Continuous Depth Map Reconstruction From Light Fields

Jianqiao Li, Minlong Lu, and Ze-Nian Li

***Abstract*— In this paper, we investigate how the recently emerged photography technology—the light field—can benefit depth map estimation, a challenging computer vision problem. A novel framework is proposed to reconstruct continuous depth maps from light field data. Unlike many traditional methods for the stereo matching problem, the proposed method does not need to quantize the depth range. By making use of the structure information amongst the densely sampled views in light field data, we can obtain dense and relatively reliable local estimations. Starting from initial estimations, we go on to propose an optimization method based on solving a sparse linear system iteratively with a conjugate gradient method. Two different affinity matrices for the linear system are employed to balance the efficiency and quality of the optimization. Then, a depth-assisted segmentation method is introduced so that different segments can employ different affinity matrices. Experiment results on both synthetic and real light fields demonstrate that our continuous results are more accurate, efficient, and able to preserve more details compared with discrete approaches.**

***Index Terms*— Depth estimation, light fields, sparse linear systems.**

I. Introduction

DEPTH map reconstruction, also known as disparity estimation, is a traditional challenging computer vision task which has been studied for more than three decades. The conventional way to get depth values is from stereo images. Scharstein and Szeliski gave a good survey of this topic in [1]. Most methods nowadays solve the problem by minimizing an energy function, which usually consists of a data term and a smoothness term. The two most popular models for the energy functions are the Markov Random Fields model [2] (MRFs) and variational approaches [3], while the formulation of the data term and the smoothness term can be different. For stereo matching, a lot of previous work shows that energy functions with non-convex terms model the problem better, although in fact only an approximate optimization can be found. In this case, the recovered depth values are not continuous but discrete in the depth range. A drawback of the discrete methods is that the time and memory cost of the algorithm is related to the number of quantized levels. When the level is not high enough, "stair" effects are usually noticeable in the recovered depth maps.

The newly developed technology the *light field* has brought new possibilities in reconstruction of depth maps. Compared with traditional image data, a light field contains not only accumulated colour intensities but also some information about ray directions. In general, the light field data can be seen as a set of photos captured from densely and regularly placed cameras. When the views are dense enough, there are some interesting properties that make light fields different from traditional multi-view data. One of the properties that has been studied to solve traditional computer vision problems is the fact that the projected points from one 3D point onto different views correspond to a line in the so called epipolar-plane image (EPI). The slope of this line is related to the depth of the point in the space, which transforms the problem of depth estimation into line detection on EPIs.

To estimate the orientation of the lines on EPIs, one idea is to try out all different orientations: the one with the least colour variance along the line is most likely to give the correct depth value. Several methods have been developed based on this point; different methods use different ways to measure the colour variance. Kim *et al.* employed a modified Parzen window estimation with an Epanechenikov kernel [4]. Tao *et al.*, on the other hand, used the standard deviation estimation [5]. Defocus is another clue that can be used for depth. Instead of using the colour variance, researchers try to find out by how much angle the EPIs need to be sheared to make the interest point in focus. Defocus and colour variance are used together to find the depth in [5] and [6]. An alternative approach proposed by Wanner and Goldluecke [7], is to use a structure tensor to estimate the slope of the lines on EPIs. Unlike the other methods, it does not need to try out different hypothetical depth values to find the optimal one, but at once provides an estimation as well as a certainty level from one structure tensor operation. In this work, we use the estimation from the structure tensor as a starting point, followed by a refinement step by examining the colour correspondence along the detected line from the structure tensor.

The initial estimations from light fields are denser and more reliable compared with those from traditional stereo images, which also makes the optimization step different. However, some works still follow a similar optimization method as with stereo matching, such as MRFs in [5] and functional lifting in [7]. These methods have to discretize the depth values, so lose the advantages of the dense and reliable initial estimations. Wanner and Goldluecke showed in [8]

Manuscript received September 13, 2014; revised February 17, 2015 and May 12, 2015; accepted May 17, 2015. Date of publication June 3, 2015; date of current version June 23, 2015. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Christine Guillemot.

The authors are with the School of Computing Science, Simon Fraser University, Burnaby, BC V5A 1S6, Canada (e-mail: cambridge0427@gmail.com; minlongl@sfu.ca; li@cs.sfu.ca).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TIP.2015.2440760

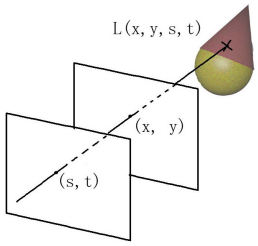


Fig. 1. Two-plane parametrization of 4D light fields.

that a simple denoising filter can generate comparable results with the discrete global optimization, since the filter keeps the continuous depth values. In Kim et al.'s work [4], by applying the local method iteratively on the EPIs with different resolutions, global estimation can be avoided. Only a median filter is applied to eliminate outliers. With the aforementioned initial estimation methods, the depth values usually come with certainty levels. Therefore the optimization problem is in essence to propagate the reliable estimations to other parts. In this work, we explore an optimization method, given by solving a linear system, which generates a smooth and globally optimized result.

The paper extends our previous work [9]. We introduce the refinement step of the initial estimations in Section III and the optimization method in Section IV. The performance of the proposed method is compared with the state-of-the-art methods in Section V.

II. Related Work

The light field concept was originally defined by physicists who interpreted the flow of light as a field. It is a plenoptic function which describes the amount of light, also known as radiance, travelling towards every direction through every point in the space. However, to measure the radiance of the light at every location towards every direction is not feasible in practice. The capture of light fields in fact relies on sampling the radiance in space and reconstructing the plenoptic function.

The 4D light field was first proposed in [10] and later widely used in light field analysis. We adopt the two-plane parametrization [10] of 4D light fields and denote it as $L(x, y, s, t)$, as shown in Figure 1. Under this parametrization, a 4D light field can be seen as a 2D array of perspective views, where (s, t) can be seen as the index of different views and (x, y) are spatial coordinates within each view (see Figure 2a).

