Performance Evaluation of Feature Extraction Techniques on Natural Image Prior in Visual Image Reconstruction

Mupparathy Babu Liby
M.Tech (Signal Processing)
Dept. of Electronics and Communication Engineering
TKM Institute of Technology
Musaliar Hills,Kollam,Kerala,India
liby.mbl@gmail.com

Jini Cheriyan
Asst. Professor
Dept. of Electronics and Communication Engineering
TKM Institute of Technology
Musaliar Hills,Kollam,Kerala,India
ssj.jinu@gmail.com

Abstract*—The functional Magnetic Resonance Imaging (fMRI) is one among the non-invasive techniques used in cognitive neuroscience, to record the activity of the brain. The fMRI reveals the functional activity caused by the Blood Oxygenated Level Dependent (BOLD) signals in the brain. The visual image reconstruction allows to translate neural brain activity pattern into the visual image (stimulus) that has caused the corresponding brain activity. The stimulus may be graphical characters, face images, handwritten characters and natural images. The proposed technique aims to develop a novel framework for visual image reconstruction of natural images from fMRI. The exact reconstruction of natural images is challenging due to its complexity. The classification of image prior plays an important role in improving the accuracy of reconstruction. Since the image prior consists of multiple categories, like Gabor wavelet transform, Scale invariant feature transform (SIFT), Speeded Up Robust Features (SURF), Local Binary Pattern (LBP), Haar feature transform, Bag of Visual Words (BoVW) were tried on image prior. The extracted features are fed into the multiclass Support Vector Machine classifier followed by k-means clustering. An analysis on reconstruction done using different feature extraction techniques revealed that the Gabor feature extraction gave the highest accuracy in final results. The reconstruction of natural images was achieved with an accuracy of 80% till now. Also 70% accuracy was achieved in identifying the category and reconstructing the test stimulus from a real time test fMRI voxel responses.The proposed work focuses on developing an accurate, less complex and automatic software technique for visual image reconstruction of natural images.

***Index Terms*—Visual Image Reconstruction, fMRI voxel responses (BOLD),Feature extraction techniques, Multiclass SVM classifier, k-means clustering, Hybrid Bayesian Framework.**

I. Introduction

In cognitive neuroscience, the functional Magnetic Resonance Imaging (fMRI) is one among the non-invasive techniques used to record the activity of the brain. Compared to other neuro imaging techniques like Electroencephelography (EEG) and Positron Emission Tomography (PET), fMRI is advantageous due to its clarity in the brain recordings and non-invasive nature. The fMRI reveals the functional activity caused by the varying oxygenated blood flow in the brain. The brain activity changes according to the persons behavioral action like viewing an image, listening to an audio, tapping his/her feet, etc. The fMRI is used for brain mapping which determines which part of the brain deals with functions like thought, speech, movements, etc. Due to its significance, encoding and decoding of fMRI has major impact on neuroscience applications. Prediction or assessment of the perceptual sensation is a major challenge in the field of neuroscience engineering. The studies related to the reconstruction methods based on fMRI are now small but in its growing stage. The visual image reconstruction allows to translate neural brain activity pattern into the visual image that has caused the corresponding brain activity.

The primary visual cortex of the brain consists of voxels carrying numerous neurons. A person's behavioural actions causes the high blood oxygenation to that particular brain area thereby resulting in the metabolic activity of the numerons. This blood oxygenation in the voxels is caused by the Blood Oxygenated Level Dependent(BOLD) signals. The neuroimaging technique, fMRI records these neuron activity as the voxel BOLD responses corresponding to the particular behavioural tasks performed by the individual. By decoding the fMRI voxel responses, one can predict the perceptual contents perceived by the individual.

