Visual Image Reconstruction from fMRI Activation Using Multi-scale Support Vector Machine Decoders

Yu Zhan[1], Jiacai Zhang[1], Sutao Song[2], and Li Yao[1,3]

[1] School of Information Science and Technology, Beijing Normal University, Beijing, China
[2] School of Education and Psychology, University of Jinan, Shandong, China
[3] State Key Laboratory of Cognitive Neuroscience and Learning, Beijing Normal University
zhanyu@mail.bnu.edu.cn, {jiacai.zhang,yaoli}@bnu.edu.cn,
Sep_songst@ujn.edu.cn

Abstract. The correspondence between the detailed contents of a person's mental state and human neuroimaging has yet to be fully explored. Previous research reconstructed contrast-defined images using combination of multi-scale local image decoders, where contrast for local image bases was predicted from fMRI activity by sparse logistic regression (SLR). The present study extends this research to probe into accurate and effective reconstruction of images from fMRI. First, support vector machine (SVM) was employed to model the relationship between contrast of local image and fMRI; second, additional 3-pixel image bases were considered. Reconstruction results demonstrated that the time consumption in modeling the local image decoder was reduced to 1% by SVM compared to SLR. Our method also improved the spatial correlation between the stimulus and reconstructed image. This finding indicated that our method could read out what a subject was viewing and reconstruct simple images from brain activity at a high speed.

Keywords: Image Reconstrucion, fMRI, Multi-scale, SVM.

1 Introduction

Functional magnetic resonance imaging (fMRI) provides a convenient tool for scientists to determine what a person perceives from his/her brain activity [1-6]. Researches about visual information reading decoded brain activity at three levels: classification, identification and reconstruction. Classification is to predict which category the present image belongs to from the brain pattern of activity [8-12]. Beyond classification, image identification established the computation model to identify the image that the subject was viewing out of a set of potential images from brain activity measurements [1, 2]. Kay and colleagues utilized a Gabor wavelet function to capture the visual stimuli characteristics related to fMRI activity, and characterized the relationship between visual stimuli and fMRI activity in early visual areas with quantitative receptive-field models. Their study showed that these receptive-field models make it possible to perform image identification [1]. More recently, researchers have moved a more forward step to reconstruct the visual image composed of flickering

M. Kurosu (Ed.): Human-Computer Interaction, Part V, HCII 2013, LNCS 8008, pp. 491–497, 2013.

checkerboard patterns or even more complicated actual natural images that were seen or movie experience, rather than simply choosing the image from a known set [4, 5].

Thirion et al. build the inverse model of the retinotopy of the visual cortex to infer the visual content of real or imaginary scenes from the brain activation patterns [6]. Another representative work of image reconstruction was the study of Miyawaki et al., they established the method of multi-scale local image decoder to directly model the relationship between the image stimulus and fMRI activity at the specific time when image was presented to subjects [3]. In their study, reconstructed images were modeled by a linear combination of local image bases. Miyawaki fixed the shape and size of image bases in his study, and he utilized local image of four scales: 1×1, 1×2, 2×1, and 2×2 patch areas [3]. These are totally 361 image bases for a 10×10 flickering checkerboard patterns. The local decoders based on sparse logistic regression approach were defined to predict the mean contrast of each local image bases. Each of the 361 local decoders was individually trained to classified fMRI data samples into a class corresponding to contrast level. After the decoders' output, a linear combination of the 361 image bases was applied to reconstruct the predicted images.

Miyawaki's research stands for the highest level of visual decoding studies that have emerged over the years. The spatial correlation between the presented stimulus and reconstruction image even came to 0.68 ± 0.16 (mean ± s.d.) for individuals. However, the time and space complexity of their method is extremely high because of the extremely laborious computation in sparse logistic regression model.

To further investigate the effective image reconstruction method in visual information decoding from brain activities, we are trying to find a method raising the training speed with little precision loss. First, the heavy work in training the contrast decoder model with SLR for each element image in Miyawaki's work was reduces by classifiers designed with SVM. As each local decoder consisted of a multi-class classifier, Instead of the sparse logistic regression method, we could use an SVM model to classify fMRI data samples into discrete contrast levels. Second, we will also use the 1×3 and 3×1 image bases, as well as the fixed image bases used in Miyawaki's work. Using Bayesian based canonical correlation analysis (CCA), Fujiwara proposed a method to automatically find a set of image bases from the fMRI data, and reconstruction results illustrated that this set of 3-pixel image bases improved the reconstruction performance [7].

