

All in Focus Image Generation based on New Focusing Measure Operators

Hossam Eldeen M. Shamardan ¹

¹ Department of Information Technology
Faculty of Computers and information
University of Helwan
Egypt

Abstract—To generate an all in focus image, the Shape-From-Focus (SFF) is used. The SFF key is finding the optimal focus depth at each pixel or area in an image within sequence of images. In this paper two new focusing measure operators are suggested to be used for SFF. The suggested operators are based on modification for the state of art tool for time-frequency analysis, the Stockwell Transform (ST). The first operator depends on Discrete Orthogonal Stockwell Transform (DOST) which represents a pared version of ST, while the other depends on Pixelwise DOST (P-DOST) which provides a local spatial frequency description. Both of the operators provides the computational complexity and memory demand efficiency compared to the operator depending on ST. A comparison between the suggested operators to operators based on ST are performed and showed that the suggested operators' performances are as analogous to that of ST.

Keywords—Focus Measure; All In Focus; Stockwell Transform; DOST

I. INTRODUCTION

SFF is an essential process to overcome a specific limitation of imaging systems which is the different depth of field (DOF) for each part in an image [1, 2, 3]. The SFF solves this problem by providing fully sharp focused image. The SFF depends on measuring the optimal focus by using the focus measure (FM) operator, the main key, for shape estimating. The usual SFF methods compute optimal focus and its depth by applying FM operator on every area in a frame within a sequence of images and seek the optimal focused part along the sequence. The fully focused image can be constructed from merging the focuses optimal parts [4, 5].

The sharp focused area is characterized by its energy for high frequency components, and The FM operator is used to measure the amount of energy in every part. Many FM operators have been proposed in the literature for both autofocus (AF) and SFF applications [5]. In [6], the performance of multiple FM operators were examined. FM operators can be classified into two broad types: space domain and, frequency domain [7] [8], and others such as compression operators [5] [9].

The ST has been used as a base for FM operator in [10] and compared to other FM operators. The suggested idea is to measure the high frequency components energy located in a region of interest in the ST domain. The ST gives good results but it suffers from high computational cost. To overcome the

problem of high cost for ST, a pared version of ST, DOST was provided in [11]. The DOST provides efficiency in both computational cost and memory usage. Another version of DOST, pixel-wise local spatial frequency description (P-DOST) was given in [12]. The P-DOST provides a tool for studying a specific frequency at specific pixel or area. Both of DOST and P-DOST have low computational cost and robustness to Gaussian noise.

In this paper two new FM operators are presented based on DOST and P-DOST for their low computational cost and memory usage.

This paper is organized as follows. In section 2, the background for ST, DOST and P-DOST are presented. The proposed algorithm and FM operators are described in section 3. Finally, experimental results and conclusion are presented in sections 4 and 5 respectively.

II. BACKGROUND

ST has been shown as a generalization of the short-time Fourier transform (STFT), and the wavelet transforms [13]. From [10] the 2D-ST, H_{st} for an image $h(x, y)$ of size $N * M$ for pixel (x, y) is given by

$$H_{st}[x, y, k_x, k_y] = \sum_{n=0}^{N-1} \sum_{m=0}^{M-1} H[n + k_x, m + k_y]^* \exp(-2\pi^2 (\frac{n^2}{k_x^2} + \frac{m^2}{k_y^2})) \exp(2\pi i(nx + my)) \quad (1)$$

Where H is the 2D Fourier transform of h , x , and y are x -coordinates and y -coordinates in space respectively and k_x , and k_y are indices in the frequency along x -axis and y -axis.

From [14], it has been demonstrated that the ST redundancy comes from the equal sampling rate for both low and high frequency bands despite the fact of Nyquist criterion which states that the sampling rate depends on the frequency of the sampled data. A detailed study for the 2D-ST computational complexity for an image of size $N * N$ is given in [12]. It has been shown that ST has a computational cost of $O(N^4 + N^4 \log N)$ and a storage requirement of $O(N^4)$.

The DOST suggested in [11] gives lower sampling rates for lower frequencies, and higher sample rates for higher

frequencies to solve the ST redundancy. It does so by building a set of N orthogonal unit-length basis vectors, each of which targets a particular region in the time-frequency domain. The regions defined by DOST are described by a set of parameters: ν specifies the center of each frequency band (voice) p , β is the width of that band, and τ specifies the location in time.

