HIERARCHICAL ADAPTIVE REGULARISATION METHOD FOR DEPTH EXTRACTION FROM PLANAR RECORDING OF 3D-INTEGRAL IMAGES

Silvia Manolache, Malcolm McCormick*

De Montfort University
Department of Engineering and Technology
The Gateway, Leicester LE1 9BH, UK

S.-Y. Kung

Princeton University
Department of Electrical Engineering
Princeton, NJ-08544, USA

ABSTRACT

The paper presents a novel algorithm for object space reconstruction from the planar (2D) recorded data set of a 3D-integral image. The integral imaging system is described and the associated point spread function is given. The space data extraction is formulated as an inverse problem, which proves ill-conditioned, and tackled by using a hierarchical multiresolution strategy and imposing additional conditions to the sought solution. The hierarchisation strategy and the two-phase adaptive constrained 3D-reconstruction algorithm based on the use of two sigmoid functions are presented. Finally, illustrative simulation results are given.

1. INTRODUCTION

The development of 3D-imaging systems has been a constant pursuit of the scientific as well as of the entertainment community in the new technological era. Integral photography was pioneered by Lippmann ([1], 1908), who used microlens arrays to create, record on photographic film, and replay integral three dimensional images. Since 1908, when it was first reported, the integral photographic technique has been improved as a result of theoretical studies, technical innovations of the optical systems, and progress in microlens manufacturing. Integral imaging resembles to holography, but it uses natural light and reproduces true colour optical models. Hence, it offers a viable alternative to other autostereoscopic systems ([2]).

The optics of an advanced form of integral imaging system – employing a two tier optical network – in which a true 3D optical reconstruction of a scene is transferred to the capture device, has been described in detail by Davies and McCormick ([3], [4]).

Extracting depth information from 3D-integral images has various applications, which range from remote inspection in robotic vision and medical imaging to combining real and computer generated 3D-pictures in a virtual studio. The aim of the present paper is to describe a method of reconstructing the composition of a 3D-object space from 2D-recorded data encoding the scene. The algorithm is applicable both to integral imaging and holography. The current work has dwelt upon the application to integral imaging. More specifically, the depth estimation from 3D-integral pictures is formulated as an inverse problem: given the image, i.e. the recorded 'effect' of the object space, find the 'cause' which had produced it, i.e. recover the intensity contours and, consequently, the composition and the depth of the object scene. The direct problem - image formation and recording - has been studied in a previous paper, and the equation of the point spread function for the lenticular integral imaging system has been derived ([5]). Inverse problems in imaging are ill-posed and their discrete correspondents are ill-conditioned due to the inherent loss of information associated with the direct process ([6]). In order to cure the ill-posedness of the problem, approximate solutions satisfying additional constraints coming from the physics of the problem are searched. The present work comprises an adaptive regularisation scheme for obtaining a constrained least squares solution of the depth extraction problem, which is hierarchically applied in order to obtain high resolution object space reconstruction.

2. 3D-INTEGRAL IMAGING SYSTEM AND ASSOCIATED POINT SPREAD FUNCTION

The optical arrangement of the 3D-integral imaging system, shown in Figure 1, comprises two macrolens arrays (MA1 and MA2) placed equidistantly behind and in front of an autocollimating transmission screen (ATS). The ATS is made up of two microlens arrays separated by their joint focal distance. The recording plane, a photographic plate whose position coincides with the focal plane of a microlens array (RA), lies within the optical model.

The paper considers a camera configuration where the ATS contains square based hemispherical lenslets and the recording array is made up of identical semicylindrical microlenses. This system is known as a *3D-lenticular integral imaging system* ([3], [5]) and produces 3D-images containing horizontal parallax.

