A New Method to Medical MRI Images Restoration with Swarm Intelligence

Mehdi Zekriyapanah Gashti, Rouhollah Habibey

Abstract—Due to the limited speed of sensors in MRI imaging, the sampling at Nyquist rate will result in elongation of the imaging time. This causes the patient's discomfort, motion induced geometric deformities, and thus, reduces the image quality. In this study, we provided a new method for reducing the image noise, in which the signal sparse representation was used to restore the degraded and noisy areas.The particle swarm optimization was also used to improve the accuracy of the sparse representation. The simulation results indicated that the proposed method has a higher efficiency than most of the popular noise removal methods both in terms of PSNR (Peak signal-to-noise ratio) parameters, MSE (Mean Square Error) and the image quality. It is also a more powerful approach in retrieving subtleties and details of the image than the most available prominent noise removal methods.

Keywords—Image Restoration, Noise Reduction Sparse Representation, Swarm Intelligence, Peak Signal-to-Noise Ratio

I. Introduction

Magnetic Resonance Imaging (MRI) is nowadays used as a common imaging modality in medical diagnostics and research projects. MRI is a non-invasive imaging method that does not use the radiation properties unlike other modalities such as Computerized Tomography, (CT). Also, unlike Positron Emission Tomography (PET) does not require the use of radioactive markers. MRI imaging is one of the important modalities in medical imaging due to high contrast in soft tissues. There is usually a compromise between imaging time, resolution, and the Signal to Noise Ratio (SNR) levels in the MRI imaging.

The acquisition of a one-dimensional MRI image was reported by Herman Carr in 1950. Working on previous research, Paul Lauterbur, an American chemist, succeeded in devising methods to produce two-dimensional and three-dimensional MRI images. Eventually, he released the first image taken based on the Nuclear magnetic Resonance (NMR) from a live rat in 1973. On the other hand, important research and developments were made for the first time at the University of Nottingham in England by Peter Mansfield in the field of imaging based on the NMR. By expanding a mathematical approach, the prominent physicist, Peter Mansfield, managed to reduce the imaging time and enhance the quality of the images compared to the method used by Paul Lauterbur. MRI was invented in the early 1970's, but the first MRI imaging devices were introduced to the market ten years later. Finally, the 2003 Nobel Prize in Medicine was awarded to Paul Lauterbur and Peter Mansfield from England for the invention of MRI [1].

The basis of MRI is the spinning motion of specific nuclei in the living tissues. This spin is resulted from individual spins of protons and neutrons inside the nucleus. The subatomic particles pairs automatically do spinning movement in opposite directions, but with the same speed. In nuclei with an even mass number, i.e., equal number of protons and neutrons, a half of the spins are in one direction and the other half are in the opposite direction; therefore, the nucleus itself does not have a pure spinning motion. In nuclei with an odd mass number, that is, where the number of neutrons is slightly more or less than the number of protons, the spin directions are not the same and the opposite. Thus, the nucleus, itself, has a pure spin or angular momentum. These are called active MR nuclei. The active MR nuclei are indicated by arranging their axis of rotation in the direction of an external magnetic field. This arrangement occurs because the active MR cores (nuclei) have an angular momentum or spin. The major active nuclei are hydrogen, carbon, nitrogen, oxygen, fluorine, sodium and phosphorus [2].

Applying a radio excitation field of B_1 to the polarization vector applies moment to it and deviates it and creates the magnetic component of the M_{xy}, which is perpendicular to the field. This magnetic component rotates by a frequency of $f_0 = \frac{\gamma}{2\pi} B_0$. In this equation, B_0 is the intensity of the static magnetic field and $\frac{\gamma}{2\pi}$ is a constant coefficient equal to $42.57^{\mathrm{MHz/T}}$. For example, a typical 15T MRI device has a frequency about 60^{MHz}. The vertical component of this vector of magnetization produces a signal, which can be received by a coil. This component represents many of the tissue characteristics. The proton density in the tissue is one of the tangible features that can be displayed in it. In fact, the signal that we are looking for in the MRI is the same component, which is an image of the spatial distribution of the vertical magnetization vector [3].

