

# Detection of Image Stretching

Yuzhen Niu<sup>1\*</sup>

Feng Liu<sup>2†</sup>

Xueqing Li<sup>1‡</sup>

Huiyun Bao<sup>1§</sup>

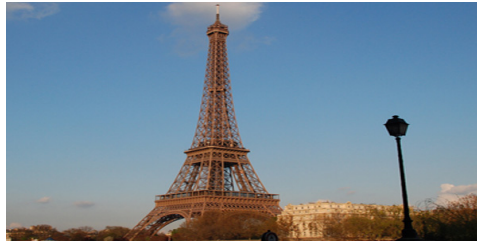
Michael Gleicher<sup>2¶</sup>

<sup>1</sup> Shandong University, China

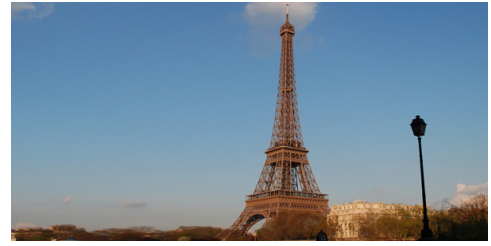
<sup>2</sup> University of Wisconsin-Madison, USA



(a) Input (aspect ratio 4:3)



(b) Horizontal stretching (aspect ratio 2:1)



(c) Seam carving (aspect ratio 2:1)



(d) Input (aspect ratio 1:1)



(e) Horizontal stretching (aspect ratio 3:2)



(f) Seam carving (aspect ratio 3:2)

**Figure 1:** Examples of image retargeting. For input (a), a horizontal stretch distorts the Eiffel Tower (b), and the content-aware seam carving method in Photoshop CS4 avoids this distortion (c). For input (d), the seam carving method introduces more undesirable artifacts (f) than stretching (e). People may not notice the stretch in the result (e).

## Abstract

Resizing images for different devices often involves changing the aspect ratio. A wide variety of approaches for resizing exist: sophisticated “content-aware” (or retargeting) approaches are built on the assumption that carefully chosen distortions are preferable to the naïve approach of uniformly stretching the image. However, there is little codified understanding of how distortions of the image, including uniform stretching or more complex warps introduced by retargeting, are perceived. In this paper, we describe experiments that explore the perception of image stretching, to establish the baseline for assessing more complex resizing methods, as well as to establish the methodology. In a series of experiments, we show that the perception of stretching is a complex phenomenon depending on a myriad of factors including the amount of distortion, the image content, the viewer’s cultural background, and the observation time. We provide a methodology for creating images that avoid unfair cues to stretching and explore issues in using online worker communities for studies. We show that even small stretches can be detected in some cases. These findings have ramifications for the design and evaluation of image retargeting, and suggest that a more thorough study of distortion perception is necessary.

\*e-mail: yuzhen@cs.wisc.edu

†e-mail: fliu@cs.wisc.edu

‡e-mail: xqli@sdu.edu.cn

§e-mail: baohuiyun@gmail.com

¶e-mail: gleicher@cs.wisc.edu

<sup>1</sup>This work was done during Yuzhen Niu’s visit to University of Wisconsin-Madison, USA.

**CR Categories:** I.4.m [Image Processing and Computer Vision]: Miscellaneous

**Keywords:** Image stretching, Image distortion, Perceptual sensitivity, Image retargeting, User study

## 1 Introduction

Images often need to be displayed at a different size and aspect ratio than they were originally created. The importance of resizing grows as devices become diverse. The naïve approach to resizing simply scales the image, potentially stretching it to the a new aspect ratio. Newer *content-aware* methods perform intelligent adaptation. Such methods are also called *image retargeting*. In the past several years, a rich literature on image retargeting has emerged, a recent survey cites dozens of papers [Niu et al. ] and commercialized versions have already appeared (for example, in Photoshop CS4).

Early retargeting methods, such as [Suh et al. 2003], applied non-distorting transformations such as cropping. However, most modern approaches, including those based on seam carving [Avidan

and Shamir 2007] or non-linear warping [Wolf et al. 2007; Wang et al. 2008], introduce complex distortions into the results. The premise of such methods is that these distortions are less objectionable than those that would be introduced by naïve resizing, for example because the content aware approaches may be able to “hide” the distortions in parts of the image where they are unlikely to be noticed. Figure 1 shows two examples. In one case, the seam carving method succeeds at hiding the stretching, while in the other, naïve stretching provides a more reasonable result. In order to understand when these methods are appropriate, we need a better understanding of how people perceive retargeting results.

In previous retargeting research, it is common to compare specific retargeting methods to others to conclude people’s viewing preference. Some retargeting methods, such as [Krähenbühl et al. 2009], conducted user studies to evaluate their methods. The goal of these kinds of evaluations is mainly to study the relative performance or desirability among their proposed methods and others. While many retargeting methods have been evaluated, there has been little examination of the fundamental principles of how distortions are perceived. Without this grounding in perceptual foundations, retargeting methods are necessarily ad hoc: we have few good principles to know when and where and what kinds of distortions should be applied. A better understanding of how people perceive distortion will lead to better retargeting methods and better metrics for evaluating approaches.

The fundamental premise of retargeting is that the complex distortions introduced by content-aware methods are less objectionable than the baseline provided by naïve resizing. However, even this baseline is not well understood. In this paper, we begin the study of the perception of image distortions by considering the stretching introduced by resizing. Our goals are to understand the types of factors that must be considered in future study, to establish methodology, and to measure how well people can notice stretching (since this is what motivates more sophisticated techniques).

