In-Camera Automation of Photographic Composition Rules

Serene Banerjee, *Member, IEEE,* and Brian L. Evans, *Senior Member, IEEE*

***Abstract*—At the time of image acquisition, professional photographers apply many rules of thumb to improve the composition of their photographs. This paper develops a joint optical-digital processing framework for automating composition rules during image acquisition for photographs with one main subject. Within the framework, we automate three photographic composition rules: repositioning the main subject, making the main subject more prominent, and making objects that merge with the main subject less prominent. The idea is to provide to the user alternate pictures obtained by applying photographic composition rules in addition to the original picture taken by the user. The proposed algorithms do not depend on prior knowledge of the indoor/outdoor setting or scene content. The proposed algorithms are also designed to be amenable to software implementation on fixed-point programmable digital signal processors available in digital still cameras.**

Index terms— main subject segmentation, photographic composition rules, raster image processing, digital signal processors, digital still cameras

I. Introduction

TO make a photograph more appealing, professional photographers apply a wealth of photographic composition rules [1]. This paper proposes a joint optical-digital framework for a digital still camera for automating selected photographic composition rules to improve the composition of pictures taken by amateur photographers. The framework would apply photographic composition rules to the user-intended picture to generate additional alternate pictures of the same scene. The camera would then provide the alternate pictures and the user-intended picture to the user for evaluation. Beyond personal use, such a smart camera system might be useful to professionals who need to take pictures for documentation, such as realtors and architects.

Within the proposed framework, photographic composition is improved by (i) segmenting the main subject(s) in the photograph, and (ii) automating selected photographic composition rules. This paper proposes an unsupervised automated method for identifying the main subject in the photograph that is assisted by optical pre-processing in the camera. Based on segmentation of the main subject, we automate three photographic composition rules: *rule-of-thirds, background blurring* and *merger mitigation.*

To place the main subject, the *rule-of-thirds* can be followed. Here the canvas is divided into three equal parts along the width and height, respectively. The center of the main subject should be placed at one of the four places: at 1/3 or 2/3 of the picture width from the left edge, and 1/3 or 2/3 of the picture height from the top edge. After segmentation, the main subject is relocated to follow the rule-of-thirds.

Background blurring is either introduced to enhance the sense of motion where the main subject is moving or decrease the depth-of-field of the picture where the main subject is stationary. After main subject segmentation, background blurring is implemented using region-of-interest filtering.

A merger occurs when equally focused foreground and background regions merge as one object in a two-dimensional picture of a three-dimensional world. Examples of mergers include a horizontal line shooting through the main subject's ears, and trees appearing to grow out of the main subject's head. Professional photographers change camera settings so that the main subject is in focus, while the objects in the background that merge with the main subject are blurred [1]. This preserves the sense of distance between the objects in the photograph. After segmentation, the background object that merges with the main subject is automatically identified and blurred. The *merger mitigation* approach could be extended to identify and blur more than one background object merging with the main subject.

The framework relies on segmentation of the main subject. In performing segmentation of the main subject, a supplementary picture is taken by the camera immediately before or immediately after the user takes the intended picture. In the supplementary picture, we assume that the auto-focus filter is focused on the main subject. During the acquisition of the supplementary picture, the shutter aperture is fully opened to allow the lens optics to blur anything not in the plane of focus, where the main subject is assumed to be. This supplementary picture has the main subject in focus, and the rest of the picture blurred with diffused light. The difference in frequency content between the main subject and the background is exploited by the proposed unsupervised digital image segmentation method, largely by applying the appropriate filtering. In this case, the optical subsystem has performed most of the computation needed to perform the segmentation. Other digital image segmentation methods are described in Section II.

In this paper, we develop algorithms for the proposed framework intended for implementation on a programmable fixed-point digital signal processor in a digital still camera. Such processors have fast 16-bit by 16-bit hardware multipliers (with single instruction throughput) but very limited on-chip program memory (32-256 kB) and data memory (32-256 kB). Algorithms developed for these processors should not only avoid using floating-point data and arithmetic, but should

S. Banerjee conducted this research while at The University of Texas at Austin, and is now with HP Research Labs, in Bangalore, India. serene.banerjee@hp.com. B. L. Evans is with the Embedded Signal Processing Laboratory, Center for Perceptual Systems, The University of Texas, Austin, TX 78712. bevans@ece.utexas.edu.

```
for each user-acquired picture do
    Acquire supplementary low depth-of-field picture;
    Compute main subject mask with supplementary
    low-depth-of-field picture;
    Perform image registration between the
    supplementary picture and user-acquired picture;
    Compute proposed alternate picture 1 that follows
    rule-of-thirds;
    Compute proposed alternate picture 2 with
    background blurring;
    Compute proposed alternate picture 3 with mitigated
    mergers;
    begin
        Segment image background based on color;
        Evaluate degree of merger for each background
        object;
        Mitigate effect of most prominently merged
        background object;
    end
    Output 4 pictures (User-acquired picture and
    proposed alternate pictures 1, 2, and 3);
end
```

Algorithm 1: Pseudo-code for in-camera automation of photographic composition rules

also avoid using fixed-point operations with large dynamic range in intermediate calculations such as eigen decomposition and singular value decomposition. Due to small amount of on-chip memory, algorithms developed for these processors should rely on neighborhood operations and not on global operations on the entire image. To reduce execution time, algorithms that perform its operations in one pass through the entire image should be favored over algorithms that perform iterative operations where the entire image must be read in each iteration. Further, in the context of a digital still camera, real-time processing means fast enough for the user not to grow impatient to see the results, which is in order of a second.

The first contribution of this paper is a joint optical-digital framework for improving photographic composition of digital pictures. The second contribution is the automation of segmentation of the main subject in a single still picture using a low-complexity digital image processing algorithm assisted by optical pre-processing. The third contribution is the automation of main subject placement, artistic background blur, and merger mitigation to improve the composition of the photograph. We show that the computational complexity of digital processing in the joint framework (main subject segmentation plus the three composition rules) is comparable with the computational complexity of performing JPEG compression and decompression. The pseudo-code combining all three contributions is given in Algo. 1.

The images used for testing the proposed algorithms in this paper were low-depth-of-field pictures downloaded from the World Wide Web or acquired with a Canon Powershot G3 camera. The shutter aperture was varied from F2 through F2.8 to make sure that the acquired images are low-depth-of-field photographs. The test set consisted of variety of pictures having human or inanimate main subjects and were taken under different lighting conditions and scene settings (indoor/outdoor). The software and color images for this paper are available at http://www.ece.utexas.edu/˜bevans/projects/dsc/

The rest of the paper is organized as follows. Section II discusses related research in main subject detection. Section III describes the main subject detection process. Section IV automates the rule-of-thirds. Section V simulates background blurring. Section VI presents merger detection and mitigation. Section VII analyzes segmentation accuracy and implementation complexity of the proposed algorithms. Section VIII concludes the paper.

II. Related Work

Previous work on main subject detection has been generally targeted towards detecting the main subject in offline settings. Luo, Etz, Singhal, and Gray [2], [3] propose a Bayes neural network to detect the main subject. Their method 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. False boundaries are removed with perceptual grouping of identifiable regions such as flesh tones, sky, and trees. Then, geometric features are extracted, including centrality, borderness, shape, and symmetry. A probability density function for the main subject location is estimated from the training data. The probability density function estimate can be applied to the unknown test set to guess what and where the main subject is. Their method requires supervised learning, which limits its ability to adapt to changing conditions in the field. Also, as this is a Bayes net based approach, the system performance will be poor if the test set is very different from the training examples. With the vast number of possibilities of scene content, scene settings, and user preferences, developing a good set of training examples in order to guarantee that the network would perform well for a varied number of circumstances is difficult.

In a wavelet-domain approach, Wang, Li, Gray, and Wiederhold [4], [5] analyze the statistics of the high-frequency wavelet coefficients to segment the focused regions in an image, thereby detecting the object of interest. Initially, the image is coarsely classified into object-of-interest and background regions by using the average intensity of each image block and the variance of wavelet coefficients in the high frequency bands. The variance is higher for the focused regions in the image. Blocks are clustered using the *k*-means algorithm [6] by noting that blocks from a homogeneous image region will have similar average intensities. Each block is further subdivided into child blocks, and a multiscale context-dependent classification is performed for further refinement. Finally, a post-processing step removes small isolated regions and smoothes the boundaries. The method uses Haar wavelets, which have transfer functions that are scaled versions of $1 + z^{-1}$ and $1 - z^{-1}$, for the lowpass and highpass filters, respectively. The Haar wavelets and feature extraction can be implemented in fixed-point arithmetic. Nonetheless, the multiscale wavelet-

domain method is computationally intensive. Section VII compares this wavelet-domain and our proposed approach.