In the earlier works, decoding of fMRI data is done to reconstruct the perceived checkerboard patterns, alphabets and graphical characters. Even several works have been tried on reconstructing the natural images. But the exact reconstruction could not be achieved till date due to their complexity. The data analysis of fMRI is a demanding research area which combines the various fields like neuroscience engineering, signal processing, machine learning, computer vision, etc. In the neuroscience, the fMRI can be decoded to detect a paralyzed and coma staged patients responses to specific images,their thoughts for further treatments. The neural decoding technique[11] developed a novel software that deals with the reconstruction of the stimulus(natural images) from the fMRI BOLD responses. This paper work is an extension of the neural decoding technique[11] which compares between different feature extraction techniques like Gabor wavelet transform, Scale invariant feature transform (SIFT), Speeded Up Robust Features (SURF), Local Binary Pattern (LBP), Haar feature transform, Bag of Visual Words (BoVW) on natural image prior and analyzes on the reconstruction accuracy.

II. Methodology

A. Natural Image Prior Classification

In the proposed method, the dataset of natural image prior consists of various natural images taken from various photographic collections and also the corresponding fMRI recordings. The image prior is assumed to be consisting of images from multiple categories. Hence a fine categorization of the images has to be carried out. The features of images are extracted are used in classifying the images using the multiclass Support Vector Machine classifier. Since the proposed method deals with natural image consisting of multiple categories, a multiple class SVM framework is used. The SVM classifier is trained with the training set consisting of image features and its respective labels. Using this training set, the SVM classifier categorizes the images depending on the feature values in each class/labels. The resulting clusters of images are again classified using the k-means clustering technique. The k-means clustering technique groups the images based on their low-level features thereby decreasing the feature description complexity. Hence the natural image priors are divided into cluster, c which consists of images having similar low-level features. The cluster mean,m_c and covariance, R_c is calculated for each cluster, c. The voxels corresponding to the images in the clusters are retrieved from the datasets and the centroid of each cluster is calculated. The Euclidean distance from the input test voxel data to each of the cluster centroid is computed. Consider $p = (p_1, p_2, ...p_n$ and $q = (q_1, q_2, ...q_n$ as two points in the Euclidean n-space, then the Euclidean distance, d from p to q is computed using:

$$d(p,q) = d(q,p) = \sqrt{q_1 - p_1)^2 + q_2 - p_2)^2 + ... = q_n - p_n)^2} \tag{1}$$

This is performed to identify the cluster that most resembles the test voxel data. The minimum distance is taken as the maximum probable category of images. This image category and their associated voxel category is chosen for designing the encoding model.

B. Encoding

A forward encoding model is designed with image x = $(x_1, x_2, ..., x_p)' \epsilon R^p$ and the associated measured fMRI brain response y = $(y_1, y_2, ..., y_q)' \epsilon R^q$:

$$y = \mathbf{B}'x + \epsilon, \epsilon\ N(0; \Sigma), \tag{2}$$

where ϵ is the zero mean normally distributed noise, **B** is the Regression coefficients and Σ is the covariance matrix. Let **X** = $(x^1, x^2, ..., x^N)' \epsilon R^{Nxp}$ denote the image matrix with x^j denoting the stimulus presented before the subject at the j-th trial. Let $y_i = (y_i^1, ..., y_i^N)$ denote the associated fMRI voxel BOLD responses in the i-th voxel. For each voxel location (i.e; i-th voxel) the voxel variance ,V_i, of the responses in the respective voxel location is calculated. The regularization parameter, λ_i is obtained from K-fold cross validation between voxels and pixels using:

$$\widehat{\lambda_i} = argmin_{\lambda \epsilon \Lambda} var(\widehat{\epsilon}_i^1(\lambda)', ..., \widehat{\epsilon}_i^K(\lambda)') \tag{3}$$

where $\widehat{\epsilon}_i^k(\lambda)' = y_i^k - X^k \widehat{\beta_i}$ are the estimated residuals. The voxels having highest voxel variance compared to the regularization parameter are considered to be carrying most significant information on pixels and hence used for reconstruction procedures. For those voxels, the regression coefficients, **B** = $\beta_1, ...\beta_i$ are estimated using the linear regression or classification of multivariate brain analysis.

$$\widehat{\beta_i} = (\mathbf{X'X} + \lambda_i \mathbf{I}_p)^{-1} \mathbf{X}' y_i \tag{4}$$

where the amount of regularization is controlled by $\lambda_i >= 0$. The covariance, $\Sigma = diag(\sigma_1^2, ..., \sigma_i^2$ is computed using:

$$\widehat{\sigma_i^2} = var(\widehat{\epsilon_i^1}(\widehat{\lambda_i})', ..., \widehat{\epsilon_i^1}(\widehat{\lambda_i})') \tag{5}$$

On the other hand, the voxels with variance lesser than the regularization parameter are considered to be less informative and hence are discarded from the reconstruction procedures.

