



HAL
open science

Real Time Dynamic Image Re-targeting Based on a Dynamic Visual Attention Model

Matthieu Perreira da Silva, Vincent Courboulay, Patrick Le Callet

► **To cite this version:**

Matthieu Perreira da Silva, Vincent Courboulay, Patrick Le Callet. Real Time Dynamic Image Re-targeting Based on a Dynamic Visual Attention Model. 2012 IEEE International Conference on Multimedia and Expo Workshops (ICMEW), Jul 2012, Melbourne, Australia. pp.653-658, 10.1109/ICMEW.2012.119 . hal-00724486

HAL Id: hal-00724486

<https://hal.science/hal-00724486v1>

Submitted on 21 Aug 2012

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Real time dynamic image re-targeting based on a dynamic visual attention model

Matthieu Perreira Da Silva*, Vincent Courboulay[†] and Patrick Le Callet*

*IRCCyN - University of Nantes, rue Christian Pauc - BP 50609, 44306 Nantes Cedex 03, France

[†] L3i - University of La Rochelle, Avenue M. Crépeau, 17042 La Rochelle Cedex 01, France

Abstract—With the constantly growing number of multimedia devices, images can now be viewed on a broad range of display of tremendously varying size. However, perceiving all the details of a multimedia content is difficult on smaller mobile devices. To solve this problem, saliency based image re-targeting algorithms have been proposed in order to generate more perceptually efficient thumbnails. In this article, we propose a new real-time visual attention based image re-targeting method which differs from previous works in that it generates dynamic (*i.e.* time changing) thumbnails.

This unique behavior is made possible by the use of a dynamic computational model of visual attention. By exploiting the singular characteristics of this model, we are able to dynamically re-target large and complex images in order to efficiently script their discovery on large and small displays. We achieved such re-targeting by integrating simulated attentional focuses over a fixed time window into a temporal heat-map. By segmenting this heat-map at each time step, we can generate the region of interest that will be displayed as dynamic preview.

In order to validate that every salient location of the image has been displayed, we compare the regions shown in the preview with mean observers eye fixations obtained during a free viewing experiment.

Index Terms—Visual attention, image re-targeting, discovery, adaptation, dynamic heat-map, mobile multimedia.

I. INTRODUCTION

Digital photographs are generally shot under the assumption that they will be displayed on a high resolution media. This is the case if these images are printed on high quality paper or displayed in full screen on a desktop PC or laptop screen. But even in these favorable conditions, screen resolution (2 million pixels for a full HD screen) is still far from original images resolution (up to more than 20 million on high-end DSLR). Additionally, there are many cases where screen resolution is much more limited (phone, portable music players, etc.) and / or a list of small image thumbnails should be displayed. In all these cases, image processing algorithms must be used in order to display the original images efficiently. The most straightforward approach is simple downscaling, but this method may result in perceived decrease in image quality since many details will be lost.

Another possibility is to perform content-aware image resizing through the use, for example, of seam carving [1]. This method removes low energy pixel paths (named *seams*) from the original image in order to reduce its size. Impressive results can be obtained on certain types of images, and for aspect ratio changes. However, for general use, these methods are difficult to use, because they can lead to huge deformations

[1].[2] has evaluated many image re-targeting algorithms both objectively and subjectively : among them, simple cropping is found as one of the three best algorithms. Hence, simple cropping based on saliency can be an efficient alternative [3]. A more complex approach is proposed by [4] which uses saliency based figure ground segmentation in order to resize background while preserving foreground figures.

In this article, we focus on extending the approach of [3] which proposes to use a heat-map (generated using eye-tracking data) or a saliency map (obtained through its own visual attention model) in order to crop the original image. The resulting cropped region contains only the most salient elements of the original image. This approach is demonstrated as efficient but if the scene is complex (e.g. it contains many salient objects) cropping is not performed and the algorithm reverts to showing the whole (downscaled) image. In this case, it would be more appropriate to discover dynamically the different salient elements of the scene.

Our article is organized as follow. In the next section, we introduce our new dynamic re-targeting method which is based on a dynamic model of visual attention. In section II, we describe the basic ideas supporting our method. In section III, we refine the method in order to exploit efficiently the dynamic properties of the real-time attention model used. Finally, in section IV, we evaluate the model with respect to eye-tracking data.

