Segmenting focused objects in visual images

D. M. Tsai and H. J. Wang
Machine Vision Lab.
Department of Industrial Engineering and Management
Yuan-Ze University, Chung-Li, Taiwan, R.O.C.
E-mail: iedmtsai@saturn.yzu.edu.tw
(Figures are not available in this report.)

1. INTRODUCTION

Image segmentation that partitions a given image into meaningful regions is an important task of image analysis for representation, interpretation and recognition. If the image consists of man-made objects in well controlled situations such as industrial environments, segmentation can be easy. In applications such as target acquisition in natural scenes, there is no control of the environment, and segmentation becomes extremely difficult.

There exists two major approaches to image segmentation : edge-based and region-based methods (Hsiao et al., 1989). The edge-based algorithms are based on the discontinuity property of gray-level values. It first detects isolated points, lines or edges in an image based on abrupt changes in gray level, and then connects these local discontinuities to form longer boundaries. The region-based methods rely on the similarity property of gray-level values. Areas of the image with homogeneous properties are conventionally found by thresholding (Otsu, 1979; Tsai and Chen, 1992), region growing (Besl and Jain, 1988; Jain et al., 1995), region splitting and merging (Horowitz and Pavlidis, 1976; Browning and Tanimoto, 1982), and clustering (Fukaka, 1980; Jain and Dubes, 1988). The success of these methods is highly dependent on the homogeneity of gray-level intensity or texture.

The detection of object in a complex background is an unresolved problem (Jain et al., 1997). In this paper, we focus on the extraction of focused objects in a visual image that contains both complicated foreground and background. The conventional edge-based or region-based algorithms cannot be directly applied to the segmentation of focused objects in a complicated image such as a car on the street or a man in front of a building. Both objects in focus and objects in the background or foreground result in high gradient magnitude at edge pixels using the edge detection operations. Texture-based segmentation (Pal and Pal, 1993) that involves identifying regions with uniform texture may not extract the complete region of the focused object since it may contain many heterogeneous textures.

The method proposed in this paper is an unsupervised edge-based segmentation. We do not need *a priopri* knowledge for the objects of interest as long as they are focused in the image. It can be observed that in the image formed by an optical system, objects at a particular distance from the lens will be focused, whereas objects at other distances will be blurred by varying degrees depending on their distances. As the distance between the imaged point and the surface of exact focus increases, the imaged object becomes progressively more defocused. The edges of focused objects will be sharp and concentrated, but the edges of background or foreground objects will be blurry and scattering. By measuring the amount of defocus (blur) for each edge pixel in the observed image, the point on the boundary of the focused object can be detected.

The basic framework of our approach is as follows. The observed gray-level image is

first converted to a gradient image using the Sobel edge operator. For every edge point of interest in the gradient image, the amount of defocus at the pixel is measured by the proportion of the edge region in a small neighborhood window using the moment-preserving method. Only the pixels with small amount of defocus, which correspond to the edge points of focused objects, are retained. This set of detected pixels seldom characterizes a boundary completely. Therefore, an edge linking procedure follows to assemble the detected edge pixels of focused objects into closed boundaries. Pixels outside the closed boundaries are declared as background. Finally, a region-filling procedure is carried out to eliminate all pixels outside the closed boundaries. The regions of focused objects in the image are isolated accordingly.

This paper is organized as follows: Section 2 overviews the optical geometry of the depth formula that determines the distance between a point object and the lens as a function of the amount of defocus. The moment-preserving algorithm for evaluating the amount of defocus at edge pixels, edge-linking and region-filling processes for isolating regions of focused objects from the background are then described. Section 3 presents experiments on applying the proposed segmentation scheme for the extraction of focused objects in a variety of visual images. The paper is concluded in Section 4.

