AN EFFICIENT DEPTH MAP ESTIMATION TECHNIQUE USING COMPLEX WAVELETS

Pankajkumar Mendapara, Aryaz Baradarani, Q.M. Jonathan Wu

Electrical and Computer Engineering
University of Windsor, Ontario, N9B3P4
Email: {mendapa, baradar, jwu}@uwindsor.ca

ABSTRACT

A new focus measure system is proposed based on complex wavelet transform and quadrature pair of steerable filters. In shape from focus (SFF), noise, illumination variation and oriented features degrade the performance of focus measure operator. This paper introduces the use of complex wavelets due to shift-invariance and directionality of the transformation suitable for detecting various types of features which plays a pivotal role in depth estimation of a scene. A quadrature pair of steerable filters is employed to measure focus by calculating the local oriented energy of the detected features. Experimental examples are provided to illustrate the effectiveness of the approach and the results compare favorably to well-documented methods in literature.

Keywords— Depth map, shape from focus, dual-tree complex wavelets, steerable filters.

1. INTRODUCTION

The reconstruction of a geometric object and to retrieve spatial information from one or multiple observation is a challenging problem in computer vision. When a 3D scene is projected into a 2D image plane, the depth information is lost. The objective of shape from focus (SFF) and depth map computation is to determine depth of each image point from the camera lens and subsequently to reconstruct the 3D shape. The depth map estimation has numerous applications such as robot guidance, collision avoidance, medical imaging, range segmentation, microscopic imaging. The term focus can be interpreted as the degree of blurring in an image. The object points on the focus plane appear sharp and blurring is increased with the distance of imaging system from the focus plane. For the scene with considerably large depth, points on the focus plane have a sharp appearance while the rest of the scene points are blurred. Practically it is impossible to have a focus plane similar to the scene depth and to obtain sharp focus for all object points in a single image.

A basic image formation geometry, when camera parameters are known, is shown in Fig. 1. Distance of an object from camera lens $u$ is required for exact 3D reconstruction of a scene.

\[ \frac{1}{f} = \frac{1}{u} + \frac{1}{v} \]

In shape from focus the focus measure technique determines the sharpness value of each image pixel. A gradient based focus measure approach, known as sum of modified Laplacian (SML), was suggested [8] to calculate the sharpness value of image pixel. Tenenbaum [10] proposed the Sobel operator based Tenangrade focus measure (Tenen). The gray level variance (GLV) [2] and mean focus measure (Mean) are discussed in [4] that utilizes the statistical approach. In [4], Curvature and $M_2$ focus measure techniques were introduced; the former is based on gray level value and the latter deals with the energy of image gradient. An exponential decaying function based on distance transform of extracted features using SUSAN operator [6], and steerable filters based method [7] are recent developments in estimating shape from focus. Unlike [7], instead of measuring orientation energy of image directly, and to avoid post-processing using median filtering which reduces accuracy, we measure orientation energy in the wavelet domain using a quadrature pair of steerable filters. When steerable filters are designed as a quadrature pair, the basis functions of are sufficient to shift Gaussian
derivatives arbitrarily in both phase and orientation. Motivated
by image denoising with dual-tree complex wavelets [1], and re-
cent work on depth map estimation [6], in this paper we propose
the use of DT-CWT for depth map estimation via multi-focus
imaging. Taking the advantage of wavelet domain, the trans-
formation provides detailed information in wavelet subbands of
each frame in the sequence; leading to more options to select
the best fitting feature among the entire input images. Steerable
filters are employed to search for the frame number which may
contain the best focused point by measuring the local oriented
energy of features. Experimental results and quantitative com-
parisons show that the proposed framework based on DT-CWT
improves the results obtained by previous methods in literature.