II. A DYNAMIC RE-TARGETING METHOD

In this section, we describe how dynamic image re-targeting can be performed using either eye-tracking data or a dynamic model of visual attention. We define dynamic re-targeting as an image cropping which evolves over time. In order to generate this time varying cropping, we need visual attention data which are also time varying. This is naturally the case for eye-tracking data, but it is not for many visual attention models [5][6][7]. We study this latter point in the following subsection.

A. Computational modeling of attention : saliency map vs. focus points

Visual attention models generally fall into two categories. *Central representation based models*, consider that attention is processed in a central location, usually named saliency map. This map represents the (bottom-up) attentional attractiveness

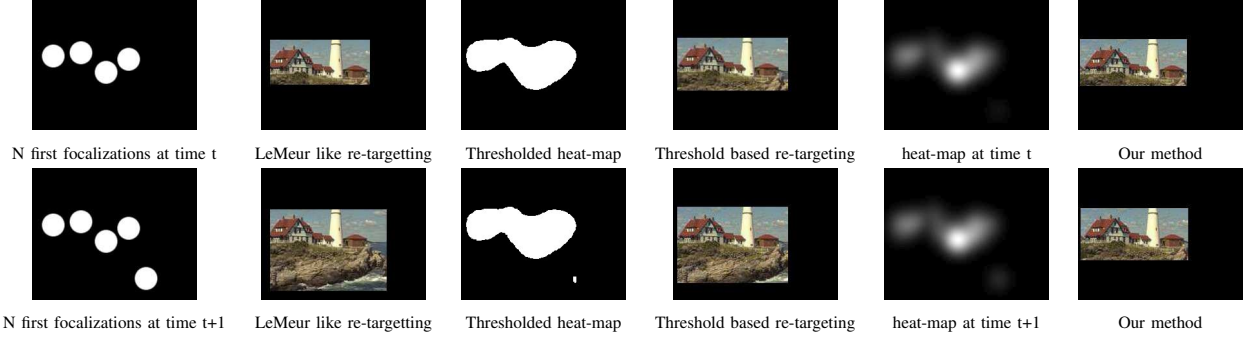


FIGURE 1: Sensitivity of re-targeting methods to new focalizations. By not using any threshold, our method is less sensitive to the addition of new focalisations: dynamic re-targetting is smoother.

of each pixel of an image. It is closely related to heat-maps which are built from the integration of all eye-fixations measured when a set of observers watches an image during a fixed period of time. Whatever their theoretical framework (hierarchical [8], spectral domain[5], information theoretic [6] or probabilistic [7], to name a few), *central representation based models* generally only output a saliency map. However, for dynamic re-targeting we need time varying output : the attention model used should be able to generate focus points. Some models can achieve this goal with the addition of *winner-takes-all* and *inhibition of return* algorithms [8][9]. But another, more natural, way of generating attentional focus is to rely on *distributed models* of visual attention [10][11] which consider that attention is a consequence of the competition between different sources of interest. In this case, attention is spread in all the visual system. These models are built to directly generate focus points and consequently provide more control possibilities. In particular, the competition based model presented in [12] allows adaptation mechanisms that we have used to improve dynamic re-targeting (see section III). Additionally it is very fast (30fps for a 640x480 image on a 2.7GHz dual core i7 processor), which allows to generate *dynamic thumbnails* in real time on a standard computer, and by extension possibly near real time on a mobile device.

B. Re-targeting for dynamic data

For static re-targeting, [3] uses the position of the N first focus points in order to define a cropping box that surround the most salient parts of an image $I(x, y)$. These focus points are processed using a *winner-takes-all* algorithm on a heat-map HM or saliency map SM . Another, even simpler method, is to binarize HM or SM using a predefined threshold $th_{coverage}$. For both methods, the cropping box B is defined as the bounding for of all non zero image pixels (figure 1). For a given still image, both maps do not evolve over time.

To extend the previous algorithms to dynamic re-targeting, we can iteratively built a heat-map HM_t at each time step :

$$HM_t(x, y) = \left(\sum_{i=1}^{N_t} \delta_{x_i, y_i}^{x_i, y_i} \right) * g_{\sigma_x, \sigma_y}(x, y) \quad (1)$$

were t is the current time step, N_t the number of focus points acquired at time t , g is a Gaussian kernel, (x_i, y_i) are the coordinates of fixation i and $\delta_{x_i, y_i}^{x_i, y_i} = \begin{cases} 1 & \text{if } (x_i, y_i) = (x_j, y_j) \\ 0 & \text{otherwise} \end{cases}$.