C. Decoding

The decoding model is designed such that, given the voxel response for testing, the stimulus (natural image) that has caused the corresponding response is reconstructed from the test fMRI voxel responses (BOLD) data, X2. The multivariate Gaussian representation of probable image's covariance and mean in canonical form is defined as:

$$mean, m = QB\Sigma^{-1}y \tag{6}$$

$$covariance, Q = (R^{-1} + B\Sigma^{-1}B^T)^{-1} \tag{7}$$

A hybrid Bayesian framework is formed using the selected voxels having most significant information of pixels and selected cluster of natural image priors. The stimulus that has caused the corresponding fMRI voxel responses is reconstructed using the hybrid Bayesian framework which consists of mixture models of natural image priors. The reconstruction of the stimulus from the centralized test fMRI voxel data is obtained using:

$$\widehat{x} = (R - RB(\Sigma + B^T RB)^{-1} B^T R) B\Sigma^{-1} X2 \tag{8}$$

The reconstructed output will be an average of all the images in the selected maximum probable category. Thus the smoothness of the reconstructed output is proportional to the number of images in the selected category. More the number of images, the more smoothened will be the reconstructed output and vice versa.

III. Results and Analysis

The input to the neural decoding technique is the fMRI voxel responses (BOLD). The voxel responses in the primary visual cortex of the brain are caused by the metabolic activity of the neurons when a person views a visual image. For better clarity of the responses, the voxel responses from the anterior occipital cortex are also taken into account. The dataset consisting of the natural images and their associated fMRI responses taken from the lower and higher brain areas are read into the MATLAB R2015b platform. The natural image (stimulus) that has caused the neural activity in the brain was decoded/ reconstructed from the fMRI voxel responses.

A. Natural Image Prior

For an accurate visual image reconstruction, the prior knowledge of the natural image is necessary. The image prior is the dataset consisting of gray scale natural images taken from various photographs. The same dataset consisting of twenty natural images from the previous work[11] is used for this work. The image prior is considered to be multimodal which contains natural images of different categories. The image prior categories may be known or unknown. For known image categories, labeling of the natural images using the supervised technique can be performed. When the image prior categories are considered to be unknown, the supervised labeling is not possible. Therefore one of the unsupervised features learning technique, k-means clustering algorithm is suited for clustering. The k-means clustering groups the images into clusters having similar low level features.

In the proposed method, the image prior is divided into group of images by extracting the features. Different feature extraction techniques Gabor wavelet transform, Scale invariant feature transform (SIFT), Speeded Up Robust Features (SURF), Local Binary Pattern (LBP), Haar feature transform, Bag of Visual Words (BoVW) were tried on the natural image prior. Each group of image is considered as set of images from the same category based on their features.

The extracted features are given to the SVM classifier. The categorized images are then grouped into clusters within that category using k-means clustering technique which will cluster the images within each category based on their low level feature information. From the final image clusters obtained, the erroneous classification happened only for very few images. The corresponding voxel responses for images in each cluster are retrieved from the database so that the final clusters of voxels are formed. Initially k-means clustering was applied on the extracted features of image prior. Since the k-means clustering technique groups the images based on their low-level features, if the images of animals/humans has mostly scenery as background, they will be grouped with scenery images thereby causing unfine classification results. Out of twenty images, the misclassification happened for 15 images. Thus the classification accuracy could not be achieved here. For better results, the SVM classifier was tried on image pixels directly which resulted in misclassification of 10 images out of 20 images. The below figures, Figure 1, Figure 2, Figure 3, Figure 4 AND Figure 5 are the clustered output using different extraction techniques.

