Human Activity Understanding using Visibility Context

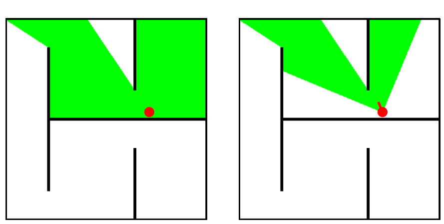
Vlad I. Morariu, V. Shiv Naga Prasad, and Larry S. Davis
Computer Vision Laboratory
University of Maryland
College Park, MD 20742
{morariu,shivnaga,lsd}@cs.umd.edu

***Abstract*— Visibility in architectural layouts affects human navigation, so a suitable representation of visibility context is useful in understanding human activity. Motivated by studies of spatial behavior, we use a set of features from visibility analysis to represent spatial context in the interpretation of human activity. An agent's goal, belief about the world, trajectory and visible layout are considered to be random variables that evolve with time during the agent's movement, and are modeled in a Bayesian framework. We design a search-based task in a sprite-world, and compare the results of our framework to those of human subject experiments. Our findings confirm that knowledge of spatial layout improves human interpretations of the trajectories (implying that visibility context is useful in this task). Since our framework demonstrates performance close to that of human subjects with knowledge of spatial layout, our findings confirm that our model makes adequate use of visibility context. In addition, the representation we use for visibility context allows our model to generalize well when presented with new scenes.**

I. Introduction

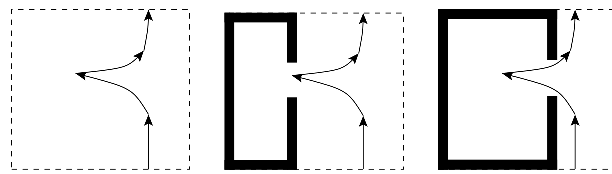
Visibility in architectural layouts affects human navigation, so a suitable representation of visibility context is useful in understanding human activity. Here, *visibility context* refers to a building's spatial layout visible to a human from various locations inside the building. See Figure 1 for an illustration. Numerous studies in psychology and architecture have underscored the significant influence of a building's layout on the manner in which people navigate through it and emotively perceive it, e.g., [6], [5]. People walk through different parts of a building depending upon its layout and their purpose (e.g. to search, hide, explore).

The context provided by spatial layout may significantly affect an observer's interpretation of an agent's trajectory. Consider a scenario in which a person is navigating through



(a) (b)

Fig. 1. Layout visibility with (a) omnidirectional and (b) directed view. The observer is denoted by the red circle, and the visible area – called an *isovist* – is the green polygon.



(a) (b) (c)

Fig. 2. Part of a person's trajectory (foot-print) while searching for an object. Depending on the spatial layout, the sharp turn may be interpreted as a search point or the location at which the subject picked up the object.

a building. The person's objective is to find and pick up an object and then place it at some location. The observer is only provided with the person's trajectory (foot-prints) on the floor plan and is assigned the task of inferring the person's actions, such as whether the person was still searching for the object at a particular time, had already located it, etc. Figure 2 shows a zoomed in portion of a hypothetical trajectory generated in this scenario. Figure 2(a) shows the trajectory in the absence of any walls. In the absence of other information, the sharp turn in the trajectory could reasonably be interpreted as the location at which the person picked up the object – the person must have deviated from an otherwise straight path for a reason. Figure 2(b) shows the same trajectory, but with walls superimposed on the image. It is now much less likely that the agent picked up an object at that point; instead, it appears more likely that the person walked to that point only to explore the closed room and then moved on after discovering that the room did not contain the object. Now consider Figure 2(c) – the same trajectory but with slightly different layout of the walls. In this case, the person walks deeper into the room. Now, it seems more likely that the person saw the object in the room and walked in to pick it up – there would be no other reason to walk into an empty room. Thus, the same trajectory can be interpreted very differently based on the visibility context! This is the principal intuition of our work – how to represent a layout's visibility context and employ it for understanding human activity.

Consider a person searching for an object in a building. Two aspects of layout visibility influence the person's movement:

1) *Vantage points:* The person would give preference to locations that provide views of large parts of the

building so that the search is efficient. The visibility context for the locations consists of features such as the field of view's area, perimeter, etc.
2) *Belief/memory:* While navigating through the building, the person builds a mental map of the areas already explored and those still to be investigated. A belief of the possible locations of the sought object is maintained and continuously updated with new information.

We present a Bayesian framework for jointly modeling the influence of visibility and belief on a person's movement. The person's goal, belief about the world, trajectory and visible layout are considered to be random variables that evolve with time during the movement. The belief/memory of the world and the visible layout constrain the person's goal. The belief and the goal together determine the sequence of actions taken by the person, which in turn determines the trajectory. Recognition is formulated as Maximum A Posteriori (MAP) estimation. The visibility context is represented with features based on behavioral studies of architecture. The features are designed to enable generalization over novel spatial layouts.

