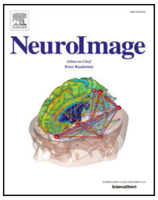
NeuroImage xxx (2016) xxx–xxx

Contents lists available at ScienceDirect

NeuroImage

journal homepage: www.elsevier.com/locate/ynimg



A method for real-time visual stimulus selection in the study of cortical object perception

Daniel D. Leeds [a,b,*], Michael J. Tarr [c,b]

[a] Fordham University, Computer and Information Science Department, Bronx, New York, USA
[b] Carnegie Mellon University, Center for the Neural Basis of Cognition, Pittsburgh, PA, USA
[c] Carnegie Mellon University, Psychology Department, Pittsburgh, PA, USA

A R T I C L E I N F O

Article history:
Received 10 September 2015
Accepted 23 February 2016
Available online xxxx

Keywords:
Real-time stimulus selection
Functional magnetic resonance imaging
Object recognition
Real-time signal processing
Computational modeling

A B S T R A C T

The properties utilized by visual object perception in the mid- and high-level ventral visual pathway are poorly understood. To better establish and explore possible models of these properties, we adopt a data-driven approach in which we repeatedly interrogate neural units using functional Magnetic Resonance Imaging (fMRI) to establish each unit's image selectivity. This approach to imaging necessitates a search through a broad space of stimulus properties using a limited number of samples. To more quickly identify the complex visual features underlying human cortical object perception, we implemented a new functional magnetic resonance imaging protocol in which visual stimuli are selected in real-time based on BOLD responses to recently shown images. Two variations of this protocol were developed, one relying on natural object stimuli and a second based on synthetic object stimuli, both embedded in feature spaces based on the complex visual properties of the objects. During fMRI scanning, we continuously controlled stimulus selection in the context of a real-time search through these image spaces in order to maximize neural responses across pre-determined 1 cm^3 rain regions. Elsewhere we have reported the patterns of cortical selectivity revealed by this approach (Leeds et al., 2014). In contrast, here our objective is to present more detailed methods and explore the technical and biological factors influencing the behavior of our real-time stimulus search. We observe that: 1) Searches converged more reliably when exploring a more precisely parameterized space of synthetic objects; 2) real-time estimation of cortical responses to stimuli is reasonably consistent; 3) search behavior was acceptably robust to delays in stimulus displays and subject motion effects. Overall, our results indicate that real-time fMRI methods may provide a valuable platform for continuing study of localized neural selectivity, both for visual object representation and beyond.

Introduction

How do humans visually recognize objects? Broadly speaking, it is held that the primate ventral occipito-temporal pathway of the human brain implements a feedforward architecture in which the features of representation progressively increase in complexity as information moves up the hierarchy (Felleman and Essen, 1991; Riesenhuber and Poggio, 1999). In almost all such models, the top layers of the hierarchy are construed as high-level *object representations* that correspond to and allow the assignment of category-level or semantic labels. Critically, there is also the presupposition that while early levels along the pathway encode information about edge locations and orientations (Hubel and Wiesel, 1968) and information about textures (Freeman et al., 2013), one or more levels, between what we think of as early vision and high-level vision, encode *intermediate* visual features. Such features, while less complex than entire objects, nonetheless capture important — and possibly compositional — object-level visual properties (Ullman et al., 2002). Remarkably, for all of the interest in biological vision, the nature of these presumed intermediate features remains frustratingly elusive. To help address this knowledge gap, we introduce new methods that leverage human fMRI to explore the intermediate properties encoded in regions of human visual cortex.

Any study investigating the visual properties employed in cortical object perception faces multiple challenges. First, the number of candidate properties present in real-world objects is large. Second, these properties are carried by millions to billions of potential stimulus images. Third, feature and image space can be parameterized by an uncountable number of potential models. Fourth, the time available in a given human fMRI experiment is limited. Scanning time for an individual subject is limited to several hours across several days. Fifth, during a given scan session, the slow evolution of the blood-flow dependent fMRI signal necessarily limits the frequency of single stimulus display trials to one every 8 to 10 s; more frequent displays produce an overlay of hemodynamic responses difficult to recover without carefully tuned

---

* Corresponding author at: Fordham University, Computer and Information Science Department, Bronx, New York, USA.
E-mail addresses: dleeds@fordham.edu (D.D. Leeds), michaeltarr@cmu.edu (M.J. Tarr).

http://dx.doi.org/10.1016/j.neuroimage.2016.02.071

Please cite this article as: Leeds, D.D., Tarr, M.J., A method for real-time visual stimulus selection in the study of cortical object perception, NeuroImage (2016), http://dx.doi.org/10.1016/j.neuroimage.2016.02.071

pre-processing or careful dissociation of temporally adjacent stimuli. Moreover, even with these considerations, the neural data recovered will be noiser and less amenable to use on a trial-by-trial basis. As such, assuming a minimum of 8 s to display each trial, at most several hundred stimuli can be displayed to a subject per an hour.

Here we suggest that dynamic stimulus selection, that is, choosing new images to present based on a subject's neural responses to recently shown images, enables a more effective investigation of visual feature coding. Our methods build on the dynamic selection of stimuli in studies of object vision in primate neurophysiology. For example Tanaka (2003), explored the minimal visual stimulus sufficient to drive a given cortical neuron at a level equivalent to the complete object. He found that individual neurons in area TE were selective for a wide variety of simple patterns and that these patterns bore some resemblance to image features embedded within the objects initially used to elicit a response. Tanaka hypothesized that this pattern-specific selectivity has a columnar structure that maps out a high-dimensional feature space for representing visual objects. In more recent neurophysiological work Yamane et al. (2008) and Hung et al. (2012), used a search procedure somewhat different from Tanaka and a highly-constrained, parameterized stimulus space to identify the contour selectivity of individual neurons in primate IT. They found that most contour-selective neurons in IT encoded a subset of the parameter space. Moreover, each 2D contour within this space mapped to specific 3D surface properties meaning that collections of these contour-selective units would be sufficient to capture the 3D appearance of an object or part.