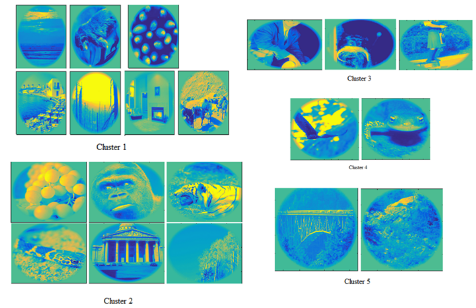

Fig. 1. Final Classified Output using SIFT technique

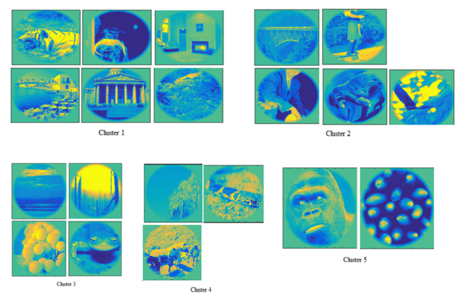

Fig. 2. Final Classified Output using SURF technique

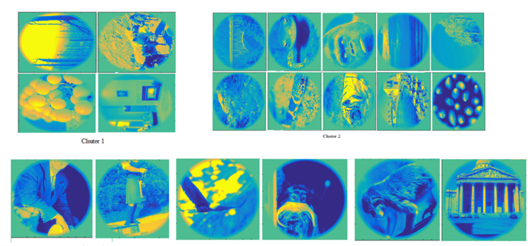

Fig. 3. Final Classified Output using LBP technique

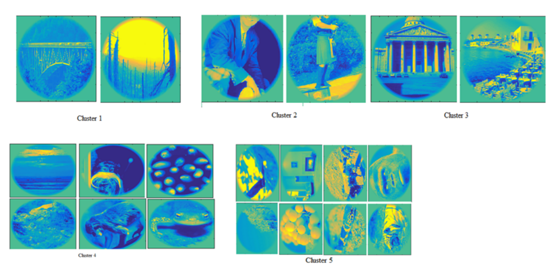

Fig. 4. Final Classified Output using HAAR technique

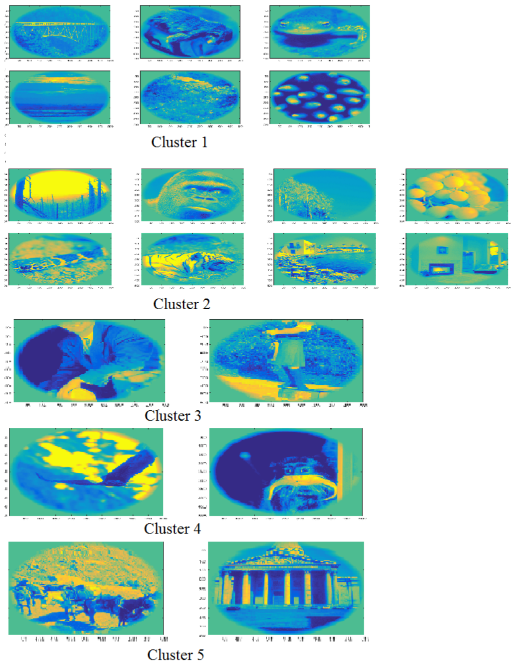

Fig. 5. Final Classified Output using Gabor filters

The table below shows an analysis on the different feature extraction techniques performed in the work.

Technique	No.of images	Reconstructed	Accuracy
Gabor	20	18	80%
Haar	20	15	75%
LBP	20	12	60%
SIFT	20	11	55%
SURF	20	11	55%
BoVW	20	10	50%

B. Stimulus and the test fMRI data

The stimulus is chosen from the given natural image prior set. The corresponding fMRI BOLD responses are pre-processed and also provided in the database. An example of both is shown in the below Figure 6. For the reconstruction of the stimulus, the evoked fMRI BOLD response corresponding to that stimulus is provided to the software as test voxel data.

