Proceedings of the 2009 IEEE International Conference on Systems, Man, and Cybernetics
San Antonio, TX, USA - October 2009

Depth Map Estimation Using Exponentially Decaying Focus Measure Based on Susan Operator

Pankajkumar Mendapara, Rashid Minhas, Q.M. Jonathan Wu
Department of Electrical Engineering
University of Windsor
Windsor, Canada
{mendapa, minhasr, jwu}@ uwindsor.ca

***Abstract*— This paper presents a novel technique for depth map estimation using a sequence of images acquired at varying focus. In depth map estimation noise, illumination variations and types of extracted features significantly affect the performance of a focus measure. This paper proposes the use of SUSAN operator, to extract features, because of its structure preserving noise filtering which plays a pivotal role in depth estimation of a scene. We introduce a new focus measure based on exponentially decaying function to use neighborhood information of an extracted feature point that assigns more weight to the closer pixel points. Experiments validate superior performance of our proposed algorithm in comparison to other well-documented methods.**

***Keywords*—focus measure, 3D shape recovery, shape from focus, exponentially decaying function, multi-focus imaging.**

I. Introduction

The technique utilized to retrieve spatial information from a sequence of images with varying focus plane is termed as shape from focus (SFF). In SFF, a sequence of images (SI) is acquired at varying relative distance between a camera lens and a scene object. Such a sequence captures well focused partial information of a scene in different images. To reconstruct a well focused image, SI acquired with varying distances is processed to extract focused points from individual image frames. Traditional SFF techniques assume convex shaped objects for accurate depth map estimation. SFF removes the inherent limitation of traditional image acquisition for its inability to capture details of a scene with a considerably large depth.

The objective of depth map estimation is to determine the depth of every object point with respect to the camera. For scenes with considerably large depth, object points present on a focus plane appear sharp in an acquired image whereas blur of imaged points increases as they move away from the focus plane.

Basic image formation geometry when camera parameters are known is shown in Fig. 1. Distance of an object from camera lens i.e. u is required for exact 3D reconstruction of a scene. Depth of a scene, distance of an object from lens, illumination conditions, camera movement, aberration effects in lens and movement in a scene can severely affect the depth map estimation. Computing distance of an object from a camera lens is simple if blur circle radius R is equal to zero. If image detector (ID) is placed at an exact distance v; sharp focused image P' of an object point P is formed. Relationship between object distance u, focal distance of lens f and ID distance v is given by Gaussian lens law.

$$1/f = 1/u + 1/v \tag{1}$$

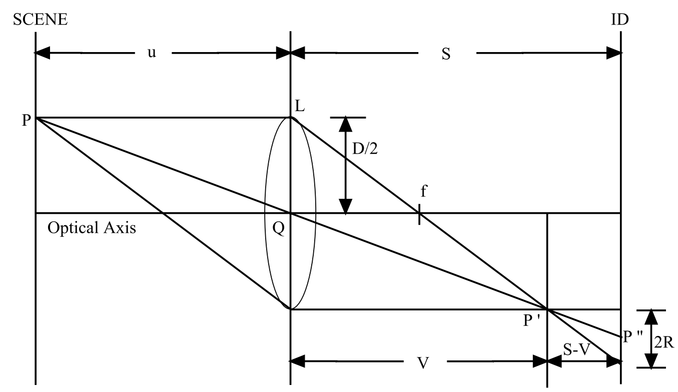


Figure 1. Image formation geometry of a 3D object

In literature [1-6,14,15] commonly used operators in SFF are sum of modified Laplacian (FM_{SML}), Tenengrade focus measure (FM_T), gray level variance focus measure (FM_{GLV}), curvature focus measure (FM_C), M2 focus measure (FM_{M2}), point focus measure (FM_P) and steerable filters based focus measure (FM_{SF}). Approximation and learning based focus measures have also been proposed [7-9] that utilize neural network, neuro fuzzy systems and dynamic programming based approaches for accurate depth map estimation. Approximation based techniques use any of the conventional aforementioned focus measures for pre-processing whereas comprehensive rule base and appropriate selection of training data restrict their application to specific domains.

In this paper a new scheme is proposed to estimate depth map by searching 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 degraded performance is observed in presence of noise, poor texture and singularities along curves.

