A Computational Algorithm for Unified Focus and Defocus Analysis for 3D Scene Recovery

Murali Subbarao and Yen-Fu Liu
murali@sbee.sunysb.edu, yfliu@sbee.sunysb.edu
State University of New York at Stony Brook
Stony Brook, New York 11794-2350, USA

ABSTRACT

The theory of Unified Focus and Defocus Analysis (UFDA) was presented by us earlier and it was extended to use both classical optimization technique and regularization approach for 3D scene recovery. In this paper we present a computational algorithm for UFDA which uses variable number of images in an optimal fashion.

UFDA is based on modeling the sensing of defocused images in a camera system. This approach unifies Image Focus Analysis (IFA) and Image Defocus Analysis (IDA), which form two extremes in a range of possible methods useful in 3D shape and focused image recovery.

The proposed computational algorithm consists of two main steps. In the first step, an initial solution is obtained by a combination of IFA, IDA, and interpolation. In the second step, the initial solution is refined by minimizing the error between the observed image data and the image data estimated using a given solution and the image formation model. A classical gradient descent or a regularization technique is used for error minimization. Our experiments indicate that the most difficult part of the algorithm is to obtain a reasonable solution for the focused image when only a few image frames are available. We employ several methods to address this part of the problem. The algorithm has been implemented and experimental results are presented.

Keywords: image focus analysis, image defocus analysis, optimization, three-dimensional scene recovery.

1 Introduction

Several methods of three-dimensional (3D) scene recovery based on image focus and defocus analysis have been proposed [1–3,5,6,8–10,13,16]. These methods can be divided into two major types – *Image Focus Analysis* (IFA) and *Image Defocus Analysis* (IDA). Image Focus Analysis methods process many image frames with different levels of focus or defocus, but Image Defocus

Analysis methods process only a few such image frames. A new theory named Unified Focus and Defocus Analysis (UFDA) [11,12] was proposed by us which provides a unified theoretical framework for IFA and IDA. A computational algorithm was proposed for UFDA [11,12] based on using a large number of image frames. Here we propose an algorithm to deal with variable number of images.

Our computational algorithm is demonstrated with two experiments on simulated image data. In the first experiment, three image frames are used to solve for the distance and focused image of a planar object. Here a classical gradient descent approach is used for error minimization. In the second experiment, five image frames and a regularization approach are used for a spherical object. Experimental results show that obtaining a reasonable solution for focused image from only a few image frames is difficult but important in obtaining a good solution for the 3D depth-map of a scene.

2 Unified Focus and Defocus Analysis (UFDA)

UFDA is a new theory that unifies IFA and IDA. This section provides a brief summary of UFDA. More details can be found in [11,12].

Consider a camera that records the image of a 3D scene. The image coordinates of a point light source are denoted by (x', y') and the image coordinates of a point where the brightness is measured on the image detector are denoted by (x, y). The camera parameters are denoted by $\mathbf{e} = (D, f, s)$ where D is the diameter of camera aperture, f is the focal length of the lens, and s is the distance between the lens and the image detector. Let d be the blur parameter (e.g. blur circle diameter) of a reference point (e.g. a point at a known distance, say infinity, on the optical axis) and let d' denote the blur parameter of an arbitrary scene point (at an unknown distance). The blur parameter d (and also d') can be continuously changed by changing the camera parameter setting $\mathbf{e} = (D, f, s)$. Let $g(x, y, d)$ denote an image volume data corresponding to the 3D scene recorded by the camera (by changing the camera parameters to change d).

