OCS-14: You Can Get Occluded in Fourteen Ways

Prithwijit Guha
TCS Innovation Labs,
New Delhi, India
prithwijit.guha@tcs.com

Amitabha Mukerjee
Dept. of Computer Sc. & Engg.
IIT Kanpur, India
amit@cse.iitk.ac.in

K. S. Venkatesh
Dept. of Electrical Engg.
IIT Kanpur, India
venkats@iitk.ac.in

Abstract

Occlusions are a central phenomenon in multi-object computer vision. However, formal analyses (*LOS14*, *ROC20*) proposed in the spatial reasoning literature ignore many distinctions crucial to computer vision, as a result of which these algebras have been largely ignored in vision applications. Two distinctions of relevance to visual computation are (a) whether the occluder is a moving object or part of the static background, and (b) whether the visible part of an object is a connected blob or fragmented. In this work, we develop a formal model of occlusion states that combines these criteria with overlap distinctions modeled in spatial reasoning to come up with a comprehensive set of fourteen occlusion states, which we define as *OCS14*. Transitions between these occlusion states are an important source of information on visual activity (e.g. splits and merges). We show that the resulting formalism is representationally complete in the sense that these states constitute a partition of all possible occlusion situations based on these criteria. Finally, we show results from implementations of this approach in a test application involving static camera based scene analysis, where occlusion state analysis and multiple object tracking can be used for two tasks – (a) identifying static occluders, and (b) modeling a class of interactions represented as transitions of occlusion states. Thus, the formalism is shown to have direct relevance to actual vision applications.

1 Introduction

When observing a scene with moving objects, a relative depth ordering is imposed on the objects and scene background structures along the lines of sight. This depth ordering leads to the partial/complete viewing obstruction of some of the objects of interest and the phenomenon is known as *occlusion* (Figure 1).

Occlusion carries information on the relative depth of objects and other aspects of a scene, and is relevant to a range of vision problems such as multi-object tracking [Yilmaz *et al.*, 2006], activity modeling [Guha *et al.*, 2007], etc. In spatial reasoning, occlusion is informative on the relative pose and motion of multiple objects. In human cognition, a sensitivity to occlusion appears to be encoded by about three months, and is crucial to acquiring concepts such as object persistence, containment and support [Baillargeon and hua Wang, 2002].

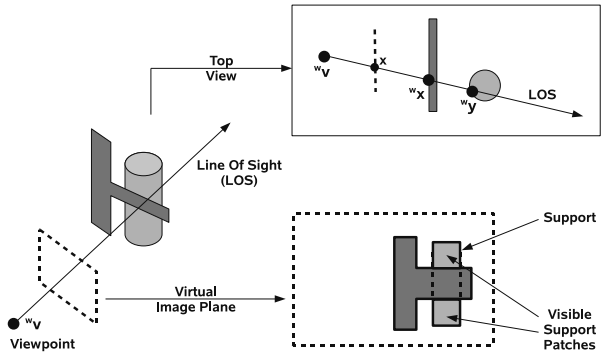

Figure 1: An object of interest (cylinder X) is occluded by object (Y), when some $3D$ world point $^w y$ on Y lies on the line of sight from viewpoint $^w v$ to some $^w x$ on X. A *relation* model of occlusion might need to say *occludes-fragmenting-while-fully-visible (Y,X)*. A state-based (unary) model might say that Y's state is *grouped-fully-visible*, while X's state might be *grouped-fragmented-visible*. The region of the image where X would be projected had there been no occluding objects is the *support(X)*.

Yet there is no comprehensive analysis of occlusion that capture distinctions that are relevant to visual analysis. Formalisms proposed for occlusion include *LOS-14* [Galton, 1994], *ROC-20* [Randell *et al.*, 2001] and *OCC-8* [Köhler, 2002], but these ignore crucial criteria such as whether the visible parts are connected or not, or whether the occluder is static or moving. Also, many aspects of importance in spatial reasoning (such as precise tangency situations) are less relevant in vision, since these can almost never be detected (though they may be inferable from transitions). As a result these formalisms have not been adopted much in the computer vision community, where implementations still adopt *ad hoc* categories, and mostly consider occlusion as a property of each individual object (*state*), as opposed to a relation between multiple objects, as in spatial reasoning. Some vision systems minimally distinguish isolation states from any

kind of occlusion; others provide a small set of occlusion situations, often confusing occluding of an object with dynamic effects like merging and splitting; while others use the term occlusion but refer only to dynamic events [Yilmaz *et al.*, 2006].