In a spatial-domain approach, Won, Pyan, and Gray [7] develop an iterative algorithm based on variance maps. A local variance map is used to measure the pixel-wise high frequency distribution in the image. This variance map has blob-like errors both in the foreground (where the image is relatively smooth) and the background (where the background is highly textured) regions. To eliminate these errors, the authors employ a block-wise maximum a posteriori image segmentation that requires recursion over image blocks and is computationally demanding. Their method yields more accurate segmentation compared to the aforementioned wavelet-based approach [4], [5]. Section VII compares this spatial-domain approach with our proposed approach.

Other recent offline region-based segmentation methods include min-cut segmentation methods [8], mean-shift analysis [9], and segmentation of blobs using expectation maximization [10]. In the min-cut method [8], the image is defined in terms of its affinity matrix. The segmentation problem is then formulated as a graph theoretic min-cut max flow problem. First the eigenvalues and eigenvectors are computed using singular value decomposition of a large matrix in floating-point arithmetic, and then seed regions are calculated. Subsequently, minimum cuts are computed between source and sink regions and the regions are later merged. The authors report segmentation times of 3 to 7 minutes per image of sizes up to 239×138 on a 2.5 GHz Pentium IV platform. In mean-shift analysis [9], the image is segmented based on the following steps: (i) it estimates the density function for the image and (ii) it shifts each feature point to the nearest stationary point along the gradient of the estimated density function. The complexity of the algorithm is quite high as it computes a Euclidean norm, a Hessian function, and a line search. In the third method [10], the image is segmented first into parts based on color analysis in the CIELab space and texture analysis at different scales, and then the parts are grouped into blobs using expectation maximization. The processing involved is a RGB to CIELab color conversion, evaluating textures using gradients at various scales and subsequently using expectation maximization algorithm for regrouping, which by itself is computationally complex. The authors report segmentation times of 5 to 7 minutes per 320×240 image on a 300 MHz Pentium II platform. As the proposed framework is intended for real-time implementation on fixed-point digital signal processors with small amounts of on-chip memory, these floating-point, high-memory usage methods are eliminated from consideration.

Previous research to detect the main subject with a computationally intensive algorithm may be appropriate for offline applications, such as image indexing for content-based retrieval, object-based image compression for image servers, and content grouping for automatic album layout. However, providing in-camera feedback to the photographer while a picture is being acquired must happen in a matter of seconds. This paper proposes to detect the main subject with a low-implementation complexity algorithm that can be implemented in fixed-point arithmetic in the digital signal processors in digital still cameras.

III. Main Subject Segmentation

To detect the main subject on the fly, we propose an in-camera main subject segmentation algorithm [11]. The proposed approach utilizes digital still camera controls, such as the auto-focus filter and the software-controlled shutter aperture and speed. Assuming the user points to the main subject, the auto-focus filter puts the main subject in focus [12]–[14]. We then open the shutter aperture all the way and the shutter speed is automatically adjusted so that the light from the out-of-focus objects does not converge as sharply as from the objects in focus and finally, we acquire a supplementary picture. The resulting blur in the out-of-focus objects is used to detect the focused main subject by using filtering, edge detection, and contour smoothing. The sole purpose of taking the supplementary picture is to determine the main subject location in the user-acquired photograph.

A. Implementation of main subject detection

In the proposed image acquisition framework, the main subject is in focus, and the background is blurred by widening the shutter aperture. So the main subject in focus will have prominent gradient features and the background that is out of focus will have blurred features. Thus, the segmentation of the main subject and the corresponding background is induced by this difference of gradient information. The goal is to obtain the intensity distribution pertaining to the high frequency component in image in contrast to the blur background component, i.e., extract the sharp features of the image.

The obtained images from the camera is first processed with an image sharpening filter as modeled in Fig. 1. So,

$$g(x,y) = I(x,y) - I_{smooth}(x,y) \tag{1}$$

and

$$I_{sharp}(x,y) = I(x,y) + k\, g(x,y) \tag{2}$$

Therefore,

$$I(x,y) = \frac{1}{k+1} I_{sharp}(x,y) + \frac{k}{k+1} I_{smooth}(x,y) \tag{3}$$

and

$$I_{sharp}(x,y) - I(x,y) = \frac{k(I_{sharp}(x,y) - I_{smooth}(x,y))}{(k+1)} \tag{4}$$

Subtracting the smoothed 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. In earlier work [11], we show that by combining the sharpening operation, difference calculation, and the edge detection for detecting the main subject into a single neighborhood operation, the implementation complexity is similar to that of a 5×5 filter.

For image sharpening, a 3×3 sharpening filter is employed:

$$\frac{1}{1+\alpha} \begin{bmatrix} -\alpha & \alpha - 1 & -\alpha \\ \alpha - 1 & \alpha + \beta & \alpha - 1 \\ -\alpha & \alpha - 1 & -\alpha \end{bmatrix} \tag{5}$$

Parameters α and β define the shape of the frequency response. We chose $\alpha = 0.2$ and $\beta = 5$. An integer implementation could choose $\alpha = 0.2$ and $\beta = 5$, remove the $\frac{1}{1+\alpha}$ factor, and scale the coefficients by 5. For the lowpass filter, a 3×3 Gaussian blur filter is used. 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. Also, the filter coefficients can be modified for implementation.

In detecting the strong edges from the optically pre-processsed and digitally filtered image, the Canny edge detector [15] gives good results 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 shown in Figs. 2 through 6 and Fig. 12. This selected hysteresis threshold depends on the amount of blurring in the acquired image, i.e. the amount of background blur obtained from the lens in the camera. So, for each camera, the hysteresis threshold could be set to work for a range of natural images. However, with any preselected threshold, the strong edge detection step would still pick background edges for some acquired images where there is not enough background blur or strong edges in the main subject.

Another popular edge detector, the Laplacian of Gaussian edge detector [16] could be tuned to preserve strong edges and suppress weak edges, but it did not perform as well as the Canny edge detector. The non-directional derivatives used in the Laplacian of Gaussian edge detector produces responses both parallel and perpendicular directions to a given edge. The drawback could have been improved by using directional first and second derivatives. Nonetheless, the Laplacian of Gaussian edge detector would still not preserve the edge direction. The Canny edge detector also performs better than Roberts, Sobel, and Prewitt edge detectors [17].

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 [18] algorithm and its direct descendants 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 for natural images. Instead, the gradient vector flow [19], [20] 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 in its ability to track image contour concavities.

In Figs. 2(a), 3(a) and 4(a), the main subjects are in focus, while the background blur is achieved by a wider shutter aperture. Figs. 2(b), 3(b) and 4(b), show that after sharpening the image and taking the difference, the main subject edges are stronger. With further edge detection, the results of locating the main subjects before contour closing are shown in Figs. 2(c), 3(c) and 4(c), respectively. Figs. 2(d), 3(d) and 4(d) show the detected main subject mask.

B. Modification of Mask Based on Difference Between Original and Supplementary Picture

One of the drawbacks of taking a supplementary picture with a shallow depth of field is that the subject or the camera could have moved while the supplementary picture is taken. This drawback could be reduced by mounting the camera on a tripod. However, to compensate for small ranges of motion, we employ a computationally simple image registration method [21], [22] to adjust the position of the main subject mask in the supplementary image to line up with the main subject in the user-acquired image.

Once the mask in the supplementary image has been generated, a difference is computed between the original and the supplementary images. Now, the difference image will contain pixels where the main subject has moved and pixels of the background that are in different focus compared to the supplementary picture. However, any change in main subject motion shows up more significantly compared to the change in focus. So, by using a threshold on the absolute value of the difference image, one can identify if the main subject has moved. A second mask is then generated that identifies the pixels in the difference image lying above this threshold. For the proposed application, this threshold is chosen to be 70 for an 8-bit image. This second mask is added to the generated main subject mask to create the mask that will be used on the original image. Thus, the generated main subject mask from the supplementary image is aligned with the main subject.

IV. Rule-of-Thirds: Automated placement of the main subject

The post-segmentation objective is to automatically place the main subject following the rule-of-thirds. The rule-of-thirds says to place the main subject in one of four places: at $1/3$ or $2/3$ of the picture width from left edge, and 1/3 or 2/3 of the picture height from the top edge. A mathematical measure is defined to check how close the picture follows the rule-of-thirds, and to reposition the main subject [23].

