3D Shape from Focus and Depth Map Computation Using Steerable Filters

Rashid Minhas, Abdul A. Mohammed, Q.M. Jonathan Wu, and Maher A. Sid-Ahmed

Department of Electrical and Computer Engineering,
Univeresity of Windsor, ON, N9B3P4, Canada
{minhasr,mohammea,jwu,ahmed}@uwindsor.ca

Abstract. The technique utilized to retrieve spatial information from a sequence of images with varying focus plane is termed as shape from focus (SFF). Traditional SFF techniques perform inadequately due to their inability to deal with images that contain high contrast variations between different regions, shadows, defocused points, noise, and oriented edges. A novel technique to compute SFF and depth map is proposed using steerable filters. Steerable filters, designed in quadrature pairs for better control over phase and orientation, have successfully been applied in many image analysis and pattern recognition schemes. Steerable filters represent architecture to synthesize filters of arbitrary orientation using linear combination of basis filters. Such synthesis is used to determine analytically the filter output as a function of orientation. SFF is computed using steerable filters on variety of image sequences. Quantitative and qualitative performance analyses validate enhanced performance of our proposed scheme.

1 Introduction

Depth map is a computation of distance between object points on focus plane and camera lens. The objective of depth map computation is to determine depth of every point on the object from the camera lens. Depth map estimation is a critical problem in computer vision with numerous applications in robot guidance, collision avoidance, 3D feature extraction and pose estimation, medical imaging, range segmentation, microscopic imaging, seismic data analysis and shape reconstruction. Time-of-flight (TOF) sensors are used to compute depth compute by sensing reflected light from scene objects. Noise is dependent on light reflected into the sensor whereas light reflected from surrounding objects also complicates depth map computation. TOF sensors are expensive, and range limitations restrict their utilization to specific applications.

Object points appear sharp in images which are present on focus plane. Blur of image points increases as they move away from the focus plane. For scenes with considerably large depth, points on the focus plane have a sharp appearance while the rest of the scene points are blurred and can be ignored during depth map estimation. It is impossible to have a focus plane similar to the scene depth and obtain sharp focus for all object points. Practically, we capture a sequence of images with varying focus plane to acquire different scene points that are well focused, whereas object portion

M. Kamel and A. Campilho (Eds.): ICIAR 2009, LNCS 5627, pp. 573–583, 2009.

away from the focus plane is blurred with low intensity variations among different image frames. Depth map of a scene is accurately estimated using the relative position of a varying focus plane. Object present on a translation stage is imaged at different distances from CCD camera. Object points present on focus plane will represent high variations in intensity. As shown in fig 1, the object touches the focus plane after p number of translations. At this point the top portion of the object is well focused and the remaining portion is blurred due to its distance from the focus plane. Translation stage is shifted up and towards the CCD camera, until translation stage touches the focus plane. Such translations are useful in capturing focused portions of an object on a planar image detector in different image frames. Translation of object and number of images acquired in a sequence depend upon available resources, required accuracy, object structure, illumination variations and reflections.

Traditional SFF techniques assume convex shaped objects for accurate depth map estimation. Images acquired with varying distances are processed to extract focused points from individual image frames and to reconstruct a well focused image. Benefits offered by SFF include batch processing mode without intelligent processing at individual sensing nodes, passive nature, cheap sensing devices and measurements close to human perception.

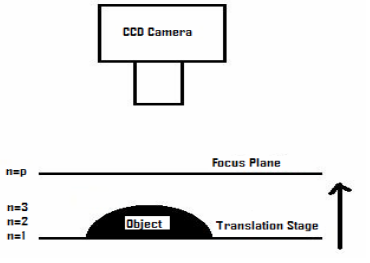


Fig. 1. Image acquisition setup for an image sequence with varying plane

Commonly used operators in SFF based techniques are sum of modified Laplacian [1,2], Tenengrade focus measure [4-6], gray level variance focus measure [7], curvature focus measure [8], M2 focus measure [5], point focus measure [8], optical focus measure [13]. Approximation and learning based focus measures have also been proposed [9, 10-12, 14] that utilize neural network, neuro fuzzy systems and dynamic programming based approaches for better depth map estimation. Approximation based techniques use any of the conventional aforementioned focus measures for preprocessing whereas comprehensive rule base and training data also restrict their application to specific problems for depth map evaluation. Recently 3D summation in local window based focus measure has been proposed by Bilal and Choi [15] which is computationally expensive. In this paper a new focus measure operator is proposed to search the frame number for the best focused object points. Most of the established focus measure operators for SFF work well for regions with dense texture only. Hence their performance deteriorates in presence of noise, poor texture and singularities along curves.

