1-1-2013

Scene-adaptive configuration of two cameras using the correspondence field function

F Safaei

University of Wollongong, farzad@uow.edu.au

Payam Mokhtarian

University of Wollongong, payam@uow.edu.au

H Shidanshidi

University of Wollongong, hooman@uow.edu.au

W Li

University of Wollongong, wanqing@uow.edu.au

Mohammad-Reza Namazi-Rad

University of Wollongong, mrad@uow.edu.au

See next page for additional authors

Follow this and additional works at: <https://ro.uow.edu.au/smartpapers>

Part of the [Engineering Commons](#), and the [Physical Sciences and Mathematics Commons](#)

Scene-adaptive configuration of two cameras using the correspondence field function

Abstract

In a free viewpoint video system, the scene is captured by a number of cameras and it would be desirable to optimize the configuration of cameras, such as their location or orientation, to improve the rendering quality. This paper introduces a mathematical representation of the multi-camera geometry, called the correspondence field (CF), which can be used to quantify the suitability of a camera configuration for a given arrangement of objects in the scene. The correspondence field describes the spatial topology of the intersecting rays of cameras, arranged as a number of layers or surfaces in the field of view of cameras. The paper derives the topology of CF for certain camera arrangements and analyzes the impact of changes in camera location or orientation on this topology. It demonstrates that CF can be used to find the optimum camera configuration for a given objective. It also presents simulation results of this method using our light field simulator.

Keywords

light field acquisition, free viewpoint video, multi-camera systems, camera configuration

Disciplines

Engineering | Physical Sciences and Mathematics

Publication Details

Safaei, F., Mokhtarian Dehkordi, P., Shidanshidi, H., Li, W., Namazi-Rad, M. & Mousavinia, A. (2013). Scene-adaptive configuration of two cameras using the correspondence field function. IEEE International Conference on Multimedia and Expo (pp. 1-6). IEEE Xplore: IEEE.

Authors

F Safaei, Payam Mokhtarian, H Shidanshidi, W Li, Mohammad-Reza Namazi-Rad, and A Mousavinia

SCENE-ADAPTIVE CONFIGURATION OF TWO CAMERAS USING THE CORRESPONDENCE FIELD FUNCTION

F. Safaei^{*}, P. Mokhtarian[#], H. Shidanshidi^{*}, W. Li^{*}, M. Namazi-Rad[#], A. Mousavinia[%]

^{*}ICT Research Institute, University of Wollongong, Australia

[#]National Institute for Applied Statistics Research Australia, University of Wollongong, Australia

[%]K. N. Toosi University, Tehran, Iran

{farzad, payam, hooman, wanqing, mrad}@uow.edu.au, moosavie@eetd.kntu.ac.ir

ABSTRACT

In a free viewpoint video system, the scene is captured by a number of cameras and it would be desirable to optimize the configuration of cameras, such as their location or orientation, to improve the rendering quality. This paper introduces a mathematical representation of the multi-camera geometry, called the correspondence field (CF), which can be used to quantify the suitability of a camera configuration for a given arrangement of objects in the scene. The correspondence field describes the spatial topology of the intersecting rays of cameras, arranged as a number of layers or surfaces in the field of view of cameras. The paper derives the topology of CF for certain camera arrangements and analyzes the impact of changes in camera location or orientation on this topology. It demonstrates that CF can be used to find the optimum camera configuration for a given objective. It also presents simulation results of this method using our light field simulator.

Index Terms— light field acquisition, free viewpoint video, multi-camera systems, camera configuration

1. INTRODUCTION

A free viewpoint video (FVV) system aims to create arbitrary views of a scene from the known samples obtained by a number of cameras. The FVV system is comprised of three components: (i) the *acquisition* component, which is responsible for the sampling of the light field [1-2] using a given configuration of cameras; (ii) *compression and transmission* of the captured or processed light field; and (iii) the *rendering* of the unknown view/ray. This paper focuses on the optimization of the acquisition component.