This *iterative heat-map* can be built from eye-tracking data, using the output of a dynamic visual attention model like [12], or by using the focus points generated by a *winner-takes-all* and inhibition mechanisms on a "standard" saliency map.

With this iterative heat-map, we can generate dynamic cropping boxes. However, as shown in figure 1, because of the use of binary data (either a thresholded heat-map or focus points), previous methods are very sensitive to the addition of new focus points. To solve this problem we propose a new method based on the mean (or centroid) and standard deviation of heat-map values. It prevents from using a threshold and modulates the influence of every pixels in the heat-map by their intensity. For a heat-map HM , centroid (c_x, c_y) and left, right, top and bottom variances v_l, v_r, v_t, v_b are calculated as follows :

$$c_x = \frac{\sum_x \sum_y x HM(x, y)}{\sum_x \sum_y HM(x, y)} \quad c_y = \frac{\sum_x \sum_y y HM(x, y)}{\sum_x \sum_y HM(x, y)}$$

and

$$v_l = \sqrt{\frac{\sum \sum \delta^-(x, c_x)^2 HM(x, y)}{\sum \sum \delta^-(x, c_x)^2}} \\ v_r = \sqrt{\frac{\sum \sum \delta^+(x)^2 HM(x, y)}{\sum \sum \delta^+(x)^2}} \\ v_t = \sqrt{\frac{\sum \sum \delta^-(y, c_y) HM(x, y)}{\sum \sum \delta^-(y, c_y)^2}} \\ v_b = \sqrt{\frac{\sum \sum \delta^+(y, c_y) HM(x, y)}{\sum \sum \delta^+(y, c_y)^2}}$$

$$\text{with } \delta^-(i, j) = \begin{cases} i - j & \text{if } i < j \\ 0 & \text{otherwise} \end{cases} \\ \delta^+(i, j) = \begin{cases} i - j & \text{if } i \geq j \\ 0 & \text{otherwise} \end{cases}$$

and the corresponding bounding box B is defined as :

$$\begin{aligned}
x_B &= c_x - \alpha v_l \\
y_B &= c_y - \alpha v_t \\
W_B &= \alpha(v_r - v_l) \\
H_B &= \alpha(v_b - v_t)
\end{aligned}$$

with (x_B, y_B) the coordinates of the top left corner of B , W_B and H_B its width and height, and α a scaling parameter. In the following, we have used a value of $\alpha = 2$. For a heat-map containing only one focus point (represented by a single Gaussian), this correspond to keeping 95% of the global saliency. This statement is true only in the case of a single focus point. For more focus points, the distribution is not Gaussian anymore and the amount of saliency represented is lower.

Thus, our system is less sensitive to sudden expansion of the cropping box when adding new focus of attention. Nevertheless, as this behavior is still sometimes problematic, we introduce another improvement by adding inertia i to the cropping box size update equations. We propose the following procedure for this iterative update:

$$\begin{aligned}
x_B(t) &= i \times x_B(t-1) + (1-i) \times (c_x - \alpha v_l) \\
y_B(t) &= i \times y_B(t-1) + (1-i) \times (c_y - \alpha v_t) \\
W_B(t) &= i \times W_B(t-1) + (1-i) \times \alpha(v_r - v_l) \\
H_B(t) &= i \times H_B(t-1) + (1-i) \times \alpha(v_b - v_t)
\end{aligned}$$

In this way, we can smooth important variations of the cropping box at the expense, however, of reactivity.

Since it is now possible to incrementally generate a heat-map as the simulation proceeds, it becomes possible to use the mechanism described above to dynamically discover the image. Some improvement are however still necessary since our dynamic cropping does not allow a real sequential discovery of the salient elements of the scene (Figure 2).

In the next subsection, we present different ways to use the dynamic and adaptive properties of [12] visual attention model in order to improve the discovery pathway.

III. ADVANCED RE-TARGETING

As mentioned before, the main objective of this section is to present two methods implemented to better control the re-targeting process.

