Pak. J. Biotechnol. Vol 15 (3) 751-755 (2018) ISSN Print: 1812-1837
www.pjbt.org ISSN Online: 2312-7791

MACHINE LEARNING BASED IMAGE PROCESSING USING UNSUPERVISED APPROACH

Dhanalakshmi Samiappan[1], S. Latha[2], Deepak Verma[3], CSA Sri Harsha[4], A. Sashank[5]

Department of ECE, SRM Institute of Science and Technology, Kattankulathur, Chennai, Tamil Nadu, India. dhanalakshmi.s@ktr.srmuniv.ac.in

Article received 12.7.2018, Revised 16.8.2018, Accepted 24.8.2018

ABSTRACT

Enhancing the visual media for the purpose of better perception has been a research topic for years. It finds its secondary application in the recognition of objects, analysis of medical images accounting the astronomical data and so on. The disintegration of an image based on its meaningful components plays a key role in many image processing applications like filtering, interpolation, image enhancement, feature variation, etc. the solution to this vary from basic segmentation techniques to advanced methods like fuzzy logic and machine learning. Through this paper, we present a novel method of image processing using machine learning algorithms. We also conduct experiments with preliminary image processing techniques and provide comparable performance measures to illustrate the success of our approach.

Index Terms – Sparse representation, Dictionary Learning, Regularization, Clustering, Unsupervised Learning

I. INTRODUCTION

Image processing as a domain is a vast topic and nonetheless to say that the wide range of applications it provides are worth the research was done in this field. This paper is focused on how the employment of machine learning approach can simplify the task of image processing both concerning the complexity of processing and quality of output. In this paper, we have used two aspects of machine learning namely regularization and clustering (unsupervised approach [Michael and Bishop 2012]) and some of the basic techniques like bilateral filtering [Tomasi and Manduchi, 1998], etc. for comparing the performance of our approach. We primarily aim at solving the problem of rain removal and secondarily the Gaussian noise. The former being the structured form of noise and latter being the unstructured noise patterns. We have successfully tried to bring out the differences from other methods used and the dominance of our approach over them.

A. Sparse Representation: Sparse modelling [Olshausen and Field, 1996, Mallat and Zhang, 1993, Bruckstein, and Donoho, 2009] finds its extensive use in the field of image processing applications. Here we employ it for the purpose of image denoising. Image denoising is achieved by minimization of the following energy function:

$$E(I) = \frac{1}{2}\|I - I'\| + R \qquad (I)$$

Where I' is a noisy image and I is the required image. R (I) is called the priori or regularization parameter. This method is based on probability. The prior can be modified based on the requirement and hence can be used for efficient modelling of images.

The two basic components of sparse modelling are a dictionary, D (will be discussed in the next section) which consist of atoms and a sparse coefficient vector β. Atoms are the linear summation of small basis functions which constitute the columns of D. Sparse representation uses the product of these two to approximate the signal of interest I. Consider a dictionary D of dimension N×K. Let K be the size of sparse coefficient vector β with L at the greatest number of non-zero elements of β such that L≪K, the Dβ=I, where I is the image signal of interest. Sparse model of a signal is very flexible and abundant in nature as it allows us to use any combination of L atoms from set of K atoms to represent our signal. Sparsity is measured using l_p norm where we count the number of non-zeros in β. Empirical results show that most favorable results are obtained when p lies between 0 and 1, which tends to idealize as p→0.

TABLE I: Performance comparisons (In terms of PSNR-Peak Signal to Noise Ratio, MAE-Mean Arithmetic Error, MSE-Mean Squared Error, RMSE- Root Mean Squared Error by varying number of atoms)

ALGORITHM	PSNR (dB)	MAE	MSE	RMSE
BILATERAL	37.2911	0.0064	12.1329	3.4832
KSVD	34.6724	0.0086	22.1730	4.7089
OUR METHOD	41.5319	0.3340	0.1058	0.3253

B. Dictionary Acquisition: Dictionary (refer Fig. 1) learning [Mairal, 2010, 2012, Elad and Aharon, 2006] is a method of training the relevant data such that it most closely summarises the required output.