A. Mathematical Formulation

Let S be the scene domain of the main subject where

$$S \in \{\mathbf{v} | \mathbf{v} \in \text{Main subject}\},\ \mathbf{v} = \{(x_1, y_1), (x_2, y_2), ..., (x_i, y_i)\} \tag{6}$$

is the set of pixel positions. Then, the center of mass is defined as the weighted sum of the components and cardinality of the scene domain. Consider that there are n main subjects. The center of mass for each of them is computed independently. A two-dimensional function $f(x, y)$ is defined such that it reaches a minimum when a center of mass is at the one-third position in the canvas both along the x and y axis. The objective will be to minimize the summation of the value of the function generated by the center of mass positions (x'_n, y'_n) of the n main subjects.

B. Implementation

For the current implementation, we assume one main subject (i.e. $n = 1$), and the function $f(x, y)$ is a product of the

Euclidean distance from the four one-third corners on the canvas. Let (x_1, y_1), (x_2, y_2), (x_3, y_3), and (x_4, y_4) be the four one-third corners. And, (x, y) is the position of the center of the mass of the main subject. Then,

$$f(x,y) = ((x-x_1)^2 + (y-y_1)^2)((x-x_2)^2 + (y-y_2)^2) \\ ((x-x_3)^2 + (y-y_3)^2)((x-x_4)^2 + (y-y_4)^2) \quad (7)$$

So, $f(x,y) \geq 0$ with $f_{min}(x,y) = 0$, and the minimum is attained when the center of mass is at one of the one-third corners. Thus, after computation of the center of mass, the image pixels are shifted so that they fall at a one-third corner.

The center of mass is computed along the rows and columns respectively. For each row (or column) if w_n is the number of "ON" pixels in the main subject mask, then the center of mass is defined as:

$$center = \frac{w_n * \text{row (or column) location}}{\Sigma w_n} \quad (8)$$

After computing the center of mass, a comparison is made as to which of the four one-third corners is closest to the current position of the center of mass. The picture is then shifted, so that the center of mass falls at the closest one-third corner.

Another approach is to crop the picture so that the center of mass of the main subject falls on one of the one-third corners. This is computationally very simple. During cropping the two competing criteria to optimize are (1) moving the center of mass as close to one of the four one-third points as possible, and (2) minimizing the number of rows and columns cropped in the picture to retain the most picture content possible, subject to the constraint that no pixels of the main subject are cropped.

The original pictures in Figs. 2–4(a) have the main subject closer to the canvas center. Figs. 2–4(d) show the detected main subject masks, the 1/3 and 2/3 lines on the canvas along the height and width, respectively, and the position of the center of mass of the detected main subject. Figs. 2–4(e) show the main subjects repositioned following the rule-of-thirds. For simulation purposes either mirror reflection (for textured boundaries) or boundary pixel extension (for smoother regions) is used for the undefined pixels. The arising artifacts could be reduced by capturing an image by using a wide-angled lens camera. The user could also be signaled to move the camera in a particular direction.

For multiple main subjects in the photograph, the proposed algorithm could be extended for automation of the *rule-of-triangles* [1]. The rule-of-triangles states that if there is more than one main subject in the picture, then their centers of mass should not lie on the same line in the canvas, but should form triangles on the canvas. This can be automated by adding a constraint during minimization so that no two center of masses lie on the same row in the canvas.

V. Motion Effects Rule: Simulating Background Blur

For simulating background blur, the original image is first masked with the main subject mask detected by the method proposed in Section III. Then region of interest filtering is performed on the masked image, so that the main subject pixels remain unaltered, whereas artistic effects can be added to the background.

We convolve the images with a motion blur filter that simulates linear and radial blurs produced by horizontal and rotational movement of the camera. The filtering involves convolving the image with a series of filters and compositing the filtered images. Figs. 2(f), 3(f) and 4(f) show simulated background blurring that could have resulted from camera panning. The current example simulates linear motion of the camera by 10 pixels. Other values of linear, radial, or zoomed motion blurs can also be simulated.

VI. Improve Perspective Rule: Merger Detection and Mitigation

The method proposed in Section III generates a main subject mask that divides the picture into foreground and background regions. Fig. 6(b) is the generated main subject mask for Fig. 6(a). The goals will be to segment the background, identify merging objects, and blur the picture. The formulation of the steps follow [24].

A. Background segmentation

The color information is used for segmentation of the background objects. The red, green, and blue (RGB) image provided by the camera is transformed to the hue channel found in the hue, saturation, value (HSV) space. In HSV space, hue corresponds to color perception, saturation provides a purity measure, and value provides the intensity. A histogram in the hue space is then utilized for segmentation of the background region. Although hue does not model the color perception of the human visual system as accurately as CIELab, it is chosen because the transformation from RGB to hue has lower implementation complexity [25]. RGB to CIELab requires calculation of cube roots. However, Linearized CIELab space [26] could be used instead of the hue space or the complexity of calculating cube roots in CIELab could be reduced by using a lookup table.

Let the hue values be on the interval $[0, 255]$ and broken into m-bins. The discrete probability distribution for hue values belonging to each bin is $P(\text{hue}_m) = \frac{c(\text{hue}_m)}{T_c}$, where $c(\text{hue}_m)$ is the count corresponding to each bin and T_c is the total count of values in all bins. By modeling the background picture as a Gaussian mixture of hue values, the task is to further segment these m-bins into n-groups, where each group will identify a different object.

The term $\frac{T_c}{m}$ gives the average of the hue values. Any hue value above this average is marked as a dominant hue. Based on the available dominant hues, the n-groups are determined automatically so that each group contains only one dominant hue. Each group boundary lies halfway between two dominant hues. This ensures that the local maxima of the probability distribution, $P(\text{hue}_m)$, is captured in each group. Pixels with hue values falling in each of the identified n-groups form different background objects.

For the proposed algorithm, m is chosen to be 64, as it is assumed that a difference in four hue levels (i.e., $256/64$

levels) would correspond to approximately the same perceived color [27]. Fig. 8 shows the color histogram for the hue values with the average and the peaks for the background of Fig. 6(a). Based on the color histogram and the average value, $n = 10$ background objects are automatically identified for Fig. 6(a). Fig. 9 shows three of these identified background objects.

B. Merger detection

Based on the background segmentation, the background image can be modeled as a linear combination of the background objects. Thus, $S_b = \bigcup_{i=1}^{n} O_i$, where S_b is the background image and O_i are the identified n background objects. Now, one or more of these background objects may merge with the main subject. We choose the background object that has the largest high frequency energy and is touching the main subject.

To automatically identify the merged object, each object O_i is transformed to a feature space representation, Ω_i, where $\Omega_i \in \Gamma$. Γ is defined as a weighted sum of the high frequencies contained in the spatial region of each object. High frequency coefficients are obtained from the first level of the two-dimensional Gaussian pyramid [28] of the intensity image. Gaussian pyramids are localized in space. The Gaussian pyramid could be replaced with a Laplacian pyramid, for the added implementation complexity of one subtraction per pixel.

The high frequency coefficients are weighted with the inverse of the distance in space to the main subject mask. To compute the inverse distance transform, the distance transform coefficients are stored as a grayscale image, and are subtracted from 255 before multiplication with the high frequency coefficients. This assigns more penalty to the higher frequencies closer to the main subject. Figs. 10(a) and (b) show the Euclidean distance transform [29]–[31] and high frequency coefficients obtained from the first level of the Gaussian pyramid, respectively. In Section VII, we will reduce the computational complexity of the inverse distance measure.

An object O_i is detected to be merged with the main subject if its feature space representation, Ω_i, is more than a threshold. This threshold could be selected by the user. This paper presents an unsupervised approach in which the object O_i yielding the maximum value of the feature space representation, Ω_i, is identified to be the merged object. This unsupervised approach detects the object producing the strongest merger and blurs the produced artifact. For Fig. 6(a), the tree object shown in Fig. 9(b), produces the maximum of the weighted sum of high frequencies, identifying that the tree merges with the main subject.

C. Selective blurring

The detected merged object, O_i^*, has feature a space representation, Ω_i^*. To reduce the effect of the merger, Ω_i^* needs to be reduced. As Ω_i is the weighted sum of the high frequencies, the high frequency coefficients are masked when the image is reconstructed from the Gaussian pyramid representation. In Fig. 6(a), the high frequency coefficients of the first level of the Gaussian pyramid are masked out using the approximate shape of the detected tree object. The resulting image is shown in Fig. 7. To increase the amount of smoothing, masking can be extended to higher levels of the Gaussian pyramid.

Cardos *et al.* [32] develop an algorithm for ranking various segmentation algorithms. In their approach, they use a measure of distance between the segmented objects to identify if they are parts of the same object or different objects. In merger detection, also we use a weighted distance measure to identify the merged object and reduce this weighted distance measure to mitigate the mergers. So, in a sense it is making two distinctly classified objects, as one, and reducing the weighted distance between them.

