# AN EFFICIENT DOCUMENT IMAGE RECONSTRUCTION AND BINARIZATION METHOD USING COMPRESSED SENSING

J V Satyanarayana, A G Ramakrishnan
*Indian Institute of Science*
*Bangalore – India*

**ABSTRACT**

In this paper, we present a novel approach of document image acquisition with reduced number of measurements, requiring lesser number of sensing elements. Reduction in the number of sensors also directly implies faster acquisition. Our approach is built around the Compressed Sensing paradigm, in which signals which are sparse can be captured and reconstructed with considerably lesser number of measurements than required by Nyquist sensing. Signal reconstruction algorithms used in Compressed Sensing are computationally intensive. A brief survey of some of the popular reconstruction algorithms is also given in the paper. In the algorithm proposed in this work, the reconstruction step associated with recovering the binary image from the Compressed Sensing measurements is pruned to take advantage of the inherent requirement of the binarization of the acquired image. Binarization of captured documents which is an essential step for post processing is built into the proposed scheme.

**KEYWORDS**

Document binarization, Compressed sensing

## 1. INTRODUCTION

Document analysis [1] is an important part of many common applications like OCR, page layout analysis of forms, digitization of official documents and library archives. The first step in Document analysis is the optical scanning to acquire an array of pixel values. When the document is in a medium unfit for scanning, the image is acquired through photographic methods. Whatever be the mode of acquisition, it is required to pre-process the raw data to be able to extract information.

Our interest in this work is confined to image acquisition and binarization steps of preprocessing. Binarization involves selection of a suitable intensity threshold on the basis of which the background and foreground information is separated, thereby, resulting in an image having only two intensity levels. In the context of document images we make two important observations which are of relevance to the acquisition process:

a) Documents are reasonably sparse in that the amount of useful foreground information is considerably lesser than the background information.
b) Since binarization is usually an unavoidable step in document analysis, a coarse measurement of the pixel intensity is sufficient during acquisition of the image in order to be able to classify the pixel as belonging to foreground or background. This, however, assumes that the histogram of the image intensities has the peaks corresponding to the foreground and background pixels not very close to each other, which is not a very unrealistic assumption in the case of document images.

By virtue of a) it is possible to capture the image by making substantially lesser number of measurements under a scheme called Compressed Sensing (CS). On account of b), the number of measurements required is further reduced when compared with that required for reconstruction of gray scale images. The consequence is an improvement in the speed of acquisition making the whole scheme suitable for systems where the image acquisition and binarization have to be done in soft real time.

Section 2 gives a brief introduction to the Compressed Sensing paradigm. Section 3 is a survey of the most popular algorithms for reconstructing the image from the CS measurements. Section 4 describes the proposed algorithm. The results on a test image are commented upon in Section 5.

## 2. COMPRESSED SENSING PARADIGM

The basic idea of Compressed Sensing is that instead of transforming a signal into a sparse basis and then reconstructing the original signal from a subset consisting of the significant basis coefficients, it is meaningful to make only limited measurements in the form of projections which are incoherent with the basis vectors and reconstruct the original signal. A signal $\boldsymbol{x_0} \in \mathbb{R}^{N\times 1}$ is said to be sparse on a basis consisting of vectors made of the rows of an $N \times N$ matrix $\boldsymbol{\Psi}$**,** if $\boldsymbol{\Psi x_0}$ is sparse, that is, $S = |\boldsymbol{\Psi x_0}|_0 \ll N$.