In vision processing, it is important to discriminate *figure* (usually objects of interest which exhibit motion, e.g. vehicles, people etc.) from *ground* (scene components that remain static through out a video, e.g. trees, posts, walls etc.). A moving car is a dynamic object (*figure*), but if it stops for a long period as at a traffic light, it may be treated as static (*ground*). A common practice in computer vision is to segment foreground objects as "blobs" (maximal connected pixel sets). When an object is occluded by other dynamic objects, their blobs will merge together into a single connected foreground blob, but when it is occluded only by a static object, the occluding object will not be part of the foreground blob. Thus this distinction becomes crucial.

In this work, we consider primarily scenes in which one or more opaque objects are moving against a relatively stable background; the set of moving objects constitute the foreground. The relatively stable background can now be used to segment and attend to the foreground portions.

We are interested in $3D$ world points ${}^w x$ lying on the surface of either the static background ${}^w B$ or on one of the dynamic (moving) objects ${}^w S_i$, where each object ${}^w S_i$ is the surface of a connected set[1]. The line joining a given viewpoint ${}^w v$ with ${}^w x$ is the *line of sight* (LOS) of ${}^w x$ (Figure 1). The $2D$ point x is the projection of ${}^w x$, if the LOS through ${}^w x$ intersects the image plane at x; this is captured by the predicate $isProjection(x, {}^w x)$. We are interested in points on the surfaces of the objects exposed to the vision sensor (${}^w S_{exp}(i)$). We define the predicate $pointOrderLOS({}^w x, {}^w y)$ indicating that ${}^w x$ and ${}^w y$ lie on the same line of sight and ${}^w x$ is nearer to the viewpoint.

We define the point occlusion and visibility as $occludes({}^w x, {}^w y) \equiv pointOrderLOS({}^w x, {}^w y)$ and $visible({}^w x) \equiv \nexists {}^w z \; occludes({}^w z, {}^w x)$. The set of points exposed to the viewpoint (not occluded by the object itself) is ${}^w S_{exp}(i) = \{{}^w x \in {}^w S_i : (\nexists {}^w z \in {}^w S_i) occludes({}^w z, {}^w x)\}$. The projection of ${}^w S_{exp}(i)$ onto the image plane, S_i, constitutes the maximal visible extent of ${}^w S_i$ or its *support*. The *support* S_i for the $3D$ point set ${}^w S_i$ (Figure 1) can be defined as $S_i = support({}^w S_i) = \{x : (\exists {}^w x \in {}^w S_i) isProjection(x, {}^w x)\}$. This is the set of pixels that would have been the object's image if there were no occluding objects. An object is *isolated* from other objects, if its support does not overlap with any other: $\forall k \neq i \; [S_i \wedge S_k = \varnothing]$.

We can now define subset of the support that is actually visible, or the *visible support* of S_i, as $V_i = \{x : (\exists {}^w x \in {}^w S_i) isProjection(x, {}^w x) \wedge visible({}^w x)\}$; $(V_i \subseteq S_i)$. The image foreground is then the $\cup V_i$, which is divided into a set of maximally connected "visible patches".

[1]Convention: points are in lower case, point sets in upper case and sets of point sets in bold upper case (x, S, $\mathbf{S}$). Also, world elements are pre-superscripted ^{w}, image elements are unmarked.

1.1 Relation vs State

Formalizations of occlusion in spatial reasoning are based on *binary relations* that involve two objects (*n-ary relations* become too verbose). Relations are helpful since they permit transitive inference and one may use tractability and other properties from formal algebras. However, occlusions in visual contexts often involve more than two objects, and even a 3-body relation system would have several hundred relations. Even for two-object relations, accounting for different types of fragmentation (Figure 3(b)) [Galton, 1998] and also the static/dynamic distinction would result in a explosion of relational distinctions. Also, existing formalizations consider objects to be separated in depth; however, non-convex objects (like two humans hugging) may be occluding each other; in k-ary situations, these relations would add many depth layers and the relations set would further explode.

