

1. Introduction

Rice is the main food and the production of rice has become a major part of the economy in most Asian countries. Rice production is beneficial not only to farmers but also to the country. Rice fields are changing different brightness patterns at different plant growth stages while they are growing. The healthy state of the plants can be known from the pattern changes of the plants. This paper creates the nutrient deficiency tracking for paddy plant healthy growing system by using the motion values from a streaming video of the paddy fields. This system can alert the states of the plants in real time.

CS theory includes the three basic components as sparse representation, encoding measuring, and reconstructing algorithm [8]. SR is a novel signal sampling method for the sparse or compressible signal and has been successfully applied to signal processing. Sparse coding allows to represent a signal as a linear combination of a few atoms of a dictionary. Dictionary learning methods determine the proper representation of data via decreased dimensionality subspaces.

Selection of extracted features plays an important role in this system. Other sparse representation systems use the sparse parameters from random frames that can cause computational complexity and needs more storage space in feature extraction. In this system, compressed domain features as Motion Vectors and residuals coefficients are extracted from video frames by partial decoding. Extracted information is representing the motion, spatial frequency, edge and color contents that can force to get accurate arguments. That can reflect the most matching edges as outliners and motion values as inliers for a specific condition of plants in creating the sparse dictionary.

The old classification systems extract HOG features and create Bag of Words model to detect objects and recognize actions. There are many drawbacks as memory complexity for storage space and decreasing the accuracy for real time system. The most important one is less of descriptors mapping from raw features to higher level feature labels. Sparse and redundant signal representations have recently used for solving existing problems as high transmission bandwidth and large storage memory allocation. CS can efficiently acquire and reconstruct a signal. The sparsity of a signal can be exploited to recover a signal from fewer samples than required in the Shannon-Nyquist sampling theorem [25]. This system creates the sparse dictionary using sparsity coefficients getting from compressed domain features for efficient tracking the healthy states of paddy plants in real time.

The rest of this paper is organized as follows: Session 2 reviewed related works concerning with this system. Session 3 will be discussed about the proposed system. Session 4 will be explained the experimental results which will be followed by conclusion in Session 5.

2. Related Work

There has been considerable effort devoted to create this system in the last decade. L.Pan, X.Shu and M.Zhang [3] proposed efficient key frame extraction algorithm that exploits Compressive Sensing and unsupervised clustering. J.Jiang et al. [4] proposed a new dimensionality reduction method called

ISBN 978-99971-0-381-9

compressive sensing with Gaussian mixture random matrix (CS-GMRM), in which a novel measurement matrix using Gaussian mixture distribution is constructed and is proved to satisfy the restricted isometry property.

X.Huang et al. [5] segmented the moving object through the robust principal component pursuit (PCP) for that the image is consisted with low-rank of the background regions and the sparsity of the foreground regions. Then, the data dictionary is created through KSVD to strengthen the sparse representation of the dictionary capabilities. S.Qaisar et al. [6] presented a brief background on the origins of the CS idea, reviews the basic mathematical foundation of the theory and then highlighted different areas of its application with a major emphasis on communications and net-work domain.

B.Kaung et al. [7] proposed an object detection model to simultaneously reconstruct the foreground, background, and video sequence using the sampled measurement. Then, they used the reconstructed video sequence to estimate a confidence map to improve the foreground reconstruction result. In paper [8], authors provided a comprehensive study and an updated review on sparse representation to supply guidance for researchers.

V.M.Patel and R.Chellappa reviewed the role of Sparse Representation (SR), Compressive Sensing (CS) and Dictionary Learning (DL) for object recognition. Algorithms to perform object recognition using these theories are reviewed [9]. In paper [10], authors addressed the problem of object detection by representing an extracted feature of an image using a sparse linear combination of chosen dictionary atoms.