This paper describes a series of experiments that study the perception of image stretching. The first experiment shows that even for these simple distortions, the sensitivity to stretching depends on several factors including the amount of distortion, the content of the image, and the viewer’s cultural background, as described in detail in our previous work [Niu et al. 2010]. Extending this experiment to use crowdsourcing to obtain a broader subject pool suggests that additional factors must be considered in the experimental design. A third experiment explores the category of images to which people are most sensitive: on images with people in them, viewers can see moderately small stretches even with brief exposures to the images.

Our work has several ramifications. First, we give a methodical process for creating images for fair testing of retargeting methods, and suggest requirements for evaluation including considering the cultural background of experiment participants. Second, we show that a wide variety of factors effect the perception of distortion, which has implications for the design of retargeting methods, suggesting that these factors are considered. Third, we show that in some cases, viewers are very sensitive to stretches, reinforcing the need for methods that effectively hide the stretching using content-aware approaches.

## 2 Phase 1: Factors affecting stretch

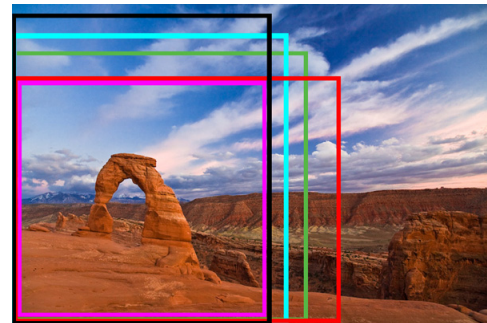
### 2.1 Study Design

Our hypothesis of this phase is that the perception of image stretch depends on many factors. To investigate this, we conducted a user study which was hosted on web to measure how well people can

notice stretching. The experiment was designed to examine the interactions of a number of effects including stretch scale, image content, and the viewer’s cultural background.

#### 2.1.1 Materials

Our first challenge was to select an appropriate set of images for the study. Our design (below) required 32 images that with sufficient quality and diversity of content. We selected 32 images from *Flickr* based on three criteria. First, this image set should cover a wide range of image categories. In our study, the image set covered all of the 14 categories used in [Torralba and Oliva 2003] and covered 19 out of 20 categories given by [Mojsilović and Rogowitz 2004]. We omitted textures and patterns because most are of artificial design. Second, all of the images in the set should be unstretched. To verify this, each image was checked to see if its aspect ratio matches the camera with which it was taken. Lastly, all selected images should be large enough so that stretching will not introduce noticeable resampling artifacts that would serve as a cue. All of the selected images were at least  $800 \times 600$  pixels.



(a) AOI (pink) and cropping windows (in other colors)



(b) 0% (cropping window in red)



(c) 20% (cropping window in green)



(d) 40% (cropping window in cyan)



(e) 60% (cropping window in black)

**Figure 2:** An input image and its four cropped and stretched results.

Next, we needed to create 4 versions of each image with varying levels of stretch. In order to keep the experiment duration reasonable in this initial study, we focused on horizontal stretch, and chose uniformly increased stretch scales, that are 0%, 20%, 40%, and 60%. These variants were created in a way that minimized possible clues of stretching, such as size and quality of composition. To do this, we created 4 variants of the same size ( $400 \times 300$ ) by cropping

appropriately sized regions and scaling them to this aspect ratio as shown in Figure 2. During the scaling process, we only stretch the cropping using bi-cubic interpolation, no other operations are used, such as seam carving or spatially-varying warping. To perform the cropping, we first manually identified an area of interest (AOI). We then randomly generated a number of appropriately sized cropping regions that included the AOI. Cropped regions contained approximately the same number of pixels to avoid having different amounts of image information and similar resampling artifacts. From these candidate croppings, we manually selected one that we considered to be the best in terms of its compositional quality, preferring ones that avoided cutting objects and that approximate compositional rules (such as centering or the rule of thirds) [Krages 2005].

The result of this preparation was a set of 32 diverse, high quality images. For each, we had 4 variants that were created using different amounts of horizontal scale, but were otherwise similar in size, composition, and resampling artifacts.

### 2.1.2 Design

In our study, the stretch scale (0%, 20%, 40%, and 60%) and image category (human, animal, architecture, man-made object, ocean, mountain and plant) are two within-participants factors, country of origin (U.S. and China) is a between-participants factor. To counterbalance stretch scales across all participants and images, a  $4 \times 4$  Latin square design was used [Box et al. 2005]. A Latin square is an  $n \times n$  table filled with  $n$  different symbols in such a way that each symbol occurs exactly once in each row and exactly once in each column. In the design of experiments, Latin squares are a special case of row-column designs for two blocking factors. The two block factors in our design are participant groups and image groups. The Latin square design is used to help eliminate from the stretch scale comparisons possible differences between the participant groups and image groups. Specifically, for every 4 participants, we assign each of them to one of the four participant groups. We also randomly assign 32 images to 4 image groups such that each image group gets 8 images. Each participant is shown all of the 4 image groups, where each group of images is stretched at a different scale (0%, 20%, 40%, and 60%). The order in which each image is shown to each participant was randomized.

