Change Blindness Images

Li-Qian Ma[1], Kun Xu[1], Tien-Tsin Wong[2], Bi-Ye Jiang[1], Shi-Min Hu[1]
[1] TNList, Department of Computer Science & Technology, Tsinghua University, Beijing
[2] Department of Computer Science & Engineering, The Chinese University of Hong Kong

Abstract—Change blindness refers to human inability to recognize large visual changes between images. In this paper, we present the first computational model of change blindness to quantify the degree of blindness between an image pair. It comprises a novel context-dependent saliency model and a measure of change, the former dependent on the site of the change, and the latter describing the amount of change. This saliency model in particular addresses the influence of background complexity, which plays an important role in the phenomenon of change blindness. Using the proposed computational model, we are able to synthesize changed images with desired degrees of blindness. User studies and comparisons to state-of-the-art saliency models demonstrate the effectiveness of our model.

Index Terms—Change blindness, image synthesis

1 Introduction

Change blindness is a psychological phenomenon that very large changes made to an image may often go unnoticed by observers. Fig. 1 shows two pairs of images, in which the two images in each pair contain a difference. Observers typically take some time to find the difference between the two images. A well-known type of game, "spot-the-difference", relies on this phenomenon.

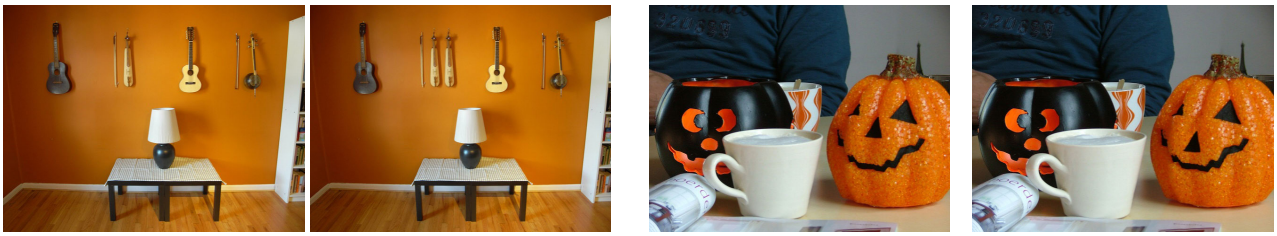
While the name "change blindness" suggests otherwise, such blindness is not due to our eyes, but to our brains. While psychological studies of change blindness continue, current research indicates that change blindness is caused by the failure to store visual information in our short-term memory [1], [2]. In order to compare two images, our brains must store at least part of one of them in order to compare it to the other. Attention is needed to see changes, however, change blindness cannot be fully explained by the mechanism of visual attention. The study of visual attention posits that while a visual stimulus is fully visible in a single image, it does not involve a comparison process as in the phenomenon of change blindness [1], [3].

Existing psychological literature on change blindness is mostly qualitative. Hence, most change blindness images are produced by hand, and their quality relies on the skill of the artist. In other words, it is hard to control the degree of change blindness. Such blindness degree depends on both how much the content has been changed, and where the change occurs (in a salient location or elsewhere). In this paper, given an input image, we propose a computational method to automatically synthesize a changed image with a desired degree of blindness.

To achieve this goal, we formulate the synthesis process as an optimization problem. To achieve a desired degree of blindness, it adjusts the location of the change and the change operators used (insertion, replacement, deletion, relocation, scaling, rotation and color-shift). Our core contribution is a novel metric to measure the degree of blindness. It takes into account both the amount of change and the saliency of the site where the change takes place. While the amount of change can be easily measured, existing saliency models are not applicable to our application due to their *lack of context dependency*. Psychologists have observed that the degree of blindness is highly dependent on the long-range neighborhood of the site where the change occurs [4]. For instance in Fig. 3, modifying the central red ball immersed in a sea of colorful balls (Fig. 3(a)) is less obvious than modifying the same ball against a simpler background (Fig. 3(e)). This indicates that the required saliency model must be context-dependent, and in particular, we find that the complexity of the surrounding background can highly influence the saliency. We thus propose a novel context-dependent saliency model to address the issue of context.

The effectiveness of our proposed metric and optimization method is supported by a user study. In summary, our major contributions include:

- A novel context-dependent saliency model that addresses the influence of long-range context complexity.
- A novel computational model for measuring the degree of change blindness for image pairs.
- An optimization-based method to synthesize change blindness images with controlled degree of blindness, which could be used e.g. to generate "spot-the-difference" games.



(a) Blindness: 0.01, average recognition time: 5 sec. (b) Blindness: 0.78, average recognition time: 44 sec.

Fig. 1. Using our change blindness model, we can synthesize changed images (right in each pair) from the original ones (left in each pair), with controllable degrees of blindness.

We believe our work is the first attempt to explicitly model the degree of change blindness, and the first one to model the context-dependent saliency arising in this context that accounts for background complexity.

2 Related Work

Change Blindness. The phenomenon of change blindness refers to a failure to notice large changes in a visual scene when there is a disruption [1], [3]. Such changes can take different forms, including shape changes [5], color changes, and object insertion, removal, or relocation [1]. The disruption can be eye movement [6], a flicker [1], 'mud splash' [7], or even real-world interactions [8]. Psychologists have made several qualitative studies [2], [9], [10] on change blindness, and it is believed that change blindness is caused by the failure to store complete visual information in our short-term memory for purposes of comparison, and hence our brains are unable to detect the changes. Psychologists have also found that change blindness is related to visual attention, and changes in locations with low saliency are less likely to be detected [1], [11]. It has also been found that change is more easily detected in a region where there is a significant saliency difference between the image pair [12]. Hence, both saliency and change in saliency are crucial factors in change blindness.

Existing literature on change blindness is mostly qualitative, and most image pairs for change blindness tasks are manually created, although Verma and McOwan [12] reported a semi-automatic method to generate change blindness images with least changed saliency for purposes of psychological study. However, the degree of blindness is not controllable. In contrast, we propose an explicit computational model for change blindness, allowing us to measure and control the degree of blindness. While most psychological studies focus on only one change at a time, Rensink [13] investigated change blindness behaviors when changes are simultaneously applied to multiple objects. For simplicity, our model considers one change at a time, although it can be empirically extended to multiple changes.

Saliency. Change blindness is highly dependent on visual attention, but unlike the latter, it involves a comparison process (requiring short-term memory) between two images. Visual attention involves two bottom-up, data-driven and top-down, goal-driven [14] processes. Many computational saliency metrics [15], [16], [17], [18], [19], [20], [21], [22], [23] have been proposed to model visual attention, and most of them are bottom-up. Next, we will briefly review some representative saliency models which are recently commonly used in image processing area [24].

Itti et al. [15] showed how to compute a saliency map by combining several multi-scale feature maps, including color, intensity, and orientation. Judd et al. [18] learned a saliency model using low-, mid-, high-level image features from eye tracking data on a thousand images. Bruce and Tsotsos [19] proposed a visual saliency model based on *attention by information maximization* (AIM), whose architecture is consistent with that observed in the visual cortex. Hou et al. [21] proposed an image feature descriptor, known as *image signature*, based on the theory of sparse signal analysis. They also developed a saliency model based on image signature. They found that the image signature is correlated to the identification time of certain types of change blindness images (deletion or insertion changes). Liu et al. [22] proposed a saliency model by learning a conditional random field from several image features, including multi-scale contrast, center-surround histogram, and color spatial distribution. Cheng et al. [23] introduced a region-wise contrast-based saliency detection algorithm. They defined the saliency of a region as the sum of global contrast differences with other regions weighted by spatial distances. In Section 6, we will compare our proposed model to the above saliency models in terms of ability to measure blindness and demonstrate the relative effectiveness of our model.

