Optimal Stereo Camera Placement Under Spatially Varying Resolution Requirements

Rahul Malik
Department of Computer Science,
UIUC, USA
rmalik4@uiuc.edu

Peter Bajcsy
National Center for Supercomputing
Applications,
UIUC, USA
pbajcsy@ncsa.uiuc.edu

ABSTRACT

With the advent of virtual spaces, there has been a need to integrate physical world with virtual spaces. The integration can be achieved by real-time 3D imaging using stereo cameras followed by fusion of virtual and physical space information. Systems that enable such information fusions over several geographically distributed locations are called tele-immersive and should be easily deployed. The optimal placement of 3D cameras becomes the key to achieving high quality 3D information about physical spaces. In this paper, we present an optimization framework for automating the placement of multiple stereo cameras in an application specific manner. The framework eliminates ad-hoc experimentations and sub-optimal camera placements for end applications by running our simulation code. The camera placement problem is formulated as optimization problem over continuous physical space with the objective function based on 3D information error and a set of constraints that generalize application specific requirements. The novelty of our work lies in developing the theoretical optimization framework under spatially varying resolution requirements and in demonstrating improved camera placements with our framework in comparison with other placement techniques.

1. INTRODUCTION

Tele-immersive systems enable fusions of 3D information from physical and virtual spaces over several geographically distributed locations. With more advanced hardware, tele-immersive systems have a high potential of being deployed in many applications. Example applications of such tele-immersive systems involve remote monitoring of physiotherapeutic patients, understanding and annotating dance movements, identification and tracking, remote learning, and so on. The main requirement for all these applications is to have high quality 3D information, where quality refers to high spatial density in 3D space, high accuracy and low uncertainty of 3D information. Camera placement has a significant impact on the quality of 3D information obtained.

It directly impacts 3D reconstruction error and spatial density (referred to as spatial resolution in 2D cross sections of 3D space). Given limited performance of imaging sensors in terms of their spatial resolution and the finite number of 3D camera available for imaging, one has to optimize the 3D camera placement with respect to application dependent priorities on accuracy and density of 3D information.

In the past, researchers assumed uniform resolution requirements in 3D physical space [6]. However, different applications have different visual requirements and thus there is a need to accommodate heterogeneous requirements on spatial resolution. For example, a tele-immersive system used for physical activities requires the coverage of a larger area with possibly lower spatial resolution requirements than for tele-conferencing activities. Even within a single visual task, there are certain regions in space where detailed (high resolution) monitoring is needed. In other areas only a low resolution coarser view of the scene is needed. One such example would be an office environment. In such an environment, a person would perform many actions that would involve his head such as talking to someone on phone, looking at someone, putting on eye glasses, etc. A finer view of the facial features would provide detailed information of these activities. A coarser low resolution view of the scene would be sufficient to provide the context about the personŠs activities. This motivates the introduction of spatially varying resolution requirements in our camera placement optimization framework.

In this paper, we address the problem of determining the placement of 3D stereo cameras in space under spatially varying resolution requirements. The input parameters include the working volume for the users, camera properties and the room dimensions. It also consists of spatially varying minimal resolution requirements in different regions of working space. The output is a determination of the camera positions and orientations in space. The novelty of our work includes taking into account heterogenous resolution requirements in space, developing the theoretical optimization framework under spatially varying resolution requirements and in demonstrating improved camera placements with our framework in comparison with other placement techniques. The next sections describe related work (Section 2), the application activity model for including the spatially varying resolution requirements (Section 3), the theoretical formulation of the optimization framework (Section 4), experimental results (Section 5) and our summary and future work (Section 6).

2. RELATED WORK

The problem of camera placement has been studied in for different applications such as pedestrian and transportation safety, industrial applications, surveillance and human-computer interaction. To a large degree, the previous works have modeled the camera placement problem as an optimization problem. However, the problem domain and goals of previous work are quite different from ours. Overall, previous work can be divided into two categories: single camera placement and placement of multiple cameras. However, the previous works focussed on 2D cameras, whereas in our case we are dealing with 3D stereo cameras. Thus, instead of focussing on metrics such as occlusion, we mainly focus on metrics related to 3D reconstruction error.

