Transactions Letters

An Efficient Algorithm for Focus Measure Computation in Constant Time

Rashid Minhas, Abdul Adeel Mohammed, and Q. M. Jonathan Wu, *Senior Member, IEEE*

***Abstract*—This letter presents an efficient algorithm for focus measure computation, in constant time, to estimate depth map using image sequences acquired at varying focus. Two major factors that complicate focus measure computation include neighborhood support and gradient detection for oriented intensity variations. We present a distinct focus measure based on steerable filters that is invariant to neighborhood size and accomplishes fast depth map estimation at a considerably faster speed compared to other well-documented methods. Steerable filters represent architecture to synthesize filters of arbitrary orientation using a linear combination of basis filters. Such synthesis is helpful to analytically determine the filter output as a function of orientation. Steerable filters remove inherent limitations of traditional gradient detection techniques which perform inadequately for oriented intensity variations and low textured regions.**

***Index Terms*—Depth map, focus measure, integral image, shape from focus, steerable filters.**

I. Introduction

In general, shape from focus (SFF) based techniques utilize a sequence of images to reconstruct a well-focused image and the shape, also termed as depth map, of an object. Each reconstruction yields a single matrix, one for each of well-focused image and depth map, similar to the size of an input frame. Please note that a well-focused image embodies a pool of the best focused pixels extracted from different frames of an input sequence, whereas, a depth map represents the relative distance between a camera lens and the focused point of an object.

SFF based techniques are mainly divided into two categories, i.e., *transform* and *non-transform* based SFF. Transform based SFF includes application of fourier domain [1], discrete cosine [2], wavelet and curvelet transform [3]–[6] to extract focused points from different images to construct a composite image and estimate depth map of a scene. Malik and Choi [7] proposed an optical focus measure (FM_O) that uses optical transfer function in Fourier domain for depth map estimation and 3-D shape recovery in the presence of noise.

Manuscript received April 9, 2010; revised August 16, 2010 and December 18, 2010; accepted March 3, 2011. Date of publication March 28, 2011; date of current version January 6, 2012. This paper was recommended by Associate Editor P. Yin.

The authors are with the Department of Electrical Engineering, University of Windsor, Windsor, ON N9B3P4, Canada (e-mail: minhasr@uwindsor.ca; mohammea@uwindsor.ca; jwu@uwindsor.ca).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TCSVT.2011.2133930

The *non-transform* based SFF is subdivided into three categories: 1) *gradient*; 2) *statistical*; and 3) *approximation* based techniques. In pioneering work, Tenenbaum proposed a Sobel operator based Tenengrad focus measure [8], (FM_T), that exploits relationship between well focused image points and their information content. Nayyar and Nakagawa [9] proposed gradient based approach, sum-of-modified-Laplacian ($\mathrm{FM}_{\mathrm{SML}}$), which was later improved using polynomial approximation, i.e., focused image surface in [10] to accurately follow the true structure of the object. In curvature focus measure (FM_C) [11] gray level values are treated as 3-D surface termed as *curvature map*, whereas M_2 focus measure (FM_{M_2}) is based on energy of image gradient. Recently, Ahmed and choi proposed a technique to extract a focused point using 3-D summation based on $\mathrm{FM}_{\mathrm{SML}}$ [12].

The subclass of *non-transform* based SFF that utilizes *statistical* approach consists of focus measures like gray level variance ($\mathrm{FM}_{\mathrm{GLV}}$) [10], mean focus measure ($\mathrm{FM}_{\mathrm{Mean}}$) [11] and entropy based fusion. The last subcategory, *approximation* based SFF comprises focus measure computation using neural networks [13], [14], neuro-fuzzy systems [15], dynamic programming [16], polynomial approximation [10], and regression [17]. Approximation based techniques use anyone of the conventional aforementioned focus measures for preprocessing, whereas, computational complexity, comprehensive rule base, and training data restrict their application to limited domains.

In this letter, a new focus computation scheme is proposed to search the frame number for the best focused object points. For proper orientation, steerable filters respond with large amplitude to high frequency intensity variations in an image. However, the steerable filters response to similar image region might be low for poorly selected orientation. Therefore, high amplitude response information of steerable filters, for an arbitrary orientation, is exploited to extract focused image points from different frames in order to construct a single well-focused image. Such an image contains all points of interest on a focus plane unlike conventional image acquisition scheme.

