# What makes an image memorable?

Phillip Isola Jianxiong Xiao Antonio Torralba Aude Oliva
{phillipi, jxiao, torralba, oliva}@mit.edu
Massachusetts Institute of Technology

## Abstract

*When glancing at a magazine, or browsing the Internet, we are continuously being exposed to photographs. Despite of this overflow of visual information, humans are extremely good at remembering thousands of pictures along with some of their visual details. But not all images are equal in memory. Some stitch to our minds, and other are forgotten. In this paper we focus on the problem of predicting how memorable an image will be. We show that memorability is a stable property of an image that is shared across different viewers. We introduce a database for which we have measured the probability that each picture will be remembered after a single view. We analyze image features and labels that contribute to making an image memorable, and we train a predictor based on global image descriptors. We find that predicting image memorability is a task that can be addressed with current computer vision techniques. Whereas making memorable images is a challenging task in visualization and photography, this work is a first attempt to quantify this useful quality of images.*

## 1. Introduction

People have a remarkable ability to remember particular images in long-term memory, even those depicting every day scenes and events [23], or the shapes of arbitrary forms [19]. We do not just remember the gist of a picture, but we are able to recognize which precise image we saw and some of its details [1, 19]. In addition to remembering particular images as icons, we also have the intuition that not all images are remembered equally. Some pictures stick in our minds whereas others fade away. The reasons why some images are remembered are varied; some pictures might contain friends, a fun event involving family members, or a particular moment during a trip. Other images might not contain any recognizable monuments or people and yet also be highly memorable [1, 12, 11, 19]. In this paper we are interested in this latter group of pictures: what are the intrinsic image features that may make an image memorable?

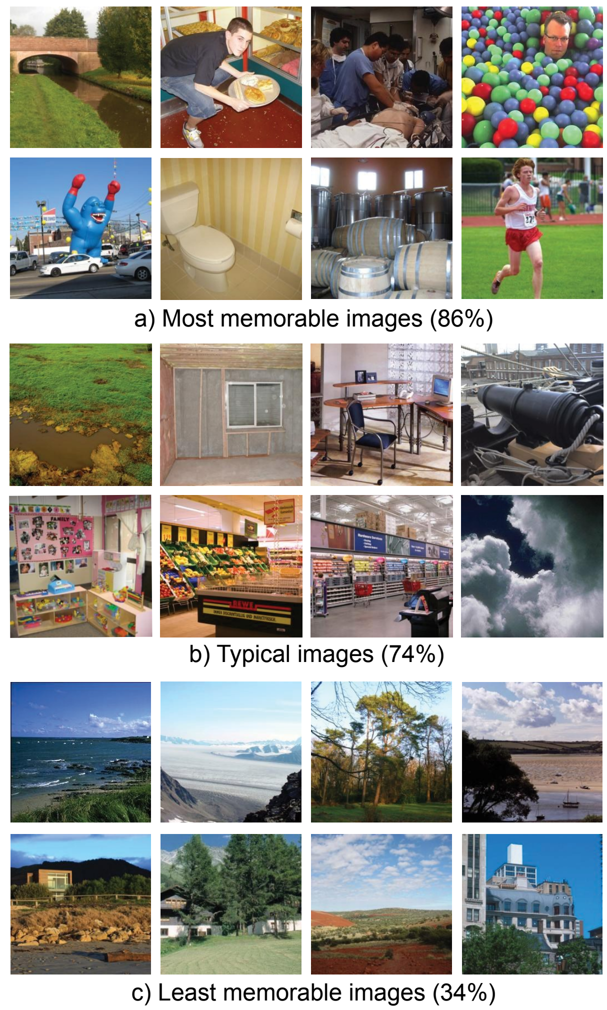



Figure 1. Each set of 8 images was selected according to one half of the participants in our study as being (a) the 8 most memorable images, (b) 8 average memorability images, and (c) the 8 least memorable images. The number in parentheses gives the percent of times that these images were remembered by an independent set of participants.

Whereas most studies on human memory have been devoted to evaluating how good average picture memory can be, no work has systematically studied individual differ-

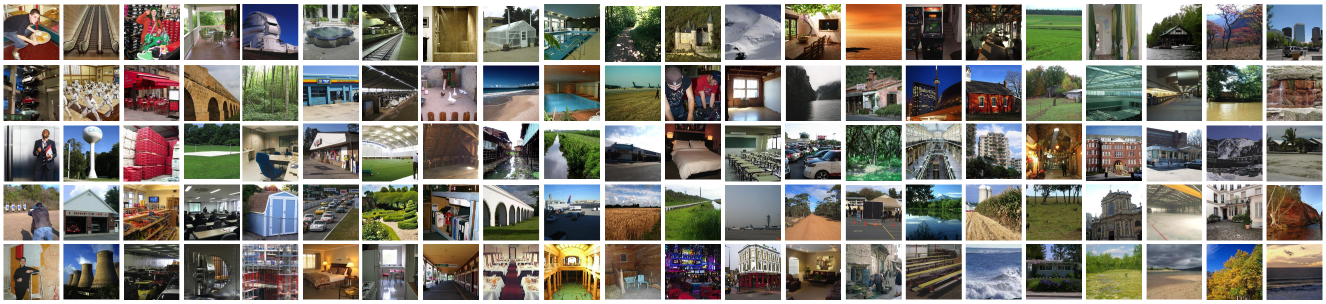


Figure 2. Sample of the database used for the memory study. The images are sorted from more memorable (left) to less memorable (right).

ences between images and if those differences are consistent across different viewers. Can a specific natural image be memorable to all of us, and can we estimate what makes it distinctive?