For proper orientation, steerable filters respond with large amplitude to high frequency intensity variations in an image. However, the steerable filters response for

same image region might be low amplitude for poorly selected orientation. Therefore, high amplitude response information of steerable filters is exploited to extract focused image points from different frames to construct a single well-focused image. Such an image contains all points of interest on focus plane unlike conventional image acquisition scheme.

This paper consists of five sections. Sections 2-3 discuss the theory of steerable filters and the proposed algorithm respectively. Section 4 presents the performance analysis of our proposed algorithm with the existing methods, followed by concluding remarks in section 5.

2 Steerable Filters

Filters with arbitrary orientations are used adaptively to examine filter response as a function of time and phase in many industrial applications. Oriented filters are used in numerous image processing and computer vision applications such as image compression, segmentation, edge detection, texture and motion analysis, and image enhancement [16-19].

The main goal of an approach to find the response of a filter for different orientations is to analyze its output at various angles. A computationally efficient approach is required to interpolate the filter response obtained using basis filters at different angles. With the application of a correct filter basis set and the use of a suitable interpolation rule it is possible to evaluate filter response at a particular orientation implicitly. Steerable filters [23] refer to randomly oriented filters synthesized using a linear combination of the basis filters. Once the basis filter responses are known, the response of the filter to an arbitrary steered angle can be determined easily. In this work we have used a 2D circularly symmetric Gaussian function. Consider the following Gaussian function G represented in Cartesian coordinates x and y as:

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

To keep things simple, scaling and normalization constants are set to unity. The directional derivative operator can be rotated as shown in [20-22]. If we denote the n^{th} derivative of a Gaussian function in the x direction as G_n. Let $(\ldots)^{\theta}$ represents the rotation operator such that $f^{\theta}(x,y)$ represents a rotated version of $f(x,y)$ at an angle θ about origin. First derivative of a Gaussian function G along x is represented by equation (2) and the same function rotated at 90° is shown by equation (3).

$$G_1^{0^\circ} = \frac{\partial}{\partial x} e^{-(x^2+y^2)} = -2xe^{-(x^2+y^2)} \tag{2}$$

$$G_1^{90^\circ} = \frac{\partial}{\partial y} e^{-(x^2+y^2)} = -2ye^{-(x^2+y^2)} \tag{3}$$

It is clearly evident that a G_1 filter with an orientation θ can be synthesized using a linear combination of $G_1^{0^\circ}$ and $G_1^{90^\circ}$ Gaussian filters using the following mathematical equation for synthesis:

$$G_1^{\theta} = \cos(\theta) G_1^{0^\circ} + \sin(\theta) G_1^{90^\circ} \tag{4}$$

$G_1^{0^\circ}$ and $G_1^{90^\circ}$ are referred to as basis functions since they span the entire set of G_1^{θ} filters and the interpolation functions used for synthesizing arbitrary steerable filters are *cos(θ)* and *sin(θ)*. Since convolution is a linear operation, an image filtered at any arbitrary orientation is synthesized by taking a linear combination of the images convolved with $G_1^{0^\circ}$ and $G_1^{90^\circ}$ respectively as below:

$$R_1^{0^\circ} = G_1^{0^\circ} * I \tag{5}$$

$$R_1^{90^\circ} = G_1^{90^\circ} * I \tag{6}$$

$$R_1^{\theta} = \cos(\theta) R_1^{0^\theta} + \sin(\theta) R_1^{90^\theta} \tag{7}$$

A Simple example of steerable filters [23] and its synthesis with respect to the equation labeled above is shown in Fig. 2. Fig. 2(a) represents $G_1^{0^\circ}$ i.e. the first derivative of the Gaussian function *G* with respect to the horizontal *x* axis and similarly Fig. 2(b) represents a 90° rotated version of $G_1^{0^\circ}$. Fig. 2 (c) represents a filter synthesized through a linear combination of basis functions, i.e. Fig. 2 (a-b), generated according to eqn. 4. Fig. 2 (d-f) show the image of the circular disk and convolution of the circular disk with $G_1^{0^\circ}$ and $G_1^{90^\circ}$ respectively. Lastly, Fig. 2(g) is a linearly weighted combination according to eqn. (7) assuming 60° orientation.