### 2.1.3 Participants

We conducted this web-based user study with two different subject pools. The U.S. pool had 62 participants (51 males and 11 females, from 21 to 57 years old, average age of 26.5), all of whom were graduate students. The China pool comprised 84 participants (54 males and 30 females, from 20 to 36 years old, average age of 26.4), all of whom were either undergraduate or graduate students. All participants were volunteers. After the study, we excluded the data from 8 U.S. participants and 15 China participants because they either did not complete all of the images or gave unreliable answers.

### 2.1.4 Procedure

The study begins with an introductory page that informs participants on the task motivation and description. The first two images in the experiment are examples (not used in analysis), one that is clearly stretched, and another that is not stretched. These examples provide a check that the participant understands the instructions. Following the examples, the 32 images are shown one by one. For each image, each participant was asked whether the image was stretched or not by selecting the radio button “Yes. I notice the stretching.” or “No. Its an original image.” At the end of the study,

Category	20%	40%	60%	post-hoc		
Image set	0.83	1.35	1.89	**	**	**
Human	1.57	2.45	3.01	**	**	**
Animal	1.14	2.39	2.89	o	**	**
Plant	0.68	1.43	2.07	o	**	**
Man-made object	0.88	1.26	1.98	**	**	**
Architecture	0.71	1.31	1.60	**	**	**
Ocean	0.38	0.36	0.94	o	o	*
Mountain	0.21	0.24	1.08	o	o	**

**Table 1:** Perceptual sensitivities for the whole image set and 7 image categories. ?% column is the sensitivity at ?% stretch scale. The last column shows the post-hoc tests following ANOVA, on all U.S. participants’ behavior at 0% and 20% scale, 0% and 40% scale, and 0% and 60% scale. Three results are separated by ||, and symbol o, \*, and \*\* mean  $p \geq 0.05$ ,  $0.01 \leq p < 0.05$ , and  $p < 0.01$  respectively. So \* and \*\* stand for people can detect the stretch, while o stands for people can’t detect the stretch.

each participant was shown how well she/he did as well as the images for which she/he gave incorrect answers. On average, it took participants about five minutes to complete this study.

## 2.2 Results

Our study shows that people’s perception of image stretch depends on a variety of factors, such as stretch scale, image content, and subjects. We performed a mixed-model analysis of variance (ANOVA) with stretch scale and image category as within-participants factors and country of origin as a between-participants factor. The results are based on all participants’ answers to the images shown to them. We found two significant main effects of stretch scale ( $F(3, 315) = 338.99$ ,  $p < 0.001$ ) and image category ( $F(6, 690) = 85.60$ ,  $p < 0.001$ ) and a marginal main effect of subject pool ( $F(1, 85) = 3.09$ ,  $p = 0.08$ ). We also identified a significant interaction between stretch scale and image category ( $F(18, 1267) = 1.62$ ,  $p < 0.001$ ). No other main effects or interaction effects were found. Post-hoc tests were conducted for stretch scales over the whole image set and each image category, comparing 20%, 40%, 60% stretch scales to 0% stretch scale, and revealed that the stretch scale had a different effect on different images. Full details of post-hoc tests can be seen in Table 1.

### 2.2.1 Stretch Scale

Our data show that people’s sensitivity to image stretch increases with the stretch scale. We calculated people’s sensitivity to image stretch based on Signal Detection Theory [Macmillan and Creelman 1991]. Sensitivity measures the discrepancy between the hit rate (H), the frequency that stretched images are correctly detected; and the false-alarm rate (F), the frequency that unstretched images are incorrectly judged as stretched ones. Perfect sensitivity means a hit rate of 1 and a false-alarm rate of 0. According to the Signal Detection Theory, sensitivity,  $d$ , is calculated in terms of  $z$  that is the inverse of the cumulative normal distribution function:

$$d = z(H) - z(F). \quad (1)$$

Figure 4 shows people’s average sensitivity over 20%, 40%, and 60% stretch scales plus the origin (zero sensitivity at 0% stretch scale). There we can see that people’s sensitivity to image stretch increases with the stretch scale, as supported by a chi-square analysis (U.S.:  $\chi^2(1, N = 3) = 70.9$ ,  $p < 0.001$ ; and China:  $\chi^2(1, N = 3) = 63.5$ ,  $p < 0.001$ ).

