Int J Comput Vis
DOI 10.1007/s11263-012-0587-7

Modeling Coverage in Camera Networks: A Survey

Aaron Mavrinac · Xiang Chen

Received: 16 June 2011 / Accepted: 9 October 2012

Abstract Modeling the coverage of a sensor network is an important step in a number of design and optimization techniques. The nature of vision sensors presents unique challenges in deriving such models for camera networks. A comprehensive survey of geometric and topological coverage models for camera networks from the literature is presented. The models are analyzed and compared in the context of their intended applications, and from this treatment the properties of a hypothetical inclusively general model of each type are derived.

Keywords Camera networks · Coverage geometry · Coverage topology · Sensor planning · Calibration

1 Introduction

Visual coverage is an important quantifiable property of camera networks, describing from a pragmatic standpoint what the system can see—that is, what visual data it is physically capable of collecting—and thus informing the most fundamental requirement of any computer vision task. Virtually all camera network applications depend on or can benefit from knowledge about the coverage of individual cameras, the coverage of the network as a whole, or the relationships between cameras in terms of their coverage.

It is therefore not surprising that there exists a significant body of literature on modeling camera network coverage, spanning back to the earliest days of camera network research. With such diverse applications as sensor planning, optimal camera placement, camera reconfiguration, camera selection, calibration, tracking correspondence, and optimal load distribution, and often broad variation of specific objectives and constraints within each, numerous structures have emerged for capturing coverage information. Given the mixed lineage of camera networks, these models are influenced by earlier work in computer vision and in sensor networks; they also exhibit their own innovations to meet the unique challenges of the field. In this correspondence, we examine the state of the art in modeling camera network coverage, with a view to how the properties of various models relate to their intended applications.

Coverage models can be classified into two different major types. *Geometric coverage models*, the focus of Sect. 2, are concerned with the physical area or volume of the scene covered by a camera network. Given some information about the camera viewpoints, the physical structure of the scene, and the task to be performed, such a model seeks to quantify whether or not a particular stimulus (minimally described as a point in $\mathbb{R}^n$) is covered, and, in some cases, how well. A set structure with geometric definitions lends itself naturally to this purpose. *Topological models* are combinatorial structures describing the relationships between cameras in a network with respect to their coverage. This survey considers two types: *coverage overlap models* in Sect. 3, which describe pairs or groups of cameras with mutual coverage of the scene, and *transition models* in Sect. 4, which describe the more abstract relationships arising from the possibility (or probability) of a moving agent transiting from one camera's region of coverage to that of another. A topological coverage model is typically formalized as a graph—or as some more general graph-like structure (e.g., simplicial complex,

A. Mavrinac · X. Chen (✉)
Department of Electrical and Computer Engineering,
University of Windsor, Windsor, ON, Canada
e-mail: xchen@uwindsor.ca

A. Mavrinac
e-mail: mavrin1@uwindsor.ca

Published online: 16 November 2012

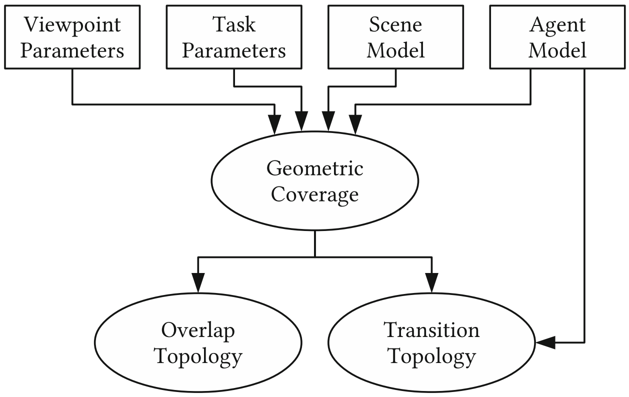


Fig. 1 Hierarchy of information in coverage models

hypergraph)—in which vertices represent the individual camera nodes, and edges indicate coverage relationships.

The ideal geometric coverage model is derived from four primary sources of information: the viewpoint parameters (including position, orientation, and intrinsic parameters) of all cameras in the network, a model of static objects in the scene, a probabilistic model of agent dynamics, and a set of independent task-specific requirements. The coverage overlap topology may be thought of as a distillation of some of this information into a combinatorial structure; ideally, it is derived from the quantified geometric overlap between individual cameras or coverage cells. Transition topology is ideally derived from both the geometric coverage (at some level of granularity) and from the agent dynamics model, capturing information not present in either the geometric coverage model or the overlap topology. Figure 1 illustrates the ideal hierarchy of information. All three types of model can be estimated directly from various forms of captured visual information in the absence of some or all of the primary sources.

We take a common approach in examining each type of coverage model. First, we describe the abstract, fundamental structure of the model class, a comprehensive generalization of all reviewed models which provides the reader with an overall understanding and consistent terminology for the ensuing discussion. Next, we introduce the various surveyed works presenting models of this type, in the context of their target applications. The models being thus introduced, we compare and contrast the realizations (or lack thereof) of each individual aspect of the general model. Finally, with respect to the ultimate goal of producing a general, accurate model of the given class, we discuss the state of the art and expose the open research questions by highlighting the most promising contributions to date.

1.1 Applications

Geometric coverage models are used in a variety of *sensor planning* applications. In single-camera vision systems, the objective is to find a viewpoint which satisfies coverage requirements, often on a well-defined target in a controlled scene. Analytic solutions are common. By contrast, the *optimal camera placement* problem in multi-camera systems is usually far less structured, with an objective of maximizing coverage under a cost constraint or minimize cost under a coverage constraint; typical approaches include optimization by linear programming or search heuristics such as genetic algorithms. In this context, the coverage of multiple cameras is modeled simultaneously within the scene, and the overall coverage performance of the network can be quantified. A special case of this problem is *camera reconfiguration*, wherein only certain viewpoint parameters of a fixed set of cameras are variable, and must be optimized for maximum coverage performance in an online context (which carries its own unique constraints). Camera coverage models can also be used for online *camera selection*; typical problem instances involve determining one or more optimal views of a given parametric target, subject to energy and other constraints.

Coverage overlap topology models have several important applications in computer vision and camera networks. The offline problem of *multi-view registration* has been approached by applying combinatorial optimizations to an equivalent structure, derived from view overlap, to control the registration process. Similarly, a number of multi-camera *calibration* algorithms, which estimate the relative poses of the cameras in the network, proceed according to optimized paths in a combinatorial representation of coverage overlap. Knowledge of overlap topology may also greatly improve the performance of *direct tracking correspondence*, in which tracked agents are matched between cameras with simultaneous coverage. This is a subproblem of the more general predictive tracking problem and is the basis for *camera handoff* across overlapping cameras. Coverage overlap models have also been applied to *scheduling* problems in camera networks with energy constraints, such as duty cycling, triggered wake-up, and load distribution.

Transition topology models are tailor-made for online *predictive tracking* applications in camera networks. The objective here is to estimate the probability and duration of an agent transition from one region of coverage (e.g., a camera) to another. In a sense, this is a generalization of the direct tracking correspondence problem to cameras which do not necessarily have coverage overlap, and is likewise used for camera handoff and other tracking tasks.

1.2 Scope of this Survey

Our interest in this survey is in geometrical and topological models of (generally) multi-camera coverage, and, where applicable, on methods of estimating their parameters. We cover posterior applications only insofar as they elucidate the motivations behind decisions in the design of the models; the inclusion of a full exposition of every application for its own sake would produce a prohibitively long work, tantamount to a survey of a considerable cross-section of the entire camera network field, and more to the point, would not be particularly germane to an understanding of the theory and state of the art of coverage models.

In order to remain focused on the stated purpose, the works surveyed here are primarily drawn from literature specifically on camera networks or other multi-camera systems. Where appropriate, the origins of certain concepts, often from the broader computer vision and sensor network fields, are mentioned to provide historical context.

The survey of geometric coverage models in Sect. 2 includes some discussion of single-camera sensor planning and next best view work from the computer vision literature, as the level of detail in some of these models provides a comparative baseline for later multi-camera work. It also includes mention of some general sensor network models where appropriate, although the focus is primarily on directional sensor networks (usually a label for camera networks).

Section 3 considers some coverage overlap models developed for multi-view registration, which is not necessarily a camera network application, as the views are typically obtained from a single camera in a video sequence or multi-image scan. However, since multiple views are theoretically equivalent to multiple cameras, the model structures and techniques are of interest here.

The transition models presented in Sect. 4 address an issue largely endemic to camera networks, and accordingly, all of the works surveyed are specific to the field.

2 Coverage Geometry

2.1 Anatomy of a Geometric Coverage Model

A *geometric coverage model* describes the coverage of a geometric space by a sensor or sensor system. Losing some generality, an intuitive example is a region of a three-dimensional Euclidean space which is imaged by a camera sensor satisfactorily for a given image processing task. A precise definition of a generalized model derived from the works surveyed follows.

2.1.1 Coverage Model

A *sensor system* is an entity which detects *stimuli* for the purpose of executing a *task*. In general, this system may physically comprise a single sensor, multiple sensors, or part of one or more sensors' ranges, with one or more sensing modalities. Stimuli are uniquely defined in a *stimulus space* $\mathcal{S}$; in most cases, stimuli are 2D or 3D points in the stimulus space $\mathbb{R}^2$ or $\mathbb{R}^3$, respectively, but exceptions exist when geometrical characteristics of the stimulus other than spatial position (e.g., direction) affect coverage as well.

A stimulus $\mathbf{p} \in \mathcal{S}$ is considered covered by a sensor system if it yields a response sufficient to achieve the given task. An ideal *coverage function*, therefore, is a mapping $C : \mathcal{S} \rightarrow \{0, 1\}$, where $C(\mathbf{p}) = 1$ indicates that a point is covered. Equivalently, one can speak of the covered volume $C \subset \mathcal{S}$. A more general definition $C : \mathcal{S} \rightarrow \mathbb{R}$ encompasses models which handle uncertainty and/or consider coverage quality with a *coverage grade*; this allows for (at least) relative assessment of coverage. Defining the function as $C : \mathcal{S} \rightarrow [0, 1]$, or equivalently, over any closed range which can be mapped linearly thereto, also allows absolute assessment; as well, extension of the subset notion is possible if one considers C a fuzzy subset of $\mathcal{S}$.

2.1.2 Coverage Criteria for Vision

All visual stimuli considered in the work reviewed herein can be reduced to *point features*, which have a single point of origin in Euclidean space and possibly other characteristics. Based on well-studied geometric imaging models (Faugeras 1993; Ma et al. 2004), various researchers have identified one or more of the criteria described here and incorporated them into their coverage models. We assume that the reader is familiar with the terminology and parameters of the standard camera model, and with basic camera optics. The collective set of intrinsic and extrinsic parameter values of a camera is termed a *viewpoint*, and the associated parameter space is the *viewpoint space* $\mathcal{V}$ (Figs. 2 and 3).

We identify three basic criteria (Tarabanis et al. 1994) which depend only on the viewpoint and a feature point in $\mathbb{R}^3$; two dimensional coverage models can be thought of as projecting these criteria onto the $\mathbb{R}^2$ plane.

- *Field of View*: The infinite subspace of $\mathbb{R}^3$ which can theoretically be imaged by the camera, determined by the horizontal and vertical apex angles (in turn, by the optics and physical image sensor size) and the pose (extrinsics) of the camera.
- *Resolution*: A constraint on the minimum [1] required resolution; translates directly into an upper limit on depth.
- *Focus*: A constraint on the acceptable sharpness of the image; given a maximum blur circle diameter, imposes upper and lower depth limits around the focus distance (this range is termed the *depth-of-field*).

[1] A maximum resolution constraint is conceivable, e.g., for privacy purposes, but we have not encountered this in the literature.