In single camera systems, work by Cown et al. [3] optimizes the camera locations from which a specified set of object features can be viewed. Sakane et al. [7] developed a system that finds possible camera positions using a 'generate and test' strategy. This is for the inspection of an object tessellated by a sphere of a given radius. For camera placement, the main task constraint considered is edge visibility. The sensor is positioned to minimize the occlusion of selected feature edges. However, as opposed to these our work focusses on placement of multiple stereo cameras.

Work in multiple camera placement systems can be further divided into observing a fixed object versus a moving target. For a fixed target, Olague and Mohr [6] have proposed uncertainty analysis for placing multiple cameras. They approximated the projective transformation of a camera using Taylor expansion, and used a scalar function of the covariance matrix as the uncertainty measure. However, they assumed uniform resolution requirements over the entire working volume and considered limited image resolution as the only cause of 3D uncertainty. Also, the computational complexity of this approach only allows solutions involving only a few cameras.

Some work has been done for observing a moving target. Cowan et al. [2] have experimented with methods to place multiple sensors and overcome the occlusion problems associated with 3D objects. Recently, Cerfontaine et al. [1] have proposed a method to determine the optimal camera alignment for a tracking system with multiple cameras by specifying the volume to be tracked and an initial camera setup. Although these works have introduced the concept of high priority 3D volumes, most of these camera placements are for person tracking applications with 2D cameras rather than 3D stereo cameras and hence are focusing more on occlusion due to field of view than on accuracy and density of reconstructed 3D information.

3. ACTIVITY MODEL

In our model, objects are placed in 2-dimensional (2D) space. We can consider 3D space to be 2D by assuming that resolution requirements are same along the height component and thus considering only the horizontal plane. In most of the immersive systems, the focus is on a person as a whole and not on some specific body part. Thus our assumption is a valid one. Our model describes how to specify spatially varying resolution requirements for regions in working space.

The remainder of this section describes two main ideas underlying the model: (1) *resolution zones* describe how the resolution requirements of different regions varies in space,

and (2) *resolution vectors* (Ψ) characterize the resolution zones. In addition, we provide a generalization of the model to account for multiple objects in space that need finer 3D detail.
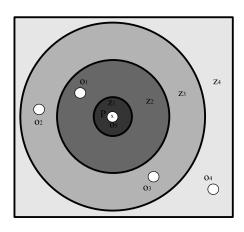
3.1 Resolution Zones

Figure 1: Resolution rings.

Within a particular visual task in space, resolution in space depends on the distance to a *pivot(P)*. It is characterized by a position in the 2D space. A pivot can be a location of a person or any other object in space. This is the region of space where a finer and detailed view is required and thus spatial resolution requirements are the highest. Figure 1 illustrates a 2D plane with objects o_1, o_2, o_3, o_4 and o_5. The pivot (o_5) is signed with a star.

By analogy with electric ($\overrightarrow{E}$) and the gravitational ($\overrightarrow{G}$) fields, a pivot generates a 'resolution field' determining the resolution of each object as a function of the distance between the object and the pivot. Thus, pivot generates *resolution zones*, iso-surfaces, ring shaped areas around the pivot, such that any objects placed within the same resolution zone would have the same minimum resolution requirements. For example, in Figure 1, pivot P is in the center of four resolution zones labeled z_i, where $1 \leq i \leq 4$. Objects o_2 and o_3 are assigned the same resolution requirement since they are in z_3.

Each resolution zone maps to a minimum resolution requirement (r_i) and forms a vector of resolution requirements. This enforces a minimum resolution requirement of r_i in zone z_i. The vector $R = < r_1, r_2, \ldots, r_n >$ is an ordered set of r_i, each specifying the resolution to be enforced within zone z_i. The property $r_i > r_{i+1}$ holds, meaning that z_i enforces higher resolution requirements than z_{i+1}. Thus, resolution zones are arranged monotonically and resolution requirements becomes weaker as the distance to P increases. In Figure 1, darker resolution zones enforce higher resolution requirements.