Creating an MR image usually requires the gathering of a set of information, which is called collection. At each collection stage, an RF stimulation creates a new vertical

Mehdi Zekriyapanah Gashti is with the Department of Computer Engineering, Payame Noor University, Tehran, Iran.
Rouhollah Habibey is with the IUF – Leibniz Research Institute, for Environmental Medicine, Düsseldorf, Germany.

magnetization, which is then sampled on a specific path in the space k [4].

In general, a complete MR image can be made with a single retrieval on a path that travels the total space k [5]. This state is most commonly used in applications such as brain activity imaging. However, these results will result in an image with inadequate resolution and sharp image artifacts for many applications. The magnetization vector is exponentially damped with time, which limits the effective collection time period. Though, the function of the gradient system and the physiological limitations limit the speed by which the space k is traveled. The combination of these two effects reduces the total number of symbols per collection product. Consequently, most MRI imaging use a sequence of restorations that each of them samples a part of the k space. Then, the information obtained is developed from a sequence of restorations to create the image.

MRI is a non-invasive method and does not use ions radiation properties unlike other imaging methods such as CT. The long time required to record Echo Resonance has led to performing many studies to speed up the imaging. Image acquisition in medical imaging devices is obtained during a process called "Image Reconstruction". Image reconstruction in MRI is slightly different from image reconstruction in other medical imaging equipment, which is due to different functional physics in it compared to other equipment.

II. Noise removal in digital images

With the increasing spread of various techniques for getting widespread information, the image processing has been widely used nowadays. However, the images resulting from the image signal producing devices always contain some noise and distortion, which reduces the image resolution. The set of operations and methods used to reduce defects and increase the quality of the image, or to recognize and compress it are called image processing [6].

Image processing involves a wide range of areas of work. However, in general, the attention has focused on four areas of the Apparent Image Enhancement Quality, Damaged Image Restoration, Image Compression, Encryption and Image Understanding by the machine. Meanwhile, the set of techniques used to reduce noise with a special and distinguished place in image quality improvement include methods such as the use of fading filter and contrast adjustment to improve the appearance and ensure the proper image representation in the destination environment [7].

Generally noise is present everywhere, and wherever a signal is measured, a noise will certainly be created on it. The minor changes are those not from the original image. Noise in the image is created in the process of image signal production by sensors or digital cameras circuits. These environment changes are created unwantedly and are unfavorable I users' eye. The noise amplitude involves a range from a good resolution image to an image that is almost completely destroyed. Every high-quality and precise experiment done in the physics world needs a lot of accuracy and precision to predict the noise of the environment and reduce its impact. The significance of the noise analysis appears entirely when the quality of the signal measured is not determined by the absolute value of the signal energy, but is determined from the signal-to-noise ratio. The research results show that the best approach to improve the signal-to-noise ratio is to reduce noise rather than to increase the signal strength. Random noise is uncontrollable by definition and its exact amount varies in different experiments [8].

Removing and reducing noise is seen as one of the most important issues in the field of image processing. An image contaminated by environmental noise has an inappropriate visual quality and is not suitable for analysis and understanding by the user. In addition, many common processing applications, including Edge Detection, Segmentation and Machine Vision are impaired in the presence of noise [9]. Thus, the removal of the noise effect added to the image seems vital in all processing areas.

Different perspectives have been proposed to remove the noise effect from the image signal, which are very diverse depending on the type and density of the noise added. A filter used to remove Gaussian noise may have not a proper result in removing the salt-and-pepper noise. On the other hand, the use of a strong filter to reduce a low-density noise will degrade the visual quality rather than just improving the image. The filters used to remove noise effects from the image signal in general are divided into two categories of spatial area filters and frequency area filters.