Authors in [11] presented the comparison of recently proposed CS several methods as Haar transforms, hybrid CS-Haar, averaging and sub-sampling, and performing recognition to compress time series either directly in the compressed domain over the reconstructed signals. Authors in [12] proposed to decompose the motion field into sparse and non-sparse components for the motion boundaries and small universal noises. By exploiting the statistics on optical flow fields dataset, authors found that gradients of flow fields come from two sources: a sparse large motion-discontinuity component and small dense Gaussian component.

S.Hou, S.Zhou and M.A.Siddique [13] proposed about CS based algorithms that are investigated for Query by Example Video Retrieval (QEVR) and a novel similarity measure approach. This system combined CS theory with the traditional discrete cosine transform (DCT), better compression efficiency for spatially sparse is achieved. X.Shu and N.Ahuja [14] proposed a three-dimensional compressive sampling (3DCS) approach to reduce the required sampling rate of the Compressive Imaging (CI) camera to a practical level. In 3DCS, a generic three dimensional sparsity measure (3DSM) is presented, which decodes a video from incomplete samples by exploiting its 3D piecewise smoothness and temporal low rank property.

S.Narayanan and A.Makur [15] proposed to use a circulant CS matrix on image frames to obtain the CS measurements and then to perform motion estimation in the measurement domain. G.Chen and D.Needell [16] introduced the compressed sensing problem as well as recent results extending the theory to the case of sparsity in tight frames and the problem of dictionary learning, its origin and applications, and existing solutions. G.Li et al. [17] proposed a robust object tracking and generative action recognition method.

All reviews supported to develop this proposed system.

3. The Proposed System

This system includes three contributions as:

(1) This system creates the first paddy plants health system using the sparse representation dictionary.
(2) This system extracts sparse values representing motion, spatial frequency, edge and color contents from compressed features that can force to get accurate arguments.
(3) This system can track and alert accurate paddy plant deficiency conditions in real time.

CS is an innovative concept that directly acquires signals in a compressed form if they are sparse in certain transform domains. This system creates the sparse dictionary using sparse representation coefficients getting from compressed domain features. Sparse representation can reduce existing noise reduction and dimensionality reduction methods because the relevance of CS that is a dimensionality reduction technique for series of sampled signals. Figure 1 shows the functional architecture of the proposed system.

This system creates the paddy plants healthy state system using the sparse dictionary for tracking in real time. Existing systems [1, 22-24] are using HOG, MBH and MFH features and creates the Bag of Words model to reconstruct and track the moving features. These old systems accuracy depends also on the code books classification methods. There are so many challenges such as increasing the memory requirements, higher complexity and less of higher visual level features mapping for deformable objects from sensing devices. Increasing features can degrade the accuracy for real time systems.

In this system, compressed domain features are firstly extracted from original video sequences. Then this system counts sparse coefficients in compressed domain by CS to create the sparse dictionary that is sparse representation coefficients classes of linear representation system. Test samples can usually represent samples from the same objects of incoming videos. This system uses the sparse dictionary to analyze different situations from new coming video shots or sensing states for efficient tracking of deficiency conditions of paddy plants.