Behavioral studies of architecture indicate that people's navigation through a building (spatial behavior) is closely coupled with the layout that is visible to them from different locations within the building. For example, Kaynar proposed that the spatial behavior of humans (in particular, their paths) in a museum can be predicted by visibility analysis [6]. The study indicated that presence or anticipation of unseen areas near a person's location is correlated with the change in the person's movement direction. Wiener and Franz showed that spatial behavior and experience can be predicted using measures derived from visibility context [5], [10]. For instance, measures of spatial qualities such as spaciousness, openness, complexity and order had significant correlations with the building's ratings given by human subjects.

Our proposed approach also relates to recent work in the robotics literature, where visibility – represented by isovists – has been used for motion planning in tasks such as exploration of unknown environments [3] and tracking a target in an environment with occlusions [2]. In the former example, a robot approximates its isovist using line-of-sight sensors and moves toward isovist boundaries that lead to unseen regions – the "inverse" of the problem we aim to solve. Rather than using visibility for motion planning, we instead use visibility to provide context in which an agent's motion can be interpreted.

Recent studies on human activity recognition have highlighted the importance of context provided by the scene, objects, etc. Our framework is closely inspired by the work of Baker et al. on the "inverse planning" problem of determining the intentions of an agent from trajectories [1]. They propose a Bayesian model for the agent's intentions based on the trajectory and the spatial layout. The agent is assumed to always know the exact position of the object. They do not consider visibility – the spatial layout affects the analysis by constraining the possible movements. In our work, the person is searching for the object and therefore has to explore the layout. Moreover, only a part of the layout may be visible to the person at any given time. There are numerous computer vision studies in activity recognition that focus on trajectory-based features – we cite only a few, e.g., [7], [8]. These approaches do not model visibility context. In other studies, the scenes are manually pre-annotated to encode spatial and semantic information (e.g., doorways, hallways, furniture [9]). The proposed visibility context complements such approaches as it does not require explicit annotations, enabling generalization to novel scenes.

We illustrate the approach with experiments in a sprite-world domain. This isolates visibility and spatial layout as the only factors affecting an agent's actions. The trajectories are generated by a human performing search-based tasks in a virtual environment similar to first-person video game interfaces. We consider 6 layouts of varying complexity. To observe the generalization over layouts, the model is trained on 5 layouts and tested on the other, in a round robin format. As part of the experiments, human observers were asked to analyze the same trajectories with and without spatial information. Their scores are compared with that of the analysis performed by the Bayesian framework. The experiments show the importance of visibility context for activity recognition in our search-based task. Moreover, the proposed framework achieves average recognition rates comparable to those of the human observers.

The paper is organized as follows. In section II we describe previous work on activity understanding and on effects that spatial and visibility constraints have on behavior. In section IV we provide the visibility features that we use. In section III we discuss our model that incorporates visibility, memory, and belief. In section V we show the results of human experiments and of our proposed approach. Finally, we provide our concluding remarks in section VI.

II. Related Work

Baker et al. [1] propose a general Bayesian framework to explain how people reason and predict the actions of an intentional agent. They call their analysis of intentional reasoning "inverse planning" since they assume that agents build plans (sequences of actions) that achieve their goals, and to infer their intentions observers need only to invert a model of how goals affect plan formation. Using experimental results on pre-verbal infants from the cognitive science literature and their own experiments on humans, the authors motivate their Bayesian framework by noting that any model of intentional reasoning should include at least primitive planning capacities with the tendency to choose plans that achieve goals as efficiently as possible and that inferences about agents' goals should be probabilistic. In addition, motivated by how humans reason with the intentional stance, the authors introduce a utility function and assume that agents will prefer actions which lead to a larger expected increase in the utility function. In their design of the Bayesian framework, they place emphasis on the ability to learn from multiple environments and generalize to new ones.

In their application of the general Bayesian framework to sprite-world inferences, Baker et al. introduce the as-