TABLE II: Performance comparisons (In terms of PSNRPeak Signal to Noise Ratio, MAE-Mean Arithmetic Error, MSE-Mean Squared Error, RMSE- Root Mean Squared Error)

Performance	1 atom		2 atoms		5 atoms	
Measures	OMP	LAR	OMP	LAR	OMP	LAR
PSNR (dB)	29.870	11.693	15.222	11.141	12.748	10.793
MAE	0.0001	0.0033	0.0011	0.0044	0.0022	0.0056
MSE	0.0009	0.8357	0.0928	1.4824	0.3689	2.3121
RMSE	0.0301	0.9142	0.3047	1.2175	0.6073	1.5205

Through dictionary learning, we try to reduce the dimensionality of the image from a very high dimension to very low dimensional space and hence it is possible to remove the noise from the image as we are using only a few atoms for the approximation of our image of interest. To find the best-fit sparse code and most favourable dictionary we need to eradicate the trivial solutions and find the apt solution to the following equation;

$$E = \arg\min_{\theta p} \frac{1}{2}\left(\|D\beta_p - I_p\|\right) + \gamma\|\beta_p\|_1 \qquad (2),$$

where I_p is the p_{th} patch of the image I. Two iterative steps are followed for the optimization of the above equation

1. In the first step I is calculated keeping D unchanged and,
2. In the second step D is calculated keeping I unchanged.

The above method is iteratively followed until a desired solution is obtained.

Fig. 1: Learnt dictionary; Training time 16.7s, using 65536 patches

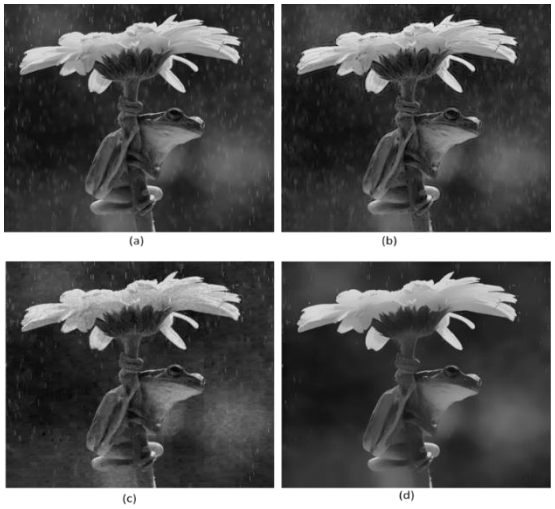


Fig. 2: (a) Reference Image (b) Bilateral Filtered Image (c) K-SVD filtered image (d) image filtered using our method

II. FRAMEWORK: BRIEF UNDERSTANDING

A. Finding a sparse solution: In most of the denoising problems using our approach our goal is to find the sparsest coefficient vector β of l_0 norm such that it minimizes the mean squared error as much as possible i.e. $\|D\beta_p - I_p\|_2^2 \leq \varepsilon^2$ and is unique in nature. But the complexity of pseudo-norm makes this task almost unsolvable or solvable with an indefinite amount of time. There are two types of basic approaches that are supported; the first one is called the relaxation method or The Basis Pursuit. In this method, the penalization factor zero of the pseudo-norm is replaced by one. This results in a convex problem which is solvable in nature at the same time avoid the time constraints. One such approach is least angle regression (LAR) or Stage wise LAR as in [Efron Bradley, 2004].

The second method is called greedy approach or matching pursuit (MP) [Mallat and Zhang, 1993]. This algorithm iterates by finding one atom at a time. Suppose I' is the image signal we are trying to approximate. The matching pursuit first traverses through the columns of the dictionary to look for the atom that is nearest to I'. In the next traversal, it searches the atom such that it minimizes the mean square error i.e. $\|D\beta - I'\|_2^2$. It continues to run until it finds the atoms which minimize the average squared error below a certain threshold determined by I. In an advanced version of MP the signal I' is projected over the entire set of atoms and an optimal solution is found by the method of least-squares which decreases the time consumption of the algorithm. It is named as Orthogonal Matching Pursuit (OMP).

III. CLUSTERING AND REGULARIZATION: AN OVERVIEW

A. Clustering: When the learning is to be done from a data set that is not labelled or classified it follows an *unsupervised learning* [Olshausen and Field, 1996] approach as the machine does not have any information about the training data or any prior relation between the noisy image and its denoised version. In such cases learning is achieved by grouping of data points or *clustering*. The data points are segregated based on similar features. Each group or cluster consists of an exemplar which best represents that particular cluster. In our case of image processing clustering algorithms are performed on the atoms. The grouping criteria of atoms is such that they contain same texture and edges is based on a similarity function;

$$s(a_i, a_j) = \left\|HOG(a_i) - HOG(a_j)\right\|^2 \qquad (3),$$

where HOG (·) is *Histogram Oriented Gradient* [Bossu, 2011] and describes the shape and context information of the atom. $s(a_i, a_j)$ describes the similarity between two atoms a_i and a_j which is calculated based on the negative mean squared

error. The task of clustering is accomplished by reducing the net similarity, S between the atoms:

$$S = \sum_{i=1}^{M}\sum_{j=1}^{M} c_{ij}\, s(a_i, a_j) - \gamma \sum_{i=1}^{M}(1 - c_{ii})\Big(\sum_{j=1}^{M} c_{ij}\Big) - \gamma \sum_{i=1}^{M}\Big|\Big(\sum_{j=1}^{M} c_{ij}\Big) - 1\Big| \quad (4)$$