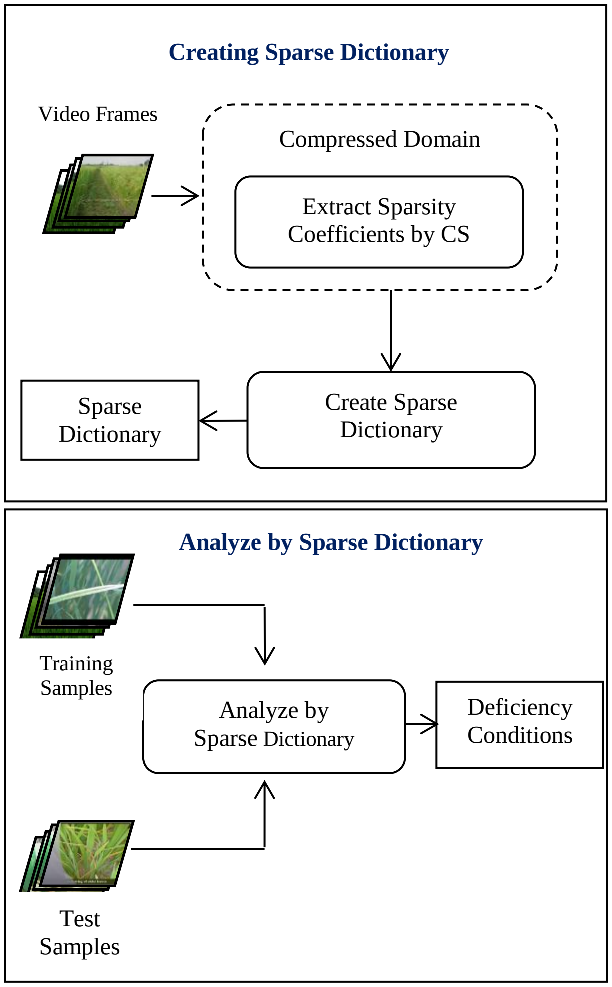

Figure 1. Functional Architecture of the Proposed System

3.1. Compressed Domain Features Extraction

The streaming video of sensing devices consists of compressed domain features as Motion fields and its intensity information of the ongoing scene. Motion fields are representing the Motion Vectors and intensity information is representing Discrete Cosine Transform (DCT) coefficients that form as residuals when motion estimation is performed.

Motion estimation is the process of estimating the best match block of current frame in the reference frame. There are three types of frames in video sequences as Intra frame (I-frame), Predicted Frame (P-frame) and Bi-directionally predicted frame (B-frame). There is also prediction error that is the result of difference between motion vectors and transforms coefficients.

Figure. 2 shows the process of Motion Estimation. Motion Vectors and its residual coefficients are complementary to each other to form an accurate action.

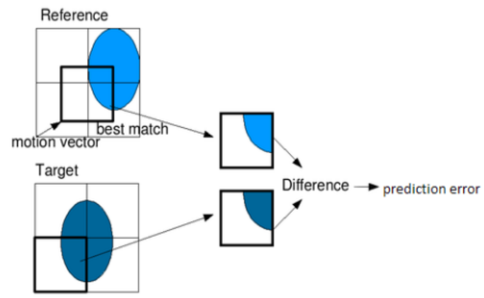

Figure 2. Motion Estimation Process [2]

This system accumulates all motion vectors and residuals of P frames from incoming video sequences as formulated in equation (1):

$$X = \sum_{i}^{n} |u_i| + |v_i| \qquad (1)$$

where X consists of total motion values from P frames in a video sequence, u and v are representing the motion vectors and residuals. Motion data that do not represent real values are filtered by Gaussian approach to be more reliable in motion segmentation. This system uses these compressed domain features that can reduce the computational complexity because coefficients are extracted by partial decoding instead of fully compression.

3.2 Sparse Representation in Compressive Sensing

Compressive sensing (CS) is a novel signal sampling method to reconstruct a signal from a series of sampling measurements by finding the solutions to underdetermined linear systems. Nyquist- Shannon sampling theorem states that if the signal's highest frequency is less than half of the sampling rate, then the signal can be reconstructed perfectly [20]. There are two types of conditions as sparsity and incoherence to recover a signal.

Sparsity: Natural signals can be stored in compressed form if a large number of projection coefficients are small enough to be ignored. Most of elements are zero in sparse matrix or sparse array. If the signal is not sparse, then recovered signal is best reconstruction getting from S largest coefficients of signal. If the total number of elements are R , total sparse elements are S and the left are dense elements $N = R - S$.