Theoretically, for perfect reconstruction of views under the assumption of a band-limited signal and linear interpolation of samples, the scene has to be sampled at the Nyquist density [3-5]. However, in most real-world scenarios, this high sampling density is not practical and, therefore, the rendering algorithm must deal with an under-sampled signal during reconstruction.

In [6] it is shown that by utilizing the concept of effective sampling density (ESD) [7], the trade-off between the depth information accuracy, the required number of cameras, and the desired rendering quality can be quantified. This trade-off is exploited to determine the minimum density of cameras required to capture the scene for a desired output quality objective. This result is derived for a regular camera grid, which is fixed in time.

In a practical FVV system, the total number of cameras is likely to be fixed for the duration of acquisition. Nevertheless, it may be feasible to alter the *configuration* of cameras to some degree as a result of changes in the scene, such as movements of players during a football match. For example, the array of cameras could be mounted on robotic platforms or supporting rails with the ability to shift their position with some degree of freedom. Each camera may also possess the ability to pan or zoom in response to commands from the control algorithm.

To our knowledge, all the existing research on light field acquisition, have assumed a fixed camera grid in time such as layered light field [8], surface light field [9], scam light field [10], pop-up light field [11], all-in-focused light field [12] and dynamic reparameterized light field [13].

To be able to optimize the camera configuration dynamically, a mathematical framework is needed to quantify the suitability of a given camera configuration for the scene. Multiple camera systems are often modeled by the epipolar geometry [14]. The current formulations of this geometry are commonly tailored to suit an image processing objective, such as correspondence matching for depth estimation at the individual pixel level. For our purpose, however, a more ‘holistic’ depiction of multi-camera geometry aimed at characterizing the *spatial* relationship between the arrangement of cameras and the objects in the scene is required.

This paper introduces an alternative representation of multi-camera geometry, which is referred to as the *correspondence field* (CF) of cameras, $f: \mathbb{R}^3 \rightarrow \mathbb{R}^3$. This function associates a vector with every point of space that is an intersection point of rays from two or more cameras. These points are referred to as *n-points*, where *n* represents

the number of cameras whose rays have intersected at this point. The vector field is the set of tangent vectors to the *correspondence n -surface* passing through these n -points. In essence, the judicious combination of n -points and their respective tangent vectors would define the n -surfaces.

In practice, each camera pixel represents the average captured light intensity of a volume of space and therefore, the ‘resolution’ of function f is limited by the cameras. In addition, sometimes it is desirable to evaluate CF at a coarser resolution by aggregating a number of neighboring rays/pixels.

The correspondence field of multiple cameras is only dependent on the camera configuration, and *not* the scene. In fact, given practical limits on the accuracy of camera re-arrangements, the number of possible configurations may be finite. For example, the translational or rotational movements of cameras may be restricted to a certain range and accuracy due to hardware limitations. This ability to calculate the correspondence field of various camera configurations during a pre-processing stage would, therefore, provide a significant practical advantage. As the scene changes in time, the acquisition system can alter its camera arrangement, perhaps by evaluating the suitability of a number of pre-calculated configurations.

The contributions of this paper are as follows:

- It introduces the correspondence field function and illustrates the properties of this function using a number of possible camera configurations for two cameras.
- It presents a mathematical framework for calculating the topology of CF for certain camera configurations.
- It demonstrates how CF could be used to find the optimum camera configuration in response to changes in the scene and for a given objective. The simulation results of this method are also presented.

The correspondence field provides other practical benefits in addition to camera configuration optimization. During the discussions in this paper some of these benefits will be alluded to, although their full exploration will have to be deferred to future publications.