2 Method

2.1 Dataset

We used the same dataset from Miyawaki et al., where fMRI signals were measured in two independent sessions. In each session, the subject viewed visual images consisting of contrast-defined 10×10 patches. In the random image session, a total of 440 flickering checkerboard spatially random pictures were presented, each stimulus block was 6 seconds long followed by 6 s rest period. In the figure image session, a total of 120 pictures were presented. Each stimulus block was 12 seconds long followed by a 12 s rest period. Stimulus pictures were geometric shapes (i.e. "square", "plus", "X") or alphabet letters ("n", "e", "u", "r", "o"). Each picture had been shown to the subject for 4 or 8 times. The Supplemental Data can be found online at http://www.neuron.org/supplemental/S0896-6273(08)00958-6.

2.2 Multi-scale Image Bases

As with Miyawaki's research, we assume that an image is represented by a linear combination of local image elements of multiple scales. The local image bases used in Miyawaki's work were 1×1, 1×2, 2×1 and 2×2 patch areas. They were placed at every location in the image with overlaps. For example, the 1×2 scale bases will cover the 9 rectangles (1 1)-(1 2), (1 2)-(1 3), … , (1 9)-(1 10) in the first row of the 10×10 image. So there are a total of 10×9 image bases of scale 1×2. Similarly, there are respectively 10×10, 9×10, 9×9 image bases of scale 1×1, 2×1 and 2×2. For each stimuli images, we counted the number of flickering grids in an image base, 1×1 scale yields value 0 for all patches staying gray and 1 for whole element image flicking. For the 2×1 scale image bases, there are three values, 0 means no flickering grid, 1 means one flickering grid, and 2 means all the 2 grid in image bases are flickering. Similarly, 1×2 yield contrast values 0, 1, 2 and 2×2 yields 0, 1, 2, 3, 4. The present study used extra image elements of 1×3 and 3×1 patch areas, thus yielded another 160 image bases (10×8 and 8×10 for 1×3 and 3×1 patch areas respectively). The flickering grids number for 3-patch image ranges from 0 to 3. The mean contrast value of each local image base was defined as the total number of patches in that local image divided by the number of flickering patches (represented as white girds).

2.3 SVM Model

Instead of applying sparse logistic regression method, we used support vector machine to predict the mean contrast value for each local element. In this study, we only used the fMRI activity of V1 area to build the SVM models. The sample features are the activation of voxels in V1 area; the sample label is the mean contrast value in each image base. All the local image decoders except those for 1×1 patches, the mean contrast level belonged to more than 2 classes, so we used multi-class SVM models. The SVM code used here was implemented by Lin Chih-Jen in Taiwan University. The source code is available online at http://www.csie.ntu.edu.tw/~cjlin/. Here, linear SVM models were trained with samples in the random image session. And we evaluated the model with test dataset from the figure image session.

2.4 Linear Combination

This procedure makes up the reconstructed image by adding all local image bases altogether. In Miyawaki's work, least square error method was used to obtain the coefficient of each image base. In our work, the prediction accuracy on the training set are 100% for all 1×1 image bases, all the 1×1 image bases will have coefficients 1. To simplify this problem, we assume that all the pixels in the 10×10 image are predicted by summing up the class labels of correlated image bases, and the coefficients of all the image bases are 1. The mainframe of this approach is shown in Fig. 1.

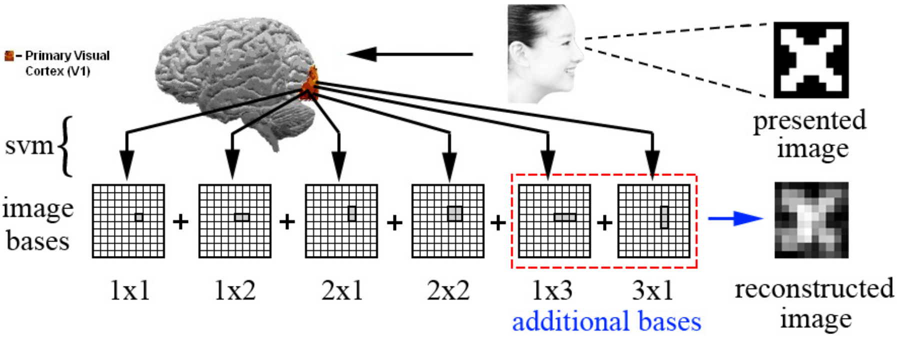

Fig. 1. The fMRI signals were measured while subject viewed the stimulus images. The SVM models were trained using V1 area data of the random image session. After the SVM assigned the class labels to each image base for the figure image session data, the predicted value of a specific pixel was calculated by summing up the class labels of all the image bases that covered the pixel.

