

Camera Placement Optimization in Object Localization Systems

Dávid Szalóki*, Sándor Kolumbán*, Kristóf Csorba,*
and Gábor Tevesz*

Abstract

This paper focuses on the placement of cameras in order to achieve the highest possible localization accuracy with a multi-camera system. The cameras have redundant fields of view. They have to be placed according to some natural constraints but user defined constraints are allowed as well. A camera model is described and the components causing the localization errors are identified. Some localization accuracy measures are defined for any number of cameras. The multi-camera placement is analytically formulated using the expanded measures for multiple cameras. An example of placing two cameras is shown and the generalizations into higher dimensional parameter spaces are examined. There are publications where camera placement algorithms are formulated or compared. We make an attempt to examine the analytical solution of this problem in case of different objective functions.

Keywords: localization accuracy, multi-camera system, optimal camera placement

1 Introduction

Object tracking is nowadays a very popular segment of computer vision. There are applications where multiple cameras are used to perform this task. The applications can be classified based on the common field of views of the cameras. The first class contains applications with relatively small overlapping in the field of views. This class contains applications like video surveillance systems where the observed area should be maximized. Using these systems only a lower localization accuracy is required and the main goal is to get the approximated trajectory of the objects like in [19]. The second class contains applications where the localization accuracy can be helpful. There are plenty of papers about the localization accuracy, its derivation and the camera placement in order to improve it.

*Department of Automation and Applied Informatics, Budapest University of Technology And Economics, H-1117 Budapest, Magyar Tudósok körútja 2/Q., Hungary, E-mail: {David.Szaloki,Sandor.Kolumban,Kristof.Csorba,Gabor.Tevesz}@aut.bme.hu

In [13, 5] the well-known pinhole camera model is used. The covariance matrix of the localization is derived and the largest eigenvalue is selected as the measure of the localization accuracy. It is stated in [10] that the complexity of camera placement belongs to the class of NP-complete problems, and a genetic algorithm is applied on the global optimization problem. In [5] the “Parisian evolutionary computation approach” is used in order to reduce the computational requirements. In our paper the pinhole camera model and the eigenvalue based measure are used as well, but in addition the analytical background of the placement optimization is examined.

A robot is observed on the ground in [6]. The cameras have to be placed on the edges of the room in order to perform the localization. A very simplified camera model is used: the mapping is represented as an orthogonal projection onto the image plane. In contrast, we use the advanced pinhole camera model and user constraints are also handled during our derivations.

An interesting problem is formulated in [7]. First, the localization accuracy is derived in case of an N-ocular vision system. A camera system containing cameras with fixed positions relative to each other is built on the bottom of a boat. This camera system maps the seafloor. The position and its accuracy of the camera system can be calculated from the overlapping images through time. With this method an image with precision data can be constructed from the seafloor. Similarly we describe the localization accuracy for multiple cameras. In [7] it is used to determine the precision of the data collected about the seafloor. In our case it is used to generate an objective function for the camera placement optimization.

In [2] a camera system with three cameras is examined. The main result of that paper is, that in order to achieve the smallest searching region, the cameras have to be placed so that their line of sights are pairwise orthogonal. This is called the “orthogonal axes stereo model”. Later in our paper a similar result is given as perpendicular facing directions are optimal in a certain sense.

The multi-camera system can also be simulated with one camera through time: one camera is moved with a robotic arm in [4]. An object has to be digitalized with this camera. The path of the robotic arm and the imaging spots are planned according to the objective function such as “Minimizing Reconstruction Uncertainty”, “Minimizing Required Motion” and “Minimizing Number of Images”.

If every camera detects the object in its own image, the position in the world can be calculated. But using such a system leads to inaccuracies which are discussed later in this paper. A simple estimation of the position of the observed object is given in [9] and [16]. They are not exactly the same but very similar. We use the estimation introduced in [16].

In [18] the algorithms capable of approximating the NP-hard camera placement problem are introduced and compared. There are two main groups: The goal of the MIN problems is to minimize the number of cameras such that a target coverage rate p can be achieved subject to other constraints. The goal of the FIX problems is to maximize the coverage of targets subject to fixed number of cameras m and other application specific constraints. We examine a problem which belongs to the second group since a fixed number of cameras are placed according to some

constraints.

In [12, 3, 11] PTZ (PanTiltZoom) cameras are used for object tracking. First, only the orientation of the cameras is optimized in order to get the best possible tracking accuracy. Second, the cameras are built on “Dubins vehicles” (ca. aerial vehicles) and their positions can be manipulated according to the dynamics of the vehicles. A Dubins vehicle is a planar vehicle that has a fixed forward velocity with a bounded turning radius. Objects with dynamics given by stochastic difference equations are tracked. It is said in [3] that the utility function is designed to optimize estimation performance over a planning horizon. Performance is quantified as a function of the expected fused geolocation information. A trace based measure is used and its advantages and drawbacks are analyzed as well. Third, the goal is to take better pictures of the objects while the provided tracking accuracy is complied. The value function is designed as an ordinal potential function, such that it can be decoupled into local objectives known to every camera. In our paper not only the trace based measure is investigated, but the more complex determinant based and the eigenvalue based ones as well. It is shown that the trace based objective function can be decoupled as described in [12, 3, 11]. Moreover, we examine the camera placement optimization in case of using the other two measures.