In the (x', y', d') space, $d' = d_v(x', y')$ is used to denote a surface named Focused Image Surface (FIS) of the 3D scene in front of the camera. This surface uniquely corresponds to the 3D shape of the scene. The focused image of the scene is denoted by $F(x', y')$ and it is related to the recorded image volume data and FIS by

$$F(x', y') = g(x', y', d_v(x', y')) \tag{1}$$

We define the Focused Image Volume (FIV) $F'(x', y', d')$ as

$$F'(x', y', d') = \begin{cases} F(x', y') & \text{if } d' = d_v(x', y') \text{ and} \\ 0 & \text{otherwise.} \end{cases} \tag{2}$$

The image formation characteristics of the camera are specified in terms of a 3D Point Spread Function (PSF) $h(x, y, d)$. Under certain weak conditions we can derive a three-dimensional convolution expression[11]:

$$g(x,y,d) \;=\; \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} F'(x',y',d')\; h(x-x',y-y',d-d')\; dx'\; dy'\; dd'. \tag{3}$$

The above equation relates the recorded image volume data $g(x,y,d)$ to the scene characteristics (specified by the 3D shape $d_v(x',y')$ and focused image $F(x',y')$) and camera characteristics (specified by the 3D PSF). Details on the 3D PSF and the derivation of the 3D convolution equation are reported in [11].

In UFDA, a sequence of image frames $g(x,y,d_i)$ are given for $i = 1,2,\cdots n$. Also, the 3D PSF of the camera is given (or obtained through calibration). The problem is to find the 3D shape of the scene specified by $d_v(x',y')$ and the focused image of the scene $F(x',y')$. This problem is formulated as an optimization problem where the difference or mean-square error E between the observed image data $g_o(x,y,d_i)$ and the estimated image data $g_e(x,y,d_i)$ is minimized:

$$E \;=\; \sum_{i=1}^{n}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} (g_o(x,y,d_i) - g_e(x,y,d_i))^2\; dx\; dy \tag{4}$$

The estimated image data $g_e(x,y,d_i)$ is obtained from Eq. (3) using the 3D PSF h and the current best known solutions to the 3D shape $d_v(x',y')$ and focused image $F(x',y')$. An initial estimation of the solutions is obtained through a combination of IFA, IDA, and interpolation methods. This solution is improved iteratively by an optimization technique.

Two optimization techniques for the UFDA have been developed by us earlier. One is a classical gradient descent approach [11] and the other is a regularization technique [12]. The first technique is simpler in both theory and implementation, but the second (regularization) technique is better and faster for smooth objects. See [11,12] for details.

3 Computational Algorithm

The computational algorithm we propose consists of two main steps. The first is the estimation of initial solution for 3D shape and focused image, and the second is the iterative improvement of the initial solution through error minimization. This algorithm extends previous research on UFDA in two respects. The first extension is a systematic way to deal with variable number of images, from minimum needed to maximum possible. The second extension is an investigation of different techniques for estimating the focused image for the case of variable number of images.

We present the computational algorithm for a specific case based on an actual camera that we use in our laboratory, but it can be generalized easily. Our camera has the following specifications: focal length f=35 mm, aperture diameter D=9 mm, width of square pixels p=0.013 mm, and lens displacement for each step of a stepper motor is approximately 0.030 mm. Let the observed image data be represented by $g(x,y,d)$ where x and y are pixel indices and d is the blur parameter index. The image data will be represented with 8 bits per pixel and therefore g will be assumed to take integer values: $0, 1, 2, \cdots 255$. We choose the blur parameter d to be the blur circle diameter of a point light source at infinity. It is given by $d = Ds(1/f - 1/s)$ where s is the distance between the lens and the image detector. Based on the actual camera in our laboratory, d will be specified in units of 0.6 times the width of a pixel (e.g.. $d = 2$ means the blur circle diameter is equal to 1.2 times the width of one square pixel). Large images are divided into many smaller subimages and processed. We take the subimages to be of size 32×32. Therefore x and y will take integer values: $0, 1, 2, \cdots 31$. Based on our camera, we will assume that d takes integer values: $0, 1, 2, \cdots 99$. The focused image surface $d_v(x, y)$ which represents the 3D shape of the scene will also take integer values in the range 0 to 99.

