Axiomatic Support for Event Perception

Jeffrey Mark Siskind*
Department of Computer Science
University of Toronto
Toronto Ontario M5S 1A4 CANADA
416/978–6114
internet: Qobi@CS.Toronto.EDU

Abstract

This paper suggests that the notions of support, contact, and attachment play a central role in specifying the truth conditions for occurrences of events described by many simple spatial motion verbs. It then describes a novel implemented method for recovering the changing support, contact, and attachment relationships between objects depicted in animated line drawings. Central to this method is the ability to efficiently determine the stability of a collection of line segments under various hypothesized ontological assumptions via a reduction to a linear programming problem.

1 Introduction

Numerous researchers (cf. Leech 1969, Miller 1972, Schank 1973, Jackendoff 1983, Pinker 1989) have long realized the role played by the causal, aspectual, and directional qualities of motion in specifying the meanings of simple spatial motion verbs. For example, part of what it means to *throw* something to someone is to cause an object to begin to move towards that person. Similarly, part of what it means to *pick* something *up* is to cause an object to begin upward motion. Researchers (cf. Herskovits 1986, Jackendoff and Landau 1991) have also realized the role played by notions such as support, contact, and attachment in specifying the meanings of spatial prepositions. For example, in some situations, part of what it means for something to be *on* something else is for one object to be in contact with, and supported by, another object. In other situations, something can be on something else by way of attachment, as in *the knob on the door*.

It is rarely noticed, however, that the notions of support, contact, and attachment also play a central role in specifying the truth conditions of many spatial motion verbs. I have argued extensively elsewhere (Siskind 1992) that this is the case. For example, causing an object to move to someone is insufficient evidence for a throwing event. That definition admits rolling and sliding events as well. Part of what it means to *throw* something is for an object to be in unsupported motion after it leaves the thrower's hand. Similarly, not all causation of upward motion constitutes picking something up. One can cause a ball to move upward by kicking it without picking it up. Part of what it means for an agent to *pick* something *up* is to change its source of support. An object previously was supported by something other than the agent's hand. Now that object is supported by virtue of being in contact with, and attached to, the agent's hand.

For several years I have been building a system called ABIGAIL (Siskind 1991, 1992, 1993) that uses these notions of support, contact, and attachment as the basis for grounding language in perception. This system is similar in intention to work done by Badler (1975), Okada (1979), Borchardt (1984), Hays (1989), Regier (1992), and others. ABIGAIL watches animated line drawings and produces semantic descriptions of the events that occur in those movies. Figure 1 depicts the key frames of one such movie presented to ABIGAIL. Given just the coordinates of the endpoints of the line segments constituting each frame of this movie as input, ABIGAIL can detect the lifting, throwing, dropping, falling, bouncing, and putting down events that occur.

My prior work has used a form of counterfactual analysis to determine support relationships. Previous versions of ABIGAIL have contained a kinematic simulator capable of projecting the short term future of an image under the force of gravity. Objects were determined to be supported if they did not fall in the imagined future of an image. Similarly, an object A was determined to support an object B if B ceased to be supported when A was removed.

This paper presents an alternate mechanism for determining the support, contact, and attachment relationships between objects that does not involve simulation. Instead of simulation, a set of axioms is proposed that constrain the possible interpretations of an image. These axioms reduce to a system of linear equalities and inequalities. One can determine whether the objects in

*Supported in part by ARO grant DAAL 03–89–C–0031, by DARPA grant N00014–90–J–1863, by NSF grant IRI 90–16592, by Ben Franklin grant 91S.3078C–1, and by the Canadian Natural Sciences and Engineering Research Council.