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

II. SUSAN OPERATOR

Smith and Brady proposed SUSAN (Smallest Univalue Segment Assimilating Nucleus) algorithm in 1997 [13]. This algorithm has three parts: edge detection, corner detection and structure preserving noise filtering. In this algorithm non-linear filtering is used to identify image sub-regions which are closely related to individual pixels. In SUSAN algorithm a circular mask is used for convolution and the brightness of each pixel within the circle is compared with the brightness of center pixel of the mask. The area of the mask that has the same brightness as the nucleus is known as USAN (Univalue Segment Assimilating Nucleus). The SUSAN filters works by taking an average of the pixels in USAN area excluding the centre pixel. The USAN area represents important information regarding structure of an image. From size, centroid and second moments of the USAN, two dimensional features and edges are detected. The SUSAN operator does not need image derivates and exhibits low computational complexity.

Let $I(r)$ denotes the gray value at pixel r, n the area of the USAN (the total no. of pixels in USAN), r_0 the nucleus, α is circular mask and σ is brightness difference threshold, then

$$n(r_0) = \sum_{r \in \alpha} c(r, r_0) \tag{2}$$

$$where, \qquad c(r, r_0) = e^{-[(I(r) - I(r_0))/\sigma]^6} \tag{3}$$

Finally, the response of Susan edge detector at pixel r_0 is given by,

$$E(r_0) = \begin{cases} GT - n(r_0) & for \quad n(r_0) < GT \\ 0 & else, \end{cases} \tag{4}$$

where, GT is called geometrical threshold. For edge detection suitable value of GT is $(3/4)n_{max}$ where n_{max} is the maximum value that n can carry.

III. PROPOSED ALGORITHM

The main goal of this paper is to estimate a depth map by ensuring the transformation of most relevant information found in source images into a new composite image. In SFF the robustness of any focus measure operator depends on its ability to calculate sharpness value of each pixel. Our proposed scheme uses exponentially decaying function with SUSAN operator to analyze sharpness of each pixel in an image sequence. Entire process of our proposed scheme is depicted in Fig. 2.

In our proposed scheme, features are extracted by applying SUSAN operator on pre-registered multi-focus image sequence. A focus measure, characterized by exponentially decaying function, is employed to compute sharpness of each pixel in an image. Such decaying function uses neighborhood information of extracted feature points assuming that intensity far from a feature point is equal to 1 and it approaches this limiting value in an exponential way.

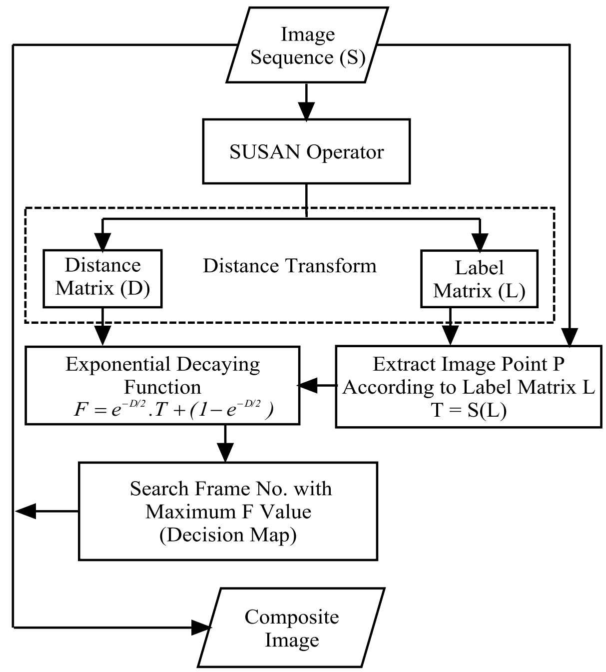


Figure 2. Different Steps of our proposed algorithm

Given an arbitrary point k and the set E of feature points, the focus measure F is estimated as:

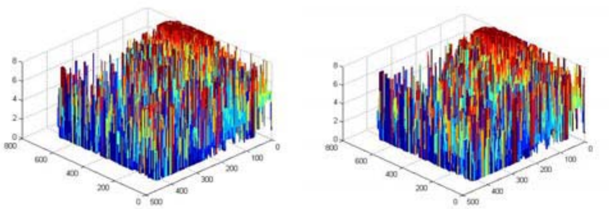
$$F(x,y) = e^{-(D(x,y)/2)}. \ T(x,y) + \left(1 - e^{-(D(x,y)/2)}\right) \tag{5}$$

where, D is the distance between point k and the nearest feature point T. Distance transformation of an output of SUSAN operator yields a distance matrix (D) and a label matrix (L). T is the actual intensity value extracted from original images using label matrix L. These matrices D and T are used in computation of focus measure through exponential function. A decision map is obtained by comparing corresponding F values of each frame. The frame with higher value of F is mapped on to corresponding pixel of decision map. Finally using decision map, pixels are extracted from original image sequence yielding a composite image.

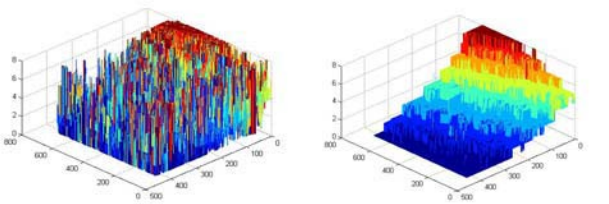
IV. EXPERIMENTAL AND RESULTS

The proposed focus measure operator is tested on pre-treatment set of 8 images each of size 640 x 460 and chess set of 29 images each of equal size i.e. 800 x 600. The sample images (pre-treatment dataset) are shown in Fig. 3. The pre-treatment dataset images are taken by moving focus plane stepwise diagonally. Fig. 4 shows estimated depth maps of pre-treatment dataset and Fig. 5 shows depth maps obtained after adding Gaussian noise of variance 0.0005 to original image dataset.

Figure 3. Sample (pre-treat) image set (obtained from Special K Software)

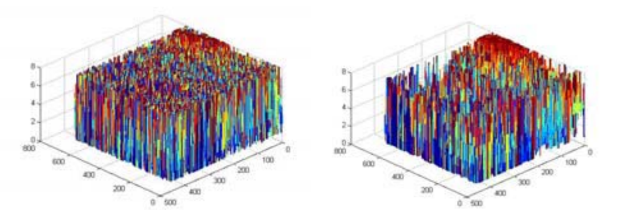


(a) FM_{SML} (b) FM_T

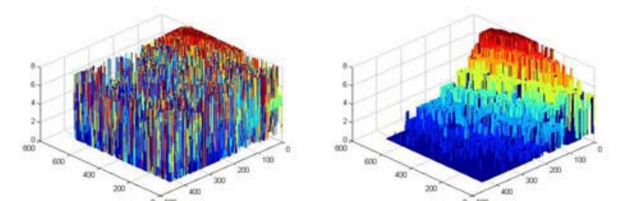


(c) FM_{GLV} (d) Proposed method

Figure 4. Depthmap for Pre-treatment dataset by various methods



(a) FM_{SML} (b) FM_T



(c) FM_{GLV} (d) Proposed method

Figure 5. Depth map for Pre-treatment dataset by various methods with added gaussian noise of variance 0.0005

Figs. 4-5 show the estimated depth maps for pre-treatment images using (a) FMSML method (b) FMT method (c) FMGLV method and (d) proposed method. As shown in Fig. 4, the depth map computed using proposed method is smoother compared to other techniques. Also, on addition of Gaussian noise of variance 0.0005 the performance of FM_{SML}, FM_T and FM_{GLV} deteriorates and the depth map is unrecognizable whereas proposed method can still track the actual depth of the scene under reference (see Fig. 5).

Fig. 6 shows two different frames of chess dataset; Fig. 7 shows depth maps estimated by (a) FM_{SML} method (b) FM_T method (c) FM_{GLV} method and (d) proposed method. Our proposed method exploits neighborhood information of identified feature points and the overall performance of our proposed scheme for depth map estimation is better than other well documented methods.