The image data $g(x, y, d)$ can be thought of as an "image volume" in the (x, y, d) space. The FIS $d_v(x, y)$ is embedded in this volume. The value of the image volume data on the FIS gives the focused image $F(x, y)$ of the scene, i.e.

$$F(x, y) = g(x, y, d_v(x, y)) \tag{5}$$

An image volume corresponds to a small and fixed field-of-view of the camera. In each image volume (or field-of-view) FIS $d_v(x, y)$ can be approximated by a piecewise constant or a planar or a smooth curved surface. The 3D shape parameters of this surface (i.e. FIS $d_v(x, y)$) and the focused image $F(x, y)$ are obtained by processing all or some of the image volume data.

A sequence of image frames consists of cross sections of the image volume $g(x, y, d)$ taken at different values of d. Let the values of d where the cross sections are taken be $d_1, d_2, \cdots d_n$.

IFA: Application of IFA to the image sequence consists of computing a focus measure in small image regions, say 8×8, for all images in the sequence and finding the value $d = d_i$ for which the focus measure is a maximum. The FIS in the 8×8 region is given by $d_v(x, y) = d_i$ and the focused image in the 8×8 region is given by $F(x, y) = g(x, y, d_i)$.

Combining IFA and Interpolation: The solution provided by IFA can be improved through interpolation by fitting smooth curves. For example, let $M(d)$ be the focus measure function of the 8×8 image region around the position $d = d_i$ estimated by IFA described above. Then a smooth curve can be fitted to the points $M(d_{i-1})$, $M(d_i)$, and $M(d_{i+1})$, and the position $d = d_f$ where $M(d)$ is a maximum can be obtained [15]. Once we find $d_v(x, y) = d_f$, the focused image $F(x, y) = g(x, y, d_f)$ can be obtained by interpolating the points $g(x, y, d_{i-1})$, $g(x, y, d_i)$, and $g(x, y, d_{i+1})$ and perhaps other nearby points.

IDA: In contrast to IFA which requires many image frames, IDA can be applied to only two image frames, say for $d = d_j$ and $d = d_k$. It can provide an estimate of $d_v(x, y)$ and the focused image $F(x, y)$.

Combining IFA and IDA: Given a sequence of images as above, IFA and IDA can be combined as follows. First IFA is applied as above and the position $d = d_i$ where the 8×8 image region is in best focus is found. This will be $g(x, y, d_i)$. Also the second best focused image close to d_i is found (at d_{i-1} or d_{i+1}) based on the computed focus measures. Let this be $g(x, y, d_{i+1})$. Then

IDA is applied to these two most focused images to get a solution for shape and focused image. If the distance measured along the d dimension between successive image frames in an image sequence is high (about 5 or more), then combining IFA and IDA is better than combining IFA and interpolation. IDA gives better results than interpolation when the image frames are far apart.

The solutions for shape and focused image in 8×8 image regions are synthesized to obtain a solution for the entire 32×32 image. The solution thus obtained can be taken as initial solution in UFDA and it is further improved through error minimization as discussed in the previous section.

If n image frames are given, we assume that these frames are roughly uniformly placed along the d dimension and cover the entire range of values taken by the FIS $d_v(x, y)$. An algorithm for obtaining the initial solution can be summarized as follows.

If the distance between image frames is only one along the d dimension, then use IFA as it will give very good results. If the inter-frame distance is about 2 to 4, then IFA combined with interpolation should be used to get good results. If the inter-frame distance is 5 or more, then the method that combines IFA and IDA described earlier should be used. In all cases, if the number of image frames given is only two–the minimum required, then only IDA can be applied.

If the images can be recorded dynamically, then first only two image frames are recorded far apart (say at $d = 20$ and $d = 50$ for our camera specified earlier). Then IDA is applied to find a rough depth map. Based on this the approximate minimum d_{min} and maximum d_{max} of the FIS $d_v(x, y)$ are found. After this, further image frames are recorded for d only in the range d_{min} to d_{max}. This avoids recording and processing of unnecessary image frames.

In the error minimization step, in each image region, it is sufficient to consider a few image frames (about 5) around the most focused image frame in that image region.