A. A forgetting heat-map

In section II-B, we described how to build a heat-map incrementally in order to generate dynamic re-targeting (equation 1). In this section we would like to introduce another change in the way heat-maps are generated. We assume that integrating a heat-map from $t = 0$ is not necessary, and that it could be more interesting to forget what happened long time ago. By the way, cropping would be more dynamic if it depended on an estimation of saliency calculated mainly from the latest focus points. To create such a *forgetting heat-map* incrementally,

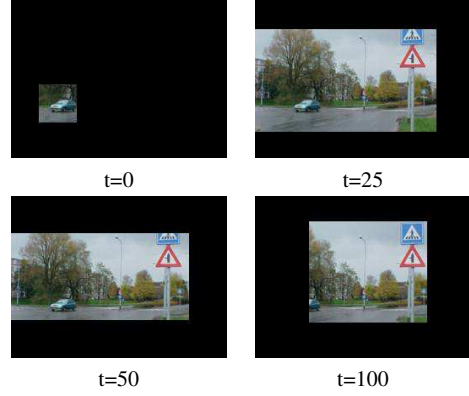


FIGURE 2: Example of dynamic re-purposing. The image is progressively discovered, but the process is too fast ("final" cropping is obtained as early as $t=50$. After that the system does not evolve much).

we just have to weight $HM_s(x, y, t-1)$ by a forgetting factor $forget \in [0, 1]$:

$$HM_s(x, y, t) = (1 - forget) \times HM_s(x, y, t-1) + (\delta_{x,y}^{x_t, y_t} * g_{\sigma_x, \sigma_y}(x, y)) \quad (2)$$

If $forget = 0$, the heat-map is computed as usual: it represents the image saliency distribution for the whole simulation duration. With a smaller value (typically $forget = 0.05$), the *forgetting heat-map* represents saliency distribution for a shorter duration: a *local time saliency*.

In the case of images cropping, the use of such an *forgetting heat-map* can make the system much more dynamic: it can display to the user parts of the image that recently *caught* the attention of the model. Thus, the cropping process is constantly evolving.

Nevertheless, this cropping behavior can still be improved. Yet, we cannot choose if it has to focus on small details or present a larger view of each image to the users. In the next section we introduce a mechanism that proposes a first solution to this question.

B. Attentional feedback

Another way of introducing intelligent dynamics is to consider a feedback factor inside the visual attention model itself. The feedback we propose, is a map R built upon a visited area map M . This latter is constructed incrementally so as to remember all the visited areas in the scene:

$$M(x, y, t) = \max(M(x, y, t-1), \frac{N - \min(\frac{d(x, y, x_f, y_f)}{blurSize}, N)}{N}) \quad (3)$$

where (x_f, y_f) are the coordinates of the simulated focus of attention at time t ; $d(x_1, y_1, x_2, y_2)$ is the Euclidean distance between (x_1, y_1) and (x_2, y_2) ; $blurSize$ the size of the retinal area (fixed to 10% of the largest image dimension; this value may be associated with human fovea size (about 2 degrees of visual field)); $N = \text{ceiling}(\log_2(\min(W, H)))$ and (W, H) the size of the input image. It guarantees that

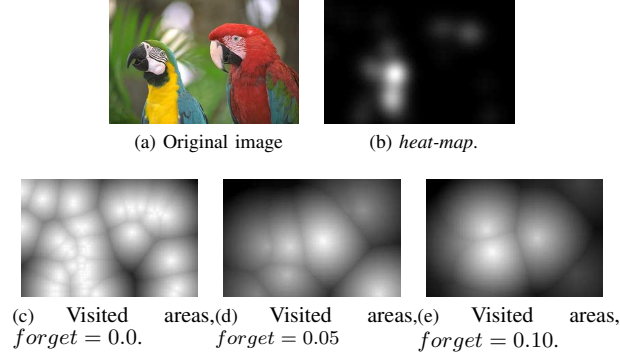


FIGURE 3: Influence of *forget* on the visited area map, after 100 attention simulation iterations.

$$M(x, y) \in [0, 1] \forall x, y.$$

Human memory is however limited, so attentional focus is most probably influenced by only the most recent focus points. As in section III-A, we have decided to introduce a forgetting factor *forget* $\in [0, 1]$ which will iteratively attenuate the role of the oldest focus points:

$$M(x, y, t) = \max\left(\frac{(1 - \text{forget}) \times M(x, y, t - 1)}{N - \min\left(\frac{d(x, y, x_f, y_f)}{\text{blurSize}}, N\right)}, 0\right) \quad (4)$$

Figure 3 shows the influence of *forget* on the visited areas map M . Another parameter, *feedback* allows modulating the influence of the visited areas map in intensity and feedback type (positive or negative):

$$R(x, y) = \begin{cases} \frac{1 + |\text{feedback}| \times M(x, y)}{1 + |\text{feedback}|} & \text{if } \text{feedback} \geq 0 \\ \frac{1 + |\text{feedback}| \times (1 - M(x, y))}{1 + |\text{feedback}|} & \text{otherwise} \end{cases} \quad (5)$$

with $R(x, y) \in [0, 1] \forall x, y$.

