



Journal of Artificial Intelligence

ISSN 1994-5450

science
alert

ANSI*net*
an open access publisher
<http://ansinet.com>



Research Article

Dictionary Properties for Sparse Representation: Implementation and Analysis

Himanshu Patel and Hiren Mewada

Department of Electronics and Communication, Charotar University of Science and Technology, Changa, Gujarat, India

Abstract

Background and Objective: Linear inverse problems emerge throughout the engineering and the mathematical sciences. Over the last two decades, sparsity constraints have emerged as a fundamental type of regularizer. This paper implements and assesses the role of dictionary miscellany in sparse representation for image processing prospective. **Materials and Methods:** A dictionary is formed by a linear basis using a mathematical model from the set of images referred as analytic dictionary or using a set of realizations of the images referred as trained dictionary. This study considers the problem of true sparsity formation and analyzes the two most commonly used algorithms-the Matching Pursuit (MP) and Orthogonal Matching Pursuit (OMP) using analytical dictionaries. These methods were compared using diverse dictionaries formation for image restoration applications. **Results:** The results were validated using peak signal to noise ratio and mean square error of the sparse approximation for the images. The different dictionaries like-discrete wavelet dictionary, Discrete Cosine Transform and Kronecker Delta dictionary and Haar Wavelet Packets and DCT dictionary had been used for implementation of these two algorithms. **Conclusion:** This experiment showed that the discrete wavelet based dictionary performs best with orthogonal matching pursuit algorithm in terms of MSE and PSNR performances. The result also shows the out performance of OMP in comparison with MP. From the experiments, it has been observed that high number of iterations and small patch size proves to be advantageous.

Key words: Dictionary learning, greedy algorithms, matching pursuit, orthogonal matching pursuit, sparse approximation

Citation: Himanshu Patel and Hiren Mewada, 2018. Dictionary properties for sparse representation: Implementation and analysis. J. Artif. Intel., 11: 1-8.

Corresponding Author: Himanshu Patel, Department of Electronics and Communication, Charotar University of Science and Technology, Changa, Gujarat, India Tel: +919427855034

Copyright: © 2018 Himanshu Patel and Hiren Mewada. This is an open access article distributed under the terms of the creative commons attribution License, which permits unrestricted use, distribution and reproduction in any medium, provided the original author and source are credited.

Competing Interest: The authors have declared that no competing interest exists.

Data Availability: All relevant data are within the paper and its supporting information files.

INTRODUCTION

Linear inverse problems emerge throughout the engineering and the mathematical sciences. These problems are undetermined in most applications, so for obtaining interesting or useful solutions one must enforce extra regularizing constraints. Sparsity constraints have developed as a fundamental type of regularizer over the last two decades¹. This approach looks for an approximate solution to a linear system by finding dictionary where condition for requirement is that the target signal (i.e., $b = b_1, b_2, \dots, b_n$) can be approximated by very few nonzero elements (i.e., x) regarding to its dimension. Thus sparse approximation can be presented as:

$$\min \{ \|x\|_0 : Ax = b \} \quad (1)$$

where, $A = \{a_j\}_{j=1}^m$ is a dictionary and x is the sparse matrix. This formulation is also known as a sparse approximation. These types of problems occur in many areas including fields of statistics, signal/image processing, approximation theory, coding theory and machine learning etc.¹.

In sparse approximation, images are divided into number of small rectangle patches². This process of breaking the image allows an easy way of dictionary learning and sparse coding with low computational complexity. To fit for the dictionary learning and sparse coding processes, each patch is reshaped as a one-dimensional (column) vector³.

The signals which are representing the columns of the matrix A in Eq. 1 are called "atoms" and "dictionary" is the set of these atoms. If $n = m$ then it generates complete dictionary and if $n > m$ then it provides over-complete dictionary⁴.

In Eq. 1, the L_0 norm is used to denote the number of nonzero components. The equation 1 is an ideal case with zero error. The problem statement which allows some approximation error for representation can be presented as:

$$\min \{ \|x\|_0 : \|Ax - b\|_p \leq \delta \} \quad (2)$$

where, δ is the threshold value. The norm p is generally 2 but it can be any value between 1 to ∞ as well depending upon requirement. Combinatorial optimization methods are used to find the exact solutions to the problem statement mentioned above. There are essentially two ways to deal with suboptimal solution. One approach is a greedy approach in which signal vector is approximated via a sequence of incremental approximations by selecting atoms suitably. This kind of

approaches are known as Matching Pursuits (MP) and Orthogonal Matching Pursuit (OMP)^{5,6}. The second is known as the Basis Pursuit (BP) which relaxes the L_0 norm condition by L_1 norm and solves the problem through linear programming⁷. Greedy algorithms matching pursuit and orthogonal matching pursuit algorithms are discussed in this study.