Steerable filters offer significant advantages in image analysis over ad-hoc methods of combining randomly oriented filters at different orientations since its synthesis is analytic and exact. Steerable filters are designed as quadrature pair filters and are extremely useful in orientation analysis, adaptive filtering, enhancement of oriented structures, and contour detection. The processing scheme does not require additional convolution after the initial pass and the contour detector utilizes quadrature pairs to mark lines and edges with a single response.

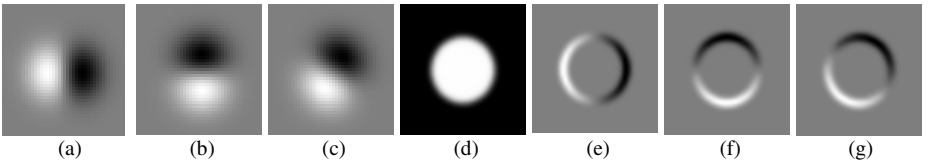


Fig. 2. Steerable filters and its application to circular disk image

3 Proposed Algorithm

Our proposed method uses steerable filters for depth map estimation and reconstruction of a well-focused image from a sequence of images acquired with varying focus plane. Steerable filters, designed for improved control over phase and orientation, perform better for images with low texture information where conventional SFF methods fail to accurately estimate depth map. Size of basis function, i.e. Gaussian function, can significantly affect the performance of steerable filters. Hence size of Gaussian function must be carefully selected to avoid over smoothing and noise effects. Images of a sequence, for construction of well focused single image, are

Table 1. Proposed algorithm using steerable filters

```
[depth_map, reconstructed_image] = SFF_SteerableFilters(Input_Sequence_of_images^M,
                                   size_of_Gaussian_filter, theta_list^N,neighborhood_window_size)
{
        Size=neighbourhood_window_size;
        for orientation_number=1 to N do;
                G^θ_orientation_number = cos(θ)G_1^0° + sin(θ)G_1^90°
        end;
        for image_number=1 to M do;
          for orientation_number=1 to N do
                intermediate_images[orientation_number]=input_sequence_of_images
                                                        [image_number]* G^θ_orientation_number
          end;
          for R=1 to Row do;
                for C=1 to Col do;
                        Processed_Image[image_number]=max(intermediate_images[R,C,:])
                end;
          end;
        end;
        for image_number=1 to M do
                for R=1 to Row do
                        for C=1 to Column do
                                Processed_Image[R,C,image_number]= Σ_{j=R-Size/2}^{R+Size/2} Σ_{k=C-Size/2}^{C+Size/2}
                                                        Processed_Image[j,k,image_number];
                        end;
                end;
        end;
        for image_number=1 to M do;
                for R=1 to Row do;
                        for C=1 to Col do;
                                Final_Image[R,C]=max(Processed_Image[R,C,:])
                        end;
                end;
        end;
        depth_map=median(Final_Image);
        reconstructed_image=Image_Reconstruction(depth_map);
}
```

assumed to be registered. SFF based techniques perform well for convex objects in a scene. Implementation of our proposed method, described in table – 1, requires a sequence of M images of same size (i.e. RowxCol), neighborhood size for 2D window sum to avoid measurement errors, and the size of basis filter along with a list of N desired orientations for steerable filters.

Section 2 discusses in detail theory of steerable filters. Initially, Gaussian coefficients for different orientations are computed, using eqns. (2)-(4) and convolved with the original image. Such convolutions result into N different gradient information obtained at different orientations. Images are convolved with steerable filters at orientations 45^0 apart from its preceding orientation i.e. {0^0, 45^0, 90^0, 135^0, 180^0, 225^0, 270^0, and 315^0}. Higher amplitude responses are obtained for edge orientations perpendicular to filter's orientation. For best focused image reconstruction and accurate depth map calculation, the highest among N amplitude responses for each pixel value

is chosen which results into RowxCol size coefficients corresponding to the highest gradient information at varying orientations. In a similar way, we compute gradient information for M input images. Next, focus measure for each frame is computed by 2D neighborhood sum on pre-computed gradient information for each pixel to avoid abrupt fluctuations, illumination variations, and measurement errors. At each pixel, the image frame that gives a maximum sharpness measure is selected. This corresponds to piecewise constant approximation of the actual focused image. Finally, to remove noise from data, non-linear median filtering is applied. Table – 1 provides step-by-step implementation of our proposed scheme.

