Optimal Selection of Camera Configurations for Multi-Camera Networks

Roopashree Shailendra[1]

Abstract— The optimal selection of camera configurations (camera locations, orientations, e.t.c) for multi-camera networks remains unsolved. Most of the previous approaches do not generalize well to large scale networks. They only focus on proposing various objective functions to achieve different tasks. To tackle with this problem a trans-dimensional simulated annealing algorithm is proposed. Comparison with the state-of-the-art method based on binary integer programming (BIP) show that the proposed approach also offers similar performance on small scale problems and better capability to deal with large scale problems.

Key words: Binary integer programming, camera placement, camera planning, multi-camera networks, optimization method, sensor planning, trans-dimensional simulate annealing

I. INTRODUCTION

A typical video surveillance installation comprises a potentially large network of cameras that are used to maintain remote observation of the environment (e.g. a car park or shopping mall). The cameras are usually placed to maximize the spatial coverage and also to ensure adequate visibility of key locations such as doorways or vulnerable sites (e.g. automatic cash-teller machines - ATM's). The selection and placement of cameras in a multi-camera system for optimal solution influence significantly the design of image processing algorithm for a particular application. The problem is addressed as selecting optimal camera configuration (camera orientation, location e.t.c) in a multi-camera network, considering a number of user specified constraints such as the maximum video coverage of the physical space.

Intelligent video surveillance (IVS) can provide automated services, such as abrupt incursion detection, shortest path recommendation using traffic jam analysis, robbery monitoring, and people counting by using network of cameras.

For these camera networks, it is of imperative importance to determine the optimal camera configuration before the deployment of cameras, as to reduce the modification cost and also the total number of cameras to accomplish the same level of utility. Optimal camera configuration of IVS system radically influences image processing algorithms.

In order to assist the design of realistic camera surveillance systems, a generalized statistical framework, capable of dealing with range of objectives, 1) full floor coverage using minimum number of cameras, 2) full floor coverage and redundant coverage of critical areas and 3) face detection potential of 100% while minimizing the number of cameras. 4) Improving existing cameras' panning and tilting angle to achieve better system utility.

The proposed algorithm Trans-Dimensional Simulated Annealing (TDSA) estimates the optimal camera parameters including the number of cameras. For small scale problems, TDSA algorithm offers similar solutions to the optimal ones produced by Binary Integer Programming (BIP). But, for larger scale problems BIP is clearly infeasible.

II. RELATED WORK

The Art Gallery Problem (AGP) is the assignment of guards to different positions in an art gallery, in order to achieve the maximum visual coverage of the walls represented by a polygon of n vertices. The general camera placement problem is first defined with assumptions that are more consistent with the capabilities of real-world cameras. The region to be observed by cameras may be volumetric, static or dynamic, and may include holes that are caused, for instance, by columns or furniture in a room that can occlude potential camera views. Given the set of all constraints, the problem is to find the optimal placements for a set of cameras in an area of interest, satisfying the constraints and minimizing a given cost function. The problem is presented as a Binary Integer Programming which is solved by a Branch and Bound algorithm. It is not clear how the optimal number of cameras is determined, which is attempted here to address.

Camera planning is formulated as a maximum a-posteriori model selection and optimization. In earlier work [9] it showed that the proposed system is capable of dealing with the objective of maximal floor coverage and here the previous work is extended to include two more scenarios which are face detection and re-planning of existing network as well as an in-depth analysis on the behavior of the proposed method with varying critical parameters.

III. PROBLEM FORMULATION

A. Generalized Framework:

In the generalized framework the camera model used is not overly restrictive. Each camera consists of n_p number of parameters $c_i = [p_1, p_2, ..., p_{n_p}]^{\mathrm{T}}$ which includes the camera location in x, y, z directions of the area to be monitored, panning, tilting angle, lens type etc. The camera parameter space is denoted as $c = R^{n_p}$ given a particular network of cameras $\theta = [c_1, c_2, ..., c_{n_c}]^T$, a set of constraints $r = \{r_j \mid r_j \in \{r_1, r_2, ..., r_{n_r}\}\}$ and a description of the area to be monitored ξ there must exist some function $L(r \mid \theta, \xi)$ that determines how well the constraints r are jointly satisfied by θ. For example, if required to achieve a frontal face capturing rate of 80%, but a particular set of cameras can only achieve 60%, then it may be said that the cameras have achieved 60/80 = 75% of the requirement.