Though sparse coding is used in large extent, the exploration of algorithm which solves the equation 1 has been inadequate in the literature. Therefore, this paper reviews these sparse approximation algorithms with blend of different dictionaries and proposed the optimum combination of dictionary with sparse approximation algorithms for optimum image restoration.

PURSUIT METHODS

Greedy algorithm is a pursuit method used to solve the problem of sparse approximation by iteratively refining the current estimate for coefficient vector x by changing one or several coefficient elements yielded a substantial improvement in approximation of a given signal. The main advantage with the greedy algorithms is that the complexity of the sparse approximation problem is reduced by ensuring that the current support is always sparse during the execution of the algorithm. This section presents simplest effective greedy algorithm, Matching Pursuit (MP) initially and later it is exploited further using Orthogonal Matching Pursuit (OMP)⁸.

MATCHING PURSUIT

Matching Pursuit (MP) is a type of greedy algorithm that iteratively solves for the sparse approximation of a given signal with the help of a series of mono-atomic approximations¹. Two steps are involved for each iteration of the algorithm: The first step being an atom selection step and the second step of the residual update step. In atom selection step, an atom is selected based on the condition that it provides the highest correlation with the current residual error, provided the correlation is measured with respect to the length of the orthogonal projection. Second step involves update in residual error by subtraction of the correlated part from it⁵.

The working principle of MP involves replacement of the simultaneous atom selection with selection of the sequential atoms³. At a time only one atom gets the chance for being selected and for each iteration, it is removed from the signal.

The steps describe the MP process are listed below:

$$\min_{x \in \mathbb{R}^m} 0.5(\|b - Ax\|_2^2) \text{ s.t. } \|x\|_0 \leq L$$

1. Initialize $x = 0$, residual $r = b$
2. While $|X|_0 < L$
3. Select the element with maximum correlation with the residual
 $\hat{I} = \operatorname{argmax}_{i=1, \dots, m} |a_i^T r|$
4. Update the coefficients and residual
 $x_i = x_i + a_i^T r$
 $r = r - (a_i^T - r) d_i$

Here, L is the threshold level and r represents residual error.

The major drawback of the MP is selection of same atom multiple times. The residual $r(i)$ is calculated by the projecting the previous residual $r(i-1)$ with the condition of orthogonality for the selected atom $A(i)$. So for any i , there are $r(i)$, $A(i)$. If 'A' do not matches the condition of orthogonality, then a correlation with a previously selected atom can be brought back by subtracting an atom. Hence, the same atom can be selected several times in Matching Pursuit. In noiseless case also, for getting a null residual, large number of iterations are required though complete support has already been recovered⁸.

ORTHOGONAL MATCHING PURSUIT

The matching pursuit algorithm is the simplest method for sparse approximation but due to the sub-optimality, it faces the problem of poor sparsity result and the slow convergence. This drawback can be overcome by using Orthogonal Matching Pursuit algorithm by projecting signal vector into the subspace which is spanned by selected atoms^{5,9}. The working principle of OMP is prevention of atom getting selected more than once by promising the residual $r(i)$ by matching the condition of orthogonality with all atoms which are previously selected¹⁰. Let the sub-dictionary $A(i) = (A_j)_{1 \leq j \leq i}$. The residual $r(i)$ is computed by projecting orthogonally to $A(i)$ ⁸. The steps describing the process are listed below.

