MIT Libraries | DSpace@MIT

MIT Open Access Articles

Modifying the Memorability of Face Photographs

The MIT Faculty has made this article openly available. ***Please share*** how this access benefits you. Your story matters.

Citation: Khosla, Aditya, Wilma A. Bainbridge, Antonio Torralba, and Aude Oliva. "Modifying the Memorability of Face Photographs." 2013 IEEE International Conference on Computer Vision (December 2013).

As Published: http://dx.doi.org/10.1109/ICCV.2013.397

Publisher: Institute of Electrical and Electronics Engineers (IEEE)

Persistent URL: http://hdl.handle.net/1721.1/90986

Version: Author's final manuscript: final author's manuscript post peer review, without publisher's formatting or copy editing

Terms of use: Creative Commons Attribution-Noncommercial-Share Alike

Massachusetts Institute of Technology

Modifying the Memorability of Face Photographs

Aditya Khosla Wilma A. Bainbridge Antonio Torralba Aude Oliva
Massachusetts Institute of Technology
Computer Science and Artificial Intelligence Laboratory
{khosla, wilma, torralba, oliva}@csail.mit.edu

Abstract

Contemporary life bombards us with many new images of faces every day, which poses non-trivial constraints on human memory. The vast majority of face photographs are intended to be remembered, either because of personal relevance, commercial interests or because the pictures were deliberately designed to be memorable. Can we make a portrait more memorable or more forgettable automatically? Here, we provide a method to modify the memorability of individual face photographs, while keeping the identity and other facial traits (e.g. age, attractiveness, and emotional magnitude) of the individual fixed. We show that face photographs manipulated to be more memorable (or more forgettable) are indeed more often remembered (or forgotten) in a crowd-sourcing experiment with an accuracy of 74%. *Quantifying and modifying the 'memorability' of a face lends itself to many useful applications in computer vision and graphics, such as mnemonic aids for learning, photo editing applications for social networks and tools for designing memorable advertisements. Additional information, data and supplemental material is available at:* http://facemem.csail.mit.edu

1. Introduction

April 2016, London. "Will this make me more memorable?" she wonders, playing with the options on her new smart phone app. Her face on the screen has gotten an instant lift, the expression is somehow more interesting. "This is for the job application, I want them to remember me."

One ubiquitous fact about people is that we cannot avoid evaluating the faces we see in daily life. In fact, we automatically tag faces with personality, social, and emotional traits within a single glance: according to [28], an emotionally neutral face is judged in the instance it is seen, on traits such as level of attractiveness, likeability, and aggressiveness. However, in this flash judgment of a face, an underlying decision is happening in the brain – should I remember this face or not? Even after seeing a picture for only half a second we can often remember it [22]. Face memorability is in fact a critical factor that dictates many of our social interactions; even if a face seems friendly or attractive, if it is easily forgotten, then it is meaningless to us. Work in computer graphics has shown how to select a candid portrait from videos [9], or how faces can be slightly adjusted to look more attractive [18]. With the rapid expansion of social networks and virtual worlds, we are becoming increasingly concerned with selecting and creating our best selves – our most remembered selves.

Figure 1: Examples of modifying the memorability of faces while keeping identity and other attributes fixed. Despite subtle changes, there is a significant impact on the memorability of the modified images.

Back to 2013: Can we make a portrait more memorable? In this work, we show that it is indeed possible to change the memorability of a face photograph. Figure 1 shows

some photographs manipulated using our method (one individual per row) such that each face is more or less memorable than the original photograph, while maintaining the identity, gender, emotions and other traits of the person. The changes are subtle, difficult to point out precisely, or even to describe in words. However, despite these subtle changes, when testing people's visual memory of faces, the modification is successful: after glancing at hundreds of faces, observers remember better seeing the faces warped towards memorability, than the ones warped away from it.

It is not immediately intuitive what qualities cause a face to be remembered. Several memory researchers [24, 26] have found that measures of distinctiveness are correlated with memorability, while familiarity is correlated with increased false memories (believing you saw something you have not seen). Another line of work has discussed the use of geometric face space models of face distinctiveness to test memorability [4]. Importantly, recent research has found that memorability is a trait intrinsic to images, regardless of the components that make up memorability [2, 12, 13]. Thus, a fresh approach looking at memorability itself rather than its individual components (such as gender or distinctiveness) allows us here to create a method to change the memorability of a face while keeping identity and other important facial characteristics (like age, attractiveness, and emotional magnitude) intact.

To overcome the complex combination of factors that determine the memorability of a face, we propose a data-driven approach to modify face memorability. In our method, we combine the representational power of features based on Active Appearance Models (AAMs) with the predictive power of global features such as Histograms of Oriented Gradients (HOG) [6], to achieve desired effects on face memorability. Our experiments show that our method can accurately modify the memorability of faces with an accuracy of 74%.