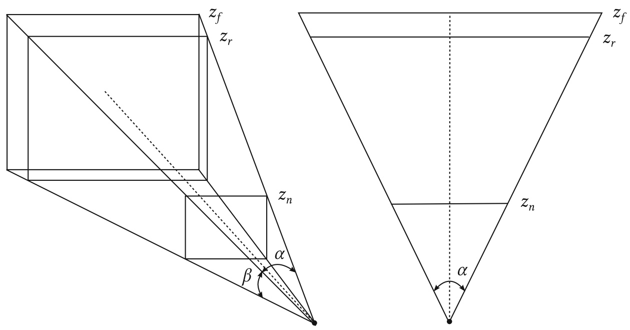


Fig. 2 Imaging criteria (Erdem et al. 2003; Tarabanis et al. 1994)—α and β are the field of view angles, z_r is the depth limit for resolution, and z_n and z_f are the near and far depth of field limits for focus

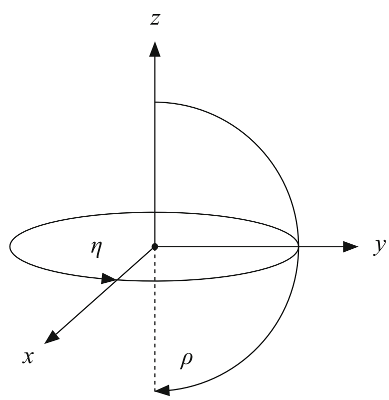


Fig. 3 Visual stimulus space—the spatial and view angle coverage criteria induce a stimulus space comprising three-dimensional position and direction

The field of view in combination with the depth constraints of resolution and/or focus are also sometimes termed the *viewing frustum*.

Considering the *view angle* to the feature—the direction of the surface normal at the feature point with respect to the camera's optical axis or the ray joining its principal point with the feature point—adds a fourth coverage criterion, and up to two additional (angular) dimensions to $\mathcal{S}$.

A feature may also be occluded, thus not covered, if the ray from the feature to the optical center of the camera is interrupted by an opaque physical object. We consider two criteria for occlusion, differing primarily in the type of information about the scene used to evaluate them: *static occlusion*, caused by static and/or deterministically dynamic objects, such as walls, and *dynamic occlusion*, caused by probabilistically dynamic objects, such as humans. Note that self-occlusion is typically handled by the view angle criterion, by imposing a maximum angle of 90°.

2.1.3 Task Definition

Some of the imaging criteria in the previous section require task-specific parameters in addition to the camera model parameters. These typically include the minimum resolution and maximum acceptable blur circle diameter for the resolution and focus criteria, respectively, as well as a maximum acceptable view angle. The scene information required by the occlusion criteria—a deterministic scene model and/or a probabilistic model of scene agent dynamics—is also part of the task definition. It should be noted that the information conveyed by the static scene model and the agent dynamics model may not be mutually independent; for example, agent motion may be constrained by the presence of walls, which also factor in static occlusion.

Besides this coverage model information, it is worth mentioning two other aspects of a task definition often encountered, both of which are known under various names.

The first is the *relevance function*, which indicates the subset of $\mathcal{S}$ which is of interest, and may be prioritized (graded). In general, this takes a form similar to that of a coverage function, $R : \mathcal{S} \to \mathbb{R}$. Depending on the task, it may indicate a large volume of the stimulus space, representing e.g., a set of rooms and hallways to observe, or a small, localized region of interest, representing e.g., a part to be inspected or a person to be tracked.

The second is the *allowable viewpoint set*, a subset of $\mathcal{V}$ encompassing all allowable viewpoints. Some dimensions may be constrained to single or discrete values dictated by hardware properties or fixed settings of the cameras, independent of the task. In general, however, such constraints are task-specific.

2.2 Geometric Coverage Models by Application

The basic function of a geometric coverage model is to evaluate the coverage of some region of the stimulus space (viz. a relevance function) by a sensor system. While not an end in itself, some form of this evaluation is a clear prerequisite for a family of camera network coverage problems.

One motivating application which predates camera networks is the offline *sensor planning* problem. The objective is to find a viewpoint which adequately covers the relevance function for a given task. This may be found via a generate-and-test search, or else the entire set of viewpoints may be solved analytically. Although the output of such methods is a subset of viewpoints in $\mathcal{V}$ which cover some $R \subset \mathcal{S}$, it is generally straightforward to invert the criteria to obtain the coverage $C \subset \mathcal{S}$ for some specific viewpoint in $\mathcal{V}$. Tarabanis et al. (1995) give an excellent survey of the topic. Typically, the target systems employ a single camera observing a relatively structured scene, and thus require (and can afford) highly accurate coverage models. Cowan and Kovesi (1988) and Tarabanis et al. (1995) present good examples.

In the multi-camera context, one encounters a similar problem most commonly known as *optimal camera* *placement*. The exact approach used in single-camera sensor planning does not scale well to multiple cameras, and there are typically additional design variables such as the number of cameras (with cost constraints), so nonlinear optimization techniques and search heuristics are the typical tools of choice, encouraging much simpler coverage models. Typically, the objective is to search for either the solution with maximum coverage given a fixed cost (or number of cameras), or the solution with minimum cost yielding some minimum coverage. The problem appears similar to the classic art gallery problem (O'Rourke 1987; González-Banos and Latombe 2001) frame it so, with their model assuming omnidirectional visibility and infinite range. Limiting visibility and range yields a more accurate model of coverage, but fundamentally changes the problem. Following the sensor network approach, Ma and Liu (2005b, 2007) propose a so-called boolean sector coverage model (derived from the common 2D disc model Wang 2010), enabling them to treat optimal camera placement similarly to a set covering problem (Tao et al. 2006; Liu et al. 2008). Qian and Qi (2008), Wang et al. (2009), and Jiang et al. (2010) further develop this direction. Erdem et al. (2003) and Erdem and Sclaroff (2006) approach the problem with a more realistic two-dimensional model; subsequent results using different coverage models and optimization techniques but similar basic method have been reported by Hörster and Lienhart (2006, 2009), Angella et al. (2007), and Zhao et al. (2008, 2009). Malik and Bajcsy (2008) similarly address optimal placement of stereo camera nodes. Yao et al. (2008) adapt this type of approach to surveillance networks with tracking and handoff tasks, adding a "safety margin" to their coverage model to enforce the necessary coverage overlap. The work of Mittal and Davis (2004, 2008) and Mittal (2006) extends the set of constraints to include dynamic occlusion, important in a significant subset of applications involving relatively high densities of moving agents.

The problem of online *camera reconfiguration* is fundamentally similar to optimal camera placement, but restricts the allowable viewpoints to those which can be arrived at by varying some set of parameters which can be controlled online (e.g., mobile platforms, pan-tilt mounts, motorized lenses), and allows for a dynamic relevance objective based on feedback from camera data. The coverage models and optimization techniques used may reflect the need for real-time online performance. Bodor et al. (2005, 2007) and Fiore et al. (2008) seek to optimize the configuration of cameras mounted on mobile robots for global scene coverage. Piciarelli et al. (2009, 2010) address reconfiguration of pan-tilt-zoom (PTZ) cameras, common in surveillance applications. Ram et al. (2006) and Erdem and Sclaroff (2006) both also touch on PTZ reconfiguration; the latter do so by introducing a time constraint to the optimal camera placement problem. Chen et al. (2010) focus on the view angle criterion in optimizing the configuration of rotating (panning) cameras.

Coverage evaluation is also useful in an online context for *camera selection*, which chooses an optimal subset of viewpoints for a localized relevance function, often subject to constraints such as energy costs. In the single-camera realm, this can be related to the *next best view* problem, approached by Reed and Allen (2000) and Chen and Li (2004) using coverage models similar to those used in sensor planning. Park et al. (2006) use a fairly realistic three-dimensional coverage model for camera selection, and acknowledge that a yet more sophisticated model could be substituted. The approach of Shen et al. (2007) is notable for assigning a scalar coverage metric to the stimulus space and for allowing task-specific weighting of the individual factors; they also touch on a version of the optimal camera placement problem. Soro and Heinzelman (2007) approach a slightly different problem: given a desired viewpoint directly, rather than a relevance function, their algorithm attempts to find the closest actual viewpoint (subject to energy costs).

For completeness, it is worth mentioning the geometric component of the topological coverage overlap model of Kulkarni et al. (2007), which differs from other geometric models surveyed here in that it is not analytically derived from a camera model. Instead, it is purely empirical: through a Monte Carlo process whereby a structured target is placed at an arbitrary number of random points in the scene, each camera with a view to the target at a given position estimates its pose, and each Voronoi cell around a target position forms a part of the geometric coverage of each camera that observed that position. In combination with the topological model, it is applied to scheduling problems. This model is discussed further in Sect. 3.

2.3 Analysis and Comparison of Geometric Models

Table 1 compares the nature and properties of a number of camera network coverage models from the literature, grouped by application. Since most of these models have been developed with specific applications in mind (indicated in the first column), it should be interpreted as a comment on the generality – not necessarily the validity or quality—of the models. The second column indicates the dimensionality of the model; a dimensionality of 2.5 indicates that the final representation is two-dimensional, but is derived from three-dimensional characteristics of the cameras and scene. The third column indicates whether the model is graded, i.e., whether it assigns to a point a scalar measure of coverage in some form (weighted, probabilistic, fuzzy, etc.); non-graded models are bivalent. The following four columns indicate which of the imaging coverage criteria (field of view, resolution, focus, view angle) are included. The final two columns indicate which occlusion criteria (static, dynamic)

Table 1 Comparison of selected geometric camera network coverage models

Model	Appl.	Properties		Imaging criteria				Occlusion	
		Dim.	Graded	FOV	Resol.	Focus	Angle	Static	Dynamic
Cowan and Kovesi (1988)	SP	● ● ●		✓	✓	✓	✓	✓	
Tarabanis et al. (1995)	SP	● ● ●	✓	✓	✓	✓		✓	
González-Banos and Latombe (2001)	OCP	● ●						✓	
Wang et al. (2009)	OCP	● ●	✓	✓					
Jiang et al. (2010)	OCP	● ●		✓				✓	
Erdem and Sclaroff (2006)	OCP	● ●		✓	✓			✓	
Hörster and Lienhart (2009)	OCP	● ●		✓	✓			✓	
Angella et al. (2007)	OCP	● ● ●		✓	✓	✓		✓	✓
Zhao et al. (2009)	OCP	● ● ●		✓			✓	✓	✓
Malik and Bajcsy (2008)	OCP	● ● ●		✓	✓		✓		
Mittal and Davis (2008)	OCP	● ● ◖	✓	✓	✓		✓	✓	✓
Bodor et al. (2007)	CR	● ● ◖	✓	✓	✓		✓		
Piciarelli et al. (2010)	CR	● ● ◖		✓				✓	
Park et al. (2006)	CS	● ● ●	✓	✓		✓			
Shen et al. (2007)	CS	● ● ◖	✓	✓			✓	✓	

SP sensor planning, *OCP* optimal camera placement, *CR* camera reconfiguration, *CS* camera selection

are included. It should be noted that, in some cases, the authors do not provide quantitative descriptions of some criteria or means of obtaining the information required to derive them.

2.3.1 Dimensionality

Although vision is an inherently three-dimensional phenomenon, many coverage models in various applications are two-dimensional. In such cases, to simplify the problem at hand, it is assumed (either implicitly or explicitly) that

– all cameras are positioned in a common plane,
– all targets are constrained to a common plane, and
– the scene consists of occluding vertical "high walls."

In models derived from the art gallery problem formulation, e.g., González-Banos and Latombe (2001), the choice reflects the fact that three-dimensional AGP is NP-hard (Marengoni et al. 2000). The vast majority of work on sensor network coverage problems (Meguerdichian et al. 2001) has employed two-dimensional disc models (Wang 2010) (although the three-dimensional case has been studied Huang et al. 2007), assuming a roughly planar environment. Some camera network models, including those of Ma and Liu (2005b, 2007), Liu et al. (2008), Wang et al. (2009), and Jiang et al. (2010), follow directly from this tradition, simply restricting the disc to a sector (Wang 2010) for directionality. Erdem and Sclaroff (2006) and Hörster and Lienhart (2009) do not appear to share this lineage, and explicitly cite the complexity of their respective optimization methods as motivating their restriction to two dimensions. The model of Yao et al. (2008) appears to be heavily influenced by that of Erdem and Sclaroff. In all of the preceding cases, the domain of camera coverage is explicitly planar.

In contrast, some two-dimensional models are not developed from the ground up as such. Bodor et al. (2005, 2007) and Mittal and Davis (2008) begin with three-dimensional analytic treatments of their respective constraints, but subsequent assumptions about the scene and viewpoint restrictions effectively reduce their models to the plane without loss of information. Shen et al. (2007) present a similar treatment of view angle—in particular, including the inclination angle between the sensor and a human subject's head with respect to the ground plane—in an otherwise two-dimensional model. Piciarelli et al. (2010) account for a three-dimensional field of view criterion by projecting the elliptical cross-section of their conical visible region onto the plane.