According to Classical Compressed Sensing theory [2, 3], if there exists a $K \times N$ matrix $\boldsymbol{\Phi}$ incoherent with $\boldsymbol{\Psi}$, where $K \ll N$ then it is possible to reconstruct $\boldsymbol{x_0}$ from the $K$ measurements

$$\mathbf{y} = \boldsymbol{\Phi\Psi x_0} \tag{1}$$

A common choice [2] of $\boldsymbol{\Phi}$ is a matrix consisting of entries from $\mathcal{N}(0,1)$ . In such a setting, with

$$K \geq CSlog(N/S) \tag{2}$$

it is possible to recover $\boldsymbol{x_0}$ accurately and sometimes exactly with the help of a host of recovery algorithms that have been proposed by various researchers. The next section brings out some of the popular reconstruction algorithms. If the entries of $\boldsymbol{\Phi}$ are chosen from the Gaussian distribution as has been done in this work, the signal can be reconstructed accurately [4], if the value of C is between 3 and 5. For our simulations we have taken C as equal to 4.

We do not present the hardware setup for making CS measurements on an image. [5] gives a detailed description of one of the ways in which the image acquisition setup can be configured employing a ‘Single pixel camera’. The focus of our work is on the reconstruction of the image from the CS measurements.

## 3. RECONSTRUCTION ALGORITHMS

In this section, we present our search for a suitable CS recovery algorithm for document images. Since, subsequent to acquisition, the document has to be anyway binarized for further analysis, we need to recover no better than the approximate intensity values of only the foreground pixels.

Signal recovery can be achieved by solving an optimization problem which finds out the sparsest signal that agrees with the measurements. In other words, the signal $\mathbf{x}$ is the solution to the $\mathbf{l_0}$ minimization,

$$\min\|\mathbf{\Psi x}\|_0 \text{ subject to } \mathbf{\Phi\Psi x} = \mathbf{y} \tag{3}$$

where $\|.\|$ indicates the number of non-zero elements or the $\mathbf{l_0}$ norm. The signal can be reconstructed with overwhelming probability as long as the number of measurements is sufficient in accordance with (2). However, this minimization is a combinatorial search and is computationally intractable when the problem dimension is large. The problem is known to be NP-hard [6]. The problem can be simplified to give a nearly good solution by solving $\mathbf{l_1}$ minimization problem,

$$\min\|\mathbf{\Psi x}\|_1 \text{ subject to } \mathbf{\Phi\Psi x} = \mathbf{y} \tag{4}$$

as long as $\boldsymbol{\Phi}$ satisfies the Restricted Isometry Property (RIP [7,8]). This optimization also known as Basis Pursuit [9], is considerably easier to solve using linear programming techniques.

While linear programming techniques promise exact recovery, with their $\Omega(\mathrm{N}^3)$ complexity they are not fast enough for problems with large dimension. Iterative greedy algorithms have been proposed, which albeit at the cost of additional measurements, can achieve a fairly accurate of the signal.

The Orthogonal Matching Pursuit (OMP) [10] is a greedy, iterative algorithm which finds the support of the sparse signal $\mathbf{x_0}$ progressively. The $\mathrm{K} \times N$ matrix $\boldsymbol{\Phi}$ is not an isometry when $\mathrm{K} < N$. However, when the entries of $\boldsymbol{\Phi}$ are random as considered in Section 3, the columns of $\boldsymbol{\Phi}$ are approximately orthogonal and the observation vector $\mathbf{v} = \boldsymbol{\Phi}^*\mathbf{y}$ is a good approximation to the original signal vector $\mathbf{x_0}$. The biggest coordinate in magnitude of $\mathbf{v}$ is thus a non-zero coordinate of $\mathbf{x_0}$, and is therefore a support of $\mathbf{x_0}$. The support is then found out progressively in steps. Many variants of the OMP algorithm have been suggested in the CS community. The Stage-wise Orthogonal Matching Pursuit algorithm (StOMP) that has been proposed by Donoho et.al [11] is an improvised version of OMP wherein each iteration recovers more than one coefficient. Further, the number of iterations is fixed and the authors recommend a value of 10. As shown by the authors, StOMP can recover the signal in $N\log N$ complexity.

