Anchoring Graph Cuts towards Accurate Depth Estimation in Integral Images

Dimitrios Zarpalas, Eleni Fotiadou, Iordanis Biperis, and Petros Daras, *Member, IEEE*

***Abstract*—Integral imaging is a three dimensional imaging technique that allows the displaying of full color images with continuous parallax. Its commercial potential has been increased, due to its ability of presenting to the viewers smooth 3-D images, with full parallax, in a wide viewing zone. Being able to extract the inherent 3-D information from the planar integral images and produce 3-D reconstructions, offers advantages in various applications of immersive entertainment and communications. On this scope, this paper addresses the problem of accurate depth estimation in integral images. The proposed method, relying on the assumption that a pixel is the projection of a 3-D imaging ray, aims to specify the first intersection of each pixel's projection ray with the 3-D scene in order to assign to it the corresponding depth value. This task is formulated as an energy optimization problem and the graph cuts approach is utilized to solve it. The energy term is twofold; its first part aims to restrict the desired solution to be close to the observed data, i.e the integral image, while the second one enforces regional smoothness in the depth estimation. This combination offers an accurate and spatially smooth scene structure. The novelty of the paper lies on the framework's formulation as one single optimization procedure and on the way that this optimization is constrained by a set of reliably estimated 3-D surface points, called the "anchor points". Anchoring the optimization results in enhanced depth estimation accuracy, while decreasing the optimization processing burden. The proposed algorithm is evaluated in both synthetic and real integral images consisting of complicated object scenes. A comparison against other state-of-the-art algorithms proves the superiority of the proposed method in terms of depth estimation accuracy.**

***Index Terms*—Integral imaging, holoscopic imaging, depth estimation, anchor points, self-similarity, graph cuts, energy optimization, microlenses, cylindrical lenses, unidirectional integral imaging, elemental images, viewpoint images.**

I. Introduction

THREE dimensional (3-D) imaging systems have attracted both commercial and scientific interest in different disciplines over the last few decades, with the main applications of them being in 3-D TV, computer games, virtual reality, immersive environments, etc. The goal of all those applications is to offer to the consumer rich, immersive experiences. One of the most important aspects on such research is how to produce free viewing 3-D displaying technologies.There are several types of 3-D imaging technologies (i.e. capturing, processing and displaying), which are broadly divided in two categories, the stereoscopic and the autostereoscopic. While in the past years most of the research in the area of 3-D imaging systems has concentrated on the stereoscopic technology [32], [10], [51], the fact that the viewer had to wear special headgear (e.g. stereoscopic glasses) in order to feel the 3-D effect, has limited the acceptance and the application of them. The autostereoscopic display systems are more comfortable for the viewer as they do not require the use of special glasses [9]. Holography [45], [43], volumetric displays [4], multi-view [56] and integral imaging are some types of autostereoscopic technology. Holographic technology offers full parallax in all directions but the need for coherent light sources and dark room conditions during recording, reduces its practical utility. Volumetric display systems often have large field of view but the difficulty to design them has limited their application. In recent years, many 3-D recording and display systems have focused on multi-view techniques. However, in the case of multi-view displays, the viewing effect depends heavily on the number of viewpoints, and capturing many views in real-time along with the use of low cost equipment is quite difficult.

Integral imaging, or holoscopic imaging, based on the Integral Photography concept proposed in 1908 by Gabriel Lippmann [23], recently re-attracted researcher interest, due to its desired properties. Integral imaging offers many advantages as opposed to the other existing 3-D sensing techniques as it uses natural light, can offer full parallax in real-time without the need of calibration and does not cause eye strain [14]. Therefore, it constitutes a promising technology for the production of real-time 3-D image capturing and displaying systems. In the past few years, as the microlens manufacturing is progressing, offering further flexibilities to integral imaging, much effort has been devoted in order to overcome some significant problems of integral imaging. Such shortcomings are the limited depth of field [27], [30], [7] and the low quality of the displayed images [16]. Researchers have focused on increasing the viewing angle of integral imaging [29], [17], [34] and on the generation of orthoscopic integral images [28] since until recently, the integral imaging systems provided pseudoscopic images (i.e reversed in depth). Having advanced the image generation procedure [31], integral imaging technology is about to become ready for massive commercialization in the next decades. However, in order this technology to be established, apart from the hardware issues, a number of image processing issues should be addressed to overcome the inherited restrictions of integral images. In terms of applications, probably the most important issue, is 3-D reconstruction, through depth estimation. Depth knowledge would benefit both coding and transmission of integral images, as already does in multi-view imaging [59], and further video games developing by appropriate mixing of real and synthetic

D. Zarpalas, E. Fotiadou, I. Biperis and P. Daras are with the Informatics and Telematics Institute, Centre for Research and Technology Hellas, 1st km Thermi-Panorama Rd, P.O. Box 60361, 57001, Thessaloniki, Greece (email: {zarpalas, fotiadou, iordanis, daras}@iti.gr)

integral images. Moreover, it would be really beneficial on interactive 3-D displays, since depth would be essential to build the interaction field. Further, integral imaging seems capable of providing promising applications on other fields as well, where depth information is essential, such as biometrics, medical imaging, robotic vision, etc.

The concept of integral imaging is to utilize an array of lenses, instead of just one lens, over a film sheet or an electronic image sensor. This configuration captures a special 2-D recording of the viewed scene [1]. Neighboring lenses capture overlapping regions of the scene and each one of them records an elemental image. The set of all the recorded elemental images constitutes an integral image. The integral image can be displayed by an optical device, such as an LCD, with a microlens array placed in front of it, in order to reconstruct the 3-D scene. Integral images are divided in two categories, the unidirectional and the omnidirectional ones. This subdivision is due to the shape of the lenses, which can be either cylindrical or spherical. Cylindrical lenses provide the viewers with horizontal parallax, while spherical lenses further offer vertical parallax. Re-arranging the pixels of the integral image leads to the formation of the viewpoint images, each of them being a rotated orthographic projection of the scene, which depicts the scene from a different direction (Fig. 3). 3-D information is embedded in integral images as each viewpoint image depicts a different perspective of the 3-D scene. Although integral images are 2-D, their rich information contains 3-D characteristics that can be replayed on specialized displays to reconstruct a true 3-D scene. Thus, extracting this inherent depth information, becomes both challenging and necessary for further processing of the integral images (e.g. for feature extraction, coding, etc.).

Reconstructing the geometry of a 3D scene, based on correspondences between several pictures depicting the same scene from different viewpoints, has been extensively studied in the multi-view stereo field [39]. The goal of multi-view stereo is to recover the geometry of the scene from a collection of images taken from scattered cameras. To achieve the latter, some approaches compute a cost function on a 3-D volume, and then extract the volume's outer surface [40], [47]. Other algorithms define a volumetric Markov random field and use max-flow [35], [42], [11] to extract the surface. Space carving [21], [3], [55], and surface evolving techniques [57], [15], [13], [53] have also been employed. All these approaches iteratively deform an initial volume by deleting some of its voxels, or even add some if needed, based on the formulated energy that is to be minimized. Other methods try to compute a set of depth maps [44], [58], [12], [37], [20] and produce the final result by either enforcing consistency constraints between them, or by merging them. Lately, graph cuts were employed to solve the multi-view stereo problem [20], [52], [48]. A related survey [39] opted for two graph cuts based methods to be among the best in this field. Graph cuts were originally used to establish correspondence between stereo image pairs [6], [18] , [5] while another survey [38] demonstrated that a method based on graph cuts was at the top of other existing stereo correspondence algorithms.

Despite the fact that extracting depth information from an integral image resembles the multi-view stereo problem, there are some basic distinctions that differentiate them. This is evident by the fact that, contrary to the plethora of multi-view stereo algorithms, the literature on depth extraction from integral images is quite limited. The multi-view framework is composed of multiple high resolution cameras, scattered at different locations observing the same scene. Difficulties arise from both the calibration of the cameras and the different light reflection on the scene's objects, which depends on the cameras' position. On the other hand, in integral imaging there is only one camera, which produces several very low resolution images of the scene, each capturing just a region of the scene from slightly different perspective. Establishing correspondences from close positioned cameras, requires sub-pixel accurate disparity estimation, which is not a trivial task when the given images are of very low resolution, as in the case of integral images.