Early coverage models used in sensor planning, such as those of Cowan and Kovesi (1988) and Tarabanis et al. (1995), are fully three-dimensional: the gains in generality and accuracy clearly outweigh the added complexity in the single-camera case. These advantages have induced a number of multi-camera coverage models across the application spectrum to follow suit. Cerfontaine et al. (2006) describe a multi-camera method employing a three-dimensional coverage model presumably derived from the pinhole camera model, but give no details on the criteria. Park et al. (2006) fully describe their model with a three-dimensional viewing frustum; the multi-camera complexity is handled by dividing the covered volume into discrete parts and generating look-up tables for coverage grade. Angella et al. (2007) employ a three-dimensional model drawing heavily on the sensor planning literature. The models of Malik and Bajcsy (2008) and Zhao et al. (2009) are also fully three-dimensional.

Fig. 4 Coverage valuation schemes—coverage may be graded either using a bivalent indicator function, or a real-valued function (without loss of generality, bounded to [0, 1])

2.3.2 Valuation

Real-world sensor planning applications typically have well-defined requirements, and the goal is simply to find any viewpoint which meets these requirements. Accordingly, models such as those of Cowan and Kovesi (1988) are bivalent: either the viewpoint is acceptable or it is not, or equivalently, either a relevance function is covered or it is not. Tarabanis et al. (1995) discuss not only this *admissibility* of a viewpoint, but also its *optimality*, proposing an overall coverage quality metric based on the robustness in individual criteria (Fig. 4).

In solving the camera selection problem, one is interested in finding the *best* view of a relevance function, to which a real-valued coverage metric clearly lends itself. In Park et al. (2006), the quality of coverage of a point $\mathbf{p}$ from a camera C_i is considered to vary inversely with the distance from $\mathbf{p}$ to the center of the viewing frustum of C_i. The authors point out that developing an accurate coverage quality metric is not their focus, and allow that a more sophisticated definition could be substituted. Shen et al. (2007) explicitly set out to define such a metric for the restricted problem case of human surveillance; theirs takes the form of a real-valued function.

Soro and Heinzelman (2007) study several coverage-based valuations of viewpoints for camera selection, but as previously mentioned, their formulation is notably different than others discussed here. Roughly speaking, each valuation can be thought of as a distance metric in $\mathcal{V}$. If one were to assign an ideal viewpoint to every $\mathbf{p} \in \mathcal{S}$, these metrics would effectively constitute a coverage grade of the form $C(\mathbf{p}) : \mathcal{S} \to \mathbb{R}^+$.

By contrast, in solving the optimal camera placement and reconfiguration problems, bivalent coverage valuations are used almost exclusively, to enable the use of various optimization techniques (e.g., binary integer programming) that would otherwise not be applicable. Wang et al. (2009) provide one counterexample, applying a multi-agent genetic algorithm over a graded coverage model simple enough to make the optimization computationally feasible. Continuous grading functions defined by Yao et al. (2008) assign reduced coverage values to the edges (i.e., regions near the limits of field of view and resolution) of a camera's model, in order to encourage their optimization process to yield solutions with a substantial margin of overlap between cameras for improved tracking and handoff. Shen et al. (2007) notably use their coverage grade as a constraint in solving a restricted case of the optimal camera placement problem using a greedy algorithm.

2.3.3 Field of View

The coverage model employed by González-Banos and Latombe (2001) is unique among those surveyed in assuming omnidirectional viewing capabilities, and thus not including a field of view criterion. The directional nature of camera coverage is a recurring key point in the literature, and field of view is the most commonly modeled constraint.

The simple sector-based models of Ma and Liu (2005b), Ma and Liu (2007), Liu et al. (2008), Qian and Qi (2008), Wang et al. (2009), and Jiang et al. (2010) describe field of view with a single angle parameter, which corresponds roughly to the horizontal apex angle. The boundary rays are symmetric about the optical axis, implying an assumption of non-oblique projection. This turns out to be a satisfactory definition in two dimensions; Erdem and Sclaroff (2006) and Hörster and Lienhart (2009) arrive at the same by way of the pinhole camera model, perhaps elucidating how its value should be determined from a given camera system.

Erdem et al. (2003) also describe the three-dimensional field of view using two apex angles. Malik and Bajcsy (2008) and Mittal and Davis (2008) handle field of view similarly.

Cowan and Kovesi (1988) and Tarabanis et al. (1995) both effectively limit field of view to the smaller of the two apex angles and assume non-oblique projection. Piciarelli et al. (2010) model the field of view as a cone, presumably with aperture angle equal to the smaller apex angle. While this representation facilitates their algorithm by projecting to a circle of constant radius on a transformation of the scene plane, it lacks accuracy and no justification is given in the context of their application.

The apex angles are derived from a more elementary characterization of the field of view. In general, a point $\mathbf{p} \in \mathbb{R}^3$ is within the field of view of a camera if its projection lies somewhere on the physical sensor surface. The field of view induced by a rectangular sensor is a pyramid bounded by the rays from each of its four corners through the optical center of the camera. Zhao et al. (2009) use this constraint directly, and can theoretically handle oblique projection. The visible pyramid volume can also be thought of as divided into an infinite set of visible "subplanes" orthogonal to the optical axis; Park et al. (2006) simply assume that the dimensions of the visible subplanes at the near and far depth of field limits are known, and that these subplanes are centered at the optical axis (implying non-oblique projection).

2.3.4 Resolution

The sector-based models proposed by Ma and Liu (2005b, 2007), Liu et al. (2008), Wang et al. (2009), and Jiang et al. (2010) have a radial sensing range limit; although there is no explicit relationship to a resolution constraint, it seems its most likely justification. Cowan and Kovesi (1988) model their resolution constraint as an arc in two dimensions and as a spherical cap in three dimensions.

In fact, this circular/spherical representation unnecessarily complicates the matter: since the projected image is planar and orthogonal to the optical axis, resolution is a function of depth along the optical axis rather than distance along the ray from the optical center (Tarabanis et al. 1994). The triangle-shaped model of Hörster and Lienhart (2009) is a more accurate two-dimensional representation of the resolution constraint, although it is not explicitly parameterized as such. Erdem and Sclaroff (2006), Bodor et al. (2007), Malik and Bajcsy (2008), Yao et al. (2008), and Mittal and Davis (2008) all use distance along the optical axis as the single parameter for the resolution constraint. The last also suggest that such a resolution criterion could be used as a "soft" constraint informing a quality measure.

2.3.5 Focus

While focus is a staple constraint in sensor planning coverage models (Tarabanis et al. 1995), it has not been included in most coverage models developed for other purposes. Angella et al. (2007) mention it, but as with their other imaging criteria, they provide no details. Park et al. (2006) are the other exception; their model is bounded in depth along the optical axis by the near and far depth of field limits.

Park et al. also use focus as part of their coverage grade computation (discussed in Sect. 2.3.2), to some extent: if the center of the viewing frustum is taken as an approximation of the focus distance, the distance of a point along the optical axis from the center varies approximately proportionally to the blur circle diameter. A similar interpretation can be applied to the valuation function of Wang et al. (2009).

2.3.6 View Angle

A constraint on view angle (Fig. 5) is present in some sensor planning coverage models (Tarabanis et al. 1995), such as that of Cowan and Kovesi (1988). In the multi-camera context, it has been included where the target task depends on view angle. For example, the task of the camera network in Zhao et al. (2009) is the identification of planar tags, the performance of which degrades with increasing view angle. Similarly, Shen et al. (2007) are interested in surveillance tasks such as face tracking, so view angle features prominently in their model. Mittal and Davis (2008), drawing on the earlier sensor planning models, include the criterion, anticipating that some tasks will have such requirements.

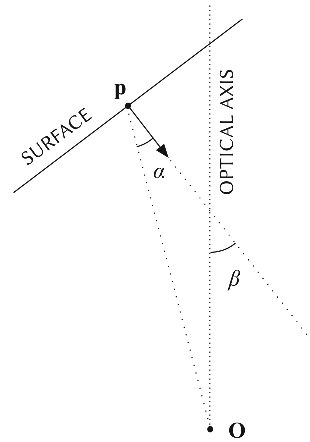


Fig. 5 View angle—the view angle to a point feature on a surface, shown as **p** with a corresponding surface normal, is measured as α in some sources and as β in others

Special cases of task view angle requirements give rise to a few alternate—but equivalent—forms of the view angle criterion. Bodor et al. (2007) are interested in observing paths, where foreshortening effects due to the view angle to a path degrade performance; their view angle criterion is based on both the angle between the path normal and the camera position, and the angle between the path center and the optical axis. Some applications, such as those of Malik and Bajcsy (2008) and Chow et al. (2007), require 360° coverage of a target, and define a maximum view angle for mutual coverage of a point by two cameras. If the view angle to a feature on an opaque surface exceeds 90°, the surface occludes the feature from view; this phenomenon is known as *self-occlusion* and is sometimes treated as a separate criterion, such as by Chen and Li (2004) and Zhao et al. (2009).

An interesting question that arises in defining this criterion is whether to measure the view angle between the feature surface normal and the optical axis, or between the feature surface normal and the line-of-sight ray from the camera's optical center. Both approaches have merit in terms of validity with respect to task requirements. The former is taken by Chen and Li (2004) and Bodor et al. (2007); the latter, by Cowan and Kovesi (1988), Shen et al. (2007), Malik and Bajcsy (2008), and Zhao et al. (2009).

Soro and Heinzelman (2007), in one of their models, grade views primarily based on view angle.

2.3.7 Static Occlusion

Occlusion by static scene objects factors heavily in most multi-camera coverage work (Fig. 6). Malik and Bajcsy (2008), whose model does not include a static occlusion criterion, assume a simple rectangular room with nonzero

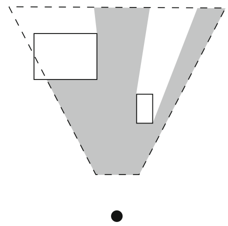


Fig. 6 Static occlusion—the two *white boxes* represent the static scene model. Coverage without the constraint is outlined, with actual coverage in gray

relevance somewhere near its center, which suits their target task, but in most multi-camera applications the scene is assumed to be more complex. The "high wall" occlusion model common in two-dimensional approaches has its origin in the art gallery problem, exemplified by González-Banos and Latombe (2001). This constraint is enforced as follows: given a scene model consisting of line segments in the plane, a point $\mathbf{p} \in \mathbb{R}^2$ is occluded (not covered) if the line of sight from the camera's optical center to $\mathbf{p}$ intersects any such line segment. Erdem and Sclaroff (2006) propose an algorithm to construct a continuous "visibility polygon" set which contains all non-occluded scene points. Hörster and Lienhart (2009), Mittal and Davis (2008), and Shen et al. (2007) simply check for line-of-sight on each discrete relevance point. Jiang et al. (2010) approximate static occlusion by simply excluding obstacle regions from the field of view of a camera; in confined spaces and using cameras with realistic field of view, this would likely result in poor performance.

The three-dimensional analog to the line segment scene model is composed of opaque surfaces. A continuous, analytic solution has been employed in sensor planning (Tarabanis et al. 1996) and next best view (Maver and Bajcsy 1993) applications. In the multi-camera context, discrete line-of-sight checking is more common, as is done by Angella et al. (2007) and Zhao et al. (2009).

Piciarelli et al. (2010) handle static occlusion directly in the relevance function. Each camera node has its own copy of the global relevance function, with all points occluded (via two-dimensional line of sight) from that camera removed from the model.

2.3.8 Dynamic Occlusion

Mittal and Davis (2008) have pioneered handling dynamic occlusion in a geometric coverage model. They use a probabilistic model of agent occupancy and some assumptions about agent height and allowable camera viewpoints to formulate a probabilistic visibility criterion, which is then integrated with their other (static) constraints. Angella et al. (2007) use this model. Chen and Davis (2008) independently propose their own probabilistic metric for dynamic occlusion, under similar assumptions about the agents and cameras. Qian and Qi (2008) also propose a probabilistic model, with targets modeled as 2D discs (analogous to Mittal and Davis' representation) and using a simple sector-type coverage model.

Zhao et al. (2009) include a "mutual occlusion" criterion in their model, which approximates worst-case dynamic occlusion by specifying a range of view angles within which a point is assumed to be occluded by another agent.

2.3.9 Combining Criteria and Multi-Camera Coverage

Cowan and Kovesi (1988) treat coverage criteria as constraints on the viewpoint, so in order to find the solution set which satisfies all constraints (i.e., the set of viewpoints which adequately cover the relevance function), it suffices to intersect the solution set for each individual criterion. Bivalent coverage models have taken much the same approach, intersecting the sets of covered points generated by each criterion, exemplified by the "feasible region" result of Erdem and Sclaroff (2006). In the multi-camera context, the overall coverage of the scene is of interest; this is usually found by taking the union of the coverage sets for each individual camera, as Erdem and Sclaroff also show.