This letter is organized into five sections. Section II presents steering theory followed by our proposed algorithm in Section III. Section IV discusses experiments and results followed by concluding remarks in the last section.

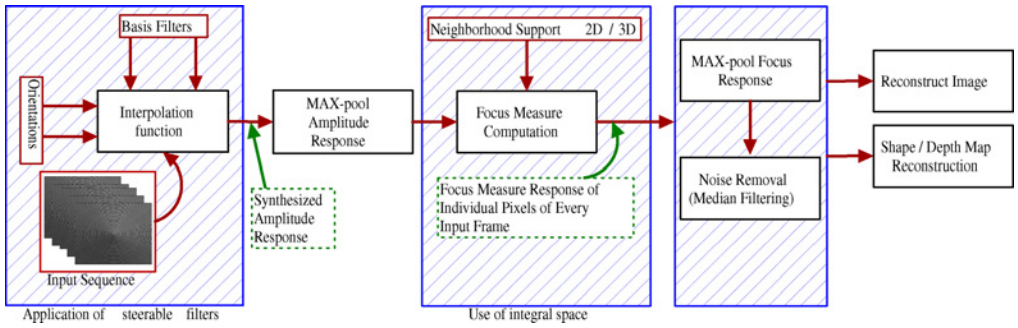

Fig. 1. System block diagram of our proposed method.

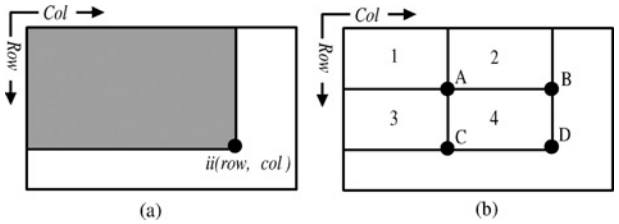

Fig. 2. Image integral to (a) calculate rectangular area (shaded), (b) compute sum of pixels within sub-window 4.

II. Steerable Filters

The term *steerable filters* represent a class of filters which are synthesized as a linear combination of a set of *basis filters*. One often needs to apply the same filter rotated at different angles for various image processing applications. Using a correct filter set and an interpolation rule, it is possible to determine the response of a filter for an arbitrary orientation without explicitly applying that filter [18]. Consider a 2-D circularly symmetric Gaussian function G written in Cartesian coordinates x and y

$$G(x, y) = e^{-(x^2+y^2)} \tag{1}$$

where scaling and normalization constants are set to 1 for convenience. Let us write the nth derivative of a Gaussian in the x direction as Gn. Let $(.)^\theta$ represent the rotation operator such that for any function $f(x, y)$, $f^\theta(x, y)$ is $f(x, y)$ rotated through an angle θ about the origin. It is straightforward that a filter G_1 at any arbitrary orientation θ can be synthesized using a linear combination of $G_1^{0^0}$ and $G_1^{90^0}$

$$G_1^\theta = cos(\theta)G_1^{0^0} + sin(\theta)G_1^{90^0} \tag{2}$$

where $G_1^{0^0}$ and $G_1^{90^0}$ are termed as *basis filters* since they span the set of G_1^θ filters. The $cos(\theta)$ and $sin(\theta)$ terms are the corresponding interpolation functions for the basis filters. The ability to interpolate such derivatives offers further freedom to attain translational and rotational invariant detection of image gradients.

III. Proposed Method

We introduce two new focus measures based on steerable filters, i.e., FM_S and 3-D-FM_S exploiting 2-D and 3-D neighborhood support, respectively. Reliable extraction of focused image points is heavily dependent upon orthogonality constraint between orientation of the feature and gradient detector. Our proposed method uses steerable filters for depth map estimation using a sequence of images acquired at varying focus plane. The basis functions for steerable filters are directional derivatives which come into different sizes and orientations.

As shown in Fig. 1, the implementation of our proposed method requires a sequence of α pre-registered images of equal size (i.e., $Row \times Col$) acquired at varying focus settings. Predefined sizes of basis filters along with β desired orientations are also input to our algorithm. To determine gradient information of an image, our steerability spans over first half of the cartesian coordinates, whereas analogous response prevents the use of additional orientations from remaining quadrants. For accurate depth map calculation, β amplitude responses for each pixel are MAX-pooled which leads to a matrix of size $Row \times Col$ corresponding to the highest gradient information at one of the desired orientations.

