Using Eye-Tracking to Assess Different Image Retargeting Methods

Susana Castillo
Universidad de Zaragoza

Tilke Judd
MIT CSAIL

Diego Gutierrez
Universidad de Zaragoza

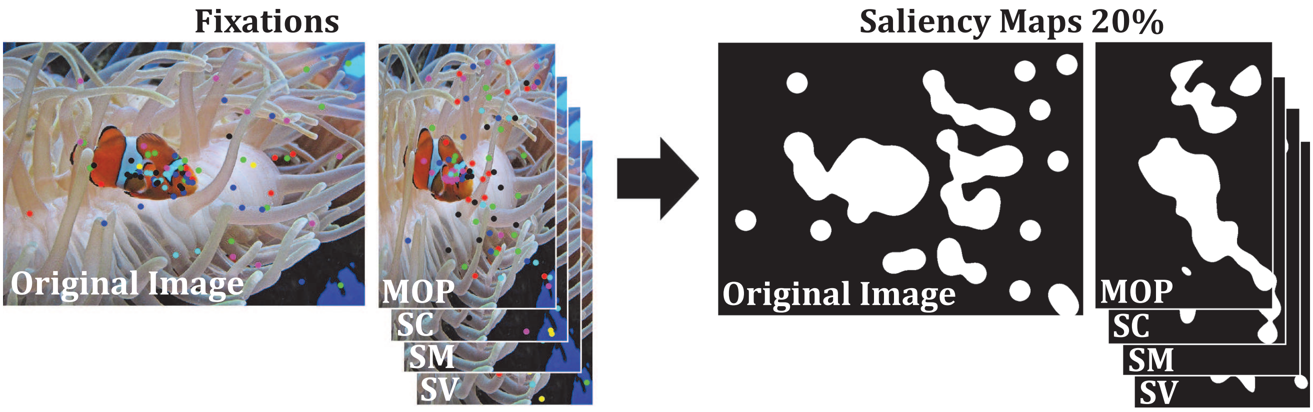

Figure 1: *In this work, we investigate the impact of the retargeting process on fixation patterns, under the assumption that viewers' fixations should not change when viewing the source image and when viewing the retargeted results. We gather the users' fixations on the images and we derive their saliency maps from these eye tracking data. We compare the saliency maps of the retargeted images to the saliency maps of the original size images with six different automatic image similarity metrics. The differences between all these saliency maps and the ones gathered by using a model of prediction of human fixations are evaluated as well, in order to analyze if this model is able to match actual eye movements in a retargeting context.*

Abstract

Assessing media retargeting results is not a trivial issue. When resizing one image to a particular percentage of its original size, some content has to be removed, which may affect the image's original meaning and/or composition. We examine the impact of the retargeting process on human fixations, by gathering eye-tracking data for a representative benchmark of retargeted images. We compute their derived saliency maps as input to a set of computational image distance metrics. When analyzing the fixations, we found that even strong artifacts may go unnoticed for areas outside the original regions of interest. We also note that the most important alterations in semantics are due to content removal. Since using an eye tracker is not always a feasible option, we additionally show how an existing model of prediction of human fixations also works sufficiently well in a retargeting context.

Keywords: media retargeting, eye tracking, image semantics

1 Introduction

When taking a picture, the photographer's intention is usually to produce an aesthetically pleasing result, or to convey a message. In some cases, the composition of the image is not casual, but seeks to transmit a particular vision of the scene and to evoke an emotional response. For this purpose, artists often rely on well-known photographic composition guidelines such as rule of thirds, diagonal dominance, visual balance or size region [Liu et al. 2010]. Due to the variety of display devices on the market and the various sizes of screens, such images may not always be viewed with their original aspect ratio. To solve this problem, there have been numerous papers on content-aware media. In this work, we focus on retargeting algorithms, aimed at smartly reducing image size. We use a subset from the *RetargetMe*[1] benchmark [Rubinstein et al. 2010] which includes a collection of images specially difficult for retargeting.

Our approach is based on analyzing saliency and fixations points both in the original and the retargeted images. Given that the environment usually contains much more information than we can process at once, visual attention mechanisms cover the necessary task of selecting the most important parts of the scene shown. The human visual system seems to be very well adapted for extracting structural information; when seeing a scene, humans first recognize its meaning and the global image features (the "gist" of the scene) and infer objects. It is just then when geometric, contextual and semantic information is progressively added [Chen 2005], [Oliva and Torralba 2001], [Oliva 2009], [Greene and Oliva 2006], [Greene and Oliva 2009], [Ehinger and Oliva 2011].

Using the gist of the scene, it is possible to detect the presence or absence of objects, as well as their locations, without running an object detector [Torralba 2003]. But, if artifacts are present, results may change [Choi et al. 2010]. These artifacts may attract the viewers' attention, as they are unexpected and may surprise them [Itti and Baldi 2009]. Thus, there is a strong interest in predicting the location of the areas where observers will focus their attention [Judd et al. 2009], since the impact of an artifact should in principle depend on the salience of image region where it appears [O. Le Meur 2009]. In fact, some retargeting methods do use a prediction of the original image saliency to guide the process.

Removing parts of the image may also substantially alter the viewers' regions of interest (RoI). In order to measure how these regions of interest of the image are affected, eye-tracking experiments can be useful [Chamaret and Le Meur 2008], [Chamaret et al. 2010]. Rubinstein et al. [2010] presented the first evaluation of retargeting algorithms, using both subjective and objective metrics. A metric to appraise quantitatively the quality of a retargeting ap-

[1] http://people.csail.mit.edu/mrub/retargetme/

proach in video context has also been recently proposed [Chamaret et al. 2010], which takes into account the preservation of visually important areas, by computing the proportion of visual fixations that fall into the cropping window. Another relevant work presents a new objective metric based on the human visual system [Liu et al. 2011]. In the same vein, we suggest a new approximation to gauge the quality of different image retargeting methods, taking into account the semantics of the image by using saliency maps from both eye tracking data and a model of saliency. Our overall goal is to check whether or not the RoI's were changed in the final result, under the assumption that a retargeting algorithm should not alter them.

