3D shape recovery from image focus using kernel regression in eigenspace

Muhammad Tariq Mahmood, Tae-Sun Choi *

Signal and Image Processing Lab, School of Information and Mechatronics, Gwangju Institute of Science and Technology, 261 Cheomdan Gwagiro, Buk-Gu, Gwangju 500-712, South Korea

1. Introduction

Recovering 3D shape of objects from their 2D image(s) is a fundamental problem in computer vision. It has numerous applications in computer vision, range segmentation, consumer video cameras, and video microscopy [1–3]. In literature, several cues including illumination, shading, texture, motion, defocus, and focus have been explored to capable machines of extraction depth information from the visual environment. Shape from focus (SFF) is one of the passive optical methods that uses degree of focus as the cue to estimate the 3D structure of the object. In SFF methods, a single camera is used to acquire a sequence of images, either by varying the distance between the object and the camera along the optical axis or by changing the focus of the camera. A focus measure is applied to each pixel in every frame to measure the focus quality. The sharpest pixel among the sequence provides depth information. An approximation method is followed to enhance the initial estimate obtained through focus measure.

Focus measure plays an important role in SFF algorithms, as it is a crucial step in the calculation of the depth map. In conventional focus measures, the sharpness is computed by aggregating the focus values within a small 2D window around each pixel. The size of window affects the depth map accuracy and computational complexity [4,5]. The accuracy of SFF methods also suffer from the errors due to translation, magnification, and vibration. Translation and magnification errors are introduced because of movement of the object away from the camera. Discreteness in focus levels results in loss of focus information between two consecutive frames. The local summation of focus values normalizes the sharpness values within a frame but enlarges the difference between adjacent frames. Thus, a coarse surface is obtained. Additionally, mostly focus measures perform poorly in the presence of noise and weak texture [6,7].

In this paper, we introduce a new algorithm (PCA-KR) for 3D shape recovery by applying kernel regression in feature space. The focus values are computed through principal component analysis (PCA) by considering a sequence of small 3D neighborhood for each object point. The initial work in this direction is presented in [8]. In our previous work, we also observed that the application of PCA on discrete cosine transform (DCT) and discrete wavelet transform (DWT) coefficients resulted in better 3D shapes [9,10]. The current work extends the idea of using PCA for computing sharpness values for image volume by integrating it with kernel regression that has, recently, gained sizeable attention in image processing [11,12]. Kernel regression is a non-parametric statistical technique to estimate the model directly from the data. It has certain advantages over conventional parametric regression methods. We apply unsupervised regression through Nadaraya and Watson Estimate (NWE) on depth values to get a refined 3D shape of the object. It reduces the effect of noise within a small surface area as well as approximates the accurate 3D shape by exploiting the depth dependencies in the neighborhood. Performance of the proposed scheme is investigated in the presence of different types of noises and textured areas. Experimental results demonstrate effectiveness of the proposed approach.
the presence of different types of noises and textures. Experimental results show that the proposed method is more accurate and robust when compared with existing approaches.

In the remainder of this paper, Section 2 describes SFF problem and algorithms previously proposed in literature. The proposed approach is explained in Section 3. Experimental results, comparative analysis, and discussions are presented in Section 4. Finally, Section 5 concludes this study.

2. Background

2.1. Problem formulation

Fig. 1 represents the basic image formation in a convex lens. The relationship between distances of object point \( P \) in object plane and its well-focused image at \( P \) is given by the lens formula.

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

where \( f \) is the focal length of the lens, \( u \) is the distance of the object point from the lens, and \( v \) is the distance of its image from the lens.

If object point is well focused in image plane, we get its clear image. Otherwise, if it is not well focused then the energy from the object point (reflection rays) through the convex lens is distributed over a circular patch of radius \( R \) on image detector (ID) and consequently, we get its blurred image. This blurred image is the linear convolution of the original image and the point spread function (PSF). In SFF techniques, the depth map is computed by determining the particular distance of the object point \( P \) from the camera for which the image detector (ID) is exactly at the point \( P \). Once, distances for all points of the object are calculated, the 3D shape can easily be recovered. In general, SFF algorithm can be divided into two parts; focus measure and approximation technique. Focus measure is applied to compute the degree of focus for each pixel in the image volume and approximation technique is used to refine the initial depth estimate.