2. PRELIMINARIES

2.1. Two-channel Complex Wavelet Filter Banks

Dual-tree complex wavelet transform (DT-CWT) which is an
enhancement to the discrete wavelet transform (DWT), pos-
sesses two key properties, i.e., the transformation is nearly shift
invariant and it has better directionality in higher-dimensional
space. As a typical DWT structure is a well known transforma-
tion, in this section, a brief review is given for DT-CWT only.
Consider a two-channel dual-tree filter bank implementation of
the complex wavelet transform. Figure 2(a) and (b) show the
primal and dual filter banks, respectively. Note that the scaling
and wavelet functions associated with the analysis side of the
primal bank are given by

\[ \phi_h(t) = 2 \sum_n h_0[n] \phi_h(2t - n) \]
\[ \psi_h(t) = 2 \sum_n h_1[n] \phi_h(2t - n). \]

The scaling function \( \phi_f \) and wavelet \( \psi_f \) in the synthesis side of
the primal bank are similarly defined with \( f_0 \) and \( f_1 \) and the
same discussion is valid for the scaling functions \( (\phi_f, \phi_f) \)
and wavelet functions \( (\psi_f, \psi_f) \) of the dual filter bank. The
dual-tree structure requires analytic complex wavelets \( \psi_h + j\psi_h \)
and \( \psi_f + j\psi_f \), i.e., analysis wavelet \( \psi_h(t) \) of the dual bank
is the Hilbert transform of the analysis wavelet \( \psi_h(t) \) of the
primal bank, and the synthesis wavelet \( \psi_f(t) \) of the primal
is the Hilbert transform of \( \psi_f(t) \). In other words \( \tilde{\Psi}_h(\omega) =
- j\text{sign}(\omega) \Psi_h(\omega) \) and \( \tilde{\Psi}_f(\omega) = - j\text{sign}(\omega) \Psi_f(\omega) \) where
\( \Psi_h(\omega), \Psi_f(\omega), \Psi_h(\omega), \) and \( \Psi_f(\omega) \) are the Fourier transforms of
wavelet functions \( \psi_h(t), \psi_f(t), \tilde{\psi}_h(t), \) and \( \tilde{\psi}_f(t) \) re-
spectively, and \( \text{sign} \) represents the signum function.

There are two traditional approaches to the design of Hilbert
pairs of complex wavelet bases. One is to design both the primal
and the dual banks simultaneously, e.g., Kingsbury’s q-shift [5]
and Selesnick’s common factor [9] solutions. The other method,
known as matching technique, is to design the dual filter bank
for an existing filter bank (See relevant references in [9]). In
general, the transformation is approximately shift-invariant and
provides directional-selective subbands in wavelet domain [5, 9]

\[ g^\theta(x, y) = \sum_{m=1}^M k_m(\theta) g^{\theta_m}(x, y) \]

where \( k_m(\theta) \) refers to interpolation functions, \( g^{\theta_m}(x, y) \) forms
basis filters and \( \theta \) is an arbitrary rotation. The second deriva-
tive of Gaussian, \( G_2(x, y) \) [3], is used as a steerable filter in
this paper. An arbitrary orientation of \( G_2(x, y) \) and its Hilbert
transform \( H_2(x, y) \) are expressed as

\[ G_2^\theta(x, y) = k_1(\theta) G_2^0(x, y) + k_2(\theta) G_2^\pi(x, y) \]
\[ + k_3(\theta) G_2^{2\pi}(x, y) \]
\[ H_2^\theta(x, y) = l_1(\theta) H_2^0(x, y) + l_2(\theta) H_2^\pi(x, y) \]
\[ + l_3(\theta) H_2^{2\pi}(x, y) \]

where normalized basis filters spaced equally between 0 and \( \pi \)
with the following interpolation functions [3] for the Gaussian,
\( k_1(\cdot) \), and the Hilbert transform, \( l_1(\cdot) \), respectively.