By fixing y and t, we can obtain a 2D (x, s) slice of a light field, as shown in Figure 2b. Similarly, 2D (y, t) slices can be obtained if x and s are fixed. These 2D slices are called *epipolar plane images* (EPIs). Any point in the 3D space can be projected to a line on EPIs. The slope of the line is shown to be related to the depth of the corresponding point in 3D space [11].

Therefore, depth values can be obtained by estimating the slope of lines in EPIs [7]. A structure tensor [12] is employed, which produces an orientation estimation at each point and the confidence level of each estimation. Figure 2c shows a depth map obtained from the orientation estimation. Figure 2d shows

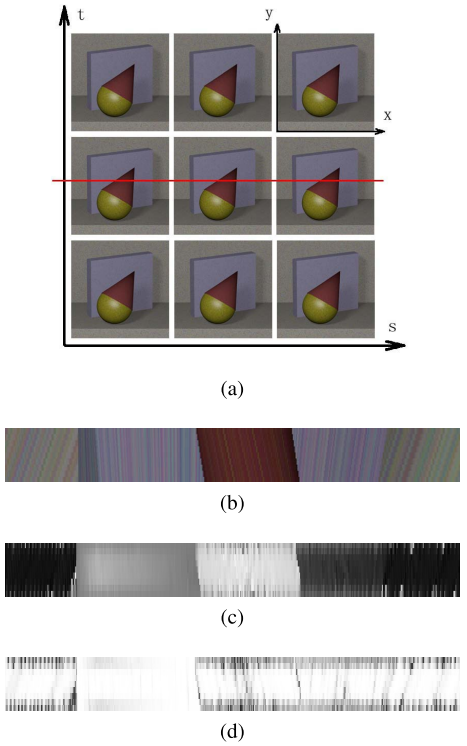


Fig. 2. Initial estimation on EPIs. (a) A visualization of 4D light field. (b) Epipolar plane image. (c) Initial depth estimation on EPI. (d) Confidence map on EPI.

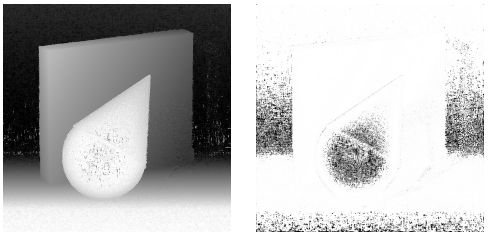


Fig. 3. An example of the initial depth estimation. (a) Initial depth estimation. (b) Confidence map.

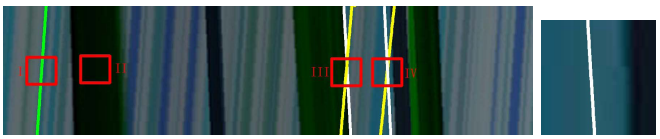
the corresponding certainty map, in which the brighter colours indicate higher certainty levels, and vice versa.

As we show in Figure 2, applying the structure tensor on an EPI, we can get the depth information on one horizontal scan line. By analyzing every 2D slice with different possible y, we can assemble the initial depth estimations and their confidence levels for the whole image from a certain view, as shown in Figure 3. Based on the initial estimation we try to construct a smooth and consistent depth map by solving linear systems instead of making use of the discrete labelling approaches.

III. Certainty Map Refinement

A. Limitations of the Structure Tensor

The structure tensor [12] provides a good way to get an initial estimation of the depth. However, it also has several limitations. The structure tensor works to find the dominant orientation in a small neighbourhood of a point. However, if there is no dominant orientation in the neighbourhood, i.e.



(a) (b)

Fig. 4. An illustration of different cases for the structure tensor. (a) An epipolar-plane image. In window I, a dominant orientation is detected, as indicated by the green line. In window II, no orientation exists. In window III, multiple orientations exist. In window IV, the real slope is indicated by the yellow line, while the estimation by the structure tensor is indicated by the white line. In Window II and III, low certainty levels are assigned. But in window IV, a high certainty level is assigned to the wrong estimation. (b) A close-up of the local window IV. The dominant orientation in the local window is indicated by the white line. But it is not the case if the whole EPI is taken into consideration.

homogeneity or multiple orientations, the structure tensor cannot give a reliable estimation. For example, in Figure 4a, the colour is uniform in the local windows II, and two dominant orientations exist in the local window III. In these cases the structure tensor can assign a low certainty level to the estimation. Areas in these cases can be fixed by the global optimization step later, because the estimations with high certainty levels will be propagated to areas with low certainty levels.

However, in areas where depth is discontinuous, the structure tensor tends to give wrong estimations but still assigns high certainty levels for them. For example the slope in the window IV of Figure 4a is along the yellow line, given the information of the whole EPI. However, the structure tensor only considers the local neighbourhood, as shown in Figure 4b, in which the dominant orientation turns to be along the white line. Therefore, in this case, the structure tensor will provide the orientation of the white line as the estimation and assign a high certainty level for it, because in the local neighbourhood Figure 4b, the white line is indeed the dominant orientation. However, in the whole EPI, it is in fact the orientation of the green area nearby. Consequently, the depth maps produced by the structure tensor tend to have "fattened" boundaries along depth discontinuities with high certainty levels assigned, as shown in Figure 6.

Assigning high certainty level to wrong estimations has an adverse effect on the future optimization step, since the optimization method works to propagate information from reliable estimations to unreliable ones. To this end, we propose a method to refine the certainty maps produced by the structure tensor in the next section.

B. The Refinement of Certainty Maps

To fix the problems discussed in last section, a variational labelling method [13] is employed to enforce visibility constraints on each EPI in Wanner and Goldluecke's paper [7]. However, since this optimization has to be applied on each 2D slice of the light fields, they need to optimize hundreds of times for each light field, which usually takes several hours. An alternative way to enhance the accuracy is to do visibility reasoning explicitly. Instead of explicit visibility reasoning, Zhang *et al.* [14] incorporate visibility into the data term of the energy function using statistical information from both colour and geometry.