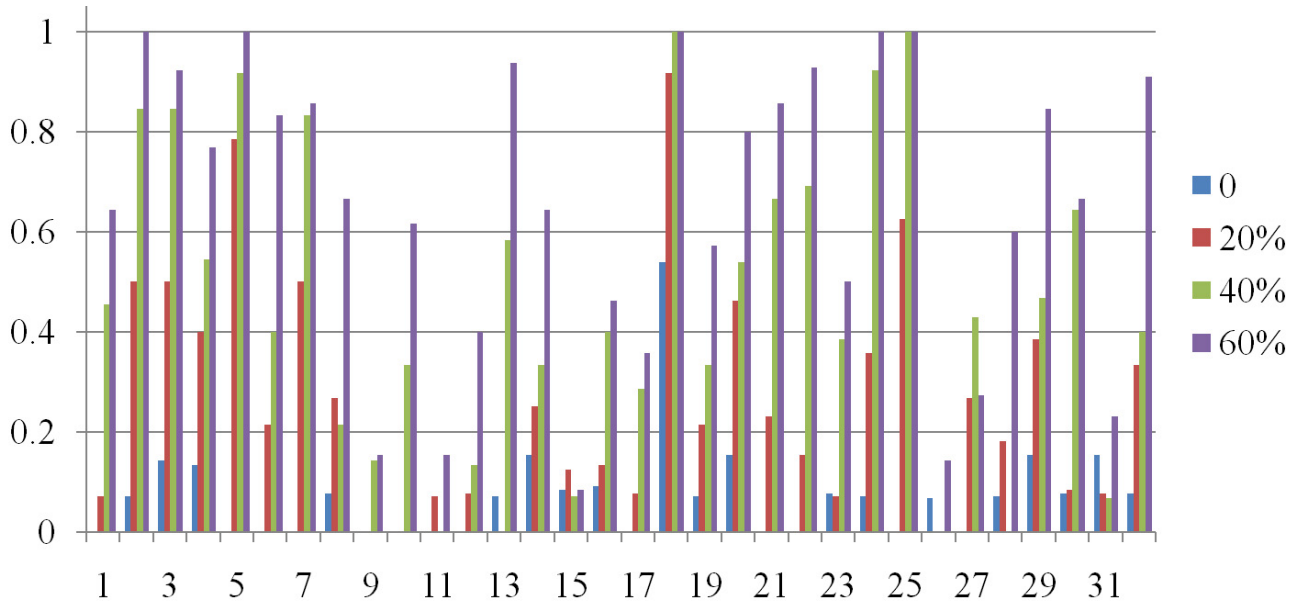


Figure 3: Perceived stretch rates of 32 images. Image number ranges from 1 to 32.

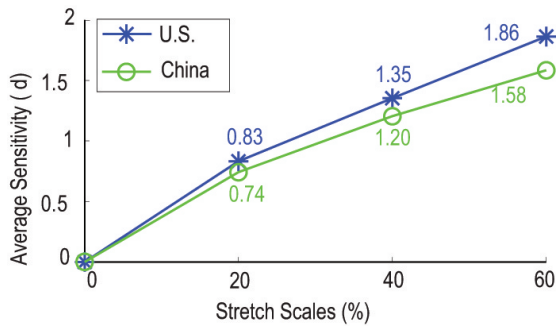


Figure 4: Average sensitivity over four stretch scales.

### 2.2.2 Image Content

We analyzed people’s perceptual sensitivity to different image categories in this section. We found that people are more sensitive to stretch in images with humans, animals and man-made objects than those with natural scenes. For clarity, here we only report the common observations we had from this study over the subject pools in U.S. and China. We take the U.S. subject pool as an example here. The difference between these two pools is reported and analyzed in the following subsection.

Figure 3 shows each image’s perceived stretch rate, the fraction of times that people thought it was stretched. Here we can see that it is easier to detect the stretch in some images than that in the others. We sorted the 32 images according to the sum of people’s sensitivity to them over all stretch scales. We find that all of the 5 images that people are most sensitive to when stretched are images with humans, animals and manmade objects. And all of the 5 images that people are least sensitive to when stretched are images with natural scenes, like oceans and mountains.

To better understand people’s perceptual sensitivity to different image categories, we show people’s average sensitivity to the whole image set and 7 categories and the post-hoc tests following ANOVA

results which show whether participants can detect 20%, 40%, and 60% stretches in Table 1. We sorted the 7 categories according to the sum of their sensitivities over four stretch scales and then arranged all categories in Table 1 by the sensitivity rank.

### 2.2.3 Viewer’s Cultural Background

While our results do not show a significant difference in the overall sensitivity between the U.S. and China subject pools ( $F(1, 367) = 1.51, p = 0.22$ ), we do find significant differences in some cases. For example, in images of man-made objects, the sensitivity of stretching is significantly different between pools. At the 20% scale, U.S. participants were more likely to notice the stretch (sensitivity: 0.88 vs. 0.07). An image that exemplifies the issue is shown in Figure 5. The stretched object appeared more consistent with Chinese participants experience based on the post-test questionnaire comments.

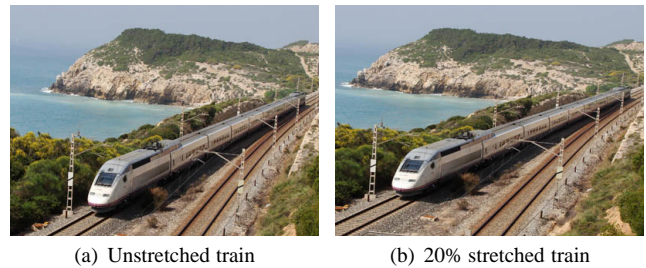


Figure 5: An example where Chinese participants is less sensitive to stretching than participants in U.S..

## 3 Phase 2: Crowdsourcing

Our first experiment suggests the need to consider a wide range of images and subjects in studying stretch perception. We were intrigued by the possibility of using an online worker community (Amazon Mechanical Turk [Amazon ]) as a cost-effective way of

getting access to a large and diverse subject pool. Using the Mechanical Turk service for user studies has a number of challenges [Kittur et al. 2008]. First, the study must be made to fit into the mechanisms of the service. Secondly, the payment structure gives peculiar incentives to the participants that can skew the results. Our second experiment tried to address these two questions.

Our second experiment adapted the first to use the Mechanical Turk. However, the answers provided were different from the first experiment, which reveal implications for our later experimental design.

### 3.1 Study Design

As in the first study, the second experiment began by giving instructions to potential participants and asking them to respond to two easy examples that make sure they understand the instructions. Participants who gave correct answers to these two examples are then assigned 32 images determined as in the first study. Each image is presented on a web page as in the first study. Each image trial was provided as a web page called a “Human Intelligence Task,” and the service allowed assigning a set of tasks to the qualified workers.