In the algorithm outlined above, the problem of estimating the focused image given an estimation of 3D shape is a very difficult one when the number of available image frames is limited. The focused image will have to be estimated from a few image frames which are in best focus through deconvolution. If a piecewise constant approach is used for 3D shape, then a Fourier domain deconvolution can be applied (e.g. Wiener Filter). In other cases a spatial domain method will have to be used. We have found that the STM method based on IDA in [10] gives good results. Even then, deconvolution in the presence of quantization and noise is found to be a very difficult problem but satisfactory results can still be obtained.

4 Experiments

We present here the results of two experiments where only a small number of image frames (3 and 5 respectively) are used.

4.1 Planar Object

We generated simulated image data for a CCD camera with the the same parameters specified earlier: focal length f=35 mm, aperture diameter D=9 mm, square pixel width p=0.013 mm, and unit of image frame distance 0.6 pixel (i.e. d=1 corresponds to 0.6 pixel). In this experiment, an image volume of 32×32×32 was considered and three defocused image frames at position $d = 5, 16, 27$ were synthesized using Eq. (3). The FIS of the scene was a planar object given by $d_v(x, y) = 13$ and the the focused image is a checker board pattern (see Figs.(1,4)). According to the proposed computational algorithm, an IFA method is first applied to these three images. Focus measures were computed over the entire 32×32 image frames without dividing them into smaller subimages. The two best focused image frames were determined based on the focus measures. These two frames were used in an IDA method (STM proposed in [8]) to find an initial solution for distance of the planar object and the focused image. This STM method uses a blur parameter σ to obtain 3D depth map and focused image. This parameter is related to the blur circle diameter d of point light sources by [6,13] $\sigma = d/(2\sqrt{2})$. The IDA method (STM) was implemented exactly as described in [8]. This involves constructing a histogram of σ obtained at each pixel in the 32×32 image and taking the mode of the histogram as the solution for FIS $d_v(x, y)$.

After estimating the FIS as above, the focused image was estimated using three different methods to investigate and evaluate their relative performance. The first was based on a spatial domain deconvolution formula (inverse S transform) derived under some weak assumptions in [8]:

$$F(x, y) = g(x, y) - \frac{\sigma^2}{4} \nabla^2 g(x, y) \tag{6}$$

where $g(x, y)$ is the least defocused image available. The second method was based on Wiener filter in the Fourier domain as described in [10]. The third was using cubic spline interpolation [?] method on the three given image frames.

The solution obtained for FIS and focused image (from each of the three methods) was used as initial solution in UFDA based on a special case of gradient descent approach presented in [11]. It is a sequential parameter search (SPS) method with only one parameter– the distance of the planar object. The estimated solutions and the 3D PSF were used to compute the estimated image data $g_e(x, y, d)$ using Eq. (3). Three images frames were estimated corresponding to the three frames of observed data $g_o(x, y, d)$. The error E between the observed image data $g_o(x, y, d)$ and the estimated image data $g_e(x, y, d)$ is computed using Eq. (4). The error was minimized iteratively using the gradient descent method for estimating FIS and the three different methods of focused image estimation.

The initial solution from the IFA method are shown in Figs.(2,4). The position of the planar object is estimated to be at position 16 and the focused image is the image data at this position. The initial solution from the IFA method followed by the IDA (STM) method are shown in Figs.(4,5). The estimated position of the planar object is 14 and the focused image is obtained from Eq. (6). The reconstructed 3D shape from the UFDA optimization with STM, Fourier method, and cubic interpolation are positions 13, 14, 14 respectively (Fig. 4). The focused image obtained using SPS and STM is shown in Fig. 3. The results from these three methods are also presented in terms of the percentage error in gray level per pixel between observed image data and estimated image data in Table 1. These results show that the STM method is better than the other two. This can be explained as follows. The checker board test image that we used will introduce error into the Fourier method because of the periodicity properity of the discrete Fourier transform. As for the cubic spline interpolation method, since there are only three observed images that are far apart, interpolation gives very poor results as expected.

