BrainReader: Effective Visualization of fMRI-based Movie Reconstruction

Natalia Bilenko, Valkyrie Savage

***Abstract*—Previous work in decoding visual experiences based on fMRI activity has been successful in reconstructing images and movies that participants viewed inside an MRI scanner. Reconstruction is done by fitting a forward model that predicts fMRI activity across the brain in response to a set of movies. The model represents brain activity as a linearized function of visual information features that capture the structure of the movies (spatiotemporal Gabor wavelet filters). The forward model is then inverted and used to decode what the subject saw based on their brain responses to a testing set of movies. Decoding is performed by fitting a maximum a posteriori function to a large library of previously unseen movie clips. The top 100 decoded movie clips are then averaged or stitched together to produce a visualization of the decoding. Though the decoding is quite precise when measured quantitatively, these visualizations do not fully reflect its accuracy. We make the visualization more coherent by combining the decoded clips in several improved ways. First, we demonstrate the change in quality gained using weighted averaging. Then, we use HOG features to select a subset clips similar to the ground truth clip and SIFT flow to find an optimal path in time. Third, we use appearance morphing to visually align the path-arranged clips. Finally, we share the decoded movies resulting from the same stimuli across different participants in the experiment.**

***Index Terms*—fMRI, decoding, visualization, computational videography.**

I. Introduction

FUNCTIONAL magnetic resonance imaging, or fMRI, data can shed light on what individuals are looking at. Specific areas in the brain are known to react strongly to particular line orientations and locations. Using activation data from these centers, such fMRI data can "reconstruct" what an individual is seeing.

The Gallant lab from UC Berkeley has worked on this problem previously, demonstrating **VS: citation needed** that they can perform this reconstruction by averaging images from a training set. The top 100 images whose recorded fMRI profiles match most closely with the recorded data are simply stacked on top of each other to create output videos.

However, this type of visualization is very "messy": while the quantified fMRI matches are quite strong, the visual output video stacks are misleadingly inaccurate. Using computational techniques, we can improve the quality of these output videos (Figure 1)

Instead of a simple averaging process for turning guesses into an output video, we used several more sophisticated approaches. First, we simply performed weighted averaging, using clip ranking as our weights. This already constituted an improvement over naïve weighting (Figure 2).

Fig. 1. a) The first generation of visualization for BrainReader clips: the top 100 guesses are simply averaged and overlaid to create an output image at each frame. This belies the accuracy of the technique, which is quite high. b) Our improved visualization **VS: obviously need image here.**.

Fig. 2. Using a weighted average improves over the naïve average seen in Figure 1a. **VS: obviously se need this image as well.**

II. Methods

Our processing pipeline takes as input two pieces of data for each second:

- The original clip stimulus presented to the subject
- The top 100 guesses (based on fMRI data) and their rankings

We first perform HOG feature extraction on both pieces of data and reject guesses whose HOG data does not match well with the original clip. We then extract SIFT features from the beginning and ending of each of the top guesses and calculate SIFT flow between them. Using cost back-propagation, we

Natalia and Valkyrie are joint first-authors on this work. They can be reached via email at *nbilenko@berkeley.edu* and *valkyrie@eecs.berkeley.edu*.
Manuscript submitted December 18th, 2014.

find the lowest-cost path through the remaining clips. **VS: Finally, we perform morphing between these clips using SIFT keypoints and output the final clip compilations.**

Clips (both original and guessed) are 1 second in length, and have 15 frames per second. We do not have data on whether there are scene breaks within a 1 second clip.

A. Pruning - HOG Features

Histogram of Oriented Gradients (HOG) features roughly indicate edges in an image as well as the orientation of those edges. We use HOG features to ensure good *spatial* alignment of guess footage with the presented clip.

We calculate the HOG features of each frame of each clip and perform an SSD with the ground truth clip's HOG features at that time step. This process is physically based in the fMRI data, as the visual processing centers of the brain react in specific ways to edges presented in particular orientations and in particular locations across the visual field. Thus we see this "pruning" step as non-essential, as we would expect that the fMRI data and subsequent ranking step (performed prior to our getting the data) is already based on these features. **VS: We should look at how well the HOG features actually correlate to rankings. I assume we basically pick the top guesses, but I don't know.**

After finding the SSD of the HOG features through each clip when compared to the original source clip, we throw away **VS: what exactly do we throw away?**.

B. Consistency - SIFT Features

Scale-Invariant Feature Transform (SIFT) features, often used in image recognition tasks, can give higher-level information about the contents of a scene. We want to minimize the key point flow (i.e., scene composition) between the last frame of one clip and the first frame of the next clip. SIFT keypoints have been used for nearest-neighbor database searches (e.g., in the SIFT flow paper), and can successfully extract, for example, a street image to match a street image, even when the *optical* flow between the two street images is large. We use this to keep a thematically consistent scene across timesteps.

We use SIFT flow to calculate costs for transitioning between one clip and the next. We simply calculate the SIFT flow between the last frame of one clip and the first frame of all potential next clips. We then use cost back-propagation, i.e., dynamic programming, to find the lowest semantic cost path through all the clips remaining after HOG pruning.

C. Visuals - Image morphing

III. Related Work

Our work relates to both basic image processing and video processing.

Appendix A
Proof of the First Zonklar Equation

Appendix one text goes here.

Appendix B

Appendix two text goes here.

Acknowledgment

The authors would like to thank...

PLACE PHOTO HERE

Natalia Bilenko Natalia is a cool grad student with purple hair that studies brains. She works in the Gallant lab in the Neuroscience department.

PLACE PHOTO HERE

Valkyrie Valkyrie is a grad student whose desk is covered in 3D printed stuff. She works for Bjoern Hartmann in the Berkeley Institute of Design.