VII. Results and Implementation Complexity

The proposed algorithm is shown in Fig. 5. After main subject detection, the post-segmentation complexity will depend on the number of rules that are automated. A digital still camera uses approximately 160 digital signal processor instruction cycles per pixel. Main subject detection, automation of the rule-of-thirds and background blurring, or merger detection and mitigation requires fewer digital signal processing cycles (as explained below). The proposed algorithms are amenable for implementation in fixed-point data types and arithmetic.

A. Main Subject Detection

1) Implementation Complexity: The RGB color image is converted to intensity by

$$I = (R + G + B)/3 \text{ or } I = (R + 2G + B)/4 \qquad (9)$$

The former step requires 2 additions and 1 multiplication, or alternately 2 multiply-accumulates, 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, to 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 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 comparisons per pixel.

Image	Resolution	Sensitivity	Specificity	Error rate
Man & child	280×350	88.0 %	97.2 %	4.1 %
Man	246×276	77.8 %	90.3 %	8.0 %
Doll	316×422	82.2 %	94.6 %	6.3 %
Merger	282×200	80.6 %	87.9 %	8.9 %

TABLE I
SEGMENTATION ACCURACY MEASURES FOR THE MAIN SUBJECT DETECTION ALGORITHM FOR IMAGES IN FIGS. 2– 6

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. This has low implementation complexity on a digital signal processor in the digital still camera.

2) Accuracy of segmentation: The accuracy of segmentation is determined by the sensitivity, specificity, and the error rate measures [33]. The sensitivity is defined as the ratio of the area of the detected main subject to the total area of the main subject in the image. The specificity is the ratio of the area of the detected background to the total area of the background in the image. Here the total area of the main subject or the background are the number of pixels that actually represent the main subject or the background, respectively, as would have been observed by a human. The error rate is the ratio of the number of pixels that are misclassified to the total area of the image. Here the segmentation ground truth was obtained by averaging the segmentation masks of three human subjects.

For the segmented images given in Figs. 2(d), 3(d), 4(d), and 6(b) the sensitivity, specificity, and the error rate are given in Table I. The inaccuracy in segmentation as seen in Table I is within the tolerable limit as a trade off for low-complexity in detecting the main subject for subsequent automation of the photographic composition rules. The results also are comparable with Wang's *et al.* [4], [5] reported values for the three quantifiable measures for low depth of field images. For their test images, sensitivity, specificity, and error rate varied from 73.7% to 97.5%, 80.1% to 97.5%, and 3.4% to 5.5%, respectively.

3) Comparison with prevalent segmentation methods: Figs. 12 compare the proposed main subject segmentation algorithm with a wavelet-based method [4], [5], and Won's *et al.* iterative method [7]. As described earlier, our proposed method takes 18 multiply-accumulates, 4 comparisons, and 6 memory accesses per pixel, and does not require any *à priori* training.

Similar studies were conducted for 30 images. The images were either downloaded from the World Wide Web in the year 2001 or acquired with a Canon Powershot G3 camera. The shutter aperture was varied from F2 through F2.8 to make sure that the acquired images are low depth-of-field photographs. The test set consisted of variety of pictures having human or inanimate main subjects were taken under different light conditions and scene settings (indoor/outdoor). The original pictures are available at: http://www.ece.utexas.edu/˜bevans/projects/dsc.

The multiscale wavelet-based method [4], [5] generates the wavelet coefficients for each stage and classifies the image based on the variance of the wavelet coefficients, using the *k*-means clustering algorithm. The process is repeated for multiple wavelet levels. Generating the wavelet coefficients involves filtering the image with low and highpass filters, respectively. Also, the computationally intensive part of the *k*-means clustering lies in computing the Euclidean distance of each point from the neighboring clusters. Taking into account all these factors, the wavelet-based method will at least be $2 \times n \times k$ more complex than the proposed method, where n is the number of wavelet levels computed and k is the number of clusters.

The iterative spatial-domain approach [7] starts from dividing the image into non-overlapping blocks. A few probability measurements are computed from the image variance to classify each block as foreground or background. The block classification is further refined into pixel-level classification using recursion and the watershed algorithm. So, if the original image in divided into $B \times B$ blocks, this method would be at least B times as complex than the proposed method. For the results in Figs. 12(d), Won *et al.* [7] substitute the grey level values from the original image onto the generated mask for visual inspection.

The proposed method generates a reasonable mask of the main subject at a much lower complexity than the aforementioned methods. Also, the proposed algorithm can be implemented in fixed-point arithmetic. A senior design student at The University of Texas at Austin, Mayank Gupta, in collaboration with the authors implemented the proposed algorithms on a Texas Instruments digital still camera chip, TMS320C55x, and was able to obtain real-time performance [34]. As the proposed pixel-based approach to detect the main subject produces comparable results with the more complex wavelet-based method [4], [5], the following subsection compares the two paradigms.

4) Comparison of Multiresolution-based (Wavelets) and Pixel-based Main Subject Detection: Comparing the multiresolution wavelet-based [4], [5] and pixel-based [7], [11], [23], [24] approaches to segment the main subject, it can be seen that any wavelet or filter-based multiresolution approach to segment an image would be better at representing regional features of the image. Depending on the filter length and the resolution which is being used for analysis, the regional properties of the image would show up in the frequency transformed domain. So, any analysis based on regional properties will have estimation errors depending on the length of the used filter and the resolution at which it is being viewed at. The pixel-based approaches however analyze the image on a pixel-by- pixel basis, and the errors will depend on how well each pixel is classified. Thus, in this research, we present a pixel-based approach that is fast and classifies the pixels with tolerable accuracy required for this application.

So, for images with substantially large smooth regions that are separated by well defined edges, both the wavelet-based or pixel-based algorithms would provide similar results. However, in images with many edges and texture, the pixel-based approach would be more accurate. Also, in Wang, Li, Gray and Wiederhold's [4], [5] wavelet-based approach, the segmentation accuracy is further reduced when the authors use a block-based approach, in which in the subsequent iterations,

the class of a subblock is switched, depending on the subblock neighborhood.

For simplicity, we choose three model images to compare the proposed pixel-based and a multiresolution approach. We use a Laplacian pyramid for the multiresolution analysis. The first image is a plain image with no edge and is generated as $f(x, y) = 0$. The second is an image with a white circle on a black background, and it has a defined strong edge. It is generated as $f(x, y) = 1$ if $x^2 + y^2 \leq r^2$ where r is the radius of the circle. The third image is a ramp modeling an image with a very blunt edge. This image is generated as $f(x, y) = \frac{\sqrt{(x-x_{mid})^2+(y-y_{mid})^2}}{\sqrt{(x_{max}-x_{mid})^2+(y_{max}-y_{mid})^2}}$, where (x_{mid}, y_{mid}) are the mid points and (x_{max}, y_{max}) are the dimensions of the image.

Both the pixel-based and multiresolution Laplacian pyramid based approaches identify that there is no edge in the first image. The 6 levels of the Laplacian pyramid decomposition are considered for the second image. Now, as this image has a sharp edge the highest frequency octave identifies the circle correctly. However, as more and more lower resolutions will be considered to segment the image, the accuracy of segmentation would reduce. But, the regional properties of the image is present across all the octaves. Similarly, for the ramp image, depending on the chosen thresholds the proposed pixel-based approach either chooses none of the image or almost the whole of the image. Here also the segmentation would depend on which levels are being considered.

Now in a natural image, the strength of the edges cannot be predetermined, and the strength of all the edges would not likely be the same. So, a multiresolution approach would be better at representing the regional properties of the image but the segmentation accuracy would depend on which frequency level is being considered for segmentation. The accuracy of the pixel-based approach on the other hand will depend on how well each pixel is classified. Post-segmentation, the image registration step requires, one subtraction, one thresholding, and one addition operations and 2 memory accesses to modify the detected main subject mask.

B. Automation of the Rule-of-Thirds

Using the main subject mask, the rule-of-thirds algorithm requires 2 multiply-accumulates, 1 comparison, and 1 or 3 memory access per pixel, plus four comparisons and one division (explained below) for the entire image. One memory access per pixel is needed to calculate the center of mass. An additional two memory accesses per pixels is needed only if the picture is shifted instead of cropped.

For automated placement of the main subject following the rule-of-thirds, the center of mass for the detected main subject mask is computed with 2 multiply-accumulates and 1 memory read per pixel, and 1 division. The closest one-third corner is computed with 4 comparisons. The next step is to alter the picture so that the center of mass is at closest one-third corner.

One approach is to crop the picture so that the center of mass of the main subject falls on one of the one-third corners. This is computationally very simple. Instead of cropping the picture, every pixel in the entire image could be shifted by the same amount so that the center of mass of the main subject occurs at one of the one-third corners. After shifting the image, many pixel values along two of the edges of the image would be undefined. These pixels could be given values through pixel replication along the boundary of known pixel values. The shifting approach requires one memory read and write per pixel.