Given the camera parameter space C, a set of user requirements r, an input environment ξ, the camera placement problem can be defined as the selection of the camera configuration that meets r while minimizing the number of cameras used. Since the environment ξ is often a constant, it is omitted in all subsequent equations. In the context of camera placement, the model of a camera network configuration is reflects the number of cameras in the configuration and each camera in the configuration is considered as a random variable over the space C.

The camera placement problem is defined by the joint posterior,

$$\phi_{opt} = \arg\max_{\phi_k \in \chi} \{p(k, \theta_k) | r\}, \tag{1}$$

where $\phi(k,\theta_k)$ denotes a camera configuration which contains a model indicator k as well as camera parameters of the model θ_k. This formulation can be interpreted as: Given there is an observation that the list of constraints and requirements have been satisfied, the problem is to find the optimal model k and the optimal parameters θ_k that are most likely to have led to this observation. For example, if the requirement is covering the maximum amount of the floor with a given number of cameras, Equation (1) can then be interpreted as finding the most likely camera configuration that has caused maximum coverage to be observed (satisfied).

Expanding Equation (1) using Bayes theorem,

$$\phi_{opt} = \arg\max_{\phi_k \in \chi} \{L(r|k,\theta_k) p(\theta_k|k) p(k)\}. \tag{2}$$

The first term $L(r|k,\theta_k)$ is the probability of satisfying the requ- irements and constraints r by the given set of camera parame- ters θ_k and is therefore called the likelihood. The second term $p(\theta_k|k)$ is termed the parameter prior since it defines the prior probability of the set of camera parameters. The prior term allows the user to set preferences on the parameters of the cameras. For example, wall locations can be preferred through assigning high prior to cameras that are located on walls and omni-directional cameras may be unwanted by associating them with low prior. In most settings where all cameras are treated equally, this term can be dropped.

The formulation (Equation (2)) can be converted to a penalized model selection problem for further investigation. An equivalent form of Equation (2) can be obtained as,

$$\phi_{opt} = \arg\min\{-\log\{L(r\,|\,k,\theta_k)\,p(\theta_k\,|\,k)\} + \log\{1/p(k)\}\} \tag{3}$$

B. Scene Models:

In this work, consider the most used scene models in practice: the floor plan of the environment of interest. Most available floor plans are in two dimensional drawings. Therefore restrict the scene models to be two dimensional. The axis is defined such that the real world x-axis is aligned with the floor plan image x-axis. The y-axis is aligned with the negative direction of the floor plan's y-axis and the z-axis points upwards from the ground plane.

C. Camera Models:

A range of camera models considered: Static perspective cameras, PTZ cameras and Omni-directional cameras in 3D space.

(1) Static Perspective Cameras: The model static perspective cameras is modeled using the perspective pinhole camera.. The model can be expressed as, $x = PX$, where $x = [x, y, z, 1]^T$ is the world coordinate, in projective 3-space (p^3). $x = [x, y, z]^T$ is the image coordinate in projective 2-space (p^2), and P is a 3×4 projection matrix that maps the world coordinates to image coordinates. The matrix P may be decomposed as $P = K[R|-Rt]$, where R is a rotation matrix representing the camera orientation and t is a translation vector representing camera center in the world coordinate frame. The matrix K is referred to as the camera calibration matrix and it consists of 6 intrinsic camera parameters: focal length (f_x, f_y), principal point (x_0, y_0) and skew k. For most modern cameras, $f_x = f_y$ and k is zero.

(2) PTZ Cameras: PTZ cameras are used to extend the FoV of static perspective cameras. They undergo panning and tilting motions only when required or as programmed. The zooming option is mostly used when the operator observes events of interest and would like to zoom in for greater details. Hence each PTZ camera is treated as an aggregation of all static perspective cameras at different pan, tilt combinations.

(3) Omni directional Cameras: Omni directional cameras are used for wide area surveillance due to their 360-degree FoV. There are mainly two types of omni directional cameras: dioptric and catadioptric. Dioptric cameras are created with a fisheye lens and catadioptric cameras are built from a pinhole camera with a parabolic mirror. The viewing regions are simplified as circular disks on the grounds which are restricted by the maximum dmax and the minimum dmin distances the camera can visualize. The heights of the cameras are taken into account by projecting $d_{\max}$ and $d_{\min}$ onto the ground.