4 Experiments

The proposed method is tested on comprehensive data sets of real and simulated objects and its performance is compared to the well documented SFF methods in literature. Extensive experiments were performed and results of three experiments are presented in this paper to analyze the performance of the proposed method with traditional SFF techniques both quantitatively and qualitatively. Distinctive focus measure operators compared in our experiments include sum of modified Laplacian operator (FM_{SML}), Tenengrad operator (FM_T), Curvature focus measure (FM_C), Gray level variance focus measure (FM_{GLV}) and M_2 focus measure (FM_{M2}). We selected above mentioned methods for comparison since they are the most widely used and exclusive focus measure operators for shape reconstruction and depth map estimation [1-6, 8, 13, 15]. Learning based focus measure approaches proposed by Asif and Choi [11], and Malik and Choi [14] utilized FM_{GLV}, FM_{SML} methods respectively for initial depth estimation; whereas approximation [7, 9, 10] and DP based [12] techniques used the Laplacian based focus measure for initial processing. Data sets with different texture properties are used to rigorously test the stability and robustness of our proposed method. For precise performance analysis in all experiments, we chose a uniform window size, i.e. 5x5, for localized search and summation to compute focus measure values. Malik and Choi [13] proved that upper bound for optimized window size is 5x5 to avoid blur, accurate calculation of depth map and 3D shape recovery.

Quantitative Performance Analysis

Fig. 3 shows images of simulated cone generated using the simulation software by varying distances between the object and the camera, and the reconstructed image using the proposed method. Fig. 3 (a-e) shows the images of the cone with varying focus planes. In each figure only some portion of the object resides on focus plane which is well focused and appears sharp in an image. However portion of the simulated cone object that does not exist on focus plane appears blurred. Reconstructed image of the simulated cone obtained using the proposed steerable filter focus measure, FM_{SF}, operator is shown in Fig. 3(f) and it is clearly evident that all the pixel locations in the image are sharp and crisp. The reconstructed image is generated by selecting pixels amongst all images of a sequence that provide a maximum response to the FM_{SF} operator. Pixel values from different frames of sequence are pooled together to regenerate a well focused image. Similarly Figs. 4-5 show the image frames of a real cone, and a slanted planar object imaged at different lens positions controlled

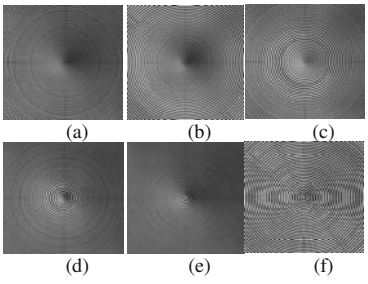


Fig. 3. (a-e) Image frames of a simulated cone (f) Reconstructed image

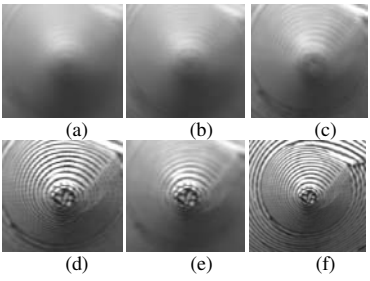


Fig. 4. (a-e) Image frames of a real cone (f) Reconstructed image

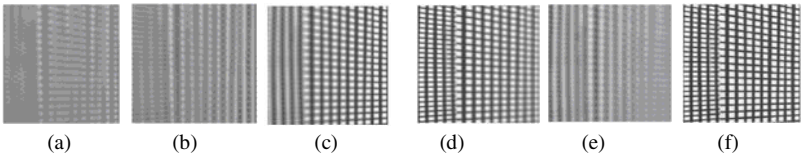
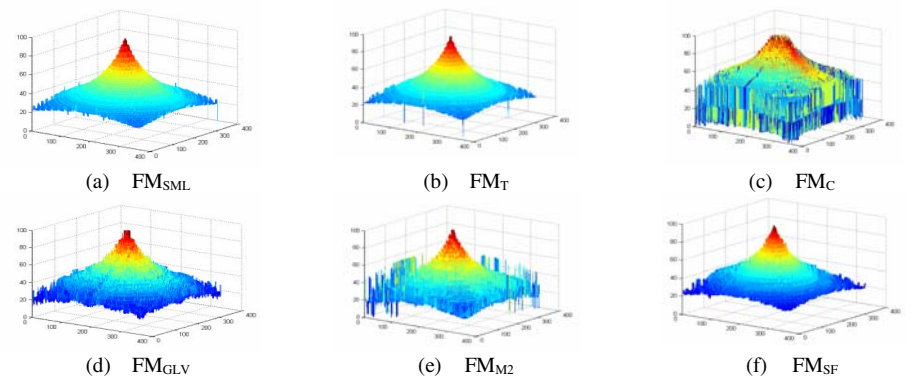


