AUTOFRAMING: A RECOMMENDATION SYSTEM FOR DETECTING UNDESIRABLE ELEMENTS AND CROPPING AUTOMATICALLY PHOTOS

Frank Nielsen *Shigeru Owada* *Yuichi Hasegawa*

Sony Computer Science Laboratories, Incorporated
Fundamental Research Laboratory
Tokyo, Japan

Sony Corporation
Digital Imaging Business Group
Tokyo, Japan

ABSTRACT

In this paper, we present a recommendation system for automatically recentering and cropping digital still pictures that exhibit capturing artefacts. Autoframing images not only yields better visual pictures but more importantly allows us to remove undesirable artefacts such as lens obstructions by fingers, cellphone straps, or back heads. We report on our real-time prototype system that is targeted to consumer digital still cameras.

1. INTRODUCTION

Collecting pictures is the ubiquituous way of keeping souvenirs by freezing past moments and recalling these moments into our memories by simply viewing them. Taking photos is both a user commodity and an art: the depiction of our everyday life surrounding world. Since the early beginning of photography, the quest has been to obtain vivid images nearly undistinguishable from the real-world. Good pictures have fine details, high contrasts, smooth gradations, rich colors, and clarity. Usually, to take a photograph, we need to:

- control the image sharpness by adjusting the lens position which sets the correct focus,
- control the exposure time for getting the right overall contrast,
- stabilize the camera using either a tripod or an optical/electronic stabilizing apparatus to avoid motion blur.[1]

Image processing technologies have been proposed and constantly developed to automatically control these settings: auto-focus (AF), auto-exposure (AE) and optical/digital stabilizer (Antishock, Auto-Stabilizer, AS). Yet, nowadays, we still need to precisely control the field of view to get a good picture.[2] This requirement is actually becoming less and less important as the sensor size of cameras is increasing (nowadays, consumer digital cameras have $\simeq$ 10 million pixels). Thus, it is acceptable to loose 20% pixels if the overall delivered picture is better. In this paper, we describe a short yet efficient autoframing method for detecting undesirable picture regions and automatically cropping pictures (see Figure 1). Ideally, autoframing would require to first determine the intended photo subject(s) and then to compute the photo center, adjust zoom, and select picture offsets so as to preserve the original aspect ratio. Detecting photo subjects[3] in a robust and efficient way is still a major unsolved problem of computer vision (see Section 4). In our system, we rather concentrate on first detecting undesirable elements in images (that can be robustly identified) to autocrop pictures accordingly.

Fig. 1. Overview of the autoframing recommendation system: Once the system has detected undesirable objects in pictures (here, a hand indicated by a bold arrow), it automatically crops the picture while preserving the aspect ratio.

Historically, autoframing has first been studied for panoramic images, also called 360^o images. Sun et al. [1] described a system for selecting region of interests (ROIs) in video panoramas based on feature extractions. Interestingly, their method works directly on compressed MPEG movies but is basically aimed at controlling smoothly a *virtual pinhole camera* in a 360^o panoramic video. Tanaka [2, 3] applied

[1] Motion blur is due to exposure time. Even for perfectly stabilized cameras, scene motion blur remains for moving objects.

[2] Note that not all cameras produce pictures with field of view (fov) perfectly matching the viewfinder fov. Single lens reflex (SLR) cameras provide viewfinders that look through the same lens that exposes the image sensor. Optical viewfinders tend to show only 95% to 98% of the final image.

[3] The photographers' intended objects.

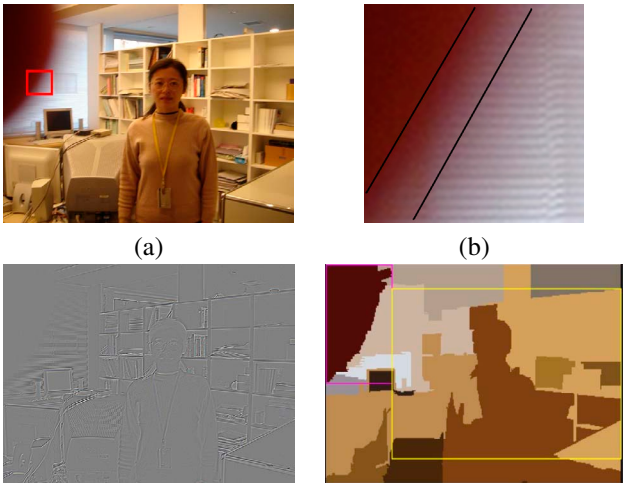


Fig. 2. Autoframing in the finger elimination mode: (a) original image, (b) a close-up showing the softedge (regions bounded by the two black strokes) of the nonfocused finger region, (c) Roberts edge filter and (d) segmentation result with finger bounding box and suggested frame.