The spatial field filters performance is based on processing on the pixel brightness level surface related to the adjacent pixels. The processing operation may focus only on the pixel brightness level. In this case, the operation will be performed bit by bit. Also, the pixels neighboring with the original pixel can be used to remove the noise effect. The second method becomes feasible based on a large similarity between the brightness levels of the images adjacent pixels [9]. Some of the most important commonly used filters in the time domain include the mean filter, median filter, adaptive median filter, maximum and minimum filters.

Contrary to the location (spatial) filters, which deal with the brightness level of one or more pixels of the image, the transmission of the image into the frequency domain by taking the two-dimensional Fourier transform of the image signal will simplify the removal of the noises that have damaged the image in a limited frequency band. In the case of adding the collective environmental noise to the image, the frequency domain filters fall into 3 categories: Band-stop filters, Band-pass filters, and Notch Filters [9].

III. The proposed method

The proposed method of repairing medical images is based on the dictionary learning approach. In the following, the issue of choosing the proper dictionary for atomic decomposition applications and its importance was discussed. Then, using the K-SVD algorithm, the dictionary learning problem was explained by employing a number of training signals. In the next step, an efficient algorithm was provided to achieve the most sparse signal representation (signal sparse coding) by using the Particle Swarm Optimization (PSO) method.

A. Step 1. Selecting the dictionary

In recent years, the sparse representation of data has been extensively used for applications such as sampling, compression, representation, retrieval and classification of images [10]. The success of the sparse representation in these applications results from the fact that most natural signals such as image or sound have sparse representation by considering specific bases. Natural signals often do not cover the entire space and are placed on a subspace of manifold.

Consider the signal of $x= [x_1,x_2,\ldots,x_n]^T \in R^n$. This signal can specify the pixels of an image. The assumption of sparsity states that the data can be expressed as a linear combination of few bases already considered. If we show the number of these nodes effective in the representation of x with k, this number should be significantly smaller than the data original space dimension ($k << n$).

The bases used to display the data are placed in a matrix. This matrix is called the Dictionary Matrix. This matrix generally covers the entire vector space related to the data. In other words, the columns of this matrix are used to construct the whole or a part of the data belonging to vector space of R^n. Each of the dictionary matrix columns, which are the bases required to model the data, is called an Atom. If the number of the dictionary atoms is as large as the vector space dimension and the bases cover the whole space, the matrix of the dictionary is called Complete. In this case, each data will have a unique representation by the dictionary atoms. For example, in the Fourier transform, the transform bases are constant and perpendicular to each other and each data will have its own specific coefficients. If the dictionary matrix is complete, the representation of each data by this matrix will be unique and not necessarily sparse. If the number of dictionary atoms is increased to be more than the complete state and a number of atoms are added to it, then, the dictionary is called over-complete in this case. If D represents the dictionary, in the relation $x=D\alpha$, the x data is written by the linear composition of the atoms and α shows this linear combination. Since the matrix is over-complete, the linear equation system will be underdetermined for determining the undefined α. This system will have countless answers that the answers are located on a vector space. We add another condition since we are looking for the sparsest representation of the existing representations. A constraint should be added to the problem to limit the number of non-zero entries of the sparse representation.

The sparse representation of signal x is shown in relation 1. In this equation, α is a vector that contains the signal sparse representation.

$$\alpha = [x_1, x_2, \ldots, x_k]^T \in R^k$$

The norm zero, $||.||_0$ counts the number of non-zero elements of the matrix in this relation. This number indicates the sparsity of the representation.

$$\text{argmin} \left||\alpha|\right| \qquad \text{s.t} \quad x = D\alpha \qquad (1)$$

The relation 1, also called P0 problem, can be expressed in other forms as well. The sparsity rate can be also stated in the form of a constraint. In relation 2, the sparsity is described in the form of a constraint and the purpose of optimizing is to reduce the error between reconstructing the signal with sparse representation and the initial signal.

$$\text{argmin} \left||x - D\alpha|\right| \qquad \text{s.t} \quad ||\alpha||_0 \le K \qquad (2)$$