We hypothesize that there are two main reasons why people change their viewing patterns when looking at an image:

- Distraction due to *artifacts* in the image.
- Alteration of the image *semantics*.

The retargeting process should add minimal alterations to these viewing patterns. Our work helps to show to what extent these two aspects affect the user perception of the image.

The main contributions of this work are as follows. First, we validate the applicability of an existing model that aims at predicting where humans look [Judd et al. 2009] to a retargeting context. Second, we propose an eye-tracking framework to analyze how image similarity measures correlate with users' RoIs. Third, we analyze the influence of the changes due to retargeting in the semantics of the image, trying to understand what is more important to viewers, preserving content or the lack of artifacts.

2 Previous Work

Image retargeting is the concept of resizing an image while preserving its important content. Simple ways to retarget an image include using a simple scaling operator or manually choosing a cropping window. Researchers have also introduced numerous methods to retarget images which can be classified as discrete or continuous [Shamir and Sorkine 2009]. Discrete approaches judiciously remove or insert pixels (or patches) to preserve content, while continuous solutions optimize a mapping (warp) from the source media size to the target size, constrained on its important regions and permissible deformations. Some of the most recent methods of retargeting include Seam-Carving (SC) [Rubinstein et al. 2008], Multi-operator (MOP) [Rubinstein et al. 2009], Shift-maps (SM) [Pritch et al. 2009], Streaming Video (SV) [Krähenbühl et al. 2009], Non-homogeneous Warping [Wolf et al. 2007], Scale-and-Stretch [Yu-Shuen Wang and Lee 2008], Energy-based Deformation [Karni et al. 2009] and Motion-based Video [Wang et al. 2010]. The latest retargeting approximations include improvements seeking a better understanding of the image semantics and visual attention, such as the inclusion of symmetry (like in the symmetry-summarization algorithm [Wu et al. 2010]) or the usage of the original image to guide the process, as it is done in Importance Filtering [Ding et al. 2011].

Since there are numerous retargeting methods, it is hard to know which one works best. Rubinstein et al. [2010] used a methodological approach to evaluate retargeting results and presented the first comprehensive perceptual study and analysis of image retargeting. They created a public available benchmark of images (*RetargetMe*) and conducted a large user study to compare a number of state-of-the-art methods. In their experiment, participants were shown two retargeted images at a time, side by side, and were asked to choose the one they like better. This was iterated across retargeting methods and images to discover an overall ranking of the retargeting methods based on user preferences. While Rubinstein et al.'s work was seminal in comparing image retargeting methods based on viewer's preferences, our aim is to more precisely measure how retargeting methods affect the way people observe images. We use a subset of images in the *RetargetMe* benchmark database and measure directly how viewers free-view the images through an eye tracking experiment.

Since the experiments of Buswell [1935] and Yarbus [1967], eye tracking has long been used to measure how people look at an image. More specifically, eye tracking measures the order and location of viewers' fixations (when the eye is stabilized at a particular point) and saccades (or discrete, rapid movements between fixations). When people fixate, they move the high resolution part of their fovea to a specific location to see it with high detail. Away from this point of gaze, the spatial resolution of the human visual system decreases dramatically so that the periphery is seen in low resolution [Anstis 1998]. Eye movements correlate with shifts in visual attention and are thought to be a consequence of optimal resource allocation of high-level tasks such as visual scene and object recognition. Often, fixations are directed to locations where viewers expect to see objects of interest [Loftus and Mackworth 1978]; [Biederman et al. 1982]; [De Graef et al. 1990]; [Henderson et al. 2006].

Chamaret et al. [2010] used an extension to the temporal dimension of a cropping algorithm of Le Meur et al. [2006]. They stated that if human fixations are known, counting the number of fixations that fall inside the retargeted sequence is easy; however, when the image is warped, it is not as easy to tell when the area of a fixation is preserved or not. How can a retargeted image be assessed when it is warped and the correlation between the fixations before and after retargeting are not as clear? This is the issue we aim to investigate.

A main objective for retargeting results is preserving the recognizability of important original media features, which is a subjective measure. Analysis of the reasons for discarding the results of a method in a pair-wise choice has shown some reasons why humans reject a retargeting result: squeezing people or faces, distortion over geometric structures and changes in proportions in the image [Rubinstein et al. 2010]. When analyzing participants' answers in subjective tests, it has also been suggested that viewers prefer the loss of content over deformation artifacts, always in a comparative framework. Moreover, in a normal context, the viewer only sees the retargeted image with no reference to the original one. In this case, the loss of content is obviously not identified. In the absence of a reference image to directly compare against, it is more difficult to recognize artifacts or to know where the information was lost.

Our work hypothesis is that human eye movements give strong evidence about the location of meaningful content in an image. When an image is retargeted, this meaningful content should be preserved without introducing arbitrary artifacts that draw attention away from important content. In our experiments, we track the observers' eye movements in a free-viewing task on retargeted images. This allows us to analyze if and how retargeting methods affect viewers' fixations. The eye tracking database on these images helps to evaluate current methods and might guide the creation of new retargeting algorithms in the future. We have made this material publicly available to the research community[2].

3 Benchmark

In the following we describe the choices for images, retargeting methods and the eye-tracking procedure to gather the saliency maps highlighting the most attractive regions.