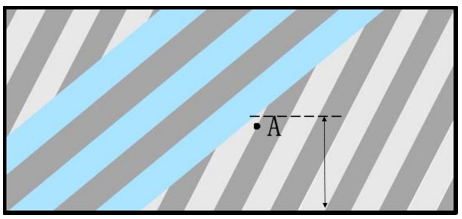


Fig. 5. Temporal Selection.

Inspired by Zhang et al.'s idea [14], we propose a method to refine the certainty map of the initial local estimation, so that wrong estimations are always assigned with low certainty. As shown in Figure 6e, the "fattened" boundaries are assigned to low certainty values after this refinement.

In an EPI $I(x, s)$, any point (x, s) can be warped to other views, given the estimated depth $d(x, s)$. Ideally, under the Lambertian assumption, the colour at the original point and at the warped point should be identical. We define a *matching distance* to measure the difference between the original point (x, s) and the one warped onto view s'.

$$\Phi(x, s, s') = ||\mathbf{I}(x, s) - \mathbf{I}(x', s')|| + \lambda(c(x', s') + c(x, s))|d(x, s) - d(x', s')|. \tag{1}$$

In the distance, $\mathbf{I}(x, s)$ and $\mathbf{I}(x', s')$ are three-dimensional vectors for colour intensities, and $d(x, s)$ and $d(x', s')$ are estimated depth values. $c(x, s)$ and $c(x', s')$ are the corresponding certainty levels. If both $d(x, s)$ and $d(x', s')$ are perfectly correct, and if the surface is Lambertion, the distance $\Phi(x, s, s')$ should be zero.

We warp point (x, s) to different available views, and accumulate the distance. Then the accumulated distance is mapped to a penalty coefficient $p(x, s)$ for certainty level $c(x, s)$. The mapping function is defined as,

$$p(x, s) = \exp\left(\frac{1}{\beta}\left(\alpha - \frac{1}{||S||}\sum_{s' \in S} \Phi(x, s, s')\right)\right) \tag{2}$$

$$c'(x, s) = p(x, s) \times c(x, s) \tag{3}$$

where S is the set of all possible views. As the accumulated distance becomes larger, the penalty coefficient goes to zero; thus the refined certainty level is also close to zero. Otherwise the estimation is considered reliable, and its certainty level is almost unchanged. Parameters α and β control the shape of the penalty function, which are empirically set as 20 and 1 respectively in the experiments.

Because of possible occlusions, it is better to only accumulate matching distances on visible views rather than go over all the views. For example, in Figure 5 point A is occluded in some views. Therefore the accumulated matching distance of A is large, even though the correct depth value is assigned. It is better to only accumulate the matching distance for the visible views of A. Thus the temporal selection scheme in [16] is employed. Instead of including all views into set S in Equation 2, only the half with lower matching distance is considered, as shown in Figure 5.

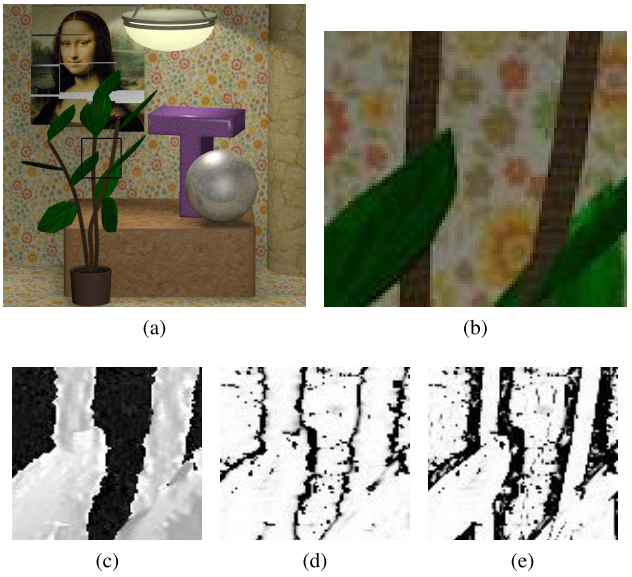


Fig. 6. A result of the certainty map refinement. (a) The center view of the light field. (b) A close-up of the center view. (c) The depth map from the structure tensor. (d) The original certainty map from the structure tensor. (e) The refined certainty map. The image (a) is from the data case "MonasRoom" in the Light Field Benchmark Dataset [15].

IV. Optimizing Depth Maps

A. Optimization by Solving a Linear System

As shown in Figure 3 and Figure 6, initial depth maps are not reliable and globally consistent. At this stage, we aim at getting an optimized depth map from the initial depth map and the corresponding confidence map.

We write the energy function in a matrix form,

$$J(\mathbf{d}) = \mathbf{d}^T L\mathbf{d} + \lambda(\mathbf{d} - \tilde{\mathbf{d}})^T C(\mathbf{d} - \tilde{\mathbf{d}}), \tag{4}$$

where $\mathbf{d}$ and $\tilde{\mathbf{d}}$ are $N \times 1$ vectors, represent optimal depth values and initial depth values respectively. N is the number of pixels in each view (i.e. $N = P \times Q$, if the resolution of the image is $P \times Q$). We want to find the optimal $\mathbf{d}$, which minimizes the energy function $J(\mathbf{d})$. In the first term, L is an affinity matrix, which enforces the points with similar colours to have similar depth values within a small neighbourhood. The second term is a data term, which makes the optimized result constrained by the initial depth estimations. C is a diagonal matrix, whose elements are confidence levels of corresponding pixels. Consequently, pixels with more reliable initial estimations are more tightly constrained by the data term. λ controls the weight of the data term.