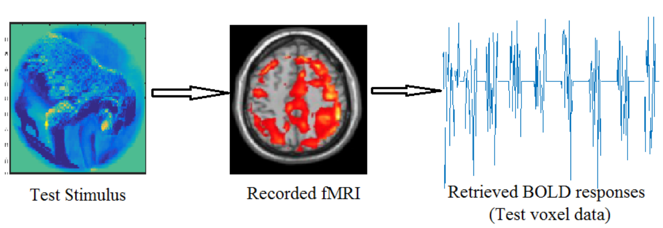

Fig. 6. An example of a stimulus and the fMRI responses

C. Selection of the maximum probable category

The centroid (mean) of each cluster of voxels is determined. The Euclidean distance from the test voxel data to centroid of each cluster is calculated. The voxel response category that has the minimum Euclidean distance from the test data is chosen to be the maximum probable category which is having high resemblance to the test voxel data. This category of voxel responses and its corresponding images are taken to design the encoding model. The Figure 7 and Figure 8 shows an example of the selected category of image prior and its associated fMRI BOLD responses. The encoding model is designed for model parameter estimation that contributes to the reconstruction.

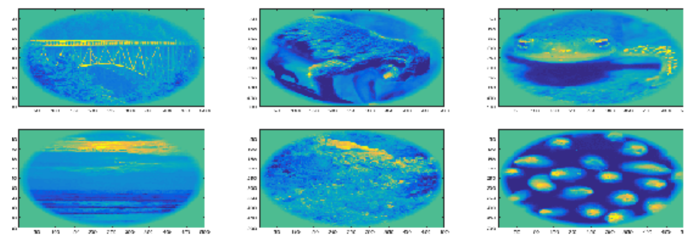

Fig. 7. An example of a selected category of image prior

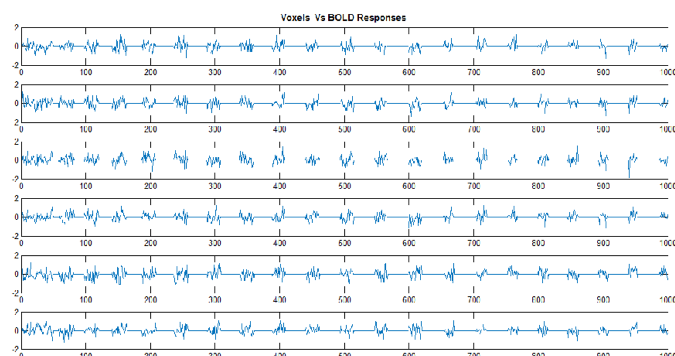

Fig. 8. An example of a selected category of BOLD responses

D. Model Parameter Estimation

The encoding model is designed for the selected category of natural images and the associated fMRI voxel responses of the brain. The model describes how the stimulus features are encoded in the brain responses. The model parameters are computed only for the voxels having significant pixel information and the remaining irrelevant voxels are discarded from the reconstruction procedures. Selection of relevant voxels are done by comparing the voxel variance, (V_i) and regularization parameter,(λ_i) at each voxel locations. The voxels having higher variance compared to the regularization parameter are considered to be having the significant pixel information. The model parameters computed for selected voxels include Regression coefficients,(β_i) and the Covariance matrix,(Σ). This automated method of calculating the model parameters in a way reduces the amount of irrelevant voxels thereby decreasing the complexity of the algorithm.

E. Reconstruction

A Hybrid Bayesian framework is formed for reconstructing the natural image from the test fMRI BOLD response. The framework is formed with selected voxels having highest variance and unimodal image cluster. Finally the reconstruction of the stimulus that has caused the test fMRI response is achieved. The actual dataset of natural image prior is having resolution of 128 x 128 pixels. Also, the dataset of fMRI voxels consists of 25000 voxels. But, due to the limitations in the memory of MATLAB platform used, the resolution size of the natural images had to be reduced to 50 X 50 pixels. Similarly the size of the voxels considered for the simulation was reduced to 1000 voxels. Hence the clarity in the reconstruction result is very poor.