Fig. 5. (a-e) Image frames of a planar object (f) Reconstructed image

by an automated motor. The reconstructed images of the real cone, and the slanted planar object using the FM_{SF} operator are shown in Figs. 4(f)- 5(f) respectively. The reconstructed images are well focused with sharp intensities and minimal blur effect.

Figs. 6-8 plot the depth maps obtained using the FM_{SF} focus measure and traditional SFF methods for image sequences of a simulated cone, a real cone, and a slanted planar object respectively. In case of an ideal simulated cone it is expected that the depth map should be smooth without spikes and must contain a sharp tip. The assumption of depth map smoothness for simulated cone is rational due to controlled lighting conditions without superfluous shadows and no measurement errors. Fig. 6(a-e) shows the depth maps obtained using the traditional SFF methods. It is quite obvious from the plots that some of the traditional SFF methods construct depth map with large fluctuations and spikes which demonstrate inconsistent and unreliable behavior of the respective methods. Depth map obtained using the proposed FM_{SF} operator is smooth with a sharp and prominent tip. Depth maps obtained from FM_{SML} and FM_{T} focus measures are very close to the depth map obtained using FM_{SF}. Fig. 7 shows the depth maps of a real cone obtained utilizing the traditional SFF methods and the proposed FM_{SF} focus measure. The depth map obtained using FM_{SF} focus measure is significantly smooth in the vertical direction and closely follows the real cone structure. The depth maps computed using traditional SFF methods have spikes which are not present on a real cone object and actual cone structure is not being tracked. Traditional SFF methods exhibit poor performance for depth map estimation of a real cone object due to superfluous shadows and bad illumination conditions. Our proposed method renders robust behavior by tracking original structure of a real cone with minimum distortion in 3D depth map. Depth maps of a slanted planar object generated using FM_{SF} focus measure and traditional SFF methods are shown in Fig. 8. The depth map computed with traditional SFF methods are shown in Fig. 8(a-e) and it is obvious that FM_{SF} focus measure clearly outperforms

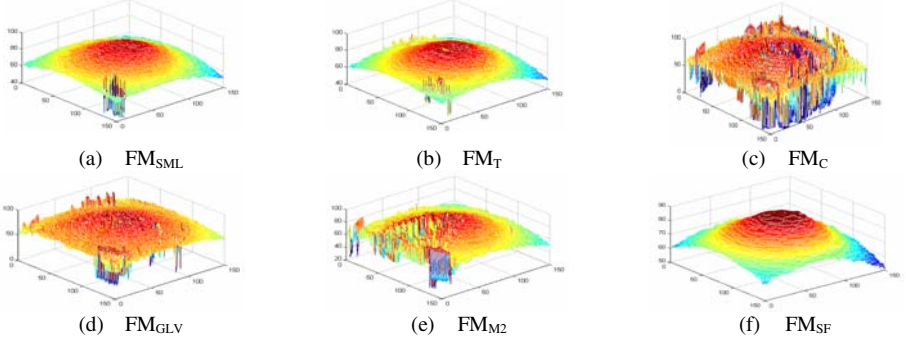
traditional SFF schemes. The depth map obtained with the FM_{SF} focus measure is very smooth, contains less number of discontinuities and closely resembles the actual structure of a slanted planar object.



