Complementary Computing for Visual Tasks: Meshing Computer Vision with Human Visual Processing

Ashish Kapoor, Desney Tan, Pradeep Shenoy† and Eric Horvitz
Microsoft Research, Redmond, WA 98052, USA
†University of Washington, Seattle, WA 98122, USA

Abstract

We explore the opportunity to harness electroencephalograph (EEG) signals generated during human visual processing to enhance computer vision systems. We review the challenging task of categorizing objects, such as faces, in images and then describe methods that can be used to combine the complementary competencies of human and machine computation to achieve improved recognition performance. We present the results of several experiments where brain signals, recorded from people examining images, are used to enhance the performance of vision systems on categorization tasks. We find that significant gains in classification accuracy can be achieved with the human-aided vision systems.

1. Introduction

Visual category recognition remains a challenging problem. Building computer vision systems that are competent at recognizing target categories of objects has typically required intensive human effort. In particular, people must provide labeled data to inform classifiers about visual categories. As this labeling process is often very expensive, recent work has focused on ways to reduce the number of labeled examples required to learn accurate models [3, 6, 15]. These systems aim to utilize human labeling effort in a most efficient manner. Other solutions to the problem of obtaining labels for visual categories include embedding the labeling task in online games [23, 24], and asking users to provide finer-grained information by selecting and labeling specific objects within images [1].

We explore here a new form of human contribution to computer vision systems–the sharing with the systems of brain signals generated while people view images and scenes. We directly measure participants' brain signals so as to provide information to the machine with little conscious effort. This approach is built on the realization that people subconsciously process different images in different ways, and that such differences are measurable by available brain-sensing methodologies, even when the user is not explicitly trying to categorize images.

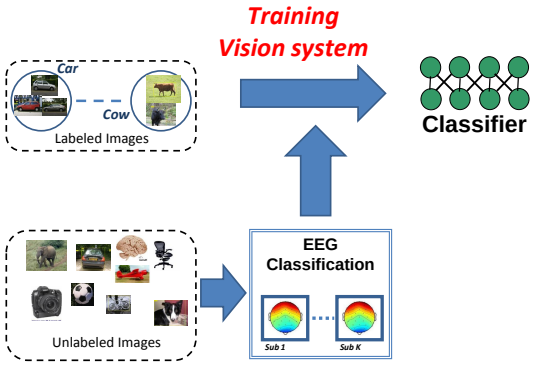

Figure 1. The proposed framework to train a computer vision system with human-brain processing for visual category recognition.

There are several advantages of developing methods that fuse human visual information processing with traditional computer vision techniques. First, the explicit collection of labels for building visual categorization systems is expensive as it involves a slow, deliberative process of viewing and assessing. In contrast to the plodding process of hand tagging images, informative brain signals, generated in response to the viewing of images, are observed even when images are displayed for only 40ms [9]. By exploiting the implicit processing in the human brain with the rapid presentation of images, we can significantly speed up the labeling process and reduce the amount of hand-labeled training data we need to collect. Second, we can take advantage of the complementary skills of computer vision and the human visual system. Current computing algorithms and the human visual processing perceive scenes very differently; the two modalities provide complementary information and such complementarity can lead to enhanced classifiers. Previous work has demonstrated the power of combining complementary methods and data sources in a co-