In the best case, the center of mass falls at one of the one-third corners so that the image does not have to be altered. In the worst case, the center of mass is at one of the corners of the picture, so that one-third of the rows and one-third of columns would be cropped or need to be given new values. In the average case, e.g. if the main subject were originally in the middle of the picture, one-sixth of the rows and columns would be cropped or be given new values.

C. Simulated Background Blurring

Using the main subject mask, background blurring realized by a 3×3 filter takes 9 multiply-accumulates and 4 memory accesses per pixel.

D. Merger Detection and Mitigation

Background segmentation starts with a conversion from RGB to hue. The hue value calculation uses an intermediate variable, H', which is in the interval $[-255, 1275]$ and can be represented by a 12-bit signed integer. The pseudo-code for the conversion follows: In the worst case, the conversion to hue requires 2 shifts, 3 adds, 6 compares, and 4 byte memory accesses per pixel. The histogram and thresholding require 1 add and 1 compare per pixel. The hue values are stored in NM pixels (or $N \times M \times 8$ bits), and a buffer of $NM \log_2 n$ bits stores the information of the n segmented objects. Now, for all practical applications, the number of segmented objects, n, will be less than 2^8. So, $n \leq 2^8$ or $\log_2 n \leq 8$ So, the information regarding the segmented objects can be stored in the buffer that originally had the hue values.

```
begin
    min = min(R, G, B);
    max = max(R, G, B);
    δ = max - min;
    if (R == max) then
        H' = G-B (within yellow & magenta);
    else
        if (G == max) then
            H' = 2δ+B-R (within cyan & yellow);
        else
            H' = 4δ+R-G (within magenta & cyan);
        end
    end
    H = (H' + 255) >> 3;
end
```

Algorithm 2: Pseudo-code for low-complexity hue determination from red, green, blue planes

The intensity Gaussian pyramid first converts the color image to an intensity image by either

$$I = (R + G + B)/3 \text{ or } I = (R + 2G + B)/4 \qquad (10)$$

The former step requiring 2 adds and 1 multiply is suitable for programmable digital signal processors. For a hardware implementation, we could use 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). Shifts can be used because the RGB values are non-negative. The intensity image is stored in NM pixels. Any level of the Gaussian pyramid can be computed by convolving the grayscale image with a 3×3 filter with power-of-two coefficients, which requires 9 shifts, 8 adds and 4 byte memory accesses per pixel. The 9 reads in image values to compute the convolution can be stored in registers in order to reduce the number of memory reads to 3 per pixel. The first level of coefficients are stored in NM pixels, and the intensity image may be overwritten in a sequential implementation of merger mitigation.

The inverse distance transform could be determined from the Euclidean distance transform [29]–[31] by subtracting its value from 255. In this case, the inverse distance transform would be computationally intensive. We propose an approximate, lower complexity, inverse distance measure. Along each row (column) the distance of each "off" pixel from the nearest "on" one is computed and a ramp function is generated. The maximum of the horizontal (row) distance and the vertical (column) distance is taken as the distance from the nearest "on" pixel. In order to assign more penalty to the high frequency coefficients close to the main subject, the pixels closer to the main subject mask have a higher weight. The weights are stored in NM pixels. This measure requires 2 adds, 1 compare, and 2 byte memory accesses per pixel.

For each background object, the intensity Gaussian pyramid coefficients are weighted by the inverse distance transform coefficients and summed. The background object with the highest sum is chosen as the background merging object, and the corresponding background object mask is the output. The background object mask can be stored in the main subject mask buffer so as to reuse memory. All totaled, 1 multiply, 1 add, and 1 compare are required per pixel.

In the final step, the color Gaussian pyramid and reconstruction only have to be applied to those pixels in the binary mask input correspondingly. For each pixel in the binary mask input, the first level of the color Gaussian pyramid transformation is calculated separably for each RGB planes. For each color plane, 9 shifts, 8 adds, and 3 byte memory accesses are required for a 3×3 filter kernel. The high frequency coefficients for the merging background object are masked with 1 compare and 1 memory access per pixel. The output (merger reduced) image takes 9 shifts, 8 adds, 1 compare, and 1 byte memory access per pixel, and would be stored in $3NM$ pixels.

The computational requirements for each block in merger mitigation are given in Table II. For merger mitigation, all the blocks, except the main subject detection and color Gaussian pyramid/reconstruction, work only on the background image. Hence, the complexity will depend on the percentage of background pixels in the image. The merger reduced image for Fig. 11(a) is shown in Fig. 11(b). The background trees merging with the bird are blurred out, thereby inducing a sense of distance.

Block	$\times$	$<<$	$+$	$\geq$	m
Segment background		2	4	7	4
Intensity Gaussian pyramid	1	9	10		4
Inverse distance transform			2	1	2
Detect merging object	1		1	1	1
Color Gaussian pyramid		27	24		9
Reconstruct pyramid	1	27	24		3
Total	**3**	**65**	**65**	**9**	**23**

TABLE II
PER PIXEL IMPLEMENTATION COMPLEXITY OF THE PROPOSED MERGER MITIGATION ALGORITHM IN NUMBER OF MULTIPLICATIONS ($\times$), SHIFTS ($<<$), ADDITIONS ($+$), COMPARISONS ($\geq$), AND BYTE MEMORY ACCESSES (m). THE LAST TWO STEPS ARE ONLY APPLIED TO THE MERGING BACKGROUND OBJECT. THE OTHER STEPS ARE APPLIED ONLY TO THE BACKGROUND.

Block	$\times$	$<<$	$+$	$\geq$	m
Main subject detection	18			4	6
Mask registration			2	1	2
Rule-of-thirds	2			1	3
Background blurring	9				4
Merger mitigation	3	65	65	9	23
Total	**32**	**65**	**67**	**15**	**38**

TABLE III
PER PIXEL IMPLEMENTATION COMPLEXITY OF MAIN SUBJECT DETECTION, AUTOMATING RULE-OF-THIRDS, BACKGROUND BLURRING, AND MERGER DETECTION AND MITIGATION IN NUMBER OF MULTIPLICATIONS ($\times$), SHIFTS ($<<$), ADDITIONS ($+$), COMPARISONS ($\geq$), AND BYTE MEMORY ACCESSES (m).

VIII. CONCLUSION

This paper proposes a joint optical-digital framework for helping the amateur photographers take pictures with better photographic composition. The framework, which is shown in Fig. 5, acquires the user-intended image as well as a supplementary image. This supplementary image is taken immediately before or after the user-intended picture and uses the same autofocus settings. In the supplementary image, however, the shutter aperture is fully opened and shutter speed is automatically adjusted so that objects not in the plane of focus are blurred by the optical subsystem. This supplementary blurry image is then digitally processed to locate the main subject. With the main subject identified, selected photographic composition rules may be automated to generate new alternate pictures with better photographic composition. Three photographic compositions rules are automated without making assumptions on scene setting or content.

This paper also moves towards the goal of implementing the framework in a digital still camera. A digital still camera implements a variety of digital image processing algorithms on a fixed-point programmable processor with little on-chip memory and relatively slow clock speeds and off-chip data transfers. We present low-complexity, non-iterative, unsupervised algorithms for automatic main subject detection and for automating three photographic composition rules: rule-of-thirds, artistic background blurring, and blurring merging background objects. With the proposed low-complexity algorithms, the entire framework can be implemented with fewer than 180 fixed-point computations and 40 memory reads/writes per pixel (as shown in Table III). This is about 50% higher than the implementation complexity of JPEG compression and decompression together.

REFERENCES

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

[2] 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.

[3] 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.

[4] 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.

[5] 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 Int.*, vol. 23, pp. 85–90, Jan. 2001.

[6] T. Kanungo, D. M. Mount, N. S. Netanyahu, C. D. Piatko, R. Silverman, and A. Y. Wu, "A Efficient k-Means Clustering Algorithm: Analysis and Implementation," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 24, pp. 881–892, July 2002.

[7] 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.

[8] F. J. Estrada, A. D. Jepson, and C. Chennubhotla, "Spectral Embedding and Min-Cut for Image Segmentation," in *Proc. British Machine Vision Conf.*, Sept. 2004.

[9] C. Yang, R. Duraiswami, D. DeMenthon, and L. Davis, "Mean-Shift Analysis Using Quasi-Newton Methods," in *Proc. IEEE Int. Conf. on Image Proc.*, vol. 2, pp. 447–450, Sept. 2003.

[10] C. Carson, S. Belongie, H. Greenspan, and J. Malik, "Blobworld: Image Segmentation Using Expectation-Maximization and Its Application to Image Querying," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 24, pp. 1026–1038, July 2002.