The Mechanical Turk service allows to limit the location of participants and can limit the work to a set number of participants. For our initial experiment, we chose participants from the U.S. so that we could compare with our results from the first experiment, and limited the experiment to the first 24 participants who completed all 32 images. We disqualified 3 participants who gave unreliable answers (e.g. answered yes or no to all images). These participants were paid only for their correct answers. We approved the per-image reward for all other participants whether their answers were correct or not.

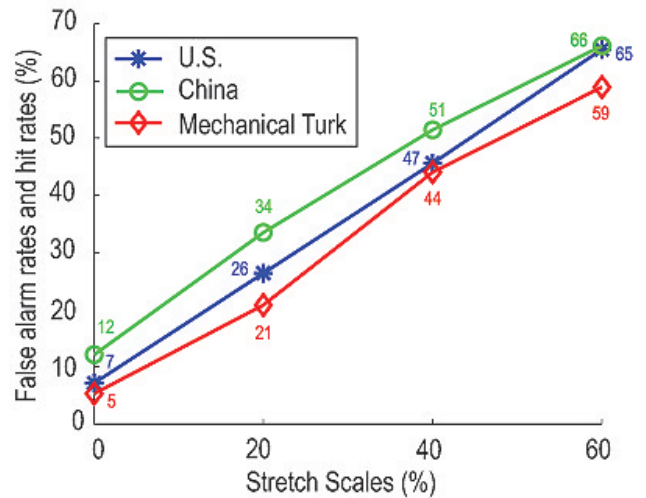
### 3.2 Results

The results from the second experiment were similar to the first. The mixed-model ANOVA showed the same significant main effects of stretch scale ( $F(3, 327) = 399.20, p < 0.001$ ) and image category ( $F(6, 741) = 96.48, p < 0.001$ ). The test also identified a significant main effect of subject pool ( $F(2, 83) = 3.5187, p = 0.03$ ).

The participants in the three subject pools, the participants in U.S. and China as described in Section 2.1.3 and the Mechanical Turk workers, had similar sensitivity, an ANOVA on all participants’ sensitivities at three stretch scales showed the different subject pools are not significantly different ( $F(2, 429) = 0.93, p = 0.39$ ).

However, the three subject pools’ average false alarm rates at 0% stretch scale and hit rates over 20%, 40%, and 60% stretch scales are different ( $F(2, 573) = 3.46, p = 0.03$ ). As shown in Figure 6, Chinese participants responded to yes the most, the U.S. pool was in the middle, and the Mechanical Turk pool was lowest. This order is supported by a chi-square analysis ( $\chi^2(1, N = 2) = 8, p = 0.018$ ). Among these three subject pools, participants in China give the most liberal answers, while Mechanical Turk workers give the most conservative answers.

These differences suggest a possible bias in the experiment, that is how the question is interpreted (what does it mean for an image to be “stretched?”). However, we feel that the similarities between the MT results and the regular U.S. subject pool are strong enough that MT can be a useful source of participants in the future, if the experimental design is improved.



**Figure 6:** Average false alarm rates at 0% stretch scale and hit rates over 20%, 40%, and 60% stretch scales for three subject pools.

## 4 Phase 3: Detection thresholds

Our third experiment aimed to more closely establish the threshold at which point stretching can be detected. Based on the previous experiment, we needed an improved design to reduce potential bias. Specifically, we used a forced-choice design where image stretching is identified relative to unstretched images. Furthermore, we made the experiment time-bounded to investigate the effect of observation time. We also wanted to consider vertical stretches (or horizontal stretches less than 100%). In order to keep the material requirements and participant time reasonable, we chose to limit ourselves to a single category of images: images with people in them.

### 4.1 Study Design

The primary aim of this study was to determine the thresholds for how much images with people must be stretched for the distortion to be reliably detected. However, in the process, we also needed to consider the amount of time the participants see the image, as well as whether horizontal or vertical stretching were different.

Our basic design is a forced-choice comparison. For each trial, the participant is shown a pair of images: one is stretched and the other is not stretched. The participant is shown two different images on each trial (see Figure 7). For each trial the participant must select which of the two images were stretched. Participants were shown the images for a fixed duration of time after which point the images were removed.

#### 4.1.1 Materials

The design requires two images per trial, which meant we needed to assemble a larger collection of images using the process described earlier. We selected images from the “human” category (pictures with one or more people in them, examples are shown in Figure 7). We were most interested in this category as it is a particularly important type of image and people had the most sensitivity.

We selected 48 images with human subjects from *Flickr* based on three criteria. First, the image set should cover a wide range of



**Figure 7:** An example set of 8 image pairs from the easy part of a session.

variety in age, gender, number (from single person to crowd), activity, expression, and shot type (from close-ups to extreme long shots). The criteria for unstretching and high quality are the same as described in Section 2.1.1.

13 versions of each image (unstretched, plus 4%, 8%, 12%, 16%, 20%, and 24% in horizontal and vertical) were created using the method described in Section 2.1.1. An input image and its 13 cropped and stretched results are shown in Figure 8.

#### 4.1.2 Design

In this experiment, the stretch scale (0%, 4%, 8%, 12%, 16%, 20%, and 24%) and stretch direction (horizontal and vertical) are two within-participants factors, image observation time (3 seconds, 6 seconds, 12 seconds) is a between-participants factor.

