Improved Image Retargeting by Distinguishing between Faces in Focus and out of Focus

Johannes Kiess, Rodrigo Garcia, Stephan Kopf, Wolfgang Effelsberg

Department of Computer Science IV
University of Mannheim
Germany
{*kiess, kopf, effelsberg*}*@informatik.uni-mannheim.de*

Abstract—The identification of relevant objects in an image is highly relevant in the context of image retargeting. Especially faces draw the attention of viewers. But the level of relevance may change between different faces depending on the size, the location, or whether a face is in focus or not. In this paper, we present a novel algorithm which distinguishes in-focus and out-of-focus faces. A face detector with multiple cascades is used first to locate initial face regions. We analyze the ratio of strong edges in each face region to classify out-of-focus faces. Finally, we use the GrabCut algorithm to segment the faces and define binary face masks. These masks can then be used as an additional input to image retargeting algorithms.

Keywords-face detection; focus detection; saliency of faces; image retargeting; image resizing; seam carving; grabcut;

I. Introduction

The fast development of display technologies in the last years paved the way for a flood of new display devices. The introduction of high definition displays contributed even more to the increasing diversity of display resolutions available, which range from the 176×132 pixels of a small mobile MP3 player up to 3840×2400 pixels of a high-end LCD monitor. At the same time, the availability of high definition displays pushed the need for content in high resolution, thus increasing the diversity of image resolutions. *Image retargeting* describes the process of adapting images to these different resolutions and aspect ratios.

Content-aware image retargeting methods make use of various metrics in order to measure the importance of different regions in an image. This analysis is used to maintain both scene structure and content when resizing the image. Therefore, the development of functions that evaluate how important these regions are to the viewer is of fundamental importance to the successful utilization of content-aware image resizing algorithms.

In 2010, a comprehensive study of image retargeting operators was published by Rubinstein et al. [1]. The study showed that viewers were highly sensitive to distortions introduced by retargeting operators, particularly when changing well-known objects, geometric structures, or the symmetry of objects. Faces play an important role in retargeting as they usually draw the attention of the viewer. But not every face is in the main focus; some may be located in the background or may be partially occluded by other objects. Current face detection algorithms do not distinguish between faces in focus and smaller, partially occluded, or unsharp faces in the background. Image retargeting operators would set a high priority to all of them. This may lead to a higher deformation of important objects as they are assigned a lower importance than an out-of-focus face.

In this paper, we present a novel enhancement for the saliency of faces which uses an improved method to classify the relevance of automatically detected faces. We use the algorithm proposed by Viola and Jones [2] with multiple cascades for the identification of faces. The image sharpness is analyzed to detect whether a face is in focus or not. Finally, the GrabCut algorithm [3] is used to segment the faces into *face masks*.

Our main contributions are as follows:

- We introduce an improvement to face detection which is able to distinguish between faces in focus and out of focus.
- GrabCut requires manually defined image regions. We propose an extension to initialize GrabCut automatically based on detected faces.
- We show the benefit of applying our novel face masks in the context of image retargeting. The quality of adapted images is illustrated by using a retargeting algorithm that combines seam carving and cropping called SeamCrop [4].

The outline of this paper is as follows: Chapter 2 gives an overview of face detection and state of the art image retargeting techniques. Our algorithm is presented in detail in Chapter 3. Chapter 4 shows our results and illustrates the application of our enhanced face detection technique in the context of image retargeting. Finally, Chapter 5 concludes the paper.

II. Related Work

A substantial amount of previous work has been done in the area of face detection. A comprehensive survey on face detection was done by Zhao et al. [5]. In the work presented

Figure 1. Haar-like rectangular features used in the Viola and Jones framework (as presented in [2]).

by Murphy-Chutorian et al. [6], a survey is conducted which focuses on head pose estimation. We use a face detection technique based on the work presented by Viola and Jones [2]. Motivated by the problem of detecting faces in an image, the authors introduced a general framework for object detection. Three main contributions can be drawn from their research. They use the concept of *integral images* to represent an image, apply a new method for building classifiers using AdaBoost [7], and combine classifiers in a cascade structure.