Context Awareness. The context of an object is usually defined as its relationship to the surrounding objects. Several works have demonstrated that contextual information can influence visual attention, object search and object recognition [4], [25]. In particular, Torralba and Oliva [26] proposed a contextual guidance model

for object search tasks, which combines bottom-up saliency and scene priors. The scene priors provide extra top-down knowledge, i.e., expected locations of the target object. Unlike their work, our saliency model does not utilize extra top-down knowledge, but incorporates a complexity value, computed for the surroundings, into our saliency definition. Goferman et al. [27] introduced a context-aware saliency model which detects a salient object together with its associated context (e.g. meaningful neighborhoods). Unlike their associated context detection, our saliency metric modulates saliency by its surrounding complexity (recall the felt ball example in Fig. 3).

Perceptually Motivated Graphics. Computer graphics researchers have already exploited the limitations of the human visual system for both rendering acceleration and recreational purposes. Image regions receiving less visual attention may be rendered more approximately with larger errors [28]. Other properties of the human visual system, such as inattentional blindness and change blindness, have also been utilized for rendering acceleration [29], [30], [31], character animation [32] and tone mapping [33]. Perceptual strategies are also useful in participating media rendering [34], virtual crowd rendering [35], and rendering of motion blur effects [36]. Ramanarayanan et al. [37] proposed the *visual equivalence predictor* for predicting visual differences between images rendered under warped illumination and images rendered under reference illumination. The *visible difference predictor* [38] is another well-known approach to determine visual distinguishability. However, both *visual equivalence predictor* and *visible difference predictor* measure small image differences that are hardly noticeable. They are not applicable to our application, as our changes (e.g. deletion, relocation, color-shift) are much more obvious and larger in size, yet remain unnoticed by a human observers for a certain length of time during the comparison.

Another stream of work focuses on recreational purposes. By exploiting the multiscale processing property of the human visual system, Oliva et al. [39] proposed a technique to synthesize *hybrid images*, which are composed of two image interpretations and appear differently when viewing distance is changed. Based on physiological and psychological knowledge of illusory motion, Chi et al. [40] presented a technique to generate self-animating images, which are static images containing certain simple color and geometric repeated pairs, however, appear to move or rotate. Mitra et al. [41] proposed a method to generate *emerging images* of 3D models, which appear noisy if looking at local parts but appear meaningful if viewing as a whole. Based on texture synthesis techniques, Chu et al. [42] introduced a method to generate *camouflage images*, where some objects are embedded in a busy, complex background, and take some time to detect. Tong et al. [43] presented a method to create *hidden images*. The form of hidden images is similar to that of camouflage images, where one or more objects are hidden in a background image, but hidden images rely on edges instead of texture details as clues for viewers. Similar to the above works, we also exploit properties of the human visual system and insights from psychological literature, however, we focus on a difference problem "change blindness". We refer readers to [44], [45], [46] for comprehensive surveys of perceptually motivated graphics.

3 Overview

Given an input image, we aim to generate a *changed* counterpart containing *one change* with a user-specified degree of blindness with respect to the given input. We formulate the problem as an optimization problem. The objective is to minimize the difference between the user-desired blindness and the measured blindness of the changed image. The input is iteratively changed by various change operators, until the objective value is optimized. Currently, the change operators supported include insertion, deletion, replacement, relocation, scale, rotation and color-shift. Our core contribution is the metric for measuring (the degree of) change blindness between an image pair, based on existing psychological findings (Section 4).

Fig. 2 overviews our system. Given an input image I, we first empirically extract certain candidate regions that are more likely to be foreground objects. To do so, instead of other image segmentation methods [47], [48], we segment the input using the novel mean shift approach [49], remove extremely large segments (larger than 30 percent of the whole image size) which are more likely to be background, discard tiny (smaller than 5x5 pixels) and long narrow segments (whose length-width-ratio is larger than 10), and group nearby segments to form larger ones based on their color properties (if the average color difference is smaller than 0.2). Next, an optional step allows users to manually refine (merge or split) the segments by drawing strokes on target segments. Fig. 8(c) shows the automatically extracted regions corresponding to the two examples in Fig. 8(a), while Fig. 8(d) gives manually refined segments based on the automatically extracted results (Fig. 8(c)). In our experiments, images with simple backgrounds usually do not need user intervention (Fig. 8(d)). For complex images that require user intervention, the correction can be done with a few strokes in seconds. The resultant disjoint segments are regarded as candidate regions, the primitives for our later change operations.

Optimization is carried out as follows. Initially, a region is randomly selected from the pool of candidate regions. Then a change operator is randomly selected and applied with a random parameter value, in order to synthesize a changed image I'. The blindness is then measured based on the image pair containing I

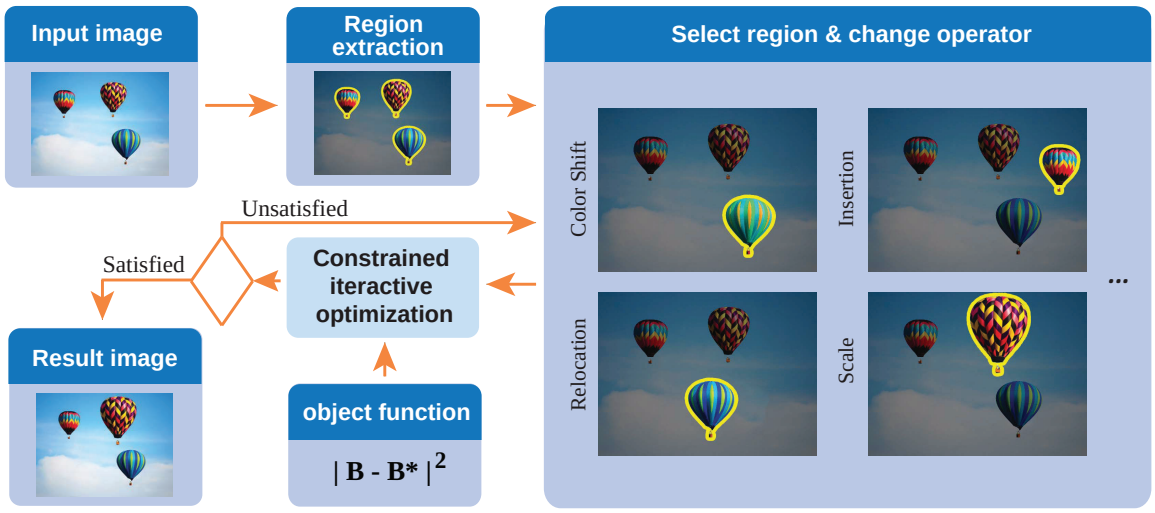


Fig. 2. The pipeline of our method to synthesize change blindness images.

and I'. The same operator is repeatedly applied to the same region with iteratively adjusted parameter until the measured blindness converges to the desired blindness within a small tolerance or until the number of iterations exceeds a predefined limit. In the latter case we randomly pick another combination of candidate region and change operator and try again. The whole optimization halts until a match is found or the total number of iterations reaches a predefined limit. Finally, the best changed image is obtained.

As tiny changes with only a few pixel differences are hard to observe with the naked eye, they are meaningless and should be avoided. To avoid such "degenerated" solutions (tiny changes), we ensure all output images to contain a *large* change by measuring the sum of squared pixel differences (SSD) between I and I'. The previously described optimization is actually performed in a constrained fashion, by constraining the SSD between I and I' to be larger than a predefined threshold. Details of the optimization is explained in Section 5.