Early works on depth estimation from integral images were based on the Point Spread Function (PSF) of the optical recording, which formulated the problem of depth extraction as an inverse problem; given the "effect", i.e the integral image, find the "cause", i.e the 3-D scene that produced it [26], [25]. However, inverse problems in imaging are ill-posed [2] and even though they work well on simulated data they are not applicable in real integral images. More recently, Cirstea et al. [8] in order to cure this ill-posedness, used two regularization methods able to provide realistic reconstructions. The first one was the Landweber method [22]. while the second was a constrained version of the Tikhonov's regularization method [46]. The two developed algorithms provided approximate solutions of the scene reconstruction and estimations of the depth of the scene.

Recently, Saavedra et al. [36] handled integral images in a novel and interesting way. The periodic nature of the integral images, due to multiple micro-lens capturing, is exploited through a Fourier filtering. After a filtering and a back-projection procedure, a 3-D image, i.e. a 3-D color function $I(x, y, z)$ is constructed. Image $I(x, y,)$ has sharp colors on coordinates (x, y, z), that belong onto the objects' surfaces, while it is blurry elsewhere. However, the extraction of the 3-D surface requires a further processing step; to segment the sharp voxels out of $I(x, y, z)$, which is a problem of its own.

Lately, the majority of latest works indicates a preference to depth-through-disparity approaches since the relation between depth and disparity estimation is straightforward. After viewpoint image extraction the disparity field between pairs of them is calculated. The first reported work in this direction was that of Wu et al. [49] where a number of correlation metrics was tested for the disparity estimation. A multi-baseline technique was also adopted taking advantage of the information recurrence between different viewpoint images. In their subsequent work [50], the authors tried to improve the accuracy of the algorithm by taking into consideration that the depth is piecewise continuous in the space, thus an additional neighborhood constraint was included. Experimental validation on both synthetic and real integral images showed the efficiency of the modified algorithm. Park et al. [33] worked along similar lines and proposed the use of a lens

array consisting of vertically long rectangular lens elements. Following the viewpoint image formation, they applied a modified correlation based technique that reduces the depth's quantization error and achieves accurate depth estimation in the case of scenes with extremely periodically patterned objects.

Another way of facing the problem is by first extracting and matching a set of feature points and then trying to fit a surface to the reconstructed features that aims to optimally connect them. On this basis, in the author's previous work [54], two depth extraction methods were presented that both used an extracted set of "strong" correspondences. Then, in order to fit a surface on them, the first method uses the 3-D integral imaging grid and tries to find surface points as a subset of the intersections between the pixels' projection rays. The disadvantages of the above method is that it produces non-smooth solutions, since it samples the 3-D scene on the imaging grid's intersections, which is a non-uniform sampling of the 3-D space, and that it neglects the piecewise nature of depth. The second method, motivated by the depth-through-disparity concept, decomposed the problem to multiple stereo problems, each composed of a pair of consecutive viewpoint images. Graph cuts were employed to fit a surface on the features set, by estimating the disparities between pixels of the given pair of viewpoint images. By doing so, multiple optimizations (i.e. one for every pair) need to be solved. In the end, all depth-maps had to be merged to produce a final depth-map. The difference among the two above techniques is that the first one treats the depth estimation problem as one 3-D optimization problem, while the second one as a merging of multiple stereo-like problems.

After thorough analysis of the advantages and disadvantages of the two methods in [54], the proposed framework was derived, which takes the advantages of both previous methods while allows for solving the disadvantages of both. As such, it proposes one surface fitting optimization that is implemented through graph cuts [6], which is constrained by a set of pre-extracted reliable features, called "anchor points". Anchor points serve for constraining the optimization procedure, which otherwise can easily get stuck to local extrema, due to the high complexity of the optimization. Anchoring the optimization results in enhanced estimation accuracy along with reduction in the optimization complexity. In the proposed formulation, graph cuts are trying to find the optimized solution of the depth of the 3-D points that correspond to each pixel, instead of the disparities between pixels of adjacent viewpoint images. The 3-D scene is scanned uniformly along the emanating projection rays of the integral image, instead of the non-uniform ray correspondences on the imaging grid. Furthermore, the proposed framework enables the modelling of the piecewise nature of depth by introducing a twofold regularization term among adjacent pixels on both the elemental and viewpoint images, contrary to none regularization term of the first method in [54], while for the second one, regularization could only be applied on viewpoint images. This twofold neighborhood handling leads to reconstructed scenes with high spatial smoothness. Moreover, measuring surface fitness in the optimization procedure utilizing the self-similarity descriptor [41], instead of a simple correlation metric, imposed valuable photo-consistency in the optimization procedure.

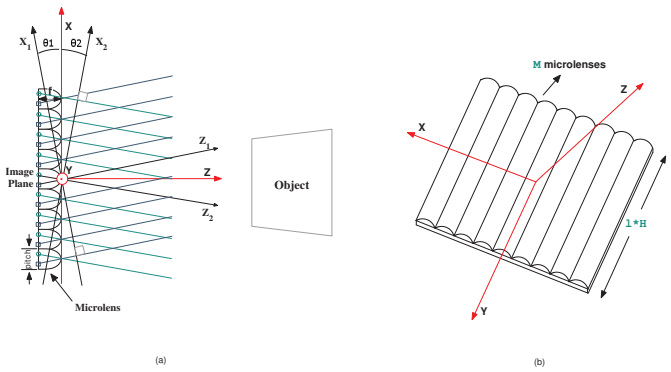

Fig. 1. (a) Lens set-up for integral image generation using a cylindrical microlens array and the local coordinate systems associated with viewpoint images. (b) The local coordinate system of the central viewpoint (l denotes the pixel pitch and H the vertical pixel resolution of the microlens).

In order to evaluate the performance of the proposed algorithm a database of 28 synthetic integral images depicting 3-D scenes was constructed and a comparison with other state-of-the-art algorithms [50], [54] is presented. Additional experiments were performed on a small set of real integral images.

The rest of the paper is organized as follows. Section II provides a brief description of integral imaging concepts, the extraction of the viewpoint images and the derivation of the projection equation. Section III presents the proposed algorithm for depth estimation. Experimental results and evaluation of the proposed method are presented in Section IV, where comparative results against [50] and [54] are also provided. Finally, conclusions are drawn in Section V.

II. Integral Imaging Projection Equations

The principle of integral imaging is the simultaneous capturing of multiple views of the 3-D scene using an array of microlenses (Fig. 1(a)). To simplify the following analysis we refer to a cylindrical microlens array and by consequence to unidirectional integral images, though the analysis is easily expandable to the omnidirectional case too. Behind each microlens an image is formed, obtained from a slightly different point of view. This image is called elemental image and is of very low resolution due to the size of the microlens.

A re-arrangement of the columns of the elemental images leads to the formation of the viewpoint images (Fig. 2). The first viewpoint image is formed by the first column of each elemental image, the second by the second and so on. As parallel rays are recorded at the same position under each microlens, each viewpoint image contains information recorded from one particular direction. The resolution of the viewpoint images is equal to the resolution of the microlens array, making the intensity matching between them easier relative to elemental images. Suppose that the integral image has (N_1 x N_2) dimensions, and is produced by a lens array of M microlenses. Thus, there will be M elemental images, each being (K x N_2), where $K = N_1/M$ is the horizontal

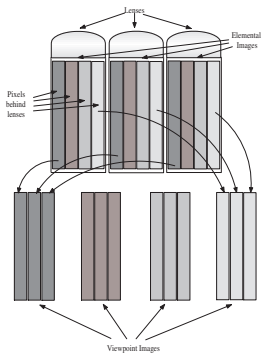

Fig. 2. Viewpoint image generation from a unidirectional integral image.