Similar to other image properties such as photographic quality, composition, and beauty, image memorability is likely to depend on the user context and is likely to be subject to some inter-subject variability [9]. However, despite this expected variability when evaluating subjective properties of images, there is also a sufficiently large degree of agreement between users that suggests it is possible to devise automatic systems to estimate these properties directly from images, ignoring user differences. As opposed to other image properties, there are no previous studies that try to quantify individual images in terms of how memorable they are, and there are no computer vision systems that try to predict image memorability. This is contrary to many other photographic properties that have been addressed in the literature such as photo quality [16], saliency [10], attractiveness [15], composition [8, 18], color harmony [4], and object importance [22]. Also, there are no databases of photographs calibrated in terms of the degree of memorability for each image.

In this paper, we characterize an image's memorability as the probability that an observer will detect a repetition of a photograph a few minutes after exposition, when presented amidst a stream of images. This setting allows us to gather memory data for a large collection of images, and determine which images left a trace in long-term memory[1]. We mine this data to identify which features of the images correlate with memorability, and we train memorability predictors on these features. Whereas further studies will be needed to see if the predictions from this data will apply to more general viewing conditions, the present work constitutes an initial benchmark. Studying what makes images memorable, and how to predict memorability from image information alone, may have many applications, including choosing a book cover, designing a logo, organizing images in a photo album, and selecting an image to decorate a website.

[1]Short-term memory typically can only hold 3 or 4 items at once [5]; since participants in our experiment had to hold many more images in memory, it is unlikely short term memory was driving performance.

## 2. Is image memorability predictable?

Although we all have the intuition that some images will capture our attention and will be easier to remember than others, quantifying this intuition has not been addressed in previous experiments. Are the images remembered by one person more likely to be remembered also by somebody else? In this section, we characterize the consistency of image memory. In order to do so, we built a database depicting a variety of images (Figure 2), and we measured the probability of remembering each image over a large population of users. These data allow us to quantify the degree to which image memory measured on one set of participants is consistent with image memory measured on an independent set of participants.

### 2.1. The Memory Game: measuring memorability

In order to measure image memorability, we presented workers on Amazon Mechanical Turk with a Visual Memory Game. In the game, participants viewed a sequence of images, each of which was displayed for 1 second, with a 1.4 second gap in between image presentations (Figure 3). Their task was to press the space bar whenever they saw an identical repeat of an image at any time in the sequence [1] [12]. Participants received feedback whenever they pressed a key (a green symbol shown at the center of the screen for correct detection, and a gray X for an error).

Image sequences were broken up into levels that consisted of 120 images each. Each level took 4.8 minutes to perform. At the end of each level, the participant saw his or her correct response average score for that level, and was allowed to take a short break. Participants could complete at most 30 levels, and were able to exit the game at any time. A total of 665 workers from Mechanical Turk ($> 95\%$ approval rate in Amazon's system) performed the game. Over 90% of our data came from 347 of these workers. We payed workers \$0.30 per level in proportion to the

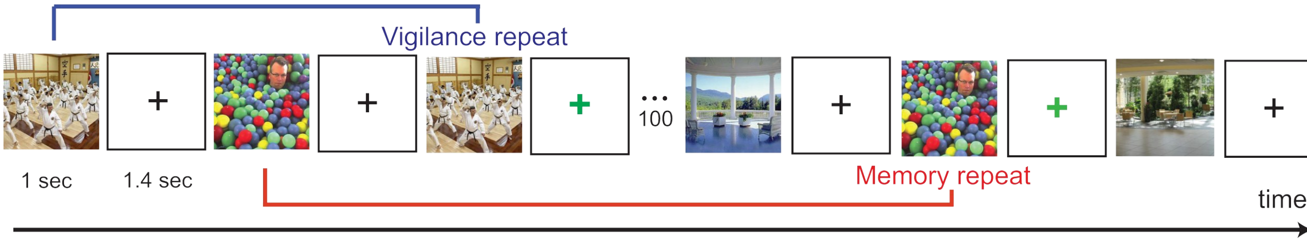



Figure 3. Mechanical Turk workers played a "Memory Game" in which they watched for repeats in a long stream of images.

amount of the level completed, plus a $0.10 bonus per fully completed level. This adds up to about $5 per hour. The average worker stayed in the game for over 13 levels.