[2]http://giga.cps.unizar.es/ diegog/projects/APGV11/apgv11.html

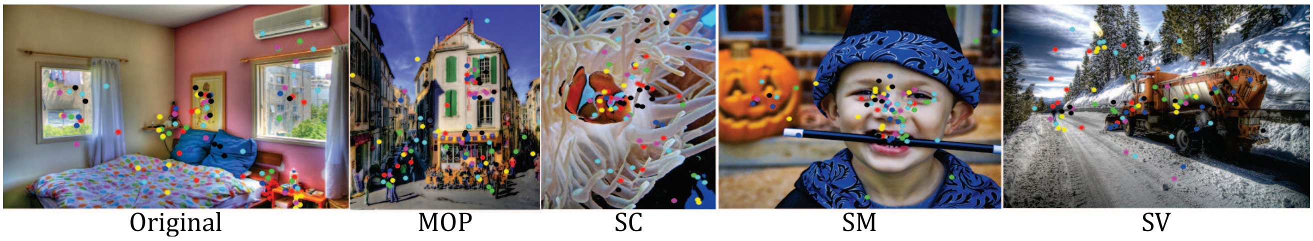

Figure 2: *A sample of the fixation locations for seven viewers on some of the images (one color for each user). We collected eye-tracking data on* 155 *images from* 35 *observers to use as ground truth data to analyze possible changes in regions of interest due to retargeting.*

3.1 Selecting Retargeting Methods and Images

In our study we use retargeted versions of 31 original images from the *RetargetMe* benchmark [Rubinstein et al. 2010]. This benchmark is a complete group of images which are specially challenging for retargeting: the amount of retargeting applied is considerable (25% or 50%) and the images contain attributes which are considered to be challenging for preserving content and structure and preventing artifacts during resizing. We got a representative set from the original database that covers its range of image attributes. The images we gathered also contain attributes which are considered attractive to humans [Judd et al. 2009]. These attributes are (the number in parentheses indicate the number of original images belonging to each set): lines and/or clear edges (23), evident foreground objects (14), texture elements or repeating patterns (7) , specific geometric structures (16), symmetry (5), people and faces (11), animals (3), vehicles (7) and textual elements (6).

Each image was shown in its original form and in four retargeted versions, yielding a total of 155 images. The four different retargeting methods were: Seam-Carving (SC) [Rubinstein et al. 2008], Shift-maps (SM) [Pritch et al. 2009], Multi-operator (MOP) [Rubinstein et al. 2009] and Streaming Video (SV) [Krähenbühl et al. 2009].

Two of these content aware methods we study, SC and SM, are classified as discrete and the other two ones, MOP and SV as continuous. The first one removes seams of pixels crossing the regions of least importance and the second one removes entire objects at a time. Both of the continuous approaches, MOP and SV, allow uniform scaling of image content. We chose these methods because of their user rankings in the study by Rubinstein and colleagues [2010]. Since we wanted them to be statistically distinguishable in terms of user preference, we chose two of the top ranked ones (SV, MOP), one middle-ranked (SM) and the lowest ranked (SC).

3.2 Participants

A total of 35 observers (20 males, 15 females, age range $18 - 40$) participated in our eye tracking study. Each reported normal or corrected-to-normal vision. They all signed a consent form and were paid $15 for their time. Each user saw a 31-image subset of the 155 images (31 original size images and the four retargeted versions of each one) and never saw the same image under different retargeting conditions. We distributed the images so that exactly seven users viewed each of the 155 images.

3.3 Procedure

All the viewers sat approximately 24 inches from a 19−inch computer screen of resolution 1280x1024 px in a semi-dark room and used a chin rest to stabilize their head. A table-mounted, video-based ETL 400 ISCAN eye tracker recorded their gaze path at 240 Hz as they viewed each image for five seconds. We used a five point calibration system, during which the coordinates of the pupil and corneal reflection were recorded for positions in the center and each corner of the screen. The average calibration error was less than one degree of visual angle (∼35pixels). During the experiment, position data was transmitted from the eye tracking computer to the presentation computer so as to ensure that the observer fixated on a cross in the center of a gray screen for 500 ms prior to the presentation of the next image. The observers were instructed simply to "look carefully at the images" for this free viewing task.

The raw data from the eye tracker consisted of time and position values for each data sample. We use the method from Torralba et al. [2006] to define saccades by a combination of velocity and distance criteria. Eye movements smaller than the predetermined criteria were considered drift within a fixation. Individual fixation durations were computed as elapsed time between saccades and the position of each fixation was computed from the average position of each data point within the fixation. We discarded the first fixation from each scanpath to avoid the trivial information from the initial fixation in the center. Figure 2 shows the fixation locations for seven different observers on some of the images used (one color for each user).

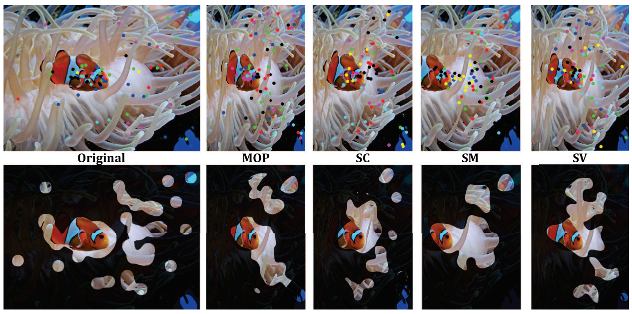

Figure 3: *The fish image and its four half sized retargeting results. Top row shows users' fixations on the image. Bottom row shows the overlap between the source images and the saliency maps we derive from these fixations.*

Once we gather all the eye-tracking data for our benchmark of images, we derive their corresponding continuous saliency maps, as done by Judd et al. [2009]. We convolve a Gaussian low pass filter (with circular boundary conditions) across the users' fixations. We get the binary maps of the top 20% salient locations of the images by thresholding these continuous saliency maps. We analyze these saliency maps of the average locations fixated by users with six different computational distance measures. We find out that the correspondence of these metrics classification to the subjective preferences is improved by using these saliency maps in place of the images themselves. See Figure 3: top row shows the fixation locations for seven different observers on the *fish* source images (the original size one and its four retargeted versions). Bottom row shows the overlap between the source images and their computed saliency maps from these fixations. When using an eye tracker is not a feasible option, using a predictive model of saliency is a desirable option. We carry out a performance analysis of one of these models, the one proposed by Judd et al. [2009].