Let λ_i be the radius of outer circumference of zone z_i. We define z_i as follows: (1) if $i = 1$, then it is a circle of radius λ_1, (2) if $i > 1$, then z_i refers to the area enclosed

between concentric circles of radius λ_i and λ_{i-1} forming a ring. So, if the pivot P is surrounded by n resolution zones, it is necessary and sufficient to specify λ_i for all i, where $1 \leq i < n$. The resolution zone z_n refers to area beyond the circumference of radius λ_{n-1}. This is represented by $\Lambda = <\lambda_1, \lambda_2, \dots, \lambda_{n-1}>$. Thus, the *spatial resolution vector* is described as $\Psi = [R, \Lambda]$. This describes the minimum resolution required at different regions of space per pivot.

We also provide a generalization of the activity model to *multi-pivot* scenarios. The multi-pivot generalization allows for more than one pivot in the space. Objects are assigned the maximum resolution requirement with respect to all the pivots.

In summary, in order to specify the activity model in the space, user needs to specify the set of pivots P_i. For each pivot P_i, resolution vector Ψ_i is described as $\Psi_i = [R_i, \Lambda_i]$. From this activity model, minimum resolution requirements in space are enforced.

4. MATHEMATICAL FORMULATION OF OPTIMIZATION FRAMEWORK

In this section, we describe briefly the formulation of objective function as well as multiple constraints. We consider stereo localization error as the key component of the objective function.

4.1 Objective Function

We study both the down-range and cross-range localization errors to include them into the objective function Localization error in the direction of viewing from a camera is called down-range error and localization error normal to that is called cross-range error. The two errors are represented as ΔR and ΔC respectively.

Kim et al.[5] have derived the relationship of the disparity error to the stereo 3D localization error. For a figure and detailed explanation, please refer to [5]. For a general case when disparities in both left and right cameras, Δp_l and Δp_r, are non-zero, the down-range error was derived as,

$$\Delta R = -(R^2 \cos \theta / fB)(\Delta p_l - \Delta p_r) \qquad (1)$$

where, B is stereo baseline, f is the focal length, θ is the angle made with normal to the camera axis and R is the distance of the feature from the camera. For the cross-range error, which is perpendicular to the line of sight, it was derived as,

$$\Delta C = (R \cos^2 \theta / f)(\Delta p_l + \Delta p_r)/2 \qquad (2)$$

4.2 Camera Placement Constraints

Identifying optimal camera placement positions might lead to placement sub-spaces where a large and unknown number of configurations have a very similar accuracy with a very different imaging geometry. The optimization framework should be well constrained to accommodate any specific user requirements, such as placement limitations or camera visibility. In our work, we incorporate constraints that assume that the working space is of arbitrary 3D shape represented by a convex hull. A user just needs to input the vertices of the working volume that is then quantized into grid points by the program. Other constraints are included as input parameters.

We impose the following constraints on camera placements: spatial resolution constraint, space dimension constraint, field

of view constraint, cutoff depth and 360° visibility constraint.

1. Spatial Resolution Constraint: Tarabanis et al.[9] analyzed the size of a linear feature in the image plane. They derived the equation defining the ratio of the size of a linear feature in the image plane (PR_{ab}) to its actual size (l). Please refer to [9] for a detailed explanation of the same. In the case of our activity model, enforcing resolution constraint leads to:

$$\frac{PR_{ab}}{l} = \frac{d|[(r_a - r_o) \times u] \times v|}{((r_a - r_o) \cdot v)((r_b - r_o) \cdot v)} > \Psi_{req} \qquad (3)$$

where, l is the length of the linear feature to be viewed having r_a and r_b as its end points, u is the unit vector along the linear feature (from r_a to r_b), PR_{ab} is the size of that feature in the image plane, d is the distance from the back nodal point of the lens to the image plane, r_o is the position vector of the frontal nodal point of the lens [1], v is the unit vector along the optical axis in the viewing direction, and Ψ_{req} is the minimum resolution required as specified by user. Also, Ψ_{req} is the minimum resolution required at that location as specified by the user. As explained previously, in case of multiple-pivots, objects are assigned the maximum resolution requirement with respect to all the pivots.

2. Space Dimension Constraint: This constraint specifies the dimensions of space where the cameras can be placed. The user specifies minimum and maximum values of the x, y dimensions. We also assume that the cameras are placed on the tripods and the camera placement height is constrained by a minimum and maximum value. So, we have,

$$x_{min} \leq x_i \leq x_{max} \qquad (4)$$

$$y_{min} \leq y_i \leq y_{max} \qquad (5)$$

$$z_{min\_placement} \leq z_i \leq z_{max\_placement} \qquad (6)$$

$$\forall i \in Cameras$$

where, x_i, y_i, z_i are the coordinates of the camera placement of ith camera.