4 The Metric

Based on psychological findings, we define a change blindness metric for measuring the blindness of an image pair. Given an input image pair I and I', we first assume that there is just a single change between them (in which region I_k is changed to region I'_k). Existing literature shows that the blindness depends on both the location of the change and the amount of change. We thus define our blindness metric B depending on both the visual saliency S (addressing the location of the change) and the amount of change D, as follows:

$$B = \exp\left(-\max\left(\|I_k\|S(I_k),\ \|I'_k\|S(I'_k)\right) \cdot D(I_k, I'_k)\right) \quad (1)$$

where B is the degree of change blindness in the range [0,1]. Higher blindness B means that the change is harder to detect; $\|I_k\|$ denotes the size of region I_k, $S(I_k)$ measures context-dependent saliency for region I (will be defined in Section 4.2); I_k and I'_k denote the region in the original image I and in the changed image I', respectively; $D(I_k, I'_k)$ denotes the amount of change from region I_k to I'_k (will be defined in Section 4.1). In Eqn. 1, we model the blindness as an exponential function of the saliency and the amount of change. We will verify this choice of exponential formulation later in Sec. 6.3.

The first term $\max\left(\|I_k\|S(I_k),\ \|I'_k\|S(I'_k)\right)$ accounts for the saliency, taking into account both sites I_k and I'_k. Taking the maximum in this saliency term is motivated by psychological findings [1], [12] which suggests that changes are easier to detect in more salient regions or in regions with higher saliency differences between image pairs. The second expression $D(I_k, I'_k)$ accounts for the amount of change between the two corresponding regions. It is worth noting that our metric is symmetric, in other words, the degree of blindness will not change if the original and changed images are swapped.

4.1 Amount of Change

The amount of change $D(I_k, I'_k)$ is defined as the sum of multiple feature differences:

$$D(I_k, I'_k) = \omega_c D_c(I_k, I'_k) + \omega_t D_t(I_k, I'_k) + \omega_s D_s(I_k, I'_k) \quad (2)$$

where D_c, D_t, and D_s are the color difference, the texture difference and the spatial difference of the two regions, respectively; ω_c, ω_t and ω_s are the corresponding weights and are determined by fitting the user statistics (Section 6.2). The color difference D_c is evaluated as the earth-mover's distance [50] between the color histograms (using $8 \times 8 \times 8$ bins)

of the two regions in *Lab* color space. The texture difference $D_{\mathbf{t}}$ is evaluated using the earth-mover's distance between Gabor wavelet features [51] of the two corresponding regions, with three scales and eight orientations. We employ the earth-mover's distance for computing differences as it is perceptually more meaningful than histogram matching techniques or Euclidean distances [50].

To compute the spatial difference $D_{\mathbf{s}}$, we first uniformly sample N points (N is usually 100) on the boundaries of both regions I_k and I'_k. Following [52], the correspondences between the two set of points are found by bipartite matching which minimizes shape context costs. The spatial difference is then evaluated as the average 2D distance between all corresponding point pairs.

When the change operator is deletion or insertion, one region does not exist. We simply regard the missing region as having the same size, shape, and position as the existing one, i.e. $D_{\mathbf{s}} = 0$. The background pixels enclosed by this imaginary region are then used to compute the color and texture differences. When the change operator is relocation, we only consider the spatial difference D_{s}, i.e. $D_{\mathrm{c}} = D_{\mathrm{t}} = 0$. In all above computations, *Lab* color values, histogram distribution, image image are normalized to $[0, 1]$.

4.2 Context-Dependent Saliency

Allman et al. [4] found that visual attention is highly context-dependent. The key is to define an effective context.

Fig. 3 (a) & (e) show two more or less identical red felt balls positioned at the image center. We observe that the visual saliency of the felt ball decreases as the background becomes more complicated. This suggests that visual saliency is *modulated by background complexity*. However, no existing saliency model attempts to explicitly quantify background complexity. State-of-the-art models, such as global contrast saliency [23], fail to model the influence of background complexity on the saliency, and the felt ball receives more or less the same saliency in Figs. 3(b) & (f).

Hence, our goal is to first quantify background complexity, and then utilize it to modulate the initial saliency. We start by defining the complexity of the whole image. Then, we derive a spatially varying image complexity based on this global one.

Global Image Complexity. A straightforward strategy for approximating global image complexity is to count the number of regions arising from segmentation [53]. However, since this strategy does not consider relative differences between regions (e.g. their appearance and positional differences), it is not very robust. Inspired by the all-pairs strategy of edit propagation [54], [55], [56], [57], we define the similarity between any two regions I_i and I_j using a Gaussian of the color difference,

$$e_{ij} = \exp(-D_{\mathrm{c}}^2(I_i, I_j)/\sigma_{\mathrm{e}}^2) \tag{3}$$

where D_{c} is the color difference defined in Eqn. 2; parameter σ_{e} controls the color range of influence and is set to 0.1 throughout our experiments. Then, the global image complexity is defined as:

$$C_{\mathrm{g}} = \sum_{i,j} w_{ij} e_{ij} / \sum_{i,j} w_{ij} \tag{4}$$

It sums the normalized similarities over all pairs of regions. The normalizing weight w_{ij} is defined as the product of the region sizes and a Gaussian of their distance,

$$w_{ij} = \|I_i\|\|I_j\| \exp(-(c_i - c_j)^2/\sigma_{\mathrm{w}}^2) \tag{5}$$

where $\|I\|$ returns the size of region I; c_i and c_j are the centroids of the two regions; σ_{w} controls the weight on spatial distance and is set to 0.4. The above complexity is high when nearby regions have large color differences. Note that all segmented regions take part in the above computation.

Spatially Varying Complexity. To define a context-dependent saliency of a region I_k, we need a spatially varying complexity that quantifies the surrounding context of I_k. Obviously, locations closer to I_k should have a higher influence. Hence, we slightly modify the weight in Eqn. 5 to incorporate the proximity influence by including the distances from I_k to I_i and I_j, as follows:

$$w'_{ij} = w_{ij} \exp\left(-\left((c_i - c_k)^2 + (c_j - c_k)^2\right)/\sigma_{\mathrm{w}}^2\right) \tag{6}$$

Then, the spatially varying complexity of region I_k can be defined as

$$C(I_k) = \sum_{i,j} w'_{ij} e_{ij} / \sum_{i,j} w'_{ij}. \tag{7}$$

Note that we define a per-region complexity instead of a per-pixel complexity. This is because regions are the primitives of change operators in our method.

Context-Modulated Saliency The context-modulated saliency of I_k is then defined as:

$$S(I_k) = S_o(I_k) C(I_k) \tag{8}$$

where S_o can be any existing saliency model. If the saliency model is computed in per-pixel basis, we can compute the saliency $S_o(I_k)$ of a region I_k by averaging the saliency values inside the region I_k. In particular, we adopt the global contrast model [23] as it is a region-based saliency model which naturally fits into our metric. The basic idea of our design is to modulate the local saliency S_o by the complexity of the surrounding background C. While the global contrast saliency returns more or less similar saliencies for the central red felt ball regardless of the background (Figs. 3(b) & (f)), our saliency model suppresses the saliency when the complexity of background is high,