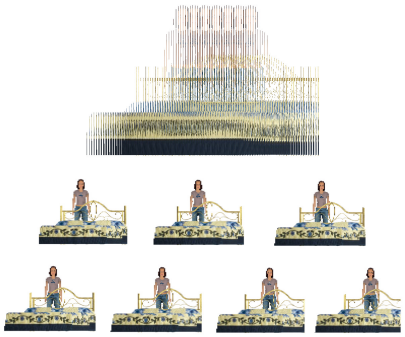

Fig. 3. A synthetic omnidirectional integral image that consists of seven viewpoint images. The viewpoint images are scaled for illustration purposes. Note the slight relative shifting from the left to the right, of the girl with respect to the bed.

pixel resolution of the microlens. Decomposing the integral image to its viewpoint images, results to K viewpoint images, each being (M x N_2). It should be noted that viewpoint images are different from common 2-D images. They are parallel projections, recording the 3-D scene from a different angle rather than perspective projection in the traditional 2-D recording. Orthographic projection is a type of parallel projection with rays perpendicular to the projection plane. Similarly, viewpoint images are orthographic projections of the 3-D scene in planes rotated, with respect to the image plane (Fig. 1(a)), by angle θ specified by the lens parameters. This angle depends on the pitch of each microlens and f (the focal length), and is ranging from $-\frac{pitch}{2f}$ to $\frac{pitch}{2f}$ (Fig. 1(a)). Figure 3 shows an example of a synthetic unidirectional integral image and the extracted viewpoint images.

The proposed algorithm involves comparison between the candidate areas of projections of scene's 3-D points onto the different parts of the integral image, i.e. the different elemental and viewpoint images. Analyzing the recording process of integral imaging, the connection between a 3-D point and its projections is established. A local coordinate system for each viewpoint image is used (Fig. 1) that is rotated by angle θ around the local coordinate system of the central viewpoint, which is the one that has zero angle with the image plane. If (i, j) are the coordinates of a pixel belonging to the k-th viewpoint image, then its position (x, y), with respect to the local coordinate system of the k-th viewpoint image, is given by the following equations:

$$x_k = \left(i - \left\lfloor \frac{M}{2} \right\rfloor\right) \cdot pitch \cdot \cos\theta_k \quad (1)$$

$$y_k = -\left(j - \frac{H}{2}\right) \cdot l \quad (2)$$

where H is the vertical pixel resolution of the microlens and l is the height of a pixel. If pixels are considered to be squared, then $l = \frac{pitch}{K}$. Viewpoint image planes are derived from the image plane with a rotation by θ rads around the y axis. The projection of a 3-D point on the k-th viewpoint image plane includes a rotation around the y-axis and a mapping from 3-D to 2-D. Using homogeneous coordinates:

$$\begin{bmatrix} x_k \\ y_k \\ 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \cos\theta_k & 0 & \sin\theta_k & 0 \\ 0 & 1 & 0 & 0 \\ -\sin\theta_k & 0 & \cos\theta_k & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix} \quad (3)$$

Denoting the first and second matrix of the right side of the projection equation as P and R_k, $k = 1, \ldots, K$, respectively, P is the orthographic projection matrix, while R_k is the k-th rotation matrix. Each 3-D object point is projected on several viewpoint planes and these projections can be used inversely to calculate the exact 3-D coordinates of the object point. Any pair of projections of the same 3-D point is sufficient to calculate its 3-D coordinates (X, Y, Z). Assuming that $(x_1, y_1), (x_2, y_2)$ are the projections on the planes with angles θ_1 and θ_2 respectively, (X, Y, Z) is given by:

$$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = \begin{bmatrix} \frac{x_2 \sin\theta_1 - x_1 \sin\theta_2}{sin(\theta_1 - \theta_2)} \\ y_1 \\ \frac{x_1 \cos\theta_2 - x_2 \cos\theta_1}{sin(\theta_1 - \theta_2)} \end{bmatrix} \quad (4)$$

Obviously, the above formula assumes a unique correspondence between $(x_1,\ y_1), (x_2,\ y_2)$, thus the depth estimation accuracy depends on the exactness of the specific correspondence. In order to produce a more reliable depth estimate, one would want to include any available correspondence pair. All pairs should contribute on the estimation procedure, thus the coordinates of the true 3-D point $(X,\ Y,\ Z)$ is the minimizer of the following system of equations in a least squares sense:

$$\begin{bmatrix} x_1 \\ y_1 \\ 1 \\ \vdots \\ x_K \\ y_K \\ 1 \end{bmatrix} = \begin{bmatrix} PR_1 & 0 & \ldots \\ 0 & \ddots & 0 \\ \ldots & 0 & PR_K \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix} \quad (5)$$

III. Depth estimation

The proposed algorithm's workflow is depicted in Fig. 4. First, the integral image is decomposed to the more meaningful viewpoint images, by appropriate sampling. Local descriptors are calculated in every viewpoint image, and as explained in section III-A, by matching the local features, correspondences

among the viewpoint images are established. Stable correspondences offered through a number of viewpoint image pairs, are gathered to form the anchor points set, i.e. a sparse set of 3-D points, with highly accurate depth estimates. In the next step,

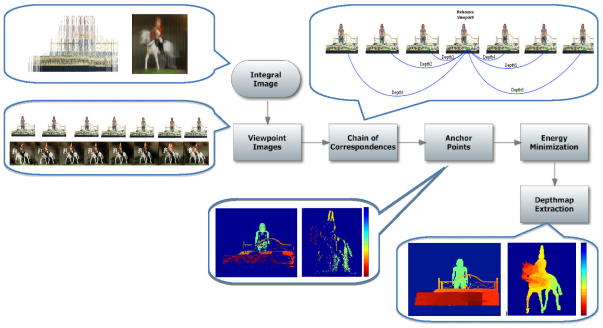

Fig. 4. Schematic representation of the proposed procedure for depth estimation.

as described in section III-B, the viewpoint images' features are provided in a 3-D surface fitting optimization procedure. The optimization problem involves assigning a label to each integral image pixel, where the label represents a depth value for the captured 3-D point. The optimization procedure is constrained by the anchor points, as explained in section III-C.

In the energy formulation two terms are used, the one provided by the data and the other by prior knowledge. The data term restricts the solution to be close to the observations, i.e the integral image, and the prior knowledge term enforces smoothness among neighboring pixels. However, the smoothness constraint is able to preserve discontinuities in order to avoid over-smoothing, especially at object boundaries where abrupt depth changes exist. The accurately estimated 3-D coordinates of anchor points contribute to the right handling of discontinuities as they lie mainly on edges. Graph cuts are utilized to solve this optimization problem, due to their capability of solving NP-hard problems efficiently and fast. The outcome of the algorithm is a reconstructed 3-D surface with high spatial smoothness. Projecting the reconstructed surface on the several viewpoint image planes, depth maps depicting the scene from a different viewpoint angle can be obtained. In the rest of this section, the three major steps of the algorithm, i.e. the anchor points detection, the energy formulation and the anchoring procedure are described in detail.

A. Anchor points detection

By construction, integral images are characterized by information abundance as multiple microlenses record the same scene's objects from slightly different angle. This captured abundant information should be exploited for depth extraction. Once an accurate correspondence between two or more viewpoint images is identified, then (5) provides an actual 3-D scene point. Several image correlation metrics are widely used for establishing correspondences between image pairs. However, in the case of viewpoint images, simple correlation metrics failed to indicate true correspondences in spite of the existence of high viewpoint images cross correlation. Instead of the deficient correlation metrics, in this work two more sophisticated local descriptors were tested, the well known SIFT descriptor [24] and the self-similarity descriptor [41]. Shechtman and Irani were the first to use local self-similarity patterns as a descriptor in the context of image and video matching. Local self-similarity descriptors have been successfully used for object detection and retrieval, while in the proposed work they are employed to extract robust correspondences, which will lead to the reliable estimation of 3-D anchor points, since they proved superior than SIFT features.

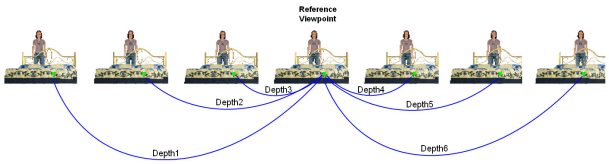

Fig. 5. A series of successive viewpoints. Correspondences, among the reference viewpoint image with the rest, are evaluated to produce the chain of pixels that share the same origin.