A. Integral Space for Focus Measure Computation in Constant Time

The concept of integral space (image/volume) [20] can be used for focus measure computation which can considerably decrease the computational burden for SFF schemes in spatial domain. The integral image at location row, col contains the sum of pixels above and to the left of row, col inclusive [shown as shaded area in Fig. 2(a)]

$$ii(row, col) = \sum_{row'=1}^{row} \sum_{col'=1}^{col} I_O(row', col') \tag{3}$$

where $ii(,)$ is the integral image and $I_O(,)$ is the original image. For fast computation of $ii(,)$ the following pair of recurrences are used:

$$cs(row, col) = cs(row, col - 1) + I_O(row, col) \tag{4}$$

$$ii(row, col) = ii(row - 1, col) + cs(row, col) \tag{5}$$

where $cs(row, col) = 0 \; \forall \; col \in \mathbb{Z}^-$ and $ii(row, col) = 0 \; \forall \; row \in \mathbb{Z}^-$. The cumulative row sum $cs(row, col)$ and the integral image can be computed in one pass over the original image. Using the integral image, any rectangular sum can be computed with four array references. In Fig. 2(b), the sum of the pixels within rectangle 4 can be computed using four array references as follows:

$$Sum = D + A - (B + C). \tag{6}$$

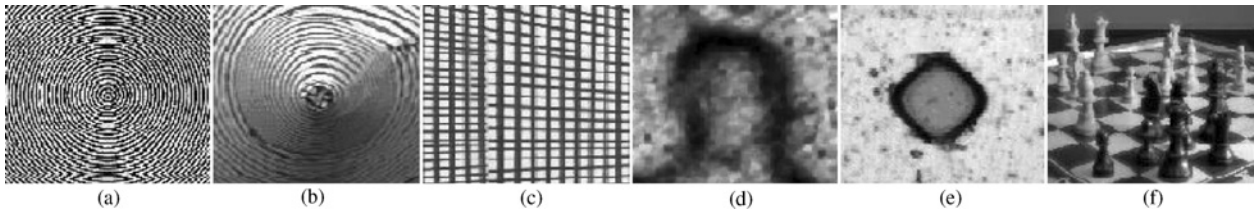

Fig. 3. Reconstructed images using our proposed method. (a) Simulated cone. (b) Real cone. (c) Slanted plane. (d) Coin. (e) LCD/TFT color filter. (f) Chess data set.

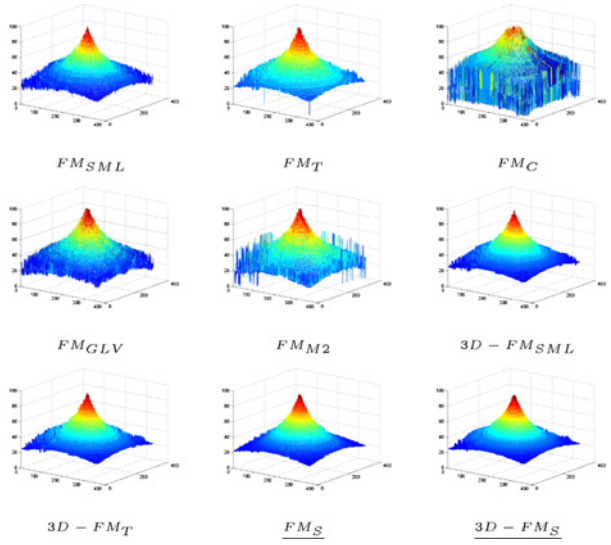

Fig. 4. Depth maps of the simulated cone.