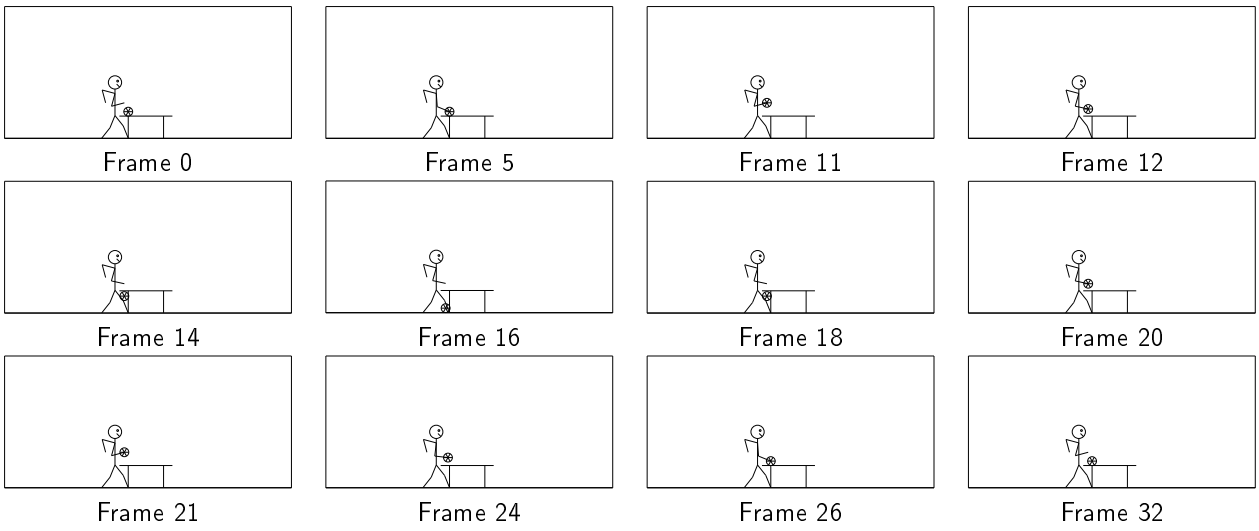

Figure 1: Several key frames from a typical movie presented as input to ABIGAIL.

an image are stable simply by determining whether or not this system of constraints is satisfiable. This is similar to the approach taken by Blum et al. (1970) and Fahlman (1974). Support, contact, and attachment relationships can be derived from the stability, or relative stability, of objects. This new approach promises to yield significantly faster event perception performance.

The remainder of the paper is outlined as follows. Section 2 describes the ontology that ABIGAIL uses to interpret images. Section 3 presents the axioms that encode this ontology. Section 4 shows how these axioms can reduce to a linear programming problem and how support, contact, and attachment relations can be determined by solving this problem. Section 5 shows how some simple event types can be recovered from these changing support, contact, and attachment relations. Section 6 concludes with a discussion of some related work.

2 Ontology

The input to ABIGAIL consists of a sequence of movie frames, each being a set of line segments.[1] These line segments are specified via the coordinates of their endpoints. In the current implementation, all movie frames have the same number of line segments. Furthermore, there is a one-to-one correspondence between line segments in adjacent movie frames, and ABIGAIL is given this correspondence as input. This implies that there is no occlusion, that line segments do not break into pieces or fuse together, that line segments do not enter or leave the field of view, and that line segments are neither created nor destroyed.

Beyond this input, ABIGAIL projects an ontology on the input movie. Informally, this ontology consists of the following five principles:

rigidity All line segments have fixed length. They do not shrink or stretch.

gravity Unsupported line segments fall.

ground plane Line segments cannot pass through the ground.

substantiality Line segments cannot pass through one another.

attachment Line segments can be attached together by joints that can optionally restrict their relative motion independently along the three possible degrees of freedom.

Collectively, the principles of rigidity, ground plane, substantiality, and attachment can offer the support necessary to counteract the principle of gravity.

Nominally, the images input to ABIGAIL are two-dimensional. Numerous substantiality violations would occur, however, if the underlying world were in fact two-dimensional. For instance, in the movie shown in figure 1, the ball would pass through the table. Thus ABIGAIL's ontology allows her to project a third dimension onto each image. This third dimension, however, is impoverished. It consists solely of a binary relation $\bowtie$ between line segments that specifies whether or not two line segments are on the same 'layer.' This relation is termed the

[1] Prior versions of ABIGAIL (Siskind 1992, 1993) allowed frames to contain circles as well as line segments. The version described here simulates circles as regular polygons. The only ontological capacity that is lost is the ability for circles to roll. So far, this capacity has not been used in ABIGAIL. There is one further subtle point however. In order for a circle that is simulated as a polygon to be stable when it rests on a horizontal surface, the polygon must contact that surface at two points rather than at a single point.

layer model. The substantiality principle applies only to line segments on the same layer.

Note that the ontological notion of 'layer' used here is very weak. The layer model is required only to be an equivalence relation. There is no further ordering or adjacency relation between layers. Thus there is no notion of a line segment being in front of or behind another line segment, and it is not possible for line segments on different layers to be in contact by virtue of being on adjacent layers.

ABIGAIL is not given the layer model as input. She must find a 'good' layer model that is consistent with the input under the above five ontological principles. Furthermore, this layer model can change from frame to frame. For example, at the beginning and end of the movie shown in figure 1, the surface of the ball might be on the same layer as the table top, to explain why the ball doesn't fall. Yet in the middle of that movie, they must be on different layers since the ball passes from above the table to the floor below and back.

In a similar fashion, the attachment relationships between line segments are specified by a *joint model*. This joint model consists of three binary relations $\leftrightarrow$, δ, and θ. The formula $f \leftrightarrow g$ specifies that f is attached to g. Line segments that are attached must intersect. If f is attached to g then that joint can be independently rigid or flexible along each of three degrees of freedom. The formula $\delta(f,g)$ specifies that the relative displacement of f along g is fixed, i.e. that the point of intersection between f and g cannot move along f. Thus the formulas $\delta(f,g)$ and $\delta(g,f)$ together specify the rigidity or flexibility of the joint along two degrees of freedom. The formula $\theta(f,g)$ specifies the rigidity or flexibility of the third degree of freedom by constraining whether or not f and g must maintain the same relative orientation. Collectively, the three relations $\leftrightarrow$, δ, and θ constitute the joint model.