To optimize $\mathbf{d}$, we can take the derivative of $J(\mathbf{d})$, and try to find the optimal $\mathbf{d}$ that makes the derivative zero. As a result, the cost function (4) can be minimized by solving a sparse linear system.

$$\frac{\partial J(\mathbf{d})}{\partial \mathbf{d}} = 2\mathbf{d}^T L + 2\lambda(\mathbf{d} - \tilde{\mathbf{d}})^T C = 0. \tag{5}$$

$$(L + \lambda C)\mathbf{d} = \lambda C\tilde{\mathbf{d}}. \tag{6}$$

By defining the affinity matrix L properly, we can make $L + \lambda C$ a symmetric positive definite matrix. Then this sparse linear system can be solved with the conjugate gradient method. Two formulations of affinity matrix are introduced, which are explained in detail in Section IV-B.

B. Affinity Matrix

A straightforward formulation of the affinity matrix L is

$$L = (I - W)^T (I - W). \tag{7}$$

Elements in W are defined as

$$W_{ij} = \begin{cases} \alpha_{ij} / \sum_{k \in Nbr(i)} \alpha_{ik} & \text{if } j \in Nbr(i) \\ 0 & \text{otherwise} \end{cases} \tag{8}$$

$$\alpha_{ij} = \max(\exp(\frac{-\Delta \mathbf{I_{ij}}}{\gamma}), \epsilon). \tag{9}$$

$Nbr(i)$ is the neighbourhood of pixel i; α_{ij} is a pair-wise weight based on the colour difference $\Delta \mathbf{I_{ij}}$ of neighbouring pixels i and j. γ and ϵ control the sharpness and the lower bound of the exponential function. With this formulation, the first term in Equation (4) is identical with the typical smoothness term in energy functions used in the area of stereo matching,

$$E_{smooth}(\mathbf{d}) = \sum_i \left(\mathbf{d_i} - \frac{\sum_{j \in N(i)} \alpha_{ij} \mathbf{d_j}}{\sum_{j \in N(i)} \alpha_{ij}} \right)^2. \tag{10}$$

To solve the linear system with the conjugate gradient method, the time complexity in each iteration is $O(Nr^2)$, where N is the total number of pixels and r is the width of the neighbourhood window. With larger window size, the depth information propagates from reliable pixels to other parts faster. However, in this way each iteration will take a longer time.

We also tried another formulation, known as the matting Laplacian matrix [17], for which a faster algorithm [18] with large window sizes is available. This matrix was originally proposed to solve matting problems, and later also used in haze removal, intrinsic images and colorization. In this formulation, the time complexity in each iteration is O(N), which is independent of the window size. As a result, a larger window size can be employed, which makes the number of iteration times much smaller and the overall time to solve the linear system much shorter.

The (i, j) element of this matting Laplacian matrix is defined as

$$\sum_{k|(i,j) \in \omega_k} (\delta_{ij} - \frac{1}{|\omega_k|}(1 + (\mathbf{I_i} - \mu_k)^T \times (\Sigma_k + \frac{\epsilon}{|\omega_k|} U)^{-1} (\mathbf{I_j} - \mu_k))), \tag{11}$$

where δ_{ij} is the Kronecker delta, μ_k and Σ_k are the mean and covariance matrix of the colours in a small local window ω_k, $|\omega_k|$ is the number of pixels in it, and U is a 3×3 identity matrix, and ϵ is a regularizing parameter. More information can be found in [17] and [18].

C. Segmentation

In order to apply different affinity matrices on different segments, and to keep the linear systems to a feasible size, we segment the reference image before the global optimization.

TABLE I

MEAN SQUARED ERRORS OF SELECTED DISPARITY ESTIMATION ALGORITHMS

	S.T. init.	Fast Denoising[8]	Constrained Opt.[21]	Global Opt.[8]	MRFs-32	MRFs-64	The Proposed Method
Buddha	0.81	0.57	0.55	0.62	0.6	0.63	0.64
Buddha2	1.22	0.87	0.87	0.89	0.58	0.67	0.53
Mona	1.15	0.9	0.82	0.93	0.73	0.73	0.64
StillLife	3.94	3.06	2.61	3.37	3.87	3.44	3.25
Horses	3.6	2.12	2.21	2.67	0.92	0.94	0.95
Medieval	1.69	1.15	1.1	1.24	2.33	2.21	2.1
Papillon	3.95	2.26	2.52	2.48	2.83	2.65	2.28
Average	**2.34**	**1.56**	**1.53**	**1.74**	**1.69**	**1.61**	**1.48**

[1] The values in the table show the average mean squared error in pixels times 100, i.e. a value of 0.81 means that the mean squared error in pixels is 0.0081.
[2] MRFs-32 is the result from MRFs method with 32 discrete depth levels, and MRFs-64 is with 64 discrete depth levels.
[3] "S.T. init" represents the initial estimations from the structure tensor without any optimization. The value in this table is generated by the 'cocolib'.

On smaller segments, the normal affinity matrix is employed, with which the window size is set as 9×9. On large segments, the matting Laplacian matrix is employed, with which the window size is set as 31×31.

The mean shift method [19] is employed for this purpose. This method does not ask for a priori knowledge about the number of segments. In addition, it is not too sensitive to the choice of parameters.

The mean shift method works in a joint domain of space and colour and thus groups pixels with similar colours and close spatial coordinates. This is suitable for general purposes in image segmentation. However, for the purpose of depth map optimization it can be improved, especially when the initial depth information is already available. For example, for a surface with limited depth variation in the space, the segmentation method may segment the surface into different pieces because of the texture variation on it. However, it is more desirable to group them into one segment since the depth values are continuous, in which the Laplacian matting matrix suffice. To this end, we modify the mean shift segmentation by taking the initial depth estimations into consideration. As shown in Figure 7, it is more desirable to group the segments on the background wall into one segment, since they are on the same depth level. It is more efficient to optimize one large piece than optimizing several some pieces separately.