3 Results

3.1 Reconstruction Performance

The stimuli images were reconstructed using the element images, whose contrast values predicted by SVM models trained with all block-average data (average of 6s or 3-volumns fMRI data, TR=2s) in the random image session. Reconstruction was performed on block-averaged data (average of 12s or 6-volumns fMRI data) in the figure image session. Although model training used only random images, the reconstructed images of test dataset showed obvious similarity between presented images and reconstructed images for stimulus shapes or letters. Using the combination of 1×1, 1×2, 2×1 and 2×2 image bases, the spatial correlation was 0.6643 ± 0.1207 (mean ± s.d. data not shown). By adding the 1×3 and 3×1 image bases, the spatial correlation increased to 0.6934 ± 0.1165 (mean ± s.d.). Reconstruction images from all trials of the figure image session are illustrated in Fig.2.

3.2 Computational Expenses

The SVM model training and testing time cost for all 6 scales are listed in Table.1. The comparison suggested that the computing complexity using SVM was far lower than that of sparse logistic regression.

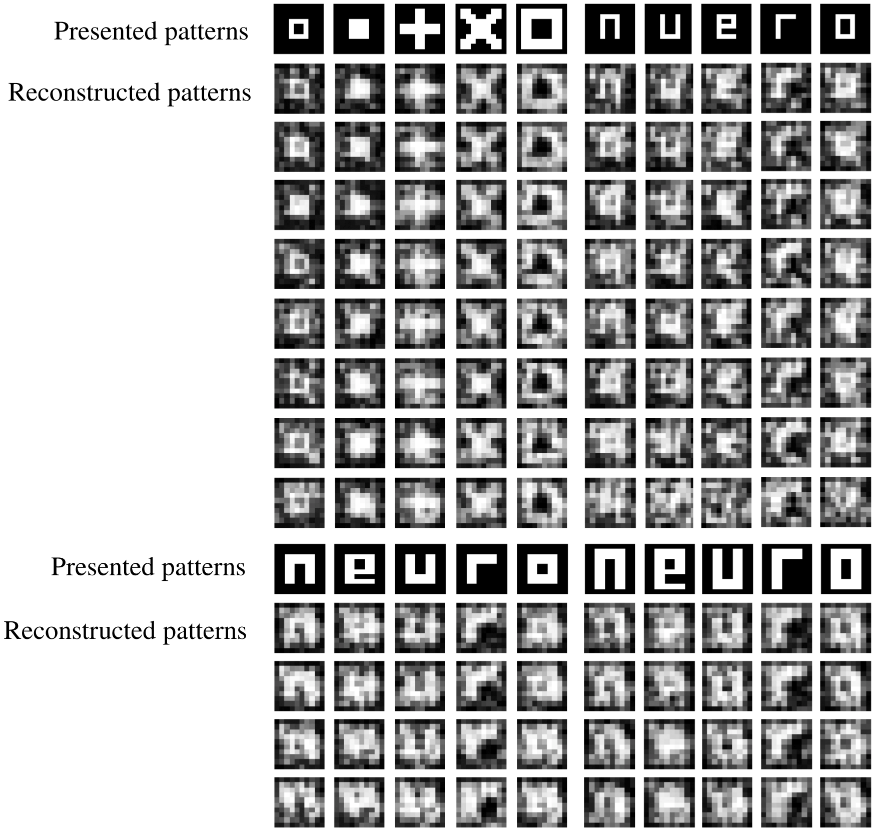

Fig. 2. Reconstructed visual images of chareacters. The reconstruction results of all the trials for the subject are shown with presented images from the figure image session (a total of 120 images). The 10 stimuli images in the first row repeated 8 times and the 10 stimuli in the 10th row repeated 4 times. The reconstructed images are sorted in descending order of the spatial correlation. No post processing was applied.

Table 1. The total time (in seconds) including training and testing for support vector machine (SVM) and sparse logistic regression (SLR) under different scales were shown below. System Information: Windows 7 SP1(64bit), CPU: Intel(R) Core(TM)2 CPU 6600@ 2.40GHz 2.39GHz, Ram: 4.00GB.

Scale	1×1	1×2	2×1	2×2	1×3	3×1
method						
SVM	**85**	**103**	**97**	**106**	**92**	**98**
SLR	**11831**	**21661**	**18739**	**58229**	**--**	**--**

4 Discussion

4.1 SVM and Over-Fitting

SVM is a pretty fast classification algorithm, which has its application in many research fields. In our study, more than 1700 voxels are used to train the local decoders. Such high dimension of features may probably cause the problem of over-fitting. As we know, SVM implements the structural risk minimization principle and it searches to minimize an upper bound of generalization error. For this reason, SVM reduces the risk of over-fitting and provides a relatively stable classification performance. In this research, the output accuracies of SVM are all beyond guess probability.