There are publications where camera placement algorithms are formulated or compared. We make an attempt to examine the analytical solution of this problem in case of different objective functions. The camera model and the derivation of the localization accuracy of a multi-camera system are introduced. The measures of the localization accuracy are defined and the analytical examination of the camera placement is described. Note that the presented algorithms and methods are part of the Smart Mobile Eyes for Localization (SMEyeL) system which is open-source and is written in C++ using the popular OpenCV [1] computer vision library. The source code, documentation and all the input data for the presented measurements are available for download from our homepage [15].

In [17] the basic problem is formulated, the camera model is introduced, the localization accuracy of multiple cameras is calculated and the placement of two cameras are examined. In [16] a localization accuracy measurement is described which corresponds to the theory in [17]. In this present paper the placement of multiple cameras are examined using several objective functions and analytical solutions are given.

2 Problem Formulation

Assume that we have a multi-camera system observing an area. Every camera is calibrated which means their parameters, positions and orientations are known. There is an object in the observed area and its location has to be determined, this process is called localization. Localization can be performed with various methods, an example with measurements is shown in [16], while another method is introduced in [9]. The main idea is that every camera detects the object in its own image and these detections determine half-lines in the world coordinate system. The object

is located at the intersection point of these half lines. As described in this paper, such a system contains some inaccuracies. This means that these half lines do not intersect in one single point, so an optimal solution has to be found. But this also means that a localization accuracy can be defined and it can be improved for example with a better camera placement.

2.1 Camera model

In order to determine the localization accuracy first the camera model has to be defined and the error sources have to be identified. Simple optical cameras are used because they are easily accessible, every person usually has one in the form of a smartphone. We use the widely known pinhole camera model as in [13, 5]. Some publications do not mention but implicitly they also use this model. Details about this camera model can be found in [8]. In this model the camera center is at its focal point and the camera has a plane imager in focal length distance from the center. The imaging is a central projection onto this plane with the focal point as center. As described in [17], there are two error sources during a localization:

- No depth information:
One camera cannot serve depth information unless the size of an object is known. This means that a camera can determine only the direction in which the object is located, it cannot say anything about the distance.
- Observation error:
One camera can detect the object in its own image only with some error. The object position in the image is derived from its contour. This means that the position is calculated as an average, which implies that this observation error has a normal distribution. The standard deviation of this error is invariable.

2.2 Covariance Matrix

Based on the calculations presented in [17] the covariance matrix of the localization performed by one camera can be constructed.

Definition 1. *The covariance matrix of the considered camera model is given by the formula:*

$$\mathbf{\Sigma}(\boldsymbol{\alpha}, d) = \mathbf{R}^T(\boldsymbol{\alpha}) \begin{bmatrix} \sigma_{\perp}^2(d) & 0 & 0 \\ 0 & \sigma_{\perp}^2(d) & 0 \\ 0 & 0 & \sigma_{\parallel}^2 \end{bmatrix} \mathbf{R}(\boldsymbol{\alpha}), \quad (1)$$

where $\mathbf{\Sigma}$ is the covariance matrix, \mathbf{R} is a rotation matrix representing the $\boldsymbol{\alpha}$ orientation of the camera, d is the distance of the object from the camera, σ_{\parallel} is the parallel component while σ_{\perp} is the perpendicular component of the standard deviation of the localization error. The camera is facing the direction of the z axis of the Cartesian coordinate system fitted to the camera center.

From now on only cameras with this form of camera model are investigated. For one camera:

$$\sigma_{\parallel} \rightarrow \infty, \quad (2)$$

$$\sigma_{\perp} \propto d. \quad (3)$$

Since the parallel component of the standard deviation goes to infinity these covariance matrices have non-real values. Fortunately the inverse of such a matrix can be calculated and contains only finite elements. If the ratio between σ_{\perp} and d is determined as

$$C = \frac{d}{\sigma_{\perp}}, \quad (4)$$

then C is invariable for a given camera, but it depends on the focal length so it can be different for multiple cameras. The inverse of the covariance matrix becomes:

$$\boldsymbol{\Sigma}^{-1}(\boldsymbol{\alpha}, d) = \mathbf{R}^T(\boldsymbol{\alpha}) \begin{bmatrix} C^2 d^{-2} & 0 & 0 \\ 0 & C^2 d^{-2} & 0 \\ 0 & 0 & 0 \end{bmatrix} \mathbf{R}(\boldsymbol{\alpha}). \quad (5)$$

Therefore in the following sections the inverse of the covariance matrix is used.

2.2.1 Merging

Now, we have one camera observing one single point and the inverse of the covariance matrix is constructed. In real-world situations multiple cameras are observing the objects, so their covariance matrices have to be merged. This combination is called *the product of Gaussian densities* [14] and it can be achieved by applying the following formulas incrementally:

$$\boldsymbol{\Sigma}_r^{-1} = \boldsymbol{\Sigma}_1^{-1} + \boldsymbol{\Sigma}_2^{-1}, \quad (6)$$

$$\boldsymbol{\mu}_r = \boldsymbol{\Sigma}_r \cdot (\boldsymbol{\Sigma}_1^{-1} \boldsymbol{\mu}_1 + \boldsymbol{\Sigma}_2^{-1} \boldsymbol{\mu}_2), \quad (7)$$

where the two original densities are $\mathcal{N}(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_1)$ and $\mathcal{N}(\boldsymbol{\mu}_2, \boldsymbol{\Sigma}_2)$, while the combined, resulting density is $\mathcal{N}(\boldsymbol{\mu}_r, \boldsymbol{\Sigma}_r)$. The visualization of this merging can be seen in Fig. 1. Two cameras represented with filled circles are observing one single point. The two larger ellipses are the individual covariance ellipses of the cameras separately. The confidence levels are chosen in order to get the best visualization. For better visualization their major axes are shrunk so that not only two pairs of parallel lines can be seen. The smaller ellipse is the combined covariance ellipse.