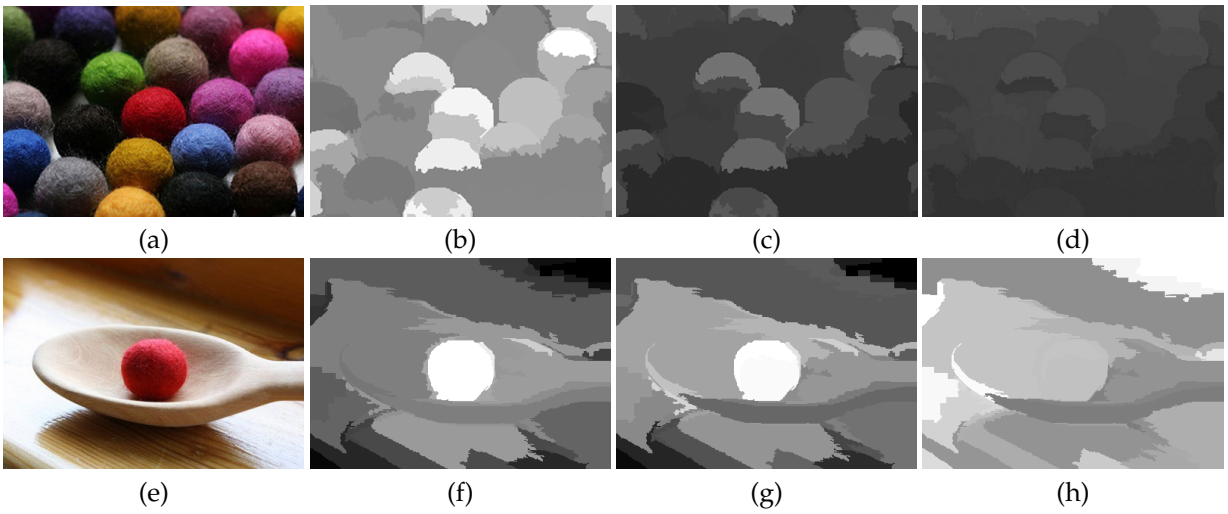


Fig. 3. Influence of background complexity on visual saliency. (a) & (e): input images; (b) & (f): global contrast saliency [23]; (c) & (g): our context-dependent saliency; (d) & (h) the complexity map C in Eqn. 7.

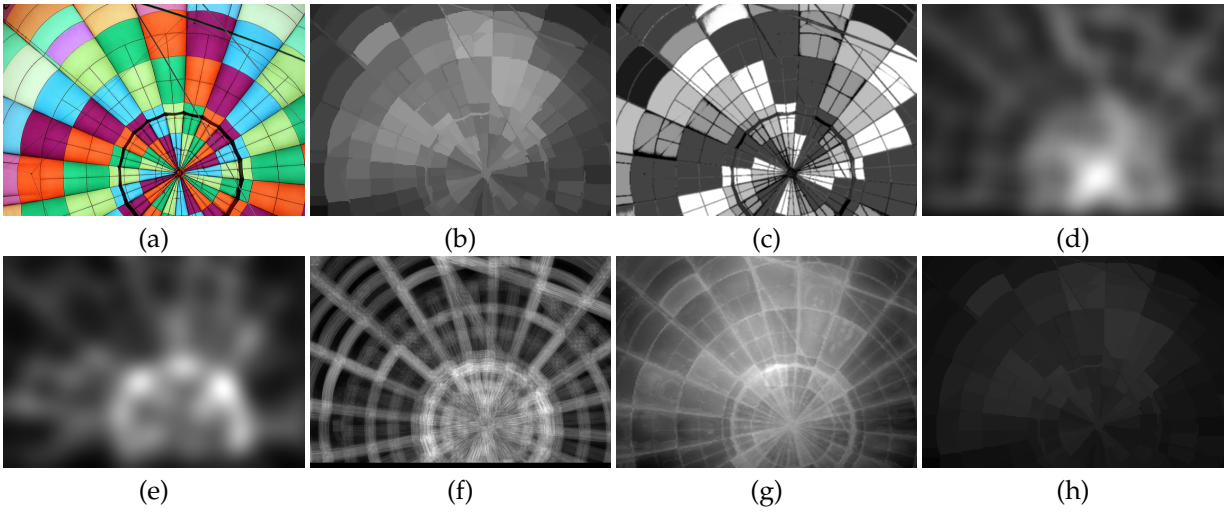


Fig. 4. Comparisons of saliency models. (a) the input image; (b) global contrast saliency [23]; (c) learning-based saliency [22]; (d) image signature [21]; (e) Itti model [15]; (f) AIM saliency [19]; (g) Judd model [18]; (h) our context-dependent saliency.

as in Fig. 3(c). Such suppression mimics human experience. Figs. 3(d) & (h) show the corresponding complexity maps C (Eqn. 7) used for modulation.

5 Optimization

With the above metric, we can optimize for a changed image by minimizing the following objective function:

$$|B - B^*|^2 + \lambda \max(M_{\min} - M, 0) \quad (9)$$

where B^* is the desired blindness and B is the current measured blindness in Eqn. 1; the right term provides a constraint, M is the sum of squared pixel difference (SSD) between the current changed image I' and the input I; $M_{\min}$ is a threshold to ensure that the changed image should at least contain $M_{\min}$ pixel changes, in order to avoid a degenerated or tiny change. In our experiments, we set $M_{\min} = 100$ and $\lambda = 1$.

Change Operators. We use seven change operators: insertion, deletion, replacement, relocation (translation), scaling, rotation, and color-shift. Insertion duplicates I_k and determines a 2D location for placement. Deletion removes I_k from the image and fills the region by inpainting. Replacement deletes I_k and duplicates another region at the same location. Relocation is based on a translation vector for relocating I_k. Scaling is based on a scaling factor to scale I_k. Rotation determines an angle to rotate I_k. Color-shift modifies the pixel colors in I_k by increasing or decreasing their hue and saturation. The parameters of all change operators are listed in Table 1. In our system, any holes in the background are filled by PatchMatch [58] and regions are composited using alpha matting [59].

Optimization Optimization is carried out as follows. Each time a region I_k is randomly selected from the

Operators	Dimension	Parameters
Insertion	2	inserting location (x, y)
Deletion	0	N/A
Replacement	0	N/A
Relocation	2	translation vector (x, y)
Scaling	1	scaling factor
Rotation	1	rotation angle
color-shift	2	hue and saturation changes

TABLE 1
Parameters of the supported change operators.

pool of candidate regions, and a change operator is randomly selected from the seven available operators. Except for the deletion and the replacement operators, parameter values of the change operation is iteratively adjusted to minimize the objective function (Eqn. 9). The best parameter values are determined by searching the parameter space in a gradient descent fashion. In detail, we first initialize the parameters by random values. In each iteration, we compute a descent direction and adjust the parameters by linearly searching the best values along the descent direction. If the objective function value does not reduce to a threshold (0.05 in our experiments) within a predefined number of iterations (30 in our experiments), we randomly pick another combination of candidate region and change operator for another round of iteration. The whole optimization halts when the optimization is satisfied, or the total number of rounds exceeds a predefined limit. The whole optimization takes between 10 seconds to 1 minute to generate a changed image.

For operators such as insertion and relocation, we need to restrict the location of I_k to a suitable space. Locations occupied by other candidate regions are avoided. In the case of relocation, if I_k is positioned at a location where its surrounding (in our case, a 3-pixel band surrounding the region) is very different from that at its original location (their color-histogram difference in terms of earth-mover's distance exceeds 0.2), this location is also prohibited.

6 Experiments and Results

6.1 User Study

We have conducted a user study to determine the best parameter values in our metric and evaluate its predictability in controlling the degree of blindness in change blindness tasks.

We download 100 real-world images from Flickr.com and generate 100 changed counterparts, resulting 100 image pairs. Each image is resized to have a resolution either 640×480 or 480×640. The changed counterparts are generated using our synthesis algorithm described in Sec. 5. Note that since the optimal values of weights ω_c, ω_t, and ω_s are obtained later after user study (in Sec. 6.2), here we use an empirical (not necessarily optimal) set of weights instead. The desired degrees of blindness are stratified sampled from a uniform distribution between [0,1]. Thirty subjects (17 males and 13 females with ages from 22 to 35) with normal vision participated in the experiment, which is carried out with the standard 'flickering' disruption [1]. In each trial, the before-change and after-change images are successively displayed to the subjects for 400ms, with a masking of 200 ms while images are switched. The subjects are asked to identify the change and the time taken is recorded, with a maximum of 60 seconds allowed. For each image pair, we compute an average recognition time of the thirty subjects. To reduce the influence of outliers, we remove 3 longest and 3 shortest recognition times before averaging.