On the other hand, if we maintain just the states of the individual objects, it provides a much more compact representation, though it is not as rich (is not representationally equivalent) to the relation based model. Thus, given the set of states of each object, the relations between them cannot be inferred, so there is some loss of information in going from relations to states. However, given that many objects may be interacting, relational algebras for modeling such interactions would be extremely large, and adopting a state-based model provides a mechanism that makes the relevant distinctions. Also, in many situations we are primarily interested in one object (in attentive focus) and hence it's states are the most relevant.

1.2 Characteristics of Occlusion

Any representation must attempt to preserve those aspects of the problem that are relevant to the task. In our analysis of occlusion in the vision literature [Yilmaz *et al.*, 2006], three characteristics appear repeatedly. As discussed above, the **static/dynamic** distinction is key to tracking the objects. Also, the visible support of an object may sometimes get split into multiple blobs (e.g. when broken up by a static object in front such as a tree or lamp-post) or the supports of different objects may merge into a single blob (dynamic occlusions). Thus, the degree of **visibility** - fully visbile, partially visible, fragmented or invisible - is an important aspect. Finally, one needs to know the state of **isolation** – is the object isolated or is it grouped with other dynamic objects? These three dimensions define the important distinctions we wish to make regarding occlusions. Of these, some aspects of visibility and isolation have been considered in other formalisms of occlusion [Galton, 1994; 1998] which we consider in Section 2.

In most visual tasks, we distinguish two computationally important concepts related to the object of interest – first, the hypothetical construct that we call *support* S_i, and second, the actual visible region - the *visible support* V_i. The occlusion relations are defined based on the support's relation to the visible patches, and the goal is to try to infer the unknown *support* based on the evidence. Given that multiple objects are visible at once, their visible supports may overlap or fragment, resulting in a set of blobs. The objective in tracking is to try to infer the hypothetical support on the image as best as

we can. The objective in activity analysis is to use the occlusion relations to infer the nature of the action.

In constructing any representation, one cannot make *all* possible distinctions. For example, fragmentation may result in a number of fragments (say k) - each k may demand a different distinction on this dimension alone. A pragmatic compromise is to argue that for the majority of occlusion judgments it is sufficient to represent only whether the object is unfragmented ($k = 1$), or is it fragmented ($k > 1$). This also permits us to consider fragmentation as a special phenomenon in the degree of visibility dimension (Section 3.1). Thus the representation proposed has three separate dimensions spanning the occlusion state space.

1. *Visibility* (4 states): invisible, fragmented-visible, unfragmented-partially-visible, and fully-visible ($v0$, vF, vP, $v1$). (Section 3.1)
2. *Static/Dynamic* (4 states): is the object unoccluded, or is the occluding object static, dynamic, or both (1, S, D, SD).
3. *Isolation* (2 states): isolation (I) or grouping (G).

While a formal analysis is presented in Section 3, we observe that though there are 32 possible combinations, many are not possible, leaving us with fourteen occlusion states (Section 4), which we call the *OCS-14* (Figure 2). We next consider the relation between our representational states and relations in qualitative spatial reasoning.

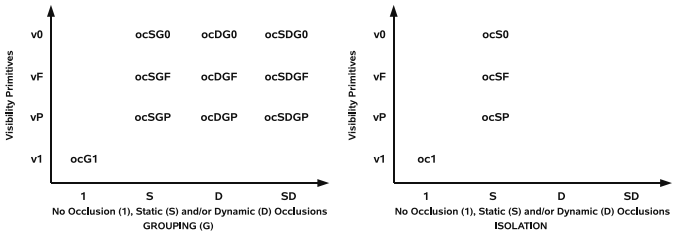

Figure 2: The three dimensions shown in two slices - Grouping (left slice) and Isolation (right slice). Static/Dynamic along horizontal axis, (1/S/D/SD), and the degree of visibility along vertical axis ($v1/vP/vF/v0$). Many distinctions are not feasible - e.g. dynamic occlusions (with object support intersections) cannot be isolated and, we show that only fourteen discriminations (*OCS-14*) are possible (Section 4).

2 Occlusions as Visual Relational Algebra

Historically, there have been two approaches towards modeling occlusion; those from the computational vision perspective have been pragmatic, and the representation has been problem dependent and idiosyncratic [Yilmaz *et al.*, 2006]. For example, since the merging and splitting of blobs is a phenomena of interest in vision, these are often lumped with occlusion phenomena such as isolated vs. partial overlap.