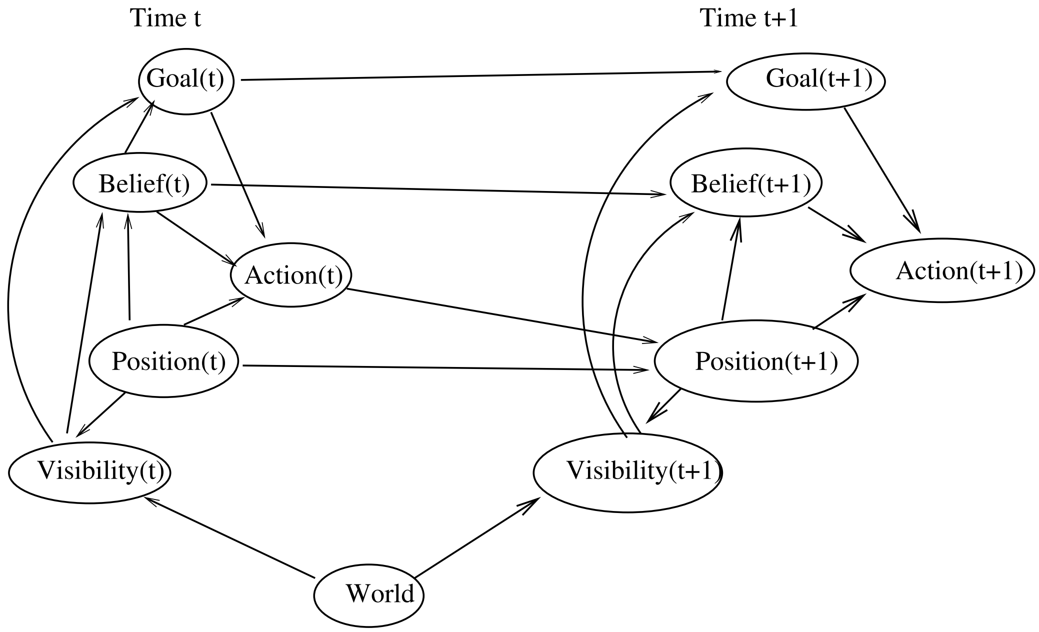


Fig. 3. Graphical model for spatial behavior.

sumptions that the world W is known to the agent and the observer (i.e. the agent knows the layout of the world and objects within it). However, agents often can only have partial observations of the world determined by what is visible from their current location, and at any point in time will know only the sections of the world that they have observed until that point. In this paper we will remove this assumption. Visibility constraints and memory are modeled in the Bayesian framework, eliminating the assumption that the agent has full knowledge of the world and incorporating spatial context into the model. Note that because spatial context is represented through visibility properties and is not represented directly by the environment, the model generalizes to new environments that have different spatial layouts.

III. Bayesian Model for Visibility in Activity Recognition

Consider an agent exploring and navigating in a world W. The agent's state at time t is defined to consist of three components:

1) Current goal, $g(t)$, for the movement. This controls the objectives guiding the agent's movements, e.g., searching for an object, approaching the object upon discovering it. An activity may in general consist of a sequence of goals, one leading to another.
2) The belief, $b(t)$, about the world. As the agent explores the world W, it continuously updates its belief about W based on the structure that is visible to it at any given time. The agent's belief consists of both memory as well as priors on the world's state.
3) The location of the agent in space, defined by $\mathbf{x}(t)$. The sequence of $\mathbf{x}(t)$'s forms the agent's trajectory. The location in space determines the substructure of the world that is visible to the agent. We denote the visible part of W by $v(t)$.

Based on its current goal, belief and position, an agent executes an action $a(t)$ with likelihood $p(a(t)|g(t), b(t), \mathbf{x}(t))$ to bring about a change in its location, $\mathbf{x}(t) \longrightarrow \mathbf{x}(t+1)$. The action's outcome is modeled with the conditional probability $p(\mathbf{x}(t+1)|a(t), \mathbf{x}(t))$. The change in location provides a novel view of the world $v(t+1)$ subject to the likelihood $p(v(t+1)|\mathbf{x}(t+1), W)$. This in turn results in an updated belief, $b(t+1)$ with probability $p(b(t+1)|v(t+1), b(t))$. The belief and current world view may lead to a change in the agent's goal, $g(t+1)$ with likelihood $p(g(t+1)|b(t+1), v(t+1), g(t))$. E.g., when the agent locates the sought object, the goal shifts from searching to that of approaching the object. The conditional probability structure is summarized in the graphical model in Figure 3. In practice, we only observe the agent's trajectory, $\mathbf{x}(t)$. All other components of the agent's state are hidden variables.

The layout's visibility is represented by the visibility-polygon, also called an *isovist*. The visibility-polygon is defined by the walls of the scene that are visible to the agent from a particular location in the world, and the occluded edges. Isovist qualities, such as area, perimeter and presence of occluded edges, determine whether a location is a good vantage point for searching (these qualities will be discussed briefly in section IV). Suppose, the agent's goal is to search for an object, and it anticipates some location to give a good view of a large part of the world, then the agent would likely navigate towards that point. Therefore, if an agent is observed to show preference to locations with high visibility, then it is assigned a high likelihood to a search goal. On the other hand, if an agent is observed to walk a direct path to a corner in a room then it has most likely located the sought object and is proceeding to pick it up.

IV. Modeling Visibility

Motivated by observations from cognitive science on visibility and architecture as discussed above, we represent visibility, $v(t)$, by using isovists and features derived from them. Figure 1(a) shows an isovist which we refer to as a *full isovist* since it includes all visible areas if viewing angle and field of view constraints are not taken into account. However, humans have a limited field of view; a *partial isovist* refers to an isovist that excludes all areas that are not in an oriented observer's field of view. Figure 1(b) shows a sample partial isovist. The edges of the isovist that do not coincide with a wall are referred to as *occluded edges*; they are formed when walls occlude an observer's view, and are often potential directions for further exploration. The isovist can be used directly for modeling what an observer sees along the trajectory, which facilitates the process of reasoning about the observer's belief of the world. However, features derived from the isovist contain additional information related to spatial layout that can further help observers reason about human behavior.

A variety of features can be computed from isovists (full and partial), many of which are discussed in [4], [5], [10]. In addition, visibility graphs (which can be used to compute shortest paths) are closely related to isovists, since each node in the visibility graph corresponds to a point in a scene and there is an edge in the graph between two nodes if they are visible (i.e. one node lies inside the other node's isovist). We calculate features using both isovists and visibility graphs.