2. THE CORRESPONDENCE FIELD

2.1. Description and terminology

A Figure 1 shows the intersection points of a number of rays associated with two cameras that are aimed at a scene. For the remainder of this paper, geometrically rectified and calibrated pinhole cameras are assumed (e.g. see [14-15]). The figure only shows the top view of a plane associated with one row of pixels and for clarity only a very small subset of rays (or aggregated rays when coarser resolution is desired) is shown.

An intersection point of two rays is called a *2-Point* and is shown as a red dot in the Figure. While it is called and shown as a point, it is understood that each pixel (or group thereof) captures the light from a contiguous volume of

space. The size of this volume increases with distance from the camera plane resulting in a corresponding decrease in the resolution of CF, as expected. CF associates a vector with each 2-point. The direction of this vector is tangential to the *2-surface* created by judiciously selecting a set of these 2-points. For example, these 2-points could be arranged as multiple layers associated with a given disparity with increasing distance from the camera plane. To emphasize again, given that each 2-point is a volume, each 2-surface has thickness and the above surfaces resemble multiple layers of an onion. By moving away from the cameras, the surfaces become larger and thicker, signifying a reduction in the resolution of the system.

Figure 1. *2-points* and *2-surfaces* associated with the correspondence field of two cameras

When there are n cameras in the acquisition system, there may also be 3-points, 4-points, ..., and n -points in the CF, which are the intersection points of rays associated with 3, 4, ..., n cameras respectively. This extension to more cameras is left for future publications.

Figure 2. *2-surfaces* and the *Spatial Extent* of CF

Figure 2 shows a number of 2-surfaces for a given configuration of two cameras. There are no 2-points or

surfaces at depths closer than d_{min} or larger than d_{max} . In addition, the surfaces are confined within a region in the field of view of cameras. Let us refer to the volume of space that contains n -surfaces as the *Spatial Extent* (SE) of the correspondence field. Objects that lie outside the SE are essentially seen only by a single camera or not seen at all. By rearranging the cameras, it is possible to alter the configuration of 2-surfaces and the spatial extent as shown in Figure 3.

Figure 3. The impact of camera rotation on CF

2.2. Derivation of CF topology

Let us refer to the spatial location and arrangement of n -points and n -surfaces in a CF as its *topology*. It is clear from the above descriptions that this topology is a function of configuration of cameras. In this section, a mathematical framework to find the topology for some camera arrangements is presented.

Consider two cameras with their centers at locations a_1 and a_2 along the y axis both aiming towards the positive direction of x axis (Figure 2). Assume that the field of view of both is θ and consider the plane associated with a row of

M pixels. Let $\delta = \frac{2 \tan(\frac{\theta}{2})}{M-1}$. Then the equation of the i^{th} ray of the first camera can be written as $y = a_1 + b_i x$ where $b_i = \tan(\frac{\theta}{2}) - (i-1)\delta$ and $i \in \{1, 2, \dots, M\}$. Similarly, the equation of the j^{th} ray of the second camera could be written and the intersection point of the two rays, denoted by $p_{ij} \equiv (x_{ij}, y_{ij})$ can be derived as:

$$x_{ij} = \frac{a_1 - a_2}{b_j - b_i}, y_{ij} = \frac{a_1 b_j - a_2 b_i}{b_j - b_i}, i \neq j, x_{ij} \geq 0 \quad (1)$$

The condition $i \neq j$ is for the intersection to take place (rays are not parallel) and $x_{ij} \geq 0$ signifies that only the intersection points in front of the cameras are relevant. The set of 2-points for this CF is therefore:

$$\mathcal{P}_2 = \{p_{ij} : i \neq j; i, j = 1, 2, \dots, M\} \quad (2)$$

There can be many ways to arrange these 2-points into 2-surfaces depending on the specific purpose that CF is

being used. An intuitive arrangement of 2-surfaces used in this paper would be based on disparity. In other words, the furthest surface from the cameras will have a disparity of one pixel, the next one a disparity of two, and so on up to a disparity of $M-1$ for the closest 2-surface. Let s_m represent the m^{th} 2-surface, then:

$$s_m = \{p_{ij} : |i - j| = m; m = 1, 2, \dots, M-1\}, \quad (3)$$

and the set of 2-surfaces for this CF would be $\mathcal{S}_2 = \{s_m\}$

The spatial extent of p_{ij} , that is, the volume of space that corresponds to this 2-point, is usually an important parameter signifying the coarseness of system resolution at this particular location. A useful measure of this would be the distance between the two adjacent 2-surfaces at this point. In this paper, the mean distance between a given 2-point and its nearest 2-points on the adjacent 2-surface is used for this purpose.

Let us denote the Euclidean distance between p_{ij} and p_{kl} as $d(ij, kl)$. Assume that p_{ij} is on s_m (i.e., $|i - j| = m$), and let w_{ij} be the desired measure of distance between this point and its nearest 2-points on s_{m+1} . (Note that s_{m+1} is closer to cameras than s_m , as m represents disparity). Along the i^{th} ray, the 2-point $p_{i+1,j}$ and along the j^{th} ray, the 2-point $p_{i,j+1}$ belong to s_{m+1} . Then w_{ij} is defined as the mean Euclidean distance between p_{ij} and these 2-points:

$$w_{ij} \equiv \frac{d(ij, (i+1)j) + d(ij, i(j+1))}{2}, 1 \leq |i - j| < M-1 \quad (4)$$

When p_{ij} is on the closest 2-surface to cameras, w_{ij} is defined as the distance to the camera plane, that is:

$$w_{ij} \equiv x_{ij}, \text{ provided } |i - j| = M-1 \quad (5)$$

So far, the 2-surfaces have been characterized as a set of discrete 2-points. In many situations, including the optimization model of this paper, this may be sufficient. However, each 2-surface is in fact continuous and it may be desirable to obtain an approximate expression for this surface. One approach would be to calculate the tangent of this surface at each 2-point and use this tangent to obtain a first order approximation of the curve between the two immediate 2-points. Let us consider 2-surface s_m and assume that point p_{ij} is on this surface and $i - j = m$. The next 2-point on this surface would be $p_{i+1,j+1}$ (or $p_{i-1,j-1}$). Let u be a continuous variable within the interval $[i, i+1]$. If the ray i from the first camera sweeps this interval, b_u would be the slopes of these rays covering the space between the two adjacent i and $i+1$ rays. Likewise, b_{u+m} would be the slopes of rays covering the interval $[j, j+1]$ from the second camera. The surface between p_{ij} and $p_{i+1,j+1}$ can then be characterized as a vector function using parameter u as follows:

$$\mathbf{s}_m(u) = f_x(u)\mathbf{i} + f_y(u)\mathbf{j}, \quad (6)$$

Where we use the boldface notation to represent vector quantities, $\mathbf{i}$ and $\mathbf{j}$ denote unit vectors in x and y directions respectively, $f_x(u) = \frac{a_1 - a_2}{b_{u+m} - b_u}$ and $f_y(u) = \frac{a_1 b_{u+m} - a_2 b_u}{b_{u+m} - b_u}$, and $i \leq u \leq i+1$.

The tangent to this curve at point p_{ij} ($u = i$) will be:

$$\mathbf{T}(i) = f'_x(i)\mathbf{i} + f'_y(i)\mathbf{j}, \quad (7)$$

where differentiation is with respect to u . If the closed form expressions for the derivatives are not obtainable and the density of 2-points is sufficiently large, this tangent can be approximated by the secant vector between the two points:

$$\mathbf{T}(i) \approx \mathbf{s}_m(i+1) - \mathbf{s}_m(i) \quad (8)$$

2.3. Changing the camera configuration

The original position and orientation of cameras in the above model results in planar 2-surfaces as shown in Figure 2, which is often used in stereo-matching studies. For this configuration, it is straightforward to calculate the depth of various 2-surfaces and the spatial extent of CF. It is also easy to show that the tangent to 2-surfaces at every point is vertical. In particular, denoting the distance between the two cameras by $k \equiv a_1 - a_2$, the depth of the m^{th} surface would be $d_m = k/m\delta$. Assuming that the motion of cameras is constrained along the y axis, then any translational movement would only alter k . This will maintain the planar topology but scale the depth of 2-surfaces accordingly because $d_m \propto k$.