The object is imaged by the input macrolenses which transmit compressed transposed images that occur at or near the central double microlens screen (ATS). The screen inverts the optical sense of each intermediary image (see Figure 2), and, simultaneously, presents these spatially reversed 3D-optical models to the corresponding output macrolenses. The output macrolenses retranspose the optical models to the correct spatial location. The final integrated image, formed by superposed optical models projected by the second macrolens array, is a true 3D optical 1:1 recon-

*The first author performed the work while at Princeton University under the sponsorship of the Exchange Visitor Program P1 180 between Princeton University and De Montfort University.

struction of the object. It is recorded as a *2D-sampled data set* possessing either horizontal parallax (if semicylindrical microlenses are used in the recording array), or continuous parallax (in the case of the spherical or square based microlens arrangement).

The integral image recorded in the focal plane of a recording microlens array as a *planar* sampled data set contains all the 3D information related to the object space. Each microlens of the recording array samples a fractional part of the scene, many microlenses recording directional information of the scene from different viewing angles. Therefore, parallax information about any particular point is spread over the recording plane. Redisplay of the full spatial model as a real 3D image can be effected by overlaying the sampled data set by an integral decoding element.

The spread function of the entire optical process combines the spread effect of both the central double microlens screen and the recording array ([5]):

$$K(X,Y,x,y,z) = \sum_{l(x,y,z)} K_{l(x,y,z)}\,, \tag{1}$$

where (X,Y) are the coordinates in the image space, (x,y,z) represent the object space, and the summation is done with respect to the index of the microlenses which 'see' the point (x,y,z). In the above equation, $K_{l(x,y,z)}$ is the point spread function component behind one microlens of the recording screen:

$$K_{l(x,y,z)} = \alpha(z)\exp\left(-\frac{(X-X_l)^2}{a_1^2}\right)\exp\left(-\frac{(Y-Y_l)^2}{a_2^2}\right) \tag{2}$$

The point (X_l, Y_l) represents the point of maximum intensity in the microimage formed behind microlens l and:

$$\begin{aligned} \alpha(z) &= \frac{v^4 w_1^2 w_2^2}{\sqrt{(v^2+w_1^2)(v^2+w_2^2)}} \\ a_1^2 &= v^2 + w_1^2\,, \qquad a_2^2 = v^2 + w_2^2 \end{aligned} \tag{3}$$

depend explicitly on lenslet parameters and point depth z ([5]).

Equation (2) indicates that the spread function is space variant and that it holds depth information in all the factors. This fact will be used in recovering three dimensional data about the object space. Due to the small values of the parameters involved in the point spread function formulae, only the significant factors should be considered in a computational approach of the depth extraction problem. When using lenticular arrays (i.e. semicylindrical microlenses), the spread on the x-dimension decreases rapidly to 0, so the x-factors in the point spread expression can be dropped.

3. FORMULATION OF THE OBJECT SPACE RECONSTRUCTION AS AN INVERSE PROBLEM

The image formation equation for the integral imaging system is:

$$g(X,Y) = \int I(x,y,z) \sum_{l(x,y,z)} K_l(X,Y,x,y,z)\,dx\,dy\,dz\,, \tag{4}$$

where $I(x,y,z)$ is the intensity at point (x,y,z).

It is noticed that the integral image formation process is not a convolution process because the critical shift invariance property does not hold. Therefore classical deconvolution methods cannot be used to tackle the problem of scene reconstruction from an integral image.

However, the point spread function provides a linear image formation operator A such that equation (4) be written as:

$$g = A\,I\,. \tag{5}$$

The reconstruction of the intensity distribution in the object is the inverse process of the image formation given by equation (5). The discrete variant of the problem has been treated considering various samplings of the object space. Numerical experiments have shown that the matrix associated with the point spread function is very ill-conditioned. Alternative methods of reconstructing the object space are thus necessary and one of them is presented in the next section.

4. HIERARCHICAL ADAPTIVE CONSTRAINED 3D-RECONSTRUCTION ALGORITHM

The ill-conditioning of the problem is determined by the existence of very small singular values of the operator A, which virtually increase the dimension of the null-space of A in numerical applications and make the solutions of the equation (5) be very unstable. The number of small singular values can be decreased by reducing the size of the matrix corresponding to A, so by considering only a low number of sampling points in the object space (lower than the number of pixels in the image). This operation has the drawback of inducing low resolution in the object reconstruction.

