A Novel Gradient Induced Main Subject Segmentation Algorithm for Digital Still Cameras

Serene Banerjee and Brian L. Evans

Embedded Signal Processing Laboratory, Center for Perceptual Systems
The University of Texas at Austin, Austin, TX 78712-1084 USA
{serene,bevans}@ece.utexas.edu

Abstract

When taking pictures, professional photographers employ a variety of composition rules. In automating these rules, it is often first necessary to detect and segment the main subject. We propose an detection and segmentation algorithm that leverages the optics in a digital still camera. Based on where the user points the camera, an auto-focus filter first puts the main subject in focus and takes a picture. Then, we open the shutter aperture to diffuse light from objects that are out-of-focus, which blurs the background, and take a second picture. Using the second picture, the resulting difference in the frequency content of the main subject and the background image is then used by the proposed algorithm to detect and segment the main subject. The algorithm does not depend on prior knowledge of the indoor/outdoor setting or scene content. Algorithm complexity is similar to that of a 5×5 *filter.*

1 Introduction

In consumer photography, several different pictures of the same subject can be taken by changing camera settings. Some of these pictures could be made more appealing by following photographic composition rules, such as *rule of thirds*, *avoidance of merges*, and *background blurring* [1]. In order to automate photographic composition rules, the main subject in the photograph needs to be segmented. Knowledge of the main subject may also be useful for image understanding and enhancement, and constrained transmission.

This paper proposes a low-complexity algorithm for automatic main subject detection during image acquisition. The algorithm leverages an auto-focus filter and software-controlled shutter aperture, which are found on digital still cameras. Given where the user is pointing the camera, the auto-focus filter puts the main subject in focus [2, 3]. Next, the shutter aperture is widened to blur the background. The blurring occurs because the light from out-of-focus objects does not converge as sharply as from objects in focus. By utilizing the significant difference in frequency content of the in-focus and background regions, the proposed algorithm detects the main subject using filtering, edge detection, and contour smoothing.

The significant contributions are: (1) automated detection of focused regions, and (2) automated segmentation of the main subject based on knowledge of the focused regions. The detection-segmentation approach does not depend on prior knowledge of the scene setting or scene content. No training is required. Matlab code for this paper can be found at

www.ece.utexas.edu/~bevans/papers/2003/stillCameras

2 Background

Luo, Etz, Singhal, and Gray [4, 5] detect the main subject using a Bayes net framework. The algorithm is performance-scalable so that it need not be reconfigured for different sets of images, and involves (a) region segmentation, (b) perceptual grouping, (c) feature extraction, and (d) probabilistic reasoning and training. An initial segmentation is obtained based on the homogeneous properties of the image such as color and texture. Then, a probability density function for the main subject location is estimated from training data. Finally, the probability density function estimate is applied to the unknown test set to guess the main subject. Their method requires training time, and has high implementation complexity.

Aizawa, Kodama, and Kubota [6, 7] propose to use two pictures for foreground segmentation. One picture has the foreground in focus, and the other has the background in focus. This pair is taken to cre-

ate an image that is focused at an arbitrary distance. Their approach requires manual operation to put the background in focus, and incurs a delay in changing the focal length.

Wang, Li, Gray, and Wiederhold [8, 9] use statistics of the high-frequency wavelet coefficients of an image to segment the focused regions from the defocused ones. Their context-dependent approach to classify individual blocks of the image is computationally intensive as it requires computing a multi-level wavelet transform, feature extraction, and postprocessing to smooth the boundary.

Won, Pyan, and Gray [10] develop an iterative algorithm based on variance maps of the image to yield a more accurate segmentation than that of the wavelet-based approach above [8, 9]. However, the implementation complexity of variance map approach is high because the optimal boundary is computed iteratively and then refined by the watershed algorithm.

3 Mathematical Formulation

Let i be an intensity value in an n-dimensional Euclidean space, $\Re^n$, such that $i \in F$, where F is the *image domain*. Let F_o and F_b represent the object and background classes, respectively, with $F_o \subset F$ and $F_b \subset F$. The objective is to separate F_o from F_b.

Suppose $i \in \Re^n$ is mapped to an n-dimensional space, Ω^n, induced by ∇i, where ∇ is the gradient operator. An n-tuple s in Ω^n is mapped such that $s \in G$, where G is the *gradient domain*. The gradient domain, G can be further partitioned as $G_H(s)$ and $G_L(s)$ domains. These can be defined as:

$$G_H(s) = \{s | s \geq \delta \text{ and } s \in G\}$$

$$G_L(s) = \{s | 0 \leq s < \delta \text{ and } s \in G\}$$

where δ is the threshold. $G_H(s)$ and $G_L(s)$ represent the high and low frequency domains, respectively.