2. Related Work

Image memorability: Recent work in computer vision [11, 12, 14] has shown that picture and face memorability are largely intrinsic to the image and importantly reproducible across people from diverse backgrounds, meaning that most of us will tend to remember and forget the same images. Furthermore, image memorability and the objects and regions that make a picture more or less memorable can be estimated using state of the art computer vision approaches [12, 13, 14] . While predicting image memorability lends itself to a wide variety of applications, no work so far has attempted to automatically predict and modify the memorability of individual face photographs.

Face modification: The major contribution of this work is modifying faces to make them more or less memorable. There has been significant work in modifying faces along other axes or attributes, such as gender [15], age [17, 23], facial expressions [1] and attractiveness [18]. However, while these works focus on intuitive and describable attributes, our work focuses on a property that is yet not well understood. Further, our method differs from existing works as it combines powerful global image features such as SIFT [19] and HOG [6] to make modifications instead of using only shape and appearance information as done in AAMs [5].

Face caricatures: Work in computer vision and psychology has also looked at face caricatures [3, 21], where the distinctive (i.e., deviant from the physical average) features of a face are accentuated to create an exaggerated face. The distinctiveness of a face is known to affect its later recognition in humans [4], so increasing the memorability of a face may caricaturize it to some degree. However, unlike face caricature work, the current study aims to maintain the realism of the faces, by preserving face identity. Recent memorability work finds that distinctiveness is not the sole predictor of face memorability [2], so the algorithm presented in this paper is likely to change the faces in more subtle ways than simply enlarging distinctive physical traits.

3. Predicting Face Memorability

As we make changes to a face, we need a model to reliably predict its memorability; if we cannot predict memorability, then we cannot hope to modify it in a predictable way. Thus, in this section, we explore various features for predicting face memorability and propose a robust memorability metric to significantly improve face memorability prediction. We also note that the task of automatically predicting the memorability of faces using computer vision features has not been explored in prior works.

In Sec. 3.1 we describe the dataset used in our experiments and the method used to measure memorability scores. Then, in Sec. 3.2, we describe our robust memorability metric that accounts for false alarms leading to significantly improved prediction performance (Sec 3.3).

3.1. Measuring Memorability

Memorability scores of images are obtained using a visual memory game, as described in [2, 12]: Amazon Mechanical Turk (AMT) workers are presented with a sequence of faces, each shown for 1.4 seconds with a 1 second blank interval, and their task is to press a key when they encounter a face they believe they have seen before. The task is structured in levels of 120 images each and includes a combination of target images (i.e. images that repeat) and filler images (i.e. images shown only once). Target image repeats are spaced $91-109$ images apart. A given target image and its repeat are shown to N unique individuals. Of the N individuals, we define H to be the number of people that correctly detected the repeat, and F as the number of people that false alarmed on the first showing. Based on this,

Metric	Face database [2]		Scene database [12]	
	Human	Pred	Human	Pred
$\frac{H}{N}$ [12]	0.68	0.33	0.75	0.43
$\frac{(H-F)}{N}$ (Ours)	0.69	**0.51**	0.73	**0.45**

Table 1: **Memorability score with false alarms:** Spearman's rank correlation (ρ) using the existing [12] and proposed (Sec. 3.2) memorability score metrics. 'Human' refers to human consistency evaluated using 25 train/test splits of data (similar to [12]), while 'prediction' refers to using support vector regression (SVR) trained on dense HOG [6] (details in Sec. 3.3).

Bainbridge et al [2] investigated two *memorability scores* for an image, (1) the proportion of correct responses , $\frac{H}{N}$, also called the hit rate and (2) the proportion of errors, $\frac{F}{N}$, also called false alarm rate. However, they do not combine the two scores into a metric that allows for a trade-off between correct responses and false alarms. We address this in the following section.

3.2. Memorability Score with False Alarms

As noted in [2], faces tend to have a significantly higher false alarm rate than scenes [12], i.e. there are certain faces that a large proportion of people believe they have seen before which they actually have not. Rather than being memorable (with high correct detections), these faces are in fact "familiar" [26] - people are more likely to report having seen them, leading to both correct detections and false alarms. To account for this effect, we propose a slight modification to the method of computing the memorability score.

To account for the *wrong* correct detections, we simply subtract the false alarms from the hit count to get an estimate of the *true* hit count of an image. Thus, the new memorability score can be computed as $\frac{H-F}{N}$, unlike $\frac{H}{N}$ as done in [12] and [2]. Note that $H, F \in [0, N]$, so $\frac{H-F}{N} \in [-1, 1]$ while $\frac{H}{N} \in [0, 1]$. The negative memorability scores can be easily adjusted to lie in the range $[0, 1]$.