Solving problem 2 in general requires solving an NP-HARD problem. Many greedy methods have been used to solve this problem. The most important and, at the same time, the easiest method available is Orthogonal Matching Pursuit (OMP) method, which has been used in the proposed method. In the OMP method, we look for a maximum K vector of the dictionary vectors set, which can display the initial signal with the lowest error. For this case, in K steps, the best vector and the proper coefficient for making the signal is chosen correspondingly step by step. Theoretically, one can provide a guarantee in certain modes for the answer given by the OMP method that would be the sparsest possible answer. These conditions depend on the dictionary matrix. The OMP method is presented to obtain the sparse representation of the simplest method. However, this method does not find the optimal sparse answer in many cases. Many methods have been provided based on the OMP and later on. But a problem with all these methods is that they solve the NP-HARD problem in a greedy way. In fact, they do not consider all possible modes for using the atoms.

Converting the problem into a convex optimization ensures the achievement of the sparsest answer using the convex optimization methods such as the Gradient Descent method. However, solving this problem and obtaining a precise answer will be very time consuming. The implementation time of this problem is one of the main challenges in the Compressed Sensing (CS) area. This time is highly important in our particular problem of classifying the images since the number of images and pieces of the image are very large and the sparse representation should be calculated for each of them.

By considering the run-time, the greedy methods still are superior to the convex optimization methods. But approximately solving the problem is another approach. Instead of exact solving of the problem (P1), another approximation problem can be defined that will have a close answer to the original answer. The relation 3 presents the approximate form of the problem (P1). In this relation, the parameter δ is a small value considered to display the difference between the data representation and the data itself.

$$\text{argmin} \; ||\alpha||_1 \qquad \text{s.t} \quad ||x - D\alpha||_2 = \delta \quad (3)$$

Relation 3 is also used to model the data noise. If we assume a Gaussian noise with a specific and small variance for the data, an appropriate response will be obtained for δ by optimizing this equation.

As stated, one of the main challenges in the area of signal processing is finding a model to represent the signal in an appropriate form by considering the objective of the problem. The easiest solution is to use the linear combination of data to get a new representation. This method is used in the Principle Component Analysis (PCA). In this method, we assume that the data has been generated by a Gaussian process. This assumption is not always true for natural data and some processes often play a role in generating the signals. The

Independent Component Analysis (ICA) method is one of the approaches provided to solve this problem. It is theoretically proven that the method is not effective enough in analyzing the data generated by a large number of processes since in both of these methods; the maximum number of independent components is limited and equal to the size of the data space dimension.

The natural signals are caused by different factors that few of them participate in the generating of each signal. Therefore, the number of constructive bases for these signals should be considered more than the space dimension. Also, the basic assumption is that the number of constructors is limited for each signal. From another perspective, the data in the previous methods are placed on a subspace. This assumption limits the space of the signals learned by the model. Observing the effectiveness of the mentioned methods, a question arises that whether a model can be provided for the data placed on more sub-spaces. Choosing these subspaces also plays an important role in displaying the data. In the dictionary learning, we look for atoms that create subspaces for displaying the data. Finding this dictionary is a very complicated task in some cases. For example, consider the image pieces. These data cannot be displayed with a linear combination and on a linear sub-space. The glossary should be determined in such way to solve these problems that the sparsity constraint is established in the representation of all training data.

Obtaining a definitive answer to the problem of learning is not possible. However, a local optimal answer can be achieved by a two-step repetitive method. In the first step, we assume that the dictionary is constant and solve the optimization problem according to the values of the sparse vectors. This stage is typically called sparse coding. In the second step, the vectors are fixed and the dictionary is changed to minimize the optimization phrase. The gradient descent method can be used for the second step. These two steps are repeated consecutively until the dictionary is converged. As the sparse coding step is very time-consuming in case of large vectors, the pieces of image used for learning are usually selected with small sizes (up to 32 × 32).