[11] S. Banerjee and B. L. Evans, "A Novel Gradient Induced Main Subject Segmentation Algorithm for Digital Still Cameras," in *Proc. IEEE Asilomar Conf. on Signals, Systems, and Computers*, pp. 1640–1644, Nov. 2003.

[12] S. K. Lim, J. Yen, and P. Wu, "Detection of Out-of-focus Digital Photographs." HP Labs, Palo Alto, Technical Report, HPL-2005-14, Jan. 2005.

[13] 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.

[14] 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.

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

[16] D. Marr and E. Hildreth, "Theory of Edge Detection," in *Proc. Royal Society of London*, vol. 207, pp. 187–217, 1980.

[17] R. C. Gonzalez and R. E. Woods, *Digital Image Processing (2nd Edition)*. Addison-Wesley Pub. Co., Jan. 2002.

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

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

[20] 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.

[21] J. L. Moigne, X. Wei, P. Chalermwat, T. El-Ghazawi, M. Mareboyana, N. Netanyahu, J. C. Tilton, W. J. Campbell, and R. P. Cromp, "First Evaluation of Automatic Registration Methods," in *Proc. IEEE Int. Symposium on Geoscience and Remote Sensing*, vol. 1, pp. 315–317, July 1998.

[22] B. Zitova and J. Flusser, "Image Registration Methods: A Survey," *Elsevier Image and Vision Computing*, vol. 21, pp. 977–1000, June 2003.

[23] S. Banerjee and B. L. Evans, "Unsupervised Automation of Photographic Composition Rules in Digital Still Cameras," in *Proc. SPIE Conf. on Sensors, Color, Cameras, and Sys. for Digital Photography VI*, pp. 364–373, Jan. 2004.

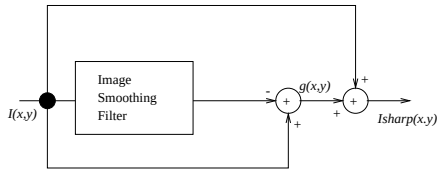


Fig. 1. Model for an image sharpening filter.

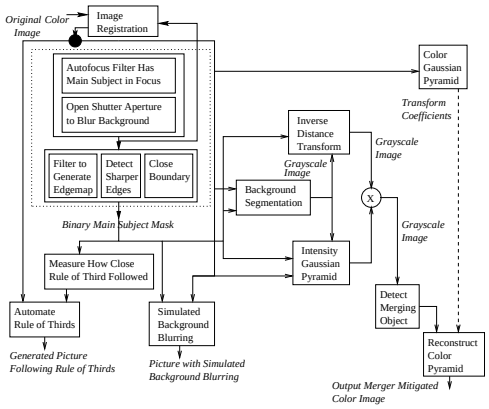


Fig. 5. Proposed automation of selected photograph composition rules for digital still cameras.

[24] S. Banerjee and B. L. Evans, "Unsupervised Merger Detection and Mitigation in Still Images Using Frequency and Color Content Analysis," in *Proc. IEEE Int. Conf. on Acoustics, Speech, and Signal Proc.*, pp. 549–552, May 2004.

[25] J. A. C. Yule and G. G. Field, *Principles of Color Reproduction*. GATFPress, Jan. 2001.

[26] T. J. Flohr, B. W. Kolpatzik, R. Balasubramanian, D. A. Carrara, C. A. Bouman, and J. P. Allebach, "Model Based Color Image Quantization," *Proc. SPIE Human Vision, Visual Proc. and Digital Display IV*, vol. 1913, pp. 270–281, 1993.

[27] C. Zhang and P. Wang, "A New Method of Color Image Segmentation Based on Intensity and Hue Clustering," in *Proc. IEEE Int. Conf. on Pattern Recognition*, vol. 3, pp. 613–616, Sept. 2000.

[28] E. P. Simoncelli, W. T. Freeman, E. H. Adelson, and D. J. Heeger, "Shiftable Multiscale Transforms," *IEEE Trans. on Information Theory*, vol. 38, pp. 587–607, Mar. 1992.

[29] F. Y. Shih and Y. T. Wu, "The Efficient Algorithms for Achieving Euclidean Distance Transformation," *IEEE Trans. on Image Processing*, vol. 13, pp. 1078–1091, Aug. 2004.

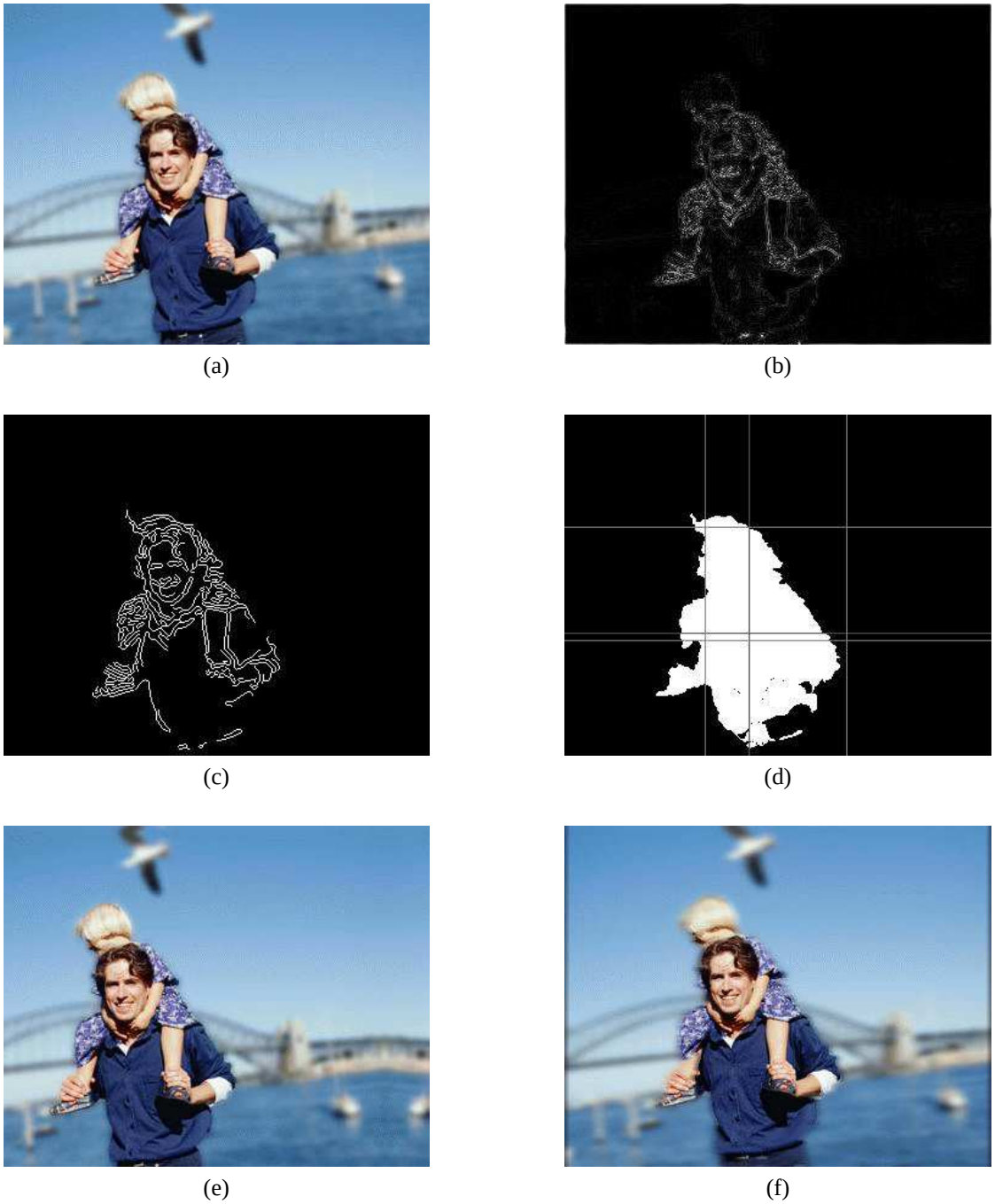
[30] H. Breu, J. Gil, D. Kirkpatrick, and M. Werman, "Linear Time Euclidean Distance Transform Algorithms," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 17, pp. 529–533, May 1995.

[31] O. Cuisenaire and B. Macq, "Fast and Exact Signed Euclidean Distance Transformation with Linear Complexity," in *Proc. IEEE Int. Conf. on Acoustics, Speech, and Signal Proc.*, vol. 6, pp. 3293–3296, Mar. 1999.

[32] J. S. Cardoso and L. Corte-Real, "Toward a Generic Evaluation of Image Segmentation," *IEEE Trans. on Image Processing*, vol. 14, pp. 1773–1782, Nov. 2005.