Just as for the layer model, ABIGAIL is not given the joint model as input. She must find a 'good' joint model that is consistent with the input and with the layer model. Like the layer model, the joint model can also change from frame to frame. For example, both at the beginning and at the end of the movie shown in figure 1, there can be no connection between the man's hand and the ball as they do not intersect. Yet when the man picks up the ball in frame 6, ABIGAIL will hypothesize the formation of an attachment relationship between the hand and the ball to explain why the ball does not fall as it is picked up.

The details of the process by which ABIGAIL hypothesizes and updates the layer and joints models is beyond the scope of this paper. The reader is referred to Siskind (1992) for a description of one method for accomplishing this task.

3 Kinematic Axioms

Within the framework of ABIGAIL's ontology it is possible to formalize the principles of rigidity, gravity, ground plane, substantiality, and attachment that were discussed informally in section 2 as a set of axioms. These axioms will apply independently to each frame in the movie. In these axioms, the symbols f, g, and h will denote line segments, and the symbol F will denote the set of all line segments in the current movie frame. In the axioms, all free occurrences of f, g, and h are assumed to be universally quantified over all singletons, pairs, and triples of distinct line segments in a given movie frame.

The symbols p, q, and r will denote points. Points will be treated as two-component vectors. I will use the terms $x(p)$ and $y(p)$ to denote the two coordinates of p, and the terms $p(f)$ and $q(f)$ to denote the two endpoints of f. The predicate $\textsc{Intersect}(f,g)$ will be true if and only if f and g intersect. If f and g intersect, I will denote their point of intersection by $\textsc{Intersection}(f,g)$. The predicate $\textsc{Touch}(p,f)$ will be true if and only if p lies on f. The predicate $\textsc{Overlap}(f,g)$ will be true if and only if f and g intersect and the point of intersection is not an endpoint of either f or g.

I will use $\overline{p}$ to denote the *conjugate* of p, the vector derived by rotating p counterclockwise 90°. More specifically, $\overline{(x,y)} = (-y,x)$. I will use $|p|$ to denote the magnitude of p. More specifically, $|p| = \sqrt{p \cdot p}$. I will use $\hat{p}$ to denote a unit vector whose orientation is the same as p. More specifically, $\hat{p} = p/|p|$. I will consider a line segment to be oriented, pointing from p to q. I will use $\vec{f}$ to denote the vector whose orientation is the same as f and whose magnitude is the same as the length of f. More specifically, $\vec{f} = q(f) - p(f)$.

If we define $U(q,f) = \textbf{signum}(\overline{\vec{f}} \cdot [q - p(f)])$ then $U(q,f)$ will be equal to $+1$ if a vector rooted at $p(f)$, pointing in a direction rotated 90° counterclockwise from the direction of f, points towards q. It will be equal to -1 if that vector points away from q and equal to 0 if q is on the line obtained by extending f. I will use $\textsc{Component}(p,f,q)$ to denote the component of p perpendicular to f in the direction away from q. Using the above, $\textsc{Component}(p,f,q) = U(q,f)p \cdot \hat{\overline{\vec{f}}}$.

The endpoint coordinates $p(f)$ and $q(f)$ for each $f \in F$, as well as the binary relations $\bowtie$, $\leftrightarrow$, δ, and θ will vary from frame to frame. While each movie frame nominally constitutes a static motionless image, the following axiomatization hypothesizes an instantaneous motion for each line segment in each frame. This motion is represented by associating with each line segment an instantaneous angular velocity $\dot{\theta}(f)$, and two vectors $\dot{p}(f)$ and $\dot{q}(f)$ denoting the instantaneous velocity of its endpoints $p(f)$ and $q(f)$ respectively.

3.1 Rigidity

In ABIGAIL's ontology line segments have fixed length. They cannot shrink or stretch. If $p(f)$ were to be fixed, and $q(f)$ were to rotate about $p(f)$ at a fixed distance, $\hat{\vec{f}}$ would be a unit vector indicating the instantaneous direction of motion of $q(f)$. In this case, $\dot{\theta}(f)|\vec{f}|\hat{\vec{f}}$ would be the tangential component of the velocity vector $\dot{q}(f)$ derived by rotating $q(f)$ about $p(f)$, at the fixed distance $|\vec{f}|$ from $p(f)$, with angular velocity $\dot{\theta}(f)$. Adding in the component of the motion of $q(f)$ due to the motion of $p(f)$ leads to the following axiom:

$$\dot{q}(f) = \dot{p}(f) + \dot{\theta}(f)|\vec{f}|\hat{\vec{f}} \tag{1}$$

3.2 Gravity

ABIGAIL's ontology is purely kinematic. Objects have no velocity, acceleration, momentum, or kinetic energy. Furthermore, gravity is the only force that can act upon objects. Under these assumptions, the potential energy of a system cannot increase. ABIGAIL adopts the assumption that the mass of a line segment is uniformly distributed along the length of that line segment. This implies that the potential energy of a line segment is the product of its mass and the height of its midpoint. ABIGAIL adopts the further assumption that all line segments have the same density. Thus the length of a line segment can be taken to be its mass. This leads to the following axiom:

$$\sum_{f \in F} |\vec{f}|\, y\left(\frac{\dot{p}(f) + \dot{q}(f)}{2}\right) \leq 0 \tag{2}$$