At the same time, there has been recent interest in real-time human neuroimaging. For example Shibata et al. (2011), used neurofeedback from visual areas V1 and V2 to control the size of a circular stimulus displayed to subjects and Ward et al. (2011) explored real-time mapping of the early visual field using Kalman filtering. Most recently Sato et al. (2013), have developed a toolbox (“FRIEND”) that implements neural feedback applications in fMRI, applying classification and connectivity analyses to study the encoding of emotion. These studies support the idea of incorporating real-time analysis and feedback into neuroimaging work to expanding fields, such as the study of object perception.

Here we explore new methods for the real-time analysis of fMRI data and the dynamic selection of stimuli. More specifically, our procedure selects new images to display based on the neural responses to previously-presented images as measured in pre-selected brain regions. Our overall objective is to maximize localized neural activity and to identify the associated complex featural selectivity within image spaces that are organized on the basis of insights from earlier studies in object perception (Leeds et al., 2013; Williams and Simons, 2000). We employ two sets of objects and their corresponding spaces — real-world objects organized based on similarities computed by the SIFT computer vision method (Lowe, 2004) and synthetic “Fribble” objects (Williams and Simons, 2000) organized based on morphs in the shapes of their component appendages (see Fig. 5).

In previously published results, we reported the nature of the cortical selectivities uncovered by this novel approach (Leeds et al., 2014). Here we study the technical and biological factors influencing the performance of our real-time stimulus search, as well as the behavior of our search across subjects and stimulus sets. In particular, using synthetic stimuli, we found that searches exhibited some convergence onto a small number of preferred visual features and consistency across repeated searches for a given brain region within an individual subject. In contrast, using real-world object stimuli, we found only weak convergence and consistency, possibly as a result of the visual diversity of the real-world stimuli included in this image space. More generally, we observe that our methods are robust to undesired actions from subjects (e.g., head motions) and program flaws (e.g., stimulus selection delays), suggesting that our methods offer an important first-step in developing effective methods for real-time human neuroimaging.

Material and methods

Stimulus selection method

Our study is unique in that it relies on the dynamic selection of stimuli in a parameterized stimulus space, choosing new images to display based on the BOLD responses to previous images within a given pre-selected brain region. More specifically, we automatically choose the next stimulus to be shown by considering a space of visual properties and probing locations in this space (corresponding to stimuli with particular visual properties) in order to efficiently identify those locations that are likely — based on prior neural responses to other stimuli in this space — to elicit maximal activity from the brain region under study. As discussed in the Defining SIFT space and Defining Fribble space sections, we employed two somewhat different representational spaces, one based on SIFT features derived from real-world images, and one based on synthetic “Fribble” objects (see Fig. 5). SIFT was used for the first group of ten subjects, while Fribbles were used for the second group of ten subjects. For both groups, each stimulus i that could be displayed is assigned a point in space p_i based on its visual properties. The measured response of a given brain region to this stimulus r_i is understood as:

$$r_i = f(p_i) + \eta \quad . \tag{1}$$

That is, a function f of the stimulus' visual properties as encoded by its location in the representational space plus a noise term η, drawn from a zero-centered Gaussian distribution. The process of displaying an image, recording the ensuing cortical activity via fMRI, and isolating the response of the brain region of interest using the preprocessing program we model as performing an evaluation under noise of the function describing the region's response. For simplicity's sake, we perform stimulus selection assuming our chosen brain region has a selectivity function f that reaches a maximum at a certain point in the representational space and falls off with increasing Euclidean distance from this point. Our assumption is consistent with prior work in primate neurophysiology, such as Tanaka (2003), Hung et al. (2012), and Yamane et al. (2008), in which stimuli were progressively adapted to maximize response of a single neural unit to converge on the single (complex) visual selectivity presumed to be associated with the unit. We also note that our assumption is consistent with recent work in human fMRI that finds that selectivity for object categories is organized in a smooth gradient across cortex whereby the amount of neural “real estate” apportioned to shared features across visually-similar categories is minimized (Huth et al., 2012). Under these assumptions, we use a modified version of the simplex simulated annealing Matlab code available from Donckels (2012), implementing the algorithm from Cardoso et al. (1996). This method seeks to identify new points (corresponding to stimuli) that evoke the highest responses from the selected cortical region. An idealized example of what a search run might look like based on this algorithm is shown in Fig. 1b. The results of our study indicate our assumption of a single peak in cortical response is not always accurate. Nonetheless, the simplex simulated annealing method achieves convergence for several real-time stimulus searches.

For each of four distinct stimulus classes — mammals, human-forms, cars, and containers for real-world objects and four classes distinguished by core body shape and appendage orientation for Fribble objects (described further in the Interleaving searches section and in Leeds et al. (2014)) — we performed searches in each of two scan sessions. To probe the consistency of our search results across different initial simplex settings, we began the search within each session at a distinct point in the relevant stimulus representational space. In the first session, the starting position was set to the origin for a given stimulus class, as specific stimulus examplars were distributed in each space relatively evenly around the origin. In the second scan session, the starting position was manually selected to be in a location opposite

Please cite this article as: Leeds, D.D., Tarr, M.J., A method for real-time visual stimulus selection in the study of cortical object perception, NeuroImage (2016), http://dx.doi.org/10.1016/j.neuroimage.2016.02.071