(a) FM_{SML} (b) FM_T (c) FM_C

(d) FM_{GLV} (e) FM_{M2} (f) FM_{SF}

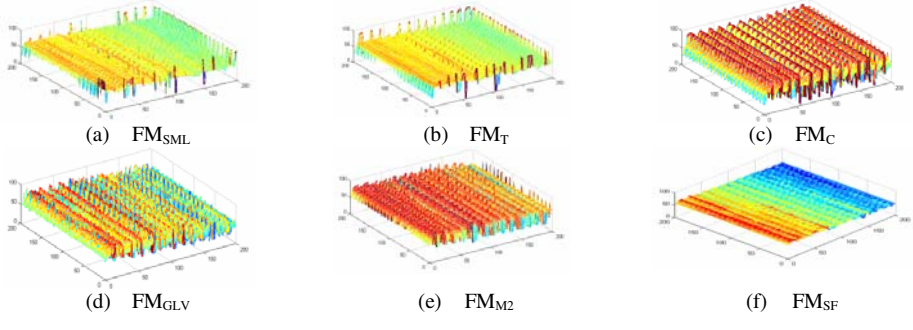
Fig. 6. Depth maps of a simulated cone object using various focus measure operators



(a) FM_{SML} (b) FM_T (c) FM_C

(d) FM_{GLV} (e) FM_{M2} (f) FM_{SF}

Fig. 7. Depth maps of a real cone object using various focus measure operators



(a) FM_{SML} (b) FM_T (c) FM_C

(d) FM_{GLV} (e) FM_{M2} (f) FM_{SF}

Fig. 8. Depth maps of a slanted planar object object using various focus measure operators

Qualitative Performance Analysis

The quality of various image reconstruction techniques depends upon how close is the reconstructed image to the reference image. Mean square error (MSE) and correlation coefficient (CC) are two celebrated metrics used to examine the qualitative performance of different techniques. Reference depth maps are used to compute MSE and CC in order to evaluate qualitative performance. MSE is a distortion metric that provides a measure of dissimilarity between the two depth maps and is evaluated as:

$$MSE = \frac{1}{R*C}\sum_{i=1}^{R}\sum_{j=1}^{C}\left|f(i,j)-f^{|}(i,j)\right|^2 \tag{8}$$

where R indicates the number of image rows and C refers to the number of columns, $f(,)$ represents the reference depth map whereas $f^{|}(,)$ refers to the computed depth map. CC provides a measure of similarity between the reference 3D depth map and the reconstructed depth map. CC is expressed as:

$$CC = \frac{\sum_{i=0}^{R}\sum_{j=0}^{C}\left(f(i,j)-\overline{f(i,j)}\right)\left(f^{|}(i,j)-\overline{f^{|}(i,j)}\right)}{\sqrt{\left(\sum_{i=0}^{R}\sum_{j=0}^{C}\left(f(i,j)-\overline{f(i,j)}\right)^2\right)\left(\sum_{i=0}^{R}\sum_{j=0}^{C}\left(f^{|}(i,j)-\overline{f^{|}(i,j)}\right)^2\right)}} \tag{9}$$

$f(,)$ and $\overline{f(,)}$ represent the reference depth map and its mean; $f^{|}(,)$ and $\overline{f^{|}(,)}$ correspond to the depth map obtained using respective focus measure operators and its mean respectively.

Table – 2 represents MSE and CC of different SFF based techniques and our proposed method. Experimental results clearly validate superior performance of our proposed method with least MSE and highest CC. In case of a simulated cone the results obtained with FM_{SF} operator are less prominent in comparison with other traditional SFF methods since the simulated cone is analogous to the ideal case with high contrast, dense texture. However, the results obtained for a real cone object, and a planar object using FM_{SF} focus measure are more pronounced and demonstrate a significant improvement over traditional SFF methods.

Table 2. Qualitative performance analysis of different focus measure methods

	Simulated Cone		Real Cone		Planar Object	
	MSE	CC	MSE	CC	MSE	CC
FM_{SML}	3.729	0.986	2.083	0.981	23.896	0.702
FM_{T}	3.388	0.987	2.218	0.979	28.079	0.631
FM_{C}	125.007	0.705	786.038	0.076	401.452	0.127
FM_{GLV}	13.987	0.955	24.541	0.804	155.935	0.388
FM_{M2}	32.984	0.898	46.472	0.678	328.083	0.287
FM_{SF}	11.432	0.957	1.144	0.989	4.319	0.924

5 Conclusion

In this paper we introduced a new method for SFF based on steerable filters. Steerable filters are applied at different orientations to the sequences of images of varying

texture properties. Steerable filters remove inherent limitations of traditional gradient detection technique which perform inadequately for oriented intensity varying scenarios. For each pixel location 2D neighborhood sum on higher amplitude response of oriented steerable filters is exploited to locate focused image points for shape reconstruction and depth map computation. Better localization, directional specificity with high amplitude response only at focused points ensures reliable results with efficient processing. In quantitative and qualitative analyses, our proposed method outperforms well documented depth estimation techniques.