3.3 Ground Plane

Objects cannot pass through the ground. In ABIGAIL's ontology this means that the vertical component of the velocity vector associated with an endpoint that is in contact with the ground cannot be negative. This leads to the following two axioms:

$$y(p(f)) = 0 \rightarrow y(\dot{p}(f)) \geq 0 \tag{3}$$

$$y(q(f)) = 0 \rightarrow y(\dot{q}(f)) \geq 0 \tag{4}$$

3.4 Substantiality

The substantiality constraint states that objects cannot pass through one another. In ABIGAIL's ontology this means that line segments that are on the same layer cannot overlap. The $\bowtie$ relation specifies whether or not two line segments are on the same layer. By definition, this is an equivalence relation and thus must be symmetric and transitive. This leads to the following two axioms:[2]

$$f \bowtie g \rightarrow g \bowtie f \tag{5}$$

[2] An axiom of reflexivity is not needed since, as stated earlier, all axioms with two or three free variables are instantiated only for pairs or triples of *distinct* line segments.

$$(f \bowtie g \wedge g \bowtie h) \rightarrow f \bowtie h \tag{6}$$

A substantiality violation can occur either statically or dynamically. Statically, observing one line segment overlap another gives evidence that they cannot be on the same layer. This leads to the following axiom:

$$f \bowtie g \rightarrow \neg\textsc{Overlap}(f, g) \tag{7}$$

Dynamically, at a given instant, a line segment f can begin to overlap another line segment g if and only if an endpoint of f touches g and the relative motion of f and g is such that they will immediately begin to overlap. There are two cases to consider, since this condition can hold for either endpoint of f. First consider the case where $p(f)$ touches g. Let r be the point on g coincident with $p(f)$ and let $\dot{r}$ be the vector denoting the instantaneous velocity of r due to the motion of g. In this case $r = p(f)$ and $\dot{r} = \dot{p}(g) + \dot{\theta}(g)\overline{r - p(g)}$. The component of $\dot{p}(f)$ perpendicular to g in the direction away from $q(f)$ must not be greater than the component of $\dot{r}$ in the same direction, or else f would immediately begin to overlap g. This leads to the following axiom:

$$\begin{array}{l}(f \bowtie g \wedge \textsc{Touch}(p(f), g)) \rightarrow \\ \textsc{Component}(\dot{p}(f), g, q(f)) \leq \\ \textsc{Component}(\dot{p}(g) + \dot{\theta}(g)\overline{p(f) - p(g)}, g, q(f))\end{array} \tag{8}$$

Now consider the case where the other endpoint $q(f)$ touches g. Let r be the point on g coincident with $q(f)$ and let $\dot{r}$ be the vector denoting the instantaneous velocity of r due to the motion of g. In this case $r = q(f)$ and $\dot{r} = \dot{p}(g) + \dot{\theta}(g)\overline{r - p(g)}$. The component of $\dot{q}(f)$ perpendicular to g in the direction away from $p(f)$ must not be greater than the component of $\dot{r}$ in the same direction, or else f would immediately begin to overlap g. This leads to the following axiom:

$$\begin{array}{l}(f \bowtie g \wedge \textsc{Touch}(q(f), g)) \rightarrow \\ \textsc{Component}(\dot{q}(f), g, p(f)) \leq \\ \textsc{Component}(\dot{p}(g) + \dot{\theta}(g)\overline{q(f) - p(g)}, g, p(f))\end{array} \tag{9}$$

3.5 Attachment

The joint model is collectively specified by the three binary relations $\leftrightarrow$, δ, and θ. There are four axioms that apply solely between these three relations. First, the $\leftrightarrow$ relation must be symmetric.

$$f \leftrightarrow g \rightarrow g \leftrightarrow f \tag{10}$$

Second, the relative displacement of two line segments can be restricted only if they are attached.

$$\delta(f, g) \rightarrow f \leftrightarrow g \tag{11}$$

Similarly, the relative rotation of two line segments can be restricted only if they are attached.

$$\theta(f, g) \rightarrow f \leftrightarrow g \tag{12}$$

Finally, the θ relation must be symmetric.