Unbeknownst to the participants, the sequence of images was composed of 'targets' (2222 images) and 'fillers' (8220 images). Target and filler images represented a random sampling of the scene categories from the SUN dataset [24]. All images were scaled and cropped about their centers to be 256x256 pixels. The role of the fillers was twofold: first, they provided spacing between the first and second repetition of a target; second, responses on repeated fillers constituted a 'vigilance task' that allowed us to continuously check that participants were attentive to the task [1, 12]. Repeats occurred on the fillers with a spacing of 1-7 images, and on the targets with a spacing of 91-109 images. Each target was sequenced to repeat exactly once, and each filler was presented at most once, unless it was a vigilance task filler, in which case it was sequenced to repeat exactly once.

Stringent criteria were used to continuously screen worker performance once they entered the game. First, the game automatically ended whenever a participant fell below a 50% success rate on the last 10 vigilance task repeats or above a 50% error rate on the last 30 non-repeat images. When this happened, all data collected on the current level was discarded. Rejection criterion reset after each level. If a participant failed any of the vigilance criteria, they were flagged. After receiving three such flags they were blocked from further participation in the experiment. Otherwise, participants were able to restart the game as many times as they wished until completing the max 30 levels. Upon each restart, the sequence was reset so that the participant would never see an image they had seen in a previous session. Finally, a qualification and training 'demo' preceeded the actual memory game levels.

After collecting the data, we assigned a 'memorability score' to each image, defined as the percentage of correct detections by participants. On average, each image was scored by 78 participants. The average memorability score was 67.5% (SD of 13.6%). Average false alarm rate was 10.7% (SD of 7.6%). As the false alarm rate was low in comparison with correct detections, correct detections are

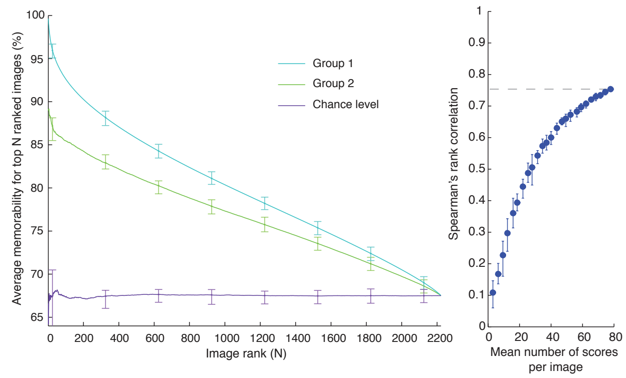



Figure 4. Measures of human consistency. Participants were split into two independent sets, group 1 and group 2. Left: Images were ranked by memorability scores from participants in group 1 and plotted against the cumulative average memorability scores given by participants in the two groups. Right: Spearman's rank correlation between subject groups 1 and 2 as a function of the mean number of scores per image. Both left and right analyses were repeated for 25 random splits and mean results are plotted. Error bars show 80% confidence intervals over the 25 trials.

unlikely to be lucky confusions. Specifically, we expect lucky confusions to account for no more than 10.7% of correct detections on average. Therefore, we believe our memorability scores are a good measure of correct memories.

## 2.2. Consistency analysis

To evaluate human consistency, we split our participant pool into two independent halves, and quantified how well image scores measured on the first half of the participants matched image scores measured on the second half of the participants. Averaging over 25 random split half trials, we calculated a Spearman's rank correlation ($\rho$) of 0.75 between these two sets of scores. We additionally examined consistency with a variant of a precision-recall task: We sorted the images by their score given by the first half of the participants and then calculated cumulative average memorability, according to the second half of the participants, as we move across this ordering (Figure 4). For example, if we select the top 100 most memorable images according to the first half of the participants (whose average memorability for this set is 93%), the second half of the participants will give an average memorability of 85% for that set. Figure 1 shows sample image sets selected and measured in this fashion. This analysis shows how high human-to-human memorability consistency can be.

Thus, our data has enough consistency that it should be possible to predict image memorability. Individual differences and random variability in the context each participant saw add noise to the estimation; nonetheless, this level of consistency suggests that information intrinsic to the images might be used by different people to remember them. In the next section, we search for this image information.

## 3. What makes an image memorable?

Among the many reasons why an image might be remembered by a viewer, we investigate first the following factors: color, simple image features, object statistics, object semantics, and scene semantics.

### 3.1. Color and simple image features

Are simple image features enough to determine whether or not an image will be memorable? We looked at the correlation between memorability and basic pixel statistics. Mean hue was weakly predictive of memory: as mean hue transitions from red to green to blue to purple, memorability tends to go down ($\rho$ = -0.16). This correlation may be due to blue and green outdoor landscapes being remembered less frequently than more warmly colored human faces and indoor scenes. Mean saturation and value, on the other hand, as well as the first three moments of the pixel intensity histogram, exhibited weaker correlations with memorability (Figure 5). These findings concord with other work that has shown that perceptual features are not retained in long term visual memory [11]. In order to make useful predictions, more descriptive features are likely necessary.