D. Camera Model Summary:

All of the camera models can be summarized into a single 6D vector: ci = $[x, y, z, \alpha, \beta, \omega]^T$, in which (x, y, z) is the camera location in the real-world coordinate system. α and β are panning and tilting angles that represent the camera poses. ω is an indicator variable for the type of the camera. The $d_{\max}$ variable is used to specify the largest range between an object of interest and the camera center at which the object is considered visible (maximum viewing distance).

IV. IMPORTANT USER REQUIREMENTS

The following objectives are considered.

A. *Coverage:*

Coverage can be stated as finding of the least number of cameras to achieve a defined level of coverage or finding of the maximum coverage given a fixed number of cameras.

Coverage of the cameras are determined by computing the intersections of their viewing frustums with the ground plane z = 0 for perspective and PTZ cameras. When the tilting angles are small enough, $d_{\max}$ is used to exclude the intersecting regions that are too far away from the camera center. For omnidirectional cameras, the coverage areas are defined by $d_{\min}$ and $d_{\max}$ directly.

If denoting the floor coverage of each camera c_i, i ∈{1,...,k} in a camera configuration $\phi=(k,\theta_k)$ as g_i, then the floor coverage of the configuration can be represented as

$$COV(\phi)=\frac{\cup_{i=1}^{k} g_i}{g_{tot}} \tag{4}$$

Where g_{tot} is the total area coverable by the entire candidate camera set. The likelihood function as requested in Equation (2) for this objective can be defined as,

$$L(r\mid\phi)=GAU(COV(\phi),r_{\mathrm{cov}},\sigma_{\mathrm{cov}}) \tag{5}$$

Where r_{cov} is the desired coverage percentage and σ_{cov} is the co-efficient controlling the width of the GAU function.

B. *Redundant Coverage:*

In real deployments of camera networks for surveillance purposes, it is necessary that a number of critical areas, such as entrances and prohibited areas, are to be monitored by one or more cameras. The total area, as usual, is required to be covered to pre-set coverage percentage rcov. The goal for this objective is to find the set of cameras that satisfy these requirements and at the same time uses the least number of cameras.

$$L(r\mid\phi)=GAU(COV_{tot}(\phi),r_{tot},\sigma_{tot}) \times GAU(COV_{cri}(\phi),r_{cri},\sigma_{cri}) \tag{6}$$

where COV_{tot} and COV_{cri} are the percentage of the total area and the percentage of critical regions that are covered by the required number of cameras. r_{tot} and r_{cri} are the desired coverage percentage which is 100%.

V. OPTIMAL CAMERA PLACEMENT

A. *Discretization:*

Preferably, the camera parameters are continuous (except the camera type). A camera can be positioned anywhere in an environment and posed at any angle. The environment maps are divided into grids to allow easier computation of cameras' coverage regions. The locations where cameras can exist are restricted to a number of location samples, which are represented as crosses in Figure 1 and Figure 1(a) and Figure 1(b). Similarly, the panning and tilting angles of the cameras (except omni-directional cameras) are also sampled. The sampling of parameters allows the construction of the set of candidate cameras (also sometimes referred to as the sampled set), from which an optimal subset is to be selected to satisfy the user constraints.

B. *Trans-Dimensional Simulated Annealing:*

The trans-dimensional simulated annealing [10] is used to solve the stochastic problem. Simulated Annealing (SA) is a class of algorithms capable of locating good near-optima of objective functions in large search spaces. The term simulated annealing derives from the interesting observation that as a heated material slowly cools down, its atoms will line up in a rigid pattern corresponding to a state of minimum energy, provided that the cooling process is sufficient slow. SA algorithms mimic this process and have been proven to converge. Trans-dimensional simulated annealing is a class of algorithms that extend the traditional simulated annealing by allowing moves that not only change the parameters of the model but as well move between plausible models.

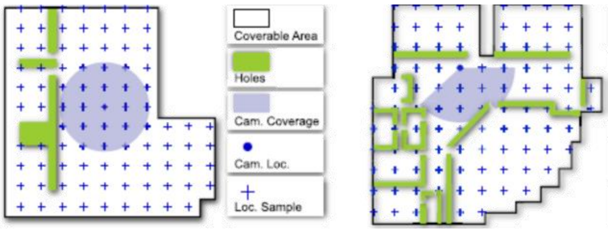

Fig.1. Floor plans used in the experiments. (a) Floor plan A (638x616px2)is adopted from [4] and (b) floor plan B (804x733px2) is modified from a real floor plan of a university building. Both floor plans consist of only polygonal areas and holes. The crosses are the location samples where cameras can be placed.