In most problems, signals are modeled by a small set of prototypes. In this system, the prototype signal representing motion values, $X = [x_1, x_2, ..., x_n] \in R^n$ is used for training the dictionary $D = [d_1, d_2, ..., d_m] \in R^{m \times n}$, which can be considered as an overcomplete basic matrix consisting of elementary signals called atoms. In the overcomplete dictionary, the number of samples is larger than the dimensions of the samples in the dictionary. This system uses the learned dictionary from compressed features matrix as:

$$x = D\alpha \tag{2}$$

The probe sample, $x \in R^{n \times m}$ that can be achieved by a linear combination of a few small number of dictionary atoms. $\alpha = [\alpha_1, \alpha_2, ..., \alpha_m]^T$ is the sparse coefficient matrix. The solution of α can be searched by the following problem:

$$\arg\min_{\alpha} \|\alpha\|_0 \quad \text{subject to} \quad x = D\alpha \tag{3}$$

where the equation (3) denotes the $l_0 - norm$ which counts the number of non-zero entries in a vector. As it is NP hard, alternative solutions are often sought. If the solution of x is sparse enough, then the sparsest solution can be recovered via l_1 minimization as shown in equation (4):

$$\arg\min_{\alpha} \|\alpha\|_1 \quad \text{subject to} \quad x = D\alpha \tag{4}$$

where, the equation (4) provides the sparsest recovery [18]. If the measurements are contaminated in the noisy setting because of the error, e which obeys $\|e\|_2 < \varepsilon$ as follows:

$$y = D\alpha + e \qquad \text{for} \qquad \|e\|_2 < \varepsilon \tag{5}$$

where ε is the allowed error tolerance. This system finds the sparsest vectors using Orthogonal Matching Pursuit [21] that has average trade-off classification accuracy and processing time comparing with other sparse representation methods [8] and this equation is presented as follows:

$$\arg\min_{\alpha} \|\alpha\|_1 \quad \text{subject to} \quad \|D\alpha - y\|_2 < \varepsilon \tag{6}$$

Intuitively, the $l_1 - norm$ is the convex relaxation function closest to the $l_0 - norm$.

3.3 Paddy Plants State Dictionary Learning

The dictionary learning (DL) method determines the proper representation of data via decreased dimensionality subspaces. An effective dictionary can lead to excellent reconstruction results and satisfactory applications, and the choice of dictionary is also significant to the success of sparse representation technique. Sparse dictionary not only provides a sparse representation but also constructs a sparse dictionary [26].

This system creates the sparse dictionary using the prototype signals representing the nutrient deficiency features as columns that are extracted from consecutive frames of a streaming video of paddy fields. This system finds the sparse representation vectors by $l_1 - norm$ and then updates the dictionary by K-SVD. These two steps are iteratively preserved until a sufficient small residue point is reached.

This system uses the feature sets samples $X = [x_1, x_2, ..., x_k]$, $x_i \in R^d$ to find the dictionary $D \in R^{d \times n}$, $D = [d_1, ..., d_n]$ and the representation matrix $R = [r_1, ..., r_k], r_i \in R^n$. This system uses K-SVD algorithm because it is one of the most simplicity and effective algorithm among existing dictionary learning algorithms [19]. It is finding the best possible codebook to represent the data samples y by nearest neighbor. $\|X - DR\|_F^2$ is minimized by equation (6) and the representation r_i are sparse enough. Then the dictionary D is updated by the following optimization problem,

$$\arg\min_{D,R} \|X - DR\|_F^2 = \sum_{i=1}^{d} \sum_{j=1}^{n} \left\| x_{i,j} - Dr_{i,j} \right\|_2^2$$

$$\text{subject to} \quad \left\| r_{i,j} \right\|_0 \le \varepsilon \tag{7}$$

where $r_{i,j}$ is the sparse representation for the $j-th$ samples of class i, and ε indicates the maximum allowed nonzero entries in $r_{i,j}$. F denotes the Frobenius norm. This system can gradually update the dictionary about the healthy status of paddy plants in real time and drastically reduce the amount of memory needed to store the huge size of dataset.