B. Step 2. The dictionary learning

The problem of dictionary learning is, in fact, very similar to Vector Quantization (VQ) problem. In the proposed method, we want to display a large number of vectors by atoms with fewer numbers. This task is indeed the equivalent of data clustering since each vector will belong to one of the atoms and the goal is to minimize the total distances of vectors and their corresponding atoms. The K-Means method is presented to solve this problem, which is one of the most successful and widely used methods for data clustering [11]. Similarly, the objective in the dictionary learning is to find atoms for the dictionary. Instead of an atom, a linear combination of a limited number of atoms can be used to display each individual vector of the data. A solution similar to the K-Means approach provided to solve this problem was similar to previous methods of the dictionary learning like MOD with the difference that in the updating stage, the dictionary changes the atoms and the coefficients of their use. At this stage, the Singular Value Decomposition (SVD) is used. Thus, this algorithm is called K-SVD. Following the process, the K-SVD algorithm and the efficient optimal encoding method, which is one of the successful methods in solving a norm 1 problem, were presented in the proposed method of the K-SVD algorithm for dictionary learning. This algorithm updates the dictionary step-by-step to eventually converge to the optimal answer with a proper number of implementation of the steps. The K-SVD algorithm consists of two phases that are run in succession. The first phase is sparse signal coding, while the second phase updates the dictionary atoms based on the codes obtained. The Particle Swarm Optimization (PSO) algorithm was used for the first phase. Both phases of this algorithm are designed to optimize the target function. These two phases are compatible with each other, and the successive implementation of these two phases converge the dictionary atoms toward the optimal answer. However, there is no proof for the convergence of this method and the algorithm may not converge in some cases.

C. Step 3. Optimization

The PSO is a metaheuristic algorithm that can search very large spaces of candidate solutions without any hypothesis or with a few assumptions about the problem under optimization. More precisely, the PSO is a pattern searching method that does not use the gradient of the problem under optimization. This suggests that unlike the classic optimization methods such as the downside gradient methods and the pseudo-Newton method, the PSO does not require the optimization problem to be differentiable. Thus, it can be used for optimization problems that are somewhat irregular, noisy, variable with time, etc. The source of inspiration for this algorithm has been the social behavior of animals such as the massive movement of birds and fish. As the PSO also begins with an initial randomized population matrix, it is similar to many other evolutionary algorithms such as continuous genetic algorithm and colonial competition algorithm. Unlike the genetic algorithm, the PSO has no evolutionary operator like mutation and coupling.

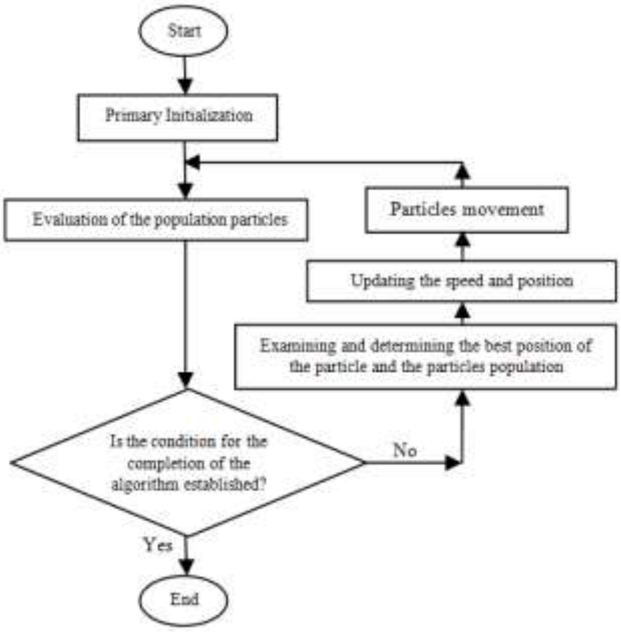

Fig. 1. PSO Algorithm

Each population element is called a particle. In fact, the PSO algorithm consists of a certain number of particles that randomly get an initial value. Two state and speed values are defined for each particle, which are modeled with a vector of space and a velocity vector, respectively.