\[ G_2^0(x, y) = 0.9213(2x^2 - 1)e^{-(x^2+y^2)} \]
\[ G_2^\pi(x, y) = 1.843x\gamma e^{-(x^2+y^2)} \]
\[ G_2^{2\pi}(x, y) = 0.9213(2y^2 - 1)e^{-(x^2+y^2)} \]
\[ k_1(\theta) = \frac{1}{3}(1 + 2\cos 2\theta) \]
\[ k_2(\theta) = \frac{1}{3}(1 + 2\cos 2(\theta - \frac{\pi}{3})) \]
\[ k_3(\theta) = \frac{1}{3}(1 + 2\cos 2(\theta - \frac{2\pi}{3})) \]

The quadrature filter of \( G_2 \) is the Hilbert transform of it,
which cannot steer. Therefore, a third-order polynomial is used
for approximation and consequently the corresponding normal-
ized basis filters of \( H_2(x, y) \) are given by
\[ H_0^2(x, y) = (-2.205x + 0.978x^3)e^{-(x^2+y^2)} \]  
\[ H_2^2(x, y) = (-0.735y + 0.978y^2)e^{-(x^2+y^2)} \]  
\[ H_2^4(x, y) = (-0.735y + 0.978xy^2)e^{-(x^2+y^2)} \]  
\[ H_2^8(x, y) = (-2.205y + 0.978y^3)e^{-(x^2+y^2)} \]  
\[ l_1(\theta) = \cos^3\theta \]  
\[ l_2(\theta) = -3\cos^2\theta \sin\theta \]  
\[ l_3(\theta) = +3 \cos \theta \sin^2 \theta \]  
\[ l_4(\theta) = -\sin^3\theta. \]  

3. PROPOSED ALGORITHM

3.1. Wavelet-based Depth Map Estimation

In SFF the success of any focus measure technique depends on its ability to calculate sharpness value of each image pixel. We take the advantage of dual-tree complex wavelet transform to decompose a pre-registered multi-focus image sequence to several directional subbands in complex wavelet domain. We then search for the best focus value using a quadrature pair of steerable filters. The decomposition in wavelet domain provides detailed information oriented in different directions. Note that decimation step is removed between levels to preserve information while decomposing data by DT-CWT. The intermediate subband with higher pixel value is computed employing the well known maximum selection rule on wavelet subbands of each image at the last level of the decomposition. After finding best feature information for each image local orientation energy \( E \) for each pixel is measured. Local orientation energy is calculated by steering a quadrature pair of steerable filters and energy for each pixel is given by

\[ E(x, y, \theta) = [I(x, y) * G_2^0(x, y)]^2 + [I(x, y) * H_2^0(x, y)]^2 \]  

Once the energy of each pixel is determined using a small window around a pixel \((x, y)\), focus measure (FM) is calculated and the value of pixel \((x, y)\) is replaced by the sum of computed values to avoid abrupt fluctuations and measurement errors. That is

\[ FM_W(i, j) = \sum_{x=i-N}^{i+N} \sum_{y=j-N}^{j+N} E(x, y, \theta) \]  

where the subscript \( W \) refers to wavelet domain. To obtain a decision map, the frame number with higher value for a corresponding pixel is mapped onto the map and therefore the decision map is constructed accordingly. Using this decision map pixels are extracted from the scaling and wavelet coefficient of each image and then inverse transform yields the fused image. The block diagram of the proposed technique for depth map estimation is depicted in Fig. 3.

3.2. Results and Discussions

The performance of the proposed wavelet-based depth map estimation technique is tested with several multi-focus image sequences. The results obtained with the proposed method are compared with other existing methods. The proposed method shows superior performance in terms of sharpness and detail preservation in the fused image.
Fig. 5. Depth map for simulated cone using various methods.

Fig. 6. Depth map for simulated cone with Gaussian noise (variance=0.001) using various methods.
sequences, namely, simulated cone, chess, clock, and Pepsi. Frames 20 and 40 are shown in Fig. 4; selected from 75 frames of size $318 \times 318$ sample frames of the simulated cone dataset.

The main goal in shape from focus is to transfer the most relevant information found in source images to a fused image as well as estimating the depth of each pixel. The resultant depth map for simulated cone dataset obtained by our proposed method and four other well-documented focus measure methods, FM$_{M_2}$, FM$_{GLV}$, FM$_{Tenen}$ and FM$_{SML}$, are shown in Fig. 5. It is seen that the depth map generated using proposed method is more accurate compared to previous techniques. Note that the results obtained using DWT and DT-CWT (Figs. 5 (e) and (f)) are free of spikes in contrast with FM$_{M_2}$ and FM$_{GLV}$ (Figs. 5 (a) and (b)) where the artifacts are clearly visible. In Fig. 6, noise is also added to the simulated cone dataset. The added noise is Gaussian with zero mean and a variance of 0.001. Figs. 6 (a)–(d) show the depth maps calculated using various methods. The results obtained by proposed method via using DWT and DT-CWT are given in Figs. 6 (e) and (f) respectively; the depth map obtained using DT-CWT clearly outperforms the previous approaches whereas M$_2$ and GLV simply fail to resists against noise. We have used structural similarity (SSIM) index [11] and objective performance metric (EI) [12] to evaluate the efficiency of the method numerically and to compare with the previous approaches in literature. These factors estimate how and what information is transferred from the input images to the fused image, i.e., assess the fusion on the basis of input-output relationship and does not need any ground-truth or reference image. The objective performance metric 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. That is

$$\text{EI} = \sum_{x=1}^{N} \sum_{y=1}^{M} Q^{AZ}(x,y)W_A(x,y) + Q^{BZ}(x,y)W_B(x,y)$$

(23)

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, and $Z$ is a composite image. SSIM image quality index is based on structural similarity, and local SSIM measures three elements; the similarity of brightness, contrast and structures. SSIM and mean SSIM (MSSIM) can be calculated as

$$\text{SSIM}(h_i, Z) = \frac{(2\mu_{h_i} \mu_Z + C_1)(2\rho_{h_iZ} + C_2)}{(\mu_{h_i}^2 + \mu_Z^2 + C_1)(\rho_{h_i}^2 + \rho_Z^2 + C_2)}$$

(24)

$$\text{MSSIM} = \sum_{h_i=1}^{N} \text{mean} (\text{SSIM}(h_i, Z))$$

(25)

where $h_i, i = 1, 2, \ldots, N$ represents input image sequence, $\mu_{h_i}$ and $\mu_Z$ are local sample means of $h_i$ and $Z$ respectively and $\rho_{h_iZ}$ 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.

To test the robustness of the method Gaussian noise with different variances is added to the datasets. For simulated cone dataset, EI and SSIM comparison without noise and in presence of noise are given in Tables 1 and 2 respectively, confirming the experimental results and visual improvements obtained. Similar demonstration is provided for the clock dataset in Tables 3 and 4. For reference, we also show the results obtained by DWT using the proposed method. It is worth pointing out that in presence of noise the proposed method still can track the actual shape of the object while the performance of the other methods deteriorates.

## 4. CONCLUSION

An efficient focus measure technique is proposed for depth map estimation based on the use of complex wavelet subbands to extract detailed feature information from a multi-focus image sequence. Local oriented energy of the detected features is calculated using quadrature pair of steerable filters. Steerable filters remove inherent limitations of traditional gradient detection based techniques which perform inadequately for oriented intensity varying scenarios. Simulation results show the performance improvement of DT-CWT based technique employing steerable filters in terms of fusion assessment factors and
visual perception. Shown in the contour plane, the boundary of depth are accurately extracted compare with previous techniques. This is due to better directional-selectivity and shift-invariance attribute of DT-CWT in general. In view of success of the presented approach in depth map estimation, it is reasonable to hope promising results, employing the proposed framework, in relevant applications such as 3D shape reconstruction and object identification.

### 5. REFERENCES