The result of applying the above metric is summarized in Tbl. 1. To show that our metric is robust, we apply it to both the face [2] and scene memorability [12] datasets. We observe that human consistency remains largely the same in both cases. This is expected as the false alarms tend to be consistent across participants in the study. Importantly, we observe that there is a significant increase in the prediction performance from rank correlation of 0.33 to 0.51 for face memorability. By using our new metric, we have effectively decreased noise in the prediction labels (memorability scores) caused by inflated memorability scores of *familiar* images. This allows the learning algorithm to better model the statistics of the data that best describe memorability. We note that the performance improvement is not as large in scenes because the human consistency of false alarms and the rate of false alarms is significantly lower, and effects of familiarity may function differently.

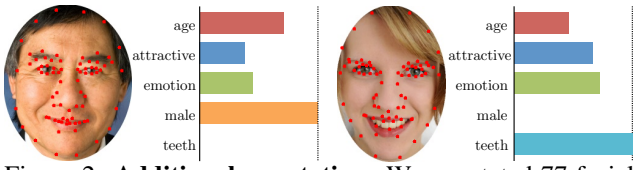


Figure 2: **Additional annotation:** We annotated 77 facial landmarks of key geometric points on the face and collected 19 demographic and facial attributes for each image in the 10k US Adult Faces Database[1].

We use our proposed memorability score that takes false alarms into consideration for the remaining experiments in this paper. Please refer to the supplemental material for additional analysis on the proposed metric.

3.3. Prediction experiments

In this section, we describe the additional annotations we collected (Sec. 3.3.1) to enable better face modification. Then we describe the setup of our experiments such as the evaluation metric and features used (Sec. 3.3.2) for the prediction of memorability and other attributes (Sec. 3.3.3).

3.3.1 Additional annotation

In order to modify faces while preserving facial attributes such as gender and emotion, we need two sets of additional annotations: (1) facial landmark locations to allow for the easy modification of faces by moving these keypoints and generating new images through warping, and (2) facial attribute (e.g., attractiveness, emotion) ratings so we can keep them fixed. Fig. 2 shows some examples of the additional annotation we collected on the 10k US Adult Faces Database [2]. Note that since we aim to modify faces instead of detect keypoints, we assume that landmark annotation is available at both train and test times. Locating these landmarks is a well-studied problem and we could obtain them using various automatic methods [5, 29] depending on the application. In our work, the landmarks were annotated by experts to ensure high consistency across the dataset.

To collect the facial attributes, we conducted a separate AMT survey similar to [16], where each of the 2222 face photographs was annotated by twelve different workers on 19 demographic and facial attributes of relevance for face memorability and face modification. We collected a variety of attributes including demographics such as gender, race and age, physical attributes such as attractiveness, facial hair and make up, and social attributes such as emotional magnitude and friendliness. These attributes are required when modifying a face so we can attempt to keep them constant or modify them jointly with memorability, as required by the user.

[1] Additional information, data and supplemental material available at: http://facemem.csail.mit.edu

3.3.2 Setup

Dataset: In our experiments, we use the 10k US Adult Faces Database [2] that consists of 2222 face photographs annotated with memorability scores.

Evaluation: The prediction performance is evaluated either using Spearman's rank correlation (ρ) for real-valued attributes, and accuracy for discrete-valued attributes. We evaluate the performance on 25 random train/test splits of the data (as done in [12]) with an equal number of images for training and testing. For memorability, the train splits are scored by one half of the participants and the test splits are scored by the other half with a human consistency of $\rho = 0.69$. This can be thought of as an upper bound on the performance we can hope to achieve.

Features: We use similar features as [14] for our experiments, namely the color naming feature [25], local binary pattern (LBP) [20], dense HOG2x2 [6] and dense SIFT [19]. For the bag-of-words features (i.e., color, HOG and SIFT), we sample descriptors at multiple scales and learn a dictionary of codewords using K-means clustering. Then we use Locality-constrained Linear Coding (LLC) [27] to assign descriptors to codewords, and apply max-pooling with 2 spatial pyramid levels to obtain the final feature vector. Similarly, we use a 2-level spatial pyramid for the non-uniform LBP descriptor. In addition, we also use the coordinates of the ground-truth landmark annotations (shown in Fig. 2) normalized by image size as 'shape' features.

Model: For discrete-valued attributes, we apply a one-vs-all linear support vector machine (SVM) [8], while for real-valued attributes, we apply support vector regression (SVR) [7]. The hyperparameters, C (for SVM/SVR) and ϵ (for SVR), are found using cross-validation.

3.3.3 Memorability and attribute prediction

Tbl. 2 summarizes the prediction performance of face memorability and other attributes when using various features. For predicting memorability, dense global features such as HOG and SIFT significantly outperform landmark-based features such as 'shape' by about 0.15 rank correlation. This implies that it is essential to use these features in our face modification algorithm to robustly predict memorability after making modifications to a face. While powerful for prediction, the dense global features tend to be computationally expensive to extract, as compared to shape. As described in Sec. 4.2, shape is used in our algorithm to parametrize faces so it essentially has zero cost of extraction for modified faces.