To determine the observation times, we conducted a pilot study. We invited 6 participants to take part in the study and asked them to complete the tasks without time constraints. Our participants took 6 seconds, on average, to examine the images. We therefore chose 3 (half), 6, and 12 (double) seconds for our trials.

Each participant was assigned an observation time. Each participant used the same observation time for all images.

We chose to make the stretch scale get progressively harder by dividing the experiment into three parts (participants did all three parts in one session). The easy part used 20% and 24% in both horizontal and vertical directions, the middle used 12% and 16% and the hard part used 4% and 8%.

We used a counterbalance design to spread the images amongst the different conditions uniformly. The 48 images were randomly assigned to 3 image groups, each with 8 pairs and got 24 pairs totally. A  $3 \times 3$  Latin square design [Box et al. 2005] was used to assign image groups to experiment parts. Pairs of participants were assigned to each square to balance which images were stretched. The stretched and unstretched images were randomly assigned left/right.

We take participant group 1 and image group 1 as an example to show how to counterbalance at the easy session. Firstly, we randomly label the 16 images as  $I_1, I_2, \dots, I_{16}$ . Then we compose 8 pairs of images using the composition shown in Table 2. We take  $I_1$  and  $I_2$  as an example, when this pair of images is shown to the first participant,  $PG_{11}$ ,  $I_1$  is not stretched and  $I_2$  is stretched 20% horizontally. But when this pair is shown to the second participant,  $PG_{12}$ , we use the unstretched image  $I_1$  as the stretched one. We also change the stretch scale and stretch direction of  $I_2$  for  $PG_{11}$ , so we stretch  $I_2$  24% vertically. And  $I_2$  is not stretched for  $PG_{12}$ . The counterbalance method is similar to the other pairs as shown in

	0 v.s. $H_{20}$	0 v.s. $H_{24}$	0 v.s. $V_{20}$	0 v.s. $V_{24}$
$PG_{11}$	$I_1$ v.s. $I_2$	$I_3$ v.s. $I_4$	$I_5$ v.s. $I_6$	$I_7$ v.s. $I_8$
	$I_9$ v.s. $I_{10}$	$I_{11}$ v.s. $I_{12}$	$I_{13}$ v.s. $I_{14}$	$I_{15}$ v.s. $I_{16}$
$PG_{12}$	$I_{16}$ v.s. $I_{15}$	$I_{14}$ v.s. $I_{13}$	$I_{12}$ v.s. $I_{11}$	$I_{10}$ v.s. $I_9$
	$I_8$ v.s. $I_7$	$I_6$ v.s. $I_5$	$I_4$ v.s. $I_3$	$I_2$ v.s. $I_1$

**Table 2:** Counterbalance for participant group 1 and image group 1 at easy part.  $PG_{11}$  and  $PG_{12}$  are the first and second participant in participant group 1.  $I_i$  is the  $i$ th image in image group, here is image group 1.

Table 2. We show an example of these 8 pairs used at easy part for participant  $PG_{11}$  in Figure 7.

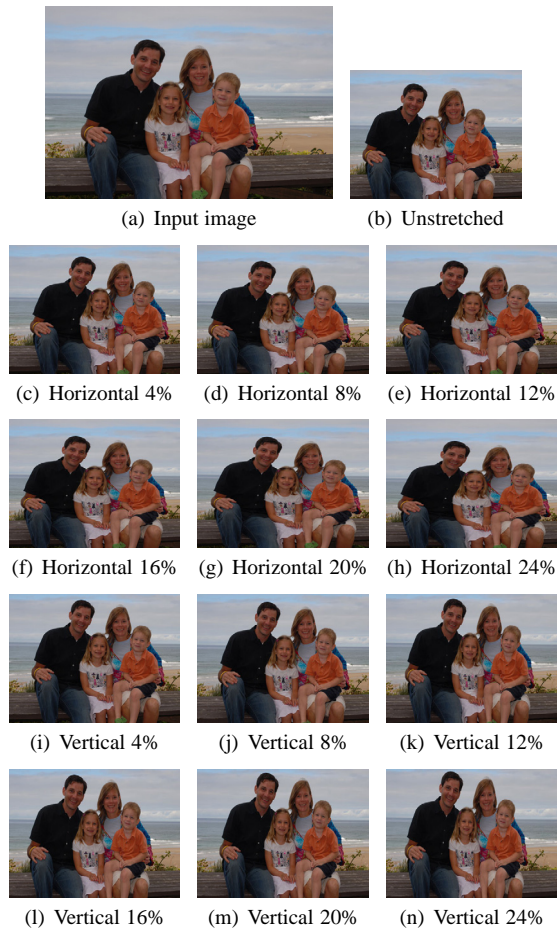
#### 4.1.3 Participants

We recruited 36 participants (20 males and 16 females, from 20 to 30 years old, average age of 25.3) in China. All of the participants are either undergraduate or graduate students and voluntarily participated in our study. After the study, we excluded the data from 5 participants because they either did not complete all of the images or gave unreliable answers. Among the remaining 31 participants, 10 were assigned 3 seconds observation times, 11 were assigned 6 seconds, and 10 were assigned 12 seconds.