B. Regularization: In machine learning, the methods of linear and logistic regression cause the problem of overfitting for high dimensional space. In simple term, the solution to the regression is so accurate that it reduces the learning efficiency and increases the complexity. When regularization in implemented on a set of data points, it adds some additional data to the set of data points hence improving the learning performance and reducing complexity. This is accomplished by regularization parameter as in (1). With the large value of γ models with high complexity are made redundant and with a low value of γ training errors are reduced. One such method of regularization is least angle regression. The LAR algorithm help in estimating which atoms to be used to get our response image. It is similar in steps to stepwise regression [Michael and Bishop 2012, Tomasi and Manduchi, 1998] but instead of including atop at every step, the approximated parameters are increased in a direction equiangular to each one's correlations with the residual.

C. Affinity Propagation: When it comes to clustering, affinity propagation is one of the most suitable algorithms because of its time and error minimization properties. Unlike conventional clustering algorithms like K-means or K-medoids, this approach does not require the number of clusters to be specified before the processing. Not only that this approach provides flexible optimization according to the needs of the user. As mentioned in Frey and Dueck [2007], affinity propagation constantly achieved lower error rate in more than two orders of time when compared to K-means.

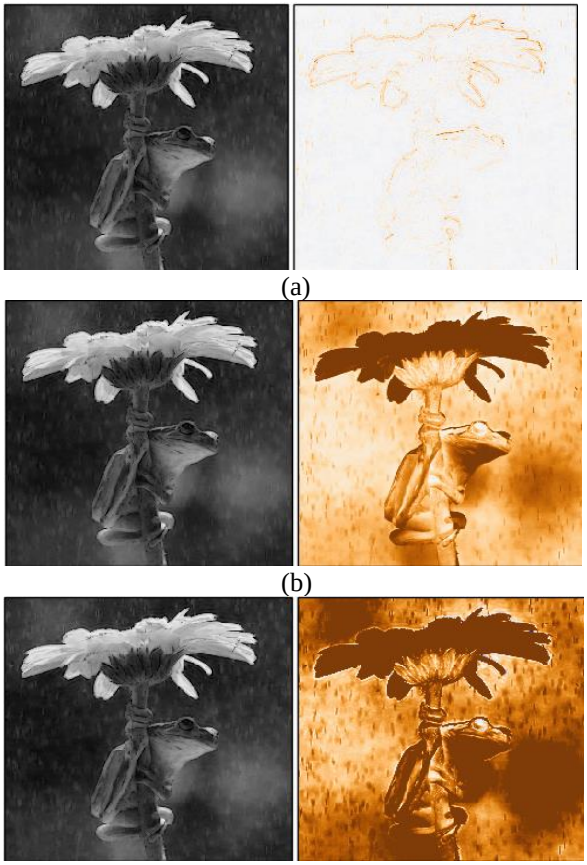
Consider a set of random data points from a_1 to a_n with each point equally potential of becoming an exemplar. Let s be similarity function as mentioned in (3). It represents the similarity or affinity of a point a_i to pointa_j. The optimization parameter or input preference in decided by the diagonal values of s i.e. s (i, i). Input preference should be chosen carefully as it shows the likelihood of a point to become exemplar and is an important factor in deciding the classes. The algorithm progresses by cycling between two message passing steps to upgrade a couple of matrices namely responsibility matrix R and availability matrix V.

1. The responsibility matrix R measures the candidature of a data point a_k to become an exemplar for a_i when compared to another point a_j in the neighborhood ofa_k.
2. The availability matrix V measures the fitness of a_i to choose a_k as its exemplar point when compared with the preferences of other neighborhood points.

In the beginning both the matrices are set to null value and the algorithm progresses through the following cyclic steps:

1) Updating of R: $r(i,k) \leftarrow s(i,k) - \max_{k' \neq k} \{a(i,k') + s(i,k')\}$.
2) Updating of V: $v(i,k) \leftarrow \min(0, r(k,k) + \sum_{i' \notin \{i,k\}} \max(o, r(i'k)))$ for i ≠ k and v(k,k) ← $\sum_{i' \notin \{i,k\}} \max(0, r(i',k))$ for i=k.