Here, Ω represents active set. Orthogonal Matching Pursuit (OMP) is a greedy algorithm. Hence, weakness of the OMP is that the process used to select the atoms is sub-optimal. The criteria for selecting an atom depend only upon the residual known at the current iteration which is being executed. Once an atom gets selected, it does not get any chance for removing it. This scenario can produce a situation of a sub-optimal atoms getting selected at an earlier stage which will never be correlated⁸:

$$\min_{x \in \mathbb{R}^m} 0.5(\|b - Ax\|_2^2) \text{ s.t. } \|x\|_0 \leq L$$

1. Initialize $x = 0$, residual $r = b$, active set $\Omega = \Phi$
2. While $|x|_0 < L$
3. Select the element with maximum correlation with the residual
 $\hat{I} = \operatorname{argmax}_{i=1, \dots, m} |a_i^T r|$
4. Update the active set, coefficients and residual
 $\Omega = \Omega \cup \hat{I}$
 $x_{\Omega} = (a_{\Omega}^T a_{\Omega})^{-1} a_{\Omega}^T r$
 $r = b - a_{\Omega} x_{\Omega}$

FORMING THE DICTIONARY

The dictionary is generalized concept to form the basis function in linear algebra. For given dictionary better localization of transformation provides better sparsity. The choice of dictionary used for sparse code can be made using either analytical approach or training approach.

In Analytic approach, choosing a predesigned transform, such as wavelet or curvelet can make the optimization faster. The disadvantage of this approach is that it is not accurate and cannot be used as a general model for each problem statement¹¹.

Training approach focuses upon training the dictionary with the help of given examples, optimizing with respect to some criteria i.e., sparsity. This approach makes the approximation slower but has advantage in terms of accuracy and can be used as a general model¹¹.

Dictionaries mentioned in first approach are called analytic dictionaries due to their characteristic of analytic formulation and they are faster to implement. While dictionaries mentioned in second approach can be more flexible in terms of ability to adapt to specific data and due to this reason being they are referred as trained dictionaries¹².

Analytic dictionary approach: An image is multidimensional signal. The complex geometry like curves and manifolds need to be trace for better representation. Thus transformation model requires localization and orientation to represent the images in sparse form. Generally tight frames are used to formulate these analytic dictionaries. Tight frames can be viewed as the main characteristics for the formulation of analytical dictionaries that means $AA^T b = b$ for all b and due to this type of characteristic the transpose of dictionary is useful to acquire a good representation by means of the dictionary. Further this approach gets advanced by resolving the behavior of the filter-set $A^T b$. This approach is quite advantageous because the operator for analysis can be viewed as an easy option than synthesis framework where the sparsity bounds must be derived. Another major advantage

is algorithm simplicity. The ability to produce a useful structure simultaneously for both of the analysis and synthesis frameworks gives a meaningful interpretation¹¹.

First of all coefficients $A^T b$ are computed and then passed through a non-linear shrinking operator. This approach has an advantage in terms simplicity and is able to give efficient sparse representations over given dictionary¹¹.

There exists a number of modeled dictionaries in literature including dictionaries based on wavelets, fourier or gabor transform based structure^{6,13,14}. Moreover, there are dictionaries with combinations of orthogonal bases¹⁵, curvelets or contourlets^{16,17}. Mentioned dictionaries are able to satisfy requirement of the target signal or data. However, the chosen model is deciding factor to specify the quality of the sparsest solution.

Dictionary in wavelet domain: One can obtain a sparse representation of the target signal with the help of the wavelet transform to some degree¹⁸. Consider the learning problem expressed by the following modification to Eq. 3:

$$\arg \min_{A,x} (\|b - W_s A x\|_2^2) \text{ s.t. } \|x\|_0 \leq L \quad (3)$$

Here, A represents the dictionary, b is the target and x is sparse vector. The wavelet synthesis W_s operator (inverse Wavelet) or equivalently the wavelet atom dictionary proposes that the data can be represented by a sparse combination of atoms, which are combinations of atoms from a fixed multi-scale core dictionary¹⁸. This problem is quiet difficult for reasonable sized data, without additional constraints or assumptions on the unknown A . Here, assumed that A has very sparse columns¹⁹. This implies that the overall dictionary atoms are linear combination of few (and arbitrary) wavelet atoms. Assuming that W_s is square and unitary (i.e., orthogonal wavelet with periodic extension), it can be formulated as:

$$\arg \min_{A,x} (\|W_A b - A x\|_2^2) \text{ s.t. } \|x\|_0 \leq L \quad (4)$$

where, W_A is a wavelet analysis operator. This formulation proposes that the training of dictionary is not in the image domain but in the analysis domain of the multi-scale decomposition operator, specifically the wavelet transform¹⁸.

The idea of learning a wavelet-dictionary had already been proposed in using a MAP approach^{20,21}. However, the effectiveness of various wavelet dictionaries is not mentioned.