The orientation of the cameras in case of using multiple cameras observing one single point can be defined in the simplest way as the camera positions are given in a spherical coordinate system. The coordinate system can be fitted to the observed point.

Assumption 1. *The cameras are facing the observed object which is located at the origin.*

Using Definition 1 and Assumption 1 the final resulting covariance matrix is:

$$\Sigma_r^{-1}(\mathbf{x}) = \sum_{i=1}^n \Sigma_i^{-1}(\mathbf{x}_i), \quad (8)$$

where n is the number of the cameras used for localization and Σ_i^{-1} is the inverse of the covariance matrix of the i^{th} camera. \mathbf{x}_i is the position of the i^{th} camera. During the calculations it will be represented as spherical coordinates. \mathbf{x} contains all the \mathbf{x}_i positions of all the cameras.

In 2D all the matrices become 2-by-2 ones, the coordinate system is a polar coordinate system and the orientation is represented by a scalar value. The resulting covariance matrix can be calculated from its inverse and it can be plotted at a given confidence level. In 3D this is an ellipsoid while in 2D this becomes an ellipse. In Fig. 1 these covariance ellipses are drawn at some localization points for a given camera configuration containing two cameras. The placement is symmetric, since all the cameras have the same parameters. It can be seen that the errors near the cameras are smaller and they have a larger variance in the y direction, while the distant ones have the larger variance in the x direction.

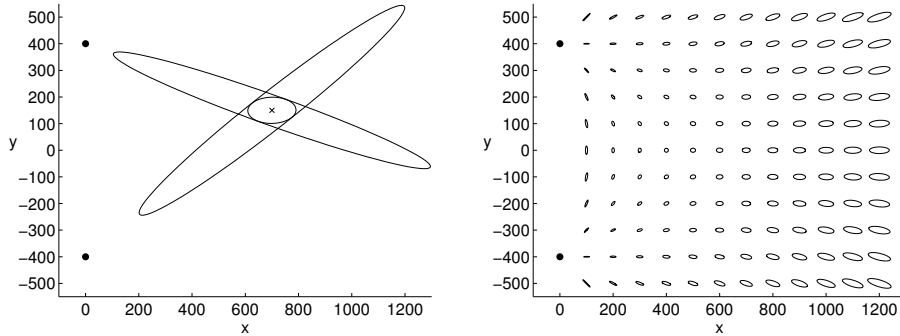


Figure 1: The derivation of the combined covariance ellipse and the resulting covariance ellipses at some localization points.

2.3 Measures

Based on the previous sections the inverse of the resulting covariance matrix is formulated in case of multiple cameras observing the origin. In order to perform an optimization of the localization accuracy, it needs to be given by one scalar value.

Definition 2. *The localization accuracy can be defined by various measures, but it is common that all the measures are derived from the inverse of the covariance*

matrix:

$$q_{\text{eig}}(\mathbf{x}) = \min(\text{eig}(\Sigma_r^{-1}(\mathbf{x}))), \quad (9)$$

$$q_{\text{det}}(\mathbf{x}) = \det(\Sigma_r^{-1}(\mathbf{x})), \quad (10)$$

$$q_{\text{trace}}(\mathbf{x}) = \text{trace}(\Sigma_r^{-1}(\mathbf{x})), \quad (11)$$

where $\Sigma_r^{-1}(\mathbf{x})$ is defined in (8).

Since the cameras are observing the origin, the measures are functions of the camera positions \mathbf{x} only. The determinant based measure is inversely proportional to the area, while the eigenvalue based measure is inversely proportional to the length of the largest axis of the covariance ellipsoid. Using the trace of the inverse of the covariance matrix can also be beneficial ($\sum \lambda_i^{-1}$). In the 2D case these measures can be easily visualized with surface plots for any camera configuration. In Fig. 2 these measures are plotted for the camera configuration defined in Fig. 1. The surface plot represents the localization accuracy with this fix camera configuration at every point if the localization was performed there. It can be seen that the

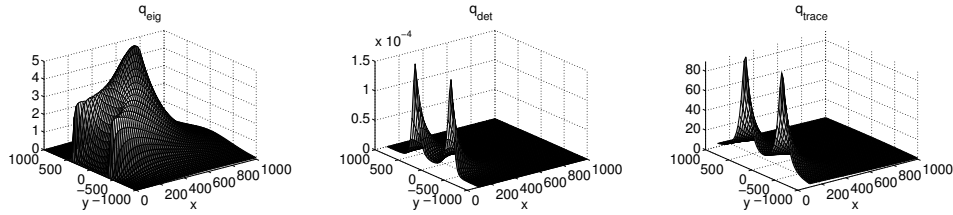


Figure 2: The surf plots of the measures defined in (9)—(11). The camera positions are the same as in Fig. 1.

trace based measure prefers the smaller distance, while with the eigenvalue based measure the inclination angle counts more. The determinant based measure is somewhere between the other two. The details of this statement are described later in this paper. The determinant and the trace based measure goes to infinity if the observed point tends to any of the cameras. With the eigenvalue based measure this is not the case. Of course the observed object cannot be placed arbitrarily close to a camera because a camera has a physical size and minimal focal distance. In real world applications these and other constraints are present.