4.2 Spherical Object

The second experiment is for an object whose FIS has a spherical shape shown in Fig. 6. In a $32\times32\times32$ image volume similar to the first experiment, the FIS has a radius of 24 (Fig. 6). The focused image is the same checker board pattern shown in Fig. 1. Five image frames at positions $d = 4, 10, 16, 22, 28$ are used. These images are processed first by an IFA method followed by an IDA method (STM) as before for the planar object. In this case, due to the curved shape of the object, the 32×32 image frames are divided into 4×4 subimages and processed separately. A combination of IFA and IDA are applied separately in 4×4 image regions separately. In applying IDA (STM), the smoothing and differentiation filters proposed by Meer and Weiss[**?**] were chosen to be of size 3×3. One estimate of the blur parameter σ (and hence $d_v(x, y)$) is obtained at each pixel by integrating over the 3×3 region around the pixel.

This initial solution is used in the UFDA optimization method based on a regularization approach proposed in [12]. The five observed image frames and the corresponding five estimated image frames are used in computing the error measure. The error was minimized iteratively subject to the smoothness of the FIS as in [12]. In each iteration, after a solution for FIS was obtained, the solution for the focused image was estimated by two different approaches. One was the STM method described earlier and the other was the cubic spline interpolation method.

The initial solution from an IFA method and an IFA followed by an IDA method are shown in Figs.(7,8,10,11). In these figures, we see that the focused image from STM is much better than that from a traditional IFA method but the 3D shape has only limited improvement. This is because the IFA and IDA methods operate on very small image regions (4×4) and STM is very sensitive to noise if the window size is too small. The solution for the regularization method after five iterations is shown in Figs.(9,12). The same results are presented in terms of percentage

Table 1: Error percentage of gray level per pixel for planar object

IFA	IFA and IDA	SPS + STM	SPS + Interpolation	SPS + Fourier method
13.8%	7.6%	6.6%	8.8%	12.1%

Table 2: Error percentage of gray level per pixel for Sphere object

IFA	IFA and IDA	Regu. + STM (5 iter.)	Regu. + Interpolation (5 iter.)
7.3%	6.7%	5.3%	11.5%

gray level error in Table 2. The results for the interpolation method were poor and and therefore are not included here. The improvement of initial solution by UFDA optimization based on regularization is limited. The main reason for this is the difficulty in obtaining good estimates of the focused image for curved objects which result in very poor initial solutions for FIS as in our experiment.

5 Conclusion

The theory of UFDA was presented by us earlier [11,12] and now we have presented a general computational algorithm for UFDA. Thus our work so far on UFDA provides a reasonably complete theoretical and implementation framework for application of UFDA in practical machine vision. The algorithm presented here deals with variable number of input images in an optimal fashion to improve accuracy and computational efficiency. The algorithm is useful in investigating the trade-off between (i) the number of image frames used, (ii) the accuracy of results, and (iii) the amount of computation used. Experimental results show that estimating focused images from a small number of defocused images (the "deconvolution" problem) is difficult. Further research is needed to obtain a good solution to this problem. This will facilitate further improvements in the performance of UFDA. However, if the number of available images is not few but many, then UFDA provides good results useful in practical applications.

6 References

[1] J. Enns and P. Lawrence, "A Matrix Based Method for Determining Depth from Focus," *Proceedings of the IEEE Computer Society Conference on CVPR*, June 1991.

[2] E. Krotkov, "Focusing," *International Journal of Computer Vision*, 1, 223-237, 1987.

[3] S. K. Nayar and Y. Nakagawa, "Shape from Focus: An Effective Approach for Rough Surfaces," *IEEE Trans. on PAMI*, 16 (8):824-831, Aug. 1994.