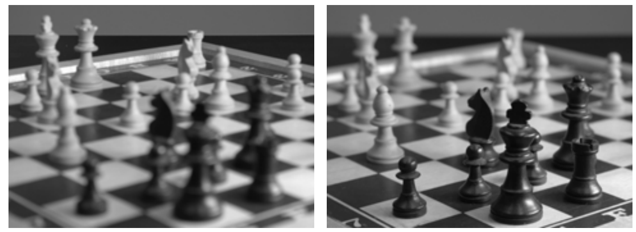
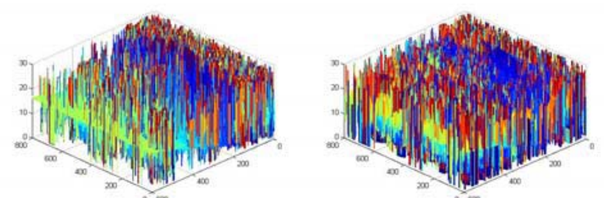
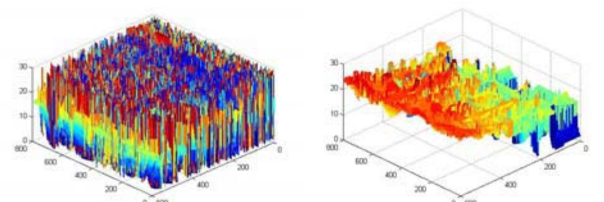


Figure 6. Two sample frames of Chess dataset



(a) FM_{SML} (b) FM_T



(c) FM_{GLV} (d) Proposed method

Figure 7. Depth map for chess dataset by various methods

For the quantitative evaluation of our proposed method, we use three different criterions: Mutual information (MI) [10], Petrovic and Xydeas metric (Q_p) [11] and Piella's quality matrix (SSIM) [12]. These methods estimate how and what information is transferred from the input images to the composite image.

Mutual information is a natural measure of dependence between random variables. In case of image fusion it is defined by the average distance between input images and a fused image. We can calculate the amount of information that composite image retains from input image sequence as:

$$MI_{ZH}(z,h_i) = \sum_{z,h_i} P_{ZH}(z,h_i)\log\frac{P_{ZH}(z,h_i)}{P_Z(z)P_H(h_i)} \tag{6}$$

$$MI_Z = \sum_{h_i=1}^{N} MI_{ZH}(z,h_i) \tag{7}$$

where h_i, $i \in \{1, 2, \ldots, N\}$ represents input image sequence and z is a composite image.

SSIM image quality index is based on structural similarity, and local SSIM measures three elements of image patches: the similarity of brightness, contrast and structures.

$$SSIM(h_i, z) = \frac{\left(2\mu_{h_i}\mu_z + C_1\right)\ \left(2\sigma_{h_i z} + C_2\right)}{\left(\mu_{h_i}^2 + \mu_z^2 + C_1\right)\ \left(\sigma_{h_i}^2 + \sigma_z^2 + C_2\right)} \tag{8}$$

$$MSSIM = \sum_{h_i=1}^{N} mean2(SSIM(h_i, z)) \tag{9}$$

where h_i, $i \in \{1, 2, \ldots, N\}$ represents input image sequence and z is a composite image, μ_{h_i} and μ_z are local sample means of h_i and z respectively and $\sigma_{h_i z}$ is the sample cross correlation of h_i and z after removing their mean. C_1 and C_2 are small positive constants used to stabilize each term so that near zero sample means, variance or correlation does not lead to numerical instability.

The objective performance matrix (Q_p) measures the amount of "edge information transferred" from a source image to the composite image and gives an estimation of the performance of a fusion algorithm.

$$Q_p^{AB|Z} = \frac{\sum_{x=1}^{N}\sum_{y=1}^{M} Q^{AZ}(x,y)W_A(x,y) + Q^{BZ}(x,y)W_B(x,y)}{\sum_{x=1}^{N}\sum_{y=1}^{M} W_A(x,y)W_B(x,y)} \tag{10}$$

where $Q^{AZ}(x, y)$ and $Q^{BZ}(x, y)$ are edge strength and orientation preservation values, $W_A(x, y)$ and $W_B(x, y)$ are the weights of edge information of image A and B respectively.