For the proposed algorithm, birth, death and update moves have been selected. The birth and death constitute a pair of reversible moves that allow the dimension of the current state of the chain to grow from k to k + 1 and decrease from k to k−1. The reverse move of the update move is itself.

Birth and Death Moves: Birth and death moves are a pair of reversible moves that facilitate model dimension changes. The birth move proposed is rather simple: for a given state φ, a new camera is created randomly and added to φ to form φ?. Similarly the death move is achieved by randomly removing a camera from the existing camera configuration.

$$\alpha_{birth}=\min\left\{1,\frac{b_T(\phi')Pm_d n_{\max}}{b_T(\phi)Pm_b(n_k+1)}\right\}, \tag{7}$$

$$\alpha_{death}=\min\left\{1,\frac{b_T(\phi')Pm_b n_k}{b_T(\phi)Pm_d n_{\max}}\right\}, \tag{8}$$

where $n_{\max}$ is a user defined maximum allowable dimension of the models. The term $n_{\max}$ was originally derived to be n_{tot}, where n_{tot} is the total number of candidate cameras from which an optimal subset is to be selected.

Update Moves: The update move is an important move that allows the estimation of a better camera configuration while preserving the dimension of the state. It starts by first randomly selecting an existing camera from the current state of the configuration, i.e. select c_i from θ_k and then update this camera with a consecutive random walk of each parameter.

$$\alpha_{update} = \min\left\{1, \frac{b_T(\phi')}{b_T(\phi)}\right\} \qquad (9)$$

Algorithm Summary: The steps of the proposed trans-dimensional simulated annealing algorithm for determining the optimal number of cameras as well as the parameters of each camera are summarized. The algorithm makes use of the following subroutines:

- c ← RANDN(a,b) one sample of a uniform distribution with range (a,b) is randomly generated.
- s ← J(φ,γ) computes the value of the objective function of a configuration φ as defined in $J(\phi) = -\log\{L(r \mid k, \theta_k \mid k)\} + \gamma n_k$.

Where γ is introduced in (equation 3).

- φ?← BIRTH(φ) the birth move is used to generate a new candidate configuration based on the current configuration φ.
- α ← ALPHAB(φ?,s) the acceptance ratio of the birth move is computed using equation (7).
- φ?← DEATH(φ) the death move to generate a new candidate configuration which is based on the current configuration φ.
- α ← ALPHAD(φ?,s) computes the acceptance ratio of the death move using Equation (8).
- φ?← UPDATE(φ) uses the birth move to generate a new candidate configuration based on the current configuration φ.
- α ← ALPHAU(φ?,s) computes the acceptance ratio of the update move using Equation (9).

The complexity of the methodology can be factored into the complexity of the candidate camera generation and optimization. First process involves heavy geometric computations. The proposed TDSA behaves differently compared to BIP, which belongs to a class of NP-Hard problems [4]. Most common implementations of BIP, such as Branch and Bound has a worst-case complexity of 2^N, where N is the total number of possible camera combinations. Instead of computing a global optimum solution, the TDSA runs for a predefined number of iterations and output the best solution found.

Algorithm 1 Trans-dimensional simulated annealing for determining the optimal camera configuration.

Function $\{\phi_{opt}, s_{\max}\} \leftarrow TDSA(T_0, T_e, p_{m_b}, p_{m_d}, \phi_0, l, \gamma, \rho)$

Input:

T_0, T_e – The initial temperature and the end temperature.

P_b, P_d, P_u – The probability of choosing the birth, death and update move respectively.

$\phi^0 = (k^0, \theta_k^0)$ – The initial state of the Markov chain, where

k^0 – The initial model order.

θ_k^0 – The initial camera configurations.

l, γ, ρ – The length of each Markov chain, model penalty parameter and cooling effect.

Output:

ϕ_{opt} – The optimal model and optimal camera parameters.