[4] T. Poggio, V. Torre, C. Koch " Computational vision and regularization theory", Nature,

Vol.317,No.6053, pp314-319, Sep 1985

[5] M. Subbarao, and T. Wei, "Depth from Defocus and Rapid Autofocusing : A practical Approach", *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, Champaign, Illinois, June 1992, pp. 773-776.

[6] M. Subbarao, "Efficient Depth Recovery through Inverse Optics", Editor: H. Freeman, *Machine Vision for Inspection and Measurement,* Academic press, Boston, pp. 101-126, 1989.

[7] M. Subbarao and M. C. Lu, "Computer Modeling and Simulation of Camera Defocus," *Machine Vision and Applications*, (1994) 7, pp. 277-289.

[8] M. Subbarao and G. Surya, "Depth from Defocus: A Spatial Domain Approach," *International Journal of Computer Vision*, 13, 3, pp. 271-294 (1994).

[9] M. Subbarao and T. S. Choi, "Accurate Recovery of Three-Dimensional Shape from Image Focus," *IEEE Transactions on PAMI,* March 1995, pp. 266-274.

[10] M. Subbarao, T. Wei, and G. Surya, "Focused Image Recovery from Two Defocused Images Recorded with Different Camera Settings", *IEEE Transactions on Image Processing*, Vol. 4, No. 12, 16 pages, Dec. 1995.

[11] M. Subbarao and Y.F. Liu, "Accurate Reconstruction of Three-dimensional Shape and Focused Image from a Sequence of Noisy Defocused Images," SPIE Vol. 2909 pp. 178-191, Boston Mass., Nov. 1996

[12] M. Subbarao and Y.F. Liu, "Analysis of Defocused Image Data for 3D Shape Recovery using a Regularization Technique," SPIE Vol. 3204, ISAM'97, Pittsburgh, Oct. 1997

[13] G. Surya, *Three-Dimensional Scene Recovery from Image Defocus* , Ph.D. Thesis, Dept. of Electrical Engineering, State University of New York, Stony Brook, Dec. 1994.

[14] Tikhonov, A.N. , Arsenin, V.Y. *Solutions of ill-posed Problems*, Winston, Washington, DC , 1977

[15] J. K. Tyan, *Analysis ofImage Focus and Defocus Analysis.......* , Ph.D. Thesis, Dept. of Electrical Engineering, State University of New York, Stony Brook, Dec. 1997.

[16] T. Wei, *Three-Dimensional Machine Vision using Image Defocus* , Ph.D. Thesis, Dept. of Electrical Engineering, State University of New York, Stony Brook, Dec. 1994.