$$\theta(f,g) \rightarrow \theta(g,f) \tag{13}$$

Two attached line segments must intersect. If they do not, then the attachment constraint between them would be violated. Like substantiality, an attachment constraint can be violated either statically or dynamically. Statically, observing two nonintersecting line segments gives evidence that they cannot be attached. This leads to the following axiom:

$$f \leftrightarrow g \rightarrow \textsc{Intersect}(f,g) \tag{14}$$

Dynamically, at a given instant, a line segment f can cease to intersect another line segment g if and only if an endpoint of f touches g and the relative motion of f and g is such that they will immediately cease to intersect. There are two cases to consider, since this condition can hold for either endpoint of f. First consider the case where $p(f)$ touches g. Let r be the point on g coincident with $p(f)$ and let $\dot{r}$ be the vector denoting the instantaneous velocity of r due to the motion of g. In this case $r = p(f)$ and $\dot{r} = \dot{p}(g) + \dot{\theta}(g)\overline{r - p(g)}$. The component of $\dot{p}(f)$ perpendicular to g in the direction away from $q(f)$ must not be less than the component of $\dot{r}$ in the same direction, or else f and g would immediately cease to intersect. This leads to the following axiom:

$$\begin{array}{l}(f \leftrightarrow g \wedge \textsc{Touch}(p(f),g)) \rightarrow \\ \textsc{Component}(\dot{p}(f),g,q(f)) \geq \\ \textsc{Component}(\dot{p}(g) + \dot{\theta}(g)\overline{p(f) - p(g)},g,q(f))\end{array} \tag{15}$$

Now consider the case where the other endpoint $q(f)$ touches g. Let r be the point on g coincident with $q(f)$ and let $\dot{r}$ be the vector denoting the instantaneous velocity of r due to the motion of g. In this case $r = q(f)$ and $\dot{r} = \dot{p}(g) + \dot{\theta}(g)\overline{r - p(g)}$. The component of $\dot{q}(f)$ perpendicular to g in the direction away from $p(f)$ must not be less than the component of $\dot{r}$ in the same direction, or else f and g would immediately cease to intersect. This leads to the following axiom:

$$\begin{array}{l}(f \leftrightarrow g \wedge \textsc{Touch}(q(f),g)) \rightarrow \\ \textsc{Component}(\dot{q}(f),g,p(f)) \geq \\ \textsc{Component}(\dot{p}(g) + \dot{\theta}(g)\overline{q(f) - p(g)},g,p(f))\end{array} \tag{16}$$

The above attachment axioms enforce only the constraint that attached line segments intersect. They do not enforce the potential rigidity of the relative displacement or rotation of two attached line segments. The following analysis can be used to derive the axiom for constraining the relative displacement of f along g. Let r be the point of intersection of f and g. Let $\dot{r}_f$ be a vector denoting the instantaneous velocity of r, taken as a point on f, due to the motion of f. Let $\dot{r}_g$ be a vector denoting the instantaneous velocity of r, taken as a point on g, due to the motion of g. Now, $r = \textsc{Intersection}(f,g)$, $\dot{r}_f = \dot{p}(f) + \dot{\theta}(f)|r - p(f)|\hat{\vec{f}}$ and $\dot{r}_g = \dot{p}(g) + \dot{\theta}(g)|r - p(g)|\hat{\vec{g}}$. If the relative displacement of f along g is fixed then the component of $\dot{r}_f$ along g must be equal to the component of $\dot{r}_g$ along g. This leads to the following axiom:

$$\begin{array}{l}\delta(f,g) \rightarrow \\ (\dot{p}(f) + \dot{\theta}(f)|\textsc{Intersection}(f,g) - p(f)|\hat{\vec{f}}) \cdot \hat{\vec{g}} = \\ (\dot{p}(g) + \dot{\theta}(g)|\textsc{Intersection}(f,g) - p(g)|\hat{\vec{g}}) \cdot \hat{\vec{g}}\end{array} \tag{17}$$

Finally, the relative rotation of two line segments is constrained by the following axiom:

$$\theta(f,g) \rightarrow \dot{\theta}(f) = \dot{\theta}(g) \tag{18}$$

4 Reduction to Linear Programming

Axioms 5, 6, 10, 11, 12, and 13 serve only to enforce the self consistency of the layer and joint models, while axioms 7 and 14 serve only to enforce the consistency between the layer and joint models and the current movie frame. For any given frame, the line segments will all be in fixed positions so the values $x(p(f))$, $y(p(f))$, $x(q(f))$, and $y(q(f))$ will all be constant for each line segment f. For a given frame and particular layer and joint models, the remaining axioms all reduce to linear equations and inequalities whose unknowns are the hypothesized angular velocities and endpoint velocity vectors. Solutions to these equations and inequalities constitute potential instantaneous motions of the line segments in the image that are consistent with given layer and joint models and with the ontology.

The particular constraints that result are as follows. First, let us define the following constants for each pair of line segments f and g:

$$(c_1,c_2) = \overline{\vec{f}} \tag{19}$$

$$(c_3,c_4) = \overline{\vec{g}} \tag{20}$$

$$(c_5,c_6) = \overline{p(f) - p(g)} \tag{21}$$

$$(c_7,c_8) = \overline{q(f) - p(g)} \tag{22}$$

$$c_9 = U(p(f),g) \tag{23}$$

$$c_{10} = U(q(f),g) \tag{24}$$

$$(c_{11},c_{12}) = |\textsc{Intersection}(f,g) - p(f)|\hat{\vec{f}} \tag{25}$$

$$(c_{13},c_{14}) = |\textsc{Intersection}(f,g) - p(g)|\hat{\vec{g}} \tag{26}$$

Given these constants, axiom 1 becomes the following two equations:

$$x(\dot{p}(f)) + c_1\dot{\theta}(f) - x(\dot{q}(f)) = 0 \tag{27}$$

$$y(\dot{p}(f)) + c_2\dot{\theta}(f) - y(\dot{q}(f)) = 0 \tag{28}$$

Axiom 2 is already a linear inequality in its current form. For a given image, axioms 3 and 4 become the following

two inequalities that are optionally instantiated depending upon the values of $y(p(f))$ and $y(q(f))$:

$$y(\dot{p}(f)) \geq 0 \quad (29)$$
$$y(\dot{q}(f)) \geq 0 \quad (30)$$

For a given image and layer model, axiom 8 becomes the following inequality that is instantiated for any pair of line segments f and g on the same layer, where $p(f)$ touches g:

$$\begin{aligned} -c_{10}c_3x(\dot{p}(f)) - c_{10}c_4y(\dot{p}(f)) \\ +c_{10}c_3x(\dot{p}(g)) + c_{10}c_4y(\dot{p}(g)) \\ +(c_{10}c_3c_5 + c_{10}c_4c_6)\dot{\theta}(g) \geq 0 \end{aligned} \quad (31)$$

Similarly, axiom 9 becomes the following inequality that is instantiated for any pair of line segments f and g on the same layer, where $q(f)$ touches g:

$$\begin{aligned} -c_9c_3x(\dot{p}(f)) - c_9c_4y(\dot{p}(f)) \\ +c_9c_3x(\dot{p}(g)) + c_9c_4y(\dot{p}(g)) \\ +(c_9c_3c_7 + c_9c_4c_8)\dot{\theta}(g) \geq 0 \end{aligned} \quad (32)$$

In an analogous fashion, axiom 15 becomes the following inequality that is instantiated for any pair of attached line segments f and g, where $p(f)$ touches g:

$$\begin{aligned} +c_{10}c_3x(\dot{p}(f)) + c_{10}c_4y(\dot{p}(f)) \\ -c_{10}c_3x(\dot{p}(g)) - c_{10}c_4y(\dot{p}(g)) \\ -(c_{10}c_3c_5 + c_{10}c_4c_6)\dot{\theta}(g) \geq 0 \end{aligned} \quad (33)$$

Likewise, axiom 16 becomes the following inequality that is instantiated for any pair of attached line segments f and g, where $q(f)$ touches g:

$$\begin{aligned} +c_9c_3x(\dot{p}(f)) + c_9c_4y(\dot{p}(f)) \\ -c_9c_3x(\dot{p}(g)) - c_9c_4y(\dot{p}(g)) \\ -(c_9c_3c_7 + c_9c_4c_8)\dot{\theta}(g) \geq 0 \end{aligned} \quad (34)$$

Axiom 17 becomes the following equation instantiated between all pairs of lines segments with rigid displacement in the current joint model:

$$\begin{aligned} c_3x(\dot{p}(f)) + c_4y(\dot{p}(f)) + (c_3c_{11} + c_4c_{12})\dot{\theta}(f) \\ -c_3x(\dot{p}(g)) - c_4y(\dot{p}(g)) - (c_3c_{13} + c_4c_{14})\dot{\theta}(g) = 0 \end{aligned} \quad (35)$$

Finally, axiom 18 becomes the following simple equation instantiated between all pairs of line segments with rigid relative rotation in the current joint model:

$$\dot{\theta}(f) - \dot{\theta}(g) = 0 \quad (36)$$

Note that none of the above constraint schemas have nonzero constant terms. This means that any system of constraints constructed out of instantiations of these schemas will always have the zero vector as a solution. This has a straightforward physical interpretation, namely that the constraints are satisfied if the image is interpreted as being static with no instantaneous motion. This would be the only solution if all objects in the image were stable. If this were not the case then there would be additional solutions for which the left hand side of axiom 2 was less than zero denoting a decrease in the potential energy of the system. Since the ontology is purely kinematic and there is no notion of gravitational acceleration, if there exists any solution that decreases the potential energy then there exist an infinite number of solutions where the left hand side of axiom 2 evaluates to any negative real number.

This leads to two distinct methods for determining the stability of a fixed image under fixed layer and joint models. First, one could formulate a linear programming problem taking the above axioms as the constraints and taking the left hand side of axiom 2 as the objective function. This linear programming problem has the zero vector as a trivial basic feasible solution. Furthermore, since each optimization step is guaranteed to increase the objective function, it is only necessary to execute a single optimization step to determine whether or not the system is stable. Alternatively, one can formulate this as a satisfiability problem instead of an optimization problem by replacing axiom 2 with a constraint that restricted the decrease in potential energy to be equal to an arbitrary fixed nonzero constant. The resulting system of equations and inequalities would have a solution if and only if the system was unstable. While the former is likely to be more efficient, the later may be more amenable to implementation within existing constraint logic programming languages such as CLP($\Re$)(Jaffar and Lassez 1987).