978-1-4244-2154-1/08/$25.00

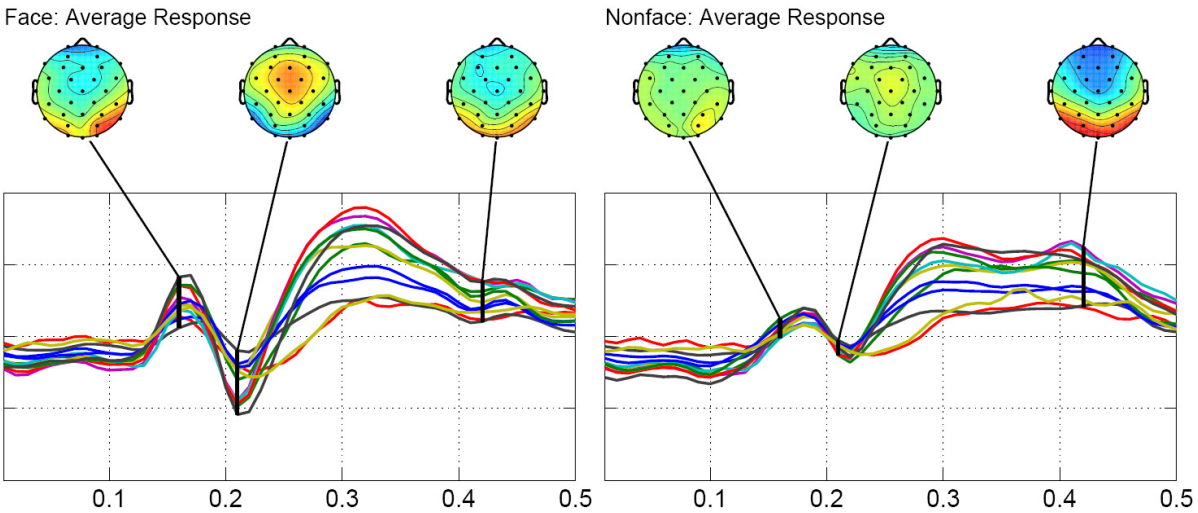

Figure 2. Average ERP responses to viewing faces (left) and non-faces (right) for one user with the controlled images. The time series for each channel is shown in multiple colored lines within each graph and the accompanying scalp plots show the spatial distribution of these signals at snapshots in time. The red in these plots signifies higher activity. The N170 (at 170 ms) face specific peak is evident in the face response but not in the non-face response.

herent manner to enhance the overall accuracy of classification [4, 22]. Finally, studying how brain signals, generated with the viewing of visual scenes, can boost traditional vision-based methods and highlight deficits in our computer vision algorithms. For example, exploring the gains that come with the addition of information drawn from human information processing can lead to insights about aspects of images and categories that are currently unmodeled or poorly modeled in computer vision. This direction of research promises to build a path toward introducing into our computer vision systems the kind of robustness and flexibility we associate with human vision. Techniques based on computer vision focus on various imaging transformations and intra-class variations and are often motivated by the specific discriminatory tasks at hand. We suspect that human information processing is less task-specific, and that human analysis employs richer features and feature ensembles, coupled with rich contextual and semantic associations that are unavailable to our vision algorithms.

We shall focus on the advantages of effectively combining information from implicit brain processing, as measured by an electroencephalograph (EEG), to build better visual categorization models. Specifically, our main contribution is a method that trains computer vision algorithms by combining information from machine computed visual image features with the information measured from the brain of a human viewing images. The overall framework is shown in Figure 1. The core idea is to exploit the informativeness of the brain signal to recover information about an unlabeled set of images. Such information about the unlabeled images can be appended to the available labeled training set, thus, effectively increasing the corpus that can be used to train a vision system.

Two key issues need to be addressed before we can achieve benefits from the framework. First, the brain signals need to be informative enough so as to be to provide useful information about the unlabeled set. Second, the information available in the channel should be complementary to the information provided by the vision system. We explore these key issues in the context of our recent work [16, 21] and show, using data from human users, that brain signals associated with human visual processing indeed can provide valuable information to the vision system.

Providing methods that can integrate information from human visual processing with computer vision may be especially valuable to researchers interested in the challenges of recognizing faces and gestures. Experiments have suggested that the brain signals associated with faces are highly informative. For example, the presentation of a human face is commonly connected with a pronounced negative drop in signal amplitude in certain channels approximately 170ms following stimulus presentation [19]. The N170 drop does not always lie exactly at 170ms because of various stimulus presentation delays as well as physiological variance. Figure 2 shows an example of the response in one of our users to face and non-face stimuli. A strong N170 face-specific response is seen (left) in EEG measured after a user has seen a face, but not when the participant is viewing images without faces (right). The N170 face-specific response is represented by a purple line that protrudes out the bottom of the series slightly after 170 ms following the stimulus presentation. Similarly, it has been noted that the responses are significantly different for categories such as animals and inanimate objects and there is enough discriminatory signal to train a classifier, indicating the discriminative power that may exist in this signal.