4 Analysis of Distance Measures

The main goal in this section of our work is to gain understanding about the impact of retargeting images on human fixations, by comparing saliency maps before and after retargeting. To do so, we carry out an analysis using six of the computational distance measures from the *RetargetMe* benchmark [Rubinstein et al. 2010]: Bidirectional Similarity (BDS) [Simakov et al. 2008], Bidirectional Similarity PatchMatch (BDS-PM) [Barnes et al. 2009], SIFT Flow (SF) [Liu et al. 2008], Earth Mover's Distance (EMD) [Pele and Werman 2009] and, from the MPEG-7 standard [MPEG-7 2002]; [Manjunath et al. 2001], Edge Histogram Descriptor (EH) [Manjunath et al. 2001] and Color Layout (CL) [Katusani and Yamada 2001].

We run these six image similarity measures on the saliency maps obtained, comparing the saliency map of each retargeted image to the saliency map of its corresponding original image (see Figure 3, bottom row), and we collect the resulting distances. We use the implementation and parameter optimization proposed by Rubinstein et al. [2010].

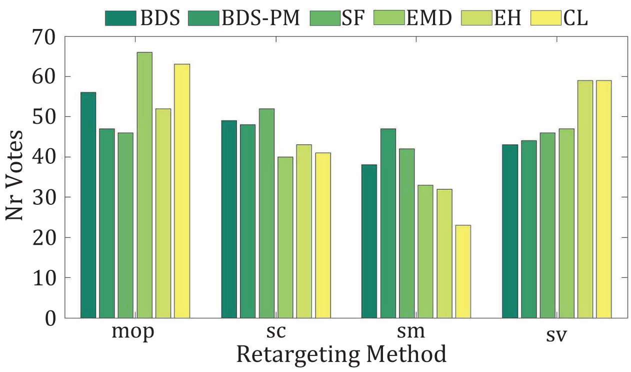

Figure 4: *All metrics' votes. Each bar represents the number of times the saliency map for a retargeted image was favored by the metric over a different retarget of the same image, collected over all images. The figure clearly shows the bias of EMD, EH and CL to MOP and SV results.*

Metric Bias. We use a voting scheme in order to examine if any given metric is biased towards an image retargeting method, or a set of methods.

For this purpose, we define $\theta =< \theta_1, ..., \theta_t >$, for $t = 4$ retargeting methods, as the objective distance vector, for a saliency map ψ, calculated by one of the objective measures. We denote by ψ_o the saliency map for the original version of the image obtained from the eye-tracking data, while ψ_i represents the saliency map after retargeting method i. Let D be a given objective measure, then $\theta_i = D(\psi_o, \psi_i)$ is the distance between ψ_o and ψ_i, according to metric D. In our context, the lower the θ_i, the better the method i is.

First, the objective vector θ is sorted in ascending order. Then we consider all pairs (i, j) in the ranking and vote for result i if $rank_{asc}(\theta_i) < rank_{asc}(\theta_j)$ and vote for j otherwise. Accumulating the votes over all images for each metric, we get an indication as to the number of times the metric has favored one result over another. Figure 4 shows the distribution of these votes among the operators for the metrics we tested.

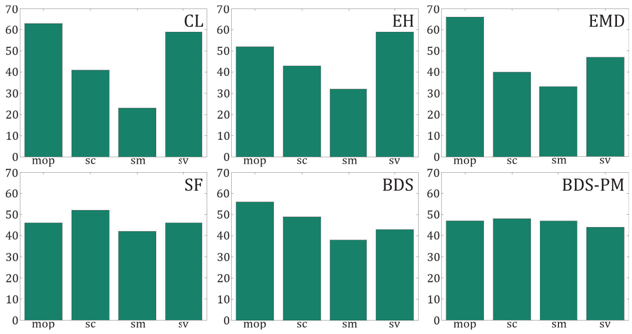

Figure 5: *Distribution of metric votes among the operators for each one of the six different metrics we tested.*

Ranking. For each one of the six metrics, Figure 5 shows the rankings for the four retargeting methods over all images. The methods are ranked by the total number of times the yielded distance for a saliency map of one retargeting method's result was smaller than the one calculated for a saliency map obtained with a different method. Similar to previous works [Gutierrez et al. 2008], [Rubinstein et al. 2010], we want to group the retargeting algorithms according to whether or not their results are statistically indistinguishable. For this purpose we develop a significance test of the score results. In order to establish if any two given retargeting methods belong to the same group or to different groups, we have to know if the differences between the obtained rankings are higher or lower, respectively, than a fixed R' [Setyawan and Lagendijk 2004]. For a given significance level α, R' should satisfy: $P(R \geq R') \leq \alpha$. David [1963] proved that this R' value can be computed by solving:

$$P\left(W_{t,\alpha} \geq (2R' - 0,5)/\sqrt{nt}\right) \quad (1)$$

where n is the number of original images and t is the number of retargeting methods. In our case, $t = 4$ and $n = 31$. $W_{t,\alpha}$ can be obtained by interpolation from published tabled data [Pearson and Hartley 1966]; we set $\alpha = 0.01$, and consequently obtain $W_{4,0.01} = 4.405$. The resulting value for R' is 24.7760, so we set $R' = 25$.