Rectangular features are used for image classification (see Figure 1). To speed up the computation of the features, an integral image representation is applied. The integral image for point (x, y) is defined as the sum of all the pixels above and to the left of this point. Integral images have the convenient property that any rectangular sum can be obtained by using only four array references. This has the advantage that any one of the features can be computed in constant time, regardless of scale or location.

Cascades are a set of classifiers which are built like a decision tree. When a first simple classifier returns a positive result, a second more complex classifier is used. This is repeated until all classifiers in the set are processed. A negative result from any classifier leads to the rejection of the sub-window. This speeds up the detection as complex classifiers do not have to be used on all possible sub-windows.

The focus detection we want to apply is best comparable to the autofocus functions in modern digital cameras. In both cases, it is necessary to classify whether a region in an image is in focus or not. Autofocus functions are either active or passive. Active autofocus uses infrared light and computes the time difference between the sending of the signal and the receiving of the signal which was reflected by the object that should be focused. Passive autofocus measures the intensity difference between adjacent pixels and tries to maximize the differences. As our retargeting algorithm uses digital images as input, we apply an approach that is similar to the passive autofocus technology.

We evaluate the quality of our face detection algorithm in the context of image retargeting which is based on seam carving. This technique was first introduced by Avidan and Shamir [8] and reduces the width or height of an image by removing connected paths of pixels called *seams* which reach either from top to bottom or from left to right. Rubinstein et al. [9] extend this approach and combine seam carving with scaling and cropping. In previous work, we

have introduced a retargeting technique called SeamCrop [4] which combines seam carving and cropping. Details of this algorithm are discussed in Section 3.4. This technique is also applicable for videos [10]. Pritch et al. [11] optimize so-called shift maps by applying the graph cut algorithm. These maps show the relative shift of each pixel between the original and the adapted image. Warping describes a technique where a mesh is put over an image and the cells of the mesh are scaled in a non-uniform way. The goal is to keep the important regions of an image unchanged and at the same time shift the distortions to homogeneous regions. Wang et al. [12] use quad cells for the mesh and perform an iterative minimization while Guo et al. [13] apply a triangular mesh and solve the retargeting as a constrained mesh parameterization problem.

All retargeting methods have in common that they use saliency maps. Such maps describe the relevance of each pixel for the human viewer. A retargeting technique then tries to preserve pixels of high relevance while removing all other pixels. By improving the saliency maps, our enhanced face detection algorithm is applicable to all image retargeting methods.

III. IN-FOCUS FACE DETECTION

As previously stated, the Viola and Jones framework [2] is used to detect face regions in images. Face regions are used as the input for the algorithm which is presented in the following. In a first step, each face is classified as in focus or as out of focus. Next, in-focus faces are automatically segmented from the background with GrabCut in order to generate face masks. Finally, these masks are encoded as binary maps which can then be used in image retargeting algorithms.

A. Face detection with multiple cascades

It is crucial that the face detection algorithm has a high detection rate while at the same time keeping the number of false positives low. For example, whenever a non-face region is classified as a face, this region would be included in the face masks, and the other areas of the image would be deformed by a resizing operator.

To achieve higher detection rates, the algorithm uses multiple cascades, which are trained with different face sets for the detection of face regions. Multiple cascades generate multiple detections of the same face. On the other hand, cascading will also increase the number of false detections.

In order to identify only one pair of coordinates for each face and eliminate probable false detections, the detected face regions are clustered. A threshold parameter δ is calculated derived from the width w and height h of both rectangles:

$$\delta = s * [min(w_1, w_2) + min(h_1, h_2)] * 0.5 \qquad (1)$$

If the absolute difference between the rectangles upper corners is both smaller than or equal to δ, the rectangles are

labeled as belonging to the same class. In case of $s = 0$, each rectangle belongs to a separate cluster, whereas $s = \infty$ aggregates all rectangles into one cluster. Considering that four cascades are used in our algorithm, a value of $s = 0.2$ provides good results. The faces in a cluster are removed if less than three rectangles have been assigned to this cluster. The four cascades we use are provided by the OpenCV SDK[1].