The first group of features are derived from the isovist at the current location along the trajectory. They include isovist *area*, *perimeter*, *occlusivity* (sum of lengths of all isovist occluded edges divided by perimeter), *openness* (ratio of length of occluded edges to that of non-occluded edges), *compactness* (square of isovist perimeter divided by area), *minimum distance to an occluded edge* and *minimum distance to a wall*. The first five correspond to the spatial qualities of spaciousness, openness, and complexity. The last two contain information about the agent's current positioning relative to the layout. Consider *area* for an example of how these features can be helpful: as the agent moves from one room to another, isovist area peaks when the agent is in the doorway between the two rooms, which can be helpful if certain events are more or less likely to occur in doorways.

The second group of features uses isovists and limited path history (such as positions at time t and $t - \Delta$), which show how agents have changed their visibility fields over time. The most straightforward use of limited path history is to approximate derivatives of an isovist field along the path. However, there are useful features that are not simply approximated derivatives. Such features include *new view area*, *lost view area*, and *deviation from the shortest path*. The new view area is the area of the isovist region at time t that does not coincide with the intersection of the isovists at times t and $t - \Delta$ (lost view area is computed similarly). Deviation from the shortest path uses visibility graphs instead of isovists, and is the additional cost relative to the shortest path that an agent must incur to travel from a start to an end point through a middle point. If the difference is large, then the middle point is a significant detour from the shortest path, and the agent likely incurred the additional cost because there was some reward for deviating from the shortest path.

Features in the third group are based on the complete history of the trajectory (e.g. the union of all areas seen by time t), and include *area seen ratio* (total area seen divided by total layout area), and *geodesic distance to unseen regions*. *Geodesic* distance is the shortest distance after taking walls into account. This is useful because a rational agent who is exploring a region or searching for an object will tend to move toward some unexplored portion of the environment, thus causing the shortest distance to unseen regions to decrease.

Figure 4 shows how some features described above change as an agent performs a search-based task. The task in this case is to first search for and "pick up" a blue cube and then to search for and "pick up" a red cube, and is described in more detail in Section V.

V. Recognition

There are several human behaviors that are significantly influenced by the structure of scene layout, e.g., searching, hiding, stalking. We use searching activities as the domain for demonstrating the importance of visibility context and the proposed Bayesian framework. The trajectories were collected by a human agent navigating in a virtual 3D environment. The interface is similar to that commonly used in first-person video games. A virtual environment allowed accurate and precise observations of ground truth, and isolated the spatial visibility features to be the only factors influencing the movement. Six scenes were constructed, shown in Figure 5. They have distinctive spatial structure, varying from cubicles as seen in offices to aisles commonly occurring in superstores.

There are a number of possible tasks that can be assigned to the human agent to investigate search activities. These can range in complexity from a very simple task, e.g., "Search for and pinpoint a stationary object", to relatively complex tasks such as "Search for an object that is trying to evade you". There can also be variations such as the degree of background clutter, a sequence of search-and-locate subtasks, etc. We chose a search task of medium complexity:

1) At the start, the human agent is "teleported" to a random location in a scene. The task is to search out a blue cube placed randomly in scene, and go touch it. Then the agent must proceed to search for a red cube, also placed randomly, and touch it. There were no other objects except the blue and red cubes in the scene.
2) The recognition task for the observer is to estimate the location of the blue cube using just the trajectory of

the agent. The recognition task was posed to human subjects as well as to the proposed Bayesian framework.

The reasoning is that:

- As the blue cube is placed completely at random, the only distinguishing feature for its location would be the trajectory of the agent before and after touching it.
- As the agent is tasked to search for the red cube after touching the blue one, the observer is forced to distinguish between searching and non-searching behavior. This is a harder recognition task than the case in which the agent is instructed to either walk to a predefined place or move around randomly. In the latter case, the purposive search for the blue cube would be easily distinguishable.

For each of the six scenes, the human agent was asked to perform the search task 15 times, generating 90 trajectories in total. Each time, the blue and red cubes were placed randomly.

To isolate and highlight the importance of visibility context, half the human subjects were asked to perform the recognition without any information about the walls present in the scene, and other half were shown the trajectories with the walls correctly superimposed. The results indicate that human recognition performance improves substantially when the context of the surrounding is provided. The Bayesian framework was assigned the recognition task in the presence of visibility context. The results indicate that the approach's performance is comparable to that of humans.

A. Human Recognition Results

For the human subject experiments, 8 subjects were presented with all 90 trajectories. The subjects were split into two groups: one group was shown the walls and trajectory (the 'walls' group) and the other was shown only the trajectory (the 'no walls' group). The subjects were informed of the agent's task and were instructed to pick the location on the trajectory where the agent most likely picked up the blue cube. As figure 6 shows, the 'walls' group performed best, detecting 72.8% of the blue cube pick up events within 1.5 meters of the ground truth (scenes are either 28m by 28m or 40m by 28m; see figure 7 for a depiction of error relative to scene size). The 'no walls' group detected only 52.8% within the 1.5m error margin. Thus, the visibility and spatial context of the walls provides significant information to humans for inferring the intention of the agent.