Similar to memorability prediction, we find that dense global features tend to outperform shape features for most attributes. However, as compared to memorability, the gap in performance between using shape features and dense features is not as large for other attributes. This might suggest why, unlike our method, existing methods [17, 18] typically use landmark-based features instead of dense global features for the modification of facial attributes.

	Color [25]	LBP [20]	HOG [6]	SIFT [19]	Shape
age (0)	0.52	0.64	0.72	**0.77**	0.68
attractive (0)	0.46	0.51	0.59	**0.62**	0.54
emoteMag (0)	0.49	0.64	0.80	0.83	**0.88**
makeup (3)	0.80	0.81	**0.84**	**0.84**	0.80
is male (2)	0.86	0.89	0.93	**0.94**	0.91
teeth (3)	0.56	0.58	0.71	0.72	**0.77**
memorability	0.27	0.23	**0.51**	0.49	0.36

Table 2: **Prediction performance of memorability and other attributes:** The number in the parantheses indicates the number of classes for the particular attribute, where 0 means that the attribute is real-valued. For real-valued attributes and memorability, we report Spearman's rank correlation (ρ), while for discrete valued attributes such as 'male', we report classification accuracy. The reported performance is averaged on 25 random train/test splits of the data. Additional results provided in supplemental material.

4. Modifying Face Memorability

In order to modify a face photograph, we must first define an expressive yet low-dimensional representation of a face. We need to parametrize a face such that we can synthesize new, realistic-looking faces. Since faces have a largely rigid structure, simple methods such as Principal Component Analysis (PCA) e.g. AAMs [5] tend to work fairly well. In Sec. 4.1, we describe a method based on AAMs where we represent faces using two distinct features, namely shape and appearance.

While the above parametrization is extremely powerful and allows us to modify a given face along various dimensions, we require a method to evaluate the modifications in order to make predictable changes to a face. Our objective is to modify the memorability score of a face, while preserving the identity and other attributes such as age, gender, emotions, etc of the individual. We encode these requirements in a cost function as described in Sec. 4.2. Specifically, our cost function consists of three terms: (1) the cost of modifying the identity of the person, (2) the cost of not achieving the desired memorability score, and (3) the cost of modifying other attributes. By minimizing this cost function, we can achieve the desired effect on the memorability of a face photograph.

As highlighted in Sec. 3.3.3, it is crucial to use dense global features when predicting face memorability. These features tend to be highly non-convex in our AAM parameter space, making it difficult to optimize the cost function

exactly. Thus, we propose a sampling-based optimization procedure in Sec. 4.3 that leverages the representational power of AAMs with the predictive power of dense global features in a computationally efficient manner.

4.1. Representation

Using facial landmark-based annotations (described in Sec. 3.3) in the form of AAMs is a common method for representing faces for modification because it provides an expressive, and low-dimensional feature space that is reversible, i.e., we can recover an image easily after moving in this feature space. AAMs typically have two components to represent a face, namely shape (x_s) and appearance (x_a).

To obtain x_s, we first compute the principal components of the normalized landmark locations[2] across the datasets. Then, for a particular image, x_s is given as the coefficients of these principal components. Similarly, to obtain x_a, we first warp all the faces to the mean shape and apply PCA to the concatenated RGB values of the resulting image[3] (resized to a common size). Then, x_a for a given image is given by the coefficients of these principal components. Thus, the full parametrization of a face, x, can be written as $x = [x_s\ x_a]$, i.e., the concatenation of shape and appearance features. We can now modify x and use the learned principal components to synthesize a new face.

When applying the above parametrization to faces, we observed that there were significant distortions when warping a face, and a high reconstruction error even without any modification (despite keeping 99.5% of the principal components). To improve the reconstruction, we cluster the normalized facial landmarks and apply PCA independently to each cluster. Further, as shown in Fig. 2, images in the 10k US Adult Faces Database contain an ellipsoid around them to minimize the background effects on memorability, so we added uniformly spaced landmarks along the boundary to constrain the warping in order to obtain more realistic results. We evaluate the effects of these modifications to AAM in Sec. 5.3.

4.2. Cost function

In order to modify a face to a desired memorability score while preserving identity and other facial attributes, we define a cost function with the following three terms:

- C_{id}: Cost of modifying the identity
- C_{mem}: Cost of *not* attaining the desired memorability
- C_{attr}: Cost of modifying facial attributes such as age, attractiveness, emotional magnitude, etc

[2]The 77 normalized landmark locations are concatenated to form a 154 dimensional shape vector.

[3]As there could be components of appearance outside the face region such as hair that we would like to be able to modify, we use the entire image instead of just the face pixels (as is typically done).