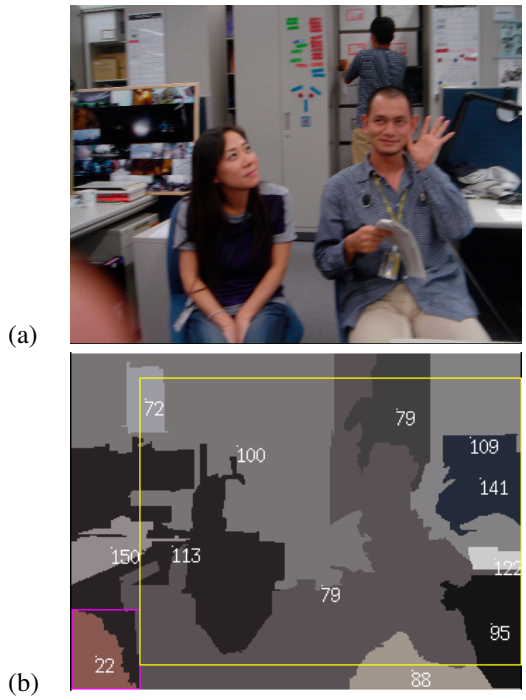


Fig. 3. Scoring segmented regions according to their blurriness: (a) original image, and (b) regions with their regions labeled with scores indicating sharpness (the higher, the sharper). The finger region has the lowest score, and therefore potentially represent an undesirable object (*i.e.*, finger).

for Japanese patents for a system detecting structural elements in images and trimming accordingly panoramas (#JP2001-126070A) or rearranging source picture regions to get more *æsthetic* pictures (#JP2000-200354A, see also [4]). Tanaka uses the edge-flow image segmentation algorithm [5] for preliminarily segmenting the image into homogeneous regions and defines some *attractiveness* evaluation procedure based on oriented Gaussian derivative filters (OGDFs — texture feature descriptor originally presented in [6]). Although ambitious, his selection and reorganization of image elements based on neural networks is supposed fairly unstable in practice. Lawrence applied for a patent [7] (#WO 02/05835 A2) for automatically cropping images using segmentation on blurred images. Segmentation is performed in the YCC color space by first growing areas that have smooth colors and intensities, and then merging regions that are separated by weak edges. Furthermore, Lawrence tries to identify horizon lines in images whenever possible, and get a set of cropping rectangles defining constraints from which the final frame is obtained. (For example, detecting horizon line yields a 1/3-2/3 height æsthetic contraint.)

2. AUTOFRAMING: A DETECTION AND RECOMMENDATION SYSTEM

2.1. Outline

Our proposed autoframing functionality is a *recommendation system*. In most cases, pictures are left as is, and need not be reframed. The recommendation system first detects for each acquired picture whether there are undesirable elements inside the pictures, or not. Undesirable elements are defined as image regions *crossing* the image border and *not in focus* (i.e., blurred). Typical such examples are lens obstructions by photographers' fingers (see Figure 3) or back heads of people in crowds. If such an element is detected the Graphical User Interface (GUI) is activated on the camera panel screen, and one or several recropped pictures are proposed (Figure 1). Note that recently, with the introduction of touch panels on digital still cameras (DSCs), one can also indicate on the camera screen which regions s/he intends to take. However, as far as the authors know, no commercial system relies yet on stable general purpose automatic image segmentation.

2.2. Detecting Obstructions

We first proceed by segmenting the color image. Image segmentation is both useful for detecting potential obstructors and for retrieving image element characteristics such as region colors, areas, textures and shapes. Figure 2(a) shows a source picture with a finger obtrusion. Figure 2(b) is a close-up of the finger part emphasizing on the fact that this segmented part is not in-focus. This area is blurred due to the point spread function characteristics implied by the current lens position. Thus, a contour-based segmentation such as edge flow [5] fails to segment the finger portion as one