Mittal and Davis (2008) integrate their probabilistic dynamic occlusion metric with their other "static" constraints to obtain an overall (graded) quality metric for each point and orientation.

Several models in the literature also provide mechanisms to compute overall k-coverage of a scene. Erdem and Sclaroff (2006) show a similar approach in their experimental figures, but none of their experimental problem statements require multi-camera coverage. Liu et al. (2008) also use an intersection-union approach in their work, which focuses specifically on k-coverage.

Mittal and Davis (2008) discuss more complex "algorithmic constraints" involving the interplay of various constraints between multiple cameras, for such tasks as stereo matching. To some extent, particularly on the view angle criterion, this is realized in the k-coverage model of Shen et al. (2007).

2.3.10 Task Parameters

A recurring motif in the literature is that the quantification of visual coverage depends as much on the task as it does on the parameters of the imaging system. Generally, given a computer vision algorithm used in a task, it is at least theoretically possible to quantify soft or hard requirements on imaging properties such as resolution, focus, and view angle. The actual values of the imaging constraints in sensor planning models, such as those of Cowan and Kovesi (1988) and Tarabanis et al. (1995), are assumed to be direct task requirements. Erdem and Sclaroff (2006) emphasize the task-specific nature of the constraints in the optimal camera placement context.

One form of the optimal camera placement problem constrains the minimum required proportion of the relevance function covered by the solution (while maximizing or minimizing some other variable, such as cost), a task-specific requirement. This is one of the four variations studied by Hörster and Lienhart (2009). The weighted form of this proportion, sometimes called the *coverage rate* (Jiang et al. 2010), may fill a similar role, as in the optimal placement problem studied by Shen et al. (2007).

2.3.11 Relevance Function

A relevance function is most commonly used in optimal camera placement and camera reconfiguration applications, where it comprises the coverage objective. Often, the nonzero relevance function is implicitly the working volume (as in the art gallery problem); in order to support a general problem definition, however, a model should separate the coverage target from other considerations such as the scene model for static occlusion and the allowable viewpoint set. Jiang et al. (2010), Hörster and Lienhart (2009), Angella et al. (2007), Zhao et al. (2009), and Malik and Bajcsy (2008) all allow specification of a relevance function, in some form, that is distinct from the scene model and/or allowable viewpoint positions.

It is also useful to allow prioritization of coverage in the relevance function. One of the experiments of Erdem and Sclaroff (2006) specifies a higher resolution requirement on certain parts of the floor plan. Hörster and Lienhart (2009) use a continuous weighted relevance function in their problem instance definition; in the actual discrete domain of their algorithm, this informs the sampling frequency of control points. Jiang et al. (2010) retain a similar continuous definition, using distinct regions with integer weights to simplify the weighted coverage computation. Piciarelli et al. (2010) define a relevance function as a mapping of discrete points to real values.

2.3.12 Allowable Viewpoint Set

Generally, explicit restrictions on viewpoints in sensor planning and optimal camera placement applications are on the position component of the viewpoint only. There are usually no restrictions on orientation; a notable exception is the work of Chen and Li (2004), where both position and orientation are constrained by kinematic reachability by the robot on which the camera is mounted. Restrictions or specifications on other aspects of the viewpoint, such as the intrinsic parameters of the camera, are usually implicit in the problem instance.

Sensor planning coverage models translate coverage criteria into constraints on the solution set of viewpoints, so the allowable viewpoint set is just a directly-defined constraint to be intersected with the rest. The "prohibited regions" constraint employed by Cowan and Kovesi (1988) is an example.

In the art gallery problem, the volume of relevance to be covered and the allowable $\mathbb{R}^n$ positions of the guards (cameras) are implicitly the same, as exemplified by González-Banos and Latombe (2001). Erdem and Sclaroff (2006) and Wang et al. (2009) also use the relevance volume as the allowable viewpoint position set. Jiang et al. (2010) specify a relevance function, but place no restriction on camera position. Malik and Bajcsy (2008) constrain camera positions to a rectangular volume which is implied to be a superset of the relevance volume. Shen et al. (2007) restrict viewpoints to the outer boundary of a rectangular relevance volume. Hörster and Lienhart (2009) specify the relevance function and the allowable viewpoint set separately, as subsets of a larger working volume.

Camera reconfiguration applications typically place tighter restrictions on allowable viewpoints. Bodor et al. (2007) allow full online control of position and orientation, as their cameras are mounted on mobile robots. Piciarelli et al. (2010) allow online control of orientation (pan and tilt) as well as some intrinsic parameters (zoom), but constrain the cameras to fixed positions. Chen et al. (2010) allow horizontal rotation (pan) only.

2.4 State of the Art and Open Problems

To date, no geometric model has fully captured the phenomenon of visual coverage in a representation suitable for the general multi-camera context. While some of the single-camera sensor planning models we have discussed are quite accurate and general enough to apply to a wide set of tasks, they are ill-suited to modeling typical systems and environments involving multiple cameras, and in their present form would likely put prohibitive computational requirements on optimizations involving even relatively small networks. Conversely, in expressly designing multi-camera models in forms suitable for specific optimization techniques, the remainder of the authors mentioned have restricted applicability to relatively specific problem classes. Mittal and Davis (2008) appear to have designed the most accurate and general model to date which is still suitable for multi-camera optimization, but it is still somewhat restricted by certain assumptions, notably its two-dimensional final representation, and its lack of a focus criterion.

The ideal geometric coverage model would not only accurately model visual coverage in a form convenient for multi-camera systems and their environments, with as few assumptions as possible and allowing for generalized task requirements, but also provide this information in a form accomodating powerful optimization techniques. It is clear from the preceding discussion that the factors involved in a model achieving the former goal would be highly complex, complicating success in the latter goal. The prevailing approach to this problem has been to design the model to be as accurate and general as possible for one specific optimization technique from the outset, but this has failed to produce the ideal model. We suggest that attempting to achieve the first goal in isolation could, at the very least, produce a tool for evaluation, but may also yield new insights into the nature of multi-camera coverage that may lend the model, or some derivative thereof, to an appropriate optimization scheme.

Most sources surveyed have assumed that the coverage model employed reflects a posteriori task performance, with little or no validation of the model itself. In order to evaluate accuracy and generality, a generic scheme for relating the coverage metric to a task performance metric should be developed and adopted. A simple statistical measure, such as the Pearson product-moment correlation coefficient, might suffice; depending on the nature of the coverage and performance metrics, other measures might be more illuminating.

3 Coverage Overlap Topology

3.1 Anatomy of a Coverage Overlap Model

A *coverage overlap model* describes the topology of a camera network in terms of coverage overlap (mutual coverage of some part of the scene). Typically, the camera node is the atomic entity, and of interest are the node-level coverage overlap relationships. It is often desirable to capture not only the fact but the degree of overlap.

In the most general form, such a model is a weighted undirected hypergraph $\mathcal{H} = (C, E, w)$, where C is a set of *coverage cells*, $E \subseteq \mathcal{P}(C)$ (where $\mathcal{P}$ denotes the power set) is a set of hyperedges, and $w : E \to \mathbb{R}^+$ is a weight function over E. A coverage cell may represent an individual camera node's coverage model or some portion thereof. The existence of a hyperedge $e \in E$ indicates that the nodes in e share mutual coverage of the scene, with a k-hyperedge corresponding to k-coverage. In a weighted model, $w(e)$ quantifies the degree of shared coverage. In an unweighted model, implicitly, $w(e) = 1$ if $e \in E$ and $w(e) = 0$ otherwise; the existence of an edge indicates sufficient mutual coverage for the given task, by some task-specific quantitative definition.

The most common form is the *vision graph* (Fig. 7), which is an ordinary graph (a 2-uniform hypergraph) and thus considers only pairwise coverage overlap.

3.2 Coverage Overlap Models by Application

The earliest examples of coverage overlap models are found in *multi-view registration* applications, including video sequence registration and 3D range image registration. Since

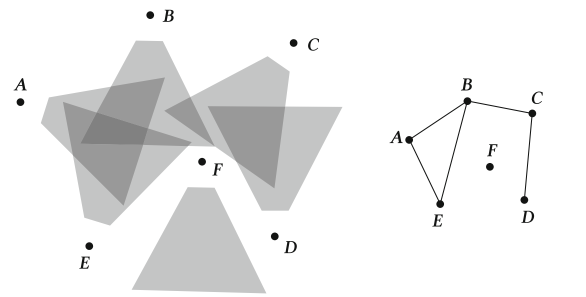


Fig. 7 Vision graph—from 2D coverage geometry (*left*) to overlap topology (*right*). Note that the pairwise overlaps of A, B, and E are represented, but their 3-overlap is not

the objective is to align visual data from multiple views, it is clearly useful to know which views overlap and thus might have some corresponding features for registration. Although this is not necessarily a camera network application, multiple views are theoretically equivalent to multiple cameras. Sawhney et al. (1998) propose a graph formalism of the coverage overlap relationships between multiple views for video sequence registration, with each frame (view) represented by a vertex, addressing the fact that frames (views) which are not temporally adjacent may still be adjacent in terms of overlap topology. Kang et al. (2000) construct a similar graph representation of the overlap topology of frames, in which edges indicate either temporal or spatial (overlap) adjacency. Their algorithm searches for an optimal path in this graph to minimize error in global registration. Huber (2001) constructs a graph for registration of partial 3D views using an overlap criterion on the range images, analyzes the registration problem through its connectivity properties, and performs reconstruction over a spanning tree. Sharp et al. (2004) also study 3D range image registration using a similar graph formalism, which they assume exists a priori. They approach the global registration problem by first considering registration over basis cycles within the graph, then merging the results using an averaging technique.

Knowledge of camera network topology in terms of coverage overlap is a useful precursor to full metric multi-camera *calibration*. Antone et al. (2002) require, as input to their calibration algorithm, a graph of node adjacency; although the criterion for edge presence is based on position (from GPS), since the algorithm targets omni-directional cameras, this is supposed to approximate coverage overlap and is thus a vision graph. Brand et al. (2004) further develop this work, using directionally-constrained graph embeddings. Devarajan and Radke (2004) name and explicitly describe the vision graph, pointing out its distinctiveness from the communication graph (a departure from sensor networks), and demonstrating its usefulness in informing a full calibration algorithm as to which camera pairs should attempt to find a homography. However, they offer no means of obtaining the vision graph automatically, instead

Table 2 Comparison of selected topological coverage overlap models

Model	Appl.	Properties				Construction data				
		Struct.[2]	Weight	*k*-View	Part.	Geom.	Reg.	Feat.	Occup.	Motion
Sawhney et al. (1998)	R	G					•			
Huber (2001)	R	G					•			
Sharp et al. (2004)	R	G	✓				•			
Cheng et al. (2007)	C	G						•		
Kurillo et al. (2008)	C	G	✓					•		
Bajramovic et al. (2009)	C	G						•		
Mavrinac et al. (2010)	C	G	✓				•			
Stauffer and Tieu (2003)	DT	G								•
Mandel et al. (2007)	DT	G			✓					•
Van Den Hengel et al. (2006)	DT	G			✓				•	
Lobaton et al. (2010)	DT	SC		✓	✓					•
Kulkarni et al. (2007)	S	HG	✓	✓					•	
Mavrinac and Chen (2011)	S	HG	✓	✓		•				

R multi-view registration, *C* calibration, *DT* direct tracking correspondence, *S* scheduling, *G* graph, *HG* - hypergraph, *SC* simplicial complex

making the temporary assumption that it is available a priori. Cheng et al. (2007) address this issue by approximating the vision graph via pairwise feature matching, and describe a full calibration algorithm also employing the feature data following the procedure of Devarajan and Radke. Kurillo et al. (2008) construct a weighted vision graph based on the number of shared calibration points, then optimize the set of calibration pairs by finding a shortest path spanning tree. Bajramovic et al. (2009) perform multi-camera calibration over connected components of their vision graph, which they construct independently using the normalized joint entropy of point correspondence probability, one of several methods described by Brückner et al. (2009). Mavrinac et al. (2010) describe the vision graph as a theoretical upper bound for the connectivity of their grouping and calibration graphs.