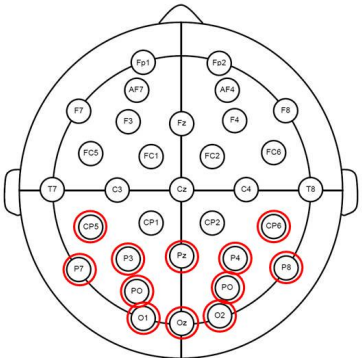

Figure 3. The figure shows a standardized layout for electrode placement in a 32-electrode EEG measurement system [12], pictured from the top, with nose and ears shown for orientation. The electrodes used for analysis are marked in red.

2. Human-Aided Computing

Electroencephalography provides neurophysiological measurement of brain activity using electrodes placed on the surface of the scalp (see e.g. [7]). Researchers often examine behavioral correlates in EEG signals by measuring the event-related potential (ERP), which represents the spatiotemporal shape of brain measurements in response to a discrete sensory stimuli (*e.g.*, flashing an image on a screen). The spatiotemporal patterns of EEG produced in response to stimuli can capture characteristic differences in the way people process information [11, 13]. These differences can be recognized by predictive models constructed from labeled training data. Consequently, the human brain can be used as a computational processor.

As an example, an event-related potential (ERP) known as the P300[1], or *recognition response*, indicates when the user has detected a stimulus of interest. This signal has been widely used in the study of brain-computer interfaces (BCI), which aim to allow users to communicate with the external world using brain signals alone [2, 5]. Similarly, Gerson and colleagues [9] exploit this P300 response in their system for "cortically coupled computer vision", in which the user intentionally performs visual search on a sequence of rapidly presented images, looking for a designated target image. The system can detect target images using the brain response alone, in certain cases faster than is possible with manual identification using button presses. This system requires that participants have an explicit intention to search for a single target or category of targets, and thus to serve as "target detectors" system, rather than as detectors for a specific category of objects. The study also did not use computer vision algorithms to enhance the EEG-based results.

[1] named for the positive amplitude change seen in certain EEG channels roughly 300ms after stimulus presentation

While work with the P300 signal requires that users explicitly be attending to a single target class, we have recently shown that it is possible to categorize a more general set of classes by inferring cognitive processing patterns from a larger set of ERPs [21]. This work demonstrates high accuracies in categorizing images displayed to subjects, even when the subjects are not explicitly trying to perform the classification task. In the rest of paper, we build upon our initial work [21] and discuss the research in the context of training computer vision systems for the purpose of object categorization.

3. Visual Category Recognition with Implicit Brain Processing

Automatic visual category recognition is a hard problem and humans typically perform better than the best available computational algorithms. Our hypothesis is that performance of computer vision system can be improved by observing implicit brain processing and obtaining and exploring the nature of complementary information that is not available to traditional computer vision algorithms.

Before moving on, we need to determine(1) if the EEG signals contain enough information about the classification task at hand and (2) if the information derived from brain signals is complementary to state-of-the-art computer vision methods for object categorization. To answer these questions, we analyze and discuss three different setups which in turn provide theoretical and empirical analysis.

3.1. Setup 1: Only EEG

We first explore a complementary system referred to as *human-aided computing* [21], in which the user is passively viewing images while performing a distracter task that does not consist of explicitly labeling or recognizing the images. The distracter task serves only to capture visual attention and cognitive processing. The participants were not told of the classification task and were not explicitly trying to perform a classification implicit brain processing. In this study, the participants viewed images categorized by whether the images contained an animal, a human face, or inanimate objects while performing a distracter task. The distractor task was counting of images that contained butterflies in them. The data set consisted of a *training* set of 60 images per class shown to each of the subjects only once, whereas the *test* set consisted of 20 images per class presented 10 times each to the subject in a block randomized fashion. Images were flashed for 150ms at about 500ms apart and EEG responses were recorded at 2 kHz from 32 channels from 14 users who wearing a cap of electrodes placed in the 10-20 standard electrode layout [12]. Figure 3 shows the configuration of the electrodes on the scalp. For processing details, see [21].

