

# Multivalued Default Logic for Identity Maintenance in Visual Surveillance \*

Vinay D. Shet, David Harwood, and Larry S. Davis

Computer Vision Laboratory, University of Maryland,  
College Park, MD, USA  
{vinay,harwood,lsd}@umiacs.umd.edu

**Abstract.** Recognition of complex activities from surveillance video requires detection and temporal ordering of its constituent “atomic” events. It also requires the capacity to robustly track individuals and maintain their identities across single as well as multiple camera views. Identity maintenance is a primary source of uncertainty for activity recognition and has been traditionally addressed via different appearance matching approaches. However these approaches, by themselves, are inadequate. In this paper, we propose a prioritized, multivalued, default logic based framework that allows reasoning about the identities of individuals. This is achieved by augmenting traditional appearance matching with contextual information about the environment and self identifying traits of certain actions. This framework also encodes qualitative confidence measures for the identity decisions it takes and finally, uses this information to reason about the occurrence of certain predefined activities in video.

---

\* We thank the U.S.Government for supporting the research described in this paper.

## 1 Introduction

The primary goal of a visual surveillance system is to help ensure safety and security by detecting the occurrence of activities of interest within an environment. This typically requires the capacity to robustly track individuals not only when they are within the field of regard of the cameras, but also when they disappear from view and later reappear. Figure 1 shows an individual marked X appearing in the scene with a bag, dropping it off in the corridor, and disappearing from view through a door. Subsequently it shows individual Y appearing in the scene through the same door and picking up the bag.



**Fig. 1.** Sequence of images showing individual X appearing in the scene with a bag, depositing it on the ground and disappearing from view. Subsequently, individual Y appears in the scene, picks up the bag and leaves.

If  $individual(X) = individual(Y)$ , the activity by itself, is probably not of interest from a security viewpoint. However, if  $individual(X) \neq individual(Y)$ , the activity observed could possibly be a theft. This example captures the general problem of automatically inferring whether two individuals observed in the video are equal or not. This problem is significant not only for camera setups where individuals routinely disappear into and reappear from pockets of the world not observed by the cameras, but also within a single field of view when tracking is lost due to a variety of reasons.

Traditionally in surveillance, the problem of identity maintenance has been addressed by appearance matching. Matching of appearances can be based on a person's color distribution and shape [1], gait [2], face [3] and other physical characteristics. All of these approaches are considered weak biometrics and, by themselves, they are inadequate for maintaining identities for recognizing complex activities.

The objectives of this paper are to provide a framework

1. **that supports reasoning about identities of individuals observed in video.** We do this by augmenting traditional appearance matching with (a) contextual information about the world and (b) self identifying traits associated with actions. In addition to stating whether or not two individuals are equal, we also qualitatively encode our confidence in it.
2. **that facilitates using this information on identities to recognize activities.** We also propagate our confidence in the identity statements to activities to which they contribute.

In the example above, if the door through which individual X disappeared leads into a closed world (a world with no other exit), we could, under some circumstances, infer that individual Y coming out of that door at a later time had to be equal to individual X

(with a high degree of confidence), regardless of whether or not he appeared similar to X.

In this work, we encode contextual information about the world and our common sense knowledge about self-identifying actions as rules in a logic programming language. Furthermore, we observe that since these rules reflect actions taking place in a real world, they can never be definite and completely correct. We therefore employ default logic as the language to specify these rules, which provides our framework the important property of nonmonotonicity (the property of retracting or disbelieving old beliefs upon acquisition of new information). We also employ a bilattice based multi-valued representation that encodes our confidence in various rules and propagates these confidence values to the identity statements and subsequently to the activities themselves. We then use prioritization over these default rules to capture the fact that different cues could provide us with different amounts of information. Finally, we use this information about identities of individuals to reason about the occurrence of activities in the video.

## 2 Motivation

Our primary motivation is to build a visual surveillance system that draws heavily upon human reasoning. While humans are very skillful in matching appearances, even we commit mistakes in this process. However, we possess the capacity to employ context and non-visual cues to aid us in recovering from these errors.

**Example 1** *Upon observing an individual, from the back and walking away from us, based on his gait and possibly body type, we tentatively conclude that the individual is Tom, a colleague at work. However, if we suddenly remember that Tom called in sick earlier in the day, we may decide that it cannot be Tom. Later still, if we observe that individual enter a Black BMW, a type of car we know Tom owns, we might conclude more strongly this time that it has to be Tom. However, before entering the car, if the individual turns around to face us and we realize that it is a person we have never seen before, we may definitely conclude that it is not Tom.*

The example demonstrates how humans employ common sense to reason about identities. Human reasoning is characterized, among other things, by [4]

1. **Its ability to err and recover** - This is important because when dealing with uncertain input, decisions or analysis made might have to be retracted upon acquisition of new information. In Example 1, we retracted our belief of the person being Tom or not several times,
2. **Its qualitative description of uncertainty** - a qualitative gradation of belief permits us to encode our confidence in decisions we make. In Example 1, our degree of belief in whether or not the person was Tom moved from slightly sure to definitely sure.
3. **Prioritization** - it is important to have a sense of how reliable our thread of reasoning is. In Example 1, based on appearance we were only slightly sure, based on vehicle information we were more sure, based on face recognition we were definitely sure etc.

### 3 Related Work

Identity maintenance in surveillance has typically only employed some form of appearance matching. [1] uses a SVM based approach to recognize individuals in indoor images based on color and shape based features. [2] employs gait as a characteristic to identify individuals while [3] performs face recognition from video. Microsoft’s *EasyLiving* project [5] employs two stereo cameras to track up to 3 people in a small room while [6] describes a multi-camera indoor people localization in a cluttered environment.

Activity recognition has traditionally been performed using statistical approaches. Hidden Markov Models have been used recognize primitive actions in [7] [8] and also complex behaviors