Overlap topology can be used to help establish *direct tracking correspondence*, a subproblem of tracking correspondence involving agents simultaneously visible in multiple cameras. This is useful for *camera handoff* among overlapping cameras (Javed et al. 2000; Khan and Shah 2003). In this context, overlap topology is usually considered to be a subset of a more general transition topology, models for which are covered in Sect. 4. Stauffer and Tieu (2003) describe a "camera graph" which identifies with the vision graph, estimating camera overlap from sets of likely correspondences between tracks. This graph is then used as feedback to improve tracking correspondence. Mandel et al. (2007) use a probabilistic approach on motion correspondence to establish overlap topology for tracking purposes. In a series of papers on the topic, Van Den Hengel et al. (2006, 2007), Detmold et al. (2007, 2008), Hill et al. (2008) describe the *exclusion* approach, whereby the vision graph begins complete and edges are removed based on contradictory occupancy observations, with a target application of tracking correspondence in surveillance networks. Lobaton et al. (2009a,b, 2010) propose a simplicial complex representation of overlap topology dubbed the *CN-complex*, primarily targeted at tracking applications. Overlap topology is employed by Song et al. (2010) as part of their consensus approach to tracking and activity recognition.

Camera networks are often composed of devices with limited computational and energy resources. Knowledge of overlap topology can help inform efficient *scheduling* of node activity. Ma and Liu (2005a) estimate the correlation between views using their previously described geometric coverage model, to improve the efficiency of video processing in camera networks with partially redundant views. However, the information used is not strictly topological, and the method applies specifically to two-camera systems. Dai and Akyildiz (2009) address the latter issue by extending the correlation problem to multiple cameras, but their model is also not strictly topological. Kulkarni et al. (2007) construct a vision graph using a Monte Carlo feature matching technique with a geometric model component, and demonstrate its use in duty cycling and triggered wake-up. Mavrinac and Chen (2011) propose a *coverage hypergraph* derived directly from their geometric coverage model, and apply it to the optimization of load distribution using a parallel machine scheduling algorithm.

3.3 Analysis and Comparison of Overlap Models

Table 2 compares the nature and properties of a selection of topological coverage overlap models from the literature, grouped by application (indicated in the first column). The second column identifies the combinatorial structure used (whether explicit or interpreted), and the following three columns indicate which additional properties are exhibited: edge weighting, k-view modeling, and modeling of partial views, respectively. The remaining five columns specify which type of data is used in constructing the model: geometric coverage information, registration results, local feature matching, occupancy correlation, or motion correlation.

3.3.1 Combinatorial Structure

Although not all of the coverage overlap models surveyed are explicitly formalized as graphs (or hypergraphs), they can be cast as cases of the general model described in Sect. 3.1 without loss of information. The original descriptions given by the authors are summarized here, and instances where ancillary information not captured by the graph representation is present are highlighted.

The vision graph as described by Devarajan and Radke (2004)—an undirected, unweighted graph with vertices representing camera nodes and edges indicating sufficient coverage overlap for the purposes of the task—is the simplest and most common combinatorial structure for models of coverage overlap topology seen in the literature. This is the explicit form of the models of Cheng et al. (2007), Bajramovic et al. (2009), Mavrinac et al. (2010), and Stauffer and Tieu (2003). The graphs of Sawhney et al. (1998) and Kang et al. (2000) describe temporal and spatial adjacency, but since in their application temporally adjacent frames are assumed to be spatially adjacent also, they are effectively describing the vision graph structure. The graphs described by Huber (2001) and Sharp et al. (2004) are also essentially vision graphs; though edges are annotated with pairwise relative pose and other relations, this information is not part of the overlap model proper. Mandel et al. (2007) and Van Den Hengel et al. (2006, 2007) do not explicitly present graph formalisms, but maintain sets of hypotheses about coverage overlap which correspond to edges in the vision graph.

Some recent models extend the captured topology from pairwise overlap to k-overlap, requiring a hypergraph or hypergraph-like structure to accomodate the relationships. Lobaton et al. (2009a,b, 2010) partially achieve this with a simplicial complex representation. This choice of representation, over a more abstract structure such as a hypergraph, seems to stem from the focus being more on geometrical properties and operations and less on combinatorial optimization. They are interested in overlap topology only up to 2-simplices (or 3-simplices in a hypothetical extension to three dimensions), so their model does not capture general k-overlap. Kulkarni et al. (2007) model the full k-overlap topology of the camera network, although they do not explicitly formalize this model in a hypergraph representation or use any combinatorial techniques. Mavrinac and Chen (2011)

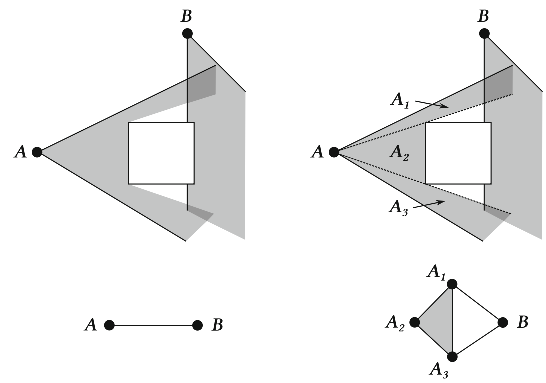


Fig. 8 Vertex granularity in the CN-complex (Lobaton et al. 2010)—the simplicial complex on the *right* more accurately describes the coverage overlap topology between cameras A and B

present an explicit hypergraph representation of k-overlap topology, with an initial scheduling application using a combinatorial algorithm.

The assingment of one vertex to each camera node is sensible for most combinatorial optimization purposes, but a few models eschew this paradigm and subdivide vertex assignment into coverage cells. Motivations for doing so vary. Van Den Hengel et al. (2006, 2007) subdivide views into an arbitrary number of windows to handle partial coverage overlap of cameras (due to the specifics of their construction method, discussed in Sect. 3.3.2). Mandel et al. (2007) divide views into regions for a similar reason. In both cases, it appears that the model of interest to the eventual application recombines the coverage cells (which are each associated with a specific camera) to the more usual granularity of one vertex per camera. Lobaton et al. (2009a,b, 2010) divide the two-dimensional geometric coverage of each camera at rays to occlusion events (which they call "bisecting lines"), allowing their model to accurately capture some geometric properties of overlap, such as static occlusions within the field of view, as shown in Fig. 8. This increased granularity is preserved, and shown to be beneficial in tracking applications.

Calibration and scheduling applications of overlap models often make use of graph optimizations related to path length. In such cases, weighting vision graph edges proportionally to the degree of coverage overlap can yield better results. Given an edge $e_{AB} = \{A, B\}$ linking cameras A and B, Kurillo et al. (2008) assign the weight $w(e_{AB}) = \frac{1}{N_{AB}}$, where N_{AB} is the number of common reference points detected by cameras A and B (see Sect. 3.3.2 for more on their graph construction method). Kulkarni et al. (2007) similarly compute the degree of k-overlap from the number of common reference points of k cameras. Both describe methods of handling non-uniform spatial distributions of reference points. Mavrinac and Chen (2011) theoretically use the volume of intersection between k cameras' geometric coverage models to weight hyperedges, but in practice, the required polytope intersection procedure is NP-hard, so they use a uniform distribution of points to compute a discrete approximation.

3.3.2 Construction from Visual Data

In theory, overlap topology is a derived property of the geometric coverage of the camera network. However, since it is often employed in applications where geometric coverage information is unavailable, especially in calibration initialization, it is often necessary to estimate it using visual data. Finding correspondences between visual data in some form among views is the obvious means—if camera A matches a piece of its own information to one from camera B, then a hypothesis of mutual coverage between A and B can be made or strengthened.

For tasks which already make use of correspondences between local image features, the same information, or some subset thereof, can be used to recover overlap topology. This is the approach originally suggested by Devarajan and Radke (2004). Because their algorithm works in an offline centralized context, Kang et al. (2000) are able to directly correlate image features to infer topology. Cheng et al. (2007), addressing the camera network calibration problem, attempt to make such an approach scalable in an online distributed context by instead sharing *feature digests* of SIFT (Lowe 2004) descriptors among camera nodes. Bajramovic et al. (2009) and Brückner et al. (2009) use the pairwise joint entropy of point correspondence probability distributions, based on SIFT feature descriptors, as a measure of overlap. Kurillo et al. (2008) use direct matching of a more sparse but more accurate set of features, obtained from a structured calibration target. A similar approach is taken by Kulkarni et al. (2007), although the structured target is only used for topology inference and is unrelated to their application. In their case, both the degree and geometry of coverage overlap are estimated using a Monte Carlo technique, whereby the target is imaged at random reference points, and the k-covered Voronoi cells around each point contribute to the estimate for each of the k cameras covering it.

Registration-based applications are typically iterative, and some overlap models are updated using new correspondence data available in each iteration. Sawhney et al. (1998) infer global overlap topology iteratively, using feedback from a local coarse registration stage to recover graph edges, and subsequently performing local fine registration on adjacent views. An analogous three-dimensional process is employed by Mavrinac et al. (2010) in a distributed calibration algorithm, with coarse registration results iteratively building a grouping graph which then informs pairwise fine registration. Huber (2001) also uses candidate registration matches to iteratively infer overlap topology.

Camera networks often have wide baselines and large rotational motion between cameras, over which local feature detectors generally have poor repeatability and matching performance (Mikolajczyk et al. 2006; Moreels and Perona 2007). Fortunately, they offer the possibility of matching online motion data instead of static features, which can be more robust under some circumstances. Stauffer and Tieu (2003) argue that the descriptiveness, spatial sparsity, temporal continuity, and linear increase in volume over time of tracking correspondences make them more reliable in matching than static features. They correlate local tracks between cameras over time, and infer a vision graph edge where the expectation of a match exceeds a threshold. Mandel et al. (2007) take a slightly different approach, detecting local motion and attempting to correlate it with motion observed in other cameras, via a distributed algorithm. Lobaton et al. (2010) automatically decompose cameras into coverage cells by locally finding “bisecting lines” at which occlusion events occur (e.g., walls), then, with a distributed algorithm, globally estimate cell overlap by matching concurrent occlusion events over time.

Van Den Hengel et al. (2006, 2007) take the reverse approach to those described thus far. Their *exclusion* algorithm begins by assuming all camera nodes have overlapping coverage, thus a complete vision graph, and eliminates edges over time using occupancy data to rule out coverage overlap. This method does not require any correspondence between observations; if camera A is occupied (currently observing an object) and camera B is unoccupied, this is evidence that A and B do not have mutual coverage, which through observation ratio calculations and thresholding contributes to the final model. Partial overlaps are handled by dividing camera coverage into an arbitrary number of coverage cells. Hill et al. (2008) describe a number of potential shortcomings in real-world operation, along with ways of mitigating the adverse effects on performance. Detmold et al. (2007, 2008) extend the approach into an online distributed context for scalability and dynamic updating of the model.

One direct route to an overlap model well-suited to the task at hand is to use the very visual data used by the task itself to estimate the model, if this data (or similar data) is available. This can clearly be seen in most cases of registration, feature-based calibration, and tracking applications in Table 2. In a distributed camera network, depending on the nature of the data and the amount of it required to establish accurate overlap estimates, there is a potential scalability issue since, initially, the data must effectively be broadcast to all other nodes. As mentioned in the preceding section, Cheng et al. (2007) address this using digests of the SIFT features to establish overlap topology, then share the substantially larger full feature data pairwise only among cameras with sufficient overlap for calibration. Kurillo et al. (2008) also use calibration feature points to estimate overlap; scalability is less of an issue because they use a structured calibration target, which yields a set of features both sparse enough to distribute among many cameras and robust enough to achieve accurate metric calibration. In the algorithm of Stauffer and Tieu (2003), overlap topology estimation is part of the closed-loop tracking correspondence task itself. The scheduling applications of Kulkarni et al. (2007) and Mavrinac and Chen (2011) use occupancy correlation and geometric coverage, respectively, in an attempt to obtain the same fundamental information, viz. the degree of content pertinent to the task in each k-view.

3.4 State of the Art and Open Problems

The vision graph is a well-established concept and theoretical tool in the multi-camera network literature. In the application classes of multi-view registration and calibration, which (in the surveyed cases) involve pairwise coverage relationships exclusively, it has proven useful in its basic form. Additional optimizations are possible with appropriate use of edge weights and related combinatorial techniques, as demonstrated by Sharp et al. (2004) and Kurillo et al. (2008) for the respective applications.