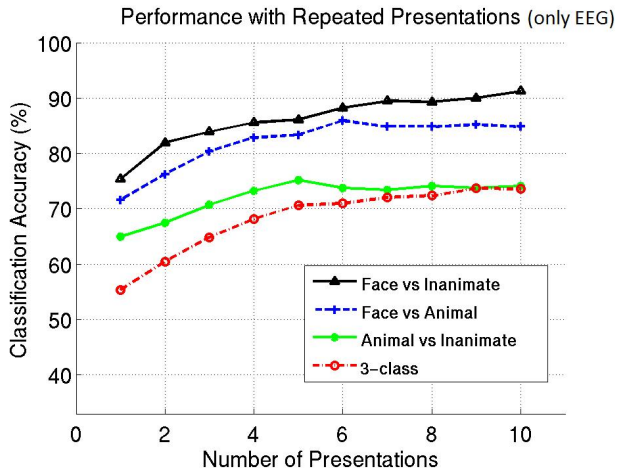

Figure 4. Relatively high accuracies for classification across multiple categories of images. Such accuracies are achieved with repeated presentations of images. This result was reported in [21].

The results (see Figure 4 from [21]) showed that passive EEG responses can be used to label images with one of three category labels, namely human faces, animals, and inanimate objects, with an average accuracy of 55.3%, using only a single presentation of an image. Furthermore, we observed that the accuracy could be boosted by using multiple presentations to one or multiple users. With up to ten presentations, the average labeling accuracy can be raised to 73.5%. Analysis of two-way classification further showed that the system was effective in recognizing faces. In particular with just a single presentation, the system was able to classify face versus inanimate objects and face versus animals at a 75.3% and 71.6% accuracy respectively. The system also demonstrated better than chance recognition performance of 65.0% for animals versus inanimate objects, which was a more difficult classification task. Further, we found that these accuracies increase significantly as more repetitions are added, rising to 91.2%, 84.8%, and 74.1% for the two-way classifications with 10 presentations of the test images to the participant.

This work highlighted the discriminatory information present in the EEG signal and demonstrated that human brain computation could in principle be used as a new modality for extracting features from images for use in an object recognition system. The results also highlighted the feasibility of harnessing signals in an *implicit manner*, potentially in-stream with background, ongoing activities, as valuable information was derived from participants who were not trying to perform classification. However, a key question remained about the value of using human brain signals in a complementary-computing solution for vision. We did not know if EEG signals provide complementary information that could extend the abilities of state of the art computer vision algorithms for object categorization-versus providing redundant information. Thus, we set out to explore the relationship of information from brain signals and current computer vision capabilities.

3.2. Setup 2: EEG + Computer Vision

The second experimental setup consisted of a combined system [16] aimed at answering the question about the value of the information gained via access to EEG signals. Specifically, we sought to explore methods for combining information from EEG responses with the state-of-the-art vision algorithms for object recognition. If the combination of system performed better than either of the individual modalities, there would be strong reasons to believe that EEG signal contains information that can be useful to the traditional computer vision algorithms.

We focus on computer vision algorithms that operate on the Pyramid Match Kernel (PMK) method [10]. Object categorization using PMK is based on analyzing sets of local features. The approach is overall tolerant to partial occlusions, object pose variation, and illumination changes [10, 17, 25, 26]. In our system, the information from the vision features and EEG are combined in a Kernel alignment framework. Specifically, we fuse information from EEG and the computer vision channel via considering a linear combination of kernels: $\mathbf{K} = \sum_i \alpha_i \mathbf{K}_i$; here, $\mathbf{K}_i$ denote kernels computed either in EEG space or using PMK. Thus, the goal of fusion algorithm is to find appropriate weights α_i so that the final kernel is aligned with the available data. Consequently, the algorithm sets a higher value for the parameter α_i if there is a lot of discriminatory information in kernel $\mathbf{K}_i$, otherwise a low weight is set. For details please refer to [16].

We found that the combined method showed significant gains over individual modalities for a battery of experiments. In particular, for studies with the same data used in the first experimental setup, we found that a combined strategy yielded superior performance. Figure 5 shows that significant gains are obtained by combining the EEG signals with computer vision features. The combination with single presentations outperforms each individual channel with an accuracy of 86.67% on the 3-way classification task. This performance is further improved to 91.67% when test images are presented multiple times. Although the vision features consistently outperform the EEG features, the combination performs better than either, suggesting that the EEG signals and the base computer vision analysis complement one another. Furthermore, the analysis of the feature weights α_i discovered by the combination algorithm highlight the complementary nature of the information provided by the human and computer vision modalities. Specifically, we look at relative weights defined as $\gamma(PMK) = \frac{\alpha_{PMK}}{\sum_i \alpha_i}$ and $\gamma(EEG) = 1 - \gamma(PMK)$. Figure 6 highlights these relatives weights averaged over 100 different runs for various amounts of training data. We can see that the vision

