Vision Research 116 (2015) 165–178

Contents lists available at ScienceDirect

Vision Research

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

Intrinsic and extrinsic effects on image memorability

Zoya Bylinskii [a,b,*], Phillip Isola [b,c], Constance Bainbridge [b], Antonio Torralba [a,b], Aude Oliva [b]

[a] Department of Electrical Engineering and Computer Science, MIT, Cambridge 02141, USA
[b] Computer Science and Artificial Intelligence Laboratory, MIT, Cambridge 02141, USA
[c] Department of Brain and Cognitive Sciences, MIT, Cambridge 02141, USA

A R T I C L E   I N F O

Article history:
Received 4 August 2014
Received in revised form 3 March 2015
Available online 20 March 2015

Keywords:
Image memorability
Eye movements
Scene dataset
Fine-grained categories
Visual distinctiveness
Context

A B S T R A C T

Previous studies have identified that images carry the attribute of memorability, a predictive value of whether a novel image will be later remembered or forgotten. Here we investigate the interplay between intrinsic and extrinsic factors that affect image memorability. First, we find that intrinsic differences in memorability exist at a finer-grained scale than previously documented. Second, we test two extrinsic factors: image context and observer behavior. Building on prior findings that images that are distinct with respect to their context are better remembered, we propose an information-theoretic model of image distinctiveness. Our model can automatically predict how changes in context change the memorability of natural images. In addition to context, we study a second extrinsic factor: where an observer looks while memorizing an image. It turns out that eye movements provide additional information that can predict whether or not an image will be remembered, on a trial-by-trial basis. Together, by considering both intrinsic and extrinsic effects on memorability, we arrive at a more complete and fine-grained model of image memorability than previously available.

1. Introduction

Recent work on image memorability has shown that independent of observer, certain images are consistently remembered and others forgotten (Bainbridge, Isola, & Oliva, 2013; Borkin et al., 2013; Isola, Parikh, et al., 2011; Isola, Xiao, et al., 2011; Isola et al., 2014), indicating that memorability is an intrinsic property of images that can be estimated with computer vision features (Isola, Parikh, et al., 2011; Isola, Xiao, et al., 2011; Isola et al., 2014; Khosla, Xiao, Torralba, et al., 2012, Khosla et al., 2013). These previous image memorability studies raise a number of questions, including: does the consistency of human memory generalize? How might extrinsic effects such as context and observer differences affect image memorability?

In this paper, we report that: (1) human consistency at remembering and forgetting images holds at a within-category level, and (2) extrinsic effects predictably affect whether an image will be later remembered or forgotten. Here we consider the effects of the context in which images are seen, as well as the observer's eye movement patterns on a trial-by-trial basis.

Previous work on image memorability has not computationally addressed either image context or trial-by-trial observer behavior. Moreover, although many decades of prior research on memory have considered context and the effects of item/image distinctiveness of memorability (Hunt & Worthen, 2006; Konkle et al., 2010; Nairne, 2006; Standing, 1973), these effects have not been rigorously quantified on large datasets of natural scenes. Prior work has relied on subjective human judgments of distinctiveness (Bainbridge et al., 2013; Konkle et al., 2010). In contrast, we provide an objective, automatic measure: we model distinctiveness as an information-theoretic property computable from raw visual data.

For our studies, we collected the *FIne-GRained Image Memorability (**FIGRIM**) dataset*[1] composed of over 9K images, which we used to test human memory performance on 21 different scene categories, each containing hundreds of images. We used this dataset to collect memorability scores for 1754 target images, whereby we systematically varied the image context. In this paper we refer to the set of images from which the experimental sequence is sampled as **image context**. We present an information-theoretic framework to quantify context differences and image distinctiveness using state-of-the-art computer vision features, and we show correlations

* Corresponding author at: 32-D542, 32 Vassar St., Cambridge, MA 02141, USA.
E-mail address: zoya@mit.edu (Z. Bylinskii).

[1] We are publicly releasing the full *FIGRIM* dataset with popular image features precomputed for all 9K images of the dataset, as well as memorability scores for each of the 1754 target images. For the target images, we provide separate memorability scores for the image presented in the context of its own scene category and different scene categories. Available at: <http://figrim.mit.edu/>.

http://dx.doi.org/10.1016/j.visres.2015.03.005
0042-6989/

with image memorability scores. We discuss which images are most affected by context to gain a better understanding of the interplay between intrinsic and extrinsic factors on image memorability.

To account for additional extrinsic effects caused by the variation in observer behavior from trial to trial, we collected eyetracking data for over 2.7 K of the *FIGRIM* images. For 630 target images and using eye movements alone we can predict, on a trial-to-trial basis, which images will be remembered and which forgotten with 66% accuracy. Thus, we demonstrate how eye movements have predictive power on a trial-by-trial basis for image memorability.