In the neural decoding technique[11], the reconstruction of natural image(stimulus) could be achieved from the test fMRI voxel responses (BOLD). In this method, different feature extraction techniques are tried and compared thereby an analysis on the reconstruction accuracy is done. Also, in our previous method[11], only the test fMRI voxel responses were provided from the database available. The below figures, Figure 9 - Figure 13 are the results reconstructed from the test fMRI voxel BOLD respose available in the dataset. In this work, we have tried providing test fMRI voxel response from outside the dataset (i.e. real time) and could identify the category to which the stimulus belong and reconstruct the same. For example, if a test fMRI voxel response of a snake is provided to the system, the system achieved in selecting the probable category consisting of a snake and reconstructed the same. The Figure 14 and Figure 15 are the results reconstructed using the test voxel BOLD response outside the dataset (real time).

In this work, 20 natural images were considered as the image prior. These images were categorized into 5 clusters. Once the probable cluster of voxels and corresponding images are determined, the reconstructed output will be an average of the images in the selected cluster.Since the selected cluster using the test voxel data in Figure 9 and Figure 10 contains more images, the output of reconstruction seems to be more smoothened. On the other hand, the selected cluster using the test voxel data in Figure 11, Figure 12 and Figure 13 consists of only two images, the reconstructed output seems to be less smoothened. Out of the 20 natural image prior used in this work, four images were wrongly reconstructed. An example of such wrong reconstruction is shown in the below Figure 16.

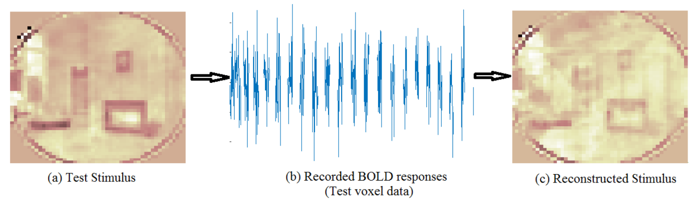

Fig. 9. Reconstruction of Stimulus; (a) Test Stimulus; (b) Recorded BOLD responses (test voxel data); (c)Reconstructed Stimulus

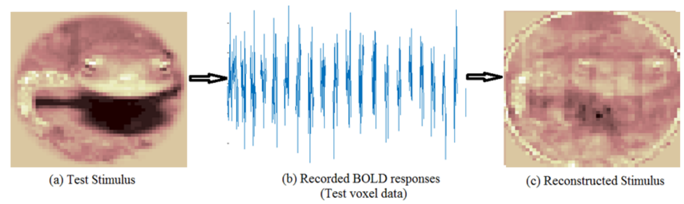

Fig. 10. Reconstruction of Stimulus; (a) Test Stimulus; (b) Recorded BOLD responses (test voxel data); (c)Reconstructed Stimulus

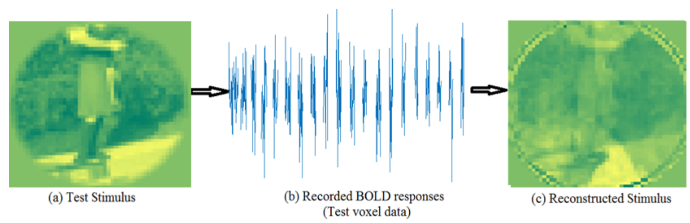

Fig. 11. Reconstruction of Stimulus; (a) Test Stimulus; (b) Recorded BOLD responses (test voxel data); (c)Reconstructed Stimulus

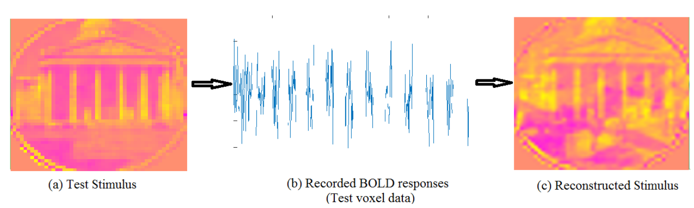

Fig. 12. Reconstruction of Stimulus; (a) Test Stimulus; (b) Recorded BOLD responses (test voxel data); (c)Reconstructed Stimulus