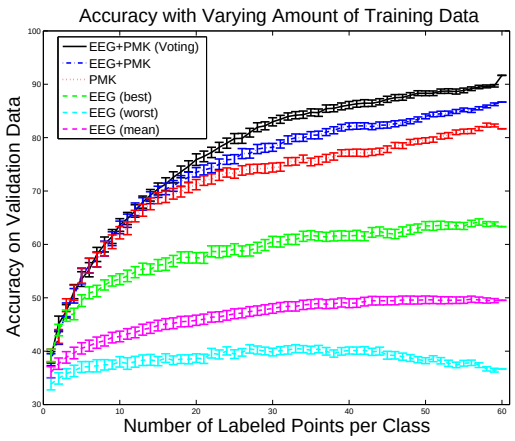

Figure 5. Performance of different modalities on the test set as the number of labeled examples are varied for a single presentation to the subject. The combined classification based on EEG and PMK significantly outperform the individual modalities. Presenting the same image multiple times to the subject and voting among those classification outcomes further improves the accuracy. The error bars represent standard deviation (reproduced from [16]).

modality has higher discriminative power overall, but that the weight of the EEG modality is highly significant and leads to significant gains in accuracy. Also, the relative contribution of EEG signal increases with data suggesting that bigger gains can be expected with increasing amounts of human training data. Overall, results from this setup suggest that indeed the EEG information contains complementary information that can be used to train the computer vision system. In the next section we describe such a strategy.

3.3. Setup 3: Training Vision Systems with EEG

With a third experimental setup, we explored whether it is possible to improve individual vision algorithms with the help of human computation. We explored how computer vision systems can use the complementary information available in the EEG signal via harnessing sets of previously unlabeled examples. Specifically, the training data for a vision system can be expanded by first showing unlabeled examples to humans and then analyzing the observed EEG signal to recover label information. The idea is described graphically in Figure 1. Intuitively, we can think of the proposed approach as a way of increasing the available training data for a vision algorithm by bootstrapping off the information available through EEG.

We note that techniques such as semi-supervised and unsupervised machine learning have been previously used to exploit information from unlabeled data. The experiments in this paper are performed in a supervised learning setting only for simplicity; extension to semi-supervised and unsupervised technique is feasible but is not explored here.

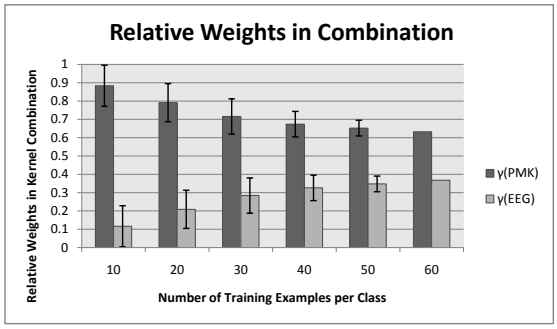

Figure 6. Comparison of relatives weights of the different modalities as we vary the number of labeled examples per class. The error bars represent represent standard deviation.

We use the Gaussian Process (GP) framework [18] for classification and the transfer of complementary information from the EEG signal. Intuitively, given a kernel $\mathbf{K}$, GP framework induces a smoothness constraint implying that two points that are similar (*i.e.,* high kernel value) should have same labels. Given this smoothness constraint, the information from training data points $\mathbf{X}_L$ with labels $\mathbf{t}_L$ is combined in a probabilistic manner to compute a posterior distribution over a class label t_u of an unlabeled point. Specifically, under a zero-mean Gaussian noise model[2] parameterized with variance σ^2 this posterior distribution takes a very simple form and can be written as a Gaussian: $p(t_u|\mathbf{X}, \mathbf{t}_L) \sim \mathcal{N}(\bar{t}_u, \Sigma_u + \sigma^2)$, where:

$$\bar{t}_u = \mathbf{k}_L(\mathbf{x}_u)^T(\sigma^2\mathbf{I} + \mathbf{K}_{LL})^{-1}\mathbf{t}_L$$
$$\Sigma_u = k(\mathbf{x}_u, \mathbf{x}_u) - \mathbf{k}_L(\mathbf{x}_u)^T(\sigma^2\mathbf{I} + \mathbf{K}_{LL})^{-1}\mathbf{k}_L(\mathbf{x}_u).$$