4. Experimental Result

This system can highly provide to manage the nutrient deficiency in paddy plants. This system gives the facts for totally eleven deficiencies as feature vectors about health paddy conditions: Phosphorus that can help in fibrous root development. If there is phosphorus deficiency, the plant is in purple or brownish red discoloration on leaves. Nitrogen encourages the vegetative growth of paddy. If nitrogen deficiency occurs, older leaves become yellow and stunt in growing plants. Calcium promotes the activities of soil bacteria. Magnesium is the essential constituent in Chlorophyll molecule. Iron is necessary for chlorophyll synthesis. Manganese helps in uptake of Nitrogen. If iron, manganese and magnesium deficiency occurs interveinal yellowing and chlorosis of young leaves.

Copper is important for panicle development. Sulphur is constituent in straw and stalk. Boron helps in fertilization. If calcium and boron deficiency occurs, leaves be-come white and tips of young leaves roll. The other two conditions are zinc toxicity and aluminum toxicity of plants. If farmers know the conditions of plants, correct measures can be applied to the fields in time.

The training dataset is being constituted using the motion values resulting from streaming frames of sensing video about eleven kinds of paddy plant deficiency states. Incoming video is sampled with 25 frames per second with totally 5356 frames of width 128 and height 128. As requirements for implementation, this system uses MATLAB implementation on the processor core i7 and 4GB RAM.

The training dataset is normalized with zero mean and unit variance. 30 percent of the training dataset is used as testing dataset to test the new incoming video or state of feature occurrence. The testing dataset is 1601 signals with total dimensions of 1296. The resulting datasets are down sampled by PCA and LDA filtering methods.

Then the dictionary is learned by K-SVD for linear representation for a given set of signals in training dataset. K-SVD can search the best dictionary that can sparsely represent each signal [19]. The learned dictionary is a matrix that contains signals, each of dimensions n (1296). The dictionary was trained 3 times. There are five iterations each time. For each iteration, the number of atoms used to represent a signal is changing as 5 atoms per signal, 10 atoms per signal, 15 atoms per signal, 30 atoms per signal and 40 atoms per signal for testing.

The comparison of average representation error (RMSE) for three group of signal is displayed in Figure 3. All sample groups are under specific representation error.

Meanwhile, this system uses the $l_1 - norm$ sparse representation method to find the precise sparse value about features of paddy plants. The sparse coefficients of signal are 1601 signals with 3755 dimensions. The average accuracy of the classification and the average speed of sparse coding for different number of signals are presented in Figure 4 and Figure 5.

When implementing this system with testing dataset, the accuracy level is about 97% on the average. According to the experimental result, average accuracy is higher when the number of samples increases although the processing time is a little longer.