### 3.2. Object statistics

Object understanding is necessary to human picture memory [13]. Using LabelMe [20], each image in our target set was segmented into object regions and each of these segments was given an object class label by a human user (e.g "person," "mountain," "stethoscope") (see [3] for details). In this section, we quantify the degree to which our data can be explained by non-semantic object statistics.

Do such statistics predict memorability? For example do the number of objects one can attach to an image determine its memorability, or is it critical that an object class takes up a large portion of an image in order for the image to stick in memory? We find the answer to be no: none of these statistics make good predictions on their own. Simple object statistics (log number of objects, log mean pixel coverage over present object classes, and log max pixel coverage

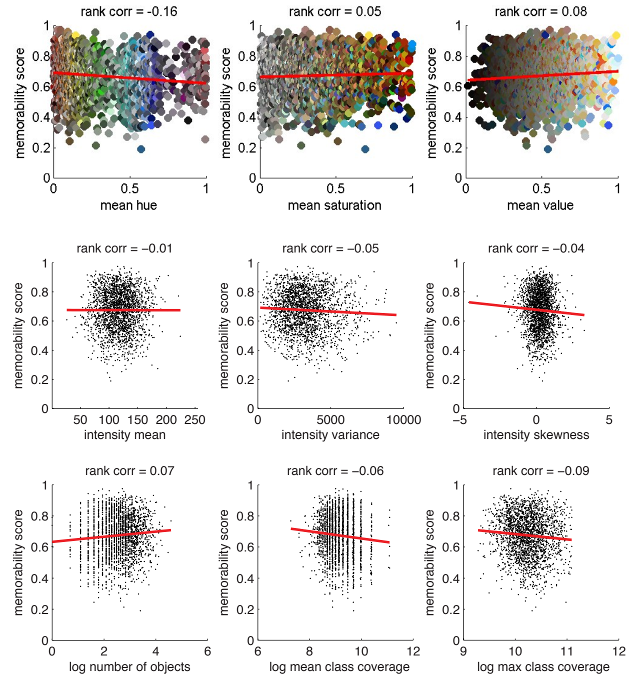



Figure 5. Simple image features, as well as non-semantic object statistics, do not correlate strongly with memorability score.

over object classes) did not correlate strongly with memorability ($\rho$ = 0.07, -0.06, and -0.09 respectively) (Figure 5).

To investigate the role of more subtle interactions between these statistics, we trained a support vector regression ($\epsilon$-SVR [2]) to map object statistics to memorability scores. For each image, we measured several object statistics: the number of objects in the image per class, and the number of pixels covered by objects of each class in the entire image as well as in each quadrant of the image. For each of these statistics, we thereby obtained joint distribution on (object class, statistic). We then marginalized across class to generate histograms of 'Object Counts', 'Object Areas', and, concatenating pixel coverage on the entire image with pixel coverage per quadrant, 'Multiscale Object Areas'. We used these histograms as features for our regression and applied histogram intersection kernels.

For each of 25 regression trials, we split both our image set and our participant set into two independent, random halves. We trained on one half of the images, which were scored by one half of the participants, and tested on the left out images, which were scored by the left out participants. During training, we performed grid search to choose cost and $\epsilon$ hyperparameters for each SVR.

We quantified the performance of our predictions similarly to how we analyzed human consistency above. First, we calculated average $\rho$ between predicted memorabilities and empirically measured memorabilities. Second, we sorted images by predicted score and selected various

| | Object Counts | Object Areas | Multiscale Object Areas | Object Label Presences | Labeled Object Counts | Labeled Object Areas | Labeled Multiscale Object Areas | Scene Category | Objects and Scenes | Other Humans |
|---|---|---|---|---|---|---|---|---|---|---|
| Top 20 | 68% | 67% | 73% | 84% | 82% | 84% | 84% | 81% | 85% | 86% |
| Top 100 | 68% | 68% | 73% | 79% | 79% | 82% | 82% | 78% | 82% | 84% |
| Bottom 100 | 67% | 64% | 64% | 57% | 57% | 56% | 56% | 57% | 55% | 47% |
| Bottom 20 | 67% | 63% | 65% | 55% | 54% | 53% | 52% | 55% | 53% | 40% |
| $\rho$ | 0.05 | 0.05 | 0.20 | 0.43 | 0.44 | 0.47 | 0.48 | 0.37 | 0.50 | 0.75 |

Table 1. Comparison of predicted versus measured memorabilities. Images are sorted into sets according to predictions made on the basis of a variety of object and scene features (denoted by column headings). Average empirically measured memorabilities are reported for each set. e.g. The "Top 20" row reports average empirical memorability over the images with the top 20 highest predicted memorabilities. $\rho$ is the Spearman rank correlation between predictions and measurements.

ranges of images in this order, examining average empirical memorability on these ranges (Table 1). As a baseline, we compared to a measure of the available consistency in our data, in which we predicted that each test set image would have the same memorability according to our test set participants as was measured by our training set participants ('Other Humans').