One can extend this technique to determine which objects in an image are supported. ABIGAIL treats certain connected sets of line segments as objects. One can determine whether or not a given object is supported by augmenting the system of constraints with an additional constraint restricting that object itself to have a nonzero decrease in potential energy. The augmented system of constraints is satisfiable if and only if that object is unsupported. Furthermore, one can determine whether an object A supports an object B by seeing whether B is supported but ceases to be if A is removed.

5 Examples

A preliminary implementation of the ideas in this paper has been written in COMMON LISP and has been demonstrated to work on small examples. An example scene processed by this new approach is illustrated in figure 2. In this example the ground is reified as the line segment g to allow objects to be attached to the ground. The configuration is unstable with an empty layer model and with

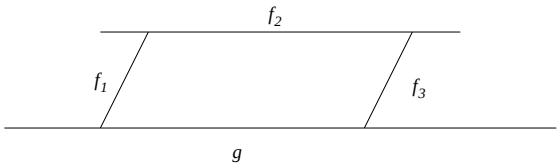

Figure 2: A sample scene whose stability can be determined by the approach discussed in this paper.

the following joint model:

$$\begin{array}{llll} f_1 \leftrightarrow f_2 & f_2 \leftrightarrow f_1 & f_2 \leftrightarrow f_3 & f_3 \leftrightarrow f_2 \\ f_1 \leftrightarrow g & g \leftrightarrow f_1 & f_3 \leftrightarrow g & g \leftrightarrow f_3 \\ \delta(f_1, f_2) & \delta(f_2, f_1) & \delta(f_2, f_3) & \delta(f_3, f_2) \\ \delta(f_1, g) & \delta(g, f_1) & \delta(f_3, g) & \delta(g, f_3) \end{array}$$

This is because the parallelogram can collapse. When the assertions $\theta(f_1, f_2)$ and $\theta(f_2, f_1)$ are added to the joint model the configuration becomes stable. Note that both $\theta(f_1, f_2)$ and its symmetric variant $\theta(f_2, f_1)$ must be added to the joint model to satisfy axiom 13.

The ability to determine the stability of a scene forms the foundation of most of ABIGAIL's perceptual processing. This allows her to construct initial layer and joint models, to update these models over time, to segment each image into objects, and to track changing support, contact, and attachment relations between these objects. Siskind (1993) illustrates how one can formulate descriptions of some simple spatial motion verbs as compound temporal logic expressions over atomic expressions denoting changing support, contact, and attachment relations. For example, one can describe *put* with the following expression:

$$\exists w \left(\left[\begin{array}{c} \textsc{Part}(w, x) \wedge \\ \left(\begin{array}{c} \textsc{Translating}(w) \wedge \\ \textsc{Contacts}(w, y) \wedge \\ \textsc{Attached}(w, y) \wedge \\ \textsc{Supports}(x, y) \wedge \\ \textsc{Translating}(y) \end{array} \right) ; \\ \exists z \left(\begin{array}{c} \textsc{Disjoint}(z, w) \wedge \\ \neg \Diamond \textsc{Translating}(y) \wedge \\ \textsc{Supported}(y) \wedge \\ \textsc{Supports}(z, y) \end{array} \right) \end{array} \right] \right)$$

This states that an agent x puts an object y on z if there exists an object w distinct from z, presumably the agent's hand, such that for some time interval, w and y are translating, w contacts and is attached to y, and x supports y, and during an immediately subsequent time interval, y is stationary, supported by z. Similarly, one can describe *throw* with the following expression:

$$\exists z \left(\left[\begin{array}{c} \neg \Diamond \textsc{Part}(y, x) \wedge \\ \textsc{Part}(z, x) \wedge \\ \left(\begin{array}{c} \textsc{Translating}(z) \wedge \\ \textsc{Contacts}(z, y) \wedge \\ \textsc{Attached}(z, y) \end{array} \right) ; \\ \left(\begin{array}{c} \neg \Diamond \textsc{Contacts}(z, y) \wedge \\ \neg \Diamond \textsc{Attached}(z, y) \wedge \\ \neg \Diamond \textsc{Supported}(y) \end{array} \right) \end{array} \right] \wedge \right.$$
$$\left. \textsc{Translating}(y) \right)$$

This states that an agent x throws an object y if y is not a part of x and there exists an object z that is a part of x, presumably the agent's hand, such that for some time interval, z is translating while contacting and being attached to y, and during the immediate subsequent time interval, z no longer contacts or is attached to y, and y is in unsupported motion. Given these definitions, ABIGAIL can detect the following events from the movie in figure 1:[3]

```
[20:26,27](PUT [JOHN-part 3] [BALL])
[6:12,13:15](THROW [JOHN-part 3] [BALL])
```

A future paper will detail the process by which the truth of the aforementioned temporal logic expressions can be determined to yield the above event detections.

6 Conclusion

Work is underway to replace the kinematic simulator component of earlier versions of ABIGAIL with the stability determination algorithm discussed here. Since the new approach uses very different internal data-structure representations than the prior approach, this task is time consuming and not yet complete. A future paper will report on the results of this effort and compare the performance of the old technique with the new technique.