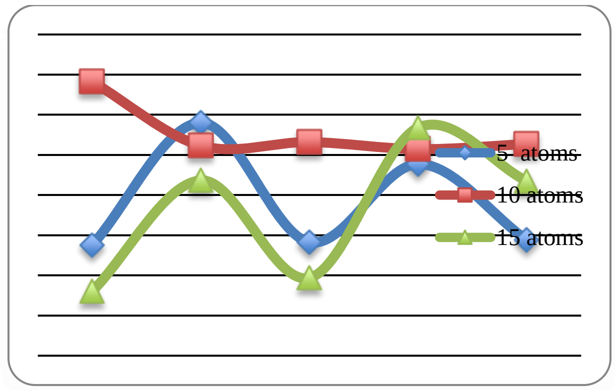

Figure 3. Comparison of RMSE of the Learned Dictionary

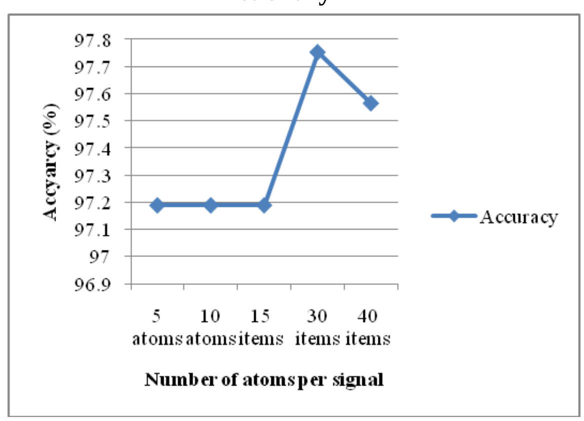

Figure 4. Average Accuracy of Different Items per Signal

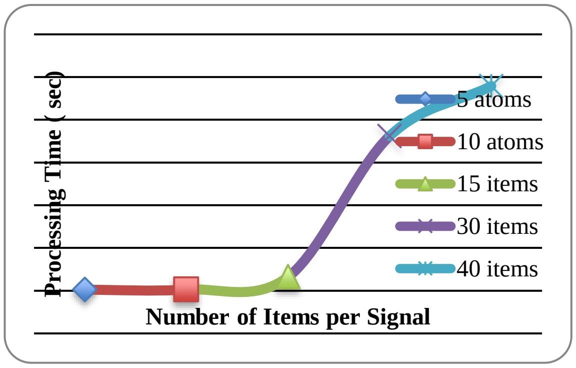

Figure 5. Average Speed of Different Items per Signal

5. Conclusion

Compressive sensing takes the advantage of the redundancy in many interesting signals that are not pure noises. The main idea of this system is that a signal can be decoded from incomplete linear measurements by seeking its sparsity in some domain. There are many ways to represent the signal, but normally the sparsest representation is preferred for simplicity and easy interpretability for sampling. This system typically starts with taking a weighted linear combination of samples about nutrient deficiencies of paddy plants from compressed domain of sensing video data to get compressive measurements. Then this system creates the sparse dictionary consisting of prototype signals about conditions of the plants that are used to express other signals for real time alerts. This system can also be applied for different object tracking systems in many areas. In the future, this system tends to expend to other related fields of study in application requirements for urban countries.

6. References

[1] Zar Zar Tun and Khin Htar Nwe, "Compressed Domain Feature Analysis For Efficient Action Recognition", *International Conference on Disruptive Innovation (ICDC)*, MUST, Malaysia, 24-25 Sept, 2016.

[2] Zar Zar Tun and Khin Htar Nwe, "Comparison of Different Motion Estimation Algorithms in Video Compression", *International Conference on Computer Applications (ICCA)*, Myanmar, 16-17 Feb, 2017.

[3] L.Pan, X.Shu and M.Zhang, "A Key Frame Extraction Algorithm Based on Clustering and Compressive Sensing", *International Journal of Multimedia and Ubiquitous Engineering*, pp. 385-396, Vol.1.0, No.11, 2015.

[4] J.Jiang, X.He, M.Gao, X.Wang and X.Wu, "Human Action Recognition via Compressive-Sensing-Based Dimensionality Reduction", *Optik - International Journal for Light and Electron Optics*, Volume 126, Issues 9–10, May 2015, Pages 882–887.

[5] X.Huang, F.Wu and P.Huang, "Moving-object Detection Based on Sparse Representation and Dictionary Learing", *AASRI Conference on Computational Intelligence and Bioinformatics*, Volume 1, 2012, Pages 492-497.

[6] S.Qaisar, R.M.Bilal, W.Iqbal, M.Naureen and S.Lee, "Compressive Sensing: From Theory to Applications, A Survey", *Journal of Communications and Networks* ", October 2013, Page: 443-456.

[7] B.Kaung, W.P.Zhu and J.Yan, "Object Detection Oriented Video Reconstruction Using Compressed Sensing", *Journal on Advances in Signal Processing 2015(1)*, November 2015.
[8] Z.Zhang, Y.Xu, J.Yang and S.Zhang , "A Survey of Sparse Representation: algorithms and Applications", *The journal for Rapid Open Access Publishing, IEEE*, May 20, 2015.