The anchor points detection procedure starts with the computation of the local self-similarity descriptors for the whole set of viewpoint images. Non-informative descriptors, which either correspond to large homogeneous image areas or they do not capture any self-similarity, are discarded. Then, a matching procedure is used between the remaining informative pixels of a reference viewpoint image, which is selected to be the central one, with pixels of the rest viewpoint images. For the matching procedure the sigmoid on the L_1 distance between the descriptors is used. For each informative pixel of the reference viewpoint image, a set is formed that contains only the strong correspondences with pixels of the rest of viewpoint images. Each correspondence pair provides the coordinates of its 3-D origin by solving equations (4). In case the set contains correspondence pairs that agree on the same origin, then these pixels are regarded as a chain. Each chain corresponds to an anchor point. Forcing the correspondence set to be from successive viewpoints, eliminates any chance of false positive anchor points. The final 3-D coordinates of the anchor point are computed by least squares fitting in that region according to (5).

The above procedure results in a dense set of points (Fig. 6) whose 3-D position estimation is of high accuracy and thus they can be considered as true scene points. As figure 6 shows, anchor points lie mainly on edges and textured regions and therefore their density depends, to a great extent, on the scene's texture. The fact that the estimated anchor points cover a wide area of the actual scene, significantly simplifies the problem of the overall scene reconstruction.

B. Energy formulation

Having a set of 3-D points whose depth estimate is very reliable, the next step is to try to fit a surface to connect them based on the information provided by the rest pixels of the

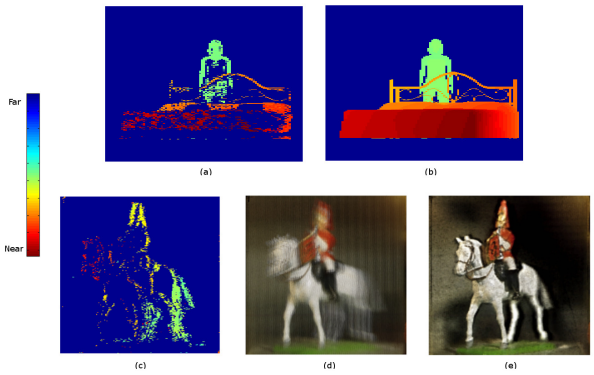

Fig. 6. Illustration of anchor points detection. (a) Anchor points depth-map of the synthetic scene "Bedroom" depicted in Fig. 3, and (b) its actual depth-map. (c) The estimated depth of the anchor points of the real integral image "horseman" (actual depth map is not available), (d) the integral image and (e) one viewpoint image.

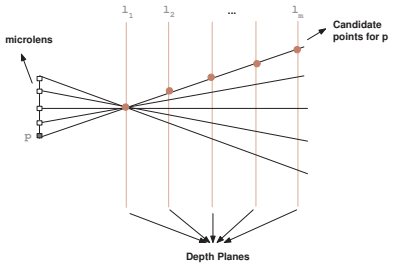

Fig. 7. For pixel's p ray, the candidate 3-D points are defined by uniformly sampling the ray with a constant depth step. Each successive depth step corresponds to the next label.

integral image. The constrained surface fitting optimization was implemented through graph cuts [6], hence the name Anchored Graph Cuts. Anchoring the optimization boosts the estimation accuracy, while reduces the optimization complexity. In the proposed formulation, the whole integral image is fed into one optimization. To accomplish the latter, the label that was assigned to each pixel was set to be its depth, rather its disparity relative to another picture, as is defined in the stereo-like concept. Thus, graph cuts are trying to find the optimized solution of the depth of the 3-D points that correspond to each pixel.

The 3-D scene is scanned uniformly along the emanating projection rays of the integral image, as depicted in figure 7. Each integral image's pixel is the projection of a scene's point along a specific ray. The rays are uniformly sampled and graph cuts evaluate those points whether they are true scene points. To facilitate this procedure we define a set of depth planes and confine the total amount of candidate 3-D scene points to the rays-planes intersections. If $\mathbf{P} = \{p_1, p_2, \ldots, p_n\}$ is the set of all pixels of the integral image and $\mathbf{L} = \{l_1, l_2, \ldots, l_m\}$ is a discrete set of labels, each one of them corresponding to m predefined depth values, a mapping $\mathbf{f} : \mathbf{P} \rightarrow \mathbf{L}$ that assigns each pixel a label is sought. Such a mapping should take into account during evaluation the depth value of neighboring pixels. In integral images a pixel's actual neighbors are the ones that the same microlens has produced. However, in the viewpoint image concept, a pixel has neighbors that belong to different microlenses. Both these sets of neighboring pixels are important and need to be taken under consideration in order to find the optimum solution for a pixel's depth value.

Graph cuts were employed to solve this problem, due to their efficiency of approximating NP-hard problems that try to find a labelling $\mathbf{f}$ which minimizes a given energy, as long as it is defined in a standard way [19]. This standard way suits the desired surface fitting optimization energy, which is defined as:

$$E(\mathbf{f}) = E_{data}(\mathbf{f}) + E_{smooth}(\mathbf{f}) \quad (6)$$

where $\mathbf{f}$ is a vector with a label for each pixel of $\mathbf{P}$. The data term $E_{data}(\mathbf{f})$ measures how well the certain depth values of $\mathbf{f}$ fit the integral image, while the smoothness term $E_{smooth}(\mathbf{f})$ measures how similar depth values neighboring pixels have. Those two terms are modelled as:

$$E(\mathbf{f}) = \sum_{\forall p_i \in P} D_{p_i}(\mathbf{f}(p_i)) + \sum_{\forall p_i, p_j \in Q} V_{p_i,p_j}(\mathbf{f}(p_i), \mathbf{f}(p_j)) \quad (7)$$

where $D_{p_i}(\mathbf{f}(p_i))$ represents the cost for assigning the label $\mathbf{f}(p_i)$ to pixel p_i, while $V_{p_i,p_j}(\mathbf{f}(p_i), \mathbf{f}(p_j))$ penalizes different labels $\mathbf{f}(p_i) \neq \mathbf{f}(p_j)$, on neighboring pixels p_i, p_j. Q defines the set of all pairs of neighboring pixels: $Q \subset \{\{p_i, p_j\} \mid p_i, p_j \in \mathbf{P}\}$. Before adding the two energy terms, they are normalized to be in the same order of magnitude.

The data term is responsible for the evaluation of candidate 3-D points, and assigns heavy cost if labelling $\mathbf{f}$ disagrees with the data (i.e the integral image), and small otherwise. The only restriction for this term is to be non-negative [6]. Taking into account that the projections of actual 3-D scene points on the several viewpoint image planes, according to (3), should lie on viewpoint image areas with similar color configurations, a photo-consistency matching cost is developed. Instead of the frequently used correlation metrics a more sophisticated approach of measuring photo-consistency is adopted. The self-similarity descriptor is again employed, due to its discriminative efficiency and the fact that descriptors have already been calculated. The descriptors that stemmed from the anchor points detection procedure are now used to form the data term of the energy function. To measure how well a certain pixel's label matches the image, the data term matches the descriptors of the pixels in the different viewpoint images that correspond to that 3-D point. Consider a pixel p_i which belongs to the i-th viewpoint and that the current label under investigation implies that p_i is the projection of P_{3D} (the 3-D point on the intersection of p_i's ray with the l-th depth plane). Projecting P_{3D} to the k-th viewpoint image, through (4), produces a candidate correspondence p_k. Having the set of $K-1$ candidate correspondences p_k with $k = 1, \ldots, K$ and $i \neq k$, the cost of assigning label l to p_i is then given by:

$$D_{p_i}(l) = 1 - median(S(SD_{p_i}, SD_{p_k})), \qquad k = [1, \ldots, K], k \neq i \quad (8)$$

where $S(SD_{p_i}, SD_{p_k})$ is the similarity between the descriptors SD_{p_i} and SD_{p_k} corresponding to pixels p_i and p_k respectively, and is in the range $\{0,1\}$. The overall similarity is selected to be the median over the $K-1$ calculated similarities to eliminate errors due to noise or occluded pixels. The cost $D_{p_i}(l)$ is thus low if the similarity between SD_{p_i} and SD_{p_k} is high and vice versa.