Acknowledgment. The work is supported in part by the Canada Research Chair program, the NSERC Discovery Grant, and the AUTO21 NCE.

References

1. Nayar, S.K., Nakagawa, Y.: Shape from focus. IEEE Trans. Pattern Anal. Mach. Intell. 16(8), 824–831 (1994)
2. Nayar, S.K., Nakagawa, Y.: Shape from focus: an effective approach for rough surfaces. CRA 2, 218–225 (1990)
3. Nayar, S.K.: Shape from focus. Carnegie Mellon University, CMU-RI-TR, 89-27 (1989)
4. Martinez-Baena, J., Garcia, J.A.: A multi channel auto focusing scheme for gray level shape scale detection. Pattern Recognition 30(10), 1769–1786 (1996)
5. Xiong, Y., Schafer, S.A.: Depth from focusing and defocusing. IEEE Computer Vision and Pattern Recognition, 68–73 (1993)
6. Tenenbaum, J.M.: Accommodation in computer vision. PhD thesis Stanford University (1970)
7. Subbarao, M., Choi, T.: Accurate recovery of three dimensional shape from image focus. IEEE Trans. Pattern Analysis and Machine Intelligence 17(3), 266–274 (1995)
8. Helmi, F.S., Scherer, S.: Adaptive shape from focus with an error estimation in light microscopy. In: Proc. of the 2nd International Symposium on Image and Signal Processing and Analysis (2001)
9. Choi, T.S., Asif, M., Yun, J.: Three-dimensional shape recovery from focused image surface. In: ICASSP, vol. 6, pp. 3269–3272 (1999)
10. Yun, J., Choi, T.S.: Accurate 3-D shape recovery using curved window focus measure. In: International Conference on Image Processing, vol. 3, pp. 910–914 (1999)
11. Asif, M., Choi, T.S.: Shape from focus using multilayer feed forward neural network. IEEE Trans. on Image Processing 10(11), 1670–1675 (2001)
12. Ahmad, M.B., Choi, T.S.: A heuristic approach for finding best focused image. IEEE Trans. on Circuits Systems and Video Technology 14(4), 566–574 (2005)
13. Malik, A.S., Choi, T.S.: A novel algorithm for estimation of depth map using image focus for 3D shape recovery in the presence of noise. Pattern Recognition 41, 2200–2225 (2008)
14. Malik, A.S., Choi, T.S.: Application of passive techniques for three dimensional cameras. IEEE Trans. On Consumer Electronics 53(2), 258–264 (2007)
15. Ahmad, M.B., Choi, T.S.: Application of three dimensional shape from image focus in LCD/TFT display manufacturing. IEEE Trans. on Consumer Electronics 53(1), 1–4 (2007)
16. Canny, J.F.: A computational approach to edge detection. IEEE Pat. Anal. Mach. Intell. 8(6), 679–698 (1986)

17. Daugman, J.G.: Complete discrete 2-d Gabor transforms by neural networks for image analysis and compression. IEEE Trans. Acoust., Speech, Signal Proc. 36(7), 1169–1179 (1988)
18. Heeger, D.J.: Optical flow using spatiotemporal filters. Intl. J. Comp. Vis. 1(4), 279–302 (1988)
19. Malik, J., Perona, P.: Preattentive texture discrimination with early vision mechanisms. J. Opt. Soc. Am. A 7, 923–931 (1990)
20. Danielsson, P., Seger, O.: Rotation invariance in gradient and higher order derivative detectors. Comp. Vis., Graphics, Image Proc. 49, 198–221 (1990)
21. Koenderink, J.J., Van Doorn, A.J.: Representation of local geometry in the visual system. Biol. Cybern. 55, 367–375 (1987)
22. Pentland, A.P.: Local shading analysis. IEEE Pat. Anal. Mach. Intell. 6(2), 170–187 (1984)
23. Freeman, W.T., Adelson, E.H.: The design and use of steerable filters. IEEE Trans on Pattern Analysis and Machine Intelligence 13(9), 891–906 (1991)