[9] V.M.Patel and R.Chellappa, "Sparse Representaions, Compressive Sensing and Dictionaries for Pattern Recognition", *Processing of IEEE*, March 2011.

[10] G.K.Vinay, S.M.Haque, R.V.Babu and K.R.Ramakrishnan, "Human Detection Using Sparse Representation", *IEEE International Conference on Intelligent Computing and Intelligent Systems*, December 2009.

[11] R.Xu, O.P.Concha and M.Piccardi "Compressive Sensing of Time Series for Human Action Recognition", *International Conference on Digital Image Computing: Techniques and Applications (DICTA)*, December 2010.

[12]Z.Chen, J.Wamg and Y.Wu "Decomposing and Regularizing Sparse/Non-sparse Components for Motion Field Estimation", *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2012.

[13] S.Hou, S.Zhou and M.A.Siddique "A Compressed Sensing Approach for Query by Example Video Retrieval", *Multimedia Tools and Applications*, Volume 72, Issue 3, October 2014, pp 3031–3044.

[14] X.Shu and N.Ahuja "Imaging via Three-dimensional Compressive Sampling (3DCS)", *IEEE International Conference on Computer Vision (ICCV)*, Nov 2011.

[15] S.Narayanan and A.Makur "Camera Motion Estimation using Circulant Compressive Sensing Matrices", *International Conference on Information, Communications & Signal Processing (ICICS)*, December 2013.

[16] G.Chen and D.Needell "*Compressed Sensing and Dictionary Learning*", 29-October- 2014.

[17] G.Li, F,Wang and W.Lei "Geneartive Human Action Tracking Based on Compressive Sensing", *International*

Journal of Signal Processing, Image Processing and Pattern Recognition, Vol.8 No.7, 2015.

[18] A.M. Bruckstein, D.L. Donoho, and M. Elad,"S*parse Solutions of Systems of Equations to Sparse Modeling of Signals and Images*", *SIAM Review*, 2009, vol. 51, no. 1, pp. 34–81.

[19] M. Aharon, M. Elad, and A.M. Bruckstein, "The K-SVD: An Algorithm for Designing of Overcomplete Dictionaries for Sparse Representation*", IEEE Trans On Signal Processing*, November 2006, Vol. 54, no. 11, pp. 4311-4322.

[20] E.J.Candes and M.B.Wakin, "An Introduction to Compressive Sampling", *IEEE Signal Processing Magazine*, March-2008.

[21] J. A. Tropp and A. C. Gilbert, "Signal recovery from random measurements via orthogonal matching pursuit", *IEEE Transactions on Information Theory*, 2007, vol. 53, no. 12, pp. 4655¨C4666.

[22] V.Kantorov and I.Laptev, "Efficient Feature Extraction, Encoding and Classification for Action Recognition*", Computer Vision and Pattern Recognition (CVPR), IEEE*, 2014.

[23] J.Miao, X.Xu, R.Mathew and H.Huang, "Residue Boundary Histograms for Action Recognition in the Compression Domain", *IEEE International Conference on Image Processing*, Sept-2015, vol.25, pp.39-52.

[24] J.Uijlings, I.C.Duta, E.Sangineto and Nicu Sebe, "*Video Classification with Densely Extracted HOG/ HOF/ MBH Features:* An Evaluation of the Accuracy/ Computational Efficiency Trade-off", *International Journal of Multimedia Information***,** March 2015, Volume 4, Issue 1, pp33-44.

[25] https://en.wikipedia.org/wiki/Compressed_sensing, last accessed 13-Oct-2017.

[26] htps://en.wikipedia.org/wiki/Sparse_dictionary_learning, last accessed 13-Oct-2017.