Now $G_H(s) \mapsto F_H(i)$ and $G_L(s) \mapsto F_L(i)$, where subscripts H and L associated with F correspond to transformation from high and low frequency domains, respectively. The mapping of G_H to F_H and G_L to F_L requires similar set of transformation and $G(s) = G_H(s) + G_L(s)$ in the intensity domain:

$$F(i) = aF_H(i) + bF_L(i) \tag{1}$$

Here, $b = ka$ and k is a constant.

Now, $i \in F$ lies in $F_H(i)$ or $F_L(i)$, with probabilities a and b, respectively, so $a + b = 1$. Possible choices of a and b could be $\frac{1}{(k+1)}$ and $\frac{k}{(k+1)}$, respectively. Then,

$$F_H(i) - F(i) = \frac{k}{k+1} \; (F_H(i) - F_L(i)) \tag{2}$$

For the digital still camera application, the main subject class, F_o, is in focus, and the background class, F_b, is blurred by widening the shutter aperture. $F_H(i) - F(i)$ will have sharper gradients around F_o and smoother gradients around F_b. Segmentation of F_o is induced by this difference of gradient information.

To generate $F_H(i)$ and $F_L(i)$ in the $\Re^n$ domain, high and lowpass filters can be designed, respectively. For the highpass filter, the criterion will be to select the frequencies, so that $\nabla i > \delta$. Similarly, the lowpass filter will have frequencies so that $\nabla i < \delta$. The choice of filter coefficients will determine its characteristics, and the filter can be designed adaptively.

4 Proposed Algorithm

For the 2-D case, the above conditions are met with an image sharpening filter as modeled in Fig. 1. So,

$$g(x, y) = I(x, y) - I_{smooth}(x, y) \tag{3}$$

and

$$I_{sharp}(x, y) = I(x, y) + k\, g(x, y) \tag{4}$$

Thereby,

$$I(x, y) = \frac{1}{k+1} I_{sharp}(x, y) + \frac{k}{k+1} I_{smooth}(x, y) \tag{5}$$

and

$$I_{sharp}(x, y) - I(x, y) = \frac{k}{k+1} \left(I_{sharp}(x, y) - I_{smooth}(x, y)\right) \tag{6}$$

Subtracting the user-intended image from the sharpened image generates an edge map in which the edges around the main subject are sharper than the background edges. Hence, the problem of segmenting the main subject reduces to separating the regions with the sharper edges from the regions with smeared ones. For this task, we employ a 3×3 sharpening filter:

$$\frac{1}{1+\alpha} \begin{bmatrix} -\alpha & \alpha - 1 & -\alpha \\ \alpha - 1 & \alpha + \beta & \alpha - 1 \\ -\alpha & \alpha - 1 & -\alpha \end{bmatrix}$$

Parameters α and β define the shape of the frequency response. We chose $\alpha = 0.2$ and $\beta = 5$. However, the filter characteristics could be adapted according to the strength of the image features. For example, an image having relatively weak edge features could be processed by a filter having a lower cut-off and greater span in the spatial domain, e.g. a 7×7 filter.

In detecting the strong edges from the filtered image, the Canny edge detector [11] gives good results

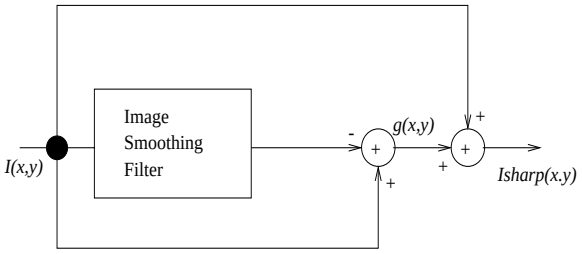


Figure 1. Model for an image sharpening filter

in identifying the strong edges, by first smoothing the difference image with a Gaussian filter and then detecting the gradient of the smoothed difference image. To separate the strong edges in the focused parts from the weak edges in the out-of-focus parts, the hysteresis threshold of 0.3 for the Canny edge detector worked well for the test images. The Laplacian of Gaussian edge detector [12] did not give as good results as the Canny edge detector, as the Laplacian of Gaussian detector picks the zero crossings of both the strong and weak edges. The Canny edge detector performs better than Roberts, Sobel, and Prewitt edge detectors.

The output of edge detection can be fed to a contour detection framework to close the boundary. To determine the closed boundary, the traditional snake [13] algorithm and its direct descendents fail to track the concavities in the contour or require the initial control points to be placed near the actual contour. This limits its automated application to natural images. Instead the gradient vector flow [14, 15] algorithm, which is guided by the diffusion of the gradient vectors from the edge map of the image, is a better choice as it requires no initialization in terms of control points and has a higher capture range with the ability to track image contour concavities.