Quantified in this way, our regressions on object statistics appear ineffective at predicting memorability (Table 1). However, predictions made on the basis of the Multiscale Object Areas did begin to show substantial correlation with measured memorability scores ($\rho$ = 0.20). Unlike the Object Counts and Object Areas, the Multiscale Object Areas are sensitive to changes across the image. As a result, these features may have been able to identify cues such as "this image has a sky," while, according to the other statistics, a sky would have been indistinguishable from a similarly large segment, such as a closeup of a face.

### 3.3. Object and scene semantics

As demonstrated above, objects without semantics are not effective at predicting memorability. This is not surprising given the large role that semantics play in picture memory [13, 11]. To investigate the role of object semantics, we performed the same regression as above, except this time using the entire joint (object class, statistic) distributions as features. This gave us histograms of 'Labeled Object Counts', 'Labeled Object Areas', 'Labeled Multiscale Object Areas', and, thresholding the labeled object counts about zero, 'Object Label Presences'. Each image was also assigned a scene category label as described in [24] ('Scene Category'). As before, we applied histogram intersection kernels to each of these features. We also tested a combination of Labeled Multiscale Object Areas and Scene Category using a kernel sum ('Objects and Scenes').

Semantics boosted performance (Table 1). Even the Object Label Presences alone, which simply convey a set of semantic labels and otherwise do not describe anything about the pixels in an image, performed well above our best unlabeled object statistic, Multiscale Object Areas ($\rho$ = 0.43 and 0.20 respectively). Moreover, Scene Category, which just gives a single label per image, appears to summarize much of what makes an image memorable ($\rho$ = 0.37), and we do best when we combine both object and scene semantic information ($\rho$ = 0.50). These performances support the idea that object and scene semantics are a primary substrate of memorability [11, 12, 13].

### 3.4. Visualizing what makes an image memorable

Since object content appears to be important in determining whether or not an image will be remembered, we further investigated the contribution of objects by visualizing object-based "memory maps" for each image. These maps shade each object according to how much the object adds to, or subtracts from, the image's predicted memorability. More precisely, to quantify the contribution of an object $i$ to an image, we take a prediction function, $f$, that maps object features to memorability scores and calculate how its prediction $m$ changes when we zero features associated with object $i$ from the current image's feature vector, $(a_1, \cdots, a_n)$. This gives us a score $s_i$ for each object in a given image:

$$m_1 = f(a_1, \cdots, a_i, \cdots, a_n) \quad (1)$$
$$m_2 = f(a_1, \cdots, 0, \cdots, a_n) \quad (2)$$
$$s_i = m_1 - m_2 \quad (3)$$

For the prediction function $f$, we use our SVR on Labeled Multiscale Object Areas, trained as above, and we plot memory maps on test set images (Figure 7). Thus, these maps show predictions as to what will make a novel image either remembered or not remembered. The validity of these maps is supported by the fact that the SVR we used to generate them (the Labeled Multiscale Object Areas regression) makes predictions that correlate relatively well with measured memory scores ($\rho$ = 0.48, see Table 1).

This visualization gives a sense of how objects contribute to the memorability of particular images. We are additionally interested in which objects are important across all images. We estimated an object's overall contribution as its contribution per image, calculated as above, averaged

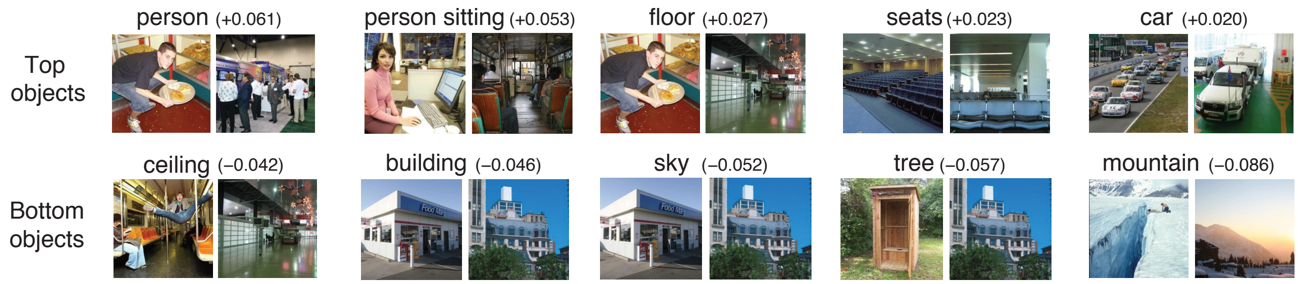



Figure 6. Objects sorted by their predicted impact on memorability. Next to each object name we report how much an image's predicted memorability will change, on average, when the object is included in the image's feature vector versus when it is not. For each object name, we also display two test set images that contain the object: on the left is the example image with the highest memorability score among all test set images that contain (over 4000 pixels of) the object. On the right is the example with the lowest score. Only objects that appear (cover over 4000 pixels) in at least 20 images in our training set are considered.

across all test set images in which it appears with substantial size (covers over 4000 pixels). This method sorts objects into an intuitive ordering: people, interiors, foregrounds, and human-scale objects tend to contribute positively to memorability; exteriors, wide angle vistas, backgrounds, and natural scenes tend to contribute negatively to memorability (Figure 6).