The recognition performance showed significant variation w.r.t. the scenes - see Figure 6(b). Scenes 1, 2 and 3 have lower complexity of wall layout compared to Scenes 4, 5 and 6. Scenes 5 and 6 are especially difficult. The complexity in scene 5 arises from the fact that the room is divided into aisles, allowing the agent to walk directly through an aisle without returning after picking up an object in the aisle. This greatly decreases the performance of the humans with and without walls. The complexity of scene 6, however, arises from the number of small rooms that must be explored.

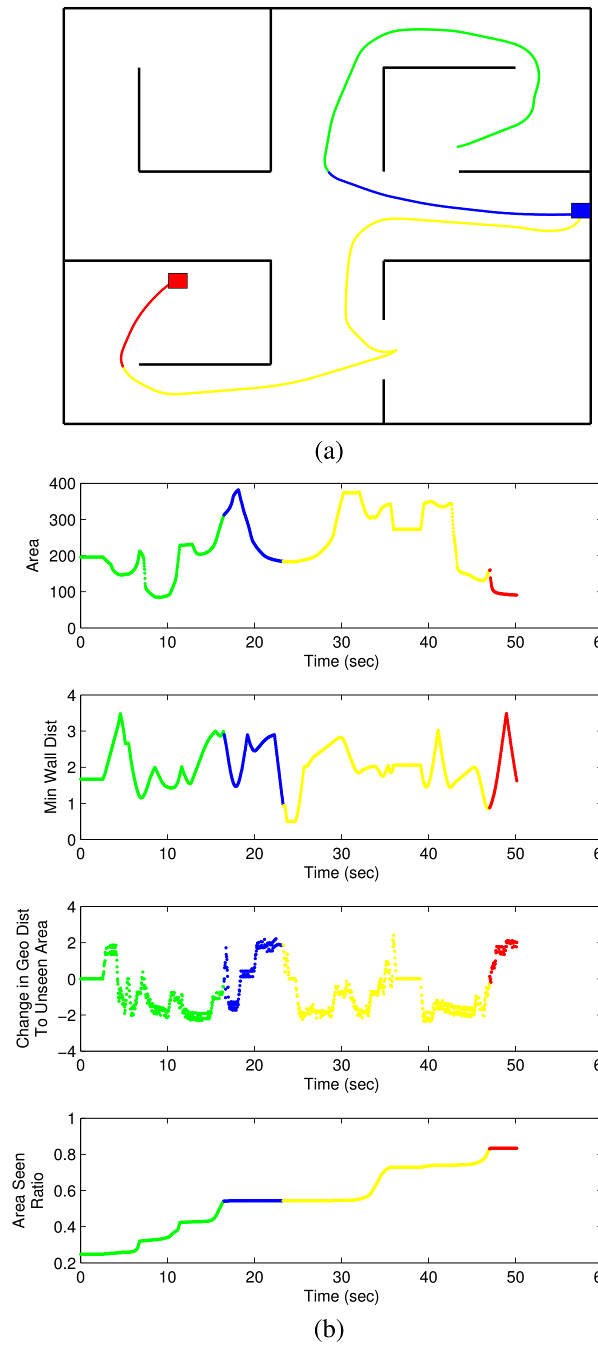


Fig. 4. Sample trajectory and corresponding visibility features. In part (a), trajectory is shown segmented by search sub-goals: green = "search for blue cube", blue = "pick up blue cube", yellow = "search for red cube", red = "pick up red cube". Part (b) shows visibility features during the trajectory as they vary with time. The coloring corresponds to trajectory coloring. Note that when the blue cube is reached, *isovist area* and *minimum distance to a wall* are close to a local minimum, the *change in geodesic distance to unseen area* is positive, and the *seen area ratio* is relatively flat.

Without layout information, subjects have no clue as to what caused all the turns in the trajectory. However, giving layout information to the subjects resulted in a much larger improvement in detection error for scene 6 compared to scene 5.