3. Field of View Constraint: The field of view constraint is concerned with determining viewpoints from which the features of interest can be seen. If the field of view (FOV) constraint is violated, then certain features will not be seen at all. In practice, the FOV is specified in terms of the angle that the extreme rays make while entering the optical system. There are two FOV angles: *horizontal field of view angle* denoted as α and the *vertical field of view angle* denoted as β.

Intuitively, this constraint can be decomposed into two constraints, i.e. the desired point must lie between two planes for both the sets of planes. Mathematically, this constraint can be formulated as follows

$$((P - [x_i \, y_i \, z_i]^T) \cdot (a_i \times b_i)) \cdot ((P - [x_i \, y_i \, z_i]^T) \cdot (c_i \times d_i)) < 0 \ (7)$$

$$((P - [x_i \, y_i \, z_i]^T) \cdot (a_i \times d_i)) \cdot ((P - [x_i \, y_i \, z_i]^T) \cdot (b_i \times c_i)) < 0 \ (8)$$

where, P is a point that we want to determine whether it lies inside FOV, $[x_i \ y_i \ z_i]^T$ are the coordinates of camera

---

[1]The front and rear nodal points have the property that a ray aimed at one of them will be refracted by the lens such that it appears to have come from the other, and with the same angle with respect to the optical axis *http://en.wikipedia.org/wiki/Magnification*

placement, a_i, b_i, c_i, d_i are the four unit vectors along the edges of the volume subtended by FOV angles.

4. Cutoff Depth Constraint: This constraint corresponds to the minimum and maximum cutoff depth in stereo cameras. In the case of stereo cameras, if the object to be viewed is too near or too far from the camera, then the reconstruction quality is poor. As a result of this, a minimum and maximum cutoff depth is defined for the stereo cameras and only the features lying within the cutoff depths are taken into account. Mathematically, this constraint can be formulated as follows

$$CLi_{min} \leq ||([x_i \ y_i \ z_i]^T - P) \cdot v_i|| \leq CLi_{max} \qquad (9)$$

$$\forall i \in Cameras$$

where, $||\cdot||$ represents norm of the vector, $CLi_{min} \& CLi_{max}$ are the minimum and maximum cutoff depths of the ith camera, $[x_i \ y_i \ z_i]^T$ are the coordinates of its placement, v_i is the unit vector along the optical axis in the viewing direction of the camera and P is the point in consideration.

5. 360° *Visibility Constraint:* This constraint ensures 360° visibility of the entire working volume. We define a point to be visible from two cameras if the angle between those cameras is less than 90°. Mathematically, this constraint can be formulated as

$$min(\theta_{i,j}) < 90° \qquad \forall i, j \in Cameras, j \neq i \qquad (10)$$

where, $\theta_{i,j}$ is the angle subtended between camera i and camera j for a point.

4.3 Methodology

Upon combining the objective functions and the camera placement constraints including the activity model, the optimization problem can be written as follows

$$\textbf{minimize:} \qquad \sum_i \sum_j (\Delta R_{ij})^2 + (\Delta C_{ij})^2$$

$$\forall i \in Cameras, \forall j \in Grid \ points$$

Subjected to:

1. *Resolution Constraint*   (11)
2. *Camera Placement Constraint*
3. *Field of View Constraint*
4. *Cutoff Depth Constraint*
5. *360° Visibility Constraint*

where, ΔR_{ij} denotes down-range error and ΔC_{ij} denotes cross-range error of ith camera at jth grid point. The optimization is performed over all grid points of the working space.

In order to solve the optimization problem, we used *Genetic Algorithms* [4] to develop an initial solution to the optimization problem. Genetic algorithms are global search techniques and thus are suitable for obtaining initial solutions. In order to further refine the initial solution, we used the *Gradient Descent Algorithm* [8]. Gradient descent is a local optimization algorithm. To find a local minimum of a function using gradient descent, one takes steps proportional to the negative of the gradient (or the approximate gradient) of the function at the current search point. We used the initial solution obtained from genetic algorithm as a starting point for gradient descent. This two-step optimization approach helps to obtain globally optimal and locally accurate results.