We parametrize the above terms using the shape and appearance based representation, x, described in Sec. 4.1. Then, our optimization objective is:

$$\min_x \quad C_{id}(x) + \lambda C_{mem}(x) + \gamma C_{attr}(x) \tag{1}$$

where λ and γ are hyperparameters to control the relative importance of the three terms in the cost function.

Before defining the above terms explicitly, we define some common terminology. Let F define the set of image features (e.g. HOG [6] or SIFT [19]) and A, the set of facial attributes (e.g., is male or are teeth visible). Then we define $m_i(x)$ as a function to predict the memorability score of an image represented by PCA coefficients x computed using feature $i \in F$. Similarly, we define $f_{i,j}(x)$ as a function to predict the value of attribute $j \in A$ of an image defined by PCA coefficients x, computed using feature $i \in F$. Note that the landmark-based features can be directly obtained from x, while there is an implicit conversion from the PCA coefficients x to an image before extracting dense global features. For brevity, we do not write this transformation explictly. In our experiments, m_i and $f_{i,j}$ are obtained using linear SVR/SVM as described in Sec. 3.3.2.

Now, given an image $\hat{I}$ that we want to modify, our goal is to synthesize a new image I that has a memorability score of M (specified by the user) and preserves the identity and other facial attributes of the original image $\hat{I}$. Representing the PCA coefficients of $\hat{I}$ by $\hat{x}$, our objective is to find the PCA coefficients x that represent I. Since landmark estimation is outside the scope of this work, we assume that the landmark annotations required to obtain $\hat{x}$ are available at both train and test time.

Based on this problem setup, we define the terms from Eqn. 1 as the following three equations (Eqn. 2, 3, 4):

$$C_{id}(x) = [w \cdot (x - \hat{x})]^2 \tag{2}$$

where w is the weight vector for preserving identity, learned using a Support Vector Machine (SVM) trained on the original image $\hat{I}$ as positive, and the remaining images in the dataset as negatives.

$$C_{mem}(x) = \sum_{i \in F} \frac{c_i}{|F|}(m_i(x) - M)^2 \tag{3}$$

where c_i represents the confidence of predictor m_i, and can be estimated using cross-validation. Since the performance of different features on memorability prediction varies significantly (Sec. 3.3), we weight the difference between M and m_i by the confidence score c_i to increase the importance of better performing features. Overall, this function penalizes the memorability score of the new image x if it does not match the desired memorability score, M.

$$C_{attr}(x) = \sum_{i \in F} \sum_{j \in A} \frac{c_{i,j}}{|F| \cdot |A|} (f_{i,j}(x) - f_{i,j}(\hat{x}))^2 \qquad (4)$$

where $c_{i,j}$ is the confidence of predictor $f_{i,j}$, and can be estimated using cross-validation. The above function computes the distance between the estimated attribute value of the original image $f_{i,j}(\hat{x})$ and the new image $f_{i,j}(x)$. Additionally, a user could easily modify the relative importance of different attributes in the above cost function.

Overall, C_{id} and C_{attr} encourage the face to remain the same as $\hat{I}$ while C_{mem} encourages the face to be modified to have the desired memorability score of M. By adjusting the hyperparameters appropriately, we can achieve the desired result on memorability.

4.3. Optimization

While the terms in the objective function defined in Eqn. 1 look deceptively simple, we note that the function is actually highly complex and non-linear because of the dense global features such as HOG or SIFT involving histograms, bag-of-words and max-pooling. Thus, typical gradient-based approaches hold little promise in this case, and finding a local minimum is the best we can hope for.

The basic idea of our algorithm is fairly straightforward: we want to find x given $\hat{x}$; since x is expected to look like $\hat{x}$, we initialize $x = \hat{x}$. Then, we randomly sample[4] some points at some fixed distance d from x. From the set of random samples, we find the sample that best minimizes the objective function and use that as our new x. We now repeat the procedure by finding samples at distance $\frac{d}{2}$ from x. We include the initial value of x as one of the samples in each iteration to ensure that the objective function always decreases. By iteratively reducing d, we can find a local minimum of our objective close to $\hat{x}$.

However, we observe that the computational cost of feature extraction differs significantly between dense global features and landmark-based features; dense global features require the synthesis of a new image based on x, and the subsequent extraction of features while landmark-based features such as shape and appearance can be trivially obtained from x. Note that the dense global features play a significant role in accurate memorability prediction (Sec. 3.3), so we must include them in our algorithm. Since it is computationally expensive to compute dense global features for all samples, this severly restricts the total number of samples considered per iteration, making it difficult to find a good solution, or even one better than $\hat{x}$.

[4]In practice, we fit a multivariate Gaussian to the PCA coefficients of the training images (similar to [18]), and obtain random samples from this distribution. Given a random sample p, we find a point at distance d from x that lies in the direction $p - x$.