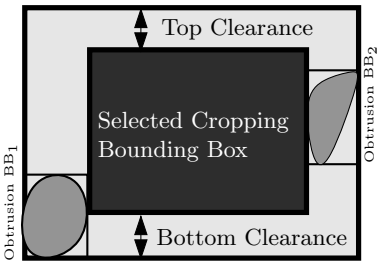


Fig. 4. Choosing the cropping bounding box by setting the top and bottom clearance identical.

single unit. We use the fast area-based linear-time statistical region growing segmentation algorithm described in [8] to segment correctly and robustly such a kind of obtrusions (see Figure 2(d) and Figure 3(b)). For each segmented area, we further evaluate its *sharpness attribute* using a standard edge detector: Roberts cross filter (see [9], pp. 221). First, the Roberts 2×2 matrix convolution filter is applied on the source image (Figure 2(c)) and for each segmented area, we select all pixels that fall within some prescribed threshold to the region boundary. We define the sharpness attributes of the regions as the mean Roberts edge values in these region "rings." Figure 3(b) displays the sharpness attribute for the source image shown in Figure 3(a). Observe that the finger has been correctly segmented as a single region, and that its sharpness value (22) is the lowest among all other regions (all others above 50).

2.3. Recentering and Cropping

Once one or several regions have been identified as undesirable, we consider their bounding boxes (BBs for short), and find a largest bounding box nonoverlapping the obtrusion BBs while preserving the original aspect ratio (usually, either $4/3$ or $16/9$). Figure 2(d) and Figure 3(b) shows the suggested frame in yellow. In those cases, we choose to center vertically the image so that the cropped image *floats equally* from the bottom/top original image border, as shown in Figure 4.

3. IMPLEMENTATION AND RESULTS

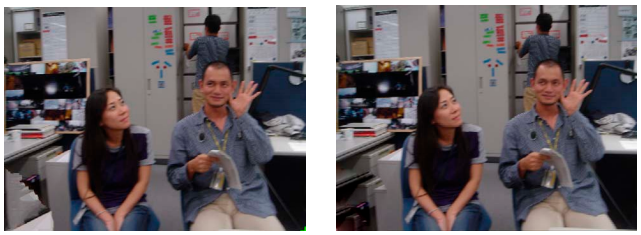
The autoframing system was implemented on a portable Sony VAIO U laptop running Linux®. Figure 5 shows the prototype. We used the fast statistical color image segmentation of [8] and Roberts' filter as a simple edge convolution kernel. A video showing the autoframing prototype in action on a few sample pictures is available on our Web site.

4. CONCLUDING REMARKS AND DISCUSSION

Prior autoframing work has either been motivated by selecting regions of interests in 360-degree cylindrical video panoramas [1], or has been devoted to picture æsthetics [2, 7]. In this paper, we presented a simple and robust recommendation system to first *detect* nonfocused picture obstructors and then *reframe* automatically images. First, our system starts by quickly identifying whether such undesirable elements are part of the raw picture, or not. In case such elements are found, the picture is reframed and presented to the photographer using a graphical user interface on the camera LCD screen. Note that instead of cropping images, we could have used inpainting/texture synthesis methods to automatically overwrite those undesirable regions with surrounding textures [10, 11] (see also [9], Chapter 8). However, observe that since obstructing elements are located on the image border, direct fast raster scan per-pixel texture synthesis fails, as shown in Figure 6. (We need to define properly the pixel ordering to bypass this problem.) Thus although smarter, image completion requires more processing time, is prone to numerous problems and still requires some fair amount of user intervention for perfect filling [11]. Another important extensions of our autoframing system is to crop images according to salient image features. Some of these features like face detection have already been plugged successfully in our recommendation system [12]. Once faces have been detected (position, scale and orientation[4]), we adjust the cropping bounding box so that head positions or sizes fit some criterion rules. We are currently considering other visual awareness or attractiveness measures to further improve the cropping rules of the recommendation system (see [13, 14]). Yet, since fully automatic segmentation or visual attention are difficult to obtain reliably in practice, we further consider to put the user in the loop by providing a convenient *user interface*: the user selects

Fig. 5. Snapshot of our autoframing prototype application demonstrating the automatic finger elimination mode for the `corridor` test image: (a) source image, (b) segmented image with finger detected and region-of-interest box highlighted, and (c) suggested cropped picture.

[4]In practice, face recognition works for faces with attitude ranging from $[-90, 90]$ degrees horizontally and $[-60, 60]$ degrees vertically.