#### 4.1.4 Procedure

On the introductory page, we inform participants the task motivation and description. We give an example of an unstretched image, its horizontally stretched version, and vertically stretched version. We also show two pairs of images. One pair consists of an unstretched image and a horizontally stretched one, the other pair is consisted of a unstretched image and a vertically stretched one. All of the images used on the introductory page are different from the 48 test images. After the participant decides to participate, they are presented with the easy part, then the middle part and then the hard part. At the end of the experiment, the participant is given a score based on their performance and was shown the pairs for which they gave incorrect answers. On average, it took participants about fifteen minutes to complete this study.

During each part, we show 8 pairs of images. For each pair of images, we first show a page displaying the positions of the two images, the question, the options to the question and ask the participant to get ready. The participant clicks a button to show the pair of images. The images are shown for the observation time and then blanked. The page also asks the participant to answer which image is stretched by selecting the radio button “The left image is stretched.” or “The right image is stretched.” The participant can



**Figure 8:** An input image and its 13 cropped and stretched results.

take a break and continue on to next pair when she/he is ready.

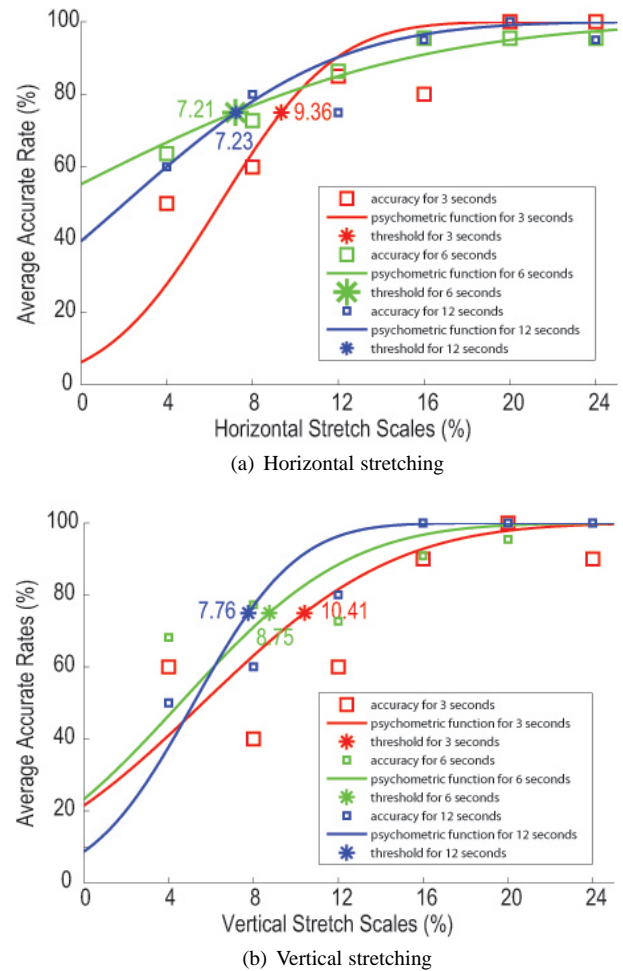
## 4.2 Results

The study in this phase shows that the observation time affects people’s perception of image stretch. We performed a mixed-model ANOVA with stretch scale and stretch direction as within-participants factors and observation time as a between-participants factor. We found two significant main effects of stretch scale ( $F(5, 140) = 26.34, p < 0.001$ ) and observation time ( $F(2, 28) = 3.71, p = 0.037$ ). No significant effect of stretch direction or significant interaction effects were identified.

Our data show that participants perform better when they are given longer time. Post-hoc pairwise tests show that all the increases in overall accuracy are significant. The increase from 3 to 6 seconds is most significant, improving from 76% to 83% ( $p < 0.001$  in post-hoc test). The increase from accuracy from 6 seconds to 12 seconds is smaller, but still statistically significant ( $p = 0.015$  in post-hoc test).

### 4.2.1 Thresholds

Our original goal for this experiment was to determine the threshold for detection of stretching, that is the amount of stretch required to expect a viewer’s performance to be better than chance. Because performance depended on other factors besides amount of stretch,



**Figure 9:** Psychometric functions and thresholds for three observation times. Red, green, and blue are used for 3 seconds, 6 seconds, and 12 seconds respectively. Squares stand for accurate rates, and stars stand for thresholds.

we would expect different thresholds for different conditions (e.g. observation times). Qualitatively, the thresholds for all three observation times seem to fall in the range of our experiments: the hard scales (4%) showed performance not too different from chance, while participants had close to perfect performance on the easier trials.

To find a more precise threshold, we use the definition of threshold given by [Ferwerda 2008]: the amount of stimulus that provides performance halfway between perfect performance (100%) and the performance that would be expected from just guessing (50%, called point of subjective equality). To compute this, we use the process from [Ferwerda 2008] where the accuracies are converted to Z-values (the inverse of the cumulative normal distribution function), a line is fit, the line is transformed back to a curve in accuracy, and the 75% point is found. We computed thresholds for both horizontal and vertical stretching, as shown in Figure 9.

The computed threshold curves fit the data poorly. We believe that this is a symptom of insufficient data: we did not have enough samples for any particular conditions to accurately determine thresholds. The expected trends, such as higher thresholds for shorter observations, appear.

## 5 Discussion and Future Work

In this paper, we presented an initial study on people’s perception of image stretch. We find that people’s perception of image stretch is a complex phenomenon. It depends on a variety of factors, such as amount of distortion, image content, the viewer’s cultural background, and the observation time. Stretch direction was not a significant effect. We have shown that it is possible to use an online worker community for performing these experiments. This will be valuable as it appears that it will take significant experimentation to determine thresholds.