The original mean shift method works via the following procedure [19]: First, a mean shift filter is applied on the image. Then pixels with similar filtered values are assigned to the same segments. At last, regions with too few pixels are eliminated.

To make it more suitable for our purpose, one extra step is added after the second step of the original mean shift method: If two neighbouring segments have similar average depth values, we consider them as on the same depth level and merge them. One example is shown in Figure 7, and more results are provided in Section V.

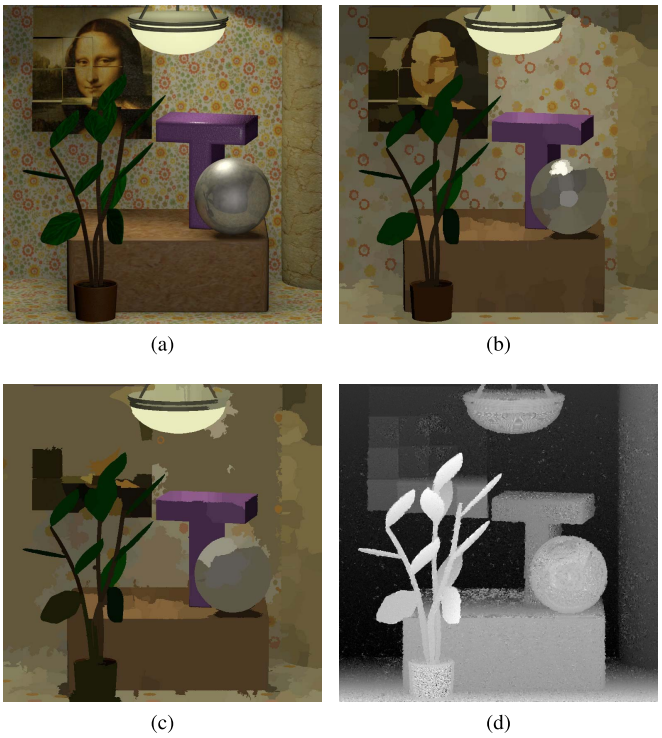


Fig. 7. An example of the depth-assisted segmentation. (a) One view of the light field; (b) Segmentation result via the mean shift; (c) Segmentation result with depth information; (d) The depth map from the initial estimation. The image (a) is from the data case "MonasRoom" in the Light Field Benchmark Dataset [15].

V. EXPERIMENTAL RESULTS

We test the proposed framework on two datasets, the Light Field Benchmark Dataset (LFBD) [15] and the Stanford light field archive [20]. The data in both datasets are four dimensional, densely sampled from 9×9 or 17×17 views. The synthesized data from LFBD comes with ground truth.

We compared our results with results reported in the Light Field Benchmark [15], including the fast denoising method in [8], global optimization with function lifting [7] and the constrained global optimization [21]. The comparative results are all generated from the open-source code 'cocolib'.

We also compare our method with a method based on the MRFs model as described in [2]. A modification is made on the data term to fit our problem. In stereo matching, the data term describes to what extent the pair of matching pixels differ with each other. In the case of optimizing the depth maps from light fields, the data term here measures to what extent the optimal result differs from the initial estimation. The certainty level is used as a weight. Then

$$E_{\text{data}} = \sum_{(x,y)} c(x,y)|d(x,y) - \tilde{d}(x,y)|, \tag{12}$$

TABLE II
PERCENTAGE OF WRONG DEPTH VALUES

	Fast Denoising[8]			Global Opt.[8]			MRFs-64			The Proposed Method		
	>5%	>1%	>0.1%	>5%	>1%	>0.1%	>5%	>1%	>0.1%	>5%	>1%	>0.1%
Buddha1	0.64	1.22	27.1	0.09	1.13	32.4	0.12	1.3	27.7	0.02	1.62	15.04
Buddha2	0.13	3.29	73.8	0.15	3.22	85.5	0.13	1.56	60.3	0.02	2.3	58.7
MonasRoom	0.4	2.18	37.7	0.44	2.18	45.3	0.34	1.6	44.1	0.07	2.45	24.4
StillLife	0	0.99	21.5	0.001	1.12	17	0	1.43	22.6	0	1.19	17.7
Horses	0.07	4.86	76.9	0.27	4.79	72.4	0	1.6	41.8	0	2	31.2
Medieval	0.29	2.03	72.7	0.31	2.22	70.74	0	0	2.38	0	0	2.3
Papillon	0	1.7	61.3	0.004	1.86	58.7	0.01	2	40.1	0	2.69	37.9
Average	**0.22**	**2.32**	**53**	**0.18**	**2.36**	**54.58**	**0.086**	**1.36**	**34.14**	**0.016**	**1.75**	**26.75**

[1] The values in the table is the number of missed pixels whose relative depth error is greater than a certain value over the total number of pixels.

TABLE III
PERCENTAGE OF WRONG DISPARITY ESTIMATIONS

	Fast Denoising [8]			Global Opt. [8]			MRFs-64			The Proposed Method		
	>1	>0.5	>0.1	>1	>0.5	>0.1	>1	>0.5	>0.1	>1	>0.5	>0.1
Buddha1	0.13	0.52	2.45	0.18	0.56	2.03	0.24	0.48	2.55	0.056	0.6	4.57
Buddha2	0.07	0.55	8.01	0.1	0.57	7.3	0.09	0.58	3.72	0.013	0.43	5.32
MonasRoom	0.29	0.86	3.95	0.32	0.86	4.33	0.31	0.63	3.97	0.033	0.85	5
StillLife	0.57	1.58	14.55	0.67	1.82	10.83	1.11	2.58	15.48	0.65	2.1	13.1
Horses	0.28	1.59	29.35	0.59	2.2	24.8	0.13	1.34	7.33	0.058	0.88	8.89
Medieval	0.28	0.89	9.97	0.31	0.93	8.65	1.29	2.24	5.56	0.57	2	7.23
Papillon	0.67	1.38	15.34	0.82	1.52	13.7	1.48	1.87	9.36	0.89	2.1	10.92
Average	**0.33**	**1.05**	**11.95**	**0.43**	**1.21**	**10.23**	**0.66**	**1.39**	**6.85**	**0.32**	**1.28**	**7.86**