The optimization methodology was implemented the algorithm using Matlab[2] and C++. We used the Genetic Algorithms and Gradient descent implementations from Matlab and Open CASCADE[3] for visualizing the camera placements.

5. EXPERIMENTAL RESULTS

In order to evaluate our optimization framework, we performed three experiments with the following room setup.

5.1 Room Setup

We considered a room with dimensions 400×250cm. The activity zone is a rectangle places in the center of the room having dimensions 200×100cm. There are two pivots P_1 and P_2 in the activity zone which are located 50cm from each side of the activity zone. The geometry of the room for camera placement is shown in Figure 2.
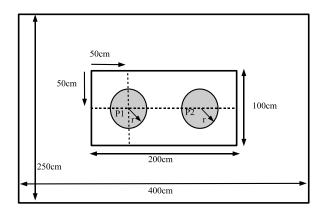

Figure 2: Room geometry for testing camera placement algorithm.

5.2 Description of Experiments

In first experiment, we compared our optimization camera placement framework under spatially varying resolution requirements (denoted as *heterogenous camera placement*) with two other placement schemes: placement under spatially homogeneous resolution requirements denoted as *homogenous camera placement* and *baseline*. In the baseline placement, the cameras are placed as far as possible in the room, touching the wall, while ensuring that the camera placement constraints are followed. This should provide the lower limit of resolution for a given set of room conditions. In homogenous camera placement algorithm, there are no pivots and no minimum resolution requirements. It tries to maximize the resolution over the entire region and simultaneously minimize 3D reconstruction error. The optimization problem for homogenous camera placement is formulated as

---

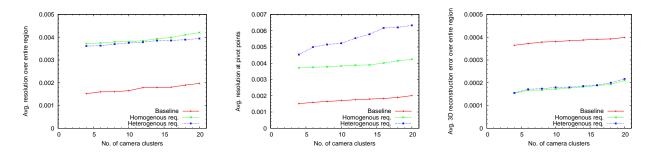[2]http://www.mathworks.com/products/matlab/
[3]http://www.opencascade.org/

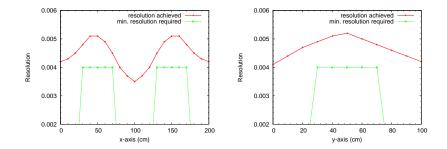**Figure 3: Comparison of heterogenous placement with other placements.**

Figure 4: Variation of resolution.

follows

minimize:

$$\frac{\gamma}{\gamma + \delta} \sum_i \sum_j ((\Delta R_{ij})^2 + (\Delta C_{ij})^2) - \frac{\delta}{\gamma + \delta} \sum_i \sum_j (PR_{ij})$$

$$\forall i \in Cameras, \forall j \in Grid\ points$$

Subjected to:

1. *Camera Placement Constraint*
2. *Field of View Constraint*
3. *Cutoff Depth Constraint*
4. *360° Visibility Constraint*

(12)

where, ΔR_{ij} denotes down-range error and ΔC_{ij} denotes cross-range error of ith camera at jth grid point and PR_{ij} represents pixel resolution from Equation 3. The negative sign in front of expression for pixel resolution denotes that it needs to be maximized instead of being minimized.

In the second experiment, we evaluated the variation of spatial resolution in space in the case of heterogenous camera placement. For these purposes, we take a horizontal and vertical line passing through a pivot point. These two lines are depicted in Figure 2 as two dotted lines. Finally, in the third experiment, we investigated the effects of varying radii and minimum resolution requirements defining the activity model (1) on total average resolution in space and (2) on average 3D reconstruction error.

5.3 Comparative Simulation Results

Figure 3 shows the results of comparison of heterogenous, homogenous and baseline placements. In terms of variation of average resolution and 3D reconstruction error over entire region, we can see that both heterogenous and homogenous placements perform much better then baseline. They provide around 150% higher resolution and around 125% lower 3D reconstruction error than baseline. This quantifies the advantages of optimal camera placement over stereo camera placement without any optimization although it does not quantify the time savings in comparison with ad-hoc camera placements. Upon comparing the three placements for average resolution at pivot points, we can see that heterogenous placement outperforms both homogenous and baseline placement. It provides around 50% higher resolution than the homogenous placement and around 200% higher resolution than the baseline placement. Another interesting fact to notice from the figure is that the average resolution provided by heterogenous placement at pivot points increases at a much higher rate than other placements.