Here, $\mathbf{k}_L(\mathbf{x}_u)$ is the vector of kernel function evaluations with n training points, and $\mathbf{K}_{LL} = \{k(\mathbf{x}_i, \mathbf{x}_j)\}$, is the training covariance, where $\mathbf{x}_i, \mathbf{x}_j$ are in the training set. One of the main advantages of using the GP framework is that, instead of receiving only a label, we get the whole posterior distribution which is very useful in the setting where we want to transfer appropriate information to the computer vision system. Intuitively, using the GP framework, we can show unlabeled images to humans and then use the EEG signals to compute a posterior distribution over class labels: $p(t_u^{eeg}|\mathbf{X}^{eeg}, \mathbf{t}_L) \sim \mathcal{N}(\bar{t}_u^{eeg}, \Sigma_u^{eeg})$. This posterior distribution also indicates the confidence of the BCI classifier; hence, we now have the option to appropriately weight the information before appending the training set. This is achieved by first considering $\bar{t}_u^{eeg}$ as the label corresponding to the unlabeled image and then assuming that the Gaussian noise model for this particular example has a variance $\Sigma_u^{eeg} + \sigma^2$. Specifically, a test image $\mathbf{x}_{test}$ can be classified by PMK based computer vision classifier as

[2]This method is referred to as least-squares classification in the literature (see Section 6.5 of [18]) and often demonstrates performance competitive with other kernel based techniques.

$f(\mathbf{x}_{test}) = Sign(t^v_{test})$, where:

$$t^v_{test} = \begin{bmatrix} \mathbf{k}_L(\mathbf{x}_{test}) \\ k(\mathbf{x}_{test}, \mathbf{x}_u) \end{bmatrix}^T \left(\begin{bmatrix} \sigma^2 \mathbf{I} & 0 \\ 0 & \Sigma^{eeg}_u + \sigma^2 \end{bmatrix} + \mathbf{K} \right)^{-1} \begin{bmatrix} \mathbf{t}_L \\ \bar{t}^{eeg}_u \end{bmatrix}$$

Here, again $\mathbf{k}(\cdot)$ and $\mathbf{K}$ represent the vector of kernel function and the kernel matrix respectively, but evaluated using the pyramid match formulation (PMK). Also, σ^2 is the variance parameter for the noise model corresponding to the PMK based classifier. When the BCI classifier has low confidence (large Σ^{eeg}_u), $\bar{t}^{eeg}_u$ will not affect the final classification; however, in light of a confident classification, $\bar{t}^{eeg}_u$ will provide useful information to the computer vision system. Thus, under the GP framework, we can incorporate the information provided by the BCI system in a manner that preserves the estimation of uncertainty about the label. Note, that the GP formulation is described for a binary classification task. We extend the binary formulation to a three-class scenario using a one-versus-all strategy. We would like to point out that, for simplicity, we described a mathematical expression for the case of analyzing a single unlabeled image. However, the same formulation readily extends to sets of unlabeled images.

For this experiment, we again look at the data set described for the first experimental setup. However, here we consider the scenario where a subset of the images has been labeled and we explore the value of presenting the rest of the unlabeled images to humans with boosting the recognition, based only on the use of the computer-vision algorithm. Thus, our aim is to train a computer-vision system *with implicit processing* (*i.e.*, without requiring manual annotation of images). Figure 7 shows the accuracy of classification for the test set as the number of images from the unlabeled pool that are presented to users are varied (starting with randomly selected set of ten images that were labeled and processed). These results are reported by averaging over 1000 runs and show performance of recognition with only the computer vision modality (*i.e.*, for test set we have only vision features). Different curves in the figure correspond to different numbers of randomly selected people contributing brain processing. We see that the accuracy of the classification using only computer vision increases even when unlabeled images are being presented to subjects. Furthermore, we see significant gains with just a single human observer and the classification accuracy also increases as we include additional participants. As observed from the figure, this gain in accuracy saturates with 7 humans, which suggests that there is correlation among EEG signals collected from different users. The experiments described in this section demonstrate that brain computation can boost the accuracy of a computer-vision classification system significantly.