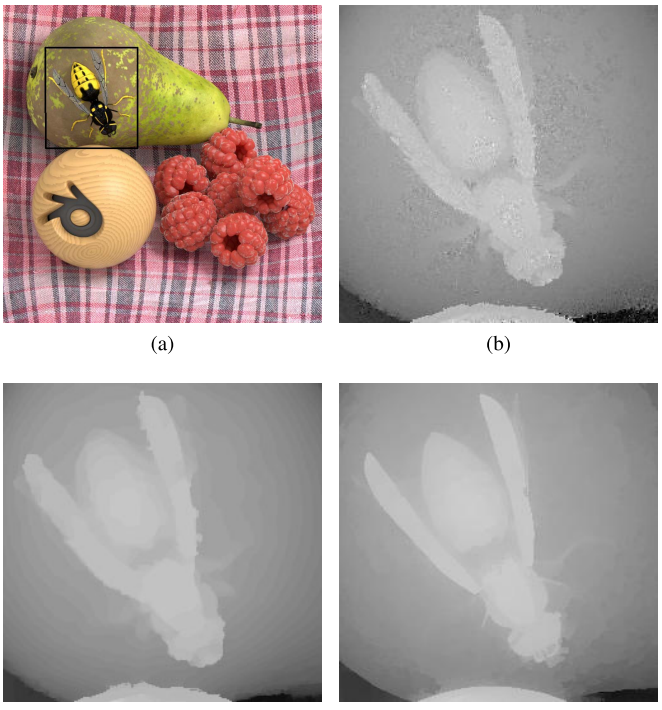
[1] The values in the table is the number of missed pixels whose relative disparity error is greater than a certain value over the total number of pixels.

where $c(x, y)$ is the certainty level, $d(x, y)$ is the optimal depth value, and $\tilde{d}(x, y)$ is the initial depth value. We made use of the code published with paper [2]. A graph-cut optimization [22] is used to minimize the energy function. Several different combination of parameters and data terms are tried, and one with best performance is chosen.

The comparison is done with three metrics: 1) the average mean squared disparity errors (MSE) in pixels; 2) the percentage of missed pixels for which the relative depth error is greater than a certain value; 3) the percentage of missed pixels for which the relative disparity error is greater than a certain value. The conversion between depth and disparity is done with the conversion formula and parameters provided by the dataset [15]. Comparisons with the three metrics are shown in Table I, Table II and Table III respectively. We only showed seven datasets from the LFBD, because the other five datasets will run out of memory using the published code 'cocolib'.

As shown in the three tables, our method outperforms the other methods in all the metrics. Moreover, although all the methods take the structure tensor as the initial estimation, the results generated with 'cocolib' is better than the implementation in our method (VIGRA library). To get a fair comparison, instead of the overall result we compared the performance improvement made by different optimization methods, in Table IV. In this case, our optimization method clearly gives a much larger improvement over the initial estimations. Due to space limitations, only the average results over all the seven datasets are shown here.

As shown in Table V, our method runs much faster compared with the global optimization method and the MRFs method. Unlike these discrete methods, the running time of our method is independent of the number of discretized levels. The denoising method in [8] is very fast, and also



(a) (b) (c) (d)

Fig. 8. Close-up results of "StillLife" from the Light Field Benchmark Dataset [15]. (a) The center view of the light field; (b) A close-up of the initial depth estimation; (c) A close-up of the result of the functional lifting; (d) A close-up of the result of the proposed method.