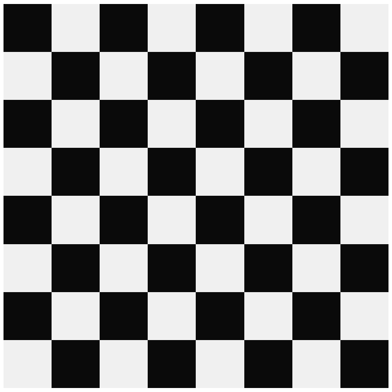

Figure 1: Original Focused Image

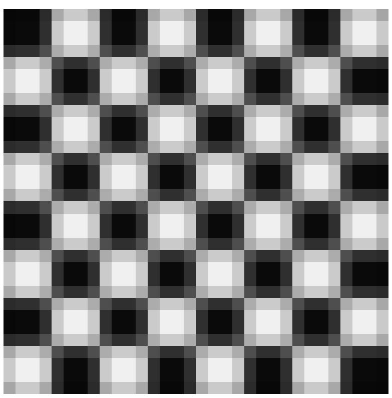

Figure 2: Focused Image from an IFA method

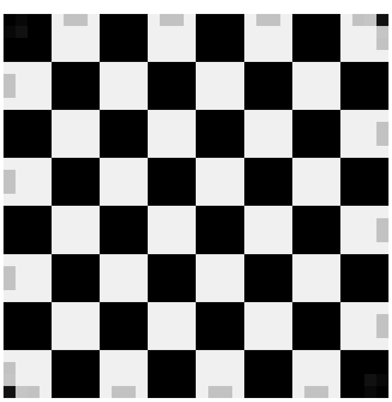

Figure 3: Focused Image by SPS with STM

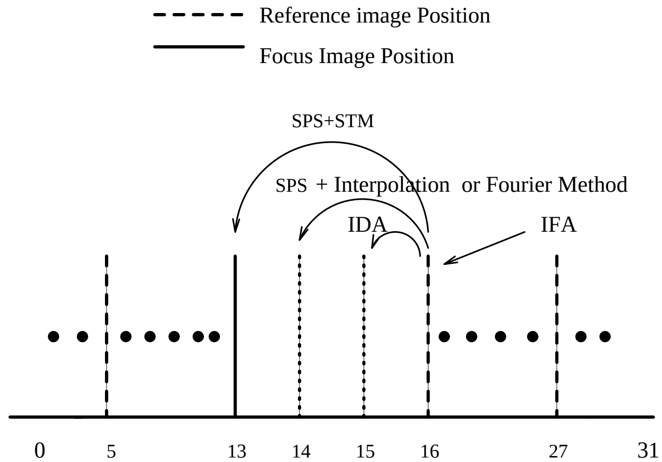

Figure 4: Focused position form IFA, IDA and SPS

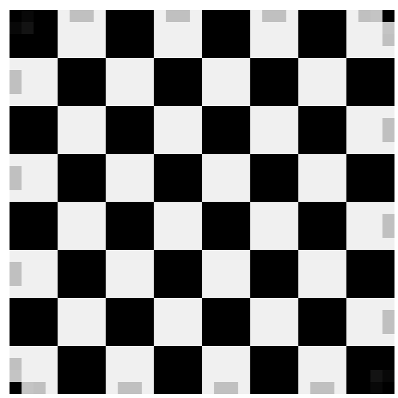

Figure 5: Focused Image by an STM method

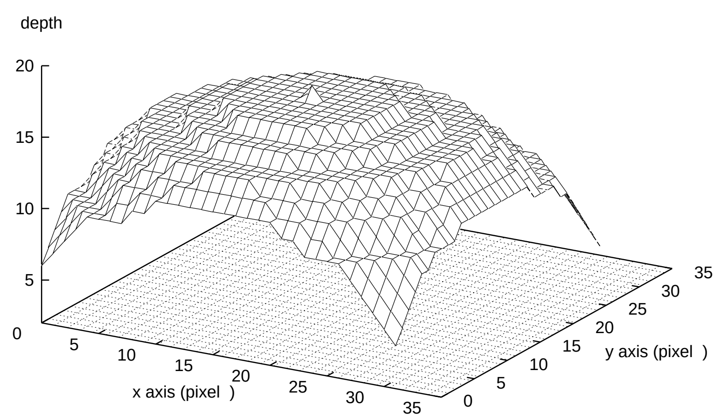

Figure 6: Original FIS of a 32x32 Hemisphere