To verify the reliability of the user study statistics, we performed a one way repeated measure analysis of variance (ANOVA). The resulting F-test value is $F(99, 2900) = 10.9$, $p \approx 0$. This measures how large inter-image variability is compared to intra-image (inter-subject) variability. Since the measured F-test value is much larger than the critical F-test value ($10.9 >> 1$), intra-image (inter-subject) variances are much smaller than inter-image variances, our statistics are meaningful. User study statistics and details, including all images, average recognition times and inter-subject variances of recognition time, can be found in the supplementary material.

Next, we divide the 100 image pairs into 2 sets: a *training set* for determining the optimal weight values in our metric, and a *validation set* for verifying the accuracy of our metric. Each set contains 50 image pairs with blindness degrees approximately uniformly distributed over the range [0,1].

6.2 Weight Determination

Recall that in Eqn. 2, there are three weights ω_c, ω_t and ω_s. Their values are determined by maximizing the predictability of our change blindness metric in Eqn. 1. To achieve this goal, we fit the three parameters by minimizing the sum of the L2-difference between the predicted blindness and the measured average recognition time of all image pairs in the training set:

$$\sum_{i=1}^{50} \left(B_i(\omega_c, \omega_t, \omega_s) - R_i\right)^2 \tag{10}$$

where $B_i(\omega_c, \omega_t, \omega_s)$ is the predicted blindness of the i-th image pair (Eqn. 1), given for certain values of ω_c,ω_t and ω_s. R_i is the measured average recognition time normalized to the range of [0,1] (In our case, the maximum allowable time is 60 seconds). We minimize Eqn. 10 using the gradient descent algorithm. A reliable initial solution is rather crucial. To obtain reliable initial values for ω_c, ω_t and ω_s, we slightly change the above minimization function in Eqn. 10

Pearson/Spearman correlation	global contrast	learning based	image signature	Itti model	AIM	Judd model
saliency along	0.44/0.47	0.38/0.41	0.34/0.27	0.42/0.46	0.42/0.44	0.43/0.44
with comp. modulation	0.51/0.53	0.39/0.42	0.34/0.29	0.47/0.49	0.44/0.47	0.43/0.46
with amount of change	0.62/0.64	0.48/0.51	0.37/0.36	0.59/0.61	0.60/0.62	0.61/0.63
comp. mod. and amt. of change	0.74/0.75	0.56/0.58	0.41/0.41	0.66/0.68	0.67/0.68	0.68/0.69

TABLE 2
Pearson and Spearman correlations between the recognition times and the predicted degrees of blindness, using different saliency models and different combinations of terms.

	Linear	Logarithm	Square Root	Exponential
Ours	0.47	0.52	0.66	0.74
Itti	0.41	0.43	0.56	0.66
AIM	0.37	0.41	0.56	0.67

TABLE 3
Pearson correlation between the recognition times and the predicted degrees of blindness, using different choices of formulation and different choices of saliency models.

by taking logarithm to both B_i and R_i:

$$\sum_{i=1}^{N} \left(\log B_i(\omega_{\mathrm{c}}, \omega_{\mathrm{t}}, \omega_{\mathrm{s}}) - \log R_i\right)^2 \qquad (11)$$

Since Eqn. 11 has a quadratic form, we can obtain the initial values of ω_{c}, ω_{t} and ω_{s} using a least square solver. Starting with these initial values, The final weight values are obtained by minimizing Eqn. 10 using gradient descent via multiple iterations. In our experiments, three iterations are sufficient. The optimal weight values are $\omega_{\mathrm{c}} = 2.96 \times 10^4, \omega_{\mathrm{t}} = 2.49 \times 10^4$, and $\omega_{\mathrm{s}} = 6.02 \times 10^2$, respectively.

It is interesting that the spatial weight ω_{s} is much smaller than the color (texture) weight ω_{c} (ω_{t}). In our experiment, we also find that position changes are harder to be detected than color or texture changes. As shown in the right example in Fig. 1, changing the location of the nose averagely takes a relatively long time, 44 seconds, to detect. This may be because that relatively small position change of an object does not significantly change the stored scene representation in human brains, making the change harder to be observed. This is consistent with the psychological finding that color is processed at a higher sampling rate than shape in memory [60].

6.3 Predictability

To evaluate the predictability of our metric, we compute the Pearson's correlation between the average recognition time and the blindness value predicted by the metric using the validation set. Our metric achieves a high correlation value of 0.74 (95% confidence interval [0.66,0.82]), which confirms the ability of our metric to control blindness.

There are no previous change blindness model to which we can make a comparison, so instead we compare our model to state-of-the-art saliency models. To do so, for each image pair in the validation set, we compute its predicted blindness using each saliency model. Note that saliency models are not designed for measuring the blindness of an image pair, so to enable a comparison we define the blindness for these saliency models as follows: we compute the sums of saliency over the changed region in both before and after images and took their maximum. Then the blindness B_m of a saliency model S_m is defined to be:

$$B_m = \exp(-\lambda_m \max\left(\|I_k\| S_m(I_k),\ \|I'_k\| S_m(I'_k)\right)). \qquad (12)$$

This takes a similar form to our blindness model in Eqn. 1. We then measure both Pearson correlation and Spearman rank correlation between the average recognition time and the blindness B_m predicted by the metric. Parameter λ_m in Eqn. 12 is adjusted for each specific metric S_m to maximize its corresponding correlation values. The correlations of different saliency models are given in the first row of Table 2. From left to right, we give correlations of the global contrast saliency model [23], learning-based saliency model [22], image signature [21], Itti model [15], AIM saliency model [19], and Judd model [18]. Note that the parameters of each saliency algorithm are also adjusted to maximize the corresponding correlation. Specifically, the parameters for the image signature algorithm are: 128 x 96 (image size) with specified blur. AIM is computed at full scale without Gaussian blur. All these saliency models are inferior to our change blindness metric (the left-bottom one) in terms of both Pearson and Speakman correlation measures.

We also examine how each component of our metric (Eqn. 1), including the context modulation term (C in Eqn. 8) and the amount of change term (D in Eqn. 2), improves the predictability. To do so, we compute correlations again using different combination of components and give them in Table 2. Starting from the second row, they are: context modulated saliency model without the amount of change, our metric (Eqn. 1) without context modulation , and our metric (Eqn. 1), which includes both context modulation and the amount of change components. Note that for each combination, the parameters are adjusted to maximize the correlation. From the results we can find that,

inclusion of either component (the second or third row) outperforms the one with saliency only (the first row). More importantly, having both components (the bottom row) gives the best predictability. This demonstrates that both context modulation and the amount of change are important terms for predicting change blindness.

We also verify the choice of formulation in defining the change blindness metric (Eqn. 1). We compare the exponential form to linear, logarithm and square root forms. The comparison is repeated with different saliency models used in our metric, including the global contrast based saliency model [23], Itti model [15], and the AIM saliency model [19], respectively. Again, for each formulation, a Pearson correlation is computed and the parameters are adjusted to maximize the corresponding correlation. The results are listed in Table 3. From the results, we can see that the exponential form gives the highest correlation among all formulations for all saliency models.

We then verify the choice for computing color and texture feature distances in Eqn. 2. We evaluate 3 different choices: Mahalanobis distance, Euler distance, and Earth-mover's distance. Pearson's correlations is computed for each choice. The correlations are 0.69 (Mahalanobis distance), 0.67 (Euler distance), and 0.74 (Earth-mover's distance), respectively. The highest correlation of Earth-mover distance justifies its appropriateness.