Rotation of cameras (pan), however, results in non-planar 2-surfaces as depicted in Figure 3. Rotation of a camera around its axis modifies the slope of rays. For example, a clockwise rotation of camera 1 by α leads to:

$$b_i^t = \tan(\tan^{-1} b_i - \alpha) \quad (9)$$

This new slope has to be incorporated into expressions for the coordinates of p_{ij} , for example, $x_{ij} = \frac{a_1 - a_2}{b_j - b_i^t}$ and $y_{ij} = a_1 + b_i^t x_{ij}$, provided $b_i^t \neq b_j$.

3. ADAPTATION OF THE CF TOPOLOGY

The CF topology is an informative representation of a given camera configuration. Assume that some information about the extent of the scene and the size of the objects of interest were available. For example, in the case of capturing a football game, this information could be obtained using depth cameras or by processing a top view obtained by a roof camera. Then, the CF topology could be tailored to match the scene based on a desired objective.

In this Section, an optimization problem is formulated to demonstrate the utility of CF for camera reconfiguration.

3.1. Problem formulation

Let us assume that the number and depth of objects in the scene were known and the objective was to choose a camera configuration (that is, CF topology) which provides the maximum ‘coverage’ of the objects with the best possible resolution. This essentially means that the correspondence field is adapted so that most or all the objects are contained:

- within the *spatial extent* of CF, to ensure that information from multiple cameras could be used to enhance the rendering quality and depth estimation;
- within CF surfaces with minimal thickness to improve resolution and the sampling density.

To formulate this problem, let us define the following:

Decision variables: Assume there are H possible camera configurations and our aim is to choose one of these. Let x_h denote a binary decision variable as defined below:

$$x_h = \begin{cases} 1 & \text{if configuration } h \text{ is selected} \\ 0 & \text{otherwise} \end{cases} \quad (10)$$

Clearly, only a single configuration can be selected, hence:

$$\sum_{h=1}^H x_h = 1 \quad (11)$$

Known parameters: For every possible configuration of cameras, it is possible to derive the topology of CF using the procedures outlined in the previous section. For a concrete example, let us consider only two cameras that are mounted on a straight rail and can be moved within a prescribed range and accuracy. Also, assume that cameras have some ability to change their orientation (pan). Using the relationships derived in the previous section, for each configuration, the set of 2-points and the associated width of layers (p_{ij} and w_{ij}) can be computed.

Assume there are B objects in the scene and the depth of these objects are known. It would then be possible to precalculate the location of the nearest 2-point to each object for each configuration. Let us use a binary indicator function to represent the closest 2-point to object b for configuration h as follows:

$$L_{ij}^{bh} = \begin{cases} 1 & \text{if } p_{ij} \text{ in config } h \text{ is the closest point to object } b \\ 0 & \text{otherwise} \end{cases} \quad (12)$$

It is also required to identify objects that are outside the spatial extent of CF. Let W be the maximum width of layers in the system, i.e., $W = \max_{i,j} w_{ij}$. If the nearest point to the object is further away than cW , where c is a tolerance factor, then the object can be considered to be outside the spatial extent. This is signified by the following:

$$R^{bh} = \begin{cases} 1 & \text{if the distance to closest point} > cW \\ 0 & \text{otherwise} \end{cases} \quad (13)$$

Objective function: The objective of the optimization would be to minimize z as defined below subject to the constraint specified by (11):

$$\text{Minimize } z = \sum_{h=1}^H \sum_{i=1}^M \sum_{j=1}^M \sum_{b=1}^B x_h (L_{ij}^{bh} w_{ij} + R^{bh} G) \quad (14)$$

G is a sufficiently large number in Equation (14) to penalize objects being outside the spatial extent of CF. This problem is a variation of the *knapsack* problem. For moderate size of H , efficient algorithms exist to obtain the solution in reasonable time [16].