Our analysis of object semantic regressions demonstrate that if a system knows which objects an image contains, it is able to predict memorability with a performance not too far from human consistency. But, how far can we go in predicting the memorability of a novel image without an object labeling stage? In other words, can we get the same utility based just off unlabeled image features?

| | Pixels | GIST | SIFT | SSIM | HOG 2x2 | All Global Features |
|---|---|---|---|---|---|---|
| Top 20 | 74% | 82% | 83% | 83% | 83% | 83% |
| Top 100 | 72% | 78% | 79% | 79% | 80% | 80% |
| Bottom 100 | 61% | 58% | 57% | 58% | 57% | 56% |
| Bottom 20 | 59% | 57% | 56% | 56% | 55% | 54% |
| $\rho$ | 0.22 | 0.38 | 0.41 | 0.43 | 0.43 | 0.46 |

Table 2. Comparison of global feature predictions versus empirically measured memory scores. Uses same measures as described in Table 1.

## 4. Predicting image memorability: The role of global features

As we have seen in the previous sections there is a significant degree of consistency between different sets of viewers on how memorable are individual images. In addition, we have seen that some of the consistency can be explained in terms of the objects present in the picture and the scene category. In order to build a predictor of how memorable an image is, we used similar approaches to works studying other subjective image properties [16, 18, 24].

As with the object regressions, we trained an SVR to map from features to memorability scores – this time using only features algorithmically extracted from the images. We tested a suite of global image descriptors that have been previously found to be effective at scene recognition tasks [24] as well as being able to predict the presence/absence of objects in images [7, 6, 21]. The facility of these features at predicting image semantics suggests that they may be able to predict, to some degree, those aspects of memorability that derive from image semantics.

These global features are GIST [17], SIFT [14], HOG2x2 [7, 6, 24], and SSIM [21]. We additionally looked at pixel histograms. We used an RBF kernel for GIST and histogram intersection kernels for the other features. Lastly, we also combined all these features with a kernel product ('All Global Features').

We evaluated performance in the same way as we evaluated the object regressions, and we found that the combination of global features performs best, achieving a rank correlation of 0.46. This correlation is less than human predictions, but close to our best predictions from labeled annotations. Figure 8 shows sample images from predicted sets. Figure 10 shows sample images on which our global features regression performed poorly.

To set a high watermark, and to get a sense of the redundancy between our image features and our annotations, we additionally trained an SVR on a kernel sum of all our global features plus Labeled Multiscale Object Areas and Scene Categories ('Global Features and Annotations'). This combination achieved a rank correlation of 0.54. See Table 2 and Figure 9 for detailed results.

## 5. Conclusions

Making memorable images is a challenging task in visualization and photography, and is generally presented as a vague concept hard to quantify. Surprisingly, there has been no previous attempt to systematically measure this property on image collections, and to apply computer vision tech-

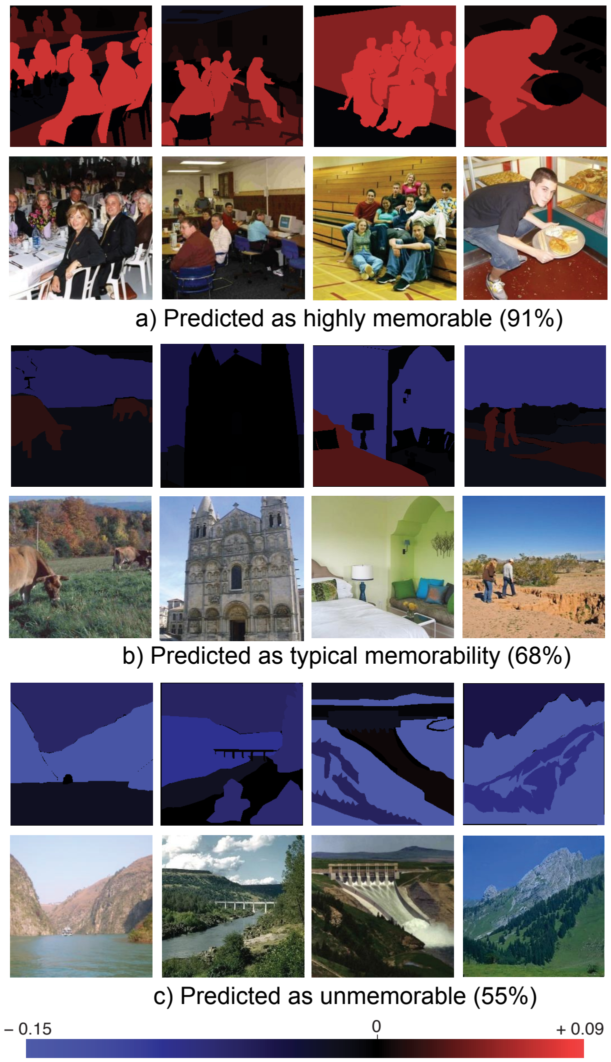



Figure 7. Visualization of how each object contributes to the memorability of sample images spanning a range of memorability predictions. We estimate contribution as the difference between predicted memorability when the object is included in the image versus when it is removed from the image. In red we show objects that contribute to higher predicted memorability and in blue are objects that contribute to lower predicted memorability. Brightness is proportional to the magnitude of the contribution. Average measured memorability of each sample set is given in parentheses.