This study presents detail analysis of three basic analytic dictionaries including discrete wavelet, DCT and haar and their performance are compared.

EXPERIMENTS RESULTS AND DISCUSSION

The sparsity has been integrated with various algorithms including support vector machine, genetic algorithm, deep learning, convolutional neural network etc^{8,9,22} to fulfill the various applications including image denoising, scene classification and pattern recognitions^{23,24}. In the experiment, the compatibility of MP and OMP algorithms is checked with the dictionaries mentioned in above section for the image enhancement application. The quality of the enhancement is generally calculated using the peak signal to noise ratio (PSNR) and mean square error (MSE). MSE and PSNR can be calculated as:

$$MSE = \frac{1}{MN} \sum_{M=0}^{M-1} \sum_{N=0}^{N-1} [I(i, j) - \hat{I}(i, j)]^2 \quad (5)$$

$$PSNR = 10 \log \left(\frac{255 \times 255}{MSE} \right) \quad (6)$$

The comparison of approximation algorithms has been given in terms of PSNR and MSE. In Fig. 1, the images of dimensions 256×256 were used for the experiment. The patch size and number of iterations used for the dictionary plays an important role in image restoration. Therefore, both methods were analyzed with reference to type of dictionary and size of patch and simulation results were presented in Table 1-3.

Table 1 depicted that for images 1-10, though PSNR of discrete wavelet was better in comparison to haar wavelet packets and DCT, the mean square error (SME) very significantly improved in haar wavelet packets and DCT. From Table 2, it has been observed that MSE and PSNR were better for haar wavelet packets and discrete wavelet in comparison with DCT based dictionary.

Table 1 and 2 presented the PSNR and MSE obtained for the defined set of images using MP and OMP method.

The orthogonal matching pursuit out performing the matching pursuit algorithm was depicted in Fig. 2 and 3. The DCT based transformation has higher MSE. The localization property of DCT affect the performance to achieve better sparsity and hence reconstruction of images.

The quality of sparse representation mainly depends on the choice of dictionary, number of iterations and patch size.



Fig. 1(a-j): Set of images used for experiment (a) Circles bright dark, (b) Barbara, (c) Manon street, (d) Lifting body, (e) Baboon, (f) Concord orthophoto, (g) Circuit, (h) Pout, (i) Cell and (j) Rice blurred

Table 1: Results for matching pursuit algorithm

Dictionary image	Discrete wavelet		DCT and Kronecker delta		Haar wavelet packets and DCT	
	PSNR (dB)	MSE	PSNR (dB)	MSE	PSNR (dB)	MSE
1	40.12	6.31	37.60	43.78	38.43	9.34
2	30.74	23.40	29.18	86.13	29.06	94.51
3	33.18	1.78	32.37	4.56	31.53	37.41
4	33.98	28.81	30.28	60.94	32.95	32.97
5	30.83	21.56	29.17	87.17	29.89	49.15
6	31.65	38.27	29.49	67.81	30.96	35.31
7	33.02	2.15	31.82	8.17	32.51	3.89
8	33.44	1.30	32.97	2.29	32.97	2.29
9	34.14	0.52	32.93	2.41	33.82	0.80
10	34.63	22.39	31.63	44.66	34.43	23.46

PSNR: Peak signal to noise ratio and MSE: Mean square error

Table 2: Results for orthogonal matching pursuit algorithm

Dictionary image	Discrete wavelet		DCT and Kronecker delta		Haar wavelet packets and DCT	
	PSNR (dB)	MSE	PSNR (dB)	MSE	PSNR (dB)	MSE
1	45.17	1.98	43.78	2.72	44.41	2.36
2	31.39	12.67	30.97	18.90	29.07	94.02
3	33.82	0.80	32.99	2.25	31.53	37.73
4	33.53	25.97	31.17	49.66	34.45	23.32
5	31.51	11.22	30.88	20.50	30.37	32.63
6	31.89	39.29	31.09	16.97	31.47	37.44
7	33.74	0.89	33.07	2.05	33.03	2.15
8	33.88	0.74	33.77	0.86	33.77	0.85
9	34.45	0.34	33.17	1.82	34.08	0.567
10	37.36	11.94	38.34	31.63	37.73	10.97

PSNR: Peak signal to noise ratio and MSE: Mean square error

Further localization property plays an important role to achieve better sparsity. Wavelet transform project the image

in both low frequency components and high frequency components. Hence, wavelet transform is more localized in