These particles move recurrently in the n-dimensional space of the problem to search the new possible options by calculating the optimal value as a measurement criterion. The dimension of the problem space is equal to the number of parameters found in the function to be optimized. A memory is assigned to store the best position of each particle in the past and a memory for storing the best position occurred among all the particles. With the experience resulting from these memories, the particles decide how to move in the next step. At each repeat, all particles move in the n-dimensional space of the problem to finally find the general optimal point. The particles update their speeds and position according to the best local and absolute answers.

The PSO algorithm updates the speed vector of each particle and then adds the new velocity value to the position or value of the particle. The speed updates are influenced by both values of the best local answer and the best absolute answer. The best local answer and the best absolute answers are the best answers that have been obtained until the moment of the algorithm implementation respectively by a particle and in the whole population. The main advantage of PSO algorithm is its simple implementation as well as its need to determine a few parameters. The PSO is also capable of optimizing complex cost functions with a large number of local minimums. Simulation & Conclusion

IV. Simulation & Conclusion

The simulation was made on a set of 8-bit medical images with gray scales, dimensions of 256 × 256, and with different and high noise density by a computer means with the characteristics given in Table 1. Each of the images was destroyed with a noise density of 25% and 50%. This number represents the percentage of degraded image pixels. Three standard images were selected as in Fig 2 from the set of generated images. The experimental images produced were noise-removed once by the proposed method and once without the proposed method.

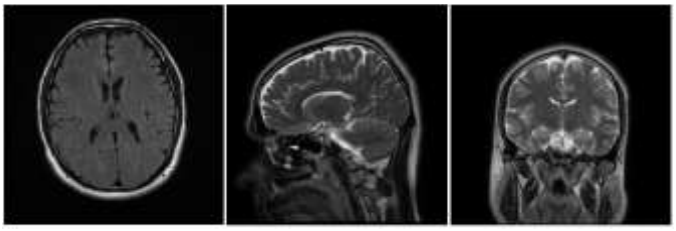

Fig. 2. An example of multiple MRI images

Three criteria of comparing visual quality, the Mean Square Error (MSE) and the Peak Signal to Noise Ratio (PSNR) were used to evaluate the performance of the proposed algorithm in the present study to remove noise from MRI images. The MSE criterion measures the difference between the original and the reconstructed images in accordance with equation (4) by dB:

$$MSE = \frac{1}{RC}\sum_{r=1}^{R}\sum_{c=1}^{C}(s[r,c]-y[r,c])^2 \quad (4)$$

Where y (r, c) and s (r, c) are respectively the brightness level of pixel (x, y) in two natural and reconstructed images. The PSNR criterion indeed represents the amount of noise power of an electric system vs. the power of the signal itself:

$$PSNR = 10\log_{10}(\frac{\max_i^2}{MSE}) \quad (5)$$

The high value of this number indicates the more closeness of the retrieved image to the original image for a certain amount of noise density. In fact, the higher this indicator, the better it is, and shows more useful signal. The qualitative measurement of the accuracy of the proposed algorithm is also done by comparing the visual quality of the original and reconstructed images.