The Tree Matching Pursuit algorithm (TMP)[12], proposes an algorithm that enables fast recovery of piecewise smooth signals, that have a distinct "connected tree" structure in the wavelet domain. The Chaining Pursuit algorithm [13] proposed by Gilbert et al seeks to approximate the signal as a set of spikes.. In each pass, the algorithm recovers a constant fraction of the remaining spikes. It then updates the residual – the difference between the given signal and the superposition of the recovered spikes. CP gives an excellent performance when the sparsity ratio of the signal is very high.

The method bases on Sudocodes proposed in [14] requires M = O(K log(N)) measurements for exact reconstruction with a worst case complexity of O(K log(K) log(N)). A method based on group testing has been proposed in [15, 16] in which subsets of signal coefficients containing at most one significant coefficient are considered from which the position and value of the coefficient are extracted using Hamming code.

Most of the methods in the survey given above and others in the field attempt to extract all the coefficients and that too with their exact values. While this is high desirable for the acquisition of a signal, there is no provision wherein the relaxation of requiring only a few non-zero coefficients and that too with limited accuracy sufficient for binarization, is rewarded in terms of either faster reconstruction or smaller set of necessary measurements. We propose in the next section an algorithm a variant of OMP, which is suitable for acquiring document images.

# 4. PRUNED ORTHOGONAL MATCHING PURSUIT ALGORITHM

Our algorithm is based on Orthogonal Matching Pursuit (OMP) for reconstruction of the image from the measurements. Let $\mathbf{I} \in \mathbb{R}^{M\times N}$ be the image of size $M$ by $N$ pixels. The image is captured row-wise, which is quite often the case with many image acquisition systems. Further, the algorithm assumes that the image consists of a light foreground on a dark background, that is foreground intensities are higher. The algorithm can be applied to images which have a dark foreground on a light background after an inversion. The output of the algorithm can again be inverted to get back the correct binary image. The measurement vector is $\mathbf{y_r} = \mathbf{\Phi I_r}$ where $\mathbf{I_r}$ $(1 \le r \le M)$ is the r$^{th}$ row of $\mathbf{I}$ and $\mathbf{\Phi}$ is the measurement matrix introduced in Section 2. The threshold for binarization is $T$. A global threshold is chosen because the acquisition and reconstruction go hand in hand. The reconstruction is carried out one row after the other from the corresponding measurement vector $\mathbf{y_r}$ as detailed by the following algorithm.

ALGORITHM

1. Let $r$ denote the row count. For row $r$, set the initial solution, $\hat{\mathbf{I}}_r^{(0)} = \mathbf{0}$ and the initial residual $\mathbf{res}^{(0)} = \mathbf{y_r}$ and the output binary vector, $\hat{\mathbf{B}}_r^{(0)}$ to zero.
2. Let $J_S$ denote the set consisting of the indices of foreground pixels of $\mathbf{I_r}$ selected until the stage $S$. Initially, that is when $S = 0$, the set $J_0$ is empty. The residual in the $S^{th}$ stage is subject to matched filtering giving a vector of residual correlations.
$$\mathbf{v}_S = \mathbf{\Phi}^*\mathbf{res}^{(S)} \tag{5}$$
3. The index $i$ which corresponds to the maximum magnitude element in $\mathbf{v}_S$ is selected, that is.
$$i = \{j: |v_S(j)| > |v_S(k)|, \forall k \neq j\}. \tag{6}$$
4. The set $J_S$ is updated.
$$J_S = J_{S-1} \cup i \tag{7}$$
5. The new approximation $\hat{\mathbf{I}}_r^{(S)}$, to $\mathbf{I_r}$ with support in the set $J_S$ is found by solving the least squares problem,
$$\hat{\mathbf{I}}_r^{(S)} = \mathrm{argmin}_{z\in\mathbb{R}^{J_S}} \|\mathbf{y_r} - \mathbf{\Phi z}\|_2. \tag{8}$$
In other words, we seek to determine a vector supported in the elements of the set $J_S$, which is closest to the measurement $\mathbf{y_r}$ in the least squares sense.