2.2. Focus measures

Focus measure determines the focus value or sharpness of image pixel. Usually high pass filtering is used to determine sharpness. In literature, many focus measures have been proposed in spatial as well as in frequency domain [13–16]. In spatial domain, gradient-based focus measures such as modified Laplacian (ML), sum modified Laplacian (SML), threshold absolute gradient (TAG) and tenenbaum focus measure (TFM) are commonly used. Gray level variance (GLV), autocorrelation, and focus measures based on statistical moments are also famous and commonly used in noisy environment. These focus measures compute focus value locally by considering a small rectangular window around each pixel of every frame in the image volume. Ahmad and Choi [17] suggested 3D window to obtain a more accurate depth map.

Among the variety of focus measure operators, some of them are reported in discrete cosine transform (DCT) and discrete wavelet transform (DWT) domains. A brief summary of these focus measures can be found in [9,10]. In these methods, the energies of high frequency coefficients or the ratio of energies of high frequency components to low frequency components is suggested as focus measures. The energies are computed by taking L1-norm or L2-norm (summation of individual energy of each coefficient) of the coefficients. The studies of these methods also have revealed that the diverse coefficients affect focus quality differently. Therefore, this norm may not produce accurate depth map. This motivated us to suggest the use of PCA to get the axis of maximum variation in the energies of the high frequency components in the transformed domain [9].

2.3. Approximation methods

An approximation method is applied to obtain a refined depth map. In the traditional method SFF-TR [6], the focus value is computed for each pixel of every image by applying \( F_{SML} \). In the second step, Gaussian model is fitted to three focus values near the peak. Subbarao and Choi [18] introduced the concept of focus image surface (FIS) and proposed a method to recover an accurate 3D shape.

In this method, a rough estimate is taken by applying \( F_{SML} \) and then the best-focused values lying on FIS are searched in second step. Choi and Yun [19] proposed an estimation of FIS through piecewise curved surface approximated through interpolation by using second order Lagrange polynomial. Asif and Choi [20] applied neural network to optimize the FIS. An initial estimate is computed by applying \( F_{GLV} \) and then the results are provided to the neural net, whereby weights are updated by using back propagation algorithm to maximize focus measure at the output layer. Ahmad and Choi [2] proposed a method SFFDP using dynamic programming (DP) to obtain an optimal FIS. In first step, a rough depth map is computed through \( F_{ML} \) and then it is optimized through DP in second step. These above discussed methods provide better results than traditional methods but at high computational cost.

3. Proposed approach

3.1. Motivation

As mentioned in Section 2, usually conventional SFF methods compute depth map in two steps: an initial estimate obtained by applying a focus measure locally, and then depth map is refined through an approximation method. The performance of these methods largely depends upon the initial estimate. Most commonly used focus measures are gradient based and are very sensitive to noise. Therefore, the error in initial estimate is propagated to the second step. Moreover, according to paraxial geometrical optics, there is one-to-one correspondence between object point and well-focused point in image plane. However, sensed images through CCD are two dimensional, so only parts of the object that are within the depth of field (DOF) have sharp images while others are blurred. In order to have 3D information of the whole object, we take a sequence of images at different focus levels. Thus, loss of focusing information between two connective frames results in inaccurate depth map. In literature, curve fitting and surface approximation techniques have been suggested to overcome this limitation [6,19]. These techniques are the special cases of parametric regression, and are based on assumption that the focus val-
ues follow some specific distribution models. However, the initial computed focus values may not follow some specific distribution. Therefore, this is another factor, which restricts the existing approaches to compute the accurate depth map. In proposed method, PCA is applied to compute the degree of sharpness and then the accurate depth map is approximated through non-parametric kernel regression.