In [13, 5] the eigenvalue based measure, while in [12, 3, 11] the trace based one is used. In the following sections all three matrix norms are used, other norms could also be applied.

3 Placement optimization

Based on (9)—(11) the localization accuracy can be determined in case of using multiple cameras observing the origin. If the localization accuracy has to be im-

proved, it can be achieved by placing the cameras in more appropriate positions. The problem of placing the cameras in optimal positions means:

$$\mathbf{x}^{(\text{opt})} = \arg \max_{\mathbf{x} \in K} q(\mathbf{x}), \quad (12)$$

where q is one of the measures defined in (9)—(11). K is the constraint domain of the optimization and it represents the area where the cameras can be placed. It is important that none of the cameras can be placed in the position of the observed point.

Assumption 2. *The observed point is excluded from the set of possible camera positions.*

3.1 Cameras with different parameters

The localization accuracy of one camera depends on the C value defined in (4). As described in Section 2.2: C is invariable for a given camera, but it depends on the focal length so it can be different for multiple cameras. It would be a large simplification if all the cameras could be transformed into equivalent ones. If for all cameras the $d' = C^{-1}d$ substitution is performed, then all cameras can be handled as if they had the same parameters: $C' = 1$. This substitution represents a homothetic transformation where the homothetic center coincides with the observed point. This means if the K constraints in (12) are transformed according to this homothetic transformation, and the problem with the new constraints and equivalent cameras is solved, then this solution is also the solution of the original problem. An important property of the transformation is that it preserves the collinearity of points which implies that it preserves the convexity of the constraints. This simplification is used in the following sections.

Assumption 3. *All the cameras have the same parameters, so for every camera $C = 1$ in Definition 1.*

3.2 Inversion of a camera through the observed point

Assume that the position of one camera is inversed through the observed point. In this case

$$\Sigma_i(\mathbf{x}_i) = \Sigma_i(-\mathbf{x}_i), \quad (13)$$

where Σ_i is the covariance matrix and \mathbf{x}_i is the position of the i^{th} camera. This statement is true since in (1) the diagonal matrix is invariable and the rotation matrices are multiplied with (-1) . Similarly to Section 3.1 this means if for some of the cameras all of their constraints are transformed according to this transformation, and the problem with the new constraints and equivalent cameras is solved, then this solution is also the solution of the original problem. Similarly this transformation also preserves the collinearity of points which implies that it preserves the convexity of the constraints. This transformation is used in the following sections.

3.3 The 2D case

In the 3D world the camera placement defined in (12) is a hard global optimization problem, so first the 2D case is managed. In the following section the \mathbf{x} camera positions are represented in the polar coordinate system with $(\boldsymbol{\alpha}, \mathbf{d}) \in ([-\pi; \pi]^n, \mathbb{R}^{+n})$.

The eigenvalues of the inverse covariance matrix can be determined as the solution of the quadratic characteristic polynomial:

$$\text{eig}(\boldsymbol{\Sigma}_r^{-1}(\boldsymbol{\alpha}, \mathbf{d})) = \frac{1}{2} \left(\sum_i \frac{1}{d_i^2} \pm \sqrt{R} \right), \quad (14)$$

where

$$R = \sum_i \frac{1}{d_i^4} + \sum_{\substack{k,l \\ k < l}} 2 \frac{1}{d_k^2} \frac{1}{d_l^2} \cos(2(\alpha_k - \alpha_l)), \quad (15)$$

and (α_i, d_i) represents the \mathbf{x}_i position of the i^{th} camera in a polar coordinate system and $(\boldsymbol{\alpha}, \mathbf{d})$ contains all the positions of all the cameras. The second summation is performed for every ordered camera pairs. It can be noted that both of the eigenvalues are non-negative values.

Using the eigenvalues the objective functions can be expanded:

$$q_{\text{eig}}(\boldsymbol{\alpha}, \mathbf{d}) = \frac{1}{2} \left(\sum_i \frac{1}{d_i^2} - \sqrt{R} \right), \quad (16)$$

$$q_{\text{det}}(\boldsymbol{\alpha}, \mathbf{d}) = \sum_{\substack{k,l \\ k < l}} \frac{1}{d_k^2} \frac{1}{d_l^2} \sin^2(\alpha_k - \alpha_l), \quad (17)$$

$$q_{\text{trace}}(\boldsymbol{\alpha}, \mathbf{d}) = \sum_i \frac{1}{d_i^2}. \quad (18)$$

It was used that the determinant is the product, while the trace is the sum of the eigenvalues.

3.3.1 Partial derivatives

The first step to examine the objective functions is to calculate their partial derivatives.

$$\frac{\partial q_{\text{eig}}}{\partial \alpha_i} = \frac{1}{\sqrt{R}} \frac{1}{d_i^2} \sum_{\substack{j \\ j \neq i}} \frac{1}{d_j^2} \sin(2(\alpha_i - \alpha_j)), \quad (19)$$

$$\frac{\partial q_{\text{det}}}{\partial \alpha_i} = \frac{1}{d_i^2} \sum_{\substack{j \\ j \neq i}} \frac{1}{d_j^2} \sin(2(\alpha_i - \alpha_j)), \quad (20)$$

$$\frac{\partial q_{\text{trace}}}{\partial \alpha_i} = 0. \quad (21)$$

The partial derivatives of the trace based measure with respect to the α_i angles are constant zero which means that this measure is independent from the angles as it can also be derived from (18). The partial derivatives with respect to the α_i angles of the eigenvalue based and the determinant based measures are more complicated. They can be negative, positive or zero and they contain trigonometric functions of inclination angles of all the camera pairs containing the i^{th} camera.