The above steps are repeated until no more changes occur in the matrices or for some fixed number of iterations.



(a) (b) (c)

Fig. 3 (a),(b),(c) OMP with 1 atom, 2 atoms and 5 atoms respectively

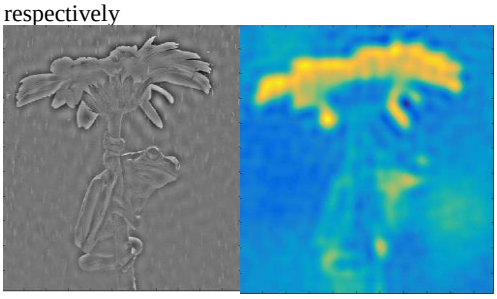


Fig. 4 (a) High frequency Image component (left) (b)Low frequency Image component (right)

IV. PROPOSED METHOD

1. We take a rainy image (refer Fig. 5), I and perform Discrete Cosine Transform (DCT) to split the image into two components namely high frequency component I_H and low frequency component I_L (refer Fig. 4).
2. From the empirical results we know that most of the structured noise is present in I_H, so we only take the high frequency component to the next stage while preserving the low frequency component.
3. For edge preservation window filtering is applied to the high frequency component.
4. The high frequency component is then given to dictionary learning; Fig. 1 for training the image data.
5. The next step is to apply a suitable clustering algorithm like affinity propagation [Mallat and Zhang, 1993] to the atoms of the dictionary to form K clusters of the high frequency component of the image.
6. Since this method is unsupervised, K is an unknown. Once the clusters are formed, process of image reconstruction is done. For K clusters, we get K high frequency image components i.e. I_H^1 toI_H^K.
7. Standard deviation is calculated for these image components to find the one with least deviation from the noisy image since that component will consist. Let it be I_H^L
8. Now I_H is constructed using K-1 components leavingI_H^L. Let it beI_H'.
9. Finally, to obtain the denoised image I_H' is added to its corresponding I_L to obtain I et al I=I_H'+I_L.

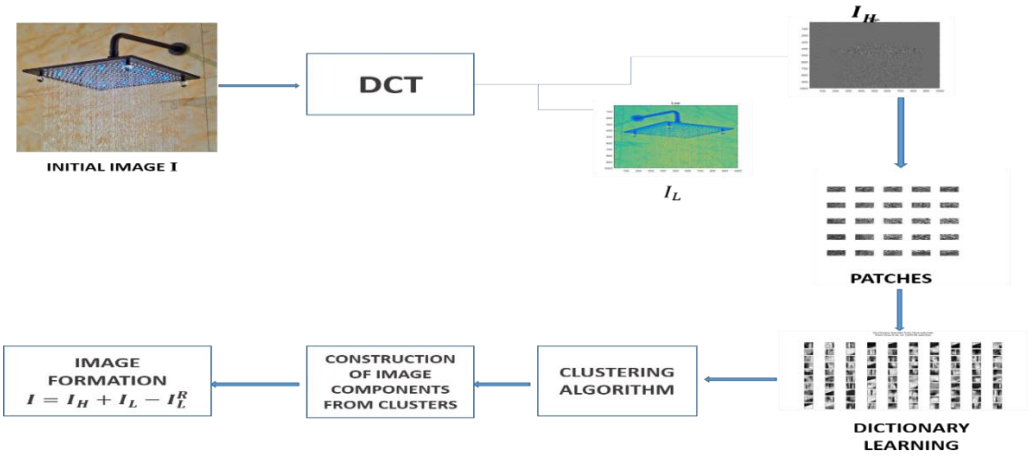


Fig. 5: Processing diagram for the proposed method for image denoising and rain removal

V. EXPERIMENTS CONDUCTED

1. First, we have used some basic decomposition and image denoising techniques using bilateral filtering [Tomasi and Manduchi, 1998], K-SVD [Elad and Aharon, 2006, Aharon, 2006] decomposition techniques (refer Fig. 2).
2. Dictionary learning was accomplished by means of LAR and OMP and affinity propagation was used for unsupervised learning. The results are displayed in Fig. 7 and Fig. 3 respectively.
3. The output response was tested for variable number of atoms and difference norm was calculated and compared. The corresponding results are plotted in Fig. 6
4. Various performance measures like Peak Signal to noise ratio were calculated and compared for different methods as in Table I and II.