In the previous user study, each image pair has a different assigned blindness level. Next, we evaluate our ability to generate multiple change images with different degrees of blindness from a single input image. We have tested 6 input images, and for each input image, we generate three changed counterparts with desired blindness values of 0.2, 0.5 and 0.8, respectively. Thirty subjects were invited to identify the changes. To avoid the bias of being familiar with the same input, subjects were asked to consider each changed image at only one difficulty level. Fig. 5(a) plots the average recognition times for three blindness levels. Fig. 6 shows two examples, each with three levels of the desired blindness.

6.4 Context-Dependent Saliency

We compare our context-dependent saliency model to state-of-the-art saliency models using a challenging example in Fig. 4(a). The competitors include global contrast saliency [23] (Fig. 4(b)), learning-based saliency [22] (Fig. 4(c)), image signature [21] (Fig. 4(d)), the Itti model [15] (Fig. 4(e)), AIM saliency [19] (Fig. 4(f)), and the Judd model [18] (Fig. 4(g)). This input image is quite crowded and, in fact, distracting. It is hard to tell which specific region draws visual attention the most, unlike the felt ball example in Fig. 3(e). As existing saliency models do not consider contextual complexity, they may predict various regions to be most salient. Note that their predictions are not consistent with each other. In contrast, our context-dependent model returns rather low saliency values for almost all regions, which matches our visual experience better.

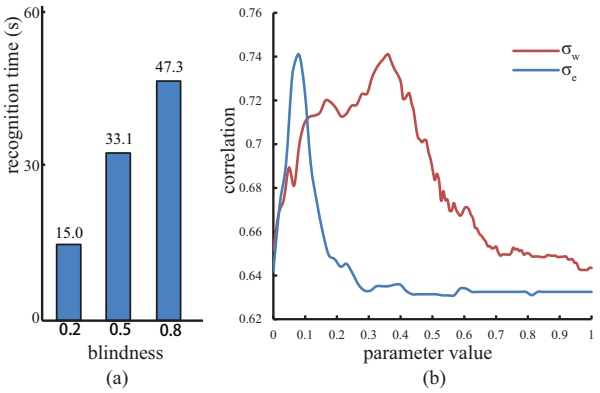


Fig. 5. (a) Average recognition times of three blindness levels for the same input images. (b) Correlation value against parameters σ_e and σ_w. The red curve: Fix $\sigma_e = 0.08$, σ_w varies from 0 to 1. The blue curve: Fix $\sigma_w = 0.4$, σ_e varies from 0 to 1.

We further validate the choices of parameters σ_e, σ_w in the context-dependent saliency model. Parameters σ_e and σ_w control the influence on color and spatial distances, respectively. In Fig. 5(b), we plot the Pearson correlation under different values of σ_e and σ_w. We find that setting $\sigma_e = 0.1, \sigma_w = 0.4$ approximately gives the highest correlation.

6.5 Visual Results

Figs. 1 and 7 show galleries of various results using different change operators. The desired blindness values are shown underneath each image pair. Fig. 6 shows two sets of input images and changed counterparts of different blindness levels (with increasing blindness from left to right).

So far, for each input image, we have applied only one change. In practice, our framework can be straightforwardly applied in multiple passes in order to obtain multiple changes. Each pass simply takes the output from the previous pass as its input, avoiding making the same change in following passes. Such results can be directly used for recreational purposes, such as spot-the-difference game. Figs. 8 (a)&(b) shows two such examples, 'poker' and 'shelf', and each of them contains 3 changes. The candidate regions used for synthesizing the changed counterparts are also shown in Figs. 8 (c)&(d). While the candidate regions in Fig. 8(c) are automatically extracted, the regions in Fig. 8(d) are manually refined from (c) in the 'shelf' example. Images with simple backgrounds normally do not require any user intervention (as in the 'poker' example). For complex images that require user intervention, refinement can be achieved by a few

(a) original image (b) $B^* = 0.2$ (c) $B^* = 0.5$ (d) $B^* = 0.8$

Fig. 6. Changed counterparts of different blindness levels.

Fig. 7. Gallery of change blindness images. The top left pair: $B = 0.05$; the top right pair: $B = 0.39$; the bottom left pair: $B = 0.61$; the bottom right pair: $B = 0.85$.

clicks within seconds. Solutions to all changed images can be found in Fig. 10.

7 Discussions

Our method offers a semi-automatic way for synthesizing changed images with desired degrees of blindness. By contrast, manual creation of change blindness images is ad hoc and lacks of control in difficulty levels.

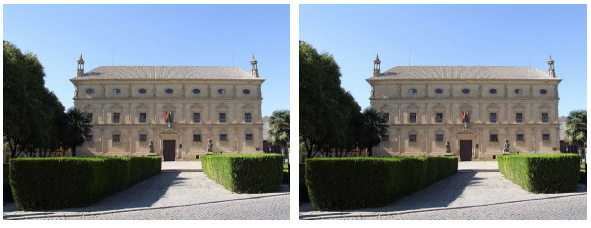
Besides the use in spot-the-difference games, our proposed metric potentially has many other applications, e.g., it can be used as a metric that measures significance of changes in an image revision control system [61]; it can also be used as a better way to find duplicated images for an image search engine. We believe the proposed model can also be applied in other areas of computer graphics, such as rendering acceleration, image retargeting, image tone mapping etc. Our proposed change blindness metric is also potential to serve as a more perceptually aligned metric for quantifying perceptual visual differences.

Our metric has certain limitations. First, since the components involved in our metric, such as image saliency and image color/texture/shape differences, are mainly low-level image features, hence, it might overlook some important semantic changes which our metric cannot model. Fig. 9 shows one such outlier. In this example, one window is removed and our metric predicts such change has a high blindness. However, humans are extremely sensitive to symmetry [62], [63].

In the future, additional semantic features, such as face and symmetry information, can be potentially integrated into our metric (Eqn. 2) in order to consider the influence of visually important semantics, e.g., face semantic can be incorporated by using a saliency model that integrates a face detector; and symmetry may be handled by first detecting repeated elements in the input image, and then penalizing changes on those repeated elements. Secondly, there is still a lot of room to further improve the predictability. Our current exponential formulation should be a good start, and more sophisticated forms may get even

(a) original image (b) changed counterpart (c) automatically extracted candidate regions (d) manually adjusted candidate regions

Fig. 8. Two examples for spot-the-difference game. Each example has been applied with 3 changes. (a) & (b) are original images and their changed counterparts. (c) are automatically extracted regions from (a). (d) is the manually adjusted result of the 'shelf' example from (c). Note that the 'poker' example in the first row does not need any user intervention.



(a) (b)

Fig. 9. A failure case. (a) original image; (b) changed counterpart.

better predictability. A possible way to find a better formulation is to deeply analyze experimental datas, e.g., analyzing different types of changes separately. Another possible direction to further improve the accuracy of the metric is to calibrate the amount of change by considering the just-noticeable-difference (JND). However, it is not trivial to quantitatively compute the JND for change blindness problems. Thirdly, since our proposed metric currently only considers bottom-up, data-driven saliency models, we cannot well predict change blindness behaviors with top-down, goal-driven controls, such as cueing [64]. Nevertheless, top-down factors, might be incorporated into our metric in the future, by replacing the used bottom-up saliency model with a top-down saliency model.

8 Conclusion

This paper presents the first computational model for change blindness, together with the first context-dependent saliency model which takes into account background complexity. The weight values in our metric are determined from user statistics. User studies demonstrate the effectiveness of our model in predicting degrees of change blindness. Our change blindness metric enables synthesis of changed images with a desired degree of blindness, which can be used in spot-the-difference games.