2. SEGMENTATION BASED ON EDGE BLUR

2.1 The optical geometry

For a convex-lens camera with a lens of focal length F, the relation between the position

of a point in the scene and the position of its focused image is given by the well-known lens law

$$\frac{1}{v}+\frac{1}{u}=\frac{1}{F} \tag{1}$$

where u is the distance of the point object from the lens and v is the distance of the focused image from the lens.

Let o be a point object on a visible surface in the scene. If o is not in focus then it gives rise to a circular image called the blur circle on the image plane (see Figure 1). Let the diameter of the blur circle be denoted by d. Pentland (1987) has shown that the relationship between the depth u of an object point and the diameter d of the blur circle is given by

$$u=\frac{Fv_0}{v_0-F-df} \qquad \text{for} \quad v_0>v \tag{2.a}$$

$$u=\frac{Fv_0}{v_0-F+df} \qquad \text{for} \quad v_0<v \tag{2.b}$$

where v_0 is the distance between the lens and the image plane, and f is the f-number (aperture) of the lens system. As the sensor displacement increases (i.e., v_0-v), the defocusing diameter d increases. If the image detector is behind the focused image (i.e., $v_0>v$), the depth u is evaluated by eq.(2.a). The object point o is in the background. If the image detector is in front of the focused image (i.e., $v_0<v$), the depth u is then evaluated by eq.(2.b). The object point o is in the foreground. For a given lens system, the parameters F, v_0 and f can be considered as constants. Therefore, eq.(2) shows that the defocus d is an unique indicator for the depth u. The depth formula of eq.(2) can be rewritten in a condensed form (Lai and Fu, 1992) as follows:

$$u = \frac{P}{Q \pm d} \tag{3}$$

where $P = Fv_0 / f$, $Q = (v_0 - F)/ f$, and P and Q are constants with respect to a given camera setting. When the point object is in the perfect focused distance u_0, the amount of defocus d should be zero, i.e., $u_0 = P/Q$. The depth formula in eq.(3) reveals that the blur circle d is gradually increasing as the point object either moves away or moves toward the lens from the focused distance u_0. Therefore, a point object in the background or foreground will have larger blur circle d than the point object at the position of perfect focus.

2.2 The measure of defocus

The conventional blur estimation algorithms (Pentland, 1987; Lai and Fu, 1992) generally model the blurred edge as the result of convolving a focused image with a point spread function that is assumed to be a Gaussian distribution with spatial parameter s. The parameter s is used as the measure of defocus. It is solved in a very complex way using iterative nonlinear search techniques. In this study, we use a more straightforward approach to find the amount of defocus by the moment-preserving technique. The observed image is initially converted into a gradient image using the Sobel edge operator so that edge pixels have large gradient magnitude, and non-edge pixels have approximately zero gradient magnitude. For each edge point of interest, the proportion of the edge region p_e (i.e., the region with high gradient magnitude) with respect to a small neighborhood window in the gradient image is computed using the moment-preserving principle. A focused edge will result in small p_e, whereas a defocused edge will yield large p_e. p_e increases as the relative distance between

the imaged point and the surface of exact focus increases. Therefore, p_e is a measure for the amount of defocus. The estimation procedure for the proportion of edge region p_e in a small window is described in detail as follows.

Let $f(x, y)$ be the gray-level of a pixel at (x, y) in the observed image. The gradient of $f(x, y)$ is given by

$$\nabla f(x, y) = \begin{bmatrix} g_x \\ g_y \end{bmatrix} \tag{4}$$

where $$g_x = \sum_{j=-1}^{1} \sum_{i=-1}^{1} f(x+i, y+j) \cdot w_x(i, j)$$

$$g_y = \sum_{j=-1}^{1} \sum_{i=-1}^{1} f(x+i, y+j) \cdot w_y(i, j)$$

$w_x(i, j)$ and $w_y(i, j)$ are the 3×3 horizontal and vertical Sobel edge operators (Gonzalez and Woods, 1992). The magnitude of the gradient is defined by