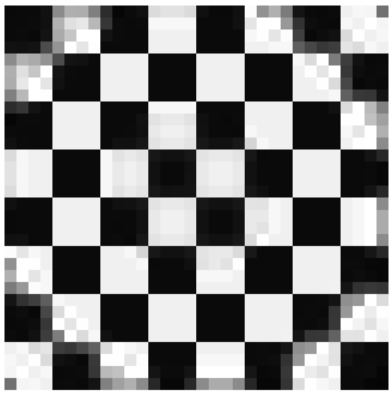

Figure 7: Focused Image from an IFA method

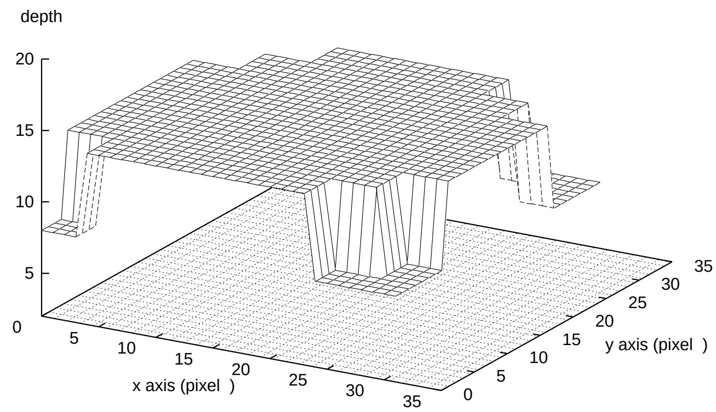

Figure 10: FIS by an IFA method

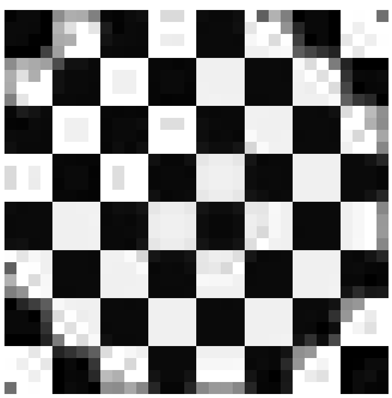

Figure 8: Focused Image from an IDA method (STM)

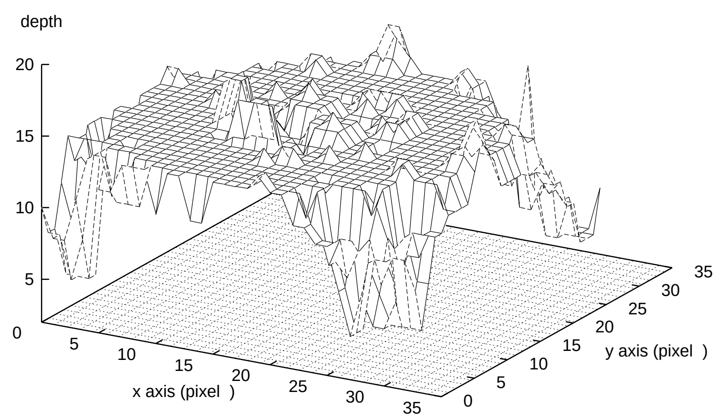

Figure 11: FIS by an IDA method (STM)

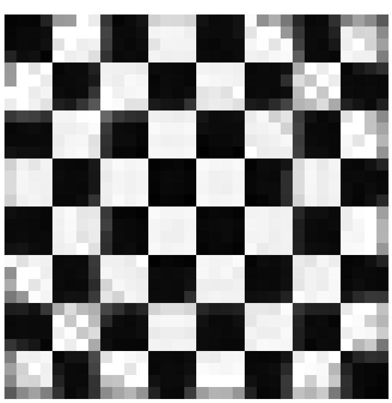

Figure 9: Focused Image by regularization(5 iters.)

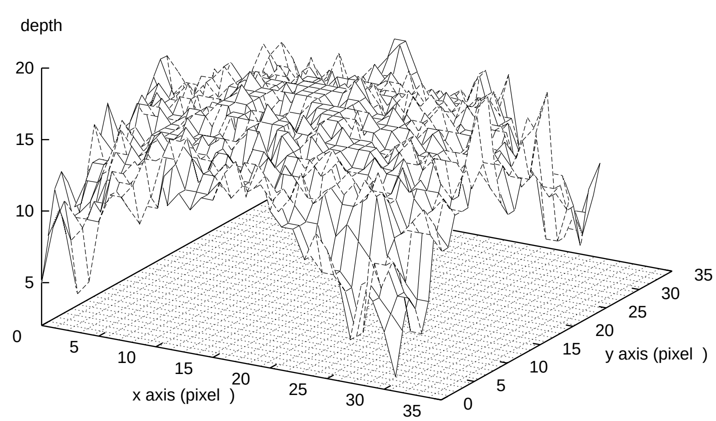

Figure 12: FIS by regularization(5 iters)