Acknowledgements. We thank the anonymous reviewers for their constructive comments. This work was supported by the National Basic Research Project of China (2011CB302205), the National High Technology Research and Development Program of China (2012AA011801) and the Natural Science Foundation of China (61170153). Tien-Tsin Wong was supported by RGC General Research Fund (Project No. CUHK 417411), and CUHK SHIAE (Project No. 8115034).

References

[1] R. A. Rensink, J. K. O'Regan, and J. J. Clark, "To see or not to see: The need for attention to perceive changes in scenes," *Psychological Science*, vol. 8, no. 5, pp. 368–373, 1997.

[2] R. A. Rensink, "Change detection," *Annual Review of Psychology*, vol. 53, pp. 245–277, 2002.

[3] D. J. Simons and D. Levin, "Change blindness," *Trends in Cognitive Sciences*, vol. 1, pp. 261–267, 1997.

[4] J. Allman, F. Miezin, and E. McGuinness, "Stimulus specific responses from beyond the classical receptive field: neurophysiological mechanisms for localvglobal comparisons in visual neurons," *Annual Review of Neuroscience*, vol. 8, pp. 407 – 430, 1985.

[5] C.-T. Yang, "Relative saliency in change signals affects perceptual comparison and decision processes in change detection," *Journal of Experimental Psychology: Human Perception and Performance*, vol. 37, no. 6, pp. 1708–1728, 2011.

[6] J. Grimes, "On the failure to detect changes in scenes across saccades," *Perception*, vol. 5, pp. 89–110, 1996.

[7] J. K. O'Regan, R. A. Rensink, and J. J. Clark, "Change blindness as a result of 'mudsplashes'," *Nature*, vol. 398, no. 34, 1999.

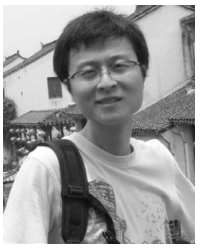
[8] D. J. Simons and D. T. Levin, "Failure to detect changes to people during a real-world interaction," *Psychonomic Bulletin & Review*, vol. 5, no. 4, pp. 644–649, 1998.

[9] D. J. Simons and R. A. Rensink, "Change blindness: Past, present and future," *Trends in Cognitive Sciences*, vol. 9, no. 1, pp. 16–20, 2005.
[10] D. J. Simons and M. S. Ambinder, "Change blindness: Theory and consequences," *Current Directions in Psychological Science*, vol. 14, no. 1, pp. 44–48, 2005.
[11] J. A. Stirk and G. Underwood, "Low-level visual saliency does not predict change detection in natural scenes," *Journal of Vision*, vol. 7, no. 10, pp. 3:1 –3:10, 2007.
[12] M. Verma and P. W. McOwan, "A semi-automated approach to balancing of bottom-up salience for predicting change detection performance," *Journal of Vision*, vol. 10, no. 6, pp. 3:1 –3:17, 2010.
[13] R. A. Rensink, "Change blindness: Implications for the nature of visual attention," in *Visual Attention*. Springer, 2001, pp. 169–188.
[14] A. M. Treisman and G. Gelade, "A feature-integration theory of attention," *Cognitive Psychology*, vol. 12, no. 1, pp. 97–136, 1980.
[15] L. Itti, C. Koch, and E. Niebur, "A model of saliency-based visual attention for rapid scene analysis," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 20, no. 11, pp. 1254–1659, 1998.
[16] Y.-F. Ma and H.-J. Zhang, "Contrast-based image attention analysis by using fuzzy growing," in *Proceedings of ACM Multimedia*, 2003, pp. 374–381.
[17] J. Harel, C. Koch, and P. Perona, "Graph-based visual saliency," in *Advances in Neural Information Processing Systems 19*, 2007, pp. 545–552.
[18] T. Judd, K. Ehinger, F. Durand, and A. Torralba, "Learning to predict where humans look," in *Proceedings of IEEE ICCV*, 2009, pp. 2106C–2113.
[19] N. D. B. Bruce and J. K. Tsotsos, "Saliency, attention, and visual search: An information theoretic approach," *Journal of Vision*, vol. 9, no. 3, pp. 5:1–5:24, 2009.
[20] R. Achanta, S. Hemami, F. Estrada, and S. Süsstrunk, "Frequency-tuned salient region detection," in *Proceedings of IEEE CVPR*, 2009, pp. 1597–1604.
[21] X. Hou, J. Harel, and C. Koch, "Image signature: Highlighting sparse salient regions," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 34, no. 1, pp. 194–201, 2011.
[22] T. Liu, Z. Yuan, J. Sun, J. Wang, N. Zheng, X. Tang, and H. Y. Shum, "Learning to detect a salient object," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 33, no. 2, pp. 353–367, 2011.
[23] M.-M. Cheng, G.-X. Zhang, N. J. Mitra, X. Huang, and S.-M. Hu, "Global contrast based salient region detection," in *Proceedings of IEEE CVPR*, 2011, pp. 409–416.
[24] S.-M. Hu, T. Chen, K. Xu, M.-M. Cheng, and R. R. Martin, "Internet visual media processing: a survey with graphics and vision applications," *The Visual Computer*, vol. 29, no. 5, pp. 393–405, 2013.
[25] A. Oliva and A. Torralba, "The role of context in object recognition," *Trends in Cognitive Sciences*, vol. 11, no. 12, pp. 520–527, 2007.
[26] A. Torralba, A. Oliva, M. S. Castelhano, and J. M. Henderson, "Contextual guidance of eye movements and attention in real-world scenes: The role of global features in object search," *Psychological Review*, vol. 113, no. 4, pp. 766–786, 2006.
[27] S. Goferman, L. Zelnik-Manor, and A. Tal, "Context-aware saliency detection," in *Proceedings of IEEE CVPR*, 2010, pp. 2376 – 2383.
[28] H. Yee, S. Pattanaik, and D. P. Greenberg, "Spatiotemporal sensitivity and visual attention for efficient rendering of dynamic environments," *ACM Transactions on Graphics*, vol. 20, no. 1, pp. 39–65, 2001.
[29] K. Cater, A. Chalmers, and P. Ledda, "Selective quality rendering by exploiting human inattentional blindness: looking but not seeing," in *Proceedings of the ACM symposium on VRST*, 2002, pp. 17–24.
[30] K. Cater, A. Chalmers, and G. Ward, "Detail to attention: exploiting visual tasks for selective rendering," in *Proceedings of the 14th Eurographics workshop on Rendering*, 2003, pp. 270–280.
[31] K. Cater, A. Chalmers, and C. Dalton, "Varying rendering fidelity by exploiting human change blindness," in *Proceedings of the 1st international conference on computer graphics and interactive techniques in Australasia and South East Asia*, 2003, pp. 39–46.
[32] J. Harrison, R. A. Rensink, and M. van de Panne, "Obscuring length changes during animated motion," *ACM Trans. Graph.*, vol. 23, no. 3, pp. 569–573, 2004.
[33] W.-C. Lin and Z.-C. Yan, "Attention-based high dynamic range imaging," *The Visual Computer*, vol. 27, no. 6–8, pp. 717–727, 2011.
[34] V. Sundstedt, D. Gutierrez, O. Anson, F. Banterle, and A. Chalmers, "Perceptual rendering of participating media," *ACM Transactions on Applied Perception*, vol. 4, no. 3, 2007.
[35] R. McDonnell, M. Larkin, B. Hernández, I. Rudomin, and C. O'Sullivan, "Eye-catching crowds: saliency based selective variation," *ACM Transactions on Graphics*, vol. 28, no. 3, pp. 55:1–55:10, 2009.
[36] F. Navarro, S. Castillo, F. J. Serón, and D. Gutierrez, "Perceptual considerations for motion blur rendering," *ACM Transactions on Applied Perception*, vol. 8, no. 3, pp. 20:1–20:15, 2011.
[37] G. Ramanarayanan, J. Ferwerda, B. Walter, and K. Bala, "Visual equivalence: towards a new standard for image fidelity," *ACM Transactions on Graphics*, vol. 26, no. 3, pp. 76:1 –76:11, 2007.
[38] S. J. Daly, "Visible differences predictor: an algorithm for the assessment of image fidelity," in *Proceedings of SPIE*, 1992, pp. 2–15.
[39] A. Oliva, A. Torralba, and P. G. Schyns, "Hybrid images," *ACM Transactions on Graphics*, vol. 25, no. 3, pp. 527–532, 2006.
[40] M.-T. Chi, T.-Y. Lee, Y. Qu, and T.-T. Wong, "Self-animating images: illusory motion using repeated asymmetric patterns," *ACM Transactions on Graphics*, vol. 27, no. 3, pp. 62:1–62:8, 2008.
[41] N. J. Mitra, H.-K. Chu, T.-Y. Lee, L. Wolf, H. Yeshurun, and D. Cohen-Or, "Emerging images," *ACM Transactions on Graphics*, vol. 28, no. 5, pp. 163:1–163:8, 2009.
[42] H.-K. Chu, W.-H. Hsu, N. J. Mitra, D. Cohen-Or, T.-T. Wong, and T.-Y. Lee, "Camouflage images," *ACM Transactions on Graphics*, vol. 29, no. 4, pp. 51:1–51:8, 2010.
[43] Q. Tong, S.-H. Zhang, S.-M. Hu, and R. R. Martin, "Hidden images," in *Proceedings of NPAR*. ACM, 2011, pp. 27–34.
[44] C. O'Sullivan, S. Howlett, Y. Morvan, R. Mcdonnell, and K. O'Conor, "Perceptually adaptive graphics," in *Proceedings of Eurographics State-of-the-Art Report*, 2004, pp. 141–164.
[45] D. Bartz, D. Cunningham, J. Fischer, and C. Wallraven, "The role of perception for computer graphics," in *Proceedings of Eurographics State-of-the-Art Report*, 2008, pp. 65–86.
[46] A. McNamara, K. Mania, and D. Gutierrez, "Perception in graphics, visualization, virtual environments and animation," in *SIGGRAPH Asia 2011 Courses*, ser. SA '11, 2011, pp. 1–137.