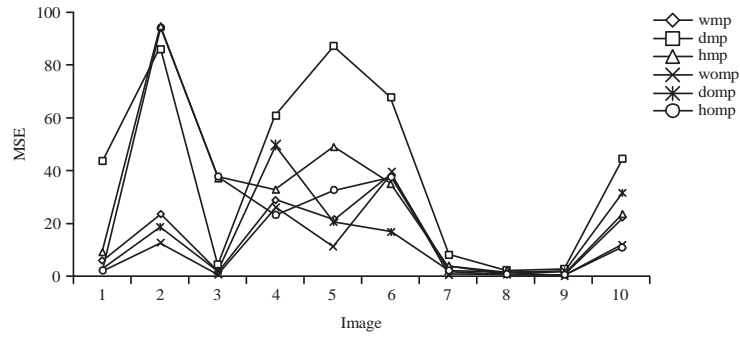


Fig. 2: MSE comparisons for image sets

MSE: Mean square error, wmp: Discrete wavelet dictionary for matching pursuit, womp: Discrete wavelet dictionary for orthogonal matching pursuit, dmp: Discrete cosine transform and Kronecker delta dictionary for matching pursuit, domp: Discrete cosine transform and Kronecker delta dictionary for orthogonal matching pursuit, hmp: Haar wavelet packets and DCT dictionary for matching pursuit

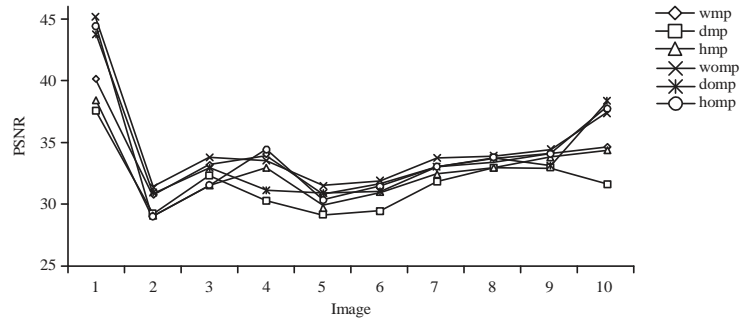


Fig. 3: PSNR comparisons for image sets

PSNR: Peak signal to noise ratio, wmp: Discrete wavelet dictionary for matching pursuit, womp: Discrete wavelet dictionary for orthogonal matching pursuit, dmp: Discrete cosine transform and Kronecker delta dictionary for matching pursuit, domp: Discrete cosine transform and Kronecker delta dictionary for orthogonal matching pursuit, hmp: Haar wavelet packets and DCT dictionary for matching pursuit

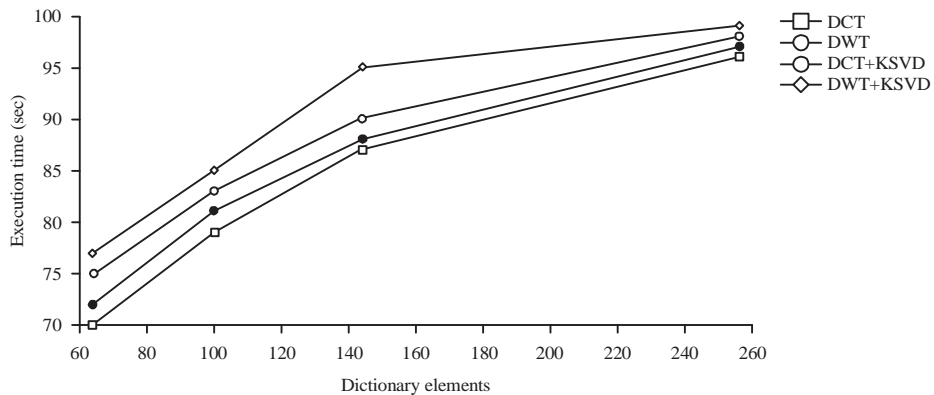


Fig. 4: Computationally time for different dictionaries using several number of sparse and redundant elements

both time and frequency domain. While DCT is fourier based approach which follows the Heisenberg uncertainty principal. Therefore, it cannot depict information both in time domain and frequency domain. Hence DCT is localized

in frequency domain only. Thus sparsity achieved by the wavelet transform was better than DCT²⁵.