TABLE I. MUTUAL INFORMATION (MI) COMPARISON

	Proposed	**FM_{SML}**	**FM_T**	**FM_{GLV}**
Chess	17.4559	15.3308	13.3002	13.3120
Pre-treat	2.9439	2.6887	2.7035	2.6076

TABLE II. STRUCTURAL SIMILARITY (MSSIM) COMPARISON

	Proposed	**FM_{SML}**	**FM_T**	**FM_{GLV}**
Chess	0.7717	0.7082	0.6757	0.7111
Pre-treat	0.5737	0.5455	0.5448	0.5414

TABLE III. EDGE INFORMATION TRANSFORMATION (Q_P) COMPARISON

	Proposed	**FM_{SML}**	**FM_T**	**FM_{GLV}**
Chess	0.3184	0.2449	0.2421	0.3172
Pre-treat	0.4034	0.3875	0.3944	0.3473

From tables 1, 2, and 3, it is evident that the performance of proposed method in terms of MI, MSSIM and Qp is superior than other schemes.

V. CONCLUSION

This paper presents a new focus measure for depth map estimation based on exponentially decaying function that exploits neighborhood information of extracted feature points identified through SUSAN operator. Structure preserving noise filtering and detection of various kinds of features (lines/corners) using SUSAN allows improved detection of well focused image points. Experimental results show the superior performance of our proposed method compared with other traditional schemes. Medical imaging, collision avoidance, shape reconstruction and image fusion are some of the areas that can potentially benefit from our proposed scheme.

ACKNOWLEDGEMENT

This research has been supported in part by Canada Research Chair Program, AUTO 21 NCE and the NSERC discovery grant.

REFERENCES

[1] A.S. Malik and T.S. Choi, "A novel algorithm for estimation of depth map using image focus for 3D shape recovery in the presence of noise," Pattern Recognition, 2008, pp. 2200-2225.

[2] S.K. Nayar and Y. Nakagawa, "Shape from focus," IEEE Trans. Pattern Anal. Mach. Intell. , 1994, vol. 16 no. 8, pp. 824-831.

[3] S.K. Nayar, Y. Nakagawa, "Shape from focus: an effective approach for rough surfaces," CRA, 1990, vol. 2, pp. 218-225.

[4] M. Subbarao, T. Choi, "Accurate recovery of three dimensional shape from image focus," IEEE Trans. Pattern Analysis and Machine Intelligence, 1995, vol. 17, no. 3, pp. 266-274.

[5] T.S. Choi, M. Asif, J. Yun, "Three-dimensional shape recovery from focused image surface," ICASSP, 1999, vol. 6, pp. 3269-3272.

[6] J. Yun, T.S. Choi, "Accurate 3D shape recovery using curved window focus measure," International Conference on Image Processing, 1999, pp. 910-914.

[7] M. Asif, T.S. Choi, "Shape from focus using multilayer feed forward neural network," IEEE Trans. on Image Processing, 2001, vol. 10, no. 11, pp. 1670-1675.

[8] A.S. Malik, T.S. Choi, "Application of passive techniques for three dimensional cameras," IEEE Trans. on Consumer Electronics, 2007, vol. 53, no. 2, pp. 258-264.

[9] M.B. Ahmad, T.S. Choi, "A heuristic approach for finding best focused image," IEEE Trans. on Circuits Systems and Video Technology, 2005, vol.14, no.4, pp. 566-574.

[10] G. Qu, D. Zhang and P. Yan, "Information measure for performance of image fusion," Electronics Letters, 2002, vol. 38, no. 7, pp. 313-315.

[11] C.S. Xydeas and. V. Petrovic , "Objective image fusion performance measure," Electronics Letters, 2000, vol. 36, no. 4, pp. 308-309.

[12] G. Piella, H. Heijmans, "A new quality metric for image fusion," Proc. IEEE Int. Conf. on Image Processing, 2003, vol.2, pp. 173-176.

[13] S.M. Smith, J.M. Brady, "SUSAN-A New approach to low level image processing," International journal of computer vision, 1997, vol.23 no. 1, pp. 45-78.

[14] R. Minhas, A.A. Mohammed, Q.M. Jonathan Wu, M.A.Sid Ahmed, "3D shape from focus and depth map computation using steerable filters," ICIAR, 2009, Halifax, Canada, in press.

[15] R. Minhas, A.A. Mohammed, Q.M. Jonathan Wu, T.S. Choi, "3D neighborhood support for accurate shape from focus and depth map estimation using steerable filters," ICIP, 2009, unpublished.