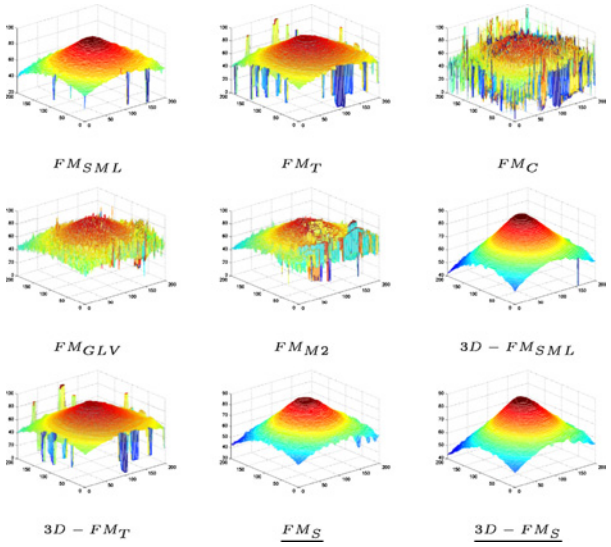

Fig. 5. Depth maps of the real cone.

The concept of integral image can also be extended to higher dimensional space, termed integral volume [21], for focus measure computation using volumetric neighborhood information around a center pixel.

B. Reconstruction of Depth Map and Focused Image

For our proposed scheme, neighborhood information is integrated in constant time using integral space. Please be reminded that the neighborhood summation (2-D and 3-D)

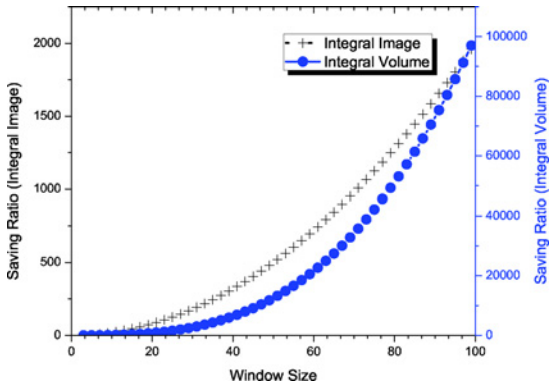

Fig. 6. Analysis of computational efficiency for focus measure computation using integral space.

in focus measure computation is applied on MAX-pooled response of steerable filters for individual images of an input sequence. The focus measure FM_S using a small neighborhood χ of size $N \times N$ around the location (row, col) is computed as

$$FM_S(row, col, i) = \sum_{row', col' \in \chi} I^i_{MP}(row', col', i) \tag{7}$$

where $I^i_{MP}(.)$ represents MAX-pooled amplitude response of ith image. Similarly, 3-D-FM_S utilizes 3-D neighborhood support from stack of $I^i_{MP}(.)$ where $1 \leq i \leq \alpha$

$$3\text{-}D\text{-}FM_S(row, col, i) = \sum_{row', col', i' \in \chi} I^i_{MP}(row', col', i'). \tag{8}$$

The output of both focus measures is the volumetric data equal to the size of input sequence, i.e., $Row \times Col \times \alpha$. Finally, reconstructions of the well-focused image, I_{FI}, and a depth map, I_{DM}, are carried out by selecting pixel locations generating highest focus measure amongst all frames

$$I_{FI}(row, col) = \max_{FM_S(row,col,i)} I_O(row, col, i) \tag{9}$$

where $1 \leq i \leq \alpha$ and $I_O(row, col, i)$ represent frame numbers and intensity values extracted from input sequence, respectively. The well-focused image is constructed using extracted pixels corresponding to the highest focus measure computed in constant time utilizing integral space. The depth map is computed as

$$I_{DM}(row, col) = \max_{FM_S(row,col,i)} i. \tag{10}$$

Finally, to remove noise from data, non-linear median filtering is applied. Fig. 1 presents system block diagram of our proposed scheme. It should be noted that the