The partial derivatives with respect to the distances d_i :

$$\frac{\partial q_{\text{eig}}}{\partial d_i} = -\frac{1}{d_i^3} + \frac{1}{\sqrt{R}} \left(\frac{1}{d_i^5} + \frac{1}{d_i^3} \frac{1}{d_j^2} \cos(2(\alpha_i - \alpha_j)) \right), \quad (22)$$

$$\frac{\partial q_{\text{det}}}{\partial d_i} = -2 \frac{1}{d_i^3} \sum_{\substack{j \\ j \neq i}} \frac{1}{d_j^2} \sin^2(\alpha_i - \alpha_j), \quad (23)$$

$$\frac{\partial q_{\text{trace}}}{\partial d_i} = -2 \frac{1}{d_i^3}. \quad (24)$$

It can be proven that for any of the previously mentioned measures the partial derivatives with respect to the distances d_i are negative. This is consistent with the natural intuition that placing any camera closer to the observed object improves the localization accuracy. Of course a camera cannot be placed arbitrarily close to the observed object. This consideration implies that the cameras of an optimal placement lie on the constraint boundary.

3.3.2 Placing 2 cameras

The simplest case is to place only two cameras since then the summations by all the camera pairs are eliminated. The whole parameter space can be partitioned so that the measures are quasi-concave functions on every domain. If these domains and the user defined constraints are convex ones, then the objective functions have only one local maximum at every domain. This can be used during the proof of optimality of a solution and it can also be useful if the local optimum values cannot be determined analytically. This means that running a local optimizer at every domain results all the local optimum values and the best one can be simply chosen. Based on the equations (19)–(24) in case of using the trace based measure the whole parameter space is one big domain, but with the other two measures the domain boundaries are determined by the hyperplanes:

$$\alpha_1 - \alpha_2 = m \cdot \frac{\pi}{2}, \quad m \in \{-4, -3, -2, -1, 0, 1, 2, 3, 4\}. \quad (25)$$

In the following section a specific case is introduced. Every coherent constraint in 2D can be approximated by a polygon. The polygon can be splitted into convex ones. Using the fact that the optimal placement lies on the boundary means that only the line segments of the polygons have to be handled. This means that we have to deal only with linear constraints.

Assumption 4. *The placing constraints of the cameras are one variable linear equality constraints.*

One of the simplest cases of camera placement with linear constraints is the problem of placing two cameras within two half planes. This problem can be parametrized in the Cartesian coordinate system with the distance of the intersection point of the half plane borderlines from the observed point (origin) at the x axis and their azimuths (γ_1, γ_2). The intersection distance of the borderlines from the origin can be chosen to unit distance based on Section 3.1. The inclination angle of the borderlines can be chosen as $0 < |\gamma_2 - \gamma_1| \leq \frac{\pi}{2}$ based on Section 3.2. Furthermore the azimuths of the borderlines can be chosen as $-\frac{\pi}{2} < \gamma_1 < 0 < \gamma_2 < \frac{\pi}{2}$ since the problem has reflectional symmetry with the x axis.

Assumption 5. *The intersection distance of the borderlines from the origin is chosen to unit distance. The inclination angle of the borderlines satisfies $0 < |\gamma_2 - \gamma_1| \leq \frac{\pi}{2}$, while the azimuths of the borderlines satisfy $-\frac{\pi}{2} < \gamma_1 < 0 < \gamma_2 < \frac{\pi}{2}$.*

It is important that none of these half planes contains the observed point. The special case when the borderlines of the half planes are parallel is discussed later in this paper. The general case is shown in Fig. 3. The two solid lines are the borderlines of the half planes. The filled circle (origin) is the observed point while the other symbols are the optimal camera positions belonging to different measures.

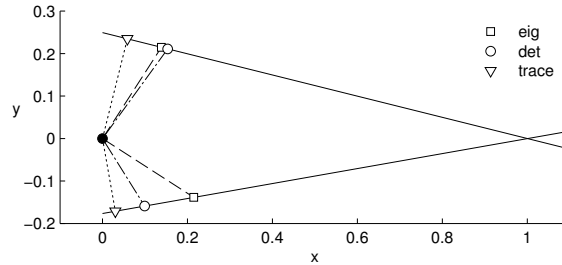


Figure 3: The optimal placements of two cameras in two half planes using the measures defined in (9)—(11).