Furthermore, the proposed framework enables the modelling of the piecewise nature of depth by introducing a twofold regularization term among adjacent pixels on both the elemental and viewpoint images. This twofold neighborhood handling leads to reconstructed scenes with high spatial smoothness.

Without underestimating the importance of the data term it is a fact that the smoothness term is really important for the algorithm's success. The design of this term is much more difficult than the data term, as it tries to enforce smoothness constraints on the labelling, which is needed when the quantity corresponding to the labelling has to be naturally smooth. However, on the object boundaries abrupt label changes occur and thus, in order the smoothness term to be efficient, it has to handle those discontinuities properly. The smoothness term includes the notion of neighborhood Q

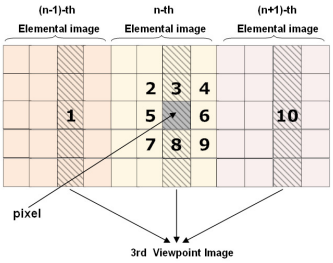

Fig. 8. The 10-neighborhood system used in the proposed algorithm.

that determines which pixels interact and probably have the same label. The neighborhood is reasonable to be assumed on the direct neighbors of a pixel in the viewpoint images. However, in this way the smoothness is achieved only in each separate viewpoint image but the whole reconstructed scene fails to be smooth. The advantage of the proposed energy formulation and the fact that labels correspond to depth values, instead of disparities among viewpoint images (as in stereo like formulation), is that it can define neighbors in the context of viewpoint and elemental images as well. Thus, a 10-neighborhood system (Fig. 8) composed of the 10 immediate neighbors, in the integral imaging context is incorporated. In this way, since neighboring pixels in elemental images correspond to neighboring viewpoint images, extra smoothing is achieved between them.

The smoothness energy is modelled through the truncated $L1$ distance between the labels of the neighbors. This choice is made because the truncated $L1$ distance is a discontinuity preserving penalty function. For a pair of neighboring pixels p_i and p_j with labels l_{p_i} and l_{p_j}, respectively, the smoothness energy is calculated by

$$V_{p_i,p_j}(l_{p_i}, l_{p_j}) = u_{\{p_i,p_j\}} min(T, |l_{p_i} - l_{p_j}|) \qquad (9)$$

T is a small constant, experimentally set to be in the range $\{2, 5\}$ and $u_{\{p_i,p_j\}}$ is a weight depending on the intensities I_{p_i} and I_{p_j} of the two involved pixels. The value of $u_{\{p_i,p_j\}}$ should be higher if the two neighbors have similar color and smaller, otherwise. The choice of $u_{\{p_i,p_j\}}$ is the one proposed in [6].

C. Anchoring Graph Cuts

As already mentioned, the set of anchor points is used as a reliable initialization of the optimization procedure. Another advantage is that they also propagate their reliable depth estimates to their neighbors through the smoothness term. In order to avoid unnecessary re-estimation of the anchor points' depth, anchor points are excluded form the optimization by setting their corresponding data term equal to zero for their known depth label, and infinite otherwise:

$$D_{p_\alpha}(f(p_\alpha)) = \begin{cases} 0, & if \quad f(p_\alpha) = l_{p_\alpha} \\ \infty, & if \quad f(p_\alpha) \neq l_{p_\alpha} \end{cases} \qquad (10)$$

where l_{p_α} is the closest label to an anchor point's depth, and p_α is any of the pixels that correspond to that anchor point α.

Setting to zero the data cost for known depth labels while assigning an infinite cost for the rest labels prevents from changing the correctly estimated depths of anchor points. Besides, it offers faster convergence of the algorithm while the known pixel labels spread over neighboring pixels, which leads to a robust solution.

IV. Experimental results

In order to be able to quantitatively evaluate the performance of the proposed algorithm and its components, a synthetic database was created which offered the advantage of knowing the ground truth, thus being able to measure the efficiency of the method and its intrinsic variations. The specifications and characteristics of the synthetic database are provided in section IV-A. Sections IV-B and IV-C describe the evaluation performed on the anchor points detection accuracy and on the graph cut efficiency, respectivelly. The overall method's accuracy and efficiency on depth estimation is provided on section IV-D, where the method's effectiveness on real application was further validated, on a small number of examples on real integral images.

A. Synthetic integral image database

To evaluate the performance of the proposed method, a virtual camera was built in order to capture a set of synthetic integral images. A database of 28 uni-directional integral images was constructed, each one of them depicting a 3-D scene with multiple objects. The objects were collected as 3-D models from the world wide web. The 3-D scene was transformed in order to fit to the field of view of the virtual array of microlenses. The texture of the objects in the scenes varies from rich to quite poor. The foreground is easily extracted, thus a foreground mask is also available for every image, in order to calculate depth only on the foreground.

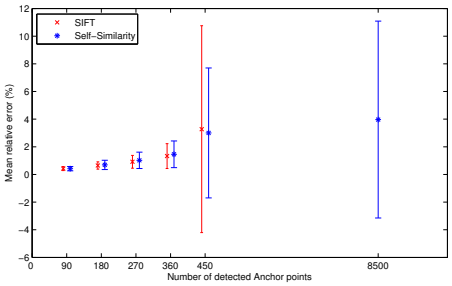

Fig. 9. Comparison of the mean relative error of the detected anchor points in the synthetic integral image database for SIFT and self-similarity. The bars show the standard deviation of the error.

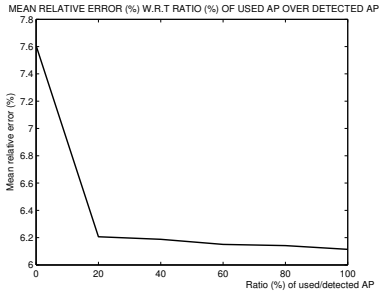

Fig. 10. Mean relative error in the synthetic integral image database with respect to the number of anchor points (AP) used per image for the proposed method. Higher density of used anchor points leads to more accurate depth estimates.

For the database construction, the lens pitch of the virtual camera was selected as $pitch = 0.5mm$, the lens height $l \cdot H = 50mm$ and the focal length $f = 4mm$. Further, the microlens array consisted of 99 cylindrical lenses with 7×700 pixels resolution. Using these parameters the resolution of the extracted integral images was 693×700 pixels.

Figure 15 shows few of the scenes captured with the virtual camera. The virtual camera and the complete dataset are publicly available at ftp://ftp.iti.gr/pub/Holoscopy/.

B. Anchor Points evaluation

For the anchor points detection procedure another descriptor was also tested before deciding to use the self-similarity descriptor; the well-known SIFT [24] that is widely used to detect local features in images. SIFT [62] detected on average 420 anchor points for every image in the database while the self-similarity was capable of producing 8500. Figure 9 demonstrates the mean relative error of the detected anchor points for both cases. The number of the anchor points increases as the restrictions in the generation procedure of the associated descriptor are relaxed. The mean relative error is defined as the absolute difference between the estimated and the actual depth over the actual depth. For the maximum number of potential SIFT based anchor points, the error is 3.27% for SIFT contrary to 3.07%, with much less variance, for the same amount of self-similarity based anchor points. However, the self-similarity is further capable of producing 8500 anchor points on average per image, with a slight increase of error to 3.98%. Although there is not a notable difference in the error, there is indeed in the error variance and in the density of the detected anchor points. Being able to produce a large number of reliable anchor points (8500 contrary to 420 on average), helps on the overall depth estimation, especially on images poor in gradients and texture.

In order to reveal and quantify the contribution of the anchor points to the depth estimation, the mean relative error, with respect to the amount of anchor points used to constrain the optimization procedure, was calculated. Figure 10 illustrates how the error decreases dramatically once APs are incorporated in the energy optimization procedure (figure 12 visualizes this error difference) and that the more the anchor points used the less the overall error. This is because the smoothness term propagates the APs' depth estimation reliability to their neighbors, thus more APs produces more reliable depth estimates.

The diagram in figure 10 also shows that in the synthetic database, the mean relative error falls very smoothly for large AP percentages. In real images though (where only visual inspection of depth accuracy is available), it becomes obvious that one would prefer to use all the available APs, as seen in figure 11.