(a) (b)

Fig. 6. Removing the finger part: (a) filling with fully automatic per-pixel texture synthesis yields noticeable artefacts. (b) manually edited using the healing brush of Adobe® Photoshop CS 2®.

on the camera touch screen a few positions by pinpointing image positions using a stylus, and the system instantly refines the segmented images and suggested cropped pictures.

Finally, let us mention that taking pictures still requires some efforts to press the shutter button at the *right time*. We would like to provide autoshutter procedures so as to minimize human/camera lags. Moreover, we would like to design a robust system that constantly acquire pictures and automatically propose photos or video clips based on our preferences. After all, except for artistic photographers, we should not take pictures but only look at them, and more importantly: appreciate them!

Acknowledgements

We would like to thank Motohiko Watanabe of Sony corporation who pointed out to us the related prior work [2, 3, 7], and Professor Richard Nock for his advice. All pictures are used with granted permission.

5. REFERENCES

[1] Xinding Sun, Jonathan Foote, Don Kimber, and B. S. Manjunath, "Region of interest extraction and virtual camera control based on panoramic video capturing," *IEEE Transactions on Multimedia*, vol. 7, no. 5, pp. 981–990, 2005.

[2] Shoji Tanaka, *Structural image information retrieval, recomposition of subjects in images and image processing apparatus*, Japan Patent Office, 1999, P2000-200354A (keywords: Neural network, dynamic symmetry and image recomposition).

[3] Shoji Tanaka, *Attractiveness extraction and automatic framing apparatus*, Japan Patent Office, 1999, P2001-126070A (keywords: Autoframing using attractiveness and panorama trimming).

[4] Shoji Tanaka, Yuichi Iwadate, and Seiji Inokuchi, "An attractiveness evaluation model based on the physical features of image regions.," in *International Conference on Pattern Recognition (ICPR)*, 2000, pp. 2793–2796, Translated from Denshi Joho Tsushin Gakkai Ronbunshi A, Vol. J83-A, No. 5, pp. 576-588, 2000.

[5] Wei-Ying Ma and B. S. Manjunath, "Edge flow: A framework of boundary detection and image segmentation," in *International Conference on Computer Vision and Pattern Recognition (CVPR)*, 1997, pp. 744–749.

[6] B. S. Manjunath and W. Y. Ma, "Texture features for browsing and retrieval of image data," *IEEE Trans. Pattern Anal. Mach. Intell. (TPAMI)*, vol. 18, no. 8, pp. 837–842, 1996, DOI 10.1109/34.531803.

[7] Richard A. Lawrence, *Automated Cropping of Electronic Images*, World Intellectual Property Organization, 2000, P2000-WO0205835A3 (keywords: segmentation, horizon line).

[8] Richard Nock and Frank Nielsen, "Statistical region merging," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 26, no. 11, pp. 1452–1458, 2004, DOI 10.1109/TPAMI.2004.110.

[9] Frank Nielsen, *Visual Computing: Geometry, Graphics, and Vision*, Charles River Media/Thomson Delmar Learning, 2005, ISBN 1584504277.

[10] Li-Yi Wei and Marc Levoy, "Fast texture synthesis using tree-structured vector quantization," in *Proceedings of the 27th annual conference on Computer graphics and interactive techniques (SIGGRAPH)*, 2000, pp. 479–488.

[11] Jian Sun, Lu Yuan, Jiaya Jia, and Heung-Yeung Shum, "Image completion with structure propagation.," *ACM Trans. Graph. (SIGGRAPH)*, vol. 24, no. 3, pp. 861–868, 2005, DOI 10.1145/1073274.

[12] Kohtaro Sabe and Ken'ichi Idai, "Real-time multi-view face detection using pixel difference feature," *Symposium on Sensing via Imaging Information (SSII)*, 2004.

[13] Song Liu, Liang-Tien Chia, and Deepu Rajan, "Attention region selection with information from professional digital camera," in *Proceedings of the 13th Annual ACM International Conference on Multimedia (MM)*, New York, NY, USA, 2005, pp. 391–394, ACM Press, DOI 10.1145/1101149.1101233.

[14] Yiqun Hu, Deepu Rajan, and Liang-Tien Chia, "Robust subspace analysis for detecting visual attention regions in images," in *Proceedings of the 13th Annual ACM International Conference on Multimedia (MM)*, New York, NY, USA, 2005, pp. 716–724, ACM Press, DOI 10.1145/1101149.1101306.