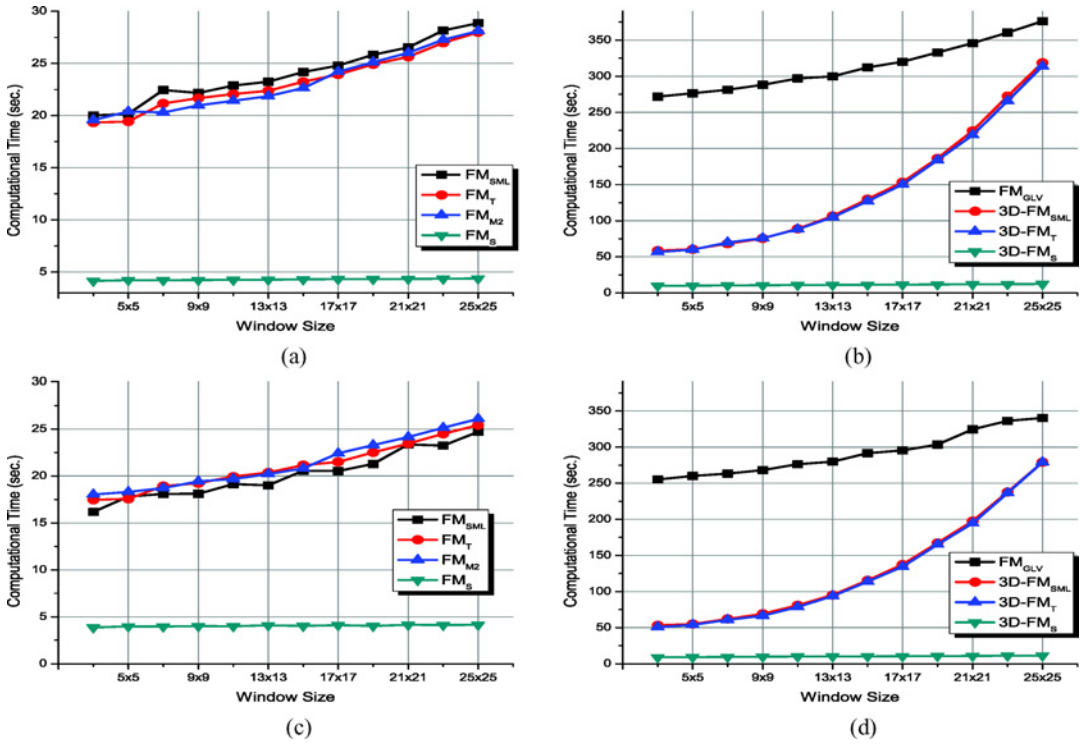

Fig. 7. Computational analysis of various methods for varying neighborhood support. (a) Real cone. (b) Real cone. (c) Slanted plane. (d) Slanted plane.

reconstruction criterion is similar to (9) and (10) for focus measure 3-D-FM_S utilizing 3-D neighborhood support.

IV. Experiments

A. Fundamental Details

The proposed method is tested on comprehensive data sets of real and simulated objects and its performance is compared against well documented SFF methods in literature. Distinctive focus measures compared in our experiments include: 1) sum-of-modified-Laplacian operator (FM_{SML}); 2) Tenengrad operator (FM_T); 3) curvature focus measure (FM_C); 4) gray level variance focus measure (FM_{GLV}); 5) M2 focus measure (FM_{M2}); 6) 3-D sum-of-modified-Laplacian focus measure (3-D-FM_{SML}); and 7) 3-D Tenengrade focus measure (3-D-FM_T). Malik and Choi [22] proved through evidence based arguments that upper bound for optimized window size is 5×5 to avoid blur and achieve improved localization. For a fair analysis, we also chose a uniform window size, i.e., 5×5 since image sequences used in our experiments and [22] are similar, for localized search and summation to compute various focus measures. Furthermore, all image sets utilized for experiments have 256 gray level values and we assume that the images of an incoming sequence are already registered. We selected six data sets (*simulated cone*, *real cone*, *slanted plane*, *coin*, *LCD/TFT*, and *chess*) with varying texture properties to rigorously test stability and robustness of our proposed method. Further details about data sets are available in [6] and [17].

B. Performance Analysis

In general, SFF techniques are used to generate well-focused images and depth maps using a sequence of images. Fig. 3 shows composite (well-focused) images reconstructed using our proposed focus measure FM_S on six different data sets. We used 2-D neighborhood information to reconstruct images shown in Fig. 3, it is evident that all pixel locations in composite images are sharp and crisp without blurring due to focusing and depth issues of a scene. Such images are generated by selecting pixels that convey highest response to FM_S amongst all frames.