3.2. Data transformation into eigenspace

We applied PCA for computing the focus value globally. PCA is a powerful stochastic approach to deal with high dimensional data and it is widely used in pattern recognition and feature extraction applications [21–23]. It transforms the given data into eigenspace such that the first few components contain more variance than others and noise is concentrated in lower order components or features. In proposed method, an image sequence \( I(x, y) \) of an object of unknown shape is obtained by moving the image detector in the optical axis direction. The sequence consists of total \( Z \) images each having \( X \times Y \) pixels. Instead of taking a neighborhood in 2D, we consider 3D neighborhood to incorporate the effect of previous and next frames and intensity values are arranged into a vector for each pixel of every frame. There are many possible combinations for 3D neighborhood selection; few of them are illustrated in [24]. For an object point, a matrix \( M(i, j) \) with dimensions \( Z \times n \) (where \( n \) is the number of pixels in the vector) is obtained by collecting all such vectors along the optical axis. PCA is then applied on this matrix \( M(i, j) = [m_{ni}] \). Mean vector \( \mu \) and covariance matrix \( C \) are computed by using Eqs. (2) and (3), respectively as:

\[
\mu = \frac{1}{Z} \sum_{k=1}^{Z} m_k
\]

(2)

\[
C = \frac{1}{Z-1} \sum_{k=1}^{Z} (m_k - \mu)(m_k - \mu)^T
\]

(3)

The eigenvalues and their corresponding eigenvectors are computed from the covariance matrix \( C \). The transformation matrix \( E \) is formed such that the first row is the eigenvector corresponding to the largest eigenvalue, and the last row is the eigenvector corresponding to the smallest eigenvalue. The transformed vectors \( f_k \) are then obtained by multiplying transformation matrix \( E \) with the mean subtracted data as follows:

\[
f_k = E(m_k - \mu)
\]

(4)

3.3. Initial depth map

An initial depth (frame number) can easily be computed for each object point by maximizing the sharpness of focus curve in the direction of optical axis:

\[
D(i, j) = \arg \max_k |f_{ik}|
\]

(5)

The analysis of the data in eigenspace shows that the first principal component provides greater discriminating power regarding focus values as compared to other features (discussed in Section 4.2.1). Therefore, it is reasonable to employ it for calculating the focus values. The first principal component is in the direction of maximum variance in the data. In our case, the data transformed through first component contains around 90% of the total variance. While, rest of the components occupy a small amount of total variance and are discarded, as they contain noisy information. The depth computed from transformed data is more accurate as it incorporates the effect of pixels from the whole sequence simultaneously. Earlier methods locally compute the focus values. In contrast, the proposed method globally computes focus values. Therefore, the depth (frame number) obtained through proposed method is more accurate as compared to the previous methods.

Initial depth values computed by Eq. (5) within a small patch on the object surface are very close to each other. However, due to the enhancement of noise related intensities, false depths may be computed and can appear as peaks in the depth map. Therefore, the obtained depth map is further refined by applying non-parametric regression on the depth values. It also reduces the error due to the discreteness of frames.

3.4. Depth refinement through kernel regression

The key idea of regression is to estimate the function from the given data points. In literature, many regression approaches have been presented. Commonly, these approaches can be divided into two categories namely parametric and non-parametric. In parametric regression methods, the formulation of functions, along with parameters to be estimated, is provided in advance. However, non-parametric regression estimates function directly from the data rather than to estimate parameters. In kernel regression, firstly, the data is mainly mapped into a higher dimensional space by using a kernel method so that a nonlinear regression problem may be simplified into a linear regression problem. The regression estimator always yields a weighted linear combination of nearby data points. In one dimensional case, for \( l \) samples of the given data \( \{(a_1, g(a_1)), (a_2, g(a_2)), (a_3, g(a_3)), \ldots, (a_l, g(a_l))\} \) can be written as \( g(a_i) = r_i(a) + u_i \). \( 0 < i < l \) so that \( E(u_i | a_i) = 0 \). Irrespective of regression order and smoothing parameters, since, the regression estimator always yields a weighted linear combination of nearby samples. So, it can be written as:

\[
r_i(a) = \sum_{i=1}^{l} W_{li}(a) \cdot g(a_i)
\]