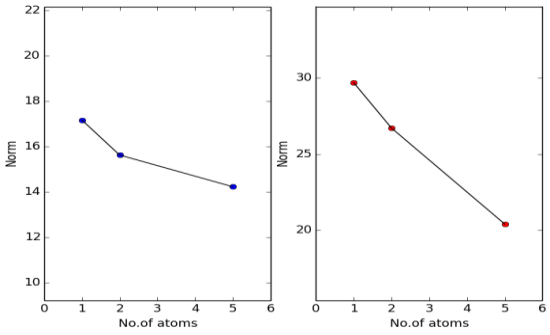
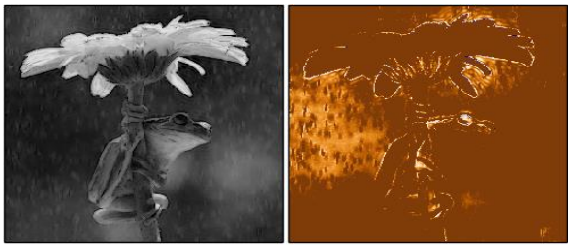
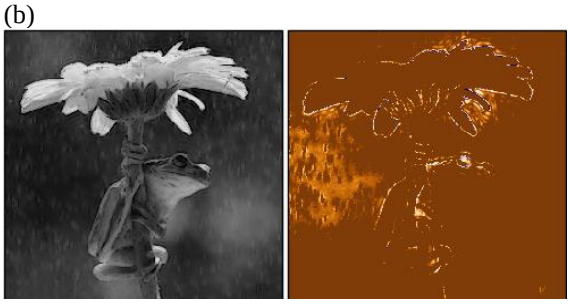


Fig. 6 Plot to depict change in norm difference with varying number of atoms

(a)



(b)



(c)

Fig. 7a, b and c: LAR with 1 atom, 2 atoms and 5 atoms respectively

VI. CONCLUSION

1. From the results displayed in Table I it is justified that OMP with 1 atom is most suitable for sparse coding of signal.
2. In Table II, though the PSNR values are high compared to OMP and LAR methods, the performance in terms of error is very less and the valuable image features are destroyed which is not desirable.

Thus, using the result displayed in Fig. 2, 3, 6 and 7 and the performance measures tabulated in Table I and II, it is reflected that the machine learning approach to image processing is advanced in terms of time complexity and accuracy. Also, it requires less human effort and is a more intelligent way for advanced signal processing.

VII. REFERENCES

Bruckstein A.M., D.L. Donoho and M. Elad, From sparse solutions of systems of equations to sparse modeling of signals and images. SIAM Rev. 51(1): 34–81 (2009).

Olshausen B.A. and D.J. Field, Emergence of simple-cell receptive field properties by learning a sparse code for natural images. Nature 381 (13): 607-609 (1996)

Frey B.J. and D. Dueck, Clustering by passing messages between data points. Science 315 (5814): 972-976 (2007).

Tomasi C. and R. Manduchi, Bilateral filtering for gray and color images. Proc. IEEE Int. Conference on Computer Vision Pp. 839-846 (1998).

Efron Bradley, Trevor Hastie, Iain Johnstone and Robert Tribshirani, Least angle regression. Annals of Statistics 32(2): 407-499 (2004).

Bossu, J., N. Hautiere and J.P. Tarel, Rain or snow detection in image sequences through use of a histogram of orientation od streaks. Int. J. Computer Vision 93(3): 348-367 (2011)

Mairal, J., F. Bach, J. Ponce and G. Sapiro, Online learning for matrix factorization and sparse coding. J. Machine Learning Res. 11: 19-60 (2010).

Mairal, J., F. Bach and J. Ponce, Task-driven dictionary learning. IEEE Trans. Pattern Anal. Mach. Itell. 34(4): 791-804 (2012).

Michael I.J. and C.M. Bishop, Neural networks. Allen B. Tucker, Computer Science Hand book 2nd Edition (Section VII: Intelligent Systems). Boca Raton, FL: Chapman & Hall/CRC Press LLC ISBN 1-58488-360-360-X (2012).

Aharon, M., M. Elad and A.M. Bruckstein, The K-SVD: An algorithm for designing of over complete dictionaries for sparse representation. IEEE Trans. Signal Processing 54(11): 4311-4322 (2006).

Elad M. and M. Aharon, Image denoising via sparse and redundant representations over learned dictionaries. IEEE Trans. Image Pro-cessing 15(12): 3736-3745 (2006).

Mallat S. and Z. Zhang, Matching pursuits with time-frequency dictionaries. IEEE Trans. Signal Processing 41(12): 3397-3415 (1993).