$$g(x, y) = |\nabla f(x, y)| = \left[g_x^2 + g_y^2\right]^{1/2} \tag{5}$$

$g(x, y)$ forms the gradient image of the observed image $f(x, y)$. Figure 2(a) demonstrates the observed gray-level image of a triangular block. The camera is focused on the table surface where the block is placed. Figure 2(b) presents the resulting gradient image of the observed image. It shows that the focused slope close to the table surface results in thin and sharp edges, and the defocused slope close to the lens yields thick and scattering edges. The width of edges increases from lower-left to upper-right in the gradient image as the slope of the triangular block is defocused progressively from the base to the top. The width of edges in the gradient image can be a description for the diameter of blur circle d.

As observed in Figure 2(b) , the gradient image can be divided into two regions, the bright region that represents the edges with high gradient magnitudes, and the dark region that represents the interior portions of objects or the background with low gradient magnitudes.

Let the gradient image $g(x, y)$ defined in a local neighborhood window be the real-world version of an ideal gradient image that consists of only two homogeneous regions, the bright region with a uniform gradient magnitude h_e, and the dark region with a uniform gradient magnitude h_b. Denote p_e and p_b by the proportions of the bright region and the dark region, respectively, in the ideal gradient image. Note that $h_e > h_b$, $0 \le p_e$, $p_b \le 1$ and $p_e + p_b = 1$. For a given edge point at (x, y), the first three moments of $g(x, y)$ are given by

$$m_j = \frac{1}{n} \sum_{(s,t) \in N(x,y)} [g(s,t)]^j \quad , \text{ j } = 1,\ 2,\ 3 \tag{6}$$

where $N(x, y)$ is the neighborhood window that consists of neighboring points around (x, y), and n is the total number of pixels in the window.

By preserving the first three moments in both real-world gradient image $g(x, y)$ and the ideal gradient image, we can obtain four equations as follows:

$$p_e \cdot h_e^1 + p_b \cdot h_b^1 = m_1 \tag{7.a}$$

$$p_e \cdot h_e^2 + p_b \cdot h_b^2 = m_2 \tag{7.b}$$

$$p_e \cdot h_e^3 + p_b \cdot h_b^3 = m_3 \tag{7.c}$$

and

$$p_e + p_b = 1 \tag{7.d}$$

There exists a closed-form solution for the four unknown variables p_e, p_b, h_e and h_b, which are given by (Tsai, 1985)

$$h_b = \frac{1}{2}\left[-c_1 - \left(c_1^2 - 4c_0\right)^{1/2}\right] \tag{8.a}$$

$$h_e = \frac{1}{2}\left[-c_1 + \left(c_1^2 - 4c_0\right)^{1/2}\right] \tag{8.b}$$

$$p_b = \begin{vmatrix} 1 & 1 \\ m_1 & h_e \end{vmatrix} \Big/ \begin{vmatrix} 1 & 1 \\ h_b & h_e \end{vmatrix} \tag{8.c}$$

$$p_e = 1 - p_b \tag{8.d}$$

where $$c_0 = \begin{vmatrix} -m_2 & m_1 \\ -m_3 & m_2 \end{vmatrix} \Big/ \left(m_2 - m_1^2\right)$$

$$c_1 = \begin{vmatrix} 1 & -m_2 \\ m_1 & -m_3 \end{vmatrix} \Big/ \left(m_2 - m_1^2\right)$$

The value of p_e, $0 \le p_e \le 1$, gives the proportion of edge region in the neighborhood window. p_e is an indicator for the diameter of blur circle d. The value of p_e increases as the amount of defocus increases.