DOST basis vectors for a particular band p and the parameters describing these basis vectors are defined in the following cases according to p where $p = 0, 1, \dots, \log_2(N) - 1$

$$\text{for } p = 0, \nu = 0, \beta = 1,$$

$$D(kT)_{[\nu, \beta, \tau]} = 1$$

$$\text{for } p = 1, \nu = 1, \beta = 1,$$

$$D(kT)_{[\nu, \beta, \tau]} = e^{(-2ik\pi/N)} \quad (2)$$

$$\text{for } p > 1, \nu = 2^{(p-1)} + 2^{(p-2)},$$

$$\beta = 2^{(p-1)}, \tau = 0, \dots, \beta - 1$$

$$D(kT)_{[\nu, \beta, \tau]} = ie^{(-i\pi k)} * \frac{e^{-2i\alpha(\nu - \beta/2 - 1/2)} - e^{-2i\alpha(\nu + \beta/2 - 1/2)}}{2\sqrt{\beta} \sin(\alpha)}$$

where $\alpha = \pi(k/N - \tau/\beta)$ is the center of the temporal window, and $k = 0, 1, \dots, N-1$ is the index of time interval. The time-frequency distribution for a family of vectors for a signal of length 16 is shown in Fig. 1 [15].

Calculating DOST is determined by taking the inner product between the basis vectors mentioned above and the input signal. By taking linear combinations of the Fourier complex sinusoids in band-limited subspaces and applying appropriate phase and frequency shifts, the 2D-DOST of $M * N$ image $h[x, y]$ is defined in [12] as follows:

$$S[x', y', \nu_x, \nu_y] = \frac{1}{\sqrt{\beta}} \sum_{m=-2^{p_x-2}}^{2^{p_x-2}-1} \sum_{n=-2^{p_y-2}}^{2^{p_y-2}-1} H(m + \nu_x, n + \nu_y) * \exp(2i\pi(\frac{mx'}{2^{p_x-1}} + \frac{ny'}{2^{p_y-1}})) \quad (3)$$

Where $\nu_x = 2^{p_x-1} + 2^{p_x-2}$ and $\nu_y = 2^{p_y-1} + 2^{p_y-2}$ are representing the horizontal and vertical voice frequencies, and $\beta = 2^{p_x p_y - 2}$ is representing the number of points in the partition, and $H(m, n)$ is the 2D Fourier transform for image $h[x, y]$.

The 2D-DOST due to using orthonormal set of basic functions as described above, has computational complexity of $O(N^2 + N^2 \log N)$ and storage requirements of $O(N^2)$ which was proved in [16]. It is obvious DOST has less computational complexity required than for the 2D-ST.

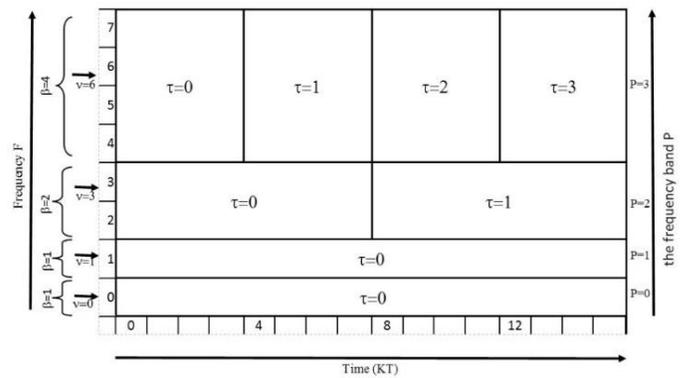


Fig. 1. Time-Frequency distribution for DOST components in positive direction

The P-DOST suggested in [12] aims to find the voice frequency distribution for a pixel or region within the image. By choosing a set of (x, y) coordinates representing a single pixel or area, all the values of $s[x, y, \nu_x, \nu_y]$ for all horizontal and vertical voices (ν_x, ν_y) can be determined for position (x, y) . Due to the variable size and limits of every DOST components at each band, the P-DOST is constructed by obtaining all components of DOST at each band (p_x, p_y) by:

$$S[x/N * 2^{p_x-1}, y/N * 2^{p_y-1}, \nu_x, \nu_y] \quad (4)$$

From (4), the P-DOST corresponding response is occupying part of the DOST response space. Hence. It only requires size of, considering the negative side, $2\log_2(N) * 2\log_2(N) (2\beta * 2\beta)$. The P-DOST response is referred to as the local domain (or spectrum). Since P-DOST selects only some components from the whole components set for DOST, a reduction will consequently go further for the computational complexity and memory demands. Consequently, the computational complexity for calculating P-DOST for single pixel is of order $O((2 * \log_2(N))^2)$. Considering the DOST calculations, it is obvious that P-DOST is less demanding for computational complexity than for DOST.