$s_{\max}$ – The optimal value of the objective function.

begin
  $T \leftarrow T_0, \phi \leftarrow \phi_0, s_{\max} \leftarrow -\infty$
  $s_c \leftarrow -J(\phi, \gamma)$
  while $T \geq T_e$ do
    for $j \in \{1,2,3,\ldots l\}$ do
      $\beta \leftarrow RAND(0,1)$
      if $\beta \leq p_b$ then
        $\phi' \leftarrow BIRTH(\phi)$
        $s_p \leftarrow -J(\phi', \gamma)$
        $\alpha \leftarrow ALPHA_B(\phi', s_p)$
      else if $\beta \leq p_b + p_d$ then
        $\phi' \leftarrow DEATH(\phi)$
        $s_p \leftarrow -J(\phi', \gamma)$
        $\alpha \leftarrow ALPHA_D(\phi', s_p)$
      else
        $\phi' \leftarrow UPDATE(\phi)$
        $s_p \leftarrow -J(\phi', \gamma)$
        $\alpha \leftarrow ALPHA_U(\phi', s_p)$
      end
      $\mu \leftarrow RAND(0,1)$
      if $\mu < \alpha$ then
        $\phi \leftarrow \phi', s_c \leftarrow s_p$
        if $s_c > s_{\max}$ then
          $s_{\max} \leftarrow s_c, \phi_{opt} \leftarrow \phi$
        end
      end
    end
    $T \leftarrow \rho T$
  end
end

VI. EVALUATION

The generalized framework and the proposed TDSA algorithm are evaluated in a number of ways. First we compare TDSA with the BIP method of Erdem and Sclaroff [4], the Greedy algorithm of Zhao et al. [6] and the Dual Sampling algorithm introduced by Horster and Lienhart [5].

A series of experiments were conducted using floorplan A and B to evaluate the performance of the proposed approach against the alternative approaches: Erdem06 [4], Zhao09 [6] and Horster09 [5] under different setups shown in Table I. The results are plotted in Figure 3(a) and Figure 3(b).

	Floorplan	Cam Model & FoV	Cam. loc. sep (px)	Orient. sep	Tilt. sep	$d_{\min}$, $d_{\max}$ (px)	Num. cai
1	A	Omni., 360°	90	NA	NA	190, 0	31
2	A	Omni., 360°	60	NA	NA	190, 0	79
3	A	Omni., 360°	30	NA	NA	190, 0	317
4	A	Pers., (47.5°, 39.6°)	90	0° : 10° : 350°	0° : 5° : 55°	190, NA	13392
5	A	Pers., (47.5°, 39.6°)	60	0° : 10° : 350°	0° : 5° : 55°	190, NA	34128
6	A	Pers., (47.5°, 39.6°)	30	0° : 10° : 350°	0° : 5° : 55°	190, NA	136944
7	B	Omni., 360°	90	NA	NA	190, 0	40
8	B	Omni., 360°	60	NA	NA	190, 0	100
9	B	Omni., 360°	30	NA	NA	190, 0	375
10	B	Pers., (47.5°, 39.6°)	90	0° : 10° : 350°	0° : 5° : 55°	190, NA	17280
11	B	Pers., (47.5°, 39.6°)	60	0° : 10° : 350°	0° : 5° : 55°	190, NA	43200
12	B	Pers., (47.5°, 39.6°)	30	0° : 10° : 350°	0° : 5° : 55°	190, NA	16200(

Fig. 2: Different Setups Used In All The Experiments

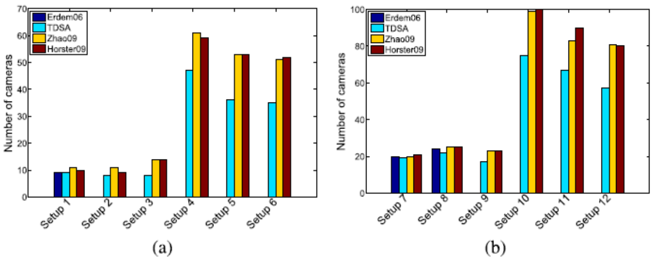

Fig. 4: Mapping nonlinear data to a higher dimensional feature space

Notes: Each setup is a combination of a specific floorplan (A or B), a specific type of camera (omnidirectional or perspective), a specific camera location seperation i.e the seperation between each valid location where a camera can be placed. For perspective cameras, it also includes the range and increment of camera orientation and tilt angles in the form of a:b:c, where a is the minimum angle, b is increment and c is the maximum angle. Each combination has resultant pool of candidate cameras from which an optimal subset is to be selected to achieve user specified tasks. The size of the candidate pool is listed as the last column of the table.computed placements strategies of Setup 1 are shown in Figure 3.