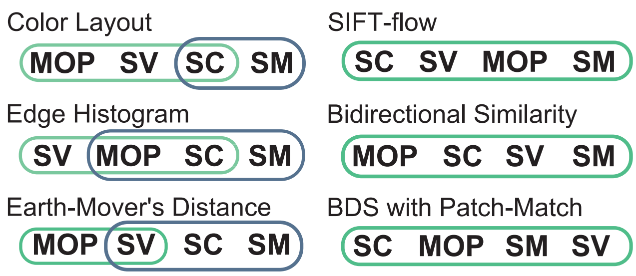

Figure 6: *Score grouping for each computational measure. Operators are descending ordered according to metric votes. Operators within a group are statistically indistinguishable in order of metric distance.*

Figure 6 shows the resulting groups for each metric. These results show that, for three of the metrics used, two algorithms (SV and MOP) usually perform better than the rest (see Figure 5), yielding statistically indistinguishable results. Another algorithm (SM) is consistently graded the lowest, also yielding results that can be considered similar in terms of ranking. Our classification seems to support the ranking classification for subjective analysis from Rubinstein et al. [2010], where SV and MOP are preferred over the rest of methods. This suggests that using the objective distance methods over saliency maps from human fixations might be a useful strategy for designing and assessing the quality of retargeting approaches.

5 Performance Analysis of a Predictive Model of Saliency

As said in Judd et al. [2009], people tend to fixate on text and faces. In their absence, they focus on animals (particularly on their faces). If none of these common objects are present in the image, human fixations are biased towards the center. According to this, the authors developed a model (SVM_{MIT}) that uses low level features (local energy of the steerable pyramid filters; intensity, orientation and color contrast; values of RGB channels and their probabilities), mid level features (horizon line detector) and two high level ones: face detector and car and person detector, as well as a feature for center prior [Judd et al. 2009].

We want to analyze if the SVM_{MIT} model is able to predict actual eye movements in a retargeting context. When resizing one image, if only the less salient parts of it are removed or distorted (as the four methods we use are supposed to do), then fixations and saliency areas should be preserved. We want to analyze how well this model works on retargeted images, which is a different context from the original work by the authors. In order to do this, we calculate the difference between the real saliency maps from eye-tracking data and the ones predicted by the SVM_{MIT} model(see Figure 7).

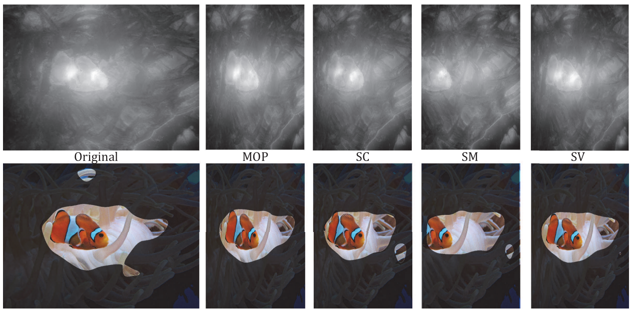

Figure 7: *Saliency maps. Top row shows the saliency maps predicted by the SVM_{MIT} model. Bottom row shows these maps (thresholded to show the top 20% salient locations) and their source images overlapped.*

If I denotes a given image from our dataset, $n = 31$ is the number of original images, V denotes one of the $v = 5$ available versions of the image (the original one and its four results for the four retargeting methods), ψ^M is the saliency map obtained with the SVM_{MIT} prediction model and ψ^E is the saliency map gathered directly from eye-tracking data, then we obtain the saliency maps for each original image and its four retargeted versions (I_k, V_l), $k = 1..n, l = 1..v$. After obtaining these saliency maps, we calculate the absolute difference for each pair $(\psi^M(I_k, V_l), \psi^E(I_k, V_l))$. All saliency maps have been thresholded to show the most salient 20% of the image. On average, SVM_{MIT} reaches more than the 80% of the way to human performance. Interestingly, it performs equally well ($\sim$82%) on retargeted versions of the image as on original pictures. Although these results are not conclusive and a much extensive analysis should be performed, viewers seem to keep fixating consistently on the same high level features, before and after retargeting: text (see Figure 2, MOP), people (specifically faces) (see Figure 2, SM), animals (specifically faces) (see Figure 2, SC), vehicles (see Figure 2, SV) and, in general, they tend towards the center of the image (see Figure 2).

Such a good performance of the SVM_{MIT} model on the retargeted images may allow future retargeting algorithms to guide their design and assess their results without the need for neither collecting eye tracking data nor performing user studies.

6 Analysis and Discussion of Artifacts.

The ranking process we carry out shows a suboptimal performance of the SC and SM operators, while it favors MOP and SV. These tendencies correlate with the viewers' preferences ranking obtained by Rubinstein et al. [2010]. The interest of a given object in an image can be affected by changes in its size, blurring level, contrast and relative position to the center of the image [V. Kadiyala and Chandler 2008]. It may be that, since MOP and SV include an overall scaling factor, the relative distances and proportions of the objects in the image correlate better with the original version and so do their corresponding saliency maps.

Nevertheless, our analysis inverts the positions for the two lowest ranked algorithms. SM which was an acceptable, middle ranked option for retargeting (according to Rubinstein et al.'s [2010] analysis of subjective preferences), becomes the lowest rating. We think this may be due to the very different nature of the artifacts that these two methods are susceptible to. On the one hand, SC removes monotonic and connected seams along the image and, thereby, its results are more likely to be affected by distortion of lines, edges and geometry, as well as by deformation of people and objects. On the other hand, the SM operator reduces the image size by removing entire objects, which may be important for users to preserve. According to our results, although distortion and deformation might be less aesthetically pleasing than cut-off content, they draw less attention away from original important content.