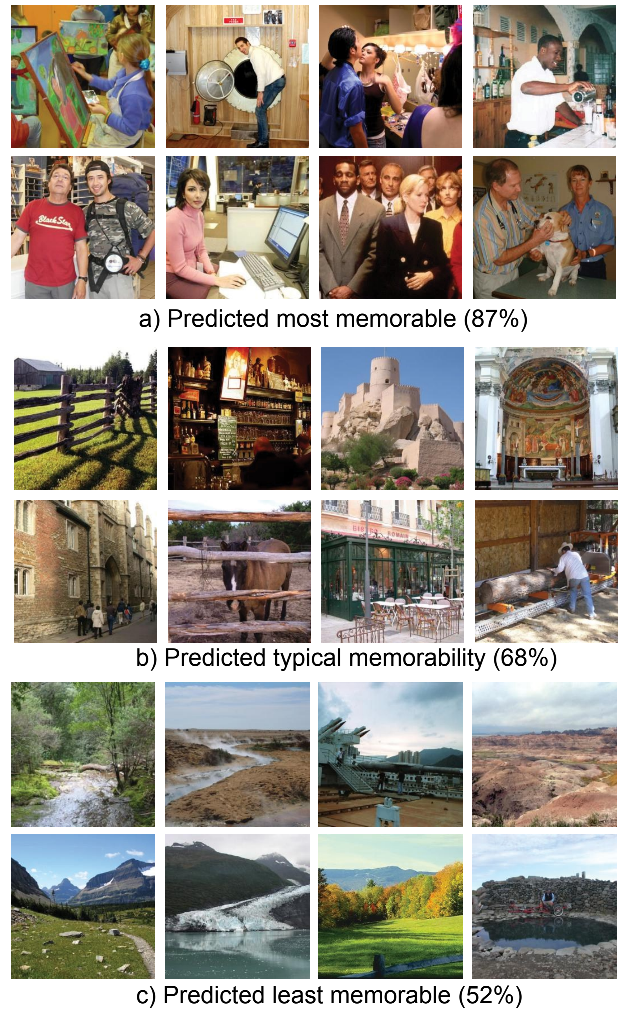



Figure 8. The 8 images predicted, on the basis of global image features, as being the most memorable out of all test set images (a), 8 images with average memorability predictions (b), and the 8 images predicted as being the least memorable of all test set images (c) . The number in parentheses gives the mean empirical memorability score for images in each set. The predictions produce clear visual distinctions, but may fail to notice more subtle cues that make certain images more memorable than others.

niques to extract memorability automatically. Measuring subjective properties of photographs is an active domain of research with numerous applications. For instance, image memorability could be used to extract, from a collection of images, the ones that are more likely to be remembered by viewers. This could be applied to selecting images for illustrations, covers, user interfaces, etc.

In this paper we have shown that predicting image memorability is a task that can be addressed with current computer vision techniques. We have measured memorability using a restricted experimental setting in order to obtain a meaningful quantity: we defined an image's memorability score as the probability that a viewer will detect a repeat of the image within a stream of pictures. We have shown that there is a large degree of consistency among different viewers, and that some images are more memorable than others even when there are no familiar elements (such as relatives or famous monuments). This work is a first attempt to quantify this useful quality of individual images. In future work it will be interesting to investigate the relationship between image memorability and other measures such as object im-

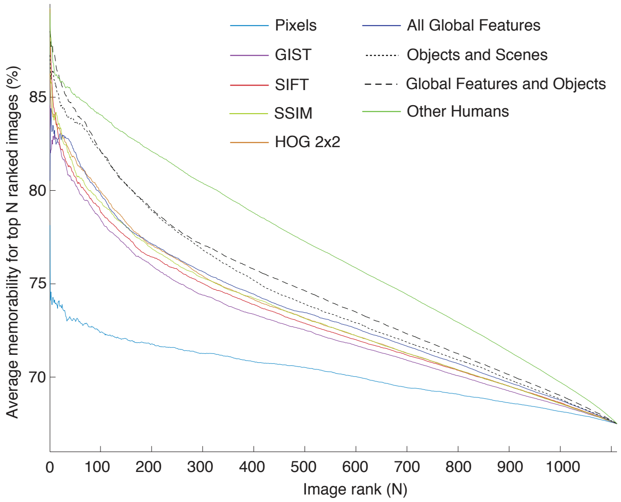
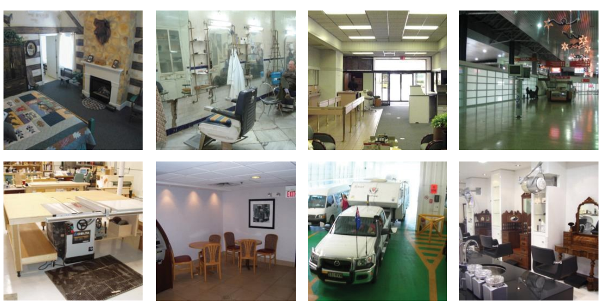