Then, we can use R in order to modulate the dynamics of the original equation of [12]. A positive *feedback* value will lead to a focusing or tracking behavior since already visited objects/locations are preferred. A negative *feedback* value will lead to an exploration behavior since unknown (unvisited) objects/locations will be favored.

In order to improve the discovery properties of our re-targeting algorithm, we propose that *feedback* gradually change from 1 (strong focus on the most salient elements) to -1 (exploration of the whole scene). Thus, users discover the image by first viewing the few most salient regions, and finally the entire scene.

C. Dealing with aspect ratio

For some applications, the thumbnails generated by the re-targeting algorithm need to have a 1:1 aspect ratio (square images). In order to deal with this requirements, two aspect ratio adaptation are possibles:

- *Cropping box extension* (as proposed by [3]) which displays more data than calculated by the cropping algorithm, in order to display all salient objects. This

method maximizes the overall saliency of the content displayed since no salient information is discarded, but it can cause problems when extension needed by the 1:1 ratio constraint generate images larger than the original one.

- *Cropping box reduction*, limits the amount of content displayed in order not to include additional non salient area. It also guaranties that the cropped image will always be smaller than the original one.

Since we want to optimize the ratio of salient vs. non salient pixels in the cropped images, we have chosen cropping box reduction.

IV. EXPERIMENTS AND RESULTS

Qualitative and quantitative assessment of non dynamic saliency based re-targeting has already been studied in [3]. Consequently, in this article, we focused only on the quantitative study of the impact of adding dynamics through three measures:

- *CA*: the percentage of original image area displayed (*cropped-area*).

$$CA = \frac{W_c \times H_c}{W \times H}$$

where W_c , H_c and W , H are respectively the width and height of the original and the cropped images;

- *FP*: the percentage of *fixated pixels* (non zero pixels in binarized ground truth) not displayed in the cropped image. It measures the amount of original salient pixels that have been lost in the cropped images.

$$FP = \frac{\sum_x \sum_y \bar{\gamma}_{x_c, y_c, W_c, H_c}(G_b(x, y), x, y)}{\sum_x \sum_y \alpha(G_b(x, y))}$$

with

$$\bar{\gamma}_{x_c, y_c, W_c, H_c}(i, x, y) = \begin{cases} 1 & \text{if } i \neq 0 \text{ and} \\ & x \notin [x_c, x_c + W_c] \text{ and} \\ & y \notin [y_c, y_c + H_c] \\ 0 & \text{otherwise} \end{cases}$$

$$\alpha(i) = \begin{cases} 1 & \text{if } i \neq 0 \\ 0 & \text{otherwise} \end{cases}$$

and G_b the binarized version of the ground truth image;

- *IC*: *image coverage* [13], which is defined as the percentage of fixated points in an image (non zero pixels in binarized ground truth). It is different from the *fixated pixels lost* ratio in that it measures the percentage of salient pixels vs. all pixels in the cropped image.

$$IC = \frac{\sum_x \sum_y \gamma_{x_c, y_c, W_c, H_c}(G_b(x, y), x, y)}{W_c \times H_c}$$

with

$$\gamma_{x_c, y_c, W_c, H_c}(i, x, y) = \begin{cases} 1 & \text{if } i \neq 0 \text{ and} \\ & x \in [x_c, x_c + W_c] \text{ and} \\ & y \in [y_c, y_c + H_c] \\ 0 & \text{otherwise} \end{cases}$$

Our experiments have been run on images extracted from the 29 undistorted reference images of the LIVE database [14] on which eye-tracking data were recorded. Table 4 summarizes the setup of these experiments.