4.2 Image Bases

By intuition, the image bases more closed to the visual recognition of human-beings the less the reconstruction error will be. There is no verdict in what shapes of image bases are more approximate to our visual system. Our results showed that using extra 3-pixel image bases produces slightly better reconstructed images, consistent with Fujiwara's conclusion that 3-pixel image bases have a high correlation with the visual cortex activities.

4.3 Least Square Method

As Miyawaki mentioned in his research, the image bases are not orthogonal which means even if all the image base decoders output the right results, we cannot reconstruct the exact image by adding all the image bases together. To deal with this problem, he used the least square method to calculate the coefficients of each image bases. We found this process is technically not necessary. First, if all the SVM decoders output right, the reconstructed images by adding all image bases altogether have an average spatial correlation of 0.9551 ± 0.0217 with the original ones. Second, the training set, which has only 440 random images, is relatively small for predicting a total of 521 least square coefficients. While it is true that using different coefficients (not LSM) may yield better performance, little improvement could be achieved. We assign unit value to each coefficient and this simple strategy seemed work well.

5 Conclusions

The results reported here provide an efficient and accurate method to reconstruct visual stimulus from fMRI signals. Furthermore, the improved spatial correlation using 3-pixel image bases suggests that these image bases may provide more supplementary visual information. The main features of the stimulus were emerged in the reconstructed images, which indicated that SVM could exactly map the activation of visual cortex (V1 area) to the contrast stimulus patterns. Here we also drew the same conclusion with Miyawaki that the outputs of local decoders in the center of an image are more accurate than that in the edges or corners, demonstrating that the visual attention is likely to be concentrated in the center of sight. Further research can be applied to investigate how to accurately predict these surrounding areas.

Acknowledgments. We thank Yoichi Miyawaki in National Institute of Information and Communications Technology, Kyoto, Japan for sharing their fMRI data used in this study. This work is supported by the National High-tech R&D Program (863 Program) under Grant Number 2012AA011600 and the Fundamental Research Funds for the Central Universities. This work is also supported by the Major Research Plan of the National Natural Science Foundation of China (NSFC) Key Program 60931003.

References

1. Kay, K., Naselaris, T., Prenger, R., Gallant, J.: Identifying natural images from human brain activity. Nature 452(7185), 352–355 (2008)
2. Mitchell, T., Shinkareva, S., Carlson, A., Chang, K., Malave, V., Mason, R., Just, M.: Predicting human brain activity associated with the meanings of nouns. Science 320(5880), 1191 (2008)
3. Miyawaki, Y., Uchida, H., Yamashita, O., Sato, M., Morito, Y., Tanabe, H., Sadato, N., Kamitani, Y.: Visual image reconstruction from human brain activity using a combination of multiscale local image decoders. Neuron 60(5), 915–929 (2008)
4. Naselaris, T., Prenger, R.J., Kay, K.N., Oliver, M., Gallant, J.L.: Bayesian reconstruction of natural images from human brain activity. Neuron 63(6), 902–915 (2009)
5. Nishimoto, S., Vu, A.T., Naselaris, T., Benjamini, Y., Yu, B., Gallant, J.L.: Reconstructing Visual Experiences from Brain Activity Evoked by Natural Movies. Current Biology 21, 1641–1646 (2011)
6. Thirion, B., Duchesnay, E., Hubbard, E., Dubois, J., Poline, J., Lebihan, D., Dehaene, S.: Inverse retinotopy: inferring the visual content of images from brain activation patterns. NeuroImage 33(4), 1104–1116 (2006)
7. Fujiwara, Y., Miyawaki, Y., Kamitani, Y.: Estimating image bases for visual image reconstruction from human brain activity. Advances in Neural Information Processing Systems 22, 576–584 (2009)
8. Haxby, J., Gobbini, M., Furey, M., Ishai, A., Schouten, J., Pietrini, P.: Distributed and overlapping representations of faces and objects in ventral temporal cortex. Science 293(5539), 2425–2430 (2001)
9. Cox, D., Savoy, R.: Functional magnetic resonance imaging (fMRI)“brain reading”: detecting and classifying distributed patterns of fMRI activity in human visual cortex. NeuroImage 19(2), 261–270 (2003)
10. Spiridon, M., Kanwisher, N.: How distributed is visual category information in human occipito-temporal cortex? An fMRI study. Neuron 35(6), 1157–1165 (2002)
11. Carlson, T., Schrater, P., He, S.: Patterns of activity in the categorical representations of objects. Journal of Cognitive Neuroscience 15(5), 704–717 (2003)
12. O’toole, A., Jiang, F., Abdi, H., Haxby, J.: Partially distributed representations of objects and faces in ventral temporal cortex. Journal of Cognitive Neuroscience 17(4), 580–590 (2005)