B. Focus detection

A focus detection algorithm is used to exclude all blurry faces. A face region is classified as an elliptical object (or blob). Each face includes characteristic edges due to nose, mouth, eyes and eyebrows, among others. Strong edges are visible when a face is in focus. A face is classified as in focus when at least one of its edges is classified as a strong edge. Figure 2 visualizes the absence of strong edges in blurry faces.

An obvious approach to the edge detection problem is to analyze luminance variations in an image. We use the gradient magnitude [14] to identify edge pixels. A threshold is applied to remove weak edge pixels with a lower edge value. The threshold is set to 12.5% of the maximum value of the gradient magnitude. Based on the maximum gradient magnitude in the whole image, different threshold values were tested: 1/16 (6.25%), 1/8 (12.5%) and 1/4 (25%) were examined in different images. A value of 12.5% empirically kept most of the in-focus edges and excluded most of the out-of-focus edges.

Border pixels between face and background may cause additional strong edge pixels. Therefore, only a small area in the center of the face corresponding to 25% of the face rectangles area is considered. A face is classified as in focus if a strong edge pixel is found in this area.

C. Creating face masks with GrabCut

To segment objects, the original GrabCut algorithm requires some user interaction [3]. The idea of the algorithm is to use Gaussian Mixture Models to specify the color distribution of background and foreground pixels. Initial labels (foreground, background, probably foreground, or probably background) are provided by a user [15]. In our case, to avoid manual classification, information provided by the face detection module is used.

A face region is defined by the center and the width and height of a rectangular region (see Figure 3(a)). The position or size of such a rectangle might not describe the exact face region. In order to obtain a more accurate segmentation of the face, labels are assigned to pixels inside and in the neighborhood of the rectangle.

The alignment or size of a detected face might be inaccurate. Therefore, we define a safety margin and add an

---

[1]opencv.willowgarage.com/

additional 40% to each border of the rectangle. Figure 3(b) shows the enlarged face region. All pixels outside of this region are labeled as background and will not be part of the final segmentation. The labels 'foreground', 'probably foreground', and 'probably background' are assigned to the pixels of the enlarged face region depending on the distance of each pixel from the center of the region. Figure 3(c) visualizes the result of the labeling step.

GrabCut is applied in the last step to identify face pixels (see Figure 3(d)). To reduce the effect of discontinuities at the borders of a segmented face a morphological dilation operation with a square 3×3 structuring element is applied to the image. This reduces minor inaccuracies and also adds a safety margin so that no face pixels are missed in the mask.

D. Image retargeting algorithm

We selected image retargeting as an application scenario and use the automatically detected face masks to improve the quality of the retargeted images. For this purpose, we use a retargeting technique which combines seam carving and cropping called SeamCrop [4].

In this algorithm, the estimated face mask is used as additional information for the energy map. Therefore, the map is defined as the sum of absolute gradients and the face mask. Both parts are normalized to a range of values so that a face alone with no energy from the gradients will still get a higher value than the maximum value for gradients.

The algorithm starts by applying seam carving to the image. Seams are iteratively removed as long as the energy of the removed pixels does not exceed a dynamic threshold. This threshold equals α percent of the total energy of the image. When the threshold is reached, the seams are only removed if they discard at least α percent of the image pixels.

If the image did not reach its target dimension after seam carving is finished cropping is done in the next step. The algorithm identifies an optimal cropping window that is equal or bigger than the target image dimensions while discarding not more than β percent of the total energy. Like in the seam carving part, the cropping is only done if it removes at least β percent of the image pixels. The two steps are repeated iteratively until the image reaches its target size.