For each image in the database, we generated :

- a *ground truth heat-map* G , generated using the eye-tracking data from the 21 observers; a binarized version of the heat-map G_b (using the same experimentally defined threshold for all images); and a static re-targeted image generated from the binarized heat-map;
- a *simulated heat-map* S , generated using [12] algorithm during 300 time steps; a binarized version of the heat-map; and a static re-targeted image generated from the binarized heat-map;
- different sets of simulated *dynamic heat-map*, generated at each time step of a 300 steps simulation; and different sequences of re-targeted images generated using the algorithms described in sections II and III. SS_t , generated using the simple version of the algorithm described in sections II using a free aspect ratio. SF_t , generated using the *forgetting heat-map* described in section III-A using a free aspect ratio. SFF_t , generated using the forgetting heat-map and attentional feedback described in section III-B using a free aspect ratio. $SFF2_t$, generated using the same algorithm as SFF_t but with a 1:1 aspect ratio.

Sample re-targeting images from a selection of these algorithms are shown in figure 5.

The results of all the measures made on these re-targeting models are shown in table I. The original (non cropped) images contain all fixated pixels, but their coverage is low (15%) : these images contain many non salient pixels. When the images are cropped using eye-tracking data (ground-truth heat-maps) mean cropped area falls to 37% whereas mean coverage drops to 44% : the thumbnails obtained are much smaller and contain more salient pixels. This method is however not usable in practice since it would require to run eye-tracking experiment on a large number of observers to be able to re-target an image. An alternative solution is simple static re-targeting based on visual attention simulation. But this approach is far less efficient since it removes 24% of salient pixels (as determined by eye tracking). This is mainly due to the fact that the attention model used (as any model) is not totally in accordance with eye-tracking data.

Dynamic re-targeting allows to overcome this limitation since the animated thumbnails allow to visit more areas and consequently display almost all salient pixels (only 1% to 5% loss). For most dynamic re-targeting algorithms this comes at the price of slightly less coverage and higher mean cropped area. The only exception is 1:1 ratio dynamic forgetting feedback cropping ($SFF2_t$) which can achieve both "high" mean coverage and low cropped area. This is mainly due to a more restrictive cropping (see section III-C) which concentrates viewing on the most salient parts of the image.

V. DISCUSSION AND CONCLUSION

In this article we have presented a new algorithm for dynamic image re-targeting on small display devices (mobile

Model	Mean CA	Mean FP	Mean IC
Original image	100%	0%	15%
GT	37%	0%	44%
S	41%	24%	31%
SS_t	55%	5%	26%
SF_t	51%	2%	27%
SFF_t	43%	1%	26%
$SFF2_t$	29%	4%	32%

Table I: Experimentation results

phones, multimedia players, etc.). This algorithm is based on first and second order statistics of a simulated *forgetting heat-map*. Different variations of the method used were described. Qualitative measures run on both static and dynamic re-targeting algorithms show that dynamic re-targeting allows to reduce the number of non-salient pixels displayed while keeping almost all salient pixels (which is not the case of classical static re-targeting methods). Of course these measures don't take into account the loss or gain in terms of quality of experience. Other experiments should be conducted in order to study this important aspect of the evaluation of re-targeting algorithms. For these experiments, dynamic cropping algorithms should probably be improved in order to limit image motion (panning and zooming) and provide a smoother more "professional video" like result.

REFERENCES

- [1] S. Avidan and A. Shamir, "Seam carving for content-aware image resizing," in *ACM Transactions on Graphics (TOG)*, vol. 26, no. 3. ACM, Jul. 2007, p. 10. [Online]. Available: <http://portal.acm.org/citation.cfm?doid=1276377.1276390> <http://portal.acm.org/citation.cfm?id=1276390>
- [2] M. Rubinstein, D. Gutierrez, O. Sorkine, and A. Shamir, "A comparative study of image retargeting," in *ACM Transactions on Graphics (TOG)*, vol. 29, no. 6. ACM, 2010, p. 160. [Online]. Available: <http://portal.acm.org/citation.cfm?id=1866186>
- [3] O. Le Meur, X. Castellan, P. Le Callet, and D. Barba, "Efficient saliency-based repurposing method," in *IEEE International Conference on Image Processing*, Atlanta, USA, 2006, pp. 421–424. [Online]. Available: <http://www.irccyn.ec-nantes.fr/~lecallet/paper/LeMeur-icip06.pdf>
- [4] T. Lechner and B. Gooch, "Images and Video for Preserving Information Saliency," *Ieee Computer Graphics And Applications*, no. October, pp. 80–88, 2007.
- [5] X. Hou and L. Zhang, "Saliency Detection: A Spectral Residual Approach," in *2007 IEEE Conference on Computer Vision and Pattern Recognition*, no. 800. Ieee, Jun. 2007, pp. 1–8. [Online]. Available: <http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=4270292>
- [6] N. D. B. Bruce and J. K. Tsotsos, "Saliency, attention, and visual search: An information theoretic approach," *Journal of Vision*, vol. 9, no. 3, p. 5, 2009.
- [7] T. Avraham and M. Lindenbaum, "Esaliency (extended saliency): meaningful attention using stochastic image modeling," *IEEE transactions on pattern analysis and machine intelligence*, vol. 32, no. 4, pp. 693–708, Apr. 2010. [Online]. Available: <http://www.ncbi.nlm.nih.gov/pubmed/20224124>
- [8] L. Itti, C. Koch, E. Niebur, and Others, "A model of saliency-based visual attention for rapid scene analysis," *IEEE Transactions on pattern analysis and machine intelligence*, vol. 20, no. 11, pp. 1254–1259, 1998. [Online]. Available: <http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.53.2366>
- [9] O. Le Meur, P. Le Callet, D. Barba, and D. Thoreau, "A coherent computational approach to model bottom-up visual attention," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 28, no. 5, pp. 802–817, 2006. [Online]. Available: <http://www.irccyn.ec-nantes.fr/~lecallet/paper/LeMeur-IEEEPAMI06.pdf>