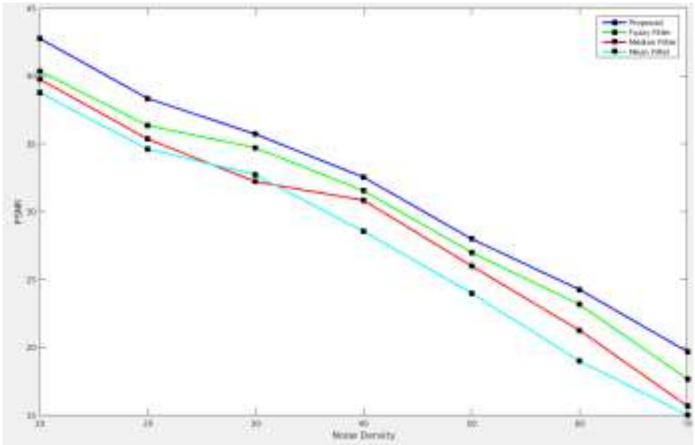

Fig. 3. Mean Square Error (MSE) graph for noise cancellation filters

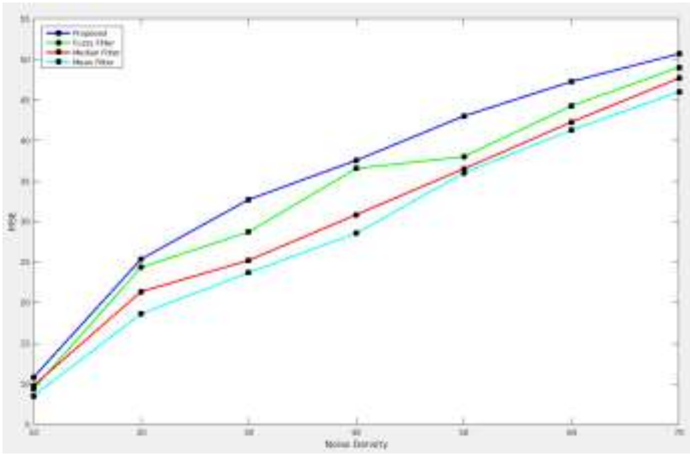

Fig. 4. Peak Signal to Noise Ratio (PSNR) criterion of noise removal filters

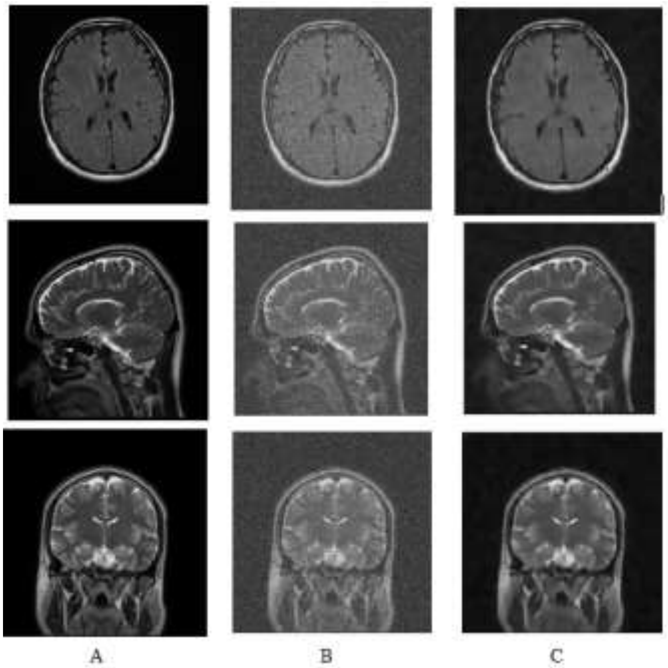

Fig. 5. Visual results of restored images of three MRI images. (A) Free-noise original image; (B) Image contaminated with Gaussian noise with a noise density of 25%; (C) Image retrieved by the proposed method