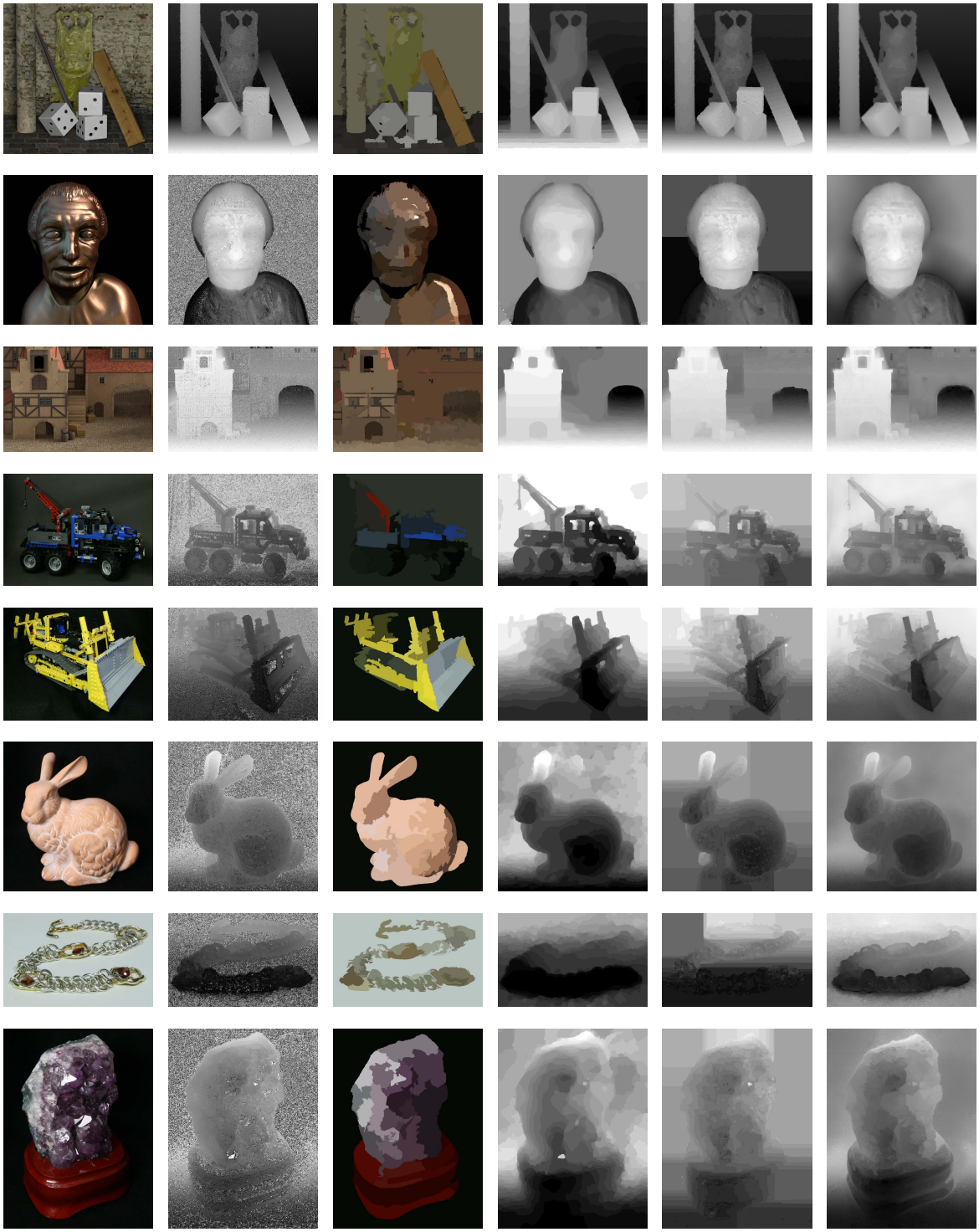


Fig. 9. Selected results. From the left to the right, the columns are the center views of the 4D light fields, the initial estimations, segmentation results, results from the global optimization with functional lifting [8], results from the MRFs and the results from the proposed method. The three rows on the top are from the Light Field Benchmark Dataset [15], and the others are from Stanford light field archive [20].

generates continuous results. However, it employs a much simpler model with L-1 norm and only enforces point-wise smoothness. Rather than an optimization step, it's more like a denoising filter. Moreover, our method achieves higher accuracy than the fast denoising method.

The running time in Table V is obtained on a computer with an Intel Core i7 CPU with 12G RAM and a NVIDIA GeForce GTX 570 graphic card. All the methods are implemented in C/C++. The 'cocolib' is implemented in parallel with CUDA.

The Stanford light field archive does not have ground truth information, so we compare the results visually in Figure 9, as well as some test cases from LFBD. A close-up comparison is shown in Figure 8. One can see that the proposed method

TABLE IV
AVERAGE PERFORMANCE OF SELECTED ALGORITHMS

		MSE	Disparity Error			Depth Error		
			>1	>0.5	>0.1	>5%	>1%	0.10%
S.T.init. ('cocolib')	-	2.34	0.55	1.62	14.04	0.25	3.63	52.9
Fast denoising[8]	-	1.56	0.33	1.05	11.95	0.22	2.32	53
	Improvement	33.33%	40.00%	35.19%	14.89%	12.00%	36.09%	-0.19%
Global Opt.[8]	-	1.74	0.43	1.21	10.23	0.18	2.36	54.58
	Improvement	25.64%	21.82%	25.31%	27.14%	28.00%	34.99%	-3.18%
S.T. init. (VIGRA)	-	4.6	0.86	3.41	23.9	0.2	4.99	46.45
MRFs-64	-	1.61	0.66	1.39	6.85	0.086	1.36	34.14
	Improvement	65.00%	23.26%	59.24%	71.34%	57.00%	72.75%	26.50%
The Proposed Method	-	1.48	0.32	1.28	7.86	0.016	1.75	26.75
	Improvement	67.83%	62.79%	62.46%	67.11%	92.00%	64.93%	42.41%

[1] This table shows the average mean squared disparity error times 100, as well as the percentage of missed pixels with disparity or depth error larger than a given quantity.
[2] The "improvement" rows for each algorithm is the relative performance improvement over the initial structure tensor result.

TABLE V
EFFICIENCY OF DIFFERENT OPTIMIZATION METHODS

	Resolution	Global Opt.-32[8]	Global Opt.-64[8]	MRFs-32	MRFs-64	The Proposed Method
Buddha1	9*9*768*768	461	917	296	1027	199
Buddha2	9*9*768*768	455	898	780	1278	203
MonasRoom	9*9*768*768	438	869	334	966	211
StillLife	9*9*768*768	474	886	214	707	338
Horses	9*9*1024*576	551	1075	212	957	252
Medieval	9*9*1024*760	590	1154	337	834	320
Papillon	9*9*768*768	465	888	686	841	182
Average	**-**	**491**	**955**	**408**	**944**	**244**

[1] The running time is in seconds.
[2] Both the global optimization in [8] and the MRFs method are tested with 32 and 64 discrete depth levels.
[3] The number shown in this table is for the optimization step only.

generates more accurate and detailed results. The global optimization method with functional lifting [8] can keep accurate and sharp boundaries, but tends to erase details. The MRFs model keeps relatively more details, but still has noticeable staircase effects.

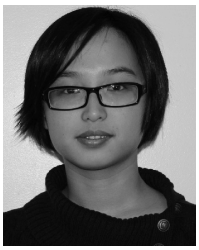
VI. CONCLUSION

In this paper, we propose a novel framework to reconstruct continuous depth maps from 4D light fields. A refinement of the initial depth estimation is introduced by checking colour consistency between different views. Based on the initial depth estimations, we construct a sparse linear system, in which two different affinity matrices are employed. In order to apply different affinity matrices on different patches, a depth-assisted segmentation method is also proposed. We compared our method to the state-of-the-art work. It generates more accurate depth values and tends to keep more details. As the running time of our method is independent of the depth levels, to achieve a similar level of smoothness and details, our method is much faster compared with other discrete optimization methods.