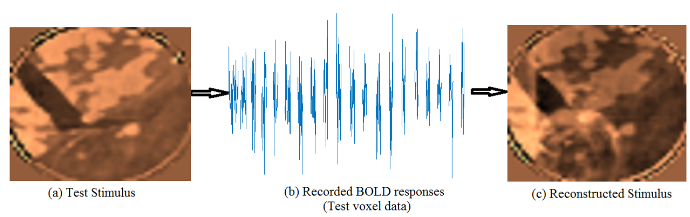

Fig. 13. Reconstruction of Stimulus; (a) Test Stimulus; (b) Recorded BOLD responses (test voxel data); (c)Reconstructed Stimulus

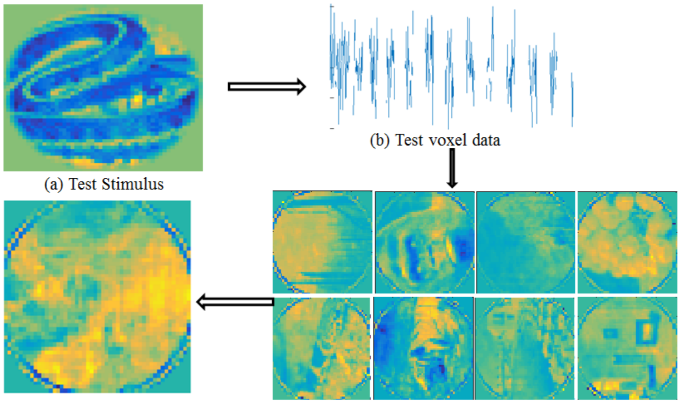

Fig. 14. Reconstruction of Stimulus; (a) Test Stimulus; (b) Recorded BOLD responses (test voxel data); (c)Reconstructed Stimulus

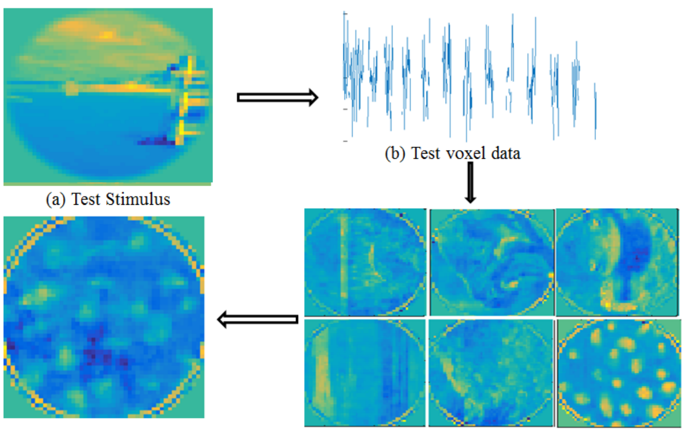

Fig. 15. Reconstruction of Stimulus; (a) Test Stimulus; (b) Recorded BOLD responses (test voxel data); (c)Reconstructed Stimulus

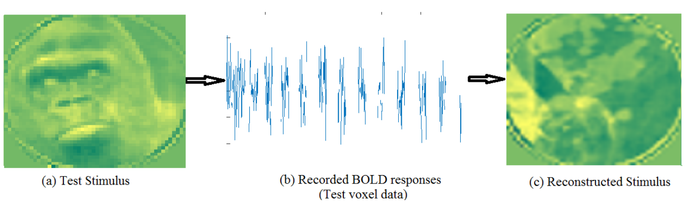

Fig. 16. Reconstruction of Stimulus; (a) Test Stimulus; (b) Recorded BOLD responses (test voxel data); (c)Reconstructed Stimulus