4.1. Hierarchical multiresolution strategy

The conflict between ill-conditioning of the imaging operator and high resolution reconstruction requirement has been solved by adopting a *hierarchical multiresolution strategy* based on zooming in the high interest regions. This approach consists of the following steps:

1. Coarse sampling: Sample a large domain $D^{(1)}$ using a coarse grid and obtain a low resolution reconstruction $\hat{I}_{D^{(1)}}$ from the equation $g = A_{D^{(1)}}\hat{I}_{D^{(1)}}$.

2. Fine sampling: Select the high interest regions D_1,.., D_m of the object space from the low resolution reconstruction $\hat{I}_{D^{(1)}}$ and consider the rest of the space as determined. Sub-sample the union $D^{(2)} = D_1 \cup \ldots \cup D_m$ and obtain a finer resolution reconstruction $\hat{I}_{D^{(2)}}$ from the equation $g = A_{D^{(2)}}\hat{I}_{D^{(2)}}$.

If necessary, the procedure can be recursively reapplied until the object space is reconstructed at the desired resolution.

The algorithm for obtaining a reliable solution for each sampled domain is given in the next subsection.

4.2. Adaptive constrained 3D-reconstruction algorithm

Choosing a low number of sampling points does not solve entirely the ill-conditioning problem, as the linear system (5) is still large, so unstable. Therefore, additional constraints coming from a priori knowledge about the object space have to be imposed upon the sought solution $\hat{I}$. More

precisely, the solution $\hat{I}$ has to be such that $A\,\hat{I}$ is as close as possible to the image g, i.e.:

$$||A\,\hat{I} - g|| = minimum\,. \tag{6}$$

Also, a condition of positivity and bounding is necessary, as there are neither negative nor infinite intensity objects:

$$\hat{I} \in [0, T]\,. \tag{7}$$

$\hat{I}$ satisfying (7) can be written as the thresholded value of arbitrary data u: $\hat{I} = f(u)$, where f is a thresholding function. The projection of the interval $[0, T]$ onto R is not differentiable, thus a sigmoid approximation of it, $f_\sigma : R^n \to [0, T]^n$, has been preferred:

$$f_\sigma(u_i) = \frac{T}{1 + exp\left(-\left(\frac{u_i}{\sigma} - k\right)\right)}\,,\ i = 1 \ldots n\,, \tag{8}$$

where n is the number of components of the vector $\hat{I}$, i.e. the number of sampling points chosen in the object space. σ is an arbitrary number which controls the slope of the thresholding curve and k is a translation term ([7]).

To sum up, the sought vector $\hat{I}$ has the form $\hat{I} = f_\sigma(u)$ and minimises the discrepancy functional $\epsilon = ||A\,I - g||$.

The gradient of ϵ with respect to σ is:

$$\nabla_\sigma \epsilon = (A\,f_\sigma(u) - g)^T\,A\;diag(f'(u_i))_{i=1\ldots n}\,. \tag{9}$$

The variation of u which leads to a constrained least-squares solution is on the decrease direction of the gradient $\nabla_\sigma \epsilon$:

$$\Delta u = -\eta \nabla_\sigma \epsilon\,, \eta > 0\,. \tag{10}$$

The algorithm starts with an arbitrary value for u, e.g. $u = 0$, and iteratively modifies u with a quantity Δu computed as in the above equation.

Choosing $\sigma = T/k$, the function f_σ becomes a quasi-linear thresholding function. Its use in simulations has proved inefficient. Instead, a two-phase scheme has been preferred:

1. Use a relatively small value of σ, e.g. $\sigma = T/(12k)$, and perform a number of iterations.

2. Use a larger σ, e.g. $\sigma = T/(6k)$, until convergence.

The first phase of the algorithm performs a polarisation of the reconstructed intensity values, whilst the second phase removes arbitrary equalisations of the reconstructed values. The final regularised solution $f_\sigma(u)$ is thresholded and only the highest values are kept. This is the resulting object space reconstruction output by the current algorithm.