IV. EVALUATION

The evaluation consists of two parts. In the first part, we discuss the results and analyze the reliability of our enhanced face saliency technique. In the second part, we show its application in the context of image retargeting.
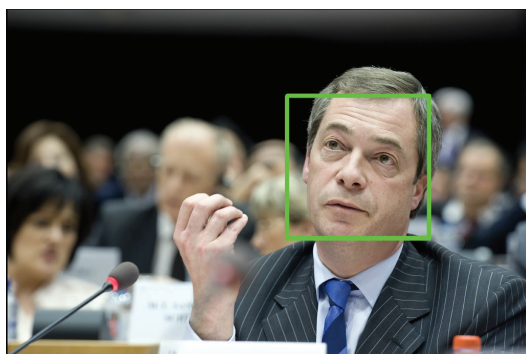
A. Focus Detection

We conducted an evaluation in order to analyze the reliability of our new algorithm. 35 images with a total of 42 in-focus and 46 out-of-focus faces are used in this study. Each image contains at least one face that is in focus

Figure 2. Distinguishing blurred faces (red) from in-focus faces (green) based on strong edges (blue).
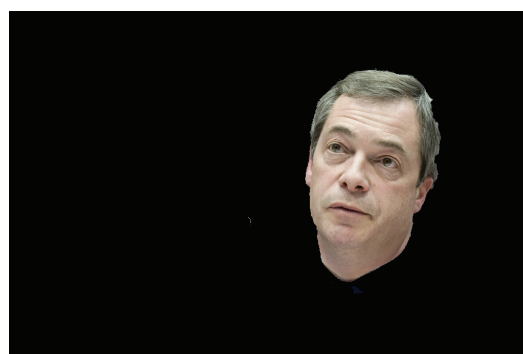

(a) Original face detection

(b) Rectangle extended with safety margin and the pixels outside the rectangle are marked as "background"

(c) Foreground (red) and probably background (blue) labels

(d) Segmented face

Figure 3. Creation of a face mask from the initial detection.

Figure 4. An example of our focus detection algorithm for faces. One face that is in focus (green) and another out-of-focus face (red) are detected.

and one that is out of focus. As the ground truth, the faces in all images were manually marked and classified as in focus or out of focus. If the decision was not clear at least three people classified such a face, and the majority of the classifications was used. We want to point out that the automatic face detection does not recognize all out-of-focus faces that can be found manually, as visualized in Figure 4.

We analyze the percentage of missed and correct faces in a first step. By using multiple cascades, the number of false detections is reduced by 43% while increasing the number of missed faces by 5% compared to using only one cascade. In a second step, all correct face regions are analyzed further. 90% of the detected faces are assigned correctly as in focus or out of focus.

Although the algorithm performs a Gaussian smoothing operation before the focus detection is applied, it is still not completely immune to noise. Areas affected by noise can still have high intensity variations which results in high values of the gradient magnitude. This is the main reason for the 10% false assignments in our the focus detection.

Figure 4 exemplary shows an image where two face regions are detected. For visualization, in-focus faces are marked with a green rectangle and out-of-focus faces are marked with a red rectangle. In this image, the woman in the foreground is clearly the center of attention. If the unsharp face of the woman in the background was also considered in the retargeting algorithm, this would preserve both faces but at the same time would cause severe distortions in other regions of the image. In case of several in-focus faces that are spread all over the image, the result of image

retargeting with our new algorithm is comparable to normal face detection.

B. Application in Image Retargeting

Threshold values of $\alpha = 1\%$ and $\beta = 15\%$ are used for our image retargeting algorithm. All images are reduced to 50% of their original width.

Figure 5 shows some results of the retargeting algorithm. In the top row (a), the shoulder and face of the man in front becomes severely distorted if the detected face in the background is treated like a normal face. With our technique, that face is classified as out of focus and therefore is allowed to be removed. In the second row (b), the face of the woman in front of the couple is unsharp. If this face region is considered in the retargeting with high priority, the target image is suboptimal. In the basketball image (c), many faces are found in the background. These faces distract the algorithm from the player in the foreground. The images with the two men in the last row (d) include detected faces but also a false detection. This incorrect region would be classified as out of focus. Focus detection also helps in this case; it would move the center of attention to the right. As our focus detection algorithm is also able to identify false positives in the background, it further improves the precision of faces that are in focus.