IV. Conclusion

The functional Magnetic Resonance Imaging (fMRI) detects the activity of neurons in the brain area that is caused by the varying Blood-Oxygenated-Level-Dependent (BOLD) signal. The neural decoding technique in [11] developed a novel framework for visual image reconstruction of natural images. The classification of image prior plays an important role in improving the accuracy of reconstruction. In this paper, the different feature extraction methods like Gabor wavelet trandform, Scale invariant feature transform (SIFT), Speeded Up Robust Features (SURF), Local Binary Pattern (LBP), Haar feature transform, Bag of Visual Words (BoVW) were tried on image prior followed by the Support Vector Machine classifier and k-means clustering. An analysis was done on the reconstruction accuracy among which the Gabor feature extraction gave the highest accuracy in reconstruction results. The reconstruction of natural images was achieved with an accuracy of 80% till now. Also an 70% accuracy was achieved in identifying the category and reconstructing the test stimulus from a real time test fMRI voxel responses.

References

[1] Sanne Schoenmakers, UmutGl, MarcelvanGerven and TomHeskes.“Gaussian mixture models and semantic gating improve reconstructions from human brain activity”.*Frontiers in Computational Neuroscience,* January 2015,Volume8,Article173.

[2] Umut Gu clu , Marcel A. J. van Gerven. “Unsupervised Feature Learning Improves Prediction of Human Brain Activity in Response to Natural Images”. *PLoS Comput Biol* 10(8): e1003724. doi:10.1371/journal.pcbi.1003724, August 2014, Volume 10, Issue 8, e1003724.

[3] Sanne Schoenmakers, Tom Heskes, Markus Barth, Marcel van Gerven. “Linear Reconstruction of Perceived Images from Human Brain Activity”.*Elsevier Inc., NeuroImage*xxx(2013) xxx - xxx Impact Factor: 6.36 DOI: 10.1016/j.neuroimage.2013.07.043, July 2013.

[4] Vincent Q. Vu, Ravikumar, P., Naselaris, T., Kay, K.N. “Encoding and decoding V1 fMRI responses to natural images with sparse nonparametric models”. *Ann Appl Stat.* 5(2B): 1159 - 1182. doi:10.1214/11-AOAS476 2011 June.

[5] Thomas Naselaris, Ryan J. Prenger, Kendrick N. Kay, Michael Oliver, and Jack L. Gallant. “Bayesian Reconstruction of Natural Images from Human Brain Activity”. *Elsevier Inc, Neuron* 60, 902 - 915, September 24, 2009.

[6] Yoichi Miyawaki,Hajime Uchida, Okito Yamashita,Masa-aki Sato, Yusuke Morito, Hiroki C.Tanabe,Norihiro Sadato,and Yukiyasu Kamitani. “Visual Image Reconstruction from Human Brain Activity using a Combination of Multiscale Local Image Decoders”.*Elsevier Inc, Neuron*60, 915 - 929, December 11, 2008.

[7] Kendrick N. Kay, Thomas Naselaris, Ryan J. Prenger, and Jack L. Gallant. “Identifying natural images from human brain activity”. *Nature* 452(7185): 352 - 355. doi:10.1038/nature 06713 March 20,2008.

[8] Martin A. Lindquist. “The Statistical Analysis of fMRI Data”. *Statistical Science* 2008, Vol.23, No. 4, 439 - 464 DOI: 10.1214/09 - STS282.

[9] Yusuke Fujiwara,Yoichi Miyawaki,Yukiyasu Kamitani,Dubois,Jean-Baptiste Poline, Denis Lebihan,and Stanislas Dehaened. “Estimating image bases for visual image reconstruction from human brain activity”. *NeuroImage* 2006.

[10] Bertrand Thirion, Edouard Duchesnay, Edward Hubbard, Jessica, Thirion et al. “Inverse retinotopy: Inferring the visual content of images from brain activation patterns”. *NeuroImage* 33 (2006) 1104 - 1116.

[11] Mupparathy Babu Liby, Jini Cheriyan. “Neural Decoding Technique to Reconstruct Stimulus from the Evoked fMRI Voxel Response“ *International Journal of Science and Research (IJSR)* Volume 5 Issue 2 February 2016.