5. SIMULATION RESULTS

The hierarchical adaptive constrained 3D-reconstruction algorithm has been tested on a set of computer generated images. These images contain either a single face with various inclination angles with respect to the Oxy-plane, or two faces having equal or different inclination angles, or a cube. For the simplicity of the presentation, planar sections parallel to the Oyz-plane of the object space reconstruction are shown. The objects considered are all perpendicular to the Oyz-plane, so their intersections with $x = x_0$ planes are similar. Hence, recovering one section means recovering the whole object. Figures 3-5 contain both the object reconstruction and the original object for comparison. The darker gray shades represent higher levels of intensity in these representations.

Figure 3 depicts the object reconstruction from an image containing a single transparent thin face, 0.4 mm wide, 8 mm long, centred at the point (y=0.975 m, z=0.459 m), perpendicular to y-axis using the adaptive constrained reconstruction algorithm. The position and dimension of the object are correctly recovered even at low resolution.

The processing of the image of the visible sides of an opaque cube yielded the scene depicted in Figure 4(a). The object recovered at the coarse sampling stage is symmetric and correctly positioned in space, and the two sides are clearly separated. The region where the cube was detected, marked by the dashed line in Figure 4(a), has been further sampled and processed. The result is shown in Figure 4(b) and provides a more accurately contoured object.

The hierarchical reconstruction of a two object scene is presented in Figure 5. The two objects are two transparent faces of unequal lengths perpendicular to each other. They are correctly positioned and separated in the reconstruction, though their shapes are slightly perturbed. A further sampling and processing is performed for the marked regions around the two objects and more accurate reconstructions are obtained (Figures 5(b)-(c)).

The adaptive constrained reconstruction algorithm has converged in 40 iterations for the single object picture and in about 150-200 iterations for the two face object and the cube at each hierarchical stage.

6. CONCLUSIONS

The present paper has approached the object reconstruction and depth extraction from 3D-integral images as an inverse problem, which proves to be ill-conditioned. In order to cure the ill-conditioning of the problem, a hierarchical adaptive constrained 3D-reconstruction algorithm has been considered. It is based on the use of two sigmoid functions to determine a bounded constrained least squares solution which provides information about the number of objects in the scene, their shape and absolute and relative position. The scheme can be hierarchically reapplied to the relevant regions of the object space in order to produce higher resolution reconstructions. The simulation results have shown that the number of objects, absolute and relative positions and shape are recovered accurately. The hierarchical scheme produces high resolution reconstructions and is computationally efficient, as only the relevant regions of the scene are oversampled.

7. REFERENCES

[1] G. Lippmann, "La Photographie integrale", Comtes Rendus, Academie des Sciences, vol.146, pp. 446-451, 1908.

[2] T. Okoshi, "Three dimensional imaging techniques", Academic Press, London, 1976.

[3] N. Davies, M. McCormick, "Holoscopic imaging with true 3D-content in full natural colour", J.Phot.Science, vol. 40, pp. 46-49, 1992.

[4] N. Davies, M. McCormick, M. Brewin, "Design and analysis of an image transfer system using microlens arrays", Opt. Eng., vol. 33-11, pp. 3624-3633, 1994.

[5] S. Manolache, A. Aggoun, M. McCormick, N. Davies, "A mathematical model of a 3D-lenticular integral recording system", Proceedings of IEEE Vision, Modeling and Visualization, pp. 51-58, Erlangen, 1999.

[6] M. Bertero, P. Boccacci, "Introduction to inverse problems in imaging", Institute of Physics Publishing, Bristol and Philadelphia, 1998.