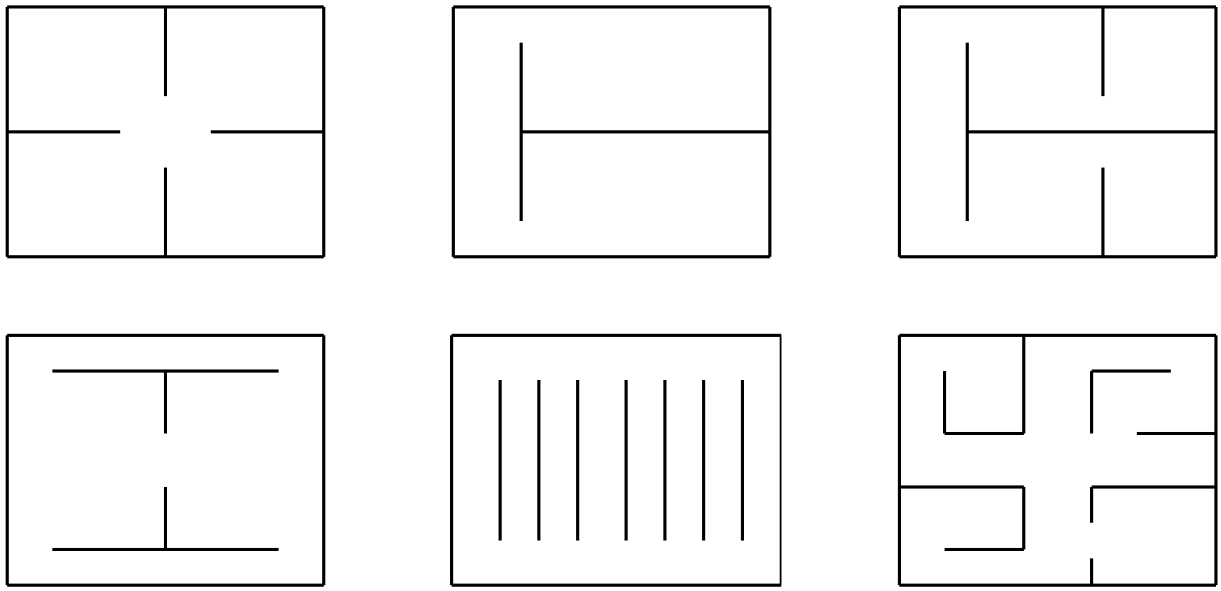


Fig. 5. Scenes. The top and bottom rows show scenes 1 to 3 and 4 to 6, respectively.

B. Recognition with the Bayes Framework

The recognition task is formulated as MAP estimation of the location of the blue cube given the human agent's trajectory and the scene. We compute a set of intermediate goal-points based on high-curvature locations in the person's trajectory. Each of these goal-points is considered to be a hypothesized location for the blue cube. The recognition result is the hypothesis with the highest likelihood, estimated using MAP. The joint likelihood of the agent's state, the sequence of executed actions and the world is

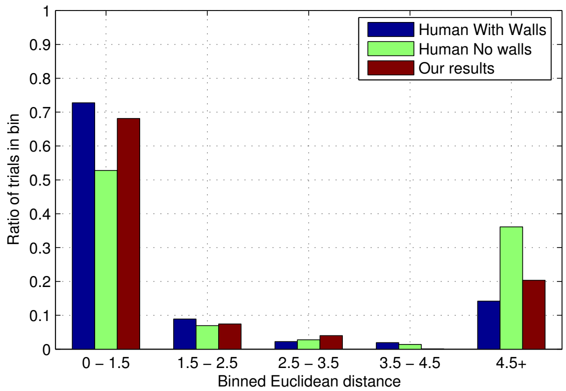
$$\begin{aligned} p(\mathbf{x}, b, g, a, W) \quad = \quad & \prod_{t=2}^{T} p(\mathbf{x}(t)|\mathbf{x}(t-1), a(t-1)) \\ & \prod_{t=1}^{T} p(v(t)|\mathbf{x}(t), W) \\ & \prod_{t=2}^{T} p(b(t)|v(t), b(t-1)) \\ & \prod_{t=2}^{T} p(g(t)|v(t), b(t), g(t-1) \\ & \prod_{t=1}^{T-1} p(a(t)|\mathbf{x}(t), b(t), g(t)) \qquad (1) \end{aligned}$$

The first 3 product-terms in the joint likelihood eq.(1) are determined from the trajectory, the scene's layout and the blue cubes hypothesized location. Thus, $p(\mathbf{x}, b, g, a, W)$, the confidence for the blue cube's location hypothesis is determined by $p(g(t)|v(t), b(t), g(t-1))$ and $p(a(t)|\mathbf{x}(t), b(t), g(t)) \propto \frac{p(g(t)|x(t),a(t),b(t))}{p(g(t)|x(t),b(t))}$ – the goodness of the goal sequence given the visibility and trajectory.

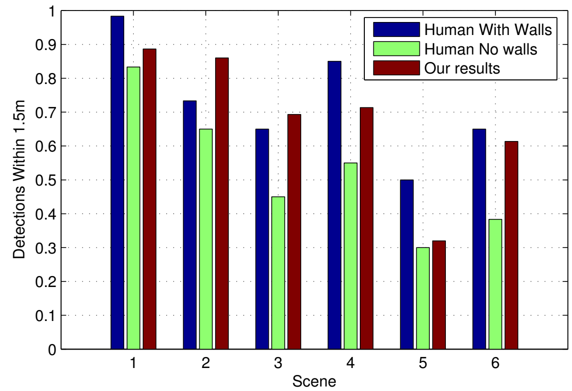
The scene and the blue cube's hypothesized location together determine the world W's state. Given the trajectory and W, the sequence of visible worlds $v(t)$ is computed. This, in turn, generates the sequences of beliefs, $b(t)$, of the human agent during the navigation (of course, conditioned on the hypothesized location). The beliefs and visible world together determine the sequence of goal states, $g(t)$, of the human agent:

- Until the time the person sights the blue cube, the goals, $g(t)$'s, can either be "search" - giving preference to high visibility areas, or "via-point" - that are just intermediate points to reach some other goal, e.g., to turn a corner.
- After sighting the blue cube and until the hypothesized time of touching the blue ball, the goal points must be "via-points". There is no searching required.
- After the hypothesized touching of the blue cube and until the sighting of the red cube, the goal points will either be "search" or "via-points".
- After sighting the red cube and until touching it (the end of the sequence), the goal points must be "via-points" because there is need for further search.