III. PROPOSED ALGORITHM

The algorithm suggests a stack of L frames $I_1 \dots I_L$ with the same size, and same scene pictures an object at different focusing depths. The frames are divided into windows $W(x, y, z)$ each of size $M * N$ located at position (x, y) in every frame z . The Suggested algorithm is described as follows:

- 1) Divide every frame $I_1 \dots I_L$ into windows. Each window $W(x, y, z)$ is located at position (x, y) in frame I_z
- 2) Apply the FM_D or FM_p for window $W_{x,y,z}$ using formulas in (5) or (6) along all the L frames.
- 3) Find $Z_{optimal}$ by using (7).
- 4) Repeat steps from 1 to 3 and merge all windows to get the generated fully focused image.

The formulas for calculating FM_D is given as follows:

$$FM_D(Z) = \sum_{v_y=v(p=-M-1)}^{v_y=v(p=M)} \sum_{v_x=v(p=-N-1)}^{v_x=v(p=N)} \xi(x, y, v_x, v_y) \quad (5)$$

where

$$\xi(x, y, v_x, v_y) = \sum_{y=0}^{M-1} \sum_{x=0}^{N-1} abs(S[x, y, v_x, v_y]) \quad (v_x, v_y) \neq (0,0)$$

And for FM_p

$$FM_P(x_0, y_0) = \sum_{v_y=v(p=-M-1)}^{v_y=v(p=M)} \sum_{v_x=v(p=-N-1)}^{v_x=v(p=N)} \xi(x_0, y_0, v_x, v_y), \quad (6)$$

where

$$\xi(x_0, y_0, v_x, v_y) = abs(S[x_0, y_0, v_x, v_y]) \quad (v_x, v_y) \neq (0,0)$$

For determining $Z_{optimal}$

$$Z_{Optimal} = \arg \max_z (FM(Z)), \quad z = 1 \dots L \quad (7)$$

The suggested FM_D and FM_p in the algorithm are based on DOST and P-DOST respectively and both of them measure the energy in high frequencies components (frequency > 0 (DC)). A formula used for the suggested operators are given in (5) and (6). The operators are used to find the optimal focused window by measuring the highest response resulted by the operator. By merging the optimal focused window a fully focused image of the scene can be reconstructed. The idea for measuring the energy is the of the energy for the components within an area reflects the sharpness and hence the focusing. For FM_p , the pixel (x_0, y_0) is selected at the center of the targeted window and used to represent the window.

IV. EXPERIMENTAL RESULTS

To evaluate the performance of the proposed algorithm and its robustness, three experiments for three sequences of images were conducted each of 256 gray levels. The first experiment contains sequence of 60 images of cone. The size for this sequence is 360*360. A simulation software was used to generate the images focused at different parts of the image Fig. 2a through 2d, the proposed algorithm has been used to construct all in focus image through FM_D and FM_p operators on this sequence. The resulting images are shown in Fig. 2e, 2f. The second experiment contains a sequence of 30 natural images, each of size 299*215. The sequence and the generated all in focus image are shown in Fig. 3. The third experiment contains also 30 natural images of size 215 * 215. Results are shown in Fig. 4.