[33] S. K. Warfield, K. H. Zou, and W. M. Wells, "Validation of Image Segmentation and Expert Quality with an Expectation-Maximization Algorithm," in *Proc. Int. Conf. on Medical Image Computing and Computer-Assisted Intervention*, pp. 298–306, Sept. 2002, Springer-Verlag.

[34] M. Gupta and B. L. Evans, *Rule-Of-Thirds Automation for Digital Still Cameras*. written in C for desktop and digital signal processors, Version 0.1 beta released, http://www.ece.utexas.edu/ ˜bevans/projects/dsc/software/RuleOfThirds0_1beta.zip, June 2004.



(a) (b) (c) (d) (e) (f)

Fig. 2. Automation of photographic composition rules by detecting the main subject, the man and the child, which are in focus: (a) Digital image with background blur from large shutter aperture; (b) Sharper edges are prominent in the filtered image; (c) Rough outline of main subject; (d) Detected main subject mask, with center of mass not following the rule-of-thirds; (e) Generated picture obeying rule-of-thirds; and (f) Simulated background blur which could result from camera panning. The mask could be also dilated before applying the motion blur filter to prevent over-blurring of edges. The full-resolution images are available at http://www.ece.utexas.edu/˜bevans/students/phd/serene_banerjee/Pictures/

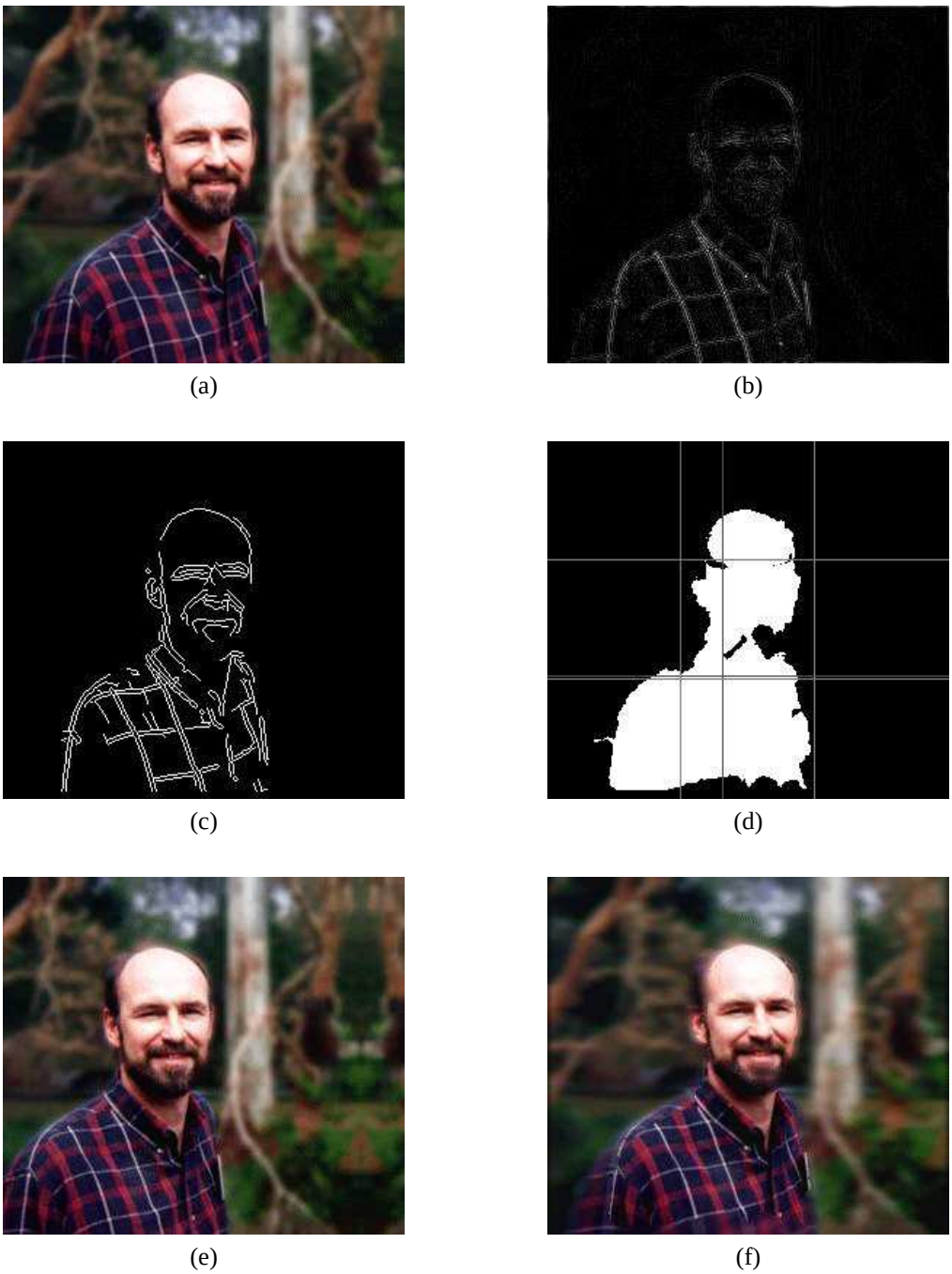


Fig. 3. Automation of photographic composition rules by detecting the main subject, the man, which is in focus: (a) Digital image with background blur from large shutter aperture; (b) Sharper edges are prominent in the filtered image; (c) Rough outline of main subject; (d) Detected main subject mask, with center of mass not following the rule-of-thirds; (e) Generated picture obeying rule-of-thirds; and (f) Simulated background blur which could result from camera panning. The mask could be also dilated before applying the motion blur filter to prevent over-blurring of edges. The full-resolution images are available at http://www.ece.utexas.edu/~bevans/students/phd/serene_banerjee/Pictures/