Our results have implications for designing image retargeting methods. Firstly, almost all existing retargeting approaches use saliency as a predictor of distortion, such as stretching, is acceptable. Our findings suggest that this is insufficient: many of the images for which stretch is difficult to notice contain salient objects. Salient objects exist in images of all categories, yet sensitivity varies across the categories. Higher level information, such as content category, may serve as a more useful importance metric. Secondly, since scaling methods, warp-based methods, and multi-operator methods will introduce stretching to result images, our thresholds should be integrated into these methods to control the stretching. Also, while faces are often identified to be preserved by retargeting methods, our results show that more generally preserving “people” would be valuable.

Our study shows the challenges in evaluating image retargeting methods. As the perception of distortion depends on many factors, the acceptance of retargeting results will have similar dependencies. Ideally studies would consider these factors, using appropriately diverse images and subjects, and controlling for observation time. Comparison of studies is difficult because of the numerous factors, for example, most studies use homogeneous (but different) subject pools.

Our initial experiments are a first step towards understanding the perception of resized images. Important extensions to this work include: performing threshold studies on more people, images, and conditions to map out the limits of distortion detection; exploring distortions beyond stretching such as various content-aware resizing methods; and exploring questions beyond detectability, for example to determine whether distortions are objectionable or misleading.

### Acknowledgements

We would like to thank the Flickr members who have kindly made their images which are used in this paper available under a Creative Commons license. All of the images used in this paper are reproduced under Creative Commons license <http://creativecommons.org/licenses/by-sa/2.0/> or <http://creativecommons.org/licenses/by/2.0/>. Table 3 gives all the links to the photographs. Limited by the space, we omit <http://www.flickr.com/photos/> for the following links. This work was supported in part by NSF grant IIS-0416284.

### References

AMAZON. Amazon mechanical turk, <https://www.mturk.com/mturk/>.

AVIDAN, S., AND SHAMIR, A. 2007. Seam carving for content-aware image resizing. *ACM Trans. Graph.* 26, 3, 267–276.

BOX, G. E. P., HUNTER, J. S., AND HUNTER, W. G. 2005. *Statistics for experimenters: design, innovation, and discovery*. Wiley-Interscience.

Photographs	Links
Figure 1 (a)~(c)	panayotis/2997274618
Figure 1 (d)~(f)	pagedooley/2429429397
Figure 2	4gott/2336126105
Figure 5	eldelinux/3525659943
Figure 7 (a)	shgmom56/3395721404, sis/121704669
Figure 7 (b)	meanestindian/254069853, thedza/3093115337
Figure 7 (c)	josefstuefer/8925835, dlkinney/56540664
Figure 7 (d)	ulikleafar/2714953737, rolandslakis/97536203
Figure 7 (e)	yourdon/3088654396, jurvetson/3217544578
Figure 7 (f)	mikebaird/2913345328, flattop341/795318584
Figure 7 (g)	yourdon/3275748024, yourdon/3140256151
Figure 7 (h)	jurvetson/3212694835, yourdon/3087716759
Figure 8	spolyak/1031569673

**Table 3:** The links to the photographs used in the paper.

FERWERDA, J. A. 2008. Psychophysics 101: how to run perception experiments in computer graphics. In *SIGGRAPH '08: ACM SIGGRAPH 2008 classes*, 1–60.

KITTUR, A., CHI, E. H., AND SUH, B. 2008. Crowdsourcing user studies with mechanical turk. In *CHI '08: Proceeding of the twenty-sixth annual SIGCHI conference on Human factors in computing systems*, ACM, New York, NY, USA, 453–456.

KRAGES, B. P. 2005. *Photography: the art of composition*. Allworth Press.

KRÄHENBÜHL, P., LANG, M., HORNUNG, A., AND GROSS, M. 2009. A system for retargeting of streaming video. *ACM Trans. Graph.* 28, 5, 1–10.

MACMILLAN, N. A., AND CREELMAN, C. D. 1991. *Detection theory: a user’s guide*. Cambridge University Press.

MOJSILOVIĆ, A., AND ROGOWITZ, B. 2004. Semantic metric for image library exploration. *IEEE Trans. on Multimedia* 6, 6, 828–838.

NIU, Y., LIU, F., AND GLEICHER, M. Multimedia retargeting bibliography. <http://sites.google.com/site/mediaretargeting/>.

NIU, Y., LIU, F., LI, X., AND GLEICHER, M. 2010. The complexity of perception of image distortion: an initial study. In *CHIEA '10: Proceedings of the 28th of the international conference extended abstracts on Human factors in computing systems*, 3235–3240.

SUH, B., LING, H., BEDERSON, B. B., AND JACOBS, D. W. 2003. Automatic thumbnail cropping and its effectiveness. In *UIST '03: Proceedings of the 16th annual ACM symposium on User Interface Software and Technology*, 95–104.

TORRALBA, A., AND OLIVA, A. 2003. Statistics of natural image categories. In *Network: Computation in Neural Systems*, 391–412.

WANG, Y.-S., TAI, C.-L., SORKINE, O., AND LEE, T.-Y. 2008. Optimized scale-and-stretch for image resizing. *ACM Trans. Graph.* 27, 5.

WOLF, L., GUTTMANN, M., AND COHEN-OR, D. 2007. Non-homogeneous content-driven video-retargeting. In *ICCV '07: Proceedings of the Eleventh IEEE International Conference on Computer Vision*, 1–6.