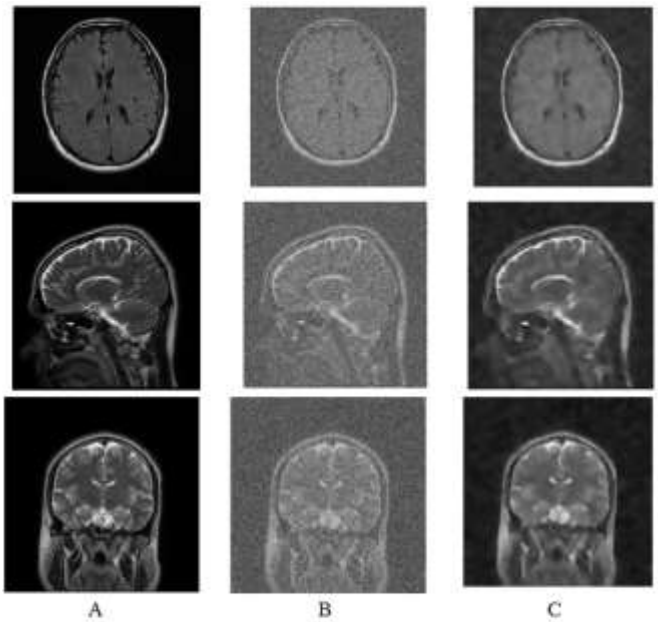

Fig. 6. Visual results of restored images of three MRI images. (A) Free-noise original image; (B) Image contaminated with Gaussian noise with a noise density of 50%; (C) Image retrieved by the proposed method

The results indicate that the use of the algorithm introduced in this study has a better performance than conventional noise removal techniques for all the noise density values.

In this paper, the concepts of MRI imaging, the principles and details of image processing, and noise removal techniques were examined. The proposed technique in this research to repair and remove noise of medical images was based on the sparse representation of the signal that the particles swarm optimization algorithm was used for its optimization. Unlike previous algorithms to remove noise from MRI images, which assume the noise density to be known or use a simple and inaccurate estimate of it as a criterion, the proposed method introduced a step-by-step and structured method for accurately estimating the image noise density, which efficiency was very suitable for noises with the density of 90%.

The suggestion for future work can be focused on an even more accurate estimation of the density of noise and identifying the noise pixels. The use of more advanced techniques for replacing detected pixels as noise will improve the noise removal of the MRI images. It is also suggested to work on removing other types of high-density noises. Other methods for continuing the work are fuzzy methods, in which, the noise removal accuracy in MRI images can be increased by using proper membership functions.

References

[1] R.M. Henkelman, Measurement of signal intensities in the presence of noise in MR images, Med. Phys. 12. 1985. PP 232–233.

[2] D. Van De Ville, M. Seghier, F. Lazeyras, T. Blu, M. Unser, WSPM: wavelet baset statistical parametric mapping, Neuroimage 37. 2007. PP1205–1217.

[3] P. Thunberg, P. Zetterberg, Noise distribution in SENSE- and GRAPPA reconstructed images: a computer simulation study, Magn. Reson. Imaging 25. 2007. PP1089–1094.

[4] A.O. Rodriguez, Principles of magnetic resonance imaging, Rev. Mex. Fis. 50. 2004. PP272–286.

[5] G. Wright, Magnetic resonance imaging, IEEE Signal Process. Mag. 14. 1997. PP56–66.

[6] Abero.E, Lightstone.M., Mitra. S.K., Arakawa.K., A New Efficient Approach for the Removal of Impulse Noise from Highly Corrupted Images, IEEE Transactions on Image Processing, 1996. Vol.5No 6.

[7] Babu.K.R., Rahul.L.A., Souri.P.V, Suneetha. A . Image Denoising in the Presence of High Level Salt and Pepper Noise using Modified Median Filter, International Journal of Computer Science & Technology. 2011. Vol. 2, SP 1.

[8] Pei-Eng.N., Kai Kuang M. A Switching Median Filter With Boundary Discriminative Noise Detection for Extremely Corrupted Images, IEEE TRANSACTIONS ON IMAGE PROCESSING. 2006. VOL. 15, NO. 6.

[9] VijayKumar V. R , Vanathi .P. T., Kanagasabapathy P, Ebenezer.D. Robust Statistics-Based Algorithm to Remove Salt and Pepper Noise in Images, International Journal of Information and Communication Engineering. 2009. Vol. 5, No. 3.

[10] J Ni, Q Qiu, R Chellappa. Subspace interpolation via dictionary learning for unsupervised domain adaptation, Computer Vision and Pattern. 2013

[11] Z Jiang, Z Lin, LS Davis, Learning a discriminative dictionary for sparse coding via label consistent K-SVD. Computer Vision and Pattern. 2011.