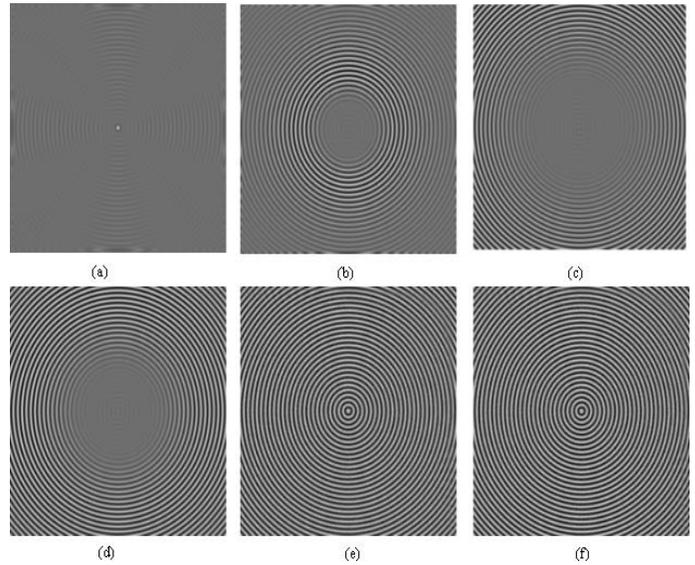


Fig. 2. Pictures from (a) to (d) are defocus cone images at different focusing level, e) SFF by using FM_D , f) SFF by using FM_p

The first sequence of images was adopted and used for next evaluating tests. To evaluate the robustness of the algorithm against noise, three different measures were adopted to measure the all in focus image quantitatively. The first measure is the rmse (root mean square error) which is defined as follows

$$rmse = \sqrt{\frac{1}{MN} \sum_{x=1}^M \sum_{y=1}^N (I(x, y) - I'(x, y))^2} \quad (8)$$

Where $I(x, y)$ the original is image, and $I'(x, y)$ is the all in focus image from the sequence of images

To test the performance of the algorithm, the algorithm was applied to the sequence with added Gaussian noise of variance ranging from 0 to 10 and the window size is ranging from 4 to 64 Fig. 5 shows the results of this test.

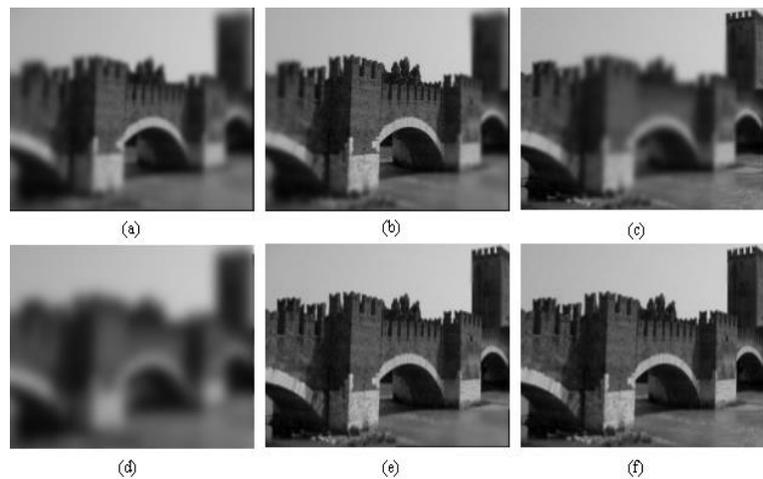


Fig. 3. Pictures from (a) to (d) are defocus cone images at different focusing level, (e) SFF by using FM_D , (f) SFF by using FM_p

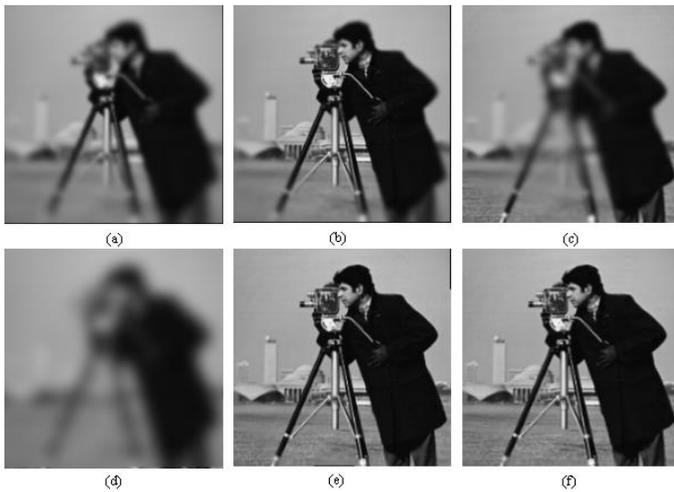


Fig. 4. Pictures from (a) to (d) are defocus cone images at different focusing level, e) SFF by using FM_D , f) SFF by using FM_P

To compare between FM_D and FM_P performance, a test was applied showing the rmse performance of the two operators against the window size and for different noise variances. The results are shown in Fig. 6. It is obvious that the FM_D is advantageous than the FM_P since it considers all the pixels rather than a single central one.