To overcome the computational bottleneck, we propose a two-step procedure: (1) for a large number of samples, evaluate the cost function ignoring the terms involving dense global features, and (2) obtain a small subset of the best scoring samples from step (1) and rescore them using the full cost function. In this way, we can obtain a better solution by pruning the space using a computationally efficient estimation of the cost function, and later rescoring a small set of samples using the full cost function to find the best sample for our final result. Please refer to the supplementary material for additional details.

5. Experiments

In this section, we describe the experimental evaluation of our memorability modification algorithm. Specifically, we describe the experimental setup in Sec. 5.1, the results obtained in Sec. 5.2 and additional analysis of our algorithm in Sec. 5.3. Overall, we find that our algorithm modifies 74% of the images accurately, and our result is statistically significant with a p-value of $< 10^{-4}$.

5.1. Setup

Our goal is to evaluate whether our algorithm is able to modify the memorability of faces in a predictable way. In order to do this, we use the face memory game [2]: we designed two balanced experiments where, for the first experiment, we increase the memorability of half the target images and decrease the memorability of the other half, and vice versa for the other (modified versions of the same target images are used in both experiments). Then we compare the memorability scores of the modified images; if the mean memorability of the set of images whose memorability was increased is higher than the decreased set, we can conclude that our algorithm is accurately modifying memorability.

Specifically, we randomly sample 505 of the 2222 target images from the 10k US Adult Faces Database and use them as targets, and the remaining as fillers in our experiments. We ensure that the set of participants in the two studies is disjoint, i.e., a participant does not see both high/low modifications of a single target. On average, we obtained 63 scores per image in each of the experiments.

Algorithmic details: We set the hyperparameters $\lambda = 10$, and $\gamma = 1$ and use the PCA features together with dense HOG2x2 [6] and SIFT [19] features as described in Sec. 3.3.2. The target images were modified to have a memorability score that differs by 0.2 from the original in the appropriate direction. To account for warping effects, the filler images are also modified by random scores in the range $[-0.1, 0.1]$ and are identical for both sets of experiments.

5.2. Memorability Modification Results

Fig. 3 summarizes the quantitative results from the memorability games described in Sec. 5.1. In Fig. 3(a), we show

the overall memorability scores of all target images after the two types of modifications (i.e., memorability increase and decrease) sorted independently. We observe that the mean memorability score 'memorability increase' images is significantly higher than that of the 'memorability decrease' images. We perform a one-tailed paired sample t-test and find that the null hypothesis that the means are equal is rejected with a p value of $< 10^{-4}$ for both sets of experiments, indicating that our results are statistically significant.

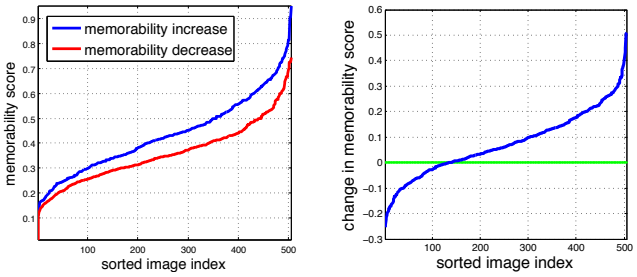
Fig. 3(b) shows the difference in memorability scores of individual images; for a given image we subtract the observed memorability of the version modified to have lower memorability image from that of the version modified to have higher memorability. We find that the expected change in memorability (> 0) occurs in about 74% of the images (chance is 50%). This is a fairly high value given our limited understanding of face memorability and the factors affecting it. We also observe that the increase in memorability scores is much larger in magnitude than the decrease.

Fig. 5 shows qualitative results of modifying images to have higher and lower memorability, together with the memorability scores obtained from our experiments. While we observe that the more memorable faces tend to be more 'interesting', there is no single modification axis such as distinctiveness, age, etc, that leads to more or less memorable faces. Essentially, our data-driven approach is effectively able to identify the subtle elements of a face that affect its memorability and apply those effects to novel faces.

5.3. Analysis

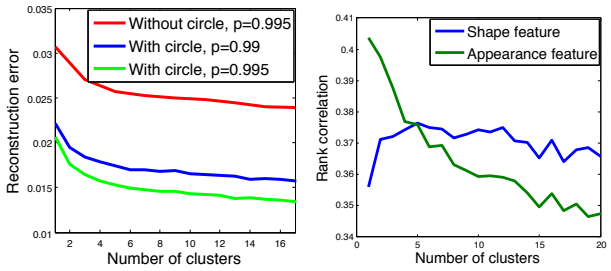
To investigate the contribution of shape and appearance features to face memorability, we conduct a second AMT study similar to the one described in Sec. 5.1, except that we only modify shape features in this case. We found that the accuracy of obtaining the expected change in memorability, as described in Sec. 5.2, dropped to 60%. In addition, the changes in memorability scores were not as significant in this case as compared to the original setting. This shows that a combination of shape and appearance features are important for modifying memorability; however, it is interesting to note that despite the limited degree of freedoms, our algorithm achieved a reasonable modification accuracy.