Looking at the fixation points on the SM results (Figure 8), it can be seen that if the removed object lies in one RoI of the original image, people turn their attention towards other regions of the image, substantially changing the saliency map of the retargeting result and altering the image semantics. Figure 8, *Umdan SM* result is a good example, several people and a dog have been removed and the attention of viewers is logically driven towards the faces of the remaining people. As a consequence, the corresponding saliency map is less scattered than the original one. Another example is Figure 8, *Marblehead Mass SM* result, where some of the houses have been removed. In this case, the relative sizes of the remaining houses have been increased and the purple facade receives more interest. More interesting results are found when the cut-off content is not a proper object by itself, but part of one, or when this removal involves extracting objects that people are aware they should be there; in those cases, fixations accumulate in that region of the image, as supported by Itti et al. surprise theory [2009] (see Figure 9, *Car SM* result, the car's back wheel is missing). On the other hand, fixation points on SC results show that, when distortions affect RoIs, viewers focus their attention on them (see Figure 9, *Brasserie L'Aficion SC*, where the text in the front of the building is absolutely distorted).

It is interesting to note that, for all of the retargeting methods in our five seconds free viewing task, fixations do not change due to the presence of artifacts if they do not fall on the original image's RoI. This suggests that even in the presence of large errors, five

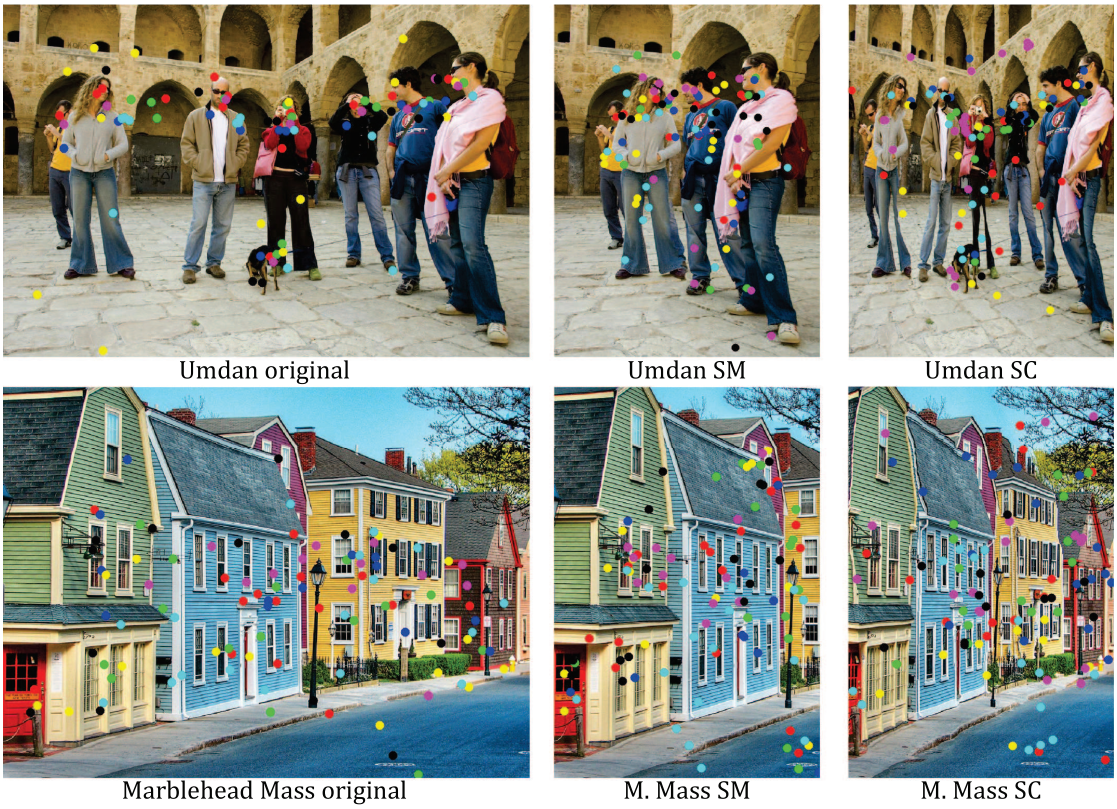

Figure 8: *How retargeting methods affect the way people observe images? This figure shows several examples of the way the fixation points change on the original versions of the images and the two lowest ranked retargeting methods (according to metric distances).*

Figure 9: *Examples of how fixation points change or keep the same according to the nature of the artifacts introduced by the retargeting methods. When an artifact is not in a RoI, it can remain unnoticed by users.*

seconds is not long enough to perceive them. See for instance Figure 9: *Johanneskirche SM*: the church's vault structural lines are badly broken; Or *Bed Room SM*, where the curtain is impossibly split into two pieces. This may be in accordance with perception theories that postulate that the gist of a scene is perceived first, and then details are slowly taken into account. Other examples are Figure 8, *Marblehead Mass SC* result: distortion is present in all the lines in the picture.

7 Conclusions and Future Work

The main contributions of this work are as follows. First, we have proposed an eye-tracking data framework, adding eye-tracking data (which we have made public) to some of the images in the *RetargetMe* benchmark. Based on this data, we have computed the corresponding saliency maps of the different results. We have then used these maps as input to a set of existing computational image distance metrics, to analyze how well they correlate with the users' preferences on retargeted images. Our results indicate that using information from eye-tracking data can improve the predicting capabilities of such computational distance measures. We think that this is a promising line of future work which may be helpful to guide the design of future retargeting operators.