The next measures are the UIQI (universal image quality index) and SSIM (structural similarity index measure) [10] and are defined as following:

$$UIQI = \frac{4\sigma_{\mu}\bar{\mu}'}{(\sigma_i^2 + \sigma_i'^2)(\bar{i}^2 + \bar{i}'^2)} \quad (9)$$

$$SSIM = \frac{(2\bar{\mu}' + c_1)(2\sigma_i\sigma_i' + c_2)}{((\bar{i})^2 + (\bar{i}')^2 + c_1)(\sigma_i^2 + \sigma_i'^2 + c_2)} \quad (10)$$

Where \bar{I} and \bar{I}' represent mean of the original and all in focus image constructed from the sequence of images respectively. The results are shown in Fig. 7,8. It is clear that both UIQI and SSIM gives similar performance for both FM_D and FM_P .

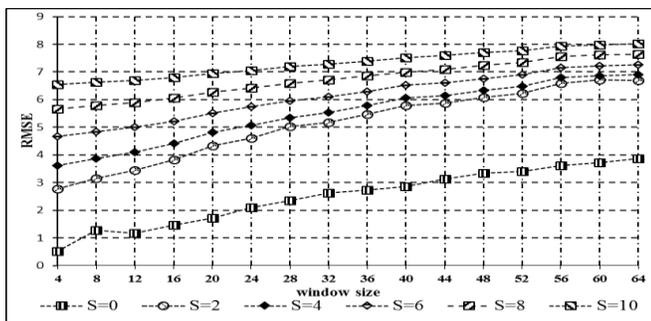


Fig. 5. rmse against window size for different variance levels of gaussian noise

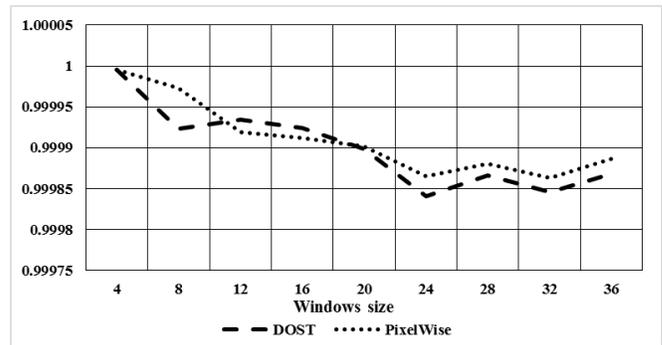


Fig. 6. Comparison between FM_D and FM_P operators performance (UIQI) versus window size

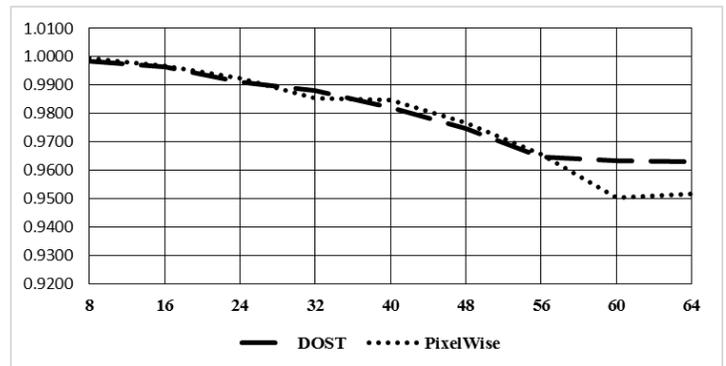


Fig. 7. Comparison between FM_D and FM_P operators performance (SSIM) versus window size

Another experiment was conducted to compare between FM_{ST} and FM_D performances for generating the cone image give in Fig. 2, assuming window is of size 4. The results showed that FM_{ST} provides performance little bit better than FM_D . The achievable rmse for the ST is 0.52 while the DOST achieved 0.72. Of course the FM_{ST} provides better performance (38%) however, the resultant performance is very minor (less than 1) out of 256 gray level. Compared to the computational complexity, the FM_{ST} is of order $O(N^4 + N^4 \log N)$ while the FM_D is of order $O(N^2 + N^2 \log N)$.