Fig. 4 shows the effect of having clusters in the AAM as described in Sec. 4.1. We find that having more clusters allows us to have better reconstructions without significant sacrifice in memorability prediction performance. Thus, we choose to use 8 clusters in our experiments. Lastly, since changes in memorability lead to unintuitive modifications to faces, in Fig. 6, we apply our algorithm to modify other attributes whose effects are better understood. Indeed, we observe that the algorithm is behaving as per our expectations. Please refer to the supplemental material for more modification results and additional analysis.



(a) Overall memorability scores (b) Individual images

Figure 3: **Quantitative results:** (a) Memorability scores of all images in the increase/decrease experimental settings, and (b) change in memorability scores of individual images.



(a) Reconstruction error (b) Memorability prediction

Figure 4: **Analysis:** Figure showing (a) reconstruction error, and (b) memorability prediction performance as we change the number of clusters in AAM. With and without circle refers to having control points on the image boundary when doing warping.

6. Conclusion

Memorability quantification and modification lends itself to several innovative applications in the fields of computer vision and graphics, marketing and entertainment, and social networking. For instance, for animated films, movies, or video games, one could imagine animators creating cartoon characters with different levels of memorability [10] or make-up artists making any actor a highly memorable protagonist surrounded by forgettable extras.

Importantly, the current results show that memorability is a trait that can be manipulated like a facial emotion, changing the whole face in subtle ways to make it look more distinctive and interesting. These memorability transformations are subtle, like an imperceptible "memory face lift." These modified faces are either better remembered or forgotten after a glance, depending on our manipulation. Here we show we can create, in a static photograph, a first impression that lasts.

Acknowledgments We thank Zoya Bylinskii, Phillip Isola and the reviewers for helpful discussions. This work is funded by a Xerox research award to A.O, Google research awards to A.O and A.T, and ONR MURI N000141010933 to A.T. W.A.B is supported by an NDSEG Fellowship. A.K. is supported by a Facebook Fellowship.

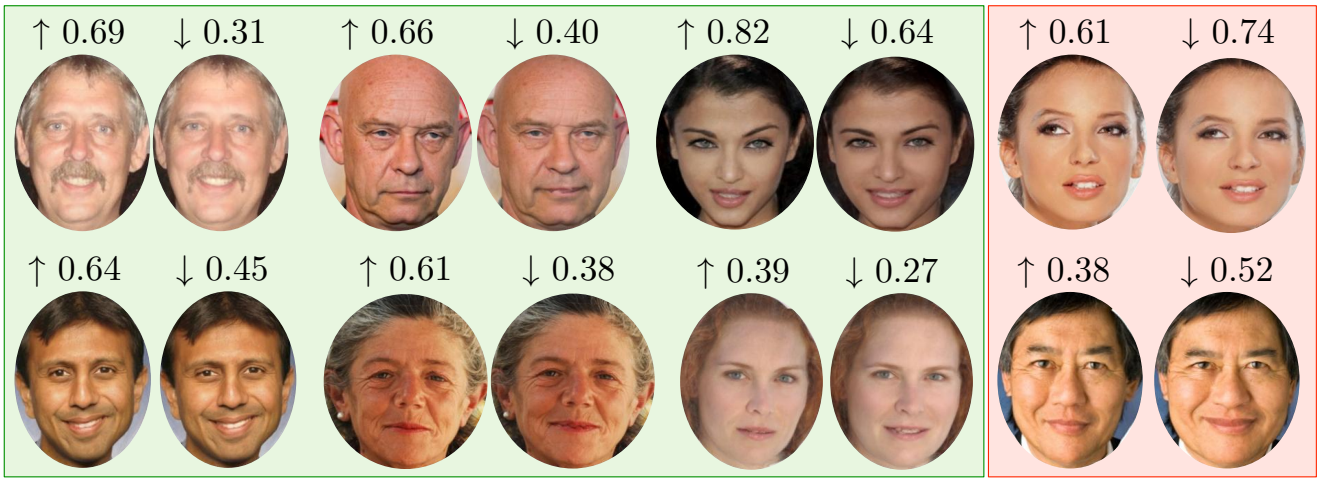


Figure 5: **Visualizing modification results:** Figure showing success (green background) and failure (red background) cases of the modification together with memorability scores from human experiments. Arrow direction indicates which face is expected to have higher or lower memorability of the two while numbers indicate the actual memorability scores.

Figure 6: **Modifying other attributes:** We increase/decrease other attributes such as age, attractiveness and friendliness.