5 Results and Complexity

The proposed algorithm is shown in Fig. 2. Figs. 3(a) and 4(a) show background blur achieved by a wider shutter aperture, while the main subjects are in focus. The results of locating the main subjects before contour closing are shown in Figs. 3(b) and 4(b), respectively. Figs. 3(c) and 4(c) show the detected main subject mask.

The RGB color image is converted to intensity by

$$I = (R + G + B)/3 \text{ or } I = (R + 2G + B)/4 \qquad (7)$$

The former step requires 2 multiply-accumulates,

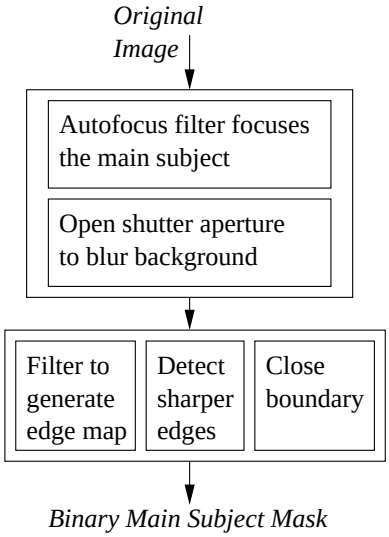


Figure 2. Proposed automated main subject detection algorithm for digital still camera

which matches a programmable digital signal processor well. The later, which requires 2 adds, a shift left by one bit (multiplication by 2) and a shift right by two bits (division by 4), matches a digital hardware implementation well. Shifts can be used here because RGB values are non-negative.

The sharpening operation convolves the image with a 3×3 filter, which would require 9 multiply-accumulates per pixel for the sharpening and difference calculation. Canny edge detection first smoothes the image in order to lower the noise sensitivity, then computes a gradient, and finally suppresses the non-maximum pixels using two thresholds. The smoothing and the gradient computation takes 9 multiply-accumulates, assuming a 3×3 pre-computed filter kernel that is the derivative of a Gaussian mask. The nonmaximum suppression step requires 2 comparisons per pixel. The two 3×3 filters can be cascaded to a 5×5 filter to reduce the number of memory accesses per pixel. This requires 5 memory reads per pixel.

As the exact implementation of the gradient vector flow algorithm to close the contour is computationally intensive, we propose to use an approximation. From the map of the detected sharper edges, the pixel position of the first "ON" pixel from the left and the right boundaries of the image is calculated. Every pixel in-between these two pixels is turned "ON". This approximation detects the convex parts correctly, but fails at the concavities in the shape of the main subject. The approximate procedure requires 2 compar-

isons per pixel. The generated mask is written back with 1 memory access operation per pixel.

Thus, the main subject mask can be generated with 18 multiply-accumulates, 4 comparisons and 6 memory accesses per pixel. As digital still cameras use approximately 160 digital signal processor instruction cycles per pixel, the main subject can be detected with minimal computational overhead.

6 Conclusion

This paper proposes an approach for detecting and segmenting the main subject during image acquisition. First, the user takes a picture as per usual, in which an auto-focus filter has put the main subject into focus. Immediately thereafter, the shutter aperture is widened to blur the background, and a second picture is taken. Finally, the background-blurred image is processed by the proposed algorithm to compute the main subject mask. In the case that the subject or camera moves during the acquisition of the two pictures, image registration may be needed before the main subject mask can be applied to the user-intended image.

The proposed algorithm is applied in the spatial domain, employs two 3×3 filters and a few threshold operations, and uses only fixed-point arithmetic. Only one pass is made over the image. Computational complexity is similar to that of a 5×5 filter.

The method in [6, 7] processes two pictures to segment the foreground, whereas the proposed method only uses one. Unlike a Bayes net approach [4, 5], the proposed method does not require training. The proposed algorithm avoids the iterations and computationally intensive wavelet transform in [8, 9]. The proposed algorithm could be transformed to segment the main subject in the compressed domain.

References

[1] Kodak, *How to Take Good Pictures: A Photo Guide by Kodak*. Ballantine, Sept. 1995.

[2] N. N. K. Chern, P. A. Neow, and J. M. H. Ang, "Practical Issues in Pixel-Based Autofocusing for Machine Vision," in *Proc. IEEE Int. Conf. on Robotics and Automation*, vol. 3, pp. 2791–2796, May 2001.

[3] C. H. Park, J. H. Paik, Y. H. You, H. K. Song, and Y. S. Cho, "Auto Focus Filter Design and Implementation Using Correlation between Filter and Auto Focus Criterion," in *Proc. IEEE Int. Conf. on Consumer Electronics*, pp. 250–251, June 2000.