[7] S. Manolache, M. McCormick, S.-Y. Kung, "3D-object space reconstruction from planar recorded data of 3D-integral images", submitted for publication in Journal of VLSI Signal Processing Systems (Kluwer).

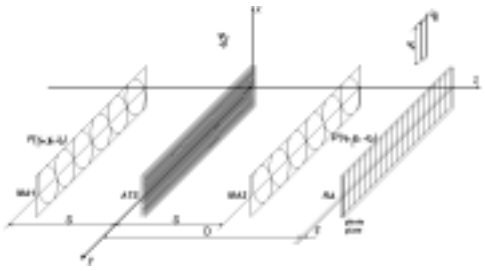

Fig. 1. 3D-lenticular integral imaging camera system

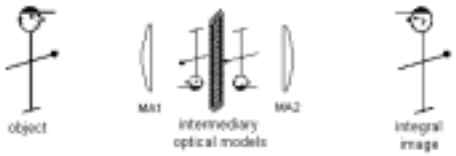

Fig. 2. Optical stages in the integral imaging

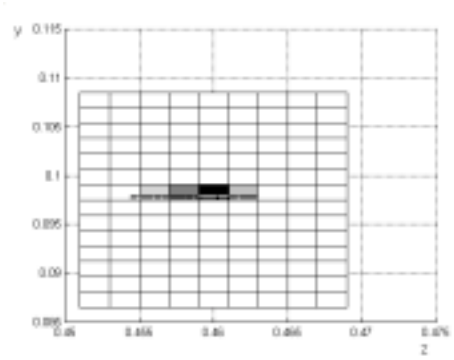

Fig. 3. Transversal section through the reconstruction of a thin face orthogonal to the recording screen. The original object is drawn in continuous line.

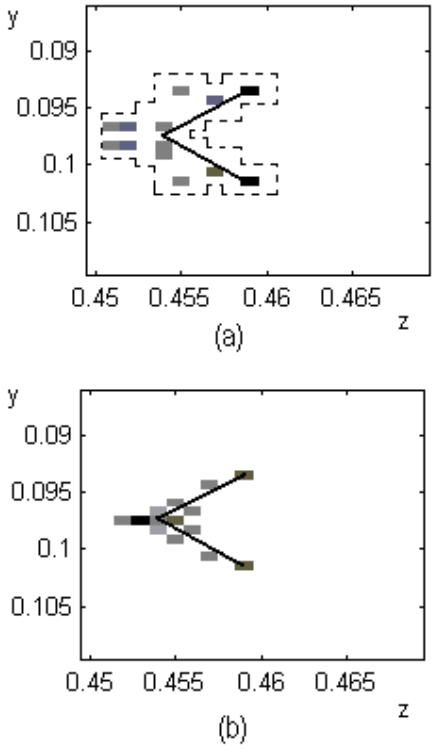

Fig. 4. Transversal sections through the hierarchical reconstructions of a cube. The original object is drawn in continuous line.

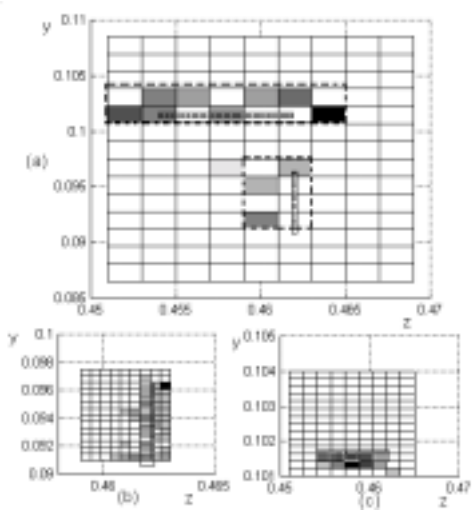

Fig. 5. Transversal sections through the hierarchical reconstructions of a pair of thin faces, one orthogonal and another parallel to the recording screen. The original objects are drawn in continuous line.