Figure 3(a) presents the original gray-level image of a doll in complex foreground and background. Figure 3(b) illustrates the corresponding gradient image of the doll image. Note that the edges of the focused doll are sharp and have thin width, whereas the edges of objects in both foreground and background are scattering and have thick width. Figure 3(c) shows the p_e values ($\times 255$) of edge pixels as an intensity function, where brightness is proportional to the magnitude of p_e. It shows that the focused edges of the doll are darker (i.e., smaller p_e

values) than the blurred edges in the foreground and background. By selecting a proper threshold T_d for the p_e values, most blurred edges in the foreground and background can be effectively eliminated, and only focused edges with p_e values less than the threshold T_d are retained, as seen in Figure 3(d).

2.3 Edge-linking processes

The set of edge pixels described by most edge detection techniques seldom characterizes a boundary completely because of noise, nonuniform illumination and other effects that introduce spurious intensity discontinuities (Gonzalez and Woods, 1992). Since the focused edges are extracted based on the edges detected in the gradient image, the resulting focused edge pixels generally cannot describe the complete boundary of a focused object, as seen in Figure 3(d). A region cannot be declared a segment unless it is completely surrounded by edge pixels. Therefore, the focused edge detection procedure must be followed by an edge-linking process to assemble focused edge pixels into closed boundaries.

While it is possible to use elegant and yet complicated edge-linking methods such as graph-theoretic techniques (Gonzalez and Woods, 1992), curve fitting (Goshtasby and Shyu, 1995), and edge following (Zhou et al., 1989; Xie 1992; Xie and Thonnat 1992), we propose a simple and straightforward edge linking procedure to evaluate the feasibility of the defocus measurement approach for focused object segmentation. The proposed edge linking scheme consists of three processes: dilation, thinning and line linking. Simple dilation (Gonzalez and Woods, 1992) of the selected edge pixels is first carried out to close small gaps before

performing the linking operation that connects edge segments in large spacing. The dilation process is performed on the thresholded p_e image (such as the one in Figure 3(d)) by a simple 3×3 structuring element for 10 iterations. The selection of 10 iterations is based on the preliminary experimental results of 25 test images. The dilation result of the thresholded p_e image in Figure 3(d) is illustrated in Figure 3(e).

Following the dilation process, we perform a simple thinning procedure (Gonzalez and Woods, 1992) so that the focused edges are 1-pixel wide. Figure 3(f) shows the thinning result of the dilated image in Figure 3(e). At this stage, the focused edge segments in small gaps have been connected. To ensure that focused objects are bounded by closed boundaries, the final edge linking procedure detects all endpoints of thinned edge lines and connects the endpoints to their neighboring edge pixels based on the current positions and the directions of the line segments.

Let (x_0, y_0) be a detected endpoint with direction $q(x_0, y_0)$ which is determined by fitting the five connected edge points in the neighborhood of (x_0, y_0) to a straight line using the least-squares method. The endpoint (x_0, y_0) then searches for its connected edge pixel in a half-circle region with search radius r, $1 < r < r_{\max}$, and search angle q^*, $q(x_0, y_0) - 90^\circ \le q^* \le q(x_0, y_0) + 90^\circ$, where $r_{\max}$ is the maximum search radius. The search procedure is as follows:

For r = 1, 2, …, $r_{\max}$

For Δq = 0, $\pm1^\circ$, $\pm2^\circ$, …, $\pm90^\circ$

$$q^* = q(x_0, y_0) + \Delta q$$

$$\begin{cases} x^* = x + r\cos q^* \\ y^* = y + r\sin q^* \end{cases}$$

If (x^*, y^*) is a focused edge point, then connect

(x_0, y_0) to (x^*, y^*) and terminate the search.

The search angle q^* is alternately incremented and decremented by 1° so that the connected line segment between (x_0, y_0) and (x^*, y^*) has the least deviation from the direction $q(x_0, y_0)$. Based on the preliminary experimental results of 25 test images of size 512×480 pixels, the maximum search radius $r_{\max}$ = 65 pixels are sufficient to form closed boundaries. Figure 3(g) shows the final result of the edge-linking for the doll image, and Figure 3(h) presents the result of superimposing the linked edges on the original doll image.