The observed point is the origin and the camera positions are represented in the polar coordinate system with the previously mentioned α_i and d_i . Using the fact that the optimal placement lies on the boundary means:

$$\gamma_1 < \alpha_1 < \gamma_1 + \pi, \quad (26)$$

$$\gamma_2 > \alpha_2 > \gamma_2 - \pi, \quad (27)$$

$$d_1 = \frac{\sin \gamma_1}{\sin(\gamma_1 - \alpha_1)}, \quad (28)$$

$$d_2 = \frac{\sin \gamma_2}{\sin(\gamma_2 - \alpha_2)}. \quad (29)$$

Claim 1. *Placing two cameras in 2D according to the Assumptions 1–5 the optimal placement for the measure defined in (16) is:*

$$\begin{bmatrix} \alpha_1 \\ \alpha_2 \end{bmatrix}^{(\text{eig})(\text{opt})} = \begin{cases} \begin{bmatrix} + \operatorname{atan} \left(\frac{1 - \operatorname{ctg} \gamma_2}{1 + \operatorname{ctg} \gamma_1} \right) \\ - \operatorname{atan} \left(\frac{1 + \operatorname{ctg} \gamma_1}{1 - \operatorname{ctg} \gamma_2} \right) \end{bmatrix}, & \text{if } \sin(\gamma_2 - \gamma_1) > \frac{\sin \gamma_2}{\sin \gamma_1} \\ \begin{bmatrix} \gamma_1 + \frac{\pi}{2} \\ \gamma_1 \end{bmatrix}, & \text{if } \sin(\gamma_2 - \gamma_1) \leq \frac{\sin \gamma_2}{\sin \gamma_1} \end{cases} \quad (30)$$

and the distances can be calculated using (28) and (29).

In case of the eigenvalue based measure after some considerations using (19) and (22) it can be proven that the optimal solution (α_1, α_2) satisfies:

$$\gamma_1 - \frac{\pi}{2} \leq \alpha_2 \leq \gamma_1 < 0 < \gamma_2 \leq \alpha_1 \leq \gamma_1 + \frac{\pi}{2}. \quad (31)$$

It can be simply proven that the solution is included by the union of the two domains defined in (25) for $m \in \{0, 1\}$ and $m \in \{1, 2\}$. It can also be shown that the solution is located strictly inside the union. We have no proof but it is our strong conjecture that the optimum lies on the common boundary of these two domains and the optimum is defined in (30).

Using the determinant or the trace based measure the optimal solutions become:

Theorem 1. *Placing two cameras in 2D according to the Assumptions 1–5 the optimal placements for the measures defined in (17) and (18) are:*

$$\begin{bmatrix} \alpha_1 \\ \alpha_2 \end{bmatrix}^{(\text{det})(\text{opt})} = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}^{-1} \begin{bmatrix} \pi + \gamma_1 \\ \pi - \gamma_2 \end{bmatrix}, \quad (32)$$

$$\begin{bmatrix} \alpha_1 \\ \alpha_2 \end{bmatrix}^{(\text{trace})(\text{opt})} = \begin{bmatrix} \gamma_1 + \frac{\pi}{2} \\ \gamma_2 - \frac{\pi}{2} \end{bmatrix}, \quad (33)$$

and the distances can be calculated using (28) and (29).

Proof. The proof of the optimal solutions for the determinant and the trace based measures.

In the determinant case similarly to the eigenvalue case it can be proven using (20) and (23) that the optimal solution (α_1, α_2) satisfies (31). The constraints (28) and (29) are substituted into the expanded objective function defined in (16) in case of $n = 2$. The roots of the partial derivatives with respect to the α_i angles are calculated. The only feasible solution is defined in (32) and it is the global maximum value.

The trace based case can be solved similar to the previous one, but there is a more simpler solution. Since this measure is independent from the α_i angles and the objective function can be decoupled, the distances has to be minimized. The optimal solution contains the orthogonal projections of the origin onto the lines. \square

The optimal solutions are shown in Fig. 3. The contour plots of the objective functions are shown in Fig. 4. The global optimum values are represented by the filled circles. A curve contained by another one has a higher localization accuracy. The dashed lines represent the domain boundaries determined in (25). It can be noticed that the objective functions have in fact only one local optimum in every domain. In the eigenvalue case the solution lies on the domain boundary, while in the other two cases the solutions lie on the user defined boundaries (borderlines of the half planes).

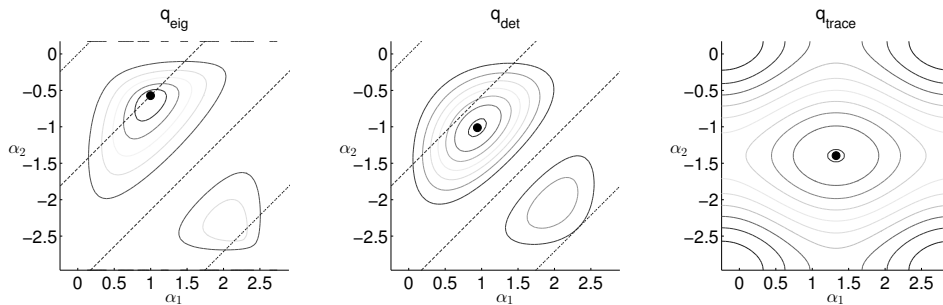


Figure 4: The contour plots of the objective functions in case of placing two cameras in two half planes where the borderlines are not parallel.

The optimal solutions introduced in (30)—(33) have some physical meaning. The trace based measure forces the cameras in positions where every camera is as close to the observed point as it can ($\min d_1, \min d_2$). The eigenvalue based measure results in camera positions where the observation angles are perpendicular and the distances from the observed object are equal ($\alpha_1 - \alpha_2 = \frac{\pi}{2}, d_1 = d_2$). If this is not possible, the observation angles are perpendicular and the d_1 distance is minimized ($\alpha_1 - \alpha_2 = \frac{\pi}{2}, \min d_1$). A similar result is described in [2] as perpendicular facing directions are optimal in a certain sense. The determinant based one is somewhere between the previous two measures which means this allows different observation inclination angle than $\frac{\pi}{2}$ in order to get closer to the observed point.