(6)

where \( r_i(a) \) is the estimated data, and \( W_{li} \in \mathbb{R} \) are non-negative weights. The Nadaraya and Watson Estimator (NWE) is the simplest non-parametric regression estimate [11], and it can be expressed as:

\[
r_i(a) = \frac{\sum_{i=1}^{l} K_h(a_i - a) g(a_i)}{\sum_{i=1}^{l} K_h(a_i - a)}
\]

(7)

\[
W_{li}(a) = \frac{K_h(a_i - a)}{\sum_{i=1}^{l} K_h(a_i - a)}
\]

(8)

where \( K_h \) is the kernel function with bandwidth \( h \). The appropriate choices of bandwidth and regression order are important factors. They can affect the smoothness and the bias and variance of the estimate. The smaller values of \( h \) result in small bias and large variance in estimate. Optimal regression order and smoothing parameter selection are discussed in [11]. Different types of kernel functions such as spline, tri-cube, local polynomials, Gaussian radial basis and sigmoid functions are commonly used. We use the Gaussian radial basis function with appropriate value for deviation parameter. It can be expressed as:

\[
K(a, a') = \exp \left(-\frac{||a - a'||^2}{2\sigma^2}\right)
\]

(9)

Two dimensional NWE on the initial estimated depth map is employed. For each depth value, an estimated depth through the regression, using a small kernel of size \( 3 \times 3 \) is obtained. The larger size of the kernel may oversmooth edges.
4. Results and discussion

4.1. Experimental setup

We conducted experiments using synthetic as well as real objects. In order to evaluate the performance quantitatively, an image sequence of simulated cone was generated synthetically by using simulation software. More details about the procedure and image generator can be found in [18,25]. In addition, to evaluate the robustness of the proposed method, we conducted experiments with noisy image sequences. The images taken through CCD may have many types of noises, however, three of them namely Gaussian, speckle, and salt and pepper noises are more notable with respect to optical imaging system. Malik and Choi [7] have discussed

![Microscopic control system](image)

Fig. 2. Microscopic control system.

![Sample images of test objects](image)

Fig. 3. Sample images of test objects: (a) simulated cone, (b) TFT–LCD color filter, (c) micro sheet, and (d) real cone.
the effect of noise on SFF methods. Image sequences of simulated cone corrupted with Gaussian, speckle, and short noise, with zero mean and nonzero variance, were obtained. An image sequence of real cone object consisting of 97 images, each of 200 $\times$ 200 dimensions, was also used in experiments. The real cone object is made of hardboard with black and white strips drawn on its surface to enrich the texture. Besides, two microscopic objects TFT–LCD color filter and micro sheet were considered for experiments. TFT–LCD color filter are thin and bright displays and are combination of cells and color filters. The micro sheet object was constructed by preparing copper solution through $\text{Cu(NO}_3\text{)}_2\cdot3\text{H}_2\text{O}$, NaOH, and distilled water. Under specific temperature, the solution was then transferred into Teflon-lined stainless steel autoclave of 100 mL capacity for specific period. The image sequences of these

![Fig. 4. (a) Original intensities for the sequence of object point (200, 200) from simulated cone with 6-pixels neighborhood and (b) transformed data into eigenspace.](#)

![Fig. 5. Focus measure performance comparisons: (a) focus curves computed for two object points from different texture samples, (b) Gaussian model fitted to the focus curves for object point (50, 50) from sample A, (c) Gaussian model fitted to the focus curves for object point (150, 150) from sample B, and (d) resolution comparisons.](#)
Fig. 6. 3D shapes reconstructed of test objects: (a and b) simulated cone, (c and d) TFT-LCD color filter, (e and f) micro sheet object, and (g and h) real cone.
objects were obtained from microscope control system (MCS) shown in Fig. 2. The system comprises of a personal computer, a frame grabber board Matrox Meteor-II, a CCD camera (SAMSUNG CAMERA SCC-341), motor driver MAC 5000 with 2.5 nm (nanometer) step length, a microscope (NIKON OPTIPHOT-100S) and image capturing software. The CCD camera is mounted on the microscope and the images were acquired by controlling the lens position through the software. Keeping the illumination conditions constant, sequences of 60 images for the both test object TFT–LCD color filter and micro sheet were obtained under 50× magnifications.