Figure 9. Comparison of regressions results averaged across 25 random split half trials. Images are ranked by predicted memorability and plotted against the cumulative average of empirically measured memorability scores. Error bars omitted for clarity.



a) Prediction too high (+919/1111)

b) Prediction too low (-948/1111)

Figure 10. The 8 images whose predicted memorability rank, on the basis of global features, most overshot empirical memorability rank (a) and most undershot empirical memorability rank (b). The mean rank error between predicted and measured ranks across each set of images is given in parentheses.

portance [22], saliency [10], and photo quality [16].

**Acknowledgments:** We would like to thank Timothy Brady and Talia Konkle for helpful discussions and advice. This work is supported by the National Science Foundation under Grant No. 1016862 to A.O, CAREER Awards No. 0546262 to A.O and No. 0747120 to A.T. A.T. was supported in part by the Intelligence Advanced Research Projects Activity via Department of the Interior contract D10PC20023, and ONR MURI N000141010933.

# References

[1] T. F. Brady, T. Konkle, G. A. Alvarez, and A. Oliva. Visual long-term memory has a massive storage capacity for object details. *Proc Natl Acad Sci, USA*, 2008. 145, 146, 147

[2] C.-C. Chang and C.-J. Lin. *LIBSVM: a library for support vector machines*, 2001. 148

[3] M. J. Choi, J. J. Lim, A. Torralba, and A. S. Willsky. Exploiting hierarchical context on a large database of object categories. *CVPR*, 2010. 148

[4] D. Cohen-Or, O. Sorkine, R. Gal, T. Leyvand, and Y. Xu. Color harmonization. *ToG*, 2006. 146

[5] N. Cowan. The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *Behavioral and Brain Sciences*, 2001. 146

[6] N. Dalal and B. Triggs. Histograms of oriented gradients for human detection. In *CVPR*, 2005. 150

[7] P. F. Felzenszwalb, R. B. Girshick, D. McAllester, and D. Ramanan. Object detection with discriminatively trained part-based models. *PAMI*, 2010. 150

[8] B. Gooch, E. Reinhard, C. Moulding, and P. Shirley. Artistic composition for image creation. In *Proc. of the 12th Eurographics workshop on Rendering Techniques*, 2001. 146

[9] R. Hunt and J. Worthen. *Distinctiveness and Memory*. New York, NY: Oxford University Press, 2006. 146

[10] L. itti, C. Koch, and E. Niebur. A model of saliency-based visual attention for rapid scene analysis. *PAMI*, 1998. 146, 152

[11] T. Konkle, T. F. Brady, G. A. Alvarez, and A. Oliva. Conceptual distinctiveness supports detailed visual long-term memory for real-world objects. *JEP:G*, 2010. 145, 148, 149

[12] T. Konkle, T. F. Brady, G. A. Alvarez, and A. Oliva. Scene memory is more detailed than you think: the role of categories in visual long-term memory. *Psych Science*, 2010. 145, 146, 147, 149

[13] W. Koutstaal and D. Schacter. Gist-based false recognition of pictures in older and younger adults. *Journal of Memory and Language*, 1997. 148, 149

[14] S. Lazebnik, C. Schmid, and J. Ponce. Beyond bags of features: Spatial pyramid matching for recognizing natural scene categories. In *CVPR*, 2006. 150

[15] T. Leyvand, D. Cohen-Or, G. Dror, and D. Lischinski. Data-driven enhancement of facial attractiveness. *ToG*, 2008. 146

[16] Y. Luo and X. Tang. Photo and video quality evaluation: Focusing on the subject. In *ECCV*, 2008. 146, 150, 152

[17] A. Oliva and A. Torralba. Modeling the shape of the scene: a holistic representation of the spatial envelope. *IJCV*, 2001. 150

[18] L. Renjie, C. L. Wolf, and D. Cohen-or. Optimizing photo composition. Technical report, Tel-Aviv University, 2010. 146, 150

[19] I. Rock and P. Englestein. A study of memory for visual form. *The American Journal of Psychology*, 1959. 145

[20] B. C. Russell, A. Torralba, K. P. Murphy, and W. T. Freeman. LabelMe: a database and web-based tool for image annotation. *IJCV*, 2008. 148

[21] E. Shechtman and M. Irani. Matching local self-similarities across images and videos. In *CVPR*, 2007. 150

[22] M. Spain and P. Perona. Some objects are more equal than others: Measuring and predicting importance. In *ECCV*, 2008. 146, 152

[23] L. Standing. Learning 10,000 pictures. *Quarterly Journal of Experimental Psychology*, 1973. 145

[24] J. Xiao, J. Hayes, K. Ehinger, A. Oliva, and A. Torralba. Sun database: Large-scale scene recognition from abbey to zoo. In *CVPR*, 2010. 147, 149, 150