The new approach has at least three potential advantages over the previous approach described in Siskind (1992). First, it should be substantially faster. Second, it should be amenable to potential parallel implementation. Third, unlike the previous approach, it can correctly determine the stability of images with closed loop kinematic chains. The ability of the current approach for dealing with closed loop kinematic chains is illustrated by the example in figure 2. On the other hand, it is more limited than the prior approach in that it can only handle images constructed solely from line segments. The previous approach could also handle circles. Nonetheless, I believe that the approach taken in this paper can be extended straightforwardly to handle not only circles, but arbitrary curved surfaces as well. This will be discussed in a future paper.

[3] This detection was performed using an older version of ABIGAIL based on kinematic simulation since the newer linear-programming-based method for determining stability has not yet been fully integrated into ABIGAIL.

It is also straightforward to extend this approach to three dimensions. I have recently begun a follow on project to ABIGAIL to investigate the feasibility of applying the techniques described in this paper to real video image sequences. The target task for this project is the recognition of simple spatial motion events like *pick up*, *put down*, *drop*, and *push* performed by a human hand on objects resting on a table. A conventional model-based vision front-end will be used to perform object recognition and localization. A three-dimensional variant of the techniques used in ABIGAIL will then be used to recover support, contact, and attachment relations and ultimately event occurrences. This project will be discussed more fully in a future paper.

Acknowledgments

I wish to thank Richard Mann for numerous discussions related to this topic and for pointing me towards Blum et al. (1971) and Fahlman (1974). Ken Anderson provided tremendous assistance in optimizing the COMMON LISP implementation of the Simplex algorithm used for testing the ideas in this paper.

References

[1] Norman I. Badler. Temporal scene analysis: Conceptual descriptions of object movements. Technical Report 80, University of Toronto Department of Computer Science, February 1975.

[2] M. Blum, A. K. Griffith, and B. Neumann. A stability test for configurations of blocks. A. I. Memo 188, M. I. T. Artificial Intelligence Laboratory, February 1970.

[3] Gary Borchardt. A computer model for the representation and identification of physical events. Technical Report T-142, Coordinated Sciences Laboratory, University of Illinois at Urbana-Champaign, May 1984.

[4] Scott Elliott Fahlman. A planning system for robot construction tasks. *Artificial Intelligence*, 5(1):1–49, 1974.

[5] Ellen M. Hays. On defining motion verbs and spatial prepositions. In Ch. Freksa and Ch. Habel, editors, *Repräsentation und Verarbeitung räumlichen Wissens*, pages 192–206. Springer-Verlag, 1989.

[6] Annette Herskovits. *Language and Spatial Cognition: An interdisciplinary study of the prepositions in English.* Cambridge University Press, 1986.

[7] Ray Jackendoff. *Semantics and Cognition.* The MIT Press, Cambridge, MA, 1983.

[8] Ray Jackendoff and Barbara Landau. Spatial language and spatial cognition. In Donna Jo Napoli and Judy Anne Kegl, editors, *Bridges Between Psychology and Linguistics: A Swarthmore Festschrift for Lila Gleitman.* Lawrence Erlbaum Associates, Hillsdale, NJ, 1991.

[9] Joxan Jaffar and Jean-Louis Lassez. Constraint logic programming. In *Proceedings of the* 14^{th} *ACM Symposium on the Principles of Programming Languages*, pages 111–119, 1987.

[10] Geoffrey N. Leech. *Towards a Semantic Description of English.* Indiana University Press, 1969.

[11] George A. Miller. English verbs of motion: A case study in semantics and lexical memory. In Arthur W. Melton and Edwin Martin, editors, *Coding Processes in Human Memory*, chapter 14, pages 335–372. V. H. Winston and Sons, Inc., Washington, DC, 1972.

[12] Naoyuki Okada. SUPP: Understanding moving picture patterns based on linguistic knowledge. In *Proceedings of the Sixth International Joint Conference on Artificial Intelligence*, pages 690–692, 1979.

[13] Steven Pinker. *Learnability and Cognition.* The MIT Press, Cambridge, MA, 1989.

[14] Terrance Philip Regier. *The Acquisition of Lexical Semantics for Spatial Terms: A Connectionist Model of Perceptual Categorization.* PhD thesis, University of California at Berkeley, 1992.

[15] Roger C. Schank. The fourteen primitive actions and their inferences. Memo AIM-183, Stanford Artificial Intelligence Laboratory, March 1973.

[16] Jeffrey Mark Siskind. Naive physics, event perception, lexical semantics and language acquisition. In *The AAAI Spring Symposium Workshop on Machine Learning of Natural Language and Ontology*, pages 165–168, March 1991.

[17] Jeffrey Mark Siskind. *Naive Physics, Event Perception, Lexical Semantics, and Language Acquisition.* PhD thesis, Massachusetts Institute of Technology, Cambridge, MA, January 1992.

[18] Jeffrey Mark Siskind. Grounding language in perception. In *Proceedings of the Annual Conference of the Society of Photo-Optical Instrumentation Engineers*, September 1993.