The motor took total 10,080 and 4260 steps. Thus, the distance between two consecutive frames of the TFT–LCD color filter was 0.42 μm (micrometer), and for the object micro sheet it was 0.1775 μm. Sample frames from the image sequences are shown in Fig. 3.

4.2. Comparative analysis

The performance of a focus measure is usually gauged based on unimodality and monotonicity of the focus curve. Zhu et al. [12] used the resolution metric for the focus measure evaluation. Malik and Choi [7] and Mahmood et al. [10] applied root mean square error (RMSE) and correlation metrics for the depth map evaluation. We employed these three performance measures to compare the performance of the proposed focus measure. Among the variety of focus measures, well-known focus measures \( F_{SML}, F_{TEN}, \) and \( F_{GLV} \) are selected for the performance comparison. The results are then analyzed qualitatively and quantitatively.

4.2.1. Qualitative analysis

To illustrate the proposed method, consider the object point (200, 200) of the test object simulated cone and the 6-pixels neighborhood, a matrix of size 97 × 7 is populated. Fig. 4a shows the original gray level values for the pixels (P1–P7) at each step length for the total 97 steps. PCA is applied on this matrix and the transformed data into eigenspace is shown in the Fig. 4b. It can be observed that the curve of the first feature \( F_1 \) is finer than the resultant curves for the remaining six features. Unimodality and monotonicity of the curve from the first feature can be compared to the curves that are obtained through other methods. The first feature contains more than 85% of the total variance, while the remaining portion is distributed among other features. Moreover, the noise is concentrated in lower order features. Therefore, only the first feature is selected for computing depth value. The location of the peak in this curve is taken as depth value for the point under consideration.

Fig. 6 shows the 3D shapes reconstructed for the test objects; simulated cone, TFT–LCD color filter, micro sheet and real cone. The tip of simulated cone must be sharp and smooth [2]. It can be seen from the right column in Fig. 6, the 3D shapes recovered by the proposed approach are better than those that are recovered by existing methods. It is also notable that the shapes reconstructed by existing approaches have coarse surfaces due to the local summation of the focus values.

Two samples from images of real cone, having different textures and illuminations, are taken. It can be observed that the upper left corner is well illuminated and has high texture contents, while the lower right corner has relatively weak texture. Sample A is taken...
from the upper left corner whereas the sample B is a part of the lower left corner of the real cone image. In Fig. 5a, focus curves for the object points (50, 50) and (175, 175) are obtained through the proposed method. We observed that the curve for the stronger textured object point has sharper peak and high focus values as compared to the object point with relatively weaker texture. However, even in the case of weak textured object point with insufficient illumination, the well-focused position can easily be found in the focus curve. Furthermore, for these object points Fig. 5b and c shows the focus curves fitted to Gaussian model from different methods. We can observe that the curves obtained by proposed method are narrower and sharper near the best-focused position as compared to other methods. Fig. 5d compares the resolution values for the above discussed object points.

Figs. 7 and 8 show the shape recovery of the real objects TFT–LCD color filter and micro sheet from their noisy image sequence. Fig. 7a and c are reconstructed from image sequences of TFT–LCD color filter corrupted with Gaussian noise with zero mean and 0.05 variance. Similarly, Fig. 7b and d are recovered from image sequences contaminated with salt and pepper noise with 0.05 densities. Fig. 8 shows the 3D shape recovered for micro sheet object from the images corrupted with speckle noise with variance 0.05. As micro sheet has some transparent patches on its surface, the recovered shapes from noisy images are comparatively more degraded. From these figures, it is observed that the shapes recovered through proposed method are better as compared to those from existing approaches.