References

[1] B. Amberg, P. Paysan, and T. Vetter. Weight, sex, and facial expressions: On the manipulation of attributes in generative 3D face models. In *Advances in Visual Computing*. 2009. 2

[2] W. A. Bainbridge, P. Isola, and A. Oliva. The intrinsic memorability of face photographs. In *Journal of Experimental Psychology: General*, 2013. 2, 3, 4, 6

[3] P. J. Benson and D. I. Perrett. Perception and recognition of photographic quality facial caricatures: Implications for the recognition of natural images. *EJCP*, 1991. 2

[4] T. A. Busey. Formal models of familiarity and memorability in face recognition. *Computational, geometric, and process perspectives on facial cognition: Contexts and challenges*, pages 147–191, 2001. 2

[5] T. F. Cootes, G. J. Edwards, and C. J. Taylor. Active appearance models. *PAMI*, 23(6):681–685, 2001. 2, 3, 4

[6] N. Dalal and B. Triggs. Histograms of oriented gradients for human detection. In *CVPR*. IEEE, 2005. 2, 3, 4, 5, 6

[7] H. Drucker, C. J. Burges, L. Kaufman, A. Smola, and V. Vapnik. Support vector regression machines. *NIPS*, pages 155–161, 1997. 4

[8] R. Fan, K. Chang, C. Hsieh, X. Wang, and C. Lin. LIBLINEAR: A library for large linear classification. *JMLR*, 9:1871–1874, 2008. 4

[9] J. Fiss, A. Agarwala, and B. Curless. Candid portrait selection from video. In *ACM TOG*, volume 30, page 128. ACM, 2011. 1

[10] B. Gooch, E. Reinhard, and A. Gooch. Human facial illustrations: creation and psychophysical evaluation. *ACM TOG*, 2004. 7

[11] P. Isola, D. Parikh, A. Torralba, and A. Oliva. Understanding the intrinsic memorability of images. In *NIPS*, 2011. 2

[12] P. Isola, J. Xiao, A. Torralba, and A. Oliva. What makes an image memorable? In *CVPR*, 2011. 2, 3, 4

[13] A. Khosla, J. Xiao, P. Isola, A. Torralba, and A. Oliva. Image memorability and visual inception. In *SIGGRAPH Asia*, 2012. 2

[14] A. Khosla, J. Xiao, A. Torralba, and A. Oliva. Memorability of image regions. In *NIPS*, 2012. 2, 4

[15] B. Knappmeyer, I. M. Thornton, and H. H. Bülthoff. The use of facial motion and facial form during the processing of identity. *Vision research*, 43(18):1921–1936, 2003. 2

[16] N. Kumar, A. C. Berg, P. N. Belhumeur, and S. K. Nayar. Attribute and simile classifiers for face verification. In *ICCV*, 2009. 3

[17] A. Lanitis, C. J. Taylor, and T. F. Cootes. Toward automatic simulation of aging effects on face images. *PAMI*, 2002. 2, 4

[18] T. Leyvand, D. Cohen-Or, G. Dror, and D. Lischinski. Data-driven enhancement of facial attractiveness. *SIGGRAPH*, 2008. 1, 2, 4, 6

[19] D. Lowe. Distinctive image features from scale-invariant keypoints. *IJCV*, 60(2):91–110, 2004. 2, 4, 5, 6

[20] T. Ojala, M. Pietikainen, and T. Maenpaa. Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. *PAMI*, 24(7):971–987, 2002. 4

[21] A. J. O'Toole, T. Vetter, H. Volz, E. M. Salter, et al. Three-dimensional caricatures of human heads: distinctiveness and the perception of facial age. *Perception*, 26:719–732, 1997. 2

[22] M. C. Potter. Meaning in visual search. *Science*, 1975. 1

[23] J. Suo, S.-C. Zhu, S. Shan, and X. Chen. A compositional and dynamic model for face aging. *PAMI*, 2010. 2

[24] T. Valentine. A unified account of the effects of distinctiveness, inversion, and race in face recognition. *QJEP*, 1991. 2

[25] J. Van De Weijer, C. Schmid, and J. Verbeek. Learning color names from real-world images. In *CVPR*, 2007. 4

[26] J. R. Vokey and J. D. Read. Familiarity, memorability, and the effect of typicality on the recognition of faces. *Memory & Cognition*, 20(3):291–302, 1992. 2, 3

[27] J. Wang, J. Yang, K. Yu, F. Lv, T. Huang, and Y. Gong. Locality-constrained linear coding for image classification. In *CVPR*. 4

[28] J. Willis and A. Todorov. First impressions making up your mind after a 100-ms exposure to a face. *Psychological science*, 2006. 1

[29] X. Zhu and D. Ramanan. Face detection, pose estimation, and landmark localization in the wild. In *CVPR*, 2012. 3