On the other hand, in Qualitative Spatial Reasoning, representationally complete models (equivalence relations sometimes called jointly exhaustive pairwise disjoint or JEPD) have been developed that treat occlusion as overlap, and pay particular attention to boundary tangency situations. Thus, the *RCC-8* calculus which makes eight distinctions for planar shapes – disconnected (DC), externally connected (EC), partial overlap (PO), parthood with or without boundary contacts (TPP, $TPPi$, $NTPP$ and $NTPPi$) and equality (EQ). The notation Ri denotes the inverse of the relation R - e.g. $A\ NTPP\ B$ - that A is a non-tangential proper part of B - is the same as saying $B\ NTPPi\ A$. (Figure 3(a)). Note that the PO relation combines a diverse set of overlap situations including fragmentation (Figure 3(b)). QSR models that try to make such distinctions (see [Galton, 1998]) quickly become too cumbersome.

Occlusions are related to these relations because projections of a pair of $3D$ objects onto the image plane would maintain these relations, and adding a depth characterization would then enable reasoning between them. This proposal is fleshed out by Galton [Galton, 1994], whose *lines of sight* (*LOS-14*) formulation (Figure 3(c)) bifurcates the six overlap situations in *RCC-8*, resulting in 14 relations. Thus, *LOS-14* extends DC to "clear visibility" (C); EC to "just clear" (JC); PO to "partial hiding" (PH and PHi); overlap ($NTPP$) to "hidden-by" and "front-of" (H, F and inverses Hi, Fi); tangential overlap TPP to (just) cases (JHi, JF and inverses JH, JFi); and EQ to exact hiding (EH and EHi). Note that the hiding and front-of relations are assumed to be disjoint - i.e. object A cannot partly occlude B and be partly occluded by it simultaneously. Permitting such non-convex occlusions results in the *ROC-20* calculus [Randell *et al.*, 2001] consisting of the mutual occlusion relations ($MuOccPO$) emerging from PO (Figure 3(d)).

We observe that relations DC, EQ, TPP and $TPPi$ involve discriminations based on tangency, which is not an immediate concern in most situations in computer vision today (though they may be important in future cognitive models). Dropping these relations we have the *RCC-5* calculus which is the basis of the *OCC-8* model of [Köhler, 2002]; here only the overlap relations (EQ, PO, $NTPP$) are bifurcated into hide/front-of, resulting in eight occlusion states.

These QSR-derived formalisms are representationally complete for degree of overlap which is similar but not the same as degree of visibility (our analysis also includes fragmented). Also, it does not consider the static-dynamic and the isolated-grouped distinction. Our state-based formalism for these three dimensions is discussed next (Section 3).

3 Occlusion and Visibility

We recall the notions of *support* S_i and *visible support* V_i introduced earlier, and consider a decomposition of the support and the notion of projection from the $3D$ world space to the $2D$ image plane. The occlusions due to which the visible support may be smaller than the support arise due to static or dynamic occluders, and constitute a partition on the support:

- ${}^wS_{stat}(i) = \{{}^wx \in {}^wS_{exp}(i) : (\exists {}^wz \in {}^wB) \wedge occludes({}^wz, {}^wx) \wedge visible({}^wz)\}$: Set of points undergoing static occlusions, i.e. by visible points belonging to the scene background.
- ${}^wS_{dyn}(i) = \{{}^wx \in {}^wS_{exp}(i) : \exists({}^wz \in {}^wS_j) \wedge (i \neq j) occludes({}^wz, {}^wx) \wedge visible({}^wz)\}$: Set of points un-