2. Background

2.1. Image memorability

Recent work in image memorability (Bainbridge et al., 2013; Borkin et al., 2013; Isola, Parikh, et al., 2011; Isola, Xiao, et al., 2011; Isola et al., 2014) has reported high consistency rates among participants in terms of which images are remembered and which forgotten, indicating that memorability is a property that is intrinsic to the image, despite individual differences between observers. The high consistency was first demonstrated for a database of images from hundreds of scene categories (Isola, Xiao, et al., 2011), and later extended to narrower classes of images – faces (Bainbridge et al., 2013) and visualizations (Borkin et al., 2013). In this paper, we show that this consistency is not a special property of the stimuli considered, and that it can not be explained away by a simple distinction between images (e.g. indoor scenes tend to be memorable, outdoor scenes forgettable). We demonstrate that the high consistencies hold within 21 different indoor and outdoor scene categories, each consisting of hundreds of instances. This is the first image memorability study to consider fine-grained scene categories. Previous studies have shown that image memorability can be computationally predicted from image features (Isola, Xiao, et al., 2011) which opens up applications such as automatically generating memorability maps for images (Khosla, Xiao, Torralba, et al., 2012), modifying image memorability (Khosla et al., 2013; Khosla, Xiao, Isola, 2012), and designing better data visualizations (Borkin et al., 2013). In this paper, we additionally model extrinsic effects on memorability, which have not yet been explored in the image memorability literature, and can open up new application areas.

2.2. Distinctiveness in visual long-term memory

Previous studies have suggested that items that stand out from (and thus do not compete with) their context are better remembered (Attneave, 1959; Eysenck, 1979; Hunt & Worthen, 2006; Konkle et al., 2010; Rawsona & Overscheldeb, 2008; Schmidt, 1985; Standing, 1973; Wiseman & Neisser, 1974; Vogt & Magnussen, 2007; von Restorff, 1933). For instance, Standing observed a large long-term memory capacity for images that depict oddities (Standing, 1973). Konkle et al. demonstrated that object categories with conceptually distinctive exemplars showed less interference in memory as the number of exemplars increased (Konkle et al., 2010). Additionally, for the specific categories of face images, studies have reported that a distinctive or atypical face (i.e., a face distant from the average) is more likely to be remembered (Bartlett, Hurry, & Thorley, 1984; Bruce, Burton, & Dench, 1994; Valentine, 1991). In the domain of data visualizations, Borkin et al. noticed that unique visualization types had significantly higher memorability scores than common graphs and that novel and unexpected visualizations were better remembered (Borkin et al., 2013). In this paper, we quantify the intuitions that distinctive images are more memorable using an information theoretic framework, and we compute the distinctiveness of images with reference to their image context (the set of images from which the experimental sequence is sampled). We steer away from subjective human ratings, and instead compute statistics over automatically-extracted image features. By systematically varying the image context across experiments, we are able to computationally model the change in context at the feature level, and predict corresponding changes in image memorability.

2.3. Memorability and visual attention

Little work has considered the intersection between image memorability and visual attention (Bulling & Roggen, 2011; Foulsham & Underwood, 2008; Mancas & Le Meur, 2013; Noton & Stark, 1971). Mancas and Le Meur (2013) used saliency features to show a slight improvement over the automatic image memorability predictions in Isola, Xiao, et al. (2011). We refer to image memorability as a **population predictor** because it ignores trial-by-trial variability, effectively averaging over a population of participants or experiments. Thus, Mancas et al. used saliency to improve a population predictor. We, instead, use eye-movements to improve the trial-by-trial predictions of memory for specific individuals (an **individual trial predictor**). Bulling and Roggen (2011) used eye movement features to predict image familiarity, classifying whether images have been seen before or not. They assumed that all images seen again are remembered, particularly due to the long exposure times (10 s) used per image, and by testing on a small dataset of 20 faces. They also used eye movement analysis as a *population predictor* to decide whether an image was *previously seen*, while we use eye movement analysis as an *individual trial predictor*, taking into account individual differences in making predictions of whether an image will be *later remembered.*

2.4. Decoding task using eye movements

Our work is also related to recent studies on the use of eye movements for decoding an observer's task (Borji & Itti, 2014; Greene, Liu, & Wolfe, 2012). These studies considered features extracted from the eye movements of individual participants to determine the task they are performing (e.g., what question they are answering about an image), modeled on the original Yarbus experiment (Yarbus, 1967). These studies utilized a very small set of images (ranging from 15 to 64) with a very constrained theme (grayscale photographs taken between 1930 and 1979 with at least two people (Greene et al., 2012); paintings depicting "an unexpected visitor" (Borji & Itti, 2014)). In our study, we measured the eye movements of participants on 630 target images sampled from 21 different indoor and outdoor scene categories. We extracted features from eye movements to determine whether or not an image is correctly encoded (measured by whether it is correctly recognized on a successive exposure). We were able to solve our decoding task using only 2 s of viewing time per image, whereas the previous studies worked with durations of 10 s (Bulling & Roggen, 2011; Greene et al., 2012), 30 s (Borji & Itti, 2014), 50 s (Tatler et al., 2010), and 60 s (Borji & Itti, 2014). For this purpose, we learned image-specific classifiers to distinguish fixations on one image versus fixations on other images.

3. Memorability experiments

3.1. FIGRIM dataset

We created a novel dataset by sampling high-resolution (at least 700 × 700 px) images from 21 different indoor and outdoor