Parallel constraints The special case when two cameras have to be placed in two half planes where the borderlines are parallel is not included in the general case. Similarly to the general case none of these half planes contains the observed point. The statements that only the borderlines have to be examined during the optimization is also true. The case when the observed point is between the two half planes is shown in this section. If this is not the case, so one of the half planes contains the other, the problem can be transformed based on Section 3.2. Similarly

to (30)—(33) the optimal solutions in this cases are:

$$\begin{bmatrix} \alpha_1 \\ \alpha_2 \end{bmatrix}^{(\text{eig})(\text{opt})} = \begin{bmatrix} + \operatorname{atan} \left(\frac{m_1}{m_2} \right) \\ - \operatorname{atan} \left(\frac{m_2}{m_1} \right) \end{bmatrix}, \quad (34)$$

$$\begin{bmatrix} \alpha_1 \\ \alpha_2 \end{bmatrix}^{(\text{det})(\text{opt})} = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}^{-1} \begin{bmatrix} \pi \\ \pi \end{bmatrix} = \begin{bmatrix} \frac{\pi}{3} \\ \frac{\pi}{3} \end{bmatrix}, \quad (35)$$

$$\begin{bmatrix} \alpha_1 \\ \alpha_2 \end{bmatrix}^{(\text{trace})(\text{opt})} = \begin{bmatrix} +\frac{\pi}{2} \\ -\frac{\pi}{2} \end{bmatrix}, \quad (36)$$

where m_1 and m_2 are the distances of the borderlines from the observed point. The physical meaning is the same as in the general case. The proof of Theorem 1 is also applicable in this case. In case of using the eigenvalue and determinant based measures the roots of the partial derivatives are determined after the substitution. With the trace based measure the distances are minimized according to the constraints.

The optimal solutions are shown in Fig. 5, while the contour plots of the objective functions are shown in Fig. 6. It can be noticed that the eigenvalue and the determinant based measures became point reflection symmetric with the center $(\frac{\pi}{2}, -\frac{\pi}{2})$. This is because the problem is reflectional symmetric and the axis of the symmetry is the y axis as it can be seen in Fig. 5.

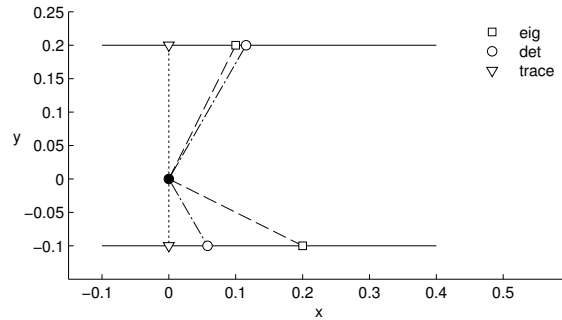


Figure 5: The optimal placements of two cameras in two half planes where the borderlines are parallel. The notations are the same as in Fig. 3.

4 Generalizations

In case of using more than two cameras some difficulties occur during generalization of the previously presented methods. Domains are required where the eigenvalue or the determinant based measures are quasi-concave functions. Based on Section 3.3.1 this means that none of these formulas can have roots inside the domains at most

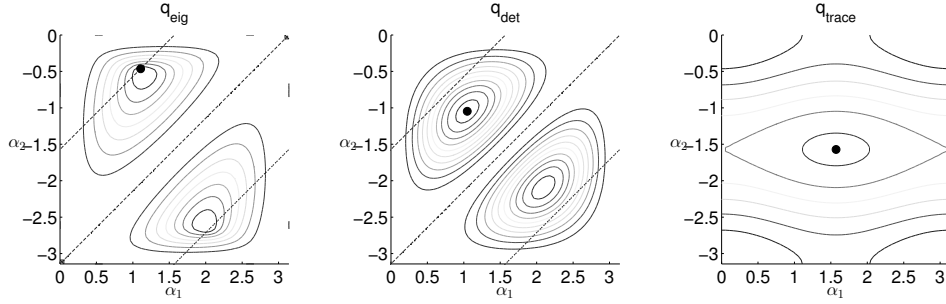


Figure 6: The contour plots of the objective functions in case of placing two cameras in two half planes where the borderlines are parallel. The notations are the same as in Fig. 4.

on the boundary:

$$\sum_{\substack{j \\ j \neq i}} \frac{1}{d_j^2} \sin(2(\alpha_i - \alpha_j)), \quad i = 1..n. \quad (37)$$

Section 3.3.1 contains a solution for $n = 2$, but for $n \geq 3$ this is a hard, analytically unsolved task. The trace based measure is a simple one. Since in this case all the cameras are independent and the whole parameter space is one big domain, this optimization can be easily performed. As described in [12, 3, 11] the objective function can be decoupled. This results camera positions where all of them are placed as close to the observed point as possible.

In case of optimizing in the 3D world some other difficulties occur during the analytical handling of the problem. In order to get the eigenvalue based measure the roots of a third order polynomial have to be determined. The determinant can be calculated but it does not have a compact form as it is the determinant of the sum of rotated 3-by-3 diagonals. The trace based measure is very simple. The trace is invariant under cyclic permutations. Using this, (5), (8), (11) and the simplification in Section 3.1, the objective function becomes

$$q_{\text{trace}}(\mathbf{x}) = \sum_{i=1}^n \frac{2}{d_i^2}, \quad (38)$$

where \mathbf{x} represents the camera positions. This form is similar to the 2D case (18).