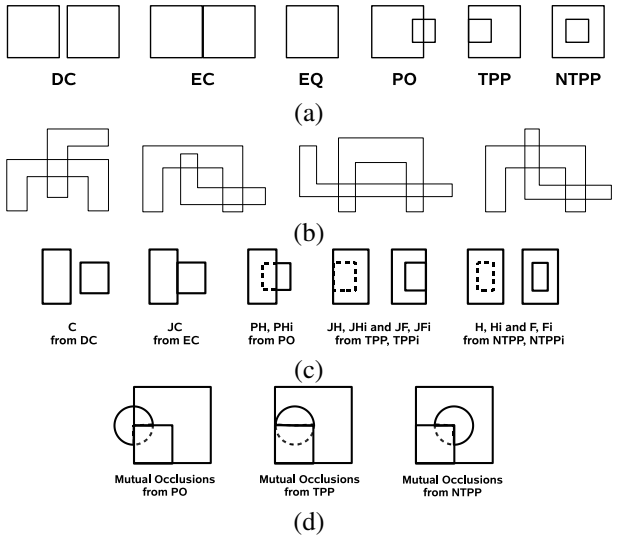

Figure 3: Illustrating the region connection relations – (a) Six of the *RCC-8* region connection relations (inverses of TPP and $NTPP$ are excluded. (b) The multiply connected spatial arrangements that are manifested by the same relation PO. (c) Schematic diagrams of the *LOS-14* relations, (excluding EH and EHi) [Galton, 1994]. (d) The Mutual occlusion relations from *ROC-20* [Randell *et al.*, 2001] (inverse relations and equality excluded).

dergoing dynamic occlusions, i.e. by visible points belonging to the objects of interest other than the i^{th} one.

- ${}^wS_{vis}(i) = {}^wS_{exp}(i) - {}^wS_{stat}(i) \cup {}^wS_{dyn}(i) = \{{}^wx : ({}^wx \in {}^wS_i) \wedge visible({}^wx)\}$: Set of unoccluded visible points.

By definition, each of these subsets of S_{exp} is disjoint, and ${}^wS_{stat}(i) \cup {}^wS_{dyn}(i) \cup {}^wS_{vis}(i) = {}^wS_{exp}(i)$. The following four situations can be derived further by using the definitions of ${}^wS_{stat}$, ${}^wS_{dyn}$ and ${}^wS_{vis}$.

- *Only Static Occlusions* ($[{}^wS_{stat}(i) \neq \varnothing] \wedge [{}^wS_{dyn}(i) = \varnothing]$),
- *Only Dynamic Occlusions* ($[{}^wS_{stat}(i) = \varnothing] \wedge [{}^wS_{dyn}(i) \neq \varnothing]$),
- *Both Static and Dynamic Occlusions* ($[{}^wS_{stat}(i) \neq \varnothing] \wedge [{}^wS_{dyn}(i) \neq \varnothing]$)
- *No Occlusion* ($[{}^wS_{stat}(i) = \varnothing] \wedge [{}^wS_{dyn}(i) = \varnothing]$) as the object stays fully visible.

These four situations constitute the four partitions on the static-dynamic criteria (along the horizontal axis in Figure 2).

3.1 Visibility in the Image Plane

An $image$ is formed at x only if x is $visible$, hence we define the predicate $isImage(x, {}^wx) \leftrightarrow isProjection(x, {}^wx) \wedge visible({}^wx)$, which leads to the notion of *visible support* as $V_i = \{x : (\exists {}^wx \in {}^wS_i) isImage(x, {}^wx)\}$ ($V_i \subseteq S_i$).

The visible support may get partitioned into several maximally connected pixel-sets or *visible patches* $P_i(r)$. We define the set of such patches as $\mathbf{V_{patch}}(i) = \{P_i(r) : \forall r' \neq r[P_i(r) \cap P_i(r') = \varnothing] \wedge [\cup_r P_i(r) = V_i] \wedge maximallyConnectedPixelSet(P_i(r))\}$.

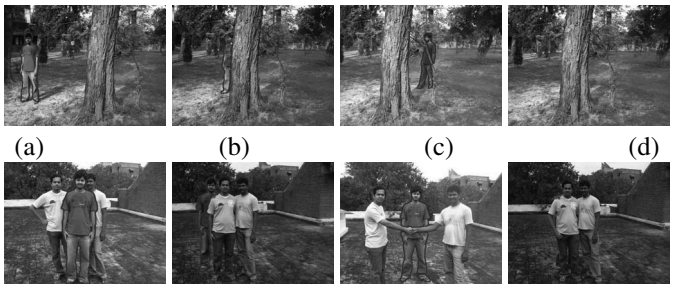

Figure 4: Visibility cases under static/dynamic occlusions. The object of interest is the person in red shirt and visible patches are outlined in blue. (a)-(d) Static occlusion by parts of tree; (e)-(f) Dynamic occlusion by other moving objects. Both (a),(e) illustrates full visibility ($v1$), but in (a) the object is Isolated, while in (e) it is grouped with others. (b),(f) shows non-fragmented partial visibility (vP), while (c),(g) portray fragmented visibility (vF). In (d),(h) the object is invisible due to complete occlusion ($v0$).