[4] S. P. Etz and J. Luo, "Ground Truth for Training and Evaluation of Automatic Main Subject Detection," in *Proc. SPIE Conf. on Human Vision and Electronic Imaging*, vol. 3959, pp. 434–442, Jan. 2000.

[5] J. Luo, S. P. Etz, A. Singhal, and R. T. Gray, "Performance-Scalable Computational Approach to Main Subject Detection in Photographs," in *Proc. SPIE Conf. on Human Vision and Electronic Imaging*, vol. 4299, pp. 494–505, Jan. 2001.

[6] K. Aizawa, K. Kodama, and A. Kubota, "Implicit 3D Approach to Image Generation: Object-Based Visual Effects by Linear Processing of Multiple Differently Focused Images," in *Proc. Int. Workshop. on Theoretical Foundations of Computer Vision*, pp. 226–237, Mar. 2000.

[7] K. Aizawa, K. Kodama, and A. Kubota, "Producing Object-Based Special Effects by Fusing Multiple Differently Focused Images," *IEEE Trans. on Circuits and Systems for Video Technology*, vol. 10, pp. 323–330, Mar. 2000.

[8] J. Li, J. Z. Wang, R. M. Gray, and G. Wiederhold, "Multiresolution Object-of-Interest Detection for Images with Low Depth of Field," in *Proc. IEEE Int. Conf. on Image Analysis and Processing*, pp. 32–37, Sept. 1999.

[9] J. Z. Wang, J. Li, R. M. Gray, and G. Wiederhold, "Unsupervised Multiresolution Segmentation for Images with Low Depth of Field," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 23, pp. 85–90, Jan. 2001.

[10] C. S. Won, K. Pyan, and R. M. Gray, "Automatic Object Segmentation in Images with Low Depth of Field," in *Proc. IEEE Int. Conf. on Image Proc.*, pp. 805–808, Sept. 2002.

[11] J. Canny, "A Computational Approach to Edge Detection," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 8, pp. 679–698, Nov. 1986.

[12] D. Marr and E. Hildreth, "Theory of Edge Detection," in *Proc. Royal Society of London*, pp. 187–217, 1980.

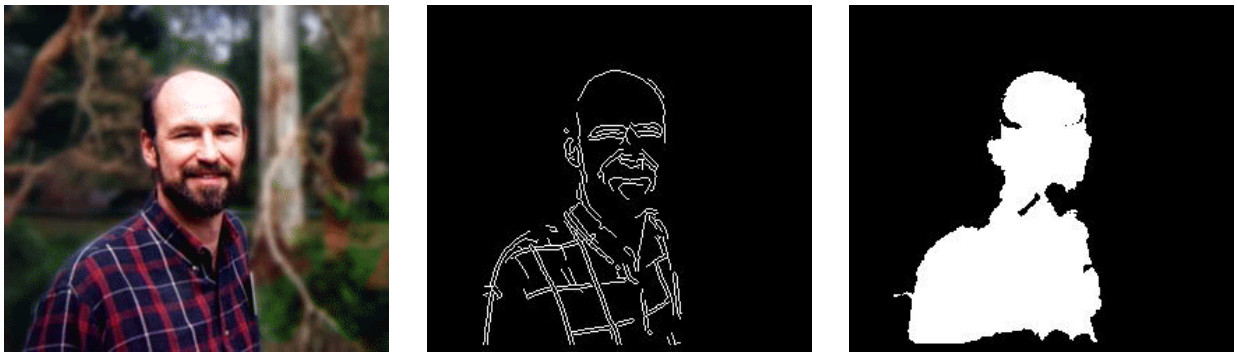
[13] M. Kass, A. Witkin, and D. Terzopoulos, "Snakes: Active Contour Models," *Int. Journal of Computer Vision*, vol. 1, pp. 321–331, 1987.

[14] C. Xu and J. L. Prince, "Snakes, Shapes, and Gradient Vector Flow," *IEEE Trans. on Image Processing*, vol. 7, pp. 359–369, Mar. 1998.

[15] C. Xu, J. A. Yezzi, and J. L. Prince, "A Summary of Geometric Level-Set Analogues for a General Class of Parametric Active Contour and Surface Models," in *Proc. IEEE Workshop on Variational and Level Set Methods in Computer Vision*, pp. 104–111, July 2001.

(a) (b) (c)

Figure 3. Detecting the main subject, the man and the child, which are in focus: (a) Digital image with background blur from large shutter aperture; (b) Rough outline of main subject; (c) Detected main subject mask.



(a) (b) (c)

Figure 4. Detecting the main subject, the man, which is in focus: (a) Digital image with background blur from large shutter aperture; (b) Rough outline of main subject; (c) Detected main subject mask