Our conjecture is that problems where the observed object is located strictly on a plane can be solved with the 2D optimization technique. This means that a robot on the floor can be localized with position optimized cameras but a flying robot will need 3D optimization. This problem requires some further research.

5 Conclusions

In this paper a multi-camera system is analytically examined. The multi-camera system is used for object localization. There are publications where algorithms are formulated [13, 5] or compared [18], but none of them investigates the analytical solution.

In this present paper the placement of multiple cameras are examined using several objective functions and analytical solutions are given. First, a camera model is shown and the components which cause the localization errors are determined. The covariance matrix of the localization is constructed for one camera and it is derived for multiple cameras too. Some localization accuracy measures are defined and their physical interpretation is given. The multi-camera placement is analytically examined. An example of placing two cameras is shown and the generalizations into higher parameter spaces are investigated. During the generalization some hard problems occur that cannot be solved analytically with the introduced specific theorems. In case of using the trace based measures all the generalizations can be done easily since in this case all the cameras are independent, the whole parameter space is one big domain and the objective function can be decoupled.

Acknowledgments

This work was partially supported by the European Union and the European Social Fund through project FuturICT.hu (grant no.: TAMOP-4.2.2.C-11/1/KONV-2012-0013) organized by VIKING Zrt. Balatonfüred.

This work was partially supported by the Hungarian Government, managed by the National Development Agency, and financed by the Research and Technology Innovation Fund (grant no.: KMR_12-1-2012-0441).

References

- [1] Bradski, G. The OpenCV Library. *Dr. Dobb's Journal of Software Tools*, 2000.
- [2] Chiou, R.-N., Chen, C.-H., Hung, K.-C., and Lee, J.-Y. The optimal camera geometry and performance analysis of a trinocular vision system. *IEEE Transactions on Systems, Man, and Cybernetics*, pages 1207–1220, 1995.
- [3] Ding, C., Morye, A. A., Farrell, J. A., and Roy-Chowdhury, A. K. Coordinated sensing and tracking for mobile camera platforms. In *American Control Conference (ACC), 2012*, pages 5114–5119, June 2012.
- [4] Dunn, E. and Olague, G. Pareto optimal camera placement for automated visual inspection. In *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 3821–3826. IEEE, 2005.

- [5] Dunn, E., Olague, G., and Lutton, E. Parisian camera placement for vision metrology. *Pattern Recognition Letters*, pages 1209–1219, 2006.
- [6] Ercan, A. O., Yang, D. B., Gamal, A. E., and Guibas, L. J. Optimal placement and selection of camera network nodes for target localization. In *IEEE International Conference on Distributed Computing in Sensor Systems (DCOSS)*, pages 389–404, 2006.
- [7] Firoozfam, P. and Negahdaripour, S. Theoretical accuracy analysis of n-ocular vision systems for scene reconstruction, motion estimation, and positioning. In *IEEE 3D Data Processing, Visualization and Transmission (3D PVT)*, pages 888–895, 2004.
- [8] Hartley, R. and Zisserman, A. *Multiple View Geometry in Computer Vision*. Cambridge University Press, New York, NY, USA, 2 edition, 2003.
- [9] Liu, W., Hu, C., He, Q., and Meng, M.Q.-H. A three-dimensional visual localization system based on four inexpensive video cameras. In *2010 IEEE International Conference on Information and Automation (ICIA)*, pages 1065–1070, June 2010.
- [10] Mason, S. Heuristic reasoning strategy for automated sensor placement. *Photogrammetric engineering and remote sensing*, pages 1093–1102, 1997.
- [11] Morye, A. A., Ding, C., Roy-Chowdhury, A. K., and Farrell, J. A. Constrained optimization for opportunistic distributed visual sensing. In *American Control Conference (ACC), 2013*, pages 6294–6301, June 2013.
- [12] Morye, A. A., Ding, C., Song, B., Roy-Chowdhury, A., and Farrell, J. A. Optimized imaging and target tracking within a distributed camera network. In *American Control Conference (ACC), 2011*, pages 474–480, June 2011.
- [13] Olague, G. and Mohr, R. Optimal camera placement for accurate reconstruction. *Pattern Recognition*, pages 927 – 944, 2002.
- [14] Petersen, K. B. and Pedersen, M. S. *The Matrix Cookbook*. Technical University of Denmark, nov 2012. Version 20121115.
- [15] SMEyeL. Smart Mobile Eyes for Localization (SMEyeL). <https://www.aut.bme.hu/Pages/ResearchEn/SMEyeL/Overview>, 2013.
- [16] Szalóki, D., Koszó, N., Csorba, K., and Tevesz, G. Marker localization with a multi-camera system. In *2013 IEEE International Conference on System Science and Engineering (ICSSE)*, pages 135–139, 2013.
- [17] Szalóki, D., Koszó, N., Csorba, K., and Tevesz, G. Optimizing camera placement for localization accuracy. In *14th IEEE International Symposium on Computational Intelligence and Informatics (CINTI)*, pages 207–212, 2013.

- [18] Zhao, J., Haws, D., Yoshida, R., and Cheung, S. S. Approximate techniques in solving optimal camera placement problems. In *2011 IEEE International Conference on Computer Vision Workshops (ICCV Workshops)*, pages 1705–1712, Nov 2011.
- [19] Zhou, Q. and Aggarwal, J. K. Object tracking in an outdoor environment using fusion of features and cameras. *Image and Vision Computing*, pages 1244 – 1255, 2006.