Table 1: Feasible visibility situations from fragmentation ($|\mathbf{V_{patch}}(i)|$) and the part-hood of V_i in S_i. The infeasible cases are marked with $\times$

	$V_i = \varnothing$	$V_i \subset S_i$	$V_i = S_i$
$\lvert\mathbf{V_{patch}}(i)\rvert = 0$	$v0(i)$	$\times$	$\times$
$\lvert\mathbf{V_{patch}}(i)\rvert = 1$	$\times$	$vP(i)$	$v1(i)$
$\lvert\mathbf{V_{patch}}(i)\rvert > 1$	$\times$	$vF(i)$	$\times$

The cardinality of $|\mathbf{V_{patch}}(i)|$ can be 0, 1, or more than 1 (since we do not distinguish higher fragments). Similarly, $[V_i = \varnothing]$, $[V_i \neq \varnothing] \wedge [V_i \subset S_i]$ and $V_i = S_i$ lead to 3 possibilities. Of these 3×3 or 9 possibilities, only 4 are however possible due to mutual constraints (table 1), which we call the *visibility primitives* (Figure 4). For example, if there are more than one patches, it is not possible that the visible support equals the hypothetical support. The four visibility primitives $v0$, vF, vP, $v1$ are disjoint by construction, and constitute a partition on the visibility dimension.

For the third dimension of isolation and grouping, we consider overlaps between the supports for the various moving objects. An object may be isolated (disjoint from other objects of interest), or it may be grouped (occluding or occluded by one or more moving object). Thus, there are only two states, *isolation* or *grouping*, and this constitutes a partition on this dimension.

4 Occlusion States

We saw that the static-dynamic analysis led to 4 cases: unoccluded, static, dynamic and both static-dynamic occlusions. Isolation/grouping results in two states. Along with the 4 cases on the visibility dimension, these lead to 32 possible state distinctions. However, as we show, only 14 of these cases are actually possible (The *OCS-14*, Figure 2).

An occluding state where the object is dynamically occluded by another object, and is visible in a fragmented way, would be described as dynamic, grouped, and fragmented

(e.g. the cylinder of Figure 1). Such a situation may be written in short as $ocDGF$. Here the notation is obtained by concatenating "oc" for occlusion with none of "S", "D" and "SD" for static/dynamic distinctions; "G" or its omission for grouping / isolated; and "1", "P" and "F", "0" for whole, part, fragmented and non-visibility respectively. Thus, $ocDGF$ indicates that the cylinder is dynamically-occluded, grouped, and that it is fragmented. On the other hand, the T-shape of Figure 1) is $oG1$ since it is not statically or dynamically occluded, it is grouped, and is wholly visible. Note that each of these states may arise from a variety of imaging situations (Figure 5).

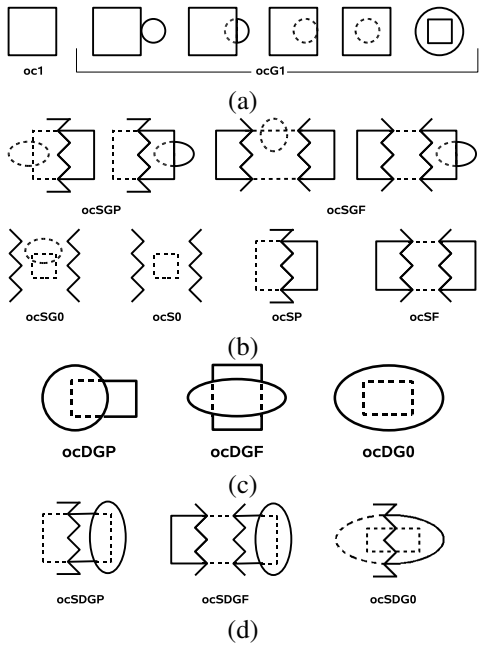

Figure 5: Illustrating the occlusion states. *Rectangle*: object under consideration. *Oval*: other dynamic object(s). *Saw-tooth*: occluding boundaries of the scene background (open-ended). (a) *No occlusion* – states $oc1$ and $ocG1$ signify full visibility, in isolation or in a group respectively. In state $ocG1$, there are many relative situations with the other object while the rectangle maintains full visibility. (b) *Static occlusions* – under static occlusions while grouping, $ocSGP$, $ocSGF$ and $ocSG0$ represent single-part, fragmented and no visibility. Many configurations exist for $ocSGP$ and $ocSGF$. Under static occlusions in isolation, states $ocSP$, $ocSF$ and $ocS0$ indicate similar distinctions. (c) *Dynamic occlusions* – A dynamically occluded object is always grouped; states "$ocDPG$", "$ocDFG$" and "$ocDG0$" signify the visibility distinctions. (d) *Both Static and Dynamic occlusions* – "$ocSDPG$", "$ocSDFG$" and "$ocSDG0$" exhibit similar distinctions.