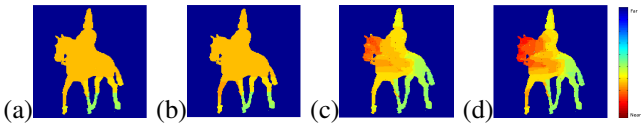

Fig. 11. Influence of the amount of anchor points in depth estimation or the real integral image "horseman". (a) Result without using anchor points, (b) 25% of the anchor points used, (c) 50% of the anchor points used, (d) 100% of the anchor points used.

In order to better understand the importance of APs and visualize the error decrease when anchoring the optimization procedure, depth estimation results in the form of depth maps are depicted in figure 12. As it can be seen without APs different objects are confused while they get more distinctive and detectable when APs constrain the optimization. Furthermore, with a close inspection, the un-anchored depth maps are more "flat" (e.g. the female figure and the handbag in the middle row and the table in the bottom row), while in the anchored ones the depth demonstrates more variations, showing that more depth details were captured. Undoubtedly, the proposed depth estimation algorithm benefits, to a great extent, from anchor points and its success is mainly attributed to the accurate estimation of their 3-D coordinates.

C. Data term evaluation

Apart from matching the self-similarity descriptor, the Pearson correlation was also investigated for formulating the data term of the energy function. The correlation between patches around the involved pixels was used in (8) instead of the self-similarity. Several sizes of patches were tested with the 11×3 producing the best results. Comparative results of using correlation and self-similarity to impose photo-consistency in the depth estimation are shown in figure 13. Obviously, this diagram suggests the use of the self-similarity descriptor for

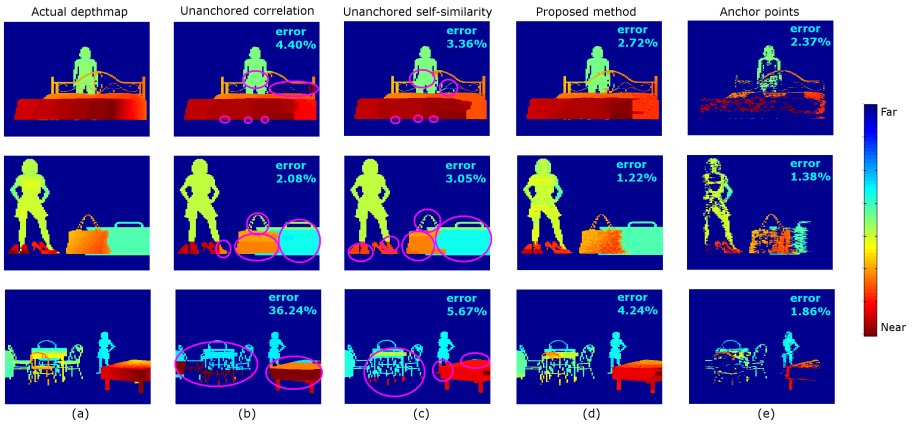

Fig. 12. Influence of the anchor points selection in the optimization procedure. The results of three integral images are shown. (a) Actual depth. (b) Obtained depth map without the use of anchor points with correlation. (c) Obtained depth map without the use of anchor points with self-similarity. (d) Depth map using constraints from anchor points (proposed method) (e) Estimated depth of anchor points. The circles depict the regions where the un-anchored cases produce obvious errors that the proposed method successfully handles.

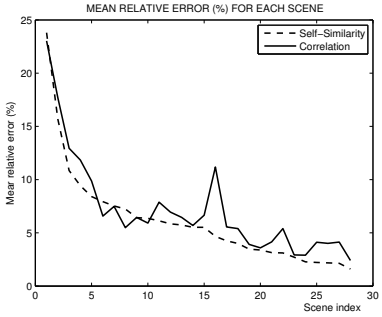

Fig. 13. Comparative results for the proposed algorithm using a correlation metric instead of self-similarity matching in the data term calculation.

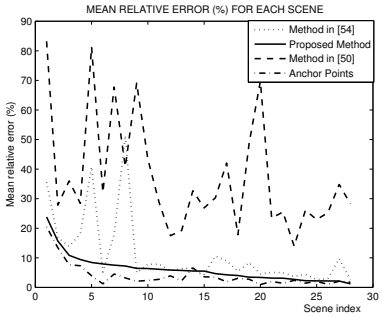

Fig. 14. Comparison of the proposed algorithm to the ones of [54] and [50]. The accuracy of anchor points is also displayed in the diagram.

the data term calculation. This choice is also computationally efficient, since the self-similarity descriptor has been already calculated for the whole integral image in the anchor point detection step. Columns (b) and (c) of figure 12, give a further insight of the differences produced in depth estimation for both cases, when it is solely based on the data term.

D. Depth accuracy

The proposed algorithm was compared against other state-of-the-art algorithms, in the synthetic database, which provides ground truth measurements, thus a reliable way for quantifying the effectiveness of each method. The proposed algorithm produces a reconstructed scene that can be projected onto any viewpoint image plane in order to obtain a depth-map. The depth-map produced for the central viewpoint image can be directly compared to the outputs of the algorithms in [50] and in [54]. The results are depicted in figure 14. Depth estimates have been calculated on the foreground only. The mean relative error is 6.13% for the proposed algorithm, 11.74% for the algorithm of [54] and 38.23% for [50]. It is obvious that the proposed method clearly outperforms the other two methods.

The results of the proposed algorithm for some integral images are illustrated in figure 15 where the estimated depth-map is shown in contrast with the scene's actual depth-map, both being estimated from the central's viewpoint angle. In all cases, the different objects are correctly perceived and differentiated from their neighboring, based on their estimated depth values. However, fine details in objects contours are not always obtained with high accuracy. This happens mainly for objects that are small, or on their narrow parts, since the distinction between them and the surrounding objects is not clear, even for the human eye in those low resolution viewpoint images. However, even in such cases, a satisfactory approximation of the object contour is obtained that makes perceivable the kind of the object.

The proposed algorithm was also tested in real integral images. The first one is the "horseman" image, captured with a cylindrical lens array, that produces an integral image of size equal to 1280×1264, has 160 cylindrical lenses, thus offering 160 elemental images of 8×1264 pixels, and 8 viewpoint images of 160×1264 pixels. As a foreground mask was not directly available for them, and since it is hard to calculate depth on large non-informative and homogeneous regions, a simple color blob segmentation technique [63] could be applied in order to separate the objects from the background. Figure 16(c) shows the foreground mask obtained for the integral image "horseman". The bigger part of the foreground area is successfully detected but a part of the object with

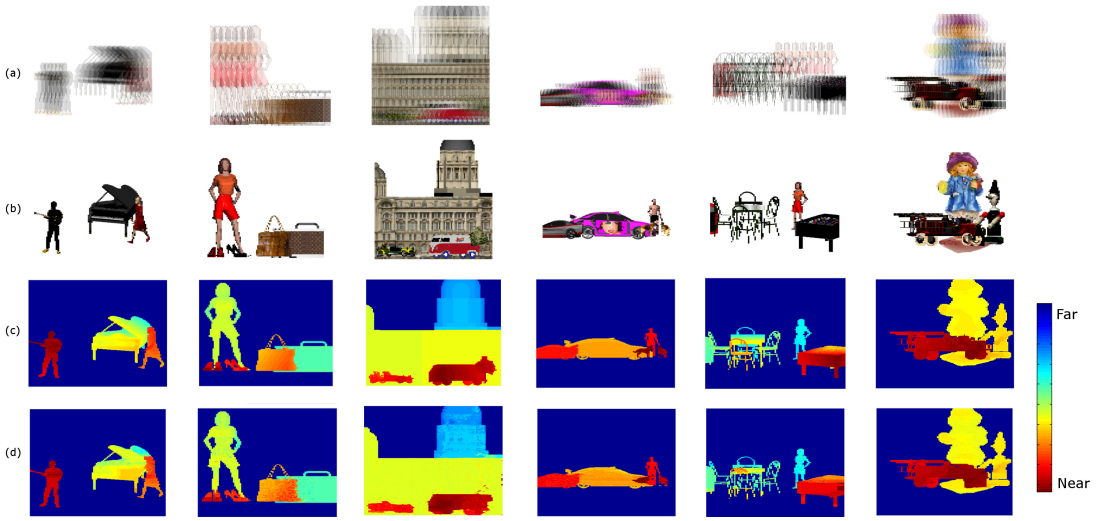

Fig. 15. Results of the proposed algorithm on synthetic integral images. (a) Integral images. (b) Corresponding viewpoint images. (c) Actual depth map of the scene. (d) Depth map estimation using the proposed method.