The objective of depth map estimation is to determine depth of every point on an object from the camera lens. For *simulated* and *real cone* data sets, Figs. 4 and 5 plot depth maps generated using various schemes. For *simulated cone* sequence, it is quite obvious from the plots that some of the traditional SFF methods construct depth map with large fluctuations and spikes, which demonstrate their inconsistent and unreliable behavior. Depth maps obtained using the proposed operators, i.e., FM_S and 3-D-FM_S, are smooth with a sharp and prominent tip. Depth maps obtained from FM_{SML} and FM_T focus measures along with their 3-D variants are very close to the depth map obtained using FM_S as well. Fig. 5 shows the depth maps of a *real cone* obtained utilizing various focus measures. The depth maps obtained using FM_S and 3-D-FM_S focus measures are significantly smooth in the vertical direction and closely follow the real cone structure. The depth maps computed using traditional SFF methods have spikes which are not present on a real cone object and thus the actual structure is not reliably tracked. Traditional SFF methods exhibit poor performance for depth map estimation of a real cone object due to superfluous shadows and bad illumination conditions.

1) *Computational Efficiency:* For neighborhood window size of ρ using the concept of integral space, we get a

minimum of $[\rho^2 - 1]/5$ and $[\rho^3 - 1]/10$ folds saving in focus measure computation of 2-D and 3-D data, respectively. The *saving ratio* is obtained as a fraction between computational loads of a focus measure using conventional and integral space methods. Intuitively, higher *saving ratio* values show improved efficiency and lower computational cost and *vice versa*. Fig. 6 shows analysis of computational saving for focus measure, FM_{SML}, computation using integral space instead of conventional summation to integrate neighborhood support. Please note that we use two vertical axes in Fig. 6 (left and right) with varying range of values and incremental steps corresponding to saving ratios of integral image (2-D) and integral volume (3-D), respectively. In Fig. 6, computational saving is drastically increasing for 3-D data compared to 2-D case. The *saving ratios* of our proposed focus measure monotonically increases and is proportional to the size of neighborhood information. It is an important fact that there is no requirement for additional memory to compute focus measure in integral space compared against traditional summation strategy.

The use of integral space allows focus measure computation in constant time. For better realization, we analyze the computational efficiency of our proposed scheme against existing methods for varying neighborhood support size. The computational analysis of various methods is presented in Fig. 7 using two inputs sequences, i.e., *real cone* and *slanted plane*. To avoid redundant information we do not include similar study for the rest of data sets which exhibit a similar trend as well. For the trials under-reference, window size to integrate neighborhood support is increased at equal intervals from 3 to 25 pixels in each dimension. For better visualization, we present our analysis into two columns format. This representation is carried out by dividing various methods into two groups based on their computation time. Such division is helpful to prominently show the trend of computational complexities of individual schemes.

In Fig. 7, it is clearly observed that FM_{GLV} is the most expensive focus measure amongst all SFF schemes, whereas FM_S requires the least resources and renders a steady behavior with an increase in the number of adjacent pixels. Furthermore, 3-D-FM_S is the second most economical focus measure with negligible variations in its computational time for expanding size of neighboring windows. Clearly, remaining focus measures can be roughly divided into two clusters based on their computational complexity: 1) for 2-D case FM_{SML}, FM_T and FM_{M2} approximately consume similar, but dwindling, amount of computational time, and 2) for 3-D case, 3-D-FM_{SML} and 3-D-FM_T require almost similar amount of computation. For different neighboring windows, the complexities of proposed focus measures stay constant whereas an exponential increase in computation is observed for other schemes. Undoubtedly, 3-D focus measures require considerably higher computational resources compared to their 2-D counterparts.

V. Conclusion

In this letter, we introduced a new method for SFF based on steerable filters. We applied steerable filters at different orientations to the sequences of images with varying texture properties. For each pixel location 2-D neighborhood sum of MAX-pooled amplitude response, generated through convolution of oriented steerable filters with the image frame, is exploited to locate focused image points for shape reconstruction. In performance analyses, our proposed method outperforms well documented depth estimation techniques and shows reliable and robust behavior in the presence of superfluous shadows, and differing illumination conditions.

References

[1] B. K. P. Horn, "Focusing," MIT Artif. Intell. Lab., Cambridge, Memo 10, 1968.

[2] M. T. Mehmood, W. J. Choi, and T.-S. Choi, "PCA based method for 3-D shape recovery of microscopic objects from image focus using discrete cosine transform," *Microsc. Res. Tech.*, vol. 71, no. 12, pp. 897–907, Dec. 2008.