4.1 How do occlusion states map to relations?

We observe that in many occlusion states, the object under consideration can be in differing occlusion relations with the interacting object (e.g. the state $ocG1$, Figure 5(a)). Consider two interacting objects A and B, and let their occlusion states be one of the four dynamic states: $ocG1$, $ocDGP$, $ocDGF$ and $ocDG0$. Now, each state combination corresponds to a finite set of relations. Considering the relation algebra ***ROC-20***, we see that if A is $ocG1$ and B is $ocDGP$ (middle figure in Figure 5(a)), then the possible relation may be either $APHB$ or $AJFB$ or AFB. Similarly, the various combinations of states result in different binary relations from ***ROC-20*** as shown in Table 2. Combinations of $ocDGP$ or $oCDGF$ is possible only for non-convex objects and result in mutual occlusion ($MuOccPO$) relations from ***ROC-20***.

Table 2: Binary relations corresponding to states. Objects A (states along row) and B (states along column) are in one of the states from $ocG1$, $ocDGP$, $ocDGF$ and $ocDG0$. The possible binary occlusion relation R, read as $A\{R\}B$, is given. For example when A is in state $ocG1$ and B in $ocDGP$, then the possible relation PH implies $A\{PH\}B$.

	ocG1	**ocDGP**	**ocDGF**	**ocDG0**
ocG1	JC	PH, JF, F	PH	H, JH, EH
ocDGP	PHi, JFi, Fi	MuOccPO	MuOccPO	×
ocDGF	PHi	MuOccPO	MuOccPO	×
ocDG0	Hi, JHi, EHi	×	×	×

Note that the objects in the same state combinations can realize different occlusion relations based on their extent of overlaps. Within these four states of dynamic occlusion we can realize all the relations of ***ROC-20*** formulation. More so, we can distinguish the cases of fragmentation or multiple connected regions in a simpler way, which otherwise would have taken a huge burden of formulating a large number of relationships if expressed by changing depth ordering in the modes of overlap formulation [Galton, 1994].

5 Applications of Occlusion States

The discourse in computer Vision circles places occlusion in a consistently negative light - as something that has to be *overcome* to obtain the correct tracking results and so on. A typical tracking application may use occlusion states to determine the update rules - different update mechanisms are used when the object is isolated vs when under static occlusion, and one may cease updates altogether once it is in a multi-group situation.

However, cognitively, occlusion may be one of the more significant sources of qualitative depth (ordering) information. Occlusion transitions are an important visual signature of the interaction between objects. Through the formal analysis presented above, it becomes possible to use data mining techniques on the occlusion transitions as they emerge in a scene, and one can gain useful abstractions about the scene and the object behaviors.

We exemplify this with two fixed-camera visual surveillance tasks on a simple image sequence (Figure 6). Even in such simple scenes, not every distinction presented in the analysis above is recoverable in the computational process. In particular, the applications outlined below discard the distinctions between the fragmented and grouped states.

In the first task, we attempt to (a) impose a depth-ordering on the background ${}^{w}B$ using the analysis, (b) mine the occlusion transitions to learn an event signature for a single action,

and (c) use occlusion characteristics to query a surveillance video and obtain all similar events. The camera is looking at a tree behind which several people are walking around (only one episode is shown). The object color distribution and support are updated only in isolated conditions to provide the tracker with most recent and confident features [Guha *et al.*, 2007]. The "walk past the tree event" is detected as a transition from the initial state of $oc1$ to $ocSP$ since there are no other dynamic objects and $ocSGP$ is ruled out. In repeated episodes, the occlusion boundary always occurs at this part of the image, and it is detected as a high-frequency static edge.