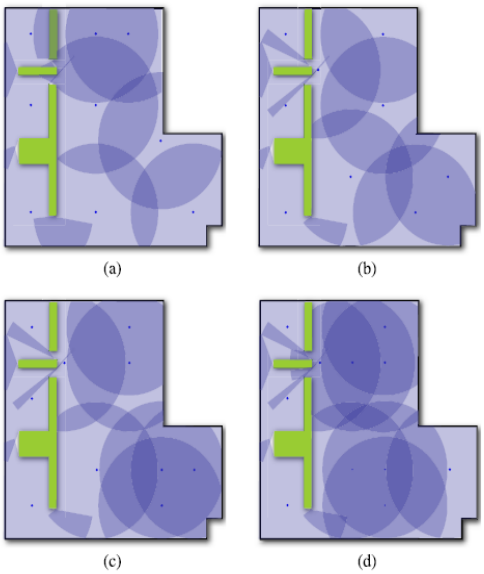

Fig. 3: Camera configurations computed by all four methods using Setup 1.

VII. CONCLUSION

this paper, the selection of optimal camera configurations in multi-camera systems has been offered. It influence significantly the subsequent design of image processing algorithm for a particular application using the multi-camera system. The solution that is much more versatile than the current available solutions. The proposed approach includes a generalized statistical formulation of the problem, taking into account a set of user constraints, the number of cameras and the parameters of the cameras. Optimal configuration is computed using a trans-dimensional simulated annealing algorithm.

The methodology has a number of advantages over the alternatives. First, it is able to pinpoint good near-optima even when the problem space is large. It's been shown that similar performance to the optimal BIP solution can be obtained in small scale problems.

TDSA is computationally intensive as time taken for our complex experiments took a few hours. By definition the camera planning problem is off-line, so the processing time taken by TDSA does not present a concern. Furthermore, if exploit the parallel processing architecture that exists in modern processor, significant improvement in processing time may be feasible.

REFERENCES

[1] P. Chen, P. Ahammad, C. Boyer, I. H. Shih, L. Leon, E. Lobaton, et al.,"Citric: A low-bandwidth wireless camera network platform," in Proc. 2nd ACM/IEEE ICDSC, 2008, pp. 1–10.

[2] T. Wark, P. Corke, J. Liu, and D. Moore, “Design and evaluation of an image analysis platform for low-power, low-bandwidth camera networks,” in Proc. Workshop Appl., Syst., Algorithms Image Sens.,
2008, pp. 7–11.
[3] A.Rowe, D. Goel, and R. Rajkumar, “Firefly mosaic: A vision-enabled wireless sensor networking system,” in Proc. 28th IEEE Int. RTSS, 2007, pp. 459–468.
[4] U. M. Erdem and S. Sclaroff, “Automated camera layout to satisfy task-specific and floor plan-specific coverage requirements,”Comput. Vis. Image Understand., vol. 103, no. 3, pp. 156–169,2006.
[5] E. Horster and R. Lienhart, “Optimal placement of multiple visual sensors,” in *Multi-Camera Networks: Concepts and Applications*. Amsterdam, The Netherlands: Elsevier, 2009.
[6] J. Zhao, S.-c. S. Cheung, and T. Nguyen, “Optimal visual sensor network configuration,” in Multi-Camera Networks: Concepts and Applications.Amsterdam, The Netherlands: Elsevier, 2009.
[7] V. Chvatal, “A combinatorial theorem in plane geometry,” J. Combinat. Theory, Ser. B, vol. 18, no. 1, pp. 39–41, 1975.
[8] M. C. Couto, P. J. de Rezende, and C. C. de Souza, “An exact algorithm for minimizing vertex guards on art galleries,” Int. Trans. Oper. Res., vol. 18, no. 4, pp. 425–448, 2011.
[9] J. Liu, C. Fookes, T. Wark, and S. Sridharan, “On the statistical determination of optimal camera configurations in large scale surveillance networks,” in *Proc. 12th Eur. Conf. Comput. Vis.*, 2012, pp. 44–57.
[10] S. P. Brooks, N. Friel, and R. King, “Classical model selection via simulated annealing,” *J. R. Statist. Soc., Ser. B, Statist. Methodol.*, vol. 65, no. 2, pp. 503–520, 2003.