The effects of variation of resolution along the horizontal and vertical axis are shown in Figure 4. We can see that in both cases, the effective resolution provided by the system is greater than the minimum specified by the user as an input. Figure 4 (left) shows that the two peaks of resolution along the x-axis are located at the two pivot points at 50cm and 150cm which is in accordance with the input specified by the user. Similarly, Figure 4 (right) illustrates that there is one peak that is located at 50cm at the pivot point along the y-axis.

Finally, Figure 5 shows the effects of varying the radii for a fixed value of minimum resolution requirement and varying the minimum resolution requirement for a fixed radius. We can see that as the radius of resolution zone is increased, the average resolution over entire region increases and the
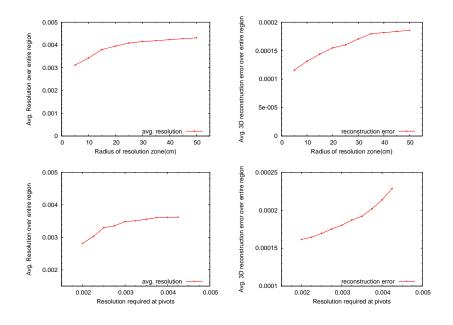
Figure 5: Effects of varying radius and minimum resolution requirement.

average 3D reconstruction error also increases. The rate of increase in average resolution is much greater for smaller values of radii than for higher values of radii. Thus, for the purpose of placement, the best value for the radius of the resolution zone is 25cm after which the average resolution nearly flattens. Upon inspecting the effect of varying the resolution requirements in the resolution zone, we can see that the average resolution over the entire region shows the same trend as shown while varying the radius. It increases steeply for lower values of requirements and then flattens out for higher values. However, the slope average 3D reconstruction error over entire region increases for higher values of resolution requirements.

6. CONCLUSIONS AND FUTURE WORK

In this paper, we have described an optimization framework for 3D stereo camera placement in 3D space under spatially varying resolution requirements. User can specify multiple resolution requirements for different regions of space in an application specific manner. The experimental results demonstrated that the heterogenous camera placement outperforms homogenous and baseline placements. In the future, we would like to perform much more extensive analyses of the optimization framework for different room dimensions, placement schemes and physical conditions. We would also like to incorporate the fact that pivots and 3D cameras might be dynamic.

Acknowledgments

7. REFERENCES

[1] P. A. Cerfontaine, M. Schirski, D. Bundgens, and T. Kuhlen. Automatic multi-camera setup optimization for optical tracking. *Proceedings of the IEEE conference on Virtual Reality*, pages 295–296, March 2006.

[2] C. Cowan, A. Bergman, and D. Nitzan. Automatic placement of vision sensors. *1990 NSF Manufacturing System Research Conference*, pages 389–395, 1990.

[3] C. Cowan and P. Kovesi. Automatic sensor placement from vision task requirements. *IEEE Trans. Pattern Analysis Machine Intelligence*, pages 407–416, 1988.

[4] D. E. Goldberg. *Genetic Algorithms in Search, Optimization and Machine Learning*. Kluwer Academic Publishers, 1989.

[5] W. S. Kim, A. I. Ansar, R. D. Steele, and R. C. Steinke. Performance analysis and validation of a stereo vision system. *IEEE international conference on systems, man and cybernetics*, 2005.

[6] G. Olague, R. Mohr, R. Venkatesh, and B. C. Lovell. Optimal camera placement to obtain accurate 3d point positions. *Fourteenth International Conference on Pattern Recognition*, 1998.

[7] S. Sakane and T. Sato. Automatic planning of light source and camera placement for an active photometric stereo system. *IEEE International Conference on Robotics and Automation*, pages 1080–1087, April 1991.

[8] J. A. Snyman. *Practical Mathematical Optimization: An Introduction to Basic Optimization Theory and Classical and New Gradient-Based Algorithms*. Springer Publishing, 2005.

[9] K. Tarabanis, R. Y. Tsai, and P. K. Allen. Analytical characterization of the feature detectability constraints resolution, focus, and field-of-view for vision sensor planning. *CVGIP: Image understanding*, 59(3):340–358, 1994.