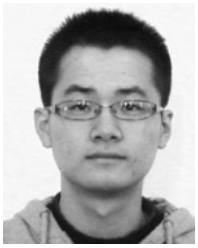
REFERENCES

[1] D. Scharstein and R. Szeliski, "A taxonomy and evaluation of dense two-frame stereo correspondence algorithms," *Int. J. Comput. Vis.*, vol. 47, nos. 1–3, pp. 7–42, 2002.
[2] R. Szeliski *et al.*, "A comparative study of energy minimization methods for Markov random fields with smoothness-based priors," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 30, no. 6, pp. 1068–1080, Jun. 2008.
[3] T. Pock, D. Cremers, H. Bischof, and A. Chambolle, "Global solutions of variational models with convex regularization," *SIAM J. Imag. Sci.*, vol. 3, no. 4, pp. 1122–1145, 2010.
[4] C. Kim, H. Zimmer, Y. Pritch, A. Sorkine-Hornung, and M. Gross, "Scene reconstruction from high spatio-angular resolution light fields," *ACM Trans. Graph.*, vol. 32, no. 4, 2013, Art. ID 73.
[5] M. W. Tao, S. Hadap, J. Malik, and R. Ramamoorthi, "Depth from combining defocus and correspondence using light-field cameras," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Dec. 2013, pp. 673–680.
[6] M.-J. Kim, T.-H. Oh, and I. S. Kweon, "Cost-aware depth map estimation for Lytro camera," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Oct. 2014, pp. 36–40.
[7] S. Wanner and B. Goldluecke, "Globally consistent depth labeling of 4D light fields," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2012, pp. 41–48.
[8] S. Wanner and B. Goldluecke, "Variational light field analysis for disparity estimation and super-resolution," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 36, no. 3, pp. 606–619, Mar. 2014.
[9] J. Li and Z.-N. Li, "Continuous depth map reconstruction from light fields," in *Proc. IEEE Conf. Multimedia Expo*, Jul. 2013, pp. 1–6.
[10] M. Levoy and P. Hanrahan, "Light field rendering," in *Proc. 23rd Annu. Conf. Comput. Graph. Interact. Techn.*, 1996, pp. 31–42.
[11] R. C. Bolles, H. H. Baker, and D. H. Marimont, "Epipolar-plane image analysis: An approach to determining structure from motion," *Int. J. Comput. Vis.*, vol. 1, no. 1, pp. 7–55, 1987.
[12] J. Bigün and G. H. Granlund, "Optimal orientation detection of linear symmetry," in *Proc. IEEE 1st Int. Conf. Comput. Vis.*, Jun. 1987, pp. 433–438.
[13] E. Strekalovskiy and D. Cremers, "Generalized ordering constraints for multilabel optimization," in *Proc. IEEE Int. Conf. Comput. Vis.*, Nov. 2011, pp. 2619–2626.
[14] G. Zhang, J. Jia, T.-S. Wong, and H. Bao, "Recovering consistent video depth maps via bundle optimization," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2008, pp. 1–8.

[15] S. Wanner, S. Meister, and B. Goldluecke, "Datasets and benchmarks for densely sampled 4D light fields," in *Proc. Int. Workshop Vis., Modelling Vis. (VMV)*, 2013.
[16] S. Kang and R. Szeliski, "Extracting view-dependent depth maps from a collection of images," *Int. J. Comput. Vis.*, vol. 58, no. 2, pp. 139–163, 2004.
[17] A. Levin, D. Lischinski, and Y. Weiss, "A closed-form solution to natural image matting," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 30, no. 2, pp. 228–242, Feb. 2008.
[18] K. He, J. Sun, and X. Tang, "Fast matting using large kernel matting Laplacian matrices," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2010, pp. 2165–2172.
[19] D. Comaniciu and P. Meer, "Mean shift: A robust approach toward feature space analysis," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 5, pp. 603–619, May 2002.
[20] *Stanford (New) Light Field Archive*. [Online]. Available: http://lightfield.stanford.edu/lfs.html, accessed Dec. 4, 2013.
[21] B. Goldluecke and S. Wanner, "The variational structure of disparity and regularization of 4D light fields," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2013, pp. 1003–1010.
[22] Y. Boykov, O. Veksler, and R. Zabih, "Fast approximate energy minimization via graph cuts," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 23, no. 11, pp. 1222–1239, Nov. 2001.



Jianqiao Li received the B.E. degree in digital media from Zhejiang University, China, in 2011, and the M.Sc. degree in computer science from Simon Fraser University, in 2013. She was with the Vision and Media Laboratory, Simon Fraser University, under the supervision of Prof. Z.-N. Li, during her graduate study. Her research interests include computer vision, multimedia, and machine learning.



Minlong Lu received the B.E. degree in computer science from Zhejiang University, China, in 2011. He is currently pursuing the Ph.D. degree with the Graduate Dual Degree Program jointly developed by Simon Fraser University and Zhejiang University under the supervision of Prof. Z.-N. Li and Prof. G. Pan. His research interests include computer vision and pattern recognition.

Ze-Nian Li received the bachelor's degree in electrical engineering from the University of Science and Technology of China, and the M.Sc. and Ph.D. degrees in computer sciences from the University of Wisconsin–Madison under the supervision of the late Prof. L. Uhr. He is currently a Professor with the School of Computing Science, Simon Fraser University, BC, Canada. He is the Co-Director of the Vision and Media Laboratory. He has co-authored a book entitled *Fundamentals of Multimedia—2nd Edition* (Springer, 2014). He has authored over 150 refereed papers in journals and conference proceedings. His current research interests include computer vision, multimedia, pattern recognition, and artificial intelligence.