3.2. Extension to more cameras

For the case of more than two cameras, it would be possible to extend the above formulation. For example, larger cost could be assigned to situations when the object is contained within 2-surfaces as opposed to 3-surfaces so that the optimization would favour placing objects within the range of more cameras.

3.3. Selecting the resolution of CF for computation

Although it is possible to derive the CF at the full resolution, i.e., calculating p_{ij} for every camera pixel, much of the overall topology of CF could be estimated rather accurately even when CF is calculated at a coarser resolution. Further discussion on this topic is deferred to future publications.

4. SIMULATION RESULTS

The aforementioned optimization was implemented in a FVV simulator described in [17]. The 3D model of a chess board was selected to simulate a complex scene with many objects and the possibility of re-arrangements and occlusions. To assess the impact of changes in the scene, two arrangements of chess pieces was selected: (i) the starting position of the board, and (ii) a position where most pieces are gathered in one part of the scene. See Figures 4(a) and 5(a) on the last page.

Two cameras with a field of view of 49° and resolution of 256×256 were assumed to be mounted on a fixed horizontal rail on the left hand side. The top and bottom cameras in the Figures could be positioned between (200, 300) and (100, 200) ranges along the rails with the accuracy of one unit. Hence, each camera could assume 100 different locations. The initial positions of the two cameras were at 300 and 100 respectively, each looking towards the positive x direction (i.e. rotation angle of 0°). Each camera could rotate $\pm 20^\circ$ with the accuracy of 1° . Consequently, the total number of configurations for this simulation is $H = 1.6 \times 10^7$. For each configuration, correspondence field of the cameras was calculated and the objective value z obtained based on Equation (14).

Figures 4(b)-(c) and 5(b)-(c) show two configurations (initial and optimal) for each scene. In each Figure, a low resolution rendition of the correspondence field is shown over the outline of the scene. It is clear from these Figures that different arrangement of objects in the scene would require substantial changes to the camera configuration if the objective function (14) were deemed desirable. For example, the optimum configuration of cameras for scene 2 occurs when the top camera is at position $a_1 = 295$ and orientation $+15^\circ$, and the second camera is at position $a_2 = 170$ and $+20^\circ$. For scene 1, the optimum position and orientation of camera 1 is at $a_1 = 290$ rotation $+20^\circ$ and for camera 2 at $a_2 = 110$ rotation -15° . It is also evident

that the optimum camera configurations result in a significant reduction in the objective value z .

5. CONCLUSIONS AND FUTURE WORK

The correspondence field of multiple cameras provides a useful mechanism to quantify the spatial relationship between the configuration of cameras and the scene. This paper demonstrated this utility for a simple scenario. Future extensions of this work include:

- Obtaining CF for an arbitrary array of cameras;
- Quantifying the impact of depth estimation error on the efficacy of camera configuration optimization and investigating the utility of CF for improving depth estimation accuracy; and
- Quantifying the impact of errors in geometry rectification algorithms on the accuracy of CF estimation.

It must be noted that the objective function developed in the optimization section of this paper, although intuitively reasonable, was for the purpose of illustrating the utility of CF. Development of a suitable objective that provides a positive impact on a desired metric (such as rendering quality or depth estimation accuracy) will also be deferred to future publications.