When used in direct tracking correspondence, the limitations become apparent. Arbitrary subdivisions of camera nodes into partial coverage cells appears to improve performance, but this is unsatisfying from a theoretical standpoint. Lobaton et al. (2009a, 2010) present an explicit departure from the graph model, allowing them to represent 2-coverage and 3-coverage in a simplicial complex; however, presumably since their application does not require it, general k-coverage modeling is absent. Kulkarni et al. (2007) and Mavrinac and Chen (2011) use more general hypergraph (or equivalent) models explicitly designed for general k-coverage, suitable for scheduling in distributed camera networks, but ignore the coverage subdivisions needed by tracking applications.

We believe that the generalized hypergraph model presented in Sect. 3.1 includes all of the information necessary to fit the needs of each of the applications covered here, and being a relatively straightforward combination of existing concepts from the literature, should be backwards-compatible with all of the reviewed sources. In the absence of task-specific geometric coverage information, it is sensible to use the task data itself to approximate the model. It remains an open question whether the nature of the information contained in edge weights, and the additional combinatorial optimizations they make possible, can be incorporated into such a unified framework.

4 Transition Topology

4.1 Anatomy of a Transition Model

A *transition model* describes the topology of a camera network in terms of the probability and/or timing of moving agents transitioning from one region of coverage to another. While an overlap model of the sort covered in Sect. 3 captures a physical topology, a transition model captures a more abstract functional topology of agent activity. Relationships may exist among camera nodes with no mutual scene coverage (non-overlapping cameras). Since the target application class is agent tracking, the granularity of the topology may extend down to subsets of camera nodes' coverage: entry and exit points and regions of overlap are often considered individually (Fig. 9).

In the most general form, such a model is a weighted directed graph $\mathcal{G} = (C, A, w)$, where C is a set of *coverage cells*, A is a set of arcs, and $w : A \to \mathbb{R}^+$ is a weight function over A. A coverage cell may represent an individual camera node's coverage model or some portion thereof, such as an entry or exit zone (note that a coverage cell may be both an entry zone and an exit zone). C may also include a special source/sink node to collectively represent the uncovered portion(s) of the scene. The existence of an arc $a \in A$ indicates that agents may transition from the tail region to the head region. In a weighted model, $w(a)$ is a quantitative metric encapsulating the probability and/or duration of the transition.

4.2 Transition Models by Application

Transition models are largely aimed at one particular application class: *predictive tracking* in (generally) non-overlapping camera networks. For a locally tracked agent leaving one coverage cell, the objective is to predict in which other coverage cell(s) the agent will reappear, possibly to inform *camera handoff*. A special case occurs when the cameras have coverage overlap, which is addressed by several models of overlap topology as the direct tracking correspondence problem (see Sect. 3.2). Javed et al. (2003) show that, in the context of non-overlapping tracking correspondence, transition probabilities and durations are dependent on individual correlations of entry and exit zones, of which each camera may have a number. Their geometric counterparts are coverage cells, and in a combinatorial transition model, they comprise the vertex set. Various techniques have been applied to this type of model to aid in tracking agents across non-overlapping views (i.e., through unobserved regions).

The model presented by Ellis et al. (2003) exemplifies this approach. Their method automatically identifies entry and exit zones in each camera ((a problem previously addressed by Stauffer 2003), then finds the transition topology by

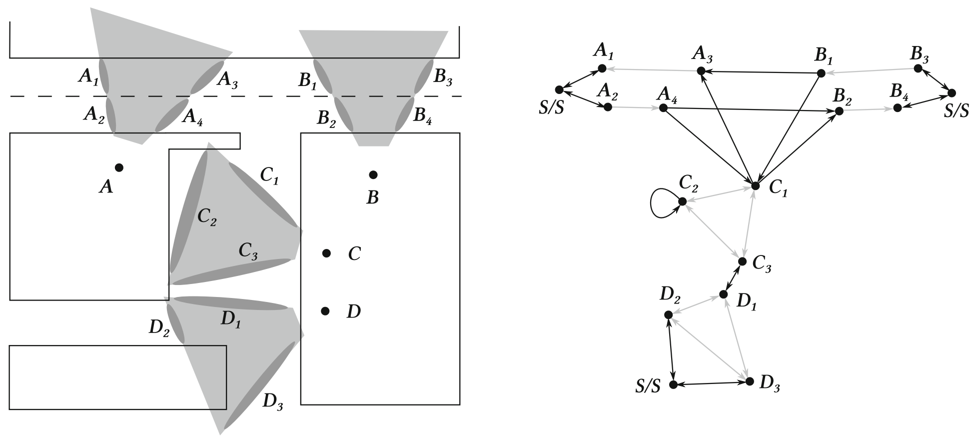


Fig. 9 Transition graph—from 2D coverage geometry (*left*) to transition topology (*right*). *Dark ellipses* denote entry and exit zones. Intra-camera transition arcs are shown in gray. The S/S vertices represent external agent source/sink

temporally correlating a large number of local trajectories between cameras, requiring no actual tracking correspondence. Makris et al. (2004) extend this method and further develop its theoretical basis. Stauffer (2005) operates on a closely related model, but presuming the availability of a coverage overlap model—Stauffer cites his own previous work with Stauffer and Tieu (2003)—treats cliques of overlapping cameras (connected components in the vision graph) as the larger coverage structure containing entry and exit zones, on the premise that the overlapping case is better handled by robust direct correspondences. The aforementioned methods ascribe to observations an implicit correspondence, and assume a unimodal statistical distribution of transitions. Tieu et al. (2005) address this with a method capable of handling multimodal distributions.

Marinakis et al. (2005); Marinakis and Dudek (2005) consider cameras with full coverage of widely-separated sections of hallways in a building, so that transitions are constrained to the hallway topology. Due to these constraints, the entry and exit zone coverage cells (transition graph vertices) are the cameras themselves, and the cameras need only be capable of detecting an agent's presence with reasonable fidelity for their method to successfully estimate the topology. Niu and Grimson (2006) target a vehicle tracking application, using appearance to match observations between, and infer the topology of, non-overlapping cameras.

Dick and Brooks (2004) approach the predictive tracking problem with a Markov model which captures transition topology after a training phase, albeit not in an explicitly combinatorial form, dividing the view into blocks over which the topology is found. The method of Gilbert and Bowden (2006) incrementally learns the topology between recursively subdivided blocks of the views; their method does not require a training phase and can adapt to changes in the camera network. Both yield a probabilistic topological model which can be used in conjunction with appearance-based matching to track across disjoint views.

Zou et al. (2007) are interested in tracking humans, and integrate appearance-based agent correspondence based on face recognition into the inference method of Ellis, Makris, and Black, for improved robustness in their target instance. Nam et al. (2007) also specifically track humans, and an appropriate appearance model is integral to their estimation method. The method of Farrell and Davis (2008) falls within this category as well, and is notable for its expressly distributed approach, which affords scalability to large, distributed surveillance networks.

Finally, it should be noted that the coverage overlap model developed by Van Den Hengel et al. (2006, 2007), Detmold et al. (2007, 2008), Hill et al. (2008) can be extended, as the authors explain, to capture non-overlapping transition topology by adding a temporal padding window to the exclusion method.

4.3 Analysis and Comparison of Transition Models

Table 3 compares the properties of a selection of topological transition models from the literature. Intepreting each model as a graph, the first and second columns indicate whether the graph is directed and weighted, respectively. The third column indicates whether vertices of the graph represent individual entry/exit points, of which each camera may have several; the implication otherwise is that the granularity is at the level of cameras only. The fourth column indicates whether the model includes an explicit source/sink vertex, for agents entering or leaving the scene. The fifth column indicates whether the graph models transitions between overlapping cameras, thus implicitly modeling coverage overlap to the extent described with direct tracking correspondence applications in Sect. 3. The final two columns specify which type of data is used in constructing the model: statistical correlation between temporal events, or correlation via an appearance model.

Table 3 Comparison of selected topological transition models

Model	Properties					Construction data	
	Directed	Weight	Entry/exit	Source/sink	Overlap	Temporal	Appearance
Ellis et al. (2003)	✓		✓		✓	•	
Makris et al. (2004)	✓	✓	✓	✓	✓	•	
Dick and Brooks (2004)	✓	✓	✓		✓		•
Marinakis et al. (2005)	✓			✓		•	
Stauffer (2005)	✓	✓	✓			•	
Tieu et al. (2005)	✓	✓	✓			•	
Niu and Grimson (2006)	✓						•
Nam et al. (2007)		✓	✓	✓	✓		•
Zou et al. (2007)	✓	✓	✓		✓	•	•
Farrell and Davis (2008)		✓					•

4.3.1 Combinatorial Structure

Relatively few of the transition models surveyed are explicitly presented as graphs resembling the generalized model described in Sect. 4.1. Marinakis et al. (2005); Marinakis and Dudek (2005) model the topology in a directed, unweighted graph, in which vertices represent camera nodes and arcs represent possible transitions. Transition probabilities and durations are captured separately in an agent model. The graph of Nam et al. (2007) also has a vertex for each camera node, but also has intermediate vertices representing either an overlapping or non-overlapping transition point and a source/sink vertex; since individual entry and exit zones are not represented, the graph is undirected. Zou et al. (2007) use essentially the same model as Ellis et al. (2003); Makris et al. (2004), but treat it explicitly as a weighted, directed graph, with vertices representing entry and exit zones and arcs indicating possible transitions. Trivially, related models, such as those of Stauffer (2005) and Tieu et al. (2005), could be treated similarly. The transition matrix of Dick and Brooks (2004) can be interpreted as an incidence matrix for the transition graph. In general, it is not difficult to apply a graph interpretation to any of the models surveyed here.

As discussed in Sect. 3.3.1, coverage overlap models typically represent each camera node as a vertex, a structure which offers useful combinatorial properties in most applications. Some transition models employ this structure as well. Marinakis et al. (2005); Marinakis and Dudek (2005) assume widely separated cameras and wish to avoid dealing with complex local tracking, so this is the sensible representation for their case. Niu and Grimson (2006) and Farrell and Davis (2008) also consider transitions only between strictly non-overlapping cameras. In scenes of even moderate complexity, however, a transition topology among individual entry and exit points is more germane to predictive tracking. This structure is described by Ellis et al. (2003); Makris et al. (2004) and used by a plurality of the models surveyed (Stauffer 2005; Tieu et al. 2005; Zou et al. 2007; Nam et al. 2007). Dick and Brooks (2004) do not automatically determine entry and exit points, but do divide the cameras into coverage cells, which would induce the vertices in a graph interpretation of their model.

Makris et al. (2004) include a source/sink vertex (which they call a "virtual node"), in addition to the entry and exit zone vertices, to handle the probabilistic paths of agents entering or leaving the overall coverage of the camera network. Marinakis and Dudek (2005) and Nam et al. (2007) also include such a vertex in their models.

Among explicit graph models with arc weights, the definition of the weighting function varies. Makris et al. (2004) annotate arcs in the graphical representation of their model according to the probability of transition, computed from the cross-correlation of the temporal sequences of departure and arrival events at each entry and exit zone (vertex), but do not operate on it as a weighted graph. Zou et al. (2007) explicitly apply this weighting to the graph representation, with a modified correlation function based on both identity and appearance (as opposed to identity only). In contrast, Nam et al. (2007) weight arcs based on the mean duration of transitions between cameras.

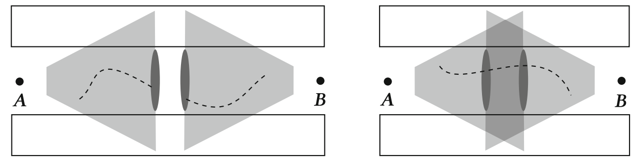
4.3.2 Construction from Visual Data

It is normally assumed that the camera network is uncalibrated and that information about the scene and agent dynamics is unavailable a priori. For the purposes of this discussion, we will approach the construction methods assuming that entry and exit zones are known, either estimated separately (Stauffer 2003; Ellis et al. 2003; Gilbert and Bowden 2006) or specified a priori, as in the case where each camera is a single entry/exit zone. If agents can be uniquely identified and reliably matched between all generally non-overlapping views, and a sequence of their arrival and departure events is obtained over a period of time, distributions of the probabilities and durations of transitions can be established. From this information, all of the parameters of the general transition model can be obtained.