The edge linking procedure generally results in many closed boundaries that include the outermost contour of a focused object and small closed boundaries within the object contour. However, the background (or foreground) region generally will not produce any closed boundary in the image. We implement a simple region-fill algorithm (Foley et al., 1996) used in computer graphics applications to eliminate all pixels belonging to the background and foreground. The region-filling algorithm needs a seed pixel to start the filling process. In this study the pixel with the largest defocus value of p_e in the image is selected as the seed for the filling algorithm. It ensures that the selected seed is a background pixel. Figure 3(i) presents the filling results of the doll image. The background region is filled with a gray-tone value and the interior region of the object contour remains white. Figure 3(j) shows the final segmentation result for the focused doll.

3. EXPERIMENTAL RESULTS

In our implementations, all algorithms are programmed in the C language and executed on a personal computer with a Pentium 166MHz processor. The image size is 512×480 pixels with 256 gray levels. 25 visual images with a variety of focused objects in complex foreground and background are experimented.

The experiments have examined three sizes of neighborhood windows $N(x,y)$ including 15×15, 21×21 and 31×31 for the computation of the defocus measure p_e. These three neighborhood windows yield similar segmentation results with different degrees of boundary raggedness for the 25 test images. Generally, too small the size of the window may not include sufficient data to estimate p_e reliably, whereas too large the size of the window may include superfluous data and increases the computational requirement. According to the experimental results on the 25 test images, the neighborhood window of size 21×21 gives the most consistent and reliable segmentations. For a 21×21 neighborhood window, the threshold T_d for discriminating focused edges and defocused edges is set to 60 ($p_e\times255$) for all 25 test images. Note that the selected threshold value of T_d is only affected by the window size, but not the contents in the image since focused edge pixels of any object will result in the same amount of defocus according to the depth formula.

Figure 4 and 5 show the segmentation results of two head images. The regions corresponding to the focused objects are correctly segmented in both images. The image in

Figure 4(a) contains two persons, the focused one is on the left and the background one is on the right. Figure 4(b) shows the result of the edge linking process that bounds the focused person in a closed boundary. As seen in Figure 4(c), the person in the background is completely eliminated from the image. Since the focused person in Figure 4 divides the background into two isolated regions, the background region shown in the lower-left corner of Figure 4(c) is retained. Note that the region-filling algorithm used in this study selects only one seed pixel to fill the background region. It is not directly applicable to segmenting focused objects that divide the background into two or more disjointed regions. A more elegant region-filling algorithm is required for such situation.

Figure 6 shows the image of two miniature cars. These two cars are identical, except that one car is in focus and the other one is in the foreground. Figure 6(c) illustrates the segmentation result for the miniature cars. The focused car has been extracted reliably from the complex image with the exception that the pole of the flag is missing. This is because the pole is a very thin object in the image, and the two vertical edges of the pole identified in the focused edge detection process are merged into a single edge in the dilation process. Using an elegant edge linking algorithm rather than simple dilation should cope with the problem.

4. CONCLUSION

Traditional segmentation techniques such as thresholding and region growing are based on gray-level or texture similarity of segmented regions. These approaches only work well for images with measurable homogeneous properties. In this paper we focus on the extraction of

focused objects in complex visual images that contain both foreground and background objects. Our segmentation approach to the focused object detection problem relies on the measurement of defocus for object edges. A given gray-level image is initially converted to a gradient image using the Sobel edge operator. For each edge pixel in the gradient image, the proportion of the blurred edge region in a small neighborhood window (p_e) is evaluated using the moment-preserving technique. The moment-preserving method provides a closed-form solution to obtain the value of p_e. The resulting value of p_e is between 0 and 1, and increases as the amount of defocus increases. Defocus measure p_e is used as a powerful cue to detect edges of focused objects. Edges of focused objects yield small values of p_e, whereas edges of objects in the foreground or background have large values of p_e. Therefore, only those edges with small amount of defocus are retained in the image. Simple edge linking scheme is then performed to connect broken edges of focused objects into closed boundaries. A region-filling procedure follows to eliminate all foreground and background pixels and retain regions of focused objects defined by the closed boundaries.