Our new face detection technique enhances the results of the SeamCrop algorithm. A comparision of SeamCrop with other state-of-the-art image retargeting techniques can be found in [4].
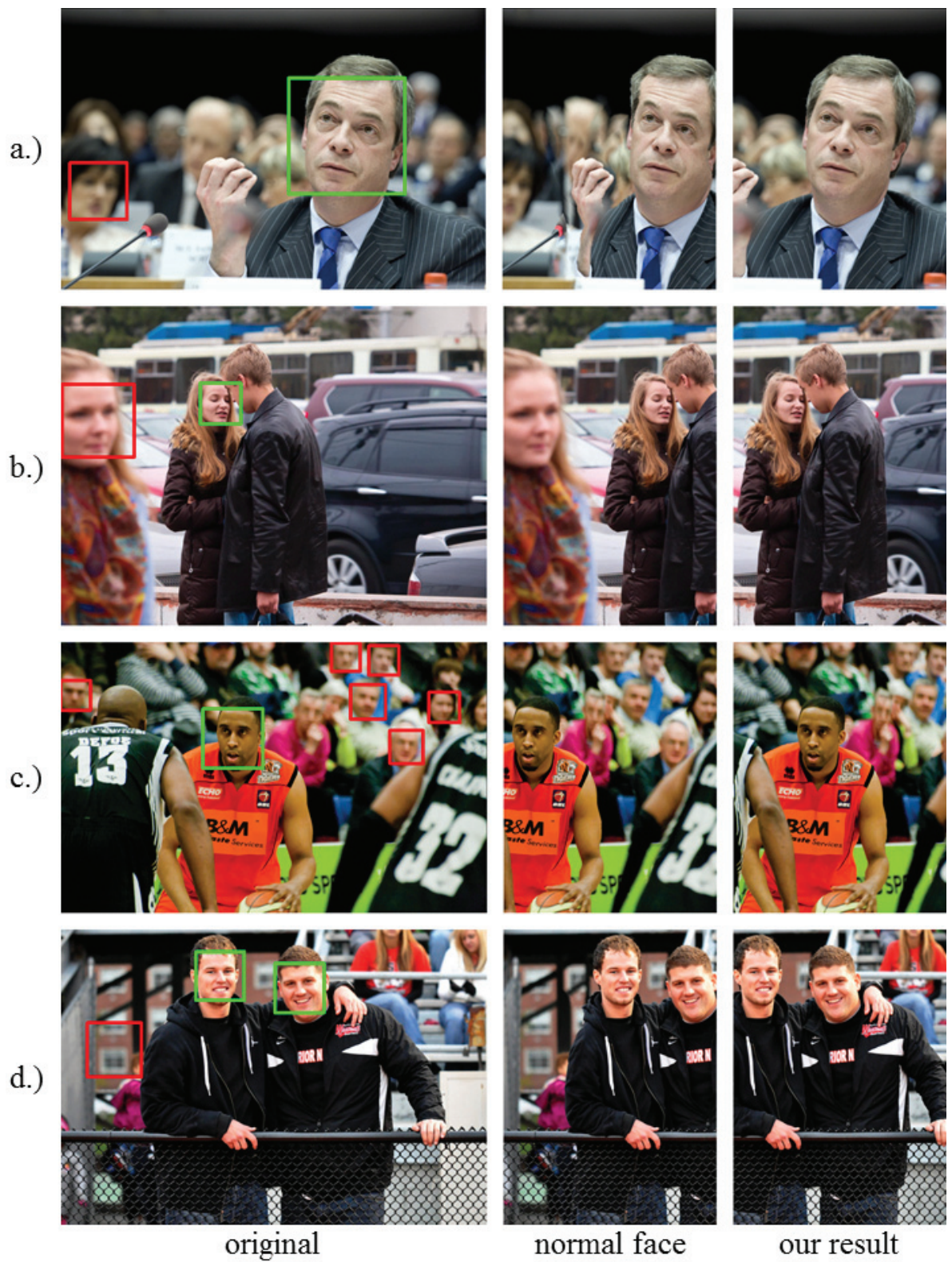
Figure 5. Results of our image retargeting technique: The width of the original image is reduced by 50%. The results of the retargeting operator are compared when using all faces (center) and only faces that are in focus (right).