Unfortunately, correspondence of agents of arbitrary appearance between generally disjoint views is notoriously difficult. Ellis et al. (2003); Makris et al. (2004) sidestep this challenge with a method of construction based on pure temporal correlation of otherwise unmatched observations. Essentially, they assume implicit correspondence between *all* pairs of arrival and departure events, and seek a single mode of temporal correlation between each pair of entry and exit zones within a time window (positive and negative); every peak above a certain threshold induces an arc in the transition graph between the associated vertices. Stauffer (2005) employs a similar method, but considers transitions between overlapping cameras separately (see Sect. 4.3.3), so the transition time window is positive only. Tieu et al. (2005) handle more general statistical dependences, capturing richer multi-modal transition distributions rather than simply a mean transition duration, and thus permitting topology estimation from more complex agent behavior. Marinakis et al. (2005); Marinakis and Dudek (2005) also avoid direct correspondence. They assume that the dynamics of an agent is a Markov process, and estimate the parameters of this process—the probabilities and durations of transitions—using a Monte Carlo Expectation Maximization method.

Methods which do rely on appearance-based agent correspondence normally have a narrower application focus. Dick and Brooks (2004) require a training phase for their Markov model which relies on colour-based correspondence. Niu and Grimson (2006) rely on correspondence of tracked vehicles using an appearance model based on colour and size. The estimation method of Nam et al. (2007) centers around correspondence based on background subtraction and a human appearance model. Farrell and Davis (2008) employ an information-theoretic appearance matching process, and infer the expected transition model from the accumulated evidence using a modified multinomial distribution. Their method is also notable for its distributed design: its "semi-localized" processing yields a scalable algorithm for which the authors demonstrate successful results in networks up to 100 nodes.

Zou et al. (2007) integrate correspondence based on face recognition into the previously described statistical method of Ellis, Makris, and Black, resulting in a hybrid approach which they claim outperforms methods based purely on either identity or appearance.



(a) Non-Overlapping **(b)** Overlapping

Fig. 10 Possible cases of transition—*dark ellipses* denote entry and exit zones, and the *dotted line* indicates the agent path

4.3.3 Transitions Between Overlapping Cameras

There is a question as to how transitions between cameras with overlapping coverage should be handled in transition models. Referring to the example agent paths in Fig. 10, it is clear how to handle the transition between non-overlapping cameras shown in Fig. 10a, as the surveyed methods unanimously agree: an arc from A or its exit zone to B or its entry zone, with a positive transit duration. However, in the transition between overlapping cameras shown in Fig. 10b, the agent passes through the entry zone of B *before* passing through the exit zone of A, and the agent is observed by one or both cameras during the entire transition. Transitions from one entry or exit zone to another within a single camera's coverage can be thought of as a special case of this scenario.

Ellis et al. (2003); Makris et al. (2004) deal with the overlapping case as with the non-overlapping case. For a given departure event at time t_1, they check for arrival events at time $t_2 \in [t_1 - T, t_1 + T]$, where T is a temporal search window. Thus, in Fig. 10a, $t_2 > t_1$, whereas in Fig. 10b, $t_2 < t_1$. The advantage of this approach is that it does not require prior estimation of overlap topology, and uses a single process to estimate transition topology for a general-case camera network with overlapping and/or non-overlapping cameras.

Stauffer (2005) argues that the overlapping case is best handled by more robust direct tracking correspondence, and proposes first estimating overlap topology (Stauffer and Tieu 2003), then treating connected components in the vision graph as single "cameras"—in general, with multiple entry and exit zone vertices—in the transition model. The advantage of this approach is improved robustness in estimating the overlapping portions of the transition topology, assuming a reliable means of finding inter-camera correspondences of agents and/or their tracks is available.

4.4 State of the Art and Open Problems

Numerous researchers have converged on the structure described in Sect. 4.1, to varying degrees. As with coverage overlap models, it is safe to say that this generalized model subsumes all existing cases; individual models have left out certain properties (arc directivity and weights, node subdivision, source/sink node) either because they are unnecessary for the particular application case or else to facilitate optimization. Given the clear focus on a single application class, future optimization efforts should adopt such a unified model, if possible, for the sake of general applicability.

Approximation of the graph from visual data is split between statistical temporal correlation and appearance-based correlation. Given the complementary strengths of both methods, the way forward seems to be a hybrid approach in the vein of Zou et al. (2007). If agent dynamics are being modeled for the purposes of probabilistic occlusion, as by Mittal and Davis (2008), this may also be informative for transition model approximation.

One point of contention, to which the answer is not yet clear, is whether the graph should model transitions strictly between non-overlapping coverage cells, with overlapping transitions handled separately as proposed by Stauffer (2005), or all transitions. If the relative reliability of the approximations for overlapping transitions is the issue, implementation of the aforementioned hybrid approximation approach may favor the latter unified model.

5 Conclusions and Future Directions

We have endeavoured to present a comprehensive and lucid exposition of the theory, state of the art, and challenges of modeling the visual coverage of camera networks. The models and estimation methods discussed in this survey represent the efforts of researchers to develop theory to describe the particular and unique properties of this relatively new type of system, drawing on concepts from computer vision, sensor networks, and other related fields.

Through an analysis of their properties in the context of specific applications, a generalized prototype model of each type has been derived, of which the major structure of the models in the surveyed works can be cast as particular instances. In addition to providing a clear overview of the three types of models individually, it is relevant at this point to explicitly expose the relationships between them in terms of the information they encapsulate.

To date, camera network coverage has not been described from the vantage of an inclusive understanding of its various applications. Researchers have developed tools to achieve particular objectives, adapting work from similar but not quite identical problems elsewhere in the landscape. Over time, this has begun to converge and evolve into a theoretical framework particular to camera networks. It is our belief that the unique challenges involved warrant a next phase in modeling camera network coverage, viz. the development of a comprehensive, rigorous, and mature theory encompassing geometric coverage as well as both notions of topology. Our generalized models and exposition of their information hierarchy are a first step in this direction, but a truly useful theory will be forged in the fire of application, and many cues can be taken from the design decisions made in the various works surveyed here. A general, analytic understanding of coverage will reduce duplicated effort and open up new possibilities in solving a large cross-section of important problems in camera networks.

Acknowledgments This research was supported in part by the Natural Sciences and Engineering Research Council of Canada.

References

Angella, F., Reithler, L., & Gallesio, F. (2007). Optimal deployment of cameras for video surveillance systems. In *Proceedings of IEEE conference on advanced video and signal based surveillance* (pp. 388–392).

Antone, M., & Teller, S. (2002). Scalable extrinsic calibration of omni-directional image networks. *International Journal of Computer Vision*, *49*(2/3), 143–174.

Bajramovic, F., Brückner, M., & Denzler, J. (2009). Using common field of view detection for multi camera calibration. In *Proceedings of vision, modeling, and visualization workshop*.

Bodor, R., Drenner, A., Janssen, M., Schrater, P., & Papanikolopoulos, N. (2005). Mobile camera positioning to optimize the observability of human activity recognition tasks. In *Proceedings of IEEE/RSJ international conference on intelligent robots* (pp. 4037–4042).

Bodor, R., Drenner, A., Schrater, P., & Papanikolopoulos, N. (2007). Optimal camera placement for automated surveillance tasks. *Journal of Intelligent and Robotic Systems*, *50*(3), 257–295.

Brand, M., Antone, M., & Teller, S. (2004). Spectral soluction of large-scale extrinsic camera calibration as a graph embedding problem. In *Proceedings of 8th European conference on computer vision* (pp. 262–273).

Brückner, M., Bajramovic, F., & Denzler, J. (2009). Geometric and probabilistic image dissimilarity measures for common field of view detection. In *Proceedings of IEEE computer society conference on computer vision and, pattern recognition* (pp. 2052–2057).

Cerfontaine, P. A., Schirski, M., Bundgens, D., & Kuhlen, T. (2006). Automatic multi-camera setup optimization for optical tracking. In *Proceedings of virtual reality conference* (pp. 295–296).

Chen, S., & Li, Y. (2004). Automatic sensor placement for model-based robot vision. *IEEE Transactions on Systems, Man, and Cybernetics*, *34*(1), 393–408.

Chen, T. S., Tsai, H. W., Chen, C. P., & Peng, J. J. (2010). Object coverage with camera rotation in visual sensor networks. In *Proceedings of 6th international wireless communications and mobile computing conference* (pp. 79–83).

Chen, X., & Davis, J. (2008). An occlusion metric for selecting robust camera configurations. *Machine Vision and Applications*, *19*(4), 217–222.

Cheng, Z., Devarajan, D., & Radke, R. J. (2007). Determining vision graphs for distributed camera networks using feature digests. *EURASIP Journal on Advances in Signal Processing*, *2007*, 1–11.

Chow, K. Y., Lui, K. S., & Lam, E. Y. (2007). Achieving 360 angle coverage with minimum transmission cost in visual sensor networks. In *Proceedings of IEEE wireless communications and networking conference* (pp. 4112–4116).

Cowan, C. K., & Kovesi, P. D. (1988). Automatic sensor placement from vision task requirements. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *10*(3), 407–416.

Dai, R., & Akyildiz, I. F. (2009). A spatial correlation model for visual information in wireless multimedia sensor networks. *IEEE Transactions on Multimedia*, *11*(6), 1148–1159.

Detmold, H., Dick, A. R., Van Den Hengel, A., Cichowski, A., Hill, R., Kocadag, E., Falkner, K., & Munro, D. S. (2007). Topology estimation for thousand-camera surveillance networks. In *Proceedings of 1st ACM/IEEE international conference on distributed smart cameras* (pp. 195–202).

Detmold, H., Dick, A. R., Van Den Hengel, A., Cichowski, A., Hill, R., Kocadag, E., Yarom, Y., Falkner, K., & Munro, D. S. (2008). Estimating camera overlap in large and growing networks. In *Proceedings of 2nd ACM/IEEE international conference on distributed smart cameras*.

Devarajan, D., & Radke, R. J. (2004). Distributed metric calibration of large camera networks. In *Proceedings of 1st workshop on broadband advanced sensor networks*.

Dick, A. R., & Brooks, M. J. (2004). A stochastic approach to tracking objects across multiple cameras. In *Proceedings Australian joint conference on artificial intelligence* (pp. 160–170).

Ellis, T. J., Makris, D., & Black, J. K. (2003). Learning a multi-camera topology. In *Proceedings of joint IEEE workshop on visual surveillance and performance evaluation of tracking and surveillance* (pp. 165–171).

Erdem, U. M., & Sclaroff, S. (2003). Automated placement of cameras in a floorplan to satisfy task-specific constraints. Tech. Report, Boston University.

Erdem, U. M., & Sclaroff, S. (2006). Automated camera layout to satisfy task-specific and floor plan-specific coverage requirements. *Computer Vision and Image Understanding*, *103*(3), 156–169.

Farrell, R., & Davis, L. S. (2008). Decentralized discovery of camera network topology. In *Proceedings of 2nd ACM/IEEE international conference on distributed smart cameras*.

Faugeras, O. (1993). *Dimensional computer vision: A geometric viewpoint*. London: MIT Press.

Fiore, L., Somasundaram, G., Drenner, A., & Papanikolopoulos, N. (2008). Optimal camera placement with adaptation to dynamic scenes. In *Proceedings of IEEE international conference on robotics and automation* (pp. 956–961).

Gilbert, A., & Bowden, R. (2006). Tracking objects across cameras by incrementally learning inter-camera colour calibration and patterns of activity. In *Proceedings of 9th European conference on computer vision* (pp. 125–136).

González-Banos, H., & Latombe, J. C. (2001). A randomized art-gallery algorithm for sensor placement. In *Proceedings of 17th annual symposium computational geometry* (pp. 232–240).

Hill, R., Dick, A. R., Van Den Hengel, A., Cichowski, A., & Detmold, H. (2008). Empirical evaluation of the exclusion approach to estimating camera overlap. In *Proceedings of 2nd ACM/IEEE international conference on distributed smart cameras*.

Hörster, E., & Lienhart, R. (2006). On the optimal placement of multiple visual sensors. In *Proceedings of 4th ACM international workshop on video surveillance and sensor, networks* (pp. 111–120).

Hörster, E., & Lienhart, R. (2009). Optimal placement of multiple visual sensors. In H. Aghajan & A. Cavallaro (Eds.), *Multi-camera networks: Principles and applications* (Chap. 5, pp. 117–138). Burlington: Academic Press.

Huang, C. F., Tseng, Y. C., & Lo, L. C. (2007). The Coverage Problem in Three-Dimensional Wireless Sensor Networks. *J. Interconnection Networks*, *8*(3), 209–227.