Category	Details	Conditions
Participants	Number	21
	Age range	18-42
	Average age	26
	Male/female	11/10
	Non-experts/experts	21/0
	Occupation	University staff / students
	Compensated	Yes
Viewing conditions	Environment	Conformant with ITU-R BT.500-11 recommendations
	Illumination	Low
	Viewing distance	70 cm
	Task	Free-viewing
Display	Make	DELL
	Type	LCD
	Size	19"
	Resolution (pixels)	1280x1024
Eye-tracker	Make	SMI iView X Hi-Speed
	Type	Infrared video-based
	Frequency	500 gaze points/sec
	Accuracy	0.25-0.5 degrees of visual angle
	Mounting	Tower with head rest
	Calibration	9 points screen
Image presentation	Order	Random
	Image duration	15 s
	Grey-screen duration	3 s
	Max/ visual angle (pixels/deg)	41.8
	Central fixation point	No

Figure 4: Overview of the eye-tracking experiments protocol.



FIGURE 5: Top row, static re-purposing (from left to right): original image, ground-truth heat-map, binarized ground-truth heat-map, cropped image from ground-truth GT , heat-map from simulation, binarized heat-map from simulation, cropping from simulation S . Bottom row : dynamic re-purposing from simulation SF_t , frames 1,50,100,150,200,250 and 300.

- [10] G. Deco, "A Neurodynamical cortical model of visual attention and invariant object recognition," *Vision Research*, vol. 44, no. 6, pp. 621–642, 2004. [Online]. Available: <http://linkinghub.elsevier.com/retrieve/pii/S0042698903006928>
- [11] J. Tsotsos, Y. Liu, J. Martineztrujillo, M. Pomplun, E. Simine, and K. Zhou, "Attending to visual motion," *Computer Vision and Image Understanding*, vol. 100, no. 1-2, pp. 3–40, 2005. [Online]. Available: <http://linkinghub.elsevier.com/retrieve/pii/S1077314205000779>
- [12] M. Perreira Da Silva, V. Courboulay, A. Prigent, and P. Estraillier, "Evaluation of preys / predators systems for visual attention simulation," in *VISAPP 2010 - International Conference on Computer Vision Theory and Applications*. Angers: INSTICC, 2010, pp. 275–282. [Online]. Available: http://hal.archives-ouvertes.fr/docs/00/48/52/53/PDF/visapp_mperreir.pdf
- [13] D. S. Wooding, "Eye movements of large populations: II. Deriving regions of interest, coverage, and similarity using fixation maps." *Behavior research methods, instruments, & computers : a journal of the Psychonomic Society, Inc*, vol. 34, no. 4, pp. 518–28, Nov. 2002. [Online]. Available: <http://www.ncbi.nlm.nih.gov/pubmed/12564556>
- [14] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: from error visibility to structural similarity." *IEEE transactions on image processing : a publication of the IEEE Signal Processing Society*, vol. 13, no. 4, pp. 600–12, Apr. 2004. [Online]. Available: <http://www.ncbi.nlm.nih.gov/pubmed/15376593>