The likelihood for a goal-point, $g(t)$, to be a "search" is determined from the visibility field. Specifically, it is determined by the newly seen area. The likelihood for a goal-point, $g(t)$, to be a "via-point" is determined by the *deviation from shortest path*. Combining the log-likelihoods of these goal-points gives the likelihood of the sequence of goals before and after the hypothesized pickup of blue cube, denoted by l. This must be combined with the likelihood of the goal-point at the blue cube location. A boosting algorithm is employed to classify correctly hypothesized blue cube locations from incorrect ones based on l and $v(t_b)$, the



(a)



(b)

Fig. 6. Experimental results: (a) Histogram of error between experimental results and ground truth, showing the proportion of detections that lie in each error range (in meters), and (b) ratio of detections with less than 1.5m error, grouped by scene.

visibility fields at the blue cube location. During training, the negative class contains sequences of goals generated by blue cube locations that are known to be incorrect, and the positive class contains the sequences that are known to be correct. Thus, the classifier is trained to recognize the MAP hypothesis using the visibility features at the blue cube location and the likelihood of the goals at all other times. During testing, the most likely hypothesis is defined to be the one with maximum confidence.

Note that we avoid the complex task of Bayesian inference on the proposed network by brute search through the hypothesis space. The number of hypothesized points (less than 50 in our experiments) was significantly smaller than the total number of points in the trajectory.

We tested the proposed approach on the same 90 trajectories described above. Unlike the human subjects (who did not require training scenes!), our algorithm was trained on 5 scenes, and tested on the remaining scene following a round robin protocol. As in the human experiments, the algorithm computed the likely location of the blue cube in each scene. Since the algorithm is blind to the test scene, the experiments test the generalization of the framework to novel scenes. Figure 6(a) shows that the results of our algorithm are very good (68.1% of detections are within 1.5m of the ground truth). Figure 6(b) shows that our algorithm performed well for all scenes (always better than the 'no walls' group), and better than the average of the 'walls' human subjects group for scenes 2 and 3!

Figure 7 shows example detections by our algorithm (red circle), 'no walls' human group (blue cross), and 'walls' group (green x). In the first row, the detections by our algorithm and by the 'walls' human are correct, but the 'no walls' human chose the wrong sharp point in the trajectory since the subjects were not provided with spatial information. In the absence of spatial information, the locations chosen by the 'no walls' humans are reasonable choices. The second row shows cases where all three groups were able to locate the blue cube location. Finally, the third row shows examples of where our algorithm failed in locating the blue cube. In the two leftmost cases, the scene is difficult to interpret because of the aisles. In the rightmost image of the third row, the point that our algorithm chose could be mistaken for the blue cube position since agent did not immediately leave the room after reaching the entrance, but instead entered the room slightly before turning around.

VI. Conclusion

We have presented a framework in which visibility context is utilized to aid in reasoning about human activity. Our experiments showed that features used to represent visibility generalize well over new scenes, and that our method resulted in a detection rate close to that of human observers in a search-based task. Future work on this framework may include applying it to more complex behavior, including joint activities of multiple agents (eg. hide-and-seek), and evaluating more precisely the generalization properties of the features by generating scenes at known 'distances' from training scenes to observe performance as a function of scene 'distance'.

Acknowledgments

We would like to thank the U.S. Government VACE program for supporting this work. Also, we would like to thank Aniruddha Kembhavi, Ryan Farrell, and other members of computer vision group for meaningful discussions and help with experiments.

References

[1] Chris Baker, Joshua B. Tenenbaum, and Rebecca Saxe. Bayesian models of human action understanding. In *NIPS*, 2005.

[2] T. Bandyopadhyay, Yuanping Li, M.H. Ang, and David Hsu. A greedy strategy for tracking a locally predictable target among obstacles. In *ICRA*, pages 2342– 2347, 2006.

[3] Tirthankar Bandyopadhyay, Zheng Liu, Marcelo H. Ang, and Windston Khoon Guan Seah. Visibility-based exploration in unknown environment containing structured obstacles. In *ICAR*, pages 484–491, 2005.

[4] Larry S. Davis and Michael L. Benedikt. Computational models of space: Isovists and isovist fields. *Computer Graphics and Image Processing*, 11(1):49–72, 1979.