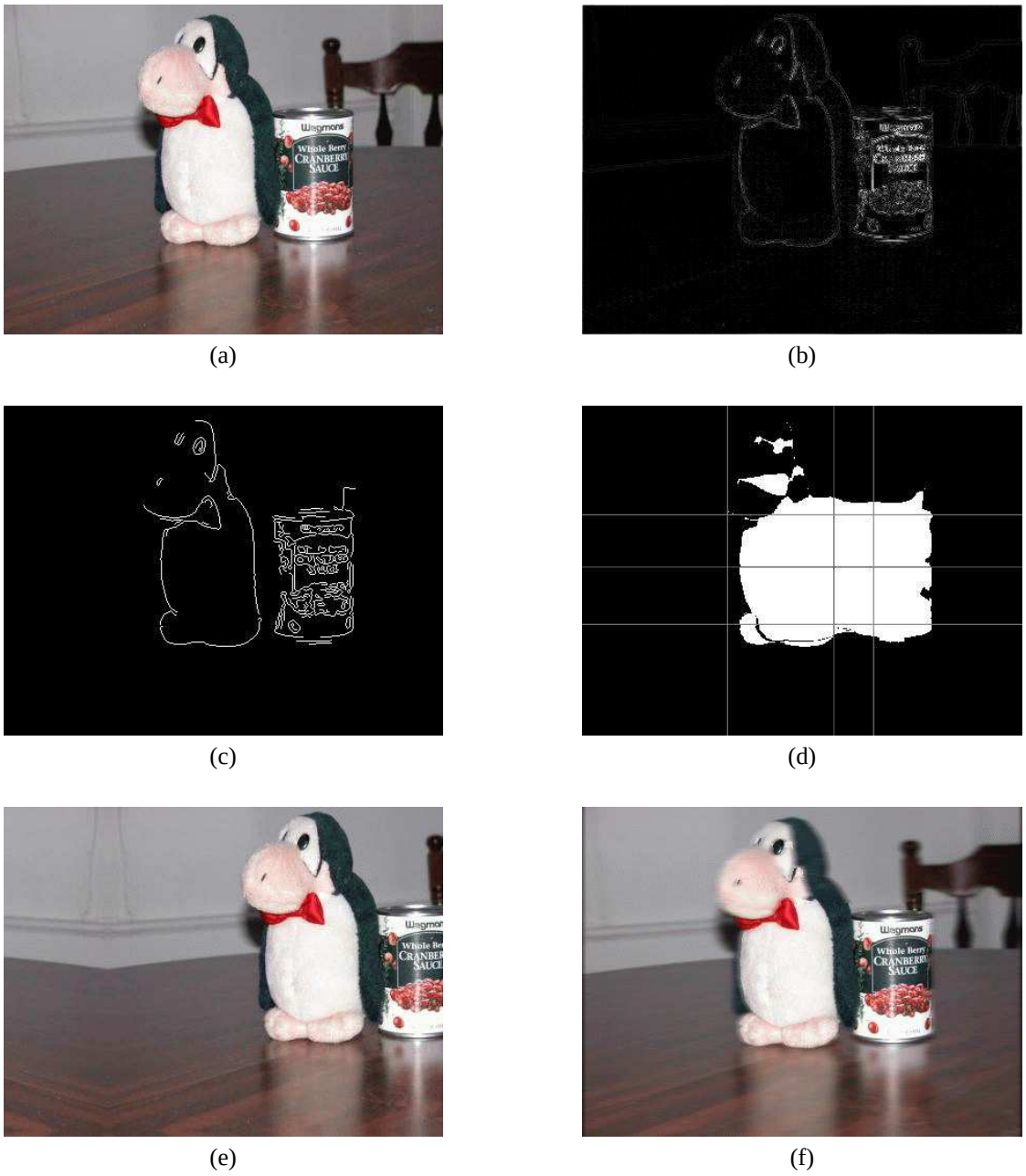


Fig. 4. Automation of photographic composition rules by detecting the main subject, the stuffed animal, which is in focus: (a) Digital image with background blur from large shutter aperture; (b) Sharper edges are prominent in the filtered image; (c) Rough outline of main subject; (d) Detected main subject mask, with center of mass not following the rule-of-thirds; (e) Generated picture obeying rule-of-thirds; and (f) Simulated background blur which could result from camera panning. The mask could be also dilated before applying the motion blur filter to prevent over-blurring of edges. The full-resolution images are available at http://www.ece.utexas.edu/~bevans/students/phd/serene_banerjee/Pictures/

PLACE PHOTO HERE

Serene Banerjee Serene Banerjee (M'04-SM'99) received the B. Tech. degree in electronics and electrical communications engineering from Indian Institute of Technology. Kharagpur, in 1999, and the MS and PhD degrees in electrical engineering from the University of Texas, Austin, TX USA, in 2001 and 2004, respectively.

From 2004 to 2006, she was a TCAD engineer at Intel, Inc., Portland and senior software engineer at Texas Instruments, Inc., Bangalore, India. Since 2006, she is a research scientist at HP research labs in Bangalore, India.

Her research interests include digital image and video processing, document image processing, and digital signal processor architectures.

PLACE PHOTO HERE

Brian Evans Brian L. Evans (M'87-SM'97) received the BS degree in electrical engineering and computer science from the Rose-Hulman Institute of Technology, Terre Haute, IN USA, in 1987, and the MS and PhD degrees in electrical engineering from the Georgia Institute of Technology, Atlanta, GA USA, in 1988 and 1993, respectively.

From 1993 to 1996, he was a post-doctoral researcher at the University of California, Berkeley. In 1996, he joined the faculty at The University of Texas at Austin. He holds the rank of Professor.

Prof. Evans has published more than 170 peer-reviewed conference and journal papers. He is an Associate Editor of the IEEE Transactions on Signal Processing and the IEEE Transactions on Image Processing. He is a member

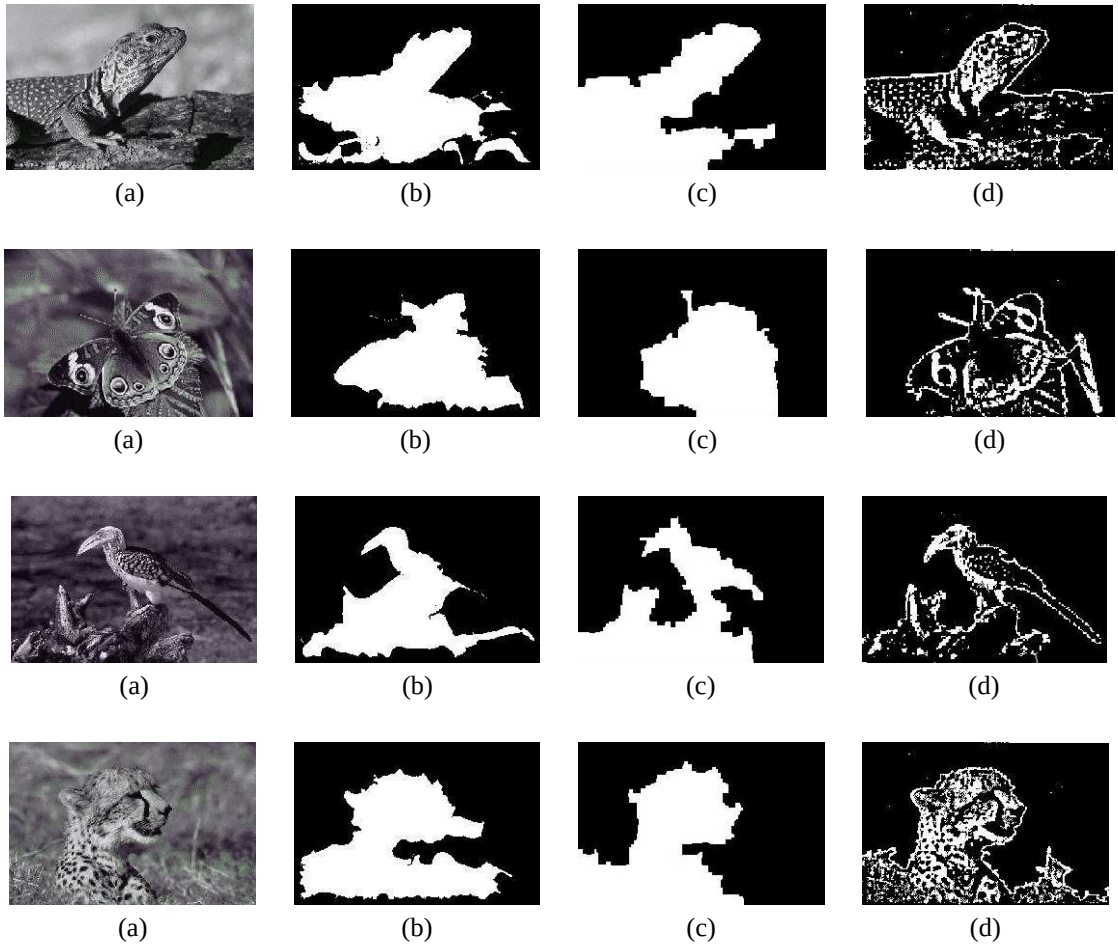


Fig. 12. Comparison of the proposed method with prevalent methods for main subject detection: (a) Original image, with the main subject (the alligator, butterfly, bird and tiger, respectively) in focus; (b) Detected mask of the main subject with the proposed low–implementation complexity one–pass algorithm; (c) Detected mask by Wang's *et. al.* multiscale wavelet-based approach [4], [5]; and (d) Detected main subject by Won's *et. al.* maximum a posteriori probability estimation approach [7] (the authors fill the segmented region with original gray levels for visual inspection).

of the Design and Implementation of Signal Processing Systems Technical Committee of the IEEE Signal Processing Society. He is the recipient of a 1997 US National Science Foundation CAREER Award.

(a) (b)

Fig. 6. Examples of (a) a merger of the main subject, the man, with the trees in the background (in color) and (b) the detected main subject mask in (a).

Fig. 7. The detected merged region is processed in the frequency domain to reduce the effect of the merger. The blurred trees induce a sense of distance.

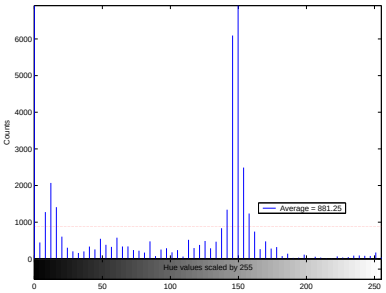


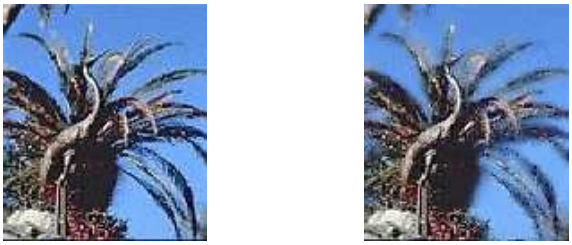
Fig. 8. Histogram of the hue values for the background of Fig. 6(a), which shows the average and peaks.

(a) Object 1 (b) Object 2 (c) Object 3

Fig. 9. Some of the background objects (segmented by color content) for Fig. 6(a) identified by the color background segmentation.

(a) (b)

Fig. 10. (a) The Euclidean distance transform coefficients and (b) the high frequency coefficients from the first level of the Gaussian pyramid for Fig. 6(a). The background object is detected to be merged if it yields the maximum of the weighted sum of (a) and (b).



(a) Original image (b) Merger reduced

Fig. 11. The proposed algorithm reduces the effect of the merger of the tree with the bird. The blurred trees in the processed image are distinguishable as a separate object from the main subject.