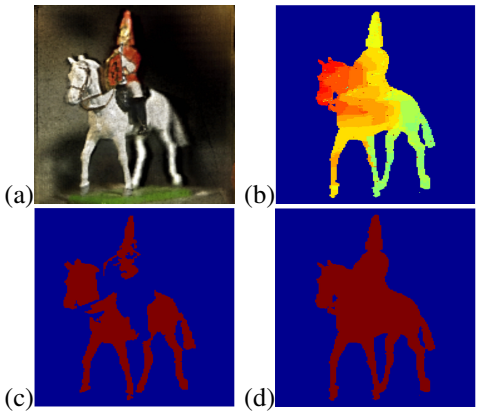

Fig. 16. (a) The central viewpoint of the real integral image "horseman", (b) obtained depth-map using the proposed method, (c) foreground mask extracted by [63], (d) corrected foreground mask.

similar color with the background is missed as it is considered as background. Since foreground detection is out of the scope of this work, the foreground mask was corrected to include that region too, as in figure 16(d). Figure 16(b) illustrates the estimated "horseman" depth-map, when incorporating the foreground mask. Despite that the actual depth-map of the image is not available, it is obvious that the objects's relative positions are correctly estimated. The head of the horse is estimated to be near the camera while the tail far, with smooth depth transitions in between, which match with the horse's actual shape. In general, the depth-map is sufficiently smooth and the existing errors are limited. In [50], the result on this image is also presented.

Figure 17 shows another example of a real integral image depicting a palm, captured from another cylindrical lens array than the one utilized in the "horseman" capturing. This configuration contains 84 cylindrical lenses, of 67 pixels width each, offering 84 elemental and 67 viewpoint images. The size of the integral image is 5628×3744, while the size of each elemental and viewpoint image is 67×3744 and 84×3744, respectively. The depth values' range is very narrow in this case, considering how flat the palm appears in front of the camera. Nonetheless, the proposed algorithm was capable, even in this case, to produce a smooth surface and to recognize correct depth transitions along each finger and between them.

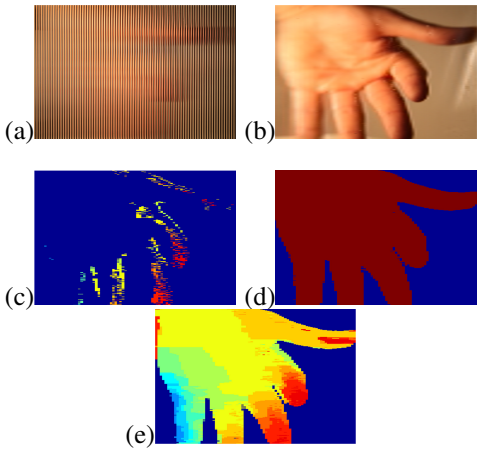

Fig. 17. Results on the real integral image "hand". (a) Integral image, (b) viewpoint image, (c) anchor points, (d) foreground mask, (e) extracted depth-map using the proposed method.

V. Conclusions

In this paper an algorithm for estimating depth from integral images was presented. Depth extraction is formulated as one energy optimization problem that is solved by anchoring the graph cuts technique. The algorithm successfully overcomes the integral imaging difficulties and produces a smooth scene structure. The use of the self-similarity descriptor both for the detection of anchor points and the imposition of photo-consistency in the optimization procedure along with the smoothness constraints was proved to be the key to success. Experiments with a ground-truth dataset showed that the proposed method clearly outperforms other state-of-the-art

algorithms, while experiments with real data further verified its efficiency on producing accurate depth-maps.

ACKNOWLEDGMENT

This work was supported by the 3D VIVANT EU funded project [60], [61]. The authors would like to thank Dr. Amar Aggoun for his suggestions on constructing the virtual camera and for providing us with the real integral images and to Dr. Peter Lanigan for the integral camera development and image capturing.

REFERENCES

[1] A. Aggoun, "3D holoscopic imaging technology for real-time volume processing and display," High-Quality Visual Experience, pp. 411428. SpringerLink, 2010.

[2] M. Bertero and P. Boccacci, *Introduction to Inverse Problems in Imaging*. Institute of Physics Publishing, 1998.

[3] R. Bhotika, D. Fleet, and K. Kutulakos, "A probabilistic theory of occupancy and emptiness," in ECCV, vol. 3, pp. 112. 132, 2002.

[4] B.G. Blundell, A.J. Schwarz, D.K. Horrell, "Volumetric three-dimensional display systems: their past, present and future," Engineering Science and Education Journal , vol.2, no.5, pp.196-200, Oct 1993.

[5] Yuri Boykov, Olga Veksler, and Ramin Zabih, "Markov random fields with efficient approximations," IEEE Conference on Computer Vision and Pattern Recognition, pages 648655, 1998.

[6] Y. Boykov, O. Veksler, and R. Zabih, "Fast approximate energy minimization via graph cuts," IEEE Trans. on Pattern Analysis and Machine Intelligence, vol. 23, no. 11, pp. 12221239, Nov. 2001.

[7] A. Castro, Y. Frauel, and B. Javidi, "Integral imaging with large depth of field using an asymmetric phase mask," Opt. Express 15, 10266-10273, 2007.

[8] S. Cirstea, A. Aggoun, M. McCormick, "Depth extraction from 3D-integral images approached as an inverse problem", Proc. IEEE Int. Symposium on Industrial Electronics, Cambridge, UK, pp 798-802, 2008.

[9] N.A. Dodgson, "Autostereo displays: 3D without glasses". In: EID: Electronic Information Displays, 1997.

[10] Sadeg M. Faris, "Novel 3D stereoscopic imaging technology", Proc. SPIE 2177, 180, 1994.

[11] Y. Furukawa and J. Ponce, "High-fidelity image-based modeling," Technical Report 2006-02, UIUC, 2006.

[12] P. Gargallo and P. Sturm, "Bayesian 3D modeling from images using multiple depth maps," in CVPR, vol. II, pp. 885. 891, 2005.

[13] C. Hernandez and F. Schmitt, "Silhouette and stereo fusion for 3D object modeling," CVIU, 96(3):367.392, 2004.

[14] W. Ijsselsteijn, H. de Ridder, and J. Vliegen, "Effects of stereoscopic filming parameters and display duration on the subjective assessment of eye strain," Proceedings of the SPIE, Stereoscopic Displays and Virtual Reality Systems VII, 3957, pp. 12-22, 2000.

[15] J. Isidoro and S. Sclaroff, "Stochastic refinement of the visual hull to satisfy photometric and silhouette consistency constraints," in ICCV, pp. 1335.1342, 2003.

[16] J.-S. Jang and B. Javidi, "Improved viewing resolution of three-dimensional integral imaging by use of nonstationary micro-optics," Opt. Lett. 27, 324326, 2002.

[17] Y. Kim, J.-H. Park, S.-W. Min, S. Jung, H. Choi, and B. Lee, "Wide-viewing-angle integral threedimensional imaging system by curving a screen and a lens array," Appl. Opt. 44, 546-552, 2005.

[18] V. Kolmogorov and R. Zabih. "Computing visual correspondence with occlusions using graph cuts," In ICCV, volume II, pages 508515, 2001.

[19] Vladimir Kolmogorov and Ramin Zabih, "What energy functions can be minimized via graph cuts?," In European Conference on Computer Vision, 2002.

[20] V. Kolmogorov and R. Zabih, "Multi-camera scene reconstruction via graph cuts," ECCV, vol. III, pp. 82.96, 2002.

[21] K. Kutulakos and S. Seitz, "A theory of shape by space carving," IJCV, 38(3):199.218, 2000.

[22] L. Landweber, "An iteration formula for Fredholm integral equations of the first kind," Amer. J. Math., vol. 73, pp. 615-624, 1951.