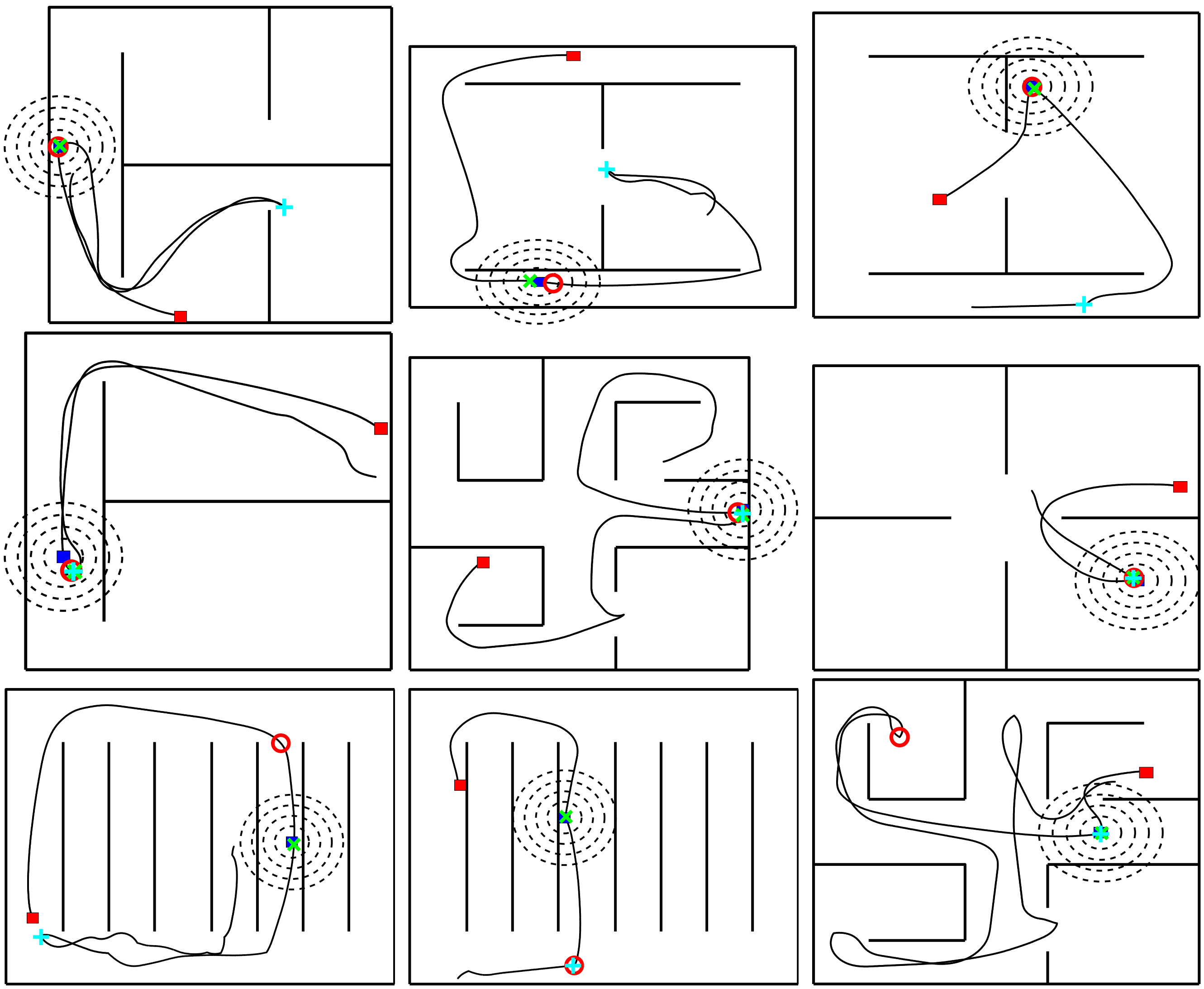


Fig. 7. Sample trajectories with ground truth, human, and algorithm detections (red circle - our algorithm, cyan plus - human no walls, green x - human with walls, blue square - blue cube ground truth). Dotted circles denote errors of 1.5m, 2.5m, 3.5m, and 4.5m from ground truth. The top row shows cases in which humans with no walls chose the wrong solutions, while our algorithm and the humans with walls were able to select the correct blue cube location. The middle row shows cases where the locations were correctly chosen by both groups of humans and our algorithm. Finally, the third row shows cases in which our algorithm failed to locate the blue cube.

[5] G. Franz and J. M. Wiener. Exploring isovist-based correlates of spatial behavior and experience. In *Proceeding of the 5th Space Syntax Symposium*, pages 503–517, 2005.

[6] Ipek Kaynar. Visibility, movement paths and preferences in open plan museums: An observational and descriptive study of the ann arbor hands-on museum. In *Proceeding of the 5th Space Syntax Symposium*, 2005.

[7] P. Ribeiro and J. Santos-Victor. Human activities recognition from video: modeling, feature selection and classification architecture. In *HAREM 2005 - in conjunction with BMVC 2005*, pages 61–70, 2005.

[8] Neil Robertson and Ian Reid. A general method for human activity recognition in video. *Comput. Vis. Image Underst.*, 104(2):232–248, 2006.

[9] V.D. Shet, D. Harwood, and L.S. Davis. Vidmap: Video monitoring of activity with prolog. In *IEEE International Conference on Advanced Video and Signal based Surveillance (AVSS)*, pages 224–229, 2005.

[10] J. M. Wiener and G. Franz. Isovists as a means to predict spatial experience and behavior. *Lecture notes in artificial intelligence*, 3343:42–57, 02 2005.