Huber, D. F. (2001). Automatic 3D modeling using range images obtained from unknown viewpoints. In *Proceedings of 3rd international conference on 3D digital imaging and modeling* (Vol. 7, pp. 153–160).

Javed, O., Khan, S., Rasheed, Z., & Shah, M. (2000). Camera handoff: Tracking in multiple uncalibrated stationary cameras. In *Proceedings of workshop on human motion* (pp. 113–118).

Javed, O., Rasheed, Z., Shafique, K., & Shah, M. (2003). Tracking across multiple cameras with disjoint views. In *Proceedings of 9th IEEE international conference on computer vision* (pp. 952–957).

Jiang, Y., Yang, J., Chen, W., & Wang, W. (2010). A coverage enhancement method of directional sensor network based on genetic algorithm for occlusion-free surveillance. In *Proceedings of international conference on computational aspects of social networks* (pp. 311–314).

Kang, E. Y., Cohen, I., & Medioni, G. G. (2000). A graph-based global registration for 2D mosaics integrated media systems center. In *Proceedings of international conference on pattern recognition* (pp. 257–260).

Khan, S., & Shah, M. (2003). Consistent labeling of tracked objects in multiple cameras with overlapping fields of view. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *25*(10), 1355–1360.

Kulkarni, P., Shenoy, P., & Ganesan, D. (2007). Approximate initialization of camera sensor networks. In *Proceedings of 4th European conference on wireless sensor networks* (pp. 67–82).

Kurillo, G., Li, Z., & Bajcsy, R. (2008). Wide-area external multi-camera calibration using vision graphs and virtual calibration object. In *Proceedings of 2nd ACM/IEEE international conference on distributed smart cameras*.

Liu, L., Ma, H., & Zhang, X. (2008). Analysis for localization-oriented coverage in camera sensor networks. In *Proceedings of wireless communications and networking conference* (pp. 2579–2584).

Liu, L., Ma, H., & Zhang, X. (2008). On directional K-coverage analysis of randomly deployed camera sensor networks. In *Proceedings of IEEE international conference on communications* (pp. 2707–2711).

Lobaton, E. J., Ahammad, P., & Sastry, S. S. (2009a). Algebraic approach for recovering topology in distributed camera networks. In *Proceedings of ACM/IEEE international conference on information processing in sensor networks*.

Lobaton, E. J., Sastry, S. S., & Ahammad, P. (2009b). Building an algebraic topological model of wireless camera networks. In H. Aghajan & A. Cavallero (Eds.), *Multi-camera networks: Principles and applications* (Chap. 4, pp. 95–115). St. Louis: Academic Press.

Lobaton, E. J., Vasudevan, R., Bajcsy, R., & Sastry, S. (2010). A distributed topological camera network representation for tracking applications. *IEEE Transactions on Image Processing*, *19*(10), 2516–2529.

Lowe, D. G. (2004). Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*, *60*(2), 91–110.

Ma, H., & Liu, Y. (2005a). Correlation based video processing in video sensor networks. In *Proceedings of internaional conference on wireless networks, communications and mobile computing* (pp. 987–992).

Ma, H., & Liu, Y. (2005b). On coverage problems of directional sensor networks. In *Proceedings of 1st international conference on mobile ad-hoc and sensor networks* (pp. 721–731).

Ma, H., & Liu, Y. (2007). Some problems of directional sensor networks. *International Journal of Sensor Networks*, *2*(1–2), 44–52.

Ma, Y., Soatto, S., Košecká, J., & Sastry, S. S. (2004). *An invitation to 3-D computer vision*. London: Springer.

Makris, D., Ellis, T. J., & Black, J. K. (2004). Bridging the gaps between cameras. In *Proceedings of IEEE computer society conference on computer vision and, pattern recognition* (pp. 205–210).

Malik, R., & Bajcsy, P. (2008). Automated placement of multiple stereo cameras. In *Proceedings of 8th ECCV workshop on omnidirectional vision, camera networks and non-classical cameras*.

Mandel, Z., Shimsoni, I., & Keren, D. (2007). Multi-camera topology recovery from coherent motion. In *Proceedings of 1st ACM/IEEE international conference distributed smart cameras* (pp. 243–250).

Marengoni, M., Draper, B. A., Hanson, A., & Sitaraman, R. (2000). A system to place observers on a polyhedral terrain in polynomial time. *Image and Vision Computing*, *18*(10), 773–780.

Marinakis, D., & Dudek, G. (2005). Topology inference for a vision-based sensor network. In *Proceedings of 2nd Canadian conference on computer and robot vision* (pp. 121–128).

Marinakis, D., Dudek, G., & Fleet, D. J. (2005). Learning sensor network topology through Monte Carlo expectation maximization. In *Proceedings of IEEE International conference on robotics and automation*.

Maver, J., & Bajcsy, R. (1993). Occlusions as a guide for planning the next view. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *15*(5), 417–433.

Mavrinac, A., & Chen, X. (2011). Optimizing load distribution in camera networks with a hypergraph model of camera topology. In *Proceedings of 5th ACM/IEEE international conference on distributed smart cameras*.

Mavrinac, A., Chen, X., & Tepe, K. (2010). An automatic calibration method for stereo-based 3D distributed smart camera networks. *Computer Vision and Image Understanding*, *114*(8), 952–962.

Meguerdichian, S., Koushanfar, F., Potkonjak, M., & Srivastava, M. B. (2001). Coverage problems in wireless ad-hoc sensor networks. In *Proceedings of 20th IEEE international conference on computer communications* (pp. 1380–1387).

Mikolajczyk, K., Tuytelaars, T., Schmid, C., Zisserman, A., Matas, J., Schaffalitzky, F., et al. (2006). A comparison of affine region detectors. *International Journal of Computer Vision*, *65*(1), 43–72.

Mittal, A. (2006). Generalized multi-sensor planning. In *Proceedings of 9th European conference on computer vision*.

Mittal, A., & Davis, L.S. (2004). Visibility analysis and sensor planning in dynamic environments. In *Proceedings of 8th European conference on computer vision*.

Mittal, A., & Davis, L. S. (2008). A general method for sensor planning in multi-sensor systems: Extension to random occlusion. *International Journal of Computer Vision*, *76*(1), 31–52.

Moreels, P., & Perona, P. (2007). Evaluation of features detectors and descriptors based on 3D objects. *International Journal of Computer Vision*, *73*(3), 263–284.

Nam, Y., Ryu, J., Choi, Y. J., & Cho, W. D. (2007). Learning spatio-temporal topology of a multi-camera network by tracking multiple people. *Proceedings of World Academy of Science, Engineering and Technology*, *24*, 175–180.

Niu, C., & Grimson, W. E. L. (2006). Recovering non-overlapping network topology using far-field vehicle tracking data. In *Proceedings of 18th international conference on pattern recognition* (pp. 944–949).

O'Rourke, J. (1987). *Art gallery theorems and algorithms*. New York: Oxford University Press.

Park, J., Bhat, P. C., & Kak, A. C. (2006). A look-up table based approach for solving the camera selection problem in large camera networks. In *Proceedings of international workshop on distributed smart cameras*.

Piciarelli, C., Micheloni, C., & Foresti, G. L. (2009). PTZ camera network reconfiguration. In *Proceedings of 3rd ACM/IEEE international conference on distributed smart cameras*.

Piciarelli, C., Micheloni, C., & Foresti, G. L. (2010). Occlusion-aware multiple camera reconfiguration. In *Proceedings of 4th ACM/IEEE international conference on distributed smart cameras*.

Qian, C., & Qi, H. (2008). Coverage estimation in the presence of occlusions for visual sensor networks. In S. Nikoletseas, B. Chlebus, D. Johnson, & B. Krishnamachari (Eds.), *Distributed computing in sensor systems* (pp. 346–356). Berlin: Springer.

Ram, S., Ramakrishnan, K. R., Atrey, P. K., Singh, V. K., & Kankanhalli, M. S. (2006). A design methodology for selection and placement of sensors in multimedia surveillance systems. In *Proceedings of 4th ACM international workshop on video surveillance and sensor networks* (pp. 121–130).

Reed, M. K., & Allen, P. K. (2000). Constraint-based sensor planning for scene modeling. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *22*(12), 1460–1467.

Sawhney, H. S., Hsu, S., & Kumar, R. (1998). Robust video mosaicing through topology inference and local to global alignment. In *Proceedings of 5th European conference on computer vision* (pp. 103–119).

Sharp, G. C., Lee, S. W., & Wehe, D. K. (2004). Multiview registration of 3D scenes by minimizing error between coordinate frames. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *26*(8), 1037–1050.

Shen, C., Zhang, C., & Fels, S. (2007). A multi-camera surveillance system that estimates quality-of-view measurement. In *Proceedings of IEEE international conference on image processing* (pp. 193–196).

Song, B., Kamal, A. T., Soto, C., Ding, C., Farrell, J. A., & Roy-Chowdhury, A. K. (2010). Tracking and activity recognition through consensus in distributed camera networks. *IEEE Transactions on Image Processing*, *19*(10), 2564–2579.

Soro, S., & Heinzelman, W. B. (2007). Camera selection in visual sensor networks. In *Proceedings of IEEE conference on advanced video and signal based surveillance* (pp. 81–86).

Stauffer, C. (2003). Estimating tracking sources and sinks. In *Proceedings of IEEE Computer Society conference on computer vision and pattern recognition* (Vol. 4).

Stauffer, C. (2005). Learning to track objects through unobserved regions. In *Proceedings of IEEE Computer Society workshop on motion and video computing* (pp. 96–102).

Stauffer, C., & Tieu, K. (2003). Automated multi-camera planar tracking correspondence modeling. In *Proceedings of IEEE Computer Society conference on computer vision and, pattern recognition* (pp. 259–266).

Tao, D., Ma, H., & Liu, L. (2006). Coverage-enhancing algorithm for directional sensor networks. In *Proceedings of 2nd international conference on mobile ad-hoc and sensor networks* (pp. 256–267).

Tarabanis, K. A., Allen, P. K., & Tsai, R. Y. (1995). A survey of sensor planning in computer vision. *IEEE Transactions on Robotics and Automation*, *11*(1), 86–104.

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

Tarabanis, K. A., Tsai, R. Y., & Allen, P. K. (1995). The MVP sensor planning system for robotic vision tasks. *IEEE Transactions on Robotics and Automation*, *11*(1), 72–85.

Tarabanis, K. A., Tsai, R. Y., & Kaul, A. (1996). Computing occlusion-free viewpoints. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *18*(3), 279–292.

Tieu, K., Dalley, G., & Grimson, W. E. L. (2005). Inference of non-overlapping camera network topology by measuring statistical dependence. In *Proceedings of 10th IEEE international conference on computer vision* (pp. 1842–1849).

Van Den Hengel, A., Dick, A. R., Detmold, H., Cichowski, A., & Hill, R. (2007). Finding camera overlap in large surveillance networks. In *Proc. 8th Asian conference on computer vision* (pp. 375–384).

Van Den Hengel, A., Dick, A. R., & Hill, R. (2006). Activity topology estimation for large networks of cameras. In *Proceedings of IEEE international conference on advanced video and signal based surveillance*.

Wang, B. (2010). *Coverage control in sensor networks*. Berlin: Springer.

Wang, C., Qi, F., & Shi, G. M. (2009). Nodes placement for optimizing coverage of visual sensor networks. In P. Muneesawang, F. Wu, I. Kumazawa, A. Roeksabutr, M. Liao, & X. Tang (Eds.), *Advances in Multimedia Information Processing—PCM 2009* (pp. 1144–1149). Berlin: Springer.

Yao, Y., Chen, C. H., Abidi, B., Page, D., Koschan, A., & Abidi, M. A. (2008). Sensor planning for automated and persistent object tracking with multiple cameras. In *Proceedings of IEEE Computer Society conference on computer vision and, pattern recognition* (pp. 1–8).

Zhao, J., Cheung, S. C., & Nguyen, T. (2008). Optimal camera network configurations for visual tagging. *Journal of Selected Topics in Signal Processing*, *2*(4), 464–479.

Zhao, J., Cheung, S. C., & Nguyen, T. (2009). Optimal visual sensor network configuration. In H. Aghajan & A. Cavallaro (Eds.), *Multi-camera networks: Principles and applications* (Chap. 6, pp. 139–162). Burlington: Academic Press.

Zou, X., Bhanu, B., Song, B., & Roy-Chowdhury, A. K. (2007). Determining topology in a distributed camera network. In *Proceedings of IEEE international conference on image processing* (pp. 133–136).