Fig. 10. Solutions to all changed images shown in our paper.

[47] Y. Liu and Y. Yu, "Interactive image segmentation based on level sets of probabilities," *IEEE Transactions on Visualization and Computer Graphics*, vol. 18, no. 2, pp. 202–213, 2012.
[48] M. J. Dahan, N. Chen, A. Shamir, and D. Cohen-Or, "Combining color and depth for enhanced image segmentation and retargeting," *The Visual Computer*, vol. 28, no. 12, pp. 1181–1193, 2012.
[49] D. Comaniciu and P. Meer, "Mean shift: A robust approach toward feature space analysis," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 5, pp. 603–619, 2002.
[50] Y. Rubner, C. Tomasi, and L. J. Guibas, "The earth mover's distance as a metric for image retrieval," *International Journal of Computer Vision*, vol. 40, no. 2, pp. 99–121, 2000.
[51] B. Manjunath and W. Ma, "Texture features for browsing and retrieval of image data," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 18, no. 8, pp. 837–842, 1996.
[52] S. Belongie, J. Malik, and J. Puzicha, "Shape matching and object recognition using shape contexts," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 4, pp. 509–522, 2002.
[53] T. Chen, M. M. Cheng, P. Tan, A. Shamir, and S. M. Hu, "Sketch2photo: Internet image montage," *ACM Transactions on Graphics*, vol. 28, no. 5, pp. 1–10, 2009.
[54] X. An and F. Pellacini, "Appprop: all-pairs appearance-space edit propagation," *ACM Transactions on Graphics*, vol. 27, no. 3, pp. 40:1–40:9, 2008.
[55] B. J. Frey and D. Dueck, "Clustering by passing messages between data points," *Science*, vol. 315, pp. 972–976, 2007.
[56] K. Xu, Y. Li, T. Ju, S.-M. Hu, and T.-Q. Liu, "Efficient affinity-based edit propagation using k-d tree," *ACM Transactions on Graphics*, vol. 28, no. 5, pp. 118:1–118:6, 2009.
[57] B. Feng, J. Cao, X. Bao, L. Bao, Y. Zhang, S. Lin, and X. Yun, "Graph-based multi-space semantic correlation propagation for video retrieval," *The Visual Computer*, vol. 27, no. 1, pp. 21–34, 2011.
[58] C. Barnes, E. Shechtman, A. Finkelstein, and D. B. Goldman, "Patchmatch: a randomized correspondence algorithm for structural image editing," *ACM Transactions on Graphics*, vol. 28, no. 3, pp. 24:1–24:11, 2009.
[59] A. Levin, D. Lischinski, and Y. Weiss, "A closed form solution to natural image matting," in *Proceedings of IEEE CVPR*, 2006, pp. 61–68.
[60] C. Kent and K. Lamberts, "The time course of perception and retrieval in matching and recognition," *Journal of Experimental Psychology: Human Perception and Performance*, vol. 32, no. 4, pp. 920–931, 2006.
[61] H.-T. Chen, L.-Y. Wei, and C.-F. Chang, "Nonlinear revision control for images," *ACM Transactions on Graphics*, vol. 30, no. 4, pp. 105:1–105:10, 2011.
[62] T. A. Kelley, "Effects of scene inversion on change detection of targets matched for visual salience," *Journal of Vision*, vol. 3, pp. 1–5, 2003.
[63] M. Wertheimer, "Untersuchungen zur lehre der gestalt ii," *Psycologische Forschung*, no. 4, pp. 301–350, 1923.
[64] V. Aginsky and M. J. Tarr, "How are different properties of a scene encoded in visual memory?" *Visual Cognition*, vol. 7, no. 1/2/3, pp. 147–162, 2000.

Li-Qian Ma is currently a graduate student in the Department of Computer Science and Technology, Tsinghua University. He received his bachelor's degree from Tsinghua University in 2010. His research interests include image/video editing and real-time rendering.



Kun Xu is an assistant professor in the Department of Computer Science and Technology, Tsinghua University. Before that, he received his bachelor and doctor's degrees from the same university in 2005 and 2009, respectively. His research interests include realistic rendering and image/video editing.

Tien-Tsin Wong received the B.Sci., M.Phil., and PhD. degrees in computer science from the Chinese University of Hong Kong in 1992, 1994, and 1998, respectively. Currently, he is a professor in the Department of Computer Science & Engineering, Chinese University of Hong Kong. His main research interest is computer graphics, including computational manga, image-based rendering, GPU techniques, natural phenomena modeling, and multimedia data compression. He received IEEE Transactions on Multimedia Prize Paper Award 2005 and Young Researcher Award 2004.

Bi-Ye Jiang is currently an undergraduate student in the Department of Computer Science and Technology, Tsinghua University. His research interests include image editing.

Shi-Min Hu received his PhD degree from Zhejiang University in 1996. He is currently a professor in the Department of Computer Science and Technology at Tsinghua University. His research interests include digital geometry processing, video processing, rendering, computer animation, and computer-aided geometric design. He is associate Editor-in-Chief of The Visual Computer, and on the editorial boards of Computer-Aided Design and Computer & Graphics. He is a member of the IEEE and ACM.