Second, we have shown that the model of saliency presented by Judd et al. [2009] is able to work sufficiently well in a retargeting context. We thus think this model can be a helpful tool to predict human fixations when using an eye tracker is not possible.

Third, we have analyzed the influence of the changes due to retargeting in the semantics of the image. When Rubinstein et al. [2010] analyzed the users' reasons for rejecting one retargeted image, the main reasons were: removing or cutting-off content for the SM operator; distorting lines and edges and deforming people and objects for SC; over-squeezing content (especially people and evident foreground objects) in the case of the MOP operator and distorting proportions for SV. In their ranking, MOP and SV achieve the best results according to users' preferences: SC is ranked the lowest and SM is middle ranked. Our results for SM operator do not agree with this classification, it is ranked as low as SC. This result seems to confirm that content preservation is important in order to keep the focal points of the image and, so, the semantics. Although SM results are aesthetically pleasing, its inherent content removal alters the RoIs, which directly affects the ranking results. Nevertheless, this is still an open research topic which we plan to investigate further.

In principle, one obvious reason to change people's visual patterns while looking at an image would be the presence of artifacts. However, we have found that, for our five-second free viewing task, even relatively large artifacts remain unnoticed, as in the case of the church's vault where structural geometry is badly broken (see Figure 9 *Johanneskirche SM*). This seems to be in accordance to the way the human visual system seems to work while recognizing scenes in images: first we extract the gist of a scene (in a few milliseconds) while progressively adding the details [Oliva and Torralba 2006]. In the church example, viewers seem to recognize the scene as a church, assume its overall coherence and correctness, and proceed to visually explore it: the five seconds allotted to our free-viewing task seems not to be long enough to notice the obvious artifacts. The situation changes if the artifacts introduced by the retargeting process do fall in one of the image's RoIs: the normal pattern of visual exploration will naturally guide the viewers' attention towards such regions, where artifact are then more easily detected, as in the case of the car, where its back wheel has been removed (see Figure 9 *Car SM*). We find this *attentional tension* between RoIs and artifacts an interesting line of future work, which again may be helpful to guide future retargeting operators.

Acknowledgements

We thank Aude Oliva for the use of her eye tracker. This research has been partially funded by a Marie Curie grant from the Seventh Framework Programme (grant agreement no.: 251415), the Spanish Ministry of Science and Technology (TIN2010-21543) and the Gobierno de Aragón (projects OTRI 2009/0411 and CTPP05/09).

References

ANSTIS, S. 1998. Picturing peripheral acuity. *Perception 27*, 817–825.

BARNES, C., SHECHTMAN, E., FINKELSTEIN, A., AND GOLDMAN, D. B. 2009. Patchmatch: A randomized correspondence algorithm for structural image editing. *ACM TOG 28, 3*.

BIEDERMAN, I., MEZZANOTTE, R. J., AND RABINOWITZ, J. C. 1982. Scene perception: Detecting and judging objects undergoing relational violations. *Cognitive Psychology 14*, 2, 143 – 177.

BUSWELL, G. T. 1935. How people look at pictures. a study of the psychology of perception in art. *Chicago: University of Chicago Press*.

CHAMARET, C., AND LE MEUR, O. 2008. Attention-based video reframing: Validation using eye-tracking. In *Pattern Recognition, 2008. ICPR 2008. 19th International Conference on*, 1–4.

CHAMARET, C., MEUR, O. L., GUILLOTEL, P., AND CHEVET, J. 2010. How to measure the relevance of a retargeting approach? In *Workshop Media retargeting, ECCV 2010*.

CHEN, L. 2005. The topological approach to perceptual organization. *Visual Cognition 12(4)*, 553–637.

CHOI, M. J., LIM, J. J., TORRALBA, A., AND WILLSKY, A. S. 2010. Exploiting hierarchical context on a large database of object categories. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*.

DAVID, H. 1963. *The Method of Paired Comparisons*. Charles Griffin and Company.

DE GRAEF, P., CHRISTIAENS, D., AND DYDEWALLE, G. 1990. Perceptual effects of scene context on object identification. *Psychological Research 52*, 317–329.

DING, Y., JING, X., AND YU, J. 2011. Importance filtering for image retargeting. In *To appear in Proceedings of IEEE CVPR 2011*.

EHINGER, K., AND OLIVA, A. 2011. Canonical views of scenes depend on the shape of the space. *Submitted manuscript*.

GREENE, M., AND OLIVA, A. 2006. Natural scene categorization from the conjunction of ecological global properties. In *Proceedings of the Twenty-Eighth Annual Conference of the Cognitive Science Society*, 291–296.

GREENE, M. R., AND OLIVA, A. 2009. Recognition of natural scenes from global properties: seeing the forest without representing the trees. *Cognitive psychology 58*, 2 (Mar.), 137–176.

GUTIERREZ, D., LOPEZ-MORENO, J., FANDOS, J., SERON, F., SANCHEZ, M., AND REINHARD, E. 2008. Depicting procedural caustics in single images. *ACM Transactions on Graphics (Proc. of SIGGRAPH Asia) 27*, 5, 120:1–120:9.

HENDERSON, J. M., BROCKMOLE, J. R., CASTELHANO, M. S., VAN R. MACKAND M. FISCHER, M., MURRAY, W., AND HILL, R. 2006. *Visual saliency does not account for eye-movements during visual search in real-world scenes*. Oxford: Elsevier.

ITTI, L., AND BALDI, P. 2009. Bayesian surprise attracts human attention. *Vision research 49*, 10 (June), 1295–1306.

JUDD, T., EHINGER, K., DURAND, F., AND TORRALBA, A. 2009. Learning to predict where humans look. In *IEEE International Conference on Computer Vision (ICCV)*.