(a) (b) (c) (d)

Figure 6: Extracting information from the spatial distributions of occlusion states. (a)-(c) Results of tracking a person (marked by a red rectangle) walking across a tree. (d) The white pixels mark the spatial distribution of the states of static occlusions. Note that it rightly captures the occluding boundaries of the tree.

For learning the event signatures, we take the temporal transition patterns of the occlusion primitives and model these using *Transition Sequence Mining* (TSM) trees which are clustered using tree-depth weighted Bhattacharya distances [Guha *et al.*, 2007]. We discover different multi-object interactions in the surveillance context. In this particular situation, there are no other objects so we focus on the walk-behind-the-tree episode. This corresponds to a pattern of $oc1 \rightarrow ocSP \rightarrow ocS0 \rightarrow ocSP \rightarrow oc1$, irrespective of which direction the person is walking from. This is a common occurrence in a large number of multi-object interaction situations - where a relatively small object walks behind a parked vehicle or a pillar or a tree and re-emerges on the other side. Querying the TSM trees with the time ordered sequences of the occlusion states detects similarities with other events where this type of sequences arose, thereby detecting episodes of the transit-across-a-large-occlusion event. Thus, in both these problems, we can see how occlusion information is actively used to obtain information about a scene.

6 Conclusion

Our main contribution lies in the identification of a representationally complete set of fourteen occlusion states (*OCS-14*) based on the three characteristics used often in existing work in vision – viz. nature of occluder. visibility, and isolation/grouping. We have argued for a state-based formalism for modeling occlusion in visual phenomena as opposed to a binary or k-ary relation algebra. We have also presented two applications which demonstrate how, unlike previous formal models, this algebra can be actually used in visual computation. The ad hoc references to occlusion situations in existing literature (mainly static camera applications in computer vision) now have a formal model for representing their occlusion states as per application necessities, which is the prime benefit of our proposed formalism.

One of the questions related to applying such formalisms in real tasks is that the distinctions highlighted by the formalism must be recoverable in the computation. As seen in the application above, this is unlikely, especially with dynamic occlusions. Nonetheless, they may provide possibilities which may be of user interest; also in future, they may become computable.

A large number of issues remain. A theoretical issue of interest is the analysis of the conceptual neighborhood for these occlusion states - these are meaningful in identifying neighboring transitions. For example, a transition such as $oc1 \rightarrow ocS0$, without going through a state such as $ocSP$, is likely to indicate some kind of error, missed state, or other problem. Such a conceptual map for *OCS-14* has been identified, but space precludes us from adding it here.

As shown by the applications, the very fact that a coherent and complete representation is available opens up many possible applications for this approach. Beyond the applications to models of tracking etc, issues of occlusion are of great importance in forming a stable conceptualization of the world. Regions of frequent occlusion clearly reveal important information about the scene depth and also about object behaviors in the given scene context. Thus, a object moving behind a screen and repeatedly appearing on the other side tells us something about these objects, and can also prime the visual attention system towards the exit point. Such models lead eventually to a stable realization of the object permanence in the visual scene, which is a key cognitive goal of vision [Baillargeon and hua Wang, 2002].

References

[Baillargeon and hua Wang, 2002] Renée Baillargeon and Su hua Wang. Event categorization in infancy. *Trends in Cognitive Sciences*, 6(2):85–93, February 2002.

[Galton, 1994] Antony Galton. Lines of sight. In *Proceedings of the Seventh Annual Conference on AI and Cognitive Science*, pages 103–113, 1994.

[Galton, 1998] Antony Galton. Modes of overlap. *Journal of Visual Languages and Computing*, 9(1):61–79, 1998.

[Guha *et al.*, 2007] Prithwijit Guha, Amitabha Mukerjee, and K.S. Venkatesh. *Pattern Recognition Technologies and Applications: Recent Advances*. Idea Group Inc., 2007.

[Köhler, 2002] Christian Köhler. The occlusion calculus. In *Workshop on Cognitive Vision*, Zürich, Switzerland, September 2002.

[Randell *et al.*, 2001] David Randell, Mark Witkowski, and Murray Shanahan. From images to bodies: Modelling and exploiting spatial occlusion and motion parallax. In *International Joint Conference on Artificial Intelligence*, pages 57–63, 2001.

[Yilmaz *et al.*, 2006] Alper Yilmaz, Omar Javed, and Mubarak Shah. Object tracking: A survey. *ACM Journal of Computing Surveys*, 38(4), 2006.