6. REFERENCES

- [1] M. Levoy and P. Hanrahan, "Light field rendering," *Proc. SIGGRAPH (ACM Trans. Graphics)*, pp. 31-42, Aug. 1996.
- [2] S. J. Gortler, R. Grzeszczuk, R. Szeliski, and M. F. Cohen, "The lumigraph," *Proc. SIGGRAPH (ACM Trans. Graphics)*, pp. 43-54, Aug. 1996.
- [3] J.X. Chai, X. Tong, S. C. Chan, and H. Y. Shum, "Plenoptic sampling," *Proc. SIGGRAPH (ACM Trans. Graphics)*, pp. 307-318, Jul. 2000.
- [4] C. Zhang and T. Chen, "Spectral analysis for sampling image-based rendering data," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 13, pp. 1038-1050, 2003.
- [5] C. Zhang and T. Chen, "Light field sampling," *Synthesis Lectures on Image, Video, and Multimedia Processing*, vol. 2, pp. 1-102, 2006.
- [6] H. Shidanshidi, F. Safaei, and W. Li, "A method for calculating the minimum number of cameras in a light field based free viewpoint video system," in *ICME*, 2013.
- [7] H. Shidanshidi, F. Safaei, and W. Li, "Objective evaluation of light field rendering methods using effective sampling density," in *MMSP*, 2011, pp. 1-6.
- [8] K. Takahashi and T. Naemura, "Layered light-field rendering with focus measurement," *Signal Processing: Image Communication*, vol. 21, pp. 519-530, 2006.
- [9] N. W. Daniel, I. A. Daniel, A. Ken, C. Brian, D. Tom, H. S. David, *et al.*, "Surface light fields for 3D

photography," Proceedings of the 27th annual conference on computer graphics and interactive techniques, 2000.

- [10] Y. Jingyi, L. McMillan, and S. Gortler, "Scam light field rendering," in *10th Pacific Conference on Computer Graphics and Applications*, 2002, pp. 137-144.
- [11] H. Y. Shum, J. Sun, S. Yamazaki, Y. Lin, and C. K. Tang, "Pop-up light field: An interactive image-based modeling and rendering system," *ACM Trans. Graphics*, vol. 23, pp. 143-162, Apr. 2004.
- [12] W. Wen, Z. Jiang Zhang, C. Yao Si, and D. Zeng, "An efficient method for all-in-focused light field rendering," in *ICCSIT, 2010*, pp. 399-404.

- [13] I. Aaron, M. Leonard, and J. G. Steven, "Dynamically reparameterized light fields," Proceedings of the 27th annual conference on Computer graphics and interactive techniques, 2000.
- [14] R. Hartley and A. Zisserman, *Multiple view geometry in computer vision* vol. 2: Cambridge Univ Press, 2000.
- [15] Z. Zhang, "A flexible new technique for camera calibration," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, pp. 1330-1334, 2000.
- [16] G. L. Nemhauser and L. A. Wolsey, "Integer and Combinatorial Optimization. Interscience Series in Discrete Mathematics and Optimization," ed: John Wiley & Sons, 1988.
- [17] H. Shidanshidi, F. Safaei, and W. Li, "A quantitative approach for comparison and evaluation of light field rendering techniques," in *ICME, 2011*, pp. 1-4.

(a) Scene 1

(b) Initial configuration: $z = 8522$
 $a_1 = 300, \theta_1 = 0^\circ; a_2 = 100, \theta_2 = 0^\circ$

(c) Optimum configuration: $z = 95$
 $a_1 = 290, \theta_1 = 20^\circ; a_2 = 120, \theta_2 = -15^\circ$

Figure 4. Initial and optimum CF for scene 1

(a) Scene 2

(b) Initial configuration: $z = \text{inf}$
 $a_1 = 300, \theta_1 = 0^\circ; a_2 = 100, \theta_2 = 0^\circ$

(c) Optimum configuration: $z = 67$
 $a_1 = 295, \theta_1 = 15^\circ; a_2 = 170, \theta_2 = 20^\circ$

Figure 5. Initial and optimum CF for scene 2