V. Conclusions

In this paper, we presented a novel technique for distinguishing between faces that are in focus and out of focus. This information is especially useful in the context of image retargeting, where detections of unimportant faces in the background may lead to visual distortions of really important objects. Faces are first detected with the use of multiple cascades. As faces with strong edges are assumed to be in focus, we use the gradient magnitude to classify the faces to be in or out of focus. GrabCut is finally used to segment the faces and create face masks which are provided as input for our image retargeting algorithm. In future work, we would like to extend the focus detection approach to derive depth from images.

References

[1] Michael Rubinstein, Diego Gutierrez, Olga Sorkine, and Ariel Shamir, "A comparative study of image retargeting," in *ACM SIGGRAPH Asia 2010 papers*, New York, NY, USA, 2010, pp. 160:1–160:10, ACM.

[2] Paul Viola and Michael Jones, "Rapid object detection using a boosted cascade of simple features," in *Proceedings of IEEE International Conference on Computer Vision and Pattern Recognition (CVPR)*, Los Alamitos, CA, USA, 2001, vol. 1, pp. 511–518, IEEE Computer Society.

[3] Carsten Rother, Vladimir Kolmogorov, and Andrew Blake, "Grabcut: interactive foreground extraction using iterated graph cuts," in *ACM SIGGRAPH 2004 Papers*, New York, NY, USA, 2004, SIGGRAPH '04, pp. 309–314, ACM.

[4] Johannes Kiess, Benjamin Guthier, Stephan Kopf, and Wolfgang Effelsberg, "SeamCrop for image retargeting," in *Multimedia on Mobile Devices 2012*. 2012, vol. 8304, SPIE.

[5] W. Zhao, R. Chellappa, P. J. Phillips, and A. Rosenfeld, "Face recognition: A literature survey," in *ACM Computing Surveys (CSUR)*. December 2003, vol. 35(4), pp. 399–458, ACM Press.

[6] E. Murphy-Chutorian and M.M. Trivedi, "Head pose estimation in computer vision: A survey," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 31, no. 4, pp. 607 –626, april 2009.

[7] Yoav Freund and Robert E. Schapire, "A decision-theoretic generalization of on-line learning and an application to boosting," in *Proceedings of the Second European Conference on Computational Learning Theory*, London, UK, 1995, pp. 23–37, Springer-Verlag.

[8] Shai Avidan and Ariel Shamir, "Seam carving for content-aware image resizing," *ACM Transactions on Graphics, SIGGRAPH 2007*, vol. 26, no. 3, 2007.

[9] Michael Rubinstein, Ariel Shamir, and Shai Avidan, "Multi-operator media retargeting," *ACM Transactions on Graphics, SIGGRAPH 2009*, vol. 28, no. 3, pp. 1–11, 2009.

[10] Johannes Kiess, Benjamin Guthier, Stephan Kopf, and Wolfgang Effelsberg, "SeamCrop: Changing the size and aspect ratio of videos," in *Proceedings of the 4th Workshop on Mobile Video*, New York, NY, USA, 2012, MoVid '12, pp. 13–18, ACM.

[11] Y. Pritch, E. Kav-Venaki, and S. Peleg, "Shift-map image editing," in *ICCV'09: Proceedings of the 2009 IEEE International Conference on Computer Vision*, Kyoto, Sept 2009, pp. 151–158.

[12] Yu-Shuen Wang, Chiew-Lan Tai, Olga Sorkine, and Tong-Yee Lee, "Optimized scale-and-stretch for image resizing," *ACM Transactions on Graphics*, vol. 27, no. 5, pp. 1–8, 2008.

[13] Yanwen Guo, Feng Liu, Zhi-Hua Zhou, and Michael Gleicher, "Image retargeting using mesh parameterization," *IEEE Transactions on Multimedia*, vol. 11, no. 5, pp. 856–867, 2009.

[14] Al Bovik, *Handbook of Image and Video Processing*, Elsevier Academic Press, San Diego, CA, USA, 2. edition, 2005.

[15] Mark A. Ruzon and Carlo Tomasi, "Alpha estimation in natural images," *Proceedings IEEE Conference on Computer Vision and Pattern Recognition CVPR 2000*, vol. 1, no. June, pp. 18–25, 2000.

---

[2]www.flickr.com

[3]http://creativecommons.org