KARNI, Z., FREEDMAN, D., AND GOTSMAN, C. 2009. Energy-based image deformation. *Computer Graphics Forum 28*, 5, 1257–1268.

KATUSANI, E., AND YAMADA, A. 2001. The mpeg-7 color layout descriptor: a compact image feature description for high speed image/video segment retrieval. In *In International Conference on Image Processing*, 674–677.

KRÄHENBÜHL, P., LANG, M., HORNUNG, A., AND GROSS, M. 2009. A system for retargeting of streaming video. *ACM Trans. Graph. 28* (December), 126:1–126:10.

LE MEUR, O., CASTELLAN, X., LE CALLET, P., AND BARBA, D. 2006. Efficient saliency-based repurposing method. In *Image Processing, 2006 IEEE International Conference on*, 421–424.

LIU, C., YUEN, J., TORRALBA, A., SIVIC, J., AND FREEMAN, W. 2008. Sift flow: Dense correspondence across different scenes. *In ECCV*, 28–42.

LIU, L., CHEN, R., WOLF, L., AND COHEN-OR, D. 2010. Optimizing photo composition. *Computer Graphic Forum (Proceedings of Eurographics) 29*, 2, 469–478.

LIU, Y.-J., LUO, X., XUAN, Y.-M., CHEN, W.-F., AND FU, X.-L. 2011. Image retargeting quality assessment. *Computer Graphics Forum (Eurographics 2011 Proceedings) 30*, 2.

LOFTUS, G. R., AND MACKWORTH, N. H. 1978. Cognitive determinants of fixation location during picture viewing. *Journal of Experimental Psychology: Human Perception and Performance 4*, 565–572.

MANJUNATH, B., OHM, J., VASUDEVAN, V., AND YAMADA, A. 2001. Color and texture descriptors. *IEEE Trans. Circuits and Systems for Video Technology 11*, 703–715.

MPEG-7. 2002. Iso/iec 15938: Multimedia content description interface.

O. LE MEUR, P. L. C. 2009. What we see is most likely to be what matters: Visual attention and applications. In *IEEE ICIP*.

OLIVA, A., AND TORRALBA, A. 2001. Modeling the shape of the scene: a holistic representation of the spatial envelope. *International Journal of Computer Vision 42*, 145–175.

OLIVA, A., AND TORRALBA, A. 2006. Building the gist of a scene: The role of global image features in recognition. *Progress in Brain Research: Visual perception 155*, 23–26.

OLIVA, A. 2009. *Visual Scene Perception*. Bruce Goldstein. Sage Edition.

PEARSON, E., AND HARTLEY, H. 1966. *Biometrika Tables for Statisticians*, vol. 3_{rd} ed., vol. 1. Cambridge University Press.

PELE, O., AND WERMAN, M. 2009. Fast and robust earth movers distances. *In ICCV'09*.

PRITCH, Y., KAV-VENAKI, E., AND PELEG, S. 2009. Shift-map image editing. In *ICCV'09*, 151–158.

RUBINSTEIN, M., SHAMIR, A., AND AVIDAN, S. 2008. Improved seam carving for video retargeting. *ACM Transactions on Graphics (Proc. SIGGRAPH Asia) 27*, 16:1–16:9.

RUBINSTEIN, M., SHAMIR, A., AND AVIDAN, S. 2009. Multi-operator media retargeting. *ACM Transactions on Graphics (Proceedings SIGGRAPH 2009) 28*, 3, 1–11.

RUBINSTEIN, M., GUTIERREZ, D., SORKINE, O., AND SHAMIR, A. 2010. A comparative study of image retargeting. *ACM Transactions on Graphics (Proc. SIGGRAPH Asia) 29*, 5.

SETYAWAN, I., AND LAGENDIJK, R. 2004. Human perception of geometric distortions in images. *In Proceedings of SPIE, Security, Steganography and Watermarking of Multimedia Contents VI*, 256–267.

SHAMIR, A., AND SORKINE, O. 2009. Visual media retargeting. In *ACM SIGGRAPH ASIA 2009 Courses*, ACM, SIGGRAPH ASIA '09, 11:1–11:13.

SIMAKOV, D., CASPI, Y., SHECHTMAN, E., AND IRANI, M. 2008. Summarizing visual data using bidirectional similarity. *In CVPR '08*.

TORRALBA, A., OLIVA, A., CASTELHANO, M. S., AND HENDERSON, J. M. 2006. Contextual guidance of eye movements and attention in real-world scenes: the role of global features in object search. *Psychological Review 113*, 4 (October), 766–786.

TORRALBA, A. 2003. Contextual priming for object detection. *International Journal of Computer Vision 53*, 2, 169–191.

V. KADIYALA, S. PINNELI, E. C. L., AND CHANDLER, D. M. 2008. Quantifying the perceived interest of objects in images: Effects of size, location, blur, and contrast. In *Proc. Human Vision and Electronic Imaging 2008*.

WANG, Y.-S., LIN, H.-C., SORKINE, O., AND LEE, T.-Y. 2010. Motion-based video retargeting with optimized crop-and-warp. *ACM Trans. Graph. 29* (July), 90:1–90:9.

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

WU, H., WANG, Y.-S., FENG, K.-C., WONG, T.-T., LEE, T.-Y., AND HENG, P.-A. 2010. Resizing by symmetry-summarization. *ACM Trans. Graph. 29* (December), 159:1–159:10.

YARBUS, A. L. 1967. Eye movements and vision. *New York: Plenum*.

YU-SHUEN WANG, CHIEW-LAN TAI, O. S., AND LEE, T.-Y. 2008. Optimized scale-and-stretch for image resizing. *ACM Trans. Graph. (Proceedings of ACM SIGGRAPH ASIA) 27*, 5.