Table 3 summarized the effect of patch size and iteration used to correlate the atoms by reducing the redundancy in

Table 3: Effect of patch size and number of iterations

Patch size	Iterations	PSNR(dB)	MSE
4×4	8	30.1854	62.3072
	16	50.8933	21.5030
	24	75.7330	9.0175
8×8	8	21.1845	495.0274
	16	24.9425	208.3698
	24	29.1205	79.6219
16×16	8	19.6955	697.4762
	16	20.3608	598.4157
	24	21.3845	472.7471

PSNR: Peak signal to noise ratio and MSE: Mean square error

sparse. Table 3 concludes that as number of iterations increases the quality of sparse representation also increases. However, an increase in patch size registers poor performance. The MP and OMP were also analyzed in terms of execution time. Figure 4 showed the variation of the computational time for various over complete dictionaries with different number of sparse elements²⁶.

As shown in Fig. 4, the DWT based dictionary was more time consuming process to calculate the sparse and redundant elements then the DCT dictionary.

The different number of atom size was implemented and the dictionary having patch size of 16×16 uses more time to adapt the sparse elements. The DCT dictionary used lesser time with all different sizes as compared to DWT. When DWT dictionary incorporated with KSVD algorithm, the computational time was little increase with the increase of accuracy. The proposed method improved the PSNR for image index 2²⁴. The choice of dictionary, number of iterations and patch size were significantly effect on the quality of sparse representation. Some of more parameters like that high number of iterations and small patch size also effect the same. This experiment clearly showed that high number of iterations and small patch size have serve better PSNR and MSE.

CONCLUSION

The quality of sparse representation mainly depends on the choice of dictionary, number of iterations and patch size. From above experiments, it has been observed that high number of iterations and small patch size proves to be advantageous. It was clear that in comparison of OMP, MP performs slightly cheaper. The reason behind this is that in matching pursuit several atoms get selected more than once. This can lead to higher MSE and lower PSNR. This can be overcome by using OMP algorithm in which, one atom is selected only once, which improves the performance measure of MSE and PSNR. Also different dictionaries

like-discrete wavelet dictionary, discrete cosine transform and Kronecker delta dictionary and haar wavelet packets and DCT dictionary have been used for implementation of these two algorithms and by observing the results, discrete wavelet dictionary has an upper hand with comparison of other two dictionaries in terms of MSE and PSNR performances.

SIGNIFICANCE STATEMENT

This paper presents study the effect of dictionary in sparse based image processing applications. The implementation of various dictionaries is analyzed using image enhancement application. Dictionary formation is the basic steps in sparse based image processing applications. This analysis will help the researchers to decide the patch size which affect the enhancement quality parameters. The comparative analysis presents that wavelet is dominant in dictionary formation in comparison with DCT.

ACKNOWLEDGMENT

Authors would like to thank management of Charotar University of Science and Technology (CHARUSAT), Changa for providing technological support to carry out research at the institute.

REFERENCES

1. Tropp, J.A. and S.J. Wright, 2010. Computational methods for sparse solution of linear inverse problems. Proc. IEEE., 98: 948-958.
2. Guo, C., Q. Ma and L. Zhang, 2008. Spatio-temporal saliency detection using phase spectrum of quaternion fourier transform. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, June 23-28, 2008, Anchorage, AK, USA., pp: 1-8.
3. Nazzal, M. and H. Ozkaramanli, 2014. Variable patch size sparse representation over learned dictionaries. Proceedings of the 22nd Signal Processing and Communications Applications Conference (SIU), April 23-25, 2014, Trabzon, Turkey, pp: 1367-1370.
4. Chen, S.S., D.L. Donoho and M.A. Saunders, 2001. Atomic decomposition by basis pursuit. Siam Rev., 43: 129-159.
5. Rath, G. and A. Sahoo, 2009. A comparative study of some greedy pursuit algorithms for sparse approximation. Proceedings of the 17th European Signal Processing Conference, August 24-28, 2009, Glasgow, UK., pp: 398-402.