[3] H. Li, B. S. Majunath, and S. K. Mitra, "Multisensor image fusion using the wavelet transform," *Graph. Models Image Process.*, vol. 57, no. 3, pp. 235–245, May 1995.

[4] M. Asif, A. S. Malik, and T.-S. Choi, "3-D shape recovery from image defocus using wavelet analysis," in *Proc. Int. Conf. Image Process.*, 2005, pp. 1025–1028.

[5] M. T. Mehmood, S. Shim, and T.-S. Choi, "Wavelet and PCA based approach for 3-D shape recovery from image focus," *Proc. SPIE*, vol. 7073, pp. 70731S–70731S-9, 2008.

[6] R. Minhas, A. A. Mohammed, and Q. M. J. Wu, "Shape from focus using fast discrete curvelet transform," *Pattern Recognit.*, vol. 44, no. 4, pp. 839–853, 2011.

[7] A. S. Malik and T.-S. Choi, "A novel algorithm for estimation of depth map using image focus for 3-D shape recovery in the presence of noise," *Pattern Recognit.*, vol. 41, no. 7, pp. 2200–2225, 2008.

[8] J. M. Tenenbaum, "Accommodations in computer vision," Ph.D. dissertation, Dept. Electr. Eng., Stanford Univ., Stanford, CA, 1970.

[9] S. K. Nayyar and Y. Nakagawa, "Shape from focus," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 16, no. 8, pp. 824–831, Aug. 1994.

[10] T.-S. Choi, M. Asif, and J. Yun, "Three-dimensional shape recovery from focused image surfaces," in *Proc. Int. Conf. Accoust. Speech Signal Process.*, vol. 6. 1999, pp. 3269–3272.

[11] F. S. Helmi and S. Scherer, "Adaptive shape from focus with an error estimation in light microscopy," in *Proc. 2nd Int. Symp. Image Signal Process. Anal.*, 2001, pp. 188–193.

[12] M. B. Ahmed and T.-S. Choi, "Application of 3-D shape from image focus in LCD-TFT display manufacturing," *IEEE Trans. Consumer Electron.*, vol. 53, no. 1, pp. 1–4, Feb. 2007.

[13] M. Asif and T.-S. Choi, "Shape from focus using multilayer feedforward neural network," *IEEE Trans. Image Process.*, vol. 10, no. 11, pp. 1670–1675, Nov. 2001.

[14] W. Huang and Z. Jing, "Multi-focus image fusion using pulse coupled neural network," *Pattern Recognit. Lett.*, vol. 28, no. 9, pp. 1123–1132, Jul. 2007.

[15] A. S. Malik and T.-S. Choi, "Application of passive techniques for 3-D cameras," *IEEE Trans. Consumer Electron.*, vol. 53, no. 2, pp. 258–264, May 2007.

[16] M. B. Ahmed and T.-S. Choi, "A heuristic approach for finding best focused image," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 15, no. 4, pp. 566–574, Apr. 2005.

[17] A. S. Malik and T.-S. Choi, "Finding best focused points using intersection of two lines," in *Proc. Int. Conf. Image Process.*, 2008, pp. 1952–1955.

[18] W. T. Freeman and E. H. Adelson, "The design and use of steerable filters," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 13, no. 9, pp. 891–906, Sep. 1991.

[19] Y. Xiong and S. A. Schafer, "Depth from focusing and defocusing," in *Proc. Int. Conf. Comput. Vis. Pattern Recognit.*, 1993, pp. 68–73.

[20] P. Viola and M. Jones, "Rapid object detection using a boosted cascade of simple features," in *Proc. Int. Conf. Comput. Vis. Pattern Recognit.*, vol. 1. 2001, pp. I-511–I-518.

[21] X. Cui, Y. Liu, S. Shan, X. Chen, and W. Gao, "3-D Haar like features for pedestrian detection," in *Proc. Int. Conf. Multimedia Expo*, 2007, pp. 1263–1266.

[22] A. S. Malik and T.-S. Choi, "Consideration of illumination effects and optimization of window size for accurate calculation of depth map for 3-D shape recovery," *Pattern Recognit.*, vol. 40, no. 1, pp. 154–170, 2007.