Experiments on 25 visual images have shown that the proposed method can achieve correct segmentation, regardless of the complexity of foreground and background in images. As seen in the figures in the experimental section, the closed boundaries produced for focused objects are ragged. This is because the edge-linking scheme implemented in this study uses only simple dilation process and employs straight line segments to connect broken edges. Using elegant edge-linking algorithms or curve fitting techniques should generate smoothly bounded regions of focused objects. Overall, the measure of edge blur is a feasible approach for segmenting focused objects in complex images.

REFERENCES

Besl, P. J. and R. C. Jain (1988). Segmentation through variable-order surface fitting. IEEE Trans. Pattern Anal. Machine Intell. 10, 167-192.

Browning, J. D. and S. L. Tanimoto (1982). Segmentation of pictures into regions with a tile by tile method. Pattern Recognition 15, 1-10.

Foley, J. D., A. van Dam, S. K. Feiner and J. F. Hughes (1996). Computer Graphics, Addison Wesley, Reading, Mass.

Fukaka, Y. (1980). Spatial clustering procedures for region analysis. Pattern Recognition 12, 395-403.

Gonzalez, R. C. and R. E. Woods (1992). Digital Image Processing. Addision Wesley, Reading, Mass.

Goshtasby, A. and H. -L. Shyu (1995). Edge detection by curve fitting. Image and Vision Computing 13, 169-177.

Horowitz, S. L. and T. Pavlidis (1976). Picture segmentation by a tree traversal algorithm. Journal of the ACM 23, 368-388.

Hsiao, J. Y. and A. A. Sawchuk (1989). Supervised textured image segmentation using feature smoothing and probabilistic relaxation techniques. IEEE Trans. Pattern Anal. Machine Intell. 11, 1279-1292.

Jain, A. K. and R. C. Dubes (1988). Algorithms for Clustering Data. Prentice-Hall, Englewood Cliffs, N. J.

Jain, R., R. Kasturi and B. G.. Schunck (1995) Machine Vision, McGrow-Hill, New York.

Jain, A. K., N. K. Ratha and S. Lakshmanan (1997). Object detection using Gabor filters. Pattern Recognition 30, 295-309.

Lai, S. -H. and C. -W. Fu (1992). A generalized depth estimation algorithm with a single image. IEEE Trans. Pattern Anal. Machine Intell. 14, 405-411.

Otsu, N. (1979). A threshold selection method for gray-level histogram. IEEE Trans. Syst. Man Cybernet. 9, 62-66.

Pal, N. R. and S. K. Pal (1993). A review on image segmentation techniques. Pattern Recognition 29, 1277-1294.

Pentland, A. P. (1987). A new sense for depth of field. IEEE Trans. Pattern Anal. Machine Intell. PAMI-9, 523-531.

Tsai, W. -H. (1985). Moment-preserving thresholding: a new approach. Computer Vision, Graphics, and Image Processing 29, 377-393.

Tsai, D. -M. and Y. -H. Chen (1992). A fast histogram-clustering approach for multi-level thresholding. Pattern Recognition Letters 13, 245-252.

Xie, M. (1992). Edge linking by using casual neighborhood window. Pattern Recognition Letters 13, 647-656.

Xie, M. and M. Thonnat (1992). An algorithm for finding close curves. Pattern Recognition Letters 13, 73-81.

Zhou, Y. T. , V. Venkateswar and R. Chellappa (1989). Edge detection and linear feature extraction using a 2-D random field model. IEEE Trans. Pattern Anal. Machine Intell. 11, 84-95.