6. Mallat, S.G. and Z. Zhang, 1993. Matching pursuits with time-frequency dictionaries. *IEEE Trans. Signal Process.*, 41: 3397-3415.
7. Pati, Y.C., R. Rezaifar and P.S. Krishnaprasad, 1993. Orthogonal matching pursuit: Recursive function approximation with applications to wavelet decomposition. *Proceedings of the 27th Asilomar Conference on Signals, Systems and Computers*, November 1-3, 1993, Pacific Grove, CA., USA., pp: 40-44.
8. Saric, M., H. Dujmic and M. Russo, 2015. Scene text extraction in IHLS color space using support vector machine. *Inform. Technol. Control*, 44: 20-29.
9. Paulinas, M. and A. Usinskas, 2007. A survey of genetic algorithms applications for image enhancement and segmentation. *Inform. Technol. Control*, 36: 278-284.
10. Cai, T.T. and L. Wang, 2011. Orthogonal matching pursuit for sparse signal recovery with noise. *IEEE Trans. Inform. Theory*, 57: 4680-4688.
11. Rubinstein, R., A.M. Bruckstein and M. Elad, 2010. Dictionaries for sparse representation modeling. *Proc. IEEE.*, 98: 1045-1057.
12. Chen, G. and D. Needell, 2016. Compressed sensing and dictionary learning. *Proc. Symposia Applied Math.*, 73: 200-241.
13. Goodwin, M.M. and M. Vetterli, 1999. Matching pursuit and atomic signal models based on recursive filter banks. *IEEE Trans. Signal Process.*, 47: 1890-1902.
14. Gribonval, R., 2001. Fast matching pursuit with a multiscale dictionary of Gaussian chirps. *IEEE Trans. Signal Process.*, 49: 994-1001.
15. Gribonval, R. and M. Nielsen, 2003. Sparse representations in unions of bases. *IEEE Trans. Inform. Theory*, 49: 3320-3325.
16. Matalon, B., M. Elad and M. Zibulevsky, 2005. Improved denoising of images using modelling of a redundant contourlet transform. *Proceedings of the SPIE Volume 5914, Wavelets XI*, September 17, 2005, San Diego, California, United States, pp: 617-628.
17. Starck, J.L., M. Elad and D.L. Donoho, 2004. Redundant multiscale transforms and their application for morphological component separation. *Adv. Imag. Electron. Phys.*, 132: 287-348.
18. Ophir, B., M. Lustig and M. Elad, 2011. Multi-scale dictionary learning using wavelets. *IEEE J. Select. Topics Signal Process.*, 5: 1014-1024.
19. Rubinstein, R., M. Zibulevsky and M. Elad, 2010. Double sparsity: Learning sparse dictionaries for sparse signal approximation. *IEEE Trans. Signal Process.*, 58: 1553-1564.
20. Olshausen, B.A., O. Sallee and M.S. Lewicki, 2001. Learning Sparse Image Codes Using Awavelet Pyramid Architecture. In: *Advances in Neural Information Processing Systems Vol. 13*, Leen, T.K., T.G. Dietterich and V. Tresp (Eds.), MIT Press, Cambridge, MA., pp: 887-893.
21. Sallee, P. and B.A. Olshausen, 2003. Learning Sparse Multiscale Image Representations. In: *Advances in Neural Information Processing Systems, Vol. 15*, Becker, S., S. Thrun and K. Obermayer (Eds.), MIT Press, Cambridge, MA., pp: 1351-1358.
22. Uktveris, T. and V. Jusas, 2017. Application of convolutional neural networks to four-class motor imagery classification problem. *Inform. Technol. Control*, 46: 260-273.
23. Elad, M. and M. Aharon, 2006. Image denoising via sparse and redundant representations over learned dictionaries. *IEEE Trans. Image Process.*, 15: 3736-3745.
24. Portilla, J. and L. Mancera, 2007. L0-based sparse approximation: Two alternative methods and some applications. *Proceedings of the SPIE Volume 6701, Wavelets XII: 67011Z*, September 20, 2007, International Society for Optics and Photonics, San Diego, California, United States.
25. Yan, S., X. Xu, D. Xu, S. Lin and X. Li, 2015. Image classification with densely sampled image windows and generalized adaptive multiple kernel learning. *IEEE Trans. Cybern.*, 45: 381-390.
26. Qayyum, A., A.S. Malik, M. Naufal, M. Saad, M. Mazher, F. Abdullah and T.A.R.B.T. Abdullah, 2015. Designing of overcomplete dictionaries based on DCT and DWT. *Proceedings of the IEEE Student Symposium on Biomedical Engineering and Sciences*, November 4, 2015, Shah Alam, Malaysia, pp: 134-139.