6. As a consequence of the previous step, if the newly found element of $\hat{I}_r^{(S)}$, that is, $\hat{I}_r^{(S)}(i)$ is lesser than the binarization threshold $T$, then all the new elements added in the subsequent iterations will also be lesser than the threshold and will correspond to background pixels. In this case, the reconstruction of the binary vector corresponding to the row $r$ is complete, hence can be terminated. The loop is then exited and $\hat{I}_r^{(S)}$ is binarized using the threshold. The resulting binary vector $\hat{\mathbf{B}}_{\mathbf{r}}$, corresponds to row $\mathbf{r}$.
7. If the condition in Step 6 is not satisfied, the residual is updated.

$$\mathbf{res}^{(S)} = \mathbf{y_r} - \mathbf{\Phi}\hat{\mathbf{I}}_r^{(S)} \tag{9}$$

This eliminates the contribution to the residual correlations $\mathbf{v_S}$ due to the foreground pixels already selected in $\mathrm{J_S}$ . The algorithm then starts a new iteration from step 2. The procedure is repeated for all the rows $1 \leq \mathrm{r} \leq \mathrm{M}$. The final output of the procedure is the binary array $\hat{\boldsymbol{B}}$. The following points concerning the algorithm are noteworthy:

a) The premature termination of the algorithm which takes place if the condition in step 6 is satisfied is what gives it an advantage with respect to the execution time as compared to the case where a gray scale image has to be recovered using OMP. The essence of this step is that it is not required to reconstruct all the non-zero pixel values, only those which lie above the threshold bear significance.
b) It is not essential to reconstruct each sample very accurately. It suffices to know if the pixel has a value above or below the threshold. This advocates the possibility of reduction in the number of measurements

## 5. RESULTS ON A TEST IMAGE

We put our algorithm to experiment on the image shown in Figure 1. This image is the signature of Louis XIV of France downloaded from Internet. The size of the image is $370 \times 590$. In order to apply our algorithm, we invert the original image to have a light foreground on a dark background. The measurement matrix $\mathbf{\Phi} \in \mathbb{R}^{K \times N}$ where $K = 308$ and $N = 590$ and the entries drawn from $\mathcal{N}(0,1)$. Also, $M$ the total number of rows is 370. In other words, only 308 measurements are made for each row. The threshold for binarization was taken as the centre gray level of 127. The global approach to binarization is not a serious limitation of the proposed algorithm for images in which the histogram has the peaks corresponding to foreground and background pixels wide apart. The result of reconstruction is shown in Figure 2. Some of the unwanted artifacts and gaps in the reconstructed image can be removed through post-processing.

## 6. CONCLUSION

A method has been proposed to recover the original image sampled through Compressed Sensing measurements. The method takes advantage of the fact that the pixel intensities of background pixels need not be found out as any way the resulting binary image will have a zero at those locations. Further, the Compressed sensing setup is benefitted by the fact that the intensity values need be found only to that accuracy where one can determine on which side of the threshold they lie. The method is suited for applications where the image pixels are captured sequentially one row after the other and the reconstruction of the image is done simultaneously. This is the case in applications like handheld portable scanners, barcode readers etc. The output binary image is constructed as the image acquisition is in progress.