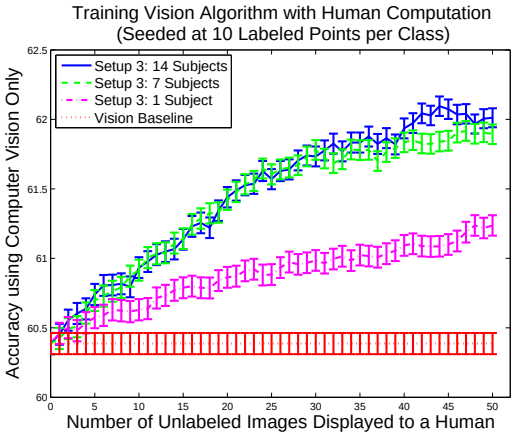

Figure 7. Results demonstrating that pure vision-based classification can benefit from human computation.

4. Discussion

We have described the evolution of our research proposing over three stages of modeling and experimentation. The efforts have led to a methodology that demonstrates how EEG signals, even when collected during ongoing activity of participants that is not necessarily directed at an explicit visual-recognition task, can augment and inform the development of machine-based object categorization systems. We have shown high accuracies for classifying faces, animate objects, and inanimate objects using this technique. It is not surprising that our best accuracies come in distinguishing faces, human or otherwise, from other objects. The human brain has been shown to have special skills at recognizing faces. Early research showed that a specific part of the brain located on the ventral surface of the temporal lobe specializes in facial recognition [20]. This area has been called the fusiform face area (FFA). More recent studies further suggest that this area may actually process categorical or even fine information about well-known objects [14]. For example, researchers have shown activity in this area when car experts were identifying cars and bird experts were identifying birds, but not vice versa [8]. Given that humans are extremely facile at recognizing and distinguishing objects such as faces, even when they are not trying, and that our best machines still struggle with this task, our approach leverages the obvious complementarity to improve the state of the art in automated object categorization.

One limitation in the current work is that we have not fully explored the granularity of categories for which we can continue to harness useful signal from the brain. This is especially true in the current experimental paradigm, in which the user is told nothing about the task beyond making sure they are looking at the images displayed. For example, we would ideally like to be able to distinguish different emotions or facial gestures within the general class

of faces. We might perhaps also be interested in distinguishing human faces from other kinds such as animals or such variants as sketched faces and other symbolic representations of faces. We are currently exploring implicitly priming paradigms in order to provide our methodology with finer classification granularity. By priming users with the classes we are interested in discriminating, for example "sad," "happy," and "surprised," we hope to influence the way that human subjects process the images, even without explicitly focusing on the specific task. Success with this would also allow us to explore a much wider range of categories than we are currently able to detect.

Along with expanding the current paradigm with implicit priming, we are also exploring how we can sense the detection of a face, or other object, within a continuous video stream. Our current processing methodology requires that we know exactly the onset of the image of interest. This is difficult to measure or infer, either in complex scenes, in which the user can be looking at any number of objects, or in continuous video in which the onset of the object is not always clear. This work represents only the first few steps towards our longer-term vision of human-aided computation, or more broadly, complementary computing, where human and machine computation are appropriately combined based on considerations of the value of the information streams as well as the cost of efforts and availability of humans and computing resources. Beyond building better vision systems, we suspect that future results may lead to new insights about the capabilities of human vision in comparison to the best computer vision methods.

On next steps, we seek to expand the set of objects that we are able to classify and also to extend the framework to analyze videos. We also aim to extend and test this system on larger, more varied data sets. In another direction of work, we are interested in the potential to use computations of the expected value of information to understand the costs and benefits of employing and sequencing different sources and types of labeling and computing efforts, given a task at hand. Such analyses for triaging attention and effort will likely be important in executing on the longer-term dream of complementary computing for visual recognition tasks.

5. Conclusion

We have presented methods for integrating human and machine computation to enhance the accuracy of performance on a visual recognition task. We have described how we can combine base computation as well as labeling effort using a soft-labeling scheme. The complementary information from the EEG observations can be used to appropriately analyze unlabeled data, which in turn helps the computer vision algorithm by providing a bigger corpus to train upon. Our empirical results demonstrate that such a combination of computer vision and human information processing can yield significant gains in accuracy for the task of object categorization. There is much to be done in exploring opportunities with human-aided computing and complementary computing. The results to date excite us about directions and prospects and we expect to see a stream of interesting future results and methods.