[23] M. G. Lippmann, "La photographie integrale," Comptes-Rendus de l'Académie des Sciences, vol. 146, pp. 446551, 1908.

[24] D. Lowe, "Distinctive image features from scale-invariant keypoints," IJCV, 60(2):91110, 2004.

[25] S. Manolache, S.-Y. Kung, M. McCormick, and A. Aggoun, "3d-object space reconstruction from planar recorded data of 3d-integral images," Journal of VLSI Signal Processing, vol. 35, no. 1, pp. 518, Aug. 2003.

[26] S. Manolache, M. McCormick, and S. Y. Kung, "Hierarchical adaptive regularization method for depth extraction from planar recording of 3-D-integral images," in Proc. ICASSP, vol. 3, pp. 14331436, 2001.

[27] M. Martnez-Corral, B. Javidi, R. Martnez-Cuenca, and G. Saavedra, "Integral imaging with improved depth of field by use of amplitude modulated microlens array," Appl. Opt. 43, 5806-5813, 2004.

[28] M. Martnez-Corral, B. Javidi, R. Martnez-Cuenca and G. Saavedra, "Formation of real, orthoscopic integral images by smart pixel mapping," Opt. Express 13, 9175-9180, 2005.

[29] R. Martnez-Cuenca, H. Navarro, G. Saavedra, B. Javidi, and M. Martinez-Corral, "Enhanced viewing-angle integral imaging by multiple-axis telecentric relay system," Opt. Express 15, 16255-16260, 2007.

[30] Ral Martnez-Cuenca, Genaro Saavedra, Manuel Martnez-Corral, and Bahram Javidi, "Enhanced depth of field integral imaging with sensor resolution constraints," Opt. Express 12, 5237-5242, 2004.

[31] R. Martinez-Cuenca, G. Saavedra, M. Martinez-Corral, and B. Javidi, "Progress in 3-D multiperspective display by integral imaging," Proc. IEEE 97, 10671077, 2009.

[32] T. Okoshi, *Three Dimensional Imaging Techniques*. Academic Press, 1976.

[33] Jae-Hyeung Park, Sungyong Jung, Heejin Choi, Yunhee Kim, and Byoungho Lee, "Depth Extraction by Use of a Rectangular Lens Array and One-Dimensional Elemental Image Modification," Appl. Opt. 43, 4882-4895, 2004.

[34] J.-H. Park, S. Jung, H. Choi, and B. Lee, "Viewing-angle-enhanced integral imaging by elemental image resizing and elemental lens switching," Appl. Opt. 41, 6875-6883, 2002.

[35] S. Roy and I. Cox, "A maximum-flow formulation of the N camera stereo correspondence problem," in ICCV, pp. 492. 499, 1998.

[36] G. Saavedra, R. Martinez-Cuenca, M. Martinez-Corral, H. Navarro, M. Daneshpanah, and B. Javidi, "Digital slicing of 3D scenes by Fourier filtering of integral images," Opt. Express 16, 17154-17160, 2008.

[37] S. Savarese, H. Rushmeier, F. Bernardini, and P. Perona, "Shadow carving," in ICCV, pp. 190.197, 2001.

[38] D. Scharstein and R. Szeliski, "A taxonomy and evaluation of dense two-frame stereo correspondence algorithms," International Journal of Computer Vision, 47(1-3):7-42, 2002.

[39] S. Seitz, B. Curless, J. Diebel, D. Scharstein, and R. Szeliski, "A comparison and evaluation of multi-view stereo reconstruction algorithms," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, volume 1, pages 519-526, 2006.

[40] S. Seitz and C. Dyer, "Photorealistic scene reconstruction by voxel coloring," IJCV, 35(2):151.173, 1999.

[41] E. Shechtman and M. Irani, "Recognition of hand gestures using range images," in IEEE Conf. on Computer Vision and Pattern Recognition, Minneapoli, MN, USA, pp. 18, Jun. 2007.

[42] ES. Sinha and M. Pollefeys, "Multi-view reconstruction using photo-consistency and exact silhouette constraints: A maximum-flow formulation," in ICCV, pp. 349.356, 2005.

[43] C. Slinger, C. Cameron and M. Stanley, "Computer-generated holography as a generic display technology," Computer, vol. 38, no. 8, pp. 46-53, 2005.

[44] R. Szeliski, "A multi-view approach to motion and stereo," in CVPR, vol. 1, pp. 157.163, 1999.

[45] Ridwan Bin Adrian Tanjung, Xuewu Xu, Xinan Liang, Sanjeev Solanki, Yuechao Pan, Farzam Farbiz, Baoxi Xu and Tow-Chong Chong, "Digital holographic three-dimensional display of 50-Mpixel holograms using a two-axis scanning mirror device", Opt. Eng. 49, 025801, Mar 02, 2010.

[46] A.N. Tikhonov, V.Y. Arsenin, "Solutions of ill-posed problems", Wiley, New York, 1977.

[47] A. Treuille, A. Hertzmann, and S. Seitz, "Example-based stereo with general BRDFs," In ECCV, vol. II, pp. 457.469, 2004.

[48] George Vogiatzis, Carlos Hernandez Esteban, Philip H. S. Torr, Roberto Cipolla, "Multiview Stereo via Volumetric Graph-Cuts and Occlusion Robust Photo-Consistency," IEEE Transactions on Pattern Analysis and Machine Intelligence, pp. 2241-2246, December, 2007.

[49] C. Wu, A. Aggoun, M. McCormick, and S. Y. Kung, "Depth extraction from unidirectional integral image using a modified multi-baseline technique," Proc. SPIE, vol. 4660, pp. 135143, 2002.

[50] C. H. Wu, M. McCormick, A. Aggoun, and S.-Y. Kung, "Depth mapping of integral images through viewpoint image extraction with a hybrid disparity analysis algorithm," Journal of Display technology, vol. 4, 2008.

[51] S. Yano, S. Ide, T. Mitsuhashi, and H. Thwaites, "A study of visual fatigue and visual comfort for 3-D HDTV/HDTV images," Displays, vol. 23, pp. 191201, 2002.
[52] Ju Yong Chang, Haesol Park, In Kyu Park, Kyoung Mu Lee, and Sang Uk Lee, "GPU-friendly multi-view stereo reconstruction using surfel representation and graph cuts," Comput. Vis. Image Underst. 115, no. 5,pp. 620-634, May 2011.
[53] T. Yu, N. Xu, and N. Ahuja, "Shape and view independent reflectance map from multiple views," in ECCV, pp. 602. 616, 2004.
[54] D. Zarpalas, I. Biperis, E. Fotiadou, E. Lyka, P. Daras, M. G. Strintzis, "Depth estimation in integral images by anchoring optimization techniques", IEEE International Conference on Multimedia & Expo (ICME), 2011.
[55] G. Zeng, S. Paris, L. Quan, and F. Sillion, "Progressive surface reconstruction from images using a local prior," In ICCV, pp. 1230.1237, 2005.
[56] Yan Zhang, Quanmin Ji, Wenshuai Zhang, "Multi-view autostereoscopic 3D display," International Conference on Optics Photonics and Energy Engineering (OPEE), 2010.
[57] L. Zhang and S. Seitz, "Image-based multiresolution shape recovery by surface deformation," in SPIE: Videometrics and Optical Methods for 3D Shape Measurement, pp. 51. 61, 2001.
[58] C. Zitnick, S.-B. Kang, M. Uyttendaele, S. Winder, and R. Szeliski, "High-quality video view interpolation using a layered representation," ACM Trans. on Graphics, 23(3):600. 608, 2004.
[59] B. Micallef, C.J. Debono, and R. Farrugia, "Exploiting depth information for efficient multi-view video coding", IEEE International Conference onMultimedia & Expo (ICME), 2011.
[60] Amar Aggoun, "3D Holoscopic video content capture, manipulation and display technologies," 9th Euro-American Workshop on Information Optics (WIO), 2010
[61] www.3dvivant.eu
[62] http://www.vlfeat.org/~vedaldi/code/sift.html
[63] http://www.mathworks.com/matlabcentral/fileexchange/25157-blobsdemo