V. CONCLUSION AND FUTURE WORK

In this paper, two new FM operators are suggested to reconstruct all-in-focus image. The two operators are built on extensions of Stockwell Transform, a space-frequency transformation tool. Stockwell Transform has been proved to suffer from excessive abundant computations and high memory requirements and to reduce the computational complexity that is the reason the FM_D and FM_P are adopted as bases for the suggested FM operators. Results have shown to

be as good as for FM_{ST} . The two suggested methods are of almost similar performance and with less computational complexity and memory demands. The future research can extend the usage of those tools to get automatic focusing and to enhance the robustness

REFERENCES

- [1] Z. Li, A. Fischer and G. Li, "Volumetric retinal fluorescence microscopic imaging with extended depth of field," in SPIE 9713, Three-Dimensional and Multidimensional Microscopy: Image Acquisition and Processing XXIII, San Francisco, 2016.
- [2] E. Anderes, B. Yu, V. Jovanovic, C. Moroney, M. Garay, A. Braverman and E. Clothiaux, "Maximum likelihood estimation of cloud height from multi-angle satellite imagery," The Annals of Applied Statistics, pp. 902-921, 5 October 2009.
- [3] Y. Song, Y. Xie, V. Malyarchuk, J. Xiao, I. Jung, K. Choi, Z. Liu, H. Park, C. Lu, R. Kim, R. Li, K. Crozier, Y. Huang and J. Rogers, "Digital cameras with designs inspired by the arthropod eye," Nature, vol. 497, no. 7447, p. 95-99, 2 May 2013.
- [4] A. Anish and T. J. Jebaseeli, "A survey on multi-focus image fusion methods," International Journal of Advanced Research in Computer Engineering & Technolog, vol. 1, no. 8, p. 319-324, October 2012.
- [5] S. Pertuz, D. Puig and M. A. Garcia, "Analysis of focus measure operators for shape-from-focus," Pattern Recognition, vol. 46, no. 5, p. 1415-1432, May 2013.
- [6] A. S. Malik and T. S. Choi, "Consideration of illumination effects and optimization of window size for accurate calculation of depth map for 3D shape recovery," Pattern Recognition, vol. 40, no. 1, pp. 154-170, 2007.
- [7] T. M. Mahmood, S. O. Shim and T. S. Choi, "Shape from focus using principal component analysis in discrete wavelet transform," Optical Engineering, vol. 48, no. 5, p. 057203, 28 May 2009.
- [8] J. Baina and J. Dublet, "Automatic focus and iris control for video cameras," in Image Processing and its Applications, 1995., Fifth International Conference on, Edinburgh, 1995.
- [9] H. Mir, P. Xu and P. V. Beek, "An extensive empirical evaluation of focus measures for digital photography," in SPIE 9023, Digital Photography X, 90230I, 2014.
- [10] T. M. Mahmood and T. S. Choi, "Focus measure based on the energy of high-frequency components in the S transform.," in Depth Map and 3D Imaging Applications: Algorithms and Technologies., vol. 35, IGI Global, 2012, pp. 189-208.
- [11] R. G. Stockwell, "A basis for efficient representation of the S-transform," Digital Signal Processing, p. 371-393, January 2007.
- [12] S. Drabycz, R. G. Stockwell and J. R. Mitchell, "Image Texture Characterization Using the Discrete Orthonormal S-Transform," Journal of Digital Imaging, vol. 22, no. 6, pp. 696-708, December 2009.
- [13] Y. Wang and J. Orchard, "On the use of the Stockwell transform for image compression," in Image Processing: Algorithms and Systems VII, 724504, 2009.
- [14] Y. Wang, Efficient stockwell transform with applications to image processing (Ph.D. thesis), Waterloo, Ontario: The University of Waterloo, 2011.
- [15] Y. Wang and J. Orchard, "Fast Discrete Orthonormal Stockwell Transform," SIAM Journal on Scientific Computing, vol. 31, no. 5, p. 4000-4012, 11 November 2009.
- [16] M. K. N. B. and P. K., "Palmprint Authentication System Based on Local and Global Feature Fusion Using DOST," Journal of Applied Mathematics, vol. 2014, p. 11, 16 December 2014.