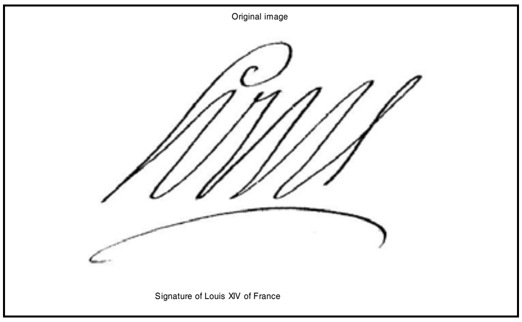


Figure 1. Original image of size 370 x 590 pixels measurements

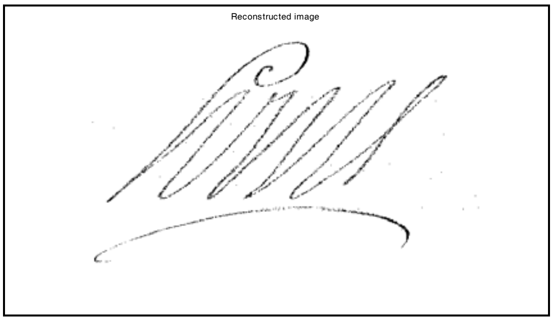


Figure 2. Image reconstructed by taking only 308 per row

# REFERENCES

[1] J. Schurmann et al, 1992. "Document Analysis – From Pixels to Contents", *Proceedings of the IEEE*, Vol. 80, No. 7, July 1992.

[2] Emmanuel Candès, 2006. "Compressive sampling", *Int. Congress of Mathematics*, 3, Madrid, Spain, pp. 1433-1452,.

[3] David Donoho, 2006. "Compressed sensing", *IEEE Trans. on Information Theory*, 52(4), pp. 1289 - 1306.

[4] E. Candès and J. Romberg, 2005. "Practical signal recovery from random projections," in *Proc. SPIE Int. Symp. Electron. Imag.*, pp.76–86.

[5] M.F. Duarte et al, March 2008. "Single Pixel Imaging via Compressive Sampling", *IEEE Signal Processing Magazine,* Special issue on Compressive Sampling. Vol. 25, No. 2, pp. 83 - 91

[6] M. Rudelson and R. Vershynin, 2005. "*Geometric* approach to error correcting codes and reconstruction of signals, Available on ArXiV preprint server: math.FA/0502299

[7] Candès, E. J., Tao, T., 2004. "Near-optimal signal recovery from random projections and universal encoding strategies*", IEEE Trans. Inform. Theory*. 52(12), pp. 5406 – 5425.

[8] Candès, E. J., Tao, T.,2005. "Decoding by linear programming*". IEEE Trans. Inform. Theory* 51(12), pp. 4203 - 4215.

[9] Scott Shaobing Chen, David L. Donoho, Michael A. Saunders, 1998. "Atomic Decomposition by basis pursuit", *SIAM Journal on Scientific Computing*.

[10] Joel Tropp and Anna Gilbert, 2007. "Signal recovery from random measurements via orthogonal matching pursuit". *IEEE Trans. on Information Theory*, 53(12) pp. 4655-4666.

[11] D. L. Donoho et al, 2006. "Sparse solution of undetermined linear equations by stagewise orthogonal matching pursuit", Preprint 2007

[12] M. F. Duarte, M. B. Wakin, and R. G. Baraniuk, 2005. "Fast reconstruction of piecewise smooth signals from random projections," in *Proc. SPARS05*, Rennes, France.

[13] A. Gilbert et al, 2006, "Algorithmic linear dimension reduction in the L1 norm for sparse vectors", submitted. Conference version in: Algorithmic Linear Dimension Reduction in the L1 Norm for Sparse Vectors, Allerton (44th *Annual Allerton Conference on Communication, Control, and Computing*).

[14] S. Sarvotham, D. Baron, and R. G. Baraniuk, 2006. "Sudocodes – fast measurement and reconstruction of sparse signals," in *Proc. International Symposium Information Theory (ISIT2006),* Seattle, WA, July 2006.

[15] G. Cormode and S. Muthukrishnan, 2005. "Towards an algorithmic theory of compressed sensing" DIMACS Technical Report TR 2005-25.

[16] G. Cormode and S. Muthukrishnan, 2005. "Combinatorial algorithms for compressed sensing" DIMACS Technical Report TR 2005-40.