References

[1] *http://labelme.csail.mit.edu/*.

[2] Special issue on brain-computer interface technology: The third international meeting. *IEEE Transactions on Neural Systen Rehabilitation Engineering*, 2006.

[3] E. Bart and S. Ullman. Cross-generalization: Learning novel classes from a single example by feature replacement. In *CVPR*, 2005.

[4] P. N. Bennett, S. T. Dumais, and E. Horvitz. The combination of text classifiers using reliability indicators. *Information Retrieval*, 8, 2005.

[5] L. A. Farwell and E. Donchin. Talking off the top of your head: toward a mental prosthesis utilizing event-related brain potentials. *Electroencephalography & Clinical Neurophysiology*, 70(6):510–23, 1988.

[6] L. Fei-Fei, R. Fergus, and P. Perona. Learning generative visual models from few training examples: an incremental bayesian Approach Tested on 101 Object Cateories. In *Workshop on Generative Model Based Vision*, 2004.

[7] B. Fisch. *Fisch & Spehlmann's EEG primer: Basic principles of digital and analog EEG*. Elsevier: Amsterdam, 2005.

[8] I. Gauthier, P. Skudlarski, J. C. Gore, and A. W. Anderson. Expertise for cars and birds recruits brain areas involved in face recognition. *Nature Neuroscience*, 3(2), 2000.

[9] A. D. Gerson, L. C. Parra, and P. Sajda. Cortically-coupled computer vision for rapid image search. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 14(2):174–179, 2006.

[10] K. Grauman and T. Darrell. The pyramid match kernel: Discriminative classification with sets of image features. In *ICCV*, 2005.

[11] K. Grill-Spector. The neural basis of object perception. *Current opinion in neurobiology*, 13:1–8, 2003.

[12] H. H. Jasper. The ten-twenty electrode system of the international federation. *Electroencephalography and Clinical Neurophysiology*, 10:371–375, 1958.

[13] J. S. Johnson and B. A. Olshausen. The earliest EEG signatures of object recognition in a cued-target task are postsensory. *Journal of Vision*, 5(4):299–312, 2005.

[14] N. Kanwisher, J. McDermott, and M. M. Chun. The fusiform face area: a module in human extrastriate cortex specialized for face perception. *Neuroscience*, 17(11), 1997.

[15] A. Kapoor, K. Grauman, R. Urtasun, and T. Darrell. Active learning with Gaussian Processes for object categorization. In *ICCV*, 2007.

[16] A. Kapoor, P. Shenoy, and D. Tan. Combining brain computer interfaces with vision for object categorization. In *CVPR*, 2008.

[17] S. Lazebnik, C. Schmid, and J. Ponce. Beyond bags of features: spatial pyramid matching for recognizing natural scene categories. In *CVPR*, 2006.

[18] C. E. Rasmusen and C. Williams. *Gaussian Processes for Machine Learning*. MIT Press, 2006.

[19] B. Roisson, I. Gauthier, J. F. Delvenne, M. Tarr, R. Bruyer, and M. Crommelinck. Does the N170 occipito-temporal component reflect a face-specific structural encoding stage? In *Object Perception and Memory 1999*, 1999.

[20] J. Sergent, S. Ohta, and B. MacDonald. Functional neuroanatomy of face and object processing. *Brain*, 115(1), 1992.

[21] P. Shenoy and D. Tan. Human-aided computing: Utilizing implicit human processing to classify images. In *ACM CHI*, 2008.

[22] K. Toyama and E. Horvitz. Bayesian modality fusion: Probabilistic integration of multiple vision algorithms for head tracking. In *ACCV*, 2000.

[23] L. von Ahn and L. Dabbish. Labeling images with a computer game. In *ACM CHI*, 2004.

[24] L. von Ahn, R. Liu, and M. Blum. Peekaboom: A game for locating objects in images. In *ACM CHI*, 2006.

[25] C. Wallraven, B. Caputo, and A. Graf. Recognition with local features: the kernel recipe. In *ICCV*, 2003.

[26] H. Zhang, A. Berg, M. Maire, and J. Malik. SVM-KNN: Discriminative nearest neighbor classification for visual category recognition. In *CVPR*, 2006.