4.2.2. Quantitative analysis

Table 1 compares the resolution values computed for the two object points (50, 50) and (175, 175) taken from real cone with different textured and illumination conditions. We fitted the obtained focus curves to the Gaussian model and scaled the function values to (0, 1). The resolution was computed by using seven positions on each side of the best focus position (peak). An object point (50, 50), which is taken from the rich textured and well-illuminated area, is well focused at the 70th position out of total 97 steps. The resolution computed for the proposed method was 0.9777, which is the smallest as compared to other methods. An object point (150, 150) is taken from the weak textured and poorly illuminated area. It has the best-focused position at 52nd step out of total 97 frames. It can be observed from the Fig. 5b and c that the fitted focus curve, for the proposed method, is narrower and sharper.

<table>
<thead>
<tr>
<th>Focus measure</th>
<th>Sample A</th>
<th>Sample B</th>
</tr>
</thead>
<tbody>
<tr>
<td>$F_{SMR}$</td>
<td>8.9163</td>
<td>5.1642</td>
</tr>
<tr>
<td>$F_{TEN}$</td>
<td>7.4568</td>
<td>5.1725</td>
</tr>
<tr>
<td>$F_{GLV}$</td>
<td>4.5528</td>
<td>2.0615</td>
</tr>
<tr>
<td>$F_{PCA-KR}$</td>
<td>3.2066</td>
<td>0.9777</td>
</tr>
</tbody>
</table>

Inherently, noise is collected in lower order features in eigen-space. We have considered only the first feature in the proposed method.
algorithm, thus the noise effect is reduced by discarding all other noisy features. Therefore, the proposed algorithm performs well even in the presence of noise. The proposed approach has been tested for image sequences corrupted with different types of noise. Figs. 9–11 compare the RMSE and the correlation values computed for depth maps constructed from image sequences corrupted with Gaussian, speckle, and salt and pepper noises. It can be observed

<table>
<thead>
<tr>
<th>Focus measure</th>
<th>RMSE</th>
<th>Correlation</th>
</tr>
</thead>
<tbody>
<tr>
<td>$F_{SML}$</td>
<td>8.3714</td>
<td>0.9349</td>
</tr>
<tr>
<td>$F_{TEN}$</td>
<td>8.2432</td>
<td>0.9385</td>
</tr>
<tr>
<td>$F_{GLV}$</td>
<td>8.2251</td>
<td>0.9363</td>
</tr>
<tr>
<td>$F_{PCA-KR}$</td>
<td>7.8020</td>
<td>0.9586</td>
</tr>
</tbody>
</table>

Fig. 9. Performance comparison in the presence of Gaussian noise (a) correlation and (b) RMSE.

Fig. 10. Performance comparison in the presence of speckle noise (a) correlation and (b) RMSE.

Fig. 11. Performance comparison in the presence of salt and pepper noise (a) correlation and (b) RMSE.
However, in case of salt and pepper noise, it also performed poorly. The second derivative based focus measure $F_{SMR}$ depreciated in a large amount, as it is very sensitive to noise. $F_{CLV}$ focus measure provided reasonable resistance against Gaussian and speckle noises, however, in case of salt and pepper noise, it also performed poorly. However, $F_{TEN}$ focus measure has shown considerable resistance against diverse types of noises.

5. Conclusion

In this paper, we have introduced a novel method for SFF based on PCA. Instead of applying a focus measure locally, a small neighborhood vector in 3D is considered. PCA is then applied on the sequence of neighborhood vectors to transform the data into eigenspace. An initial depth estimate is obtained by searching the position of the maximum absolute value from the score of the first component. Non-parametric kernel regression with a small kernel size is applied on the initial depth map to estimate an accurate 3D shape of the object. The proposed algorithm was tested using image sequences of synthetic and real objects. The performance of the proposed method was compared with existing techniques using three metrics: resolution, RMSE and correlation. The experimental results have demonstrated the effectiveness of the proposed method.

Acknowledgement

This work was supported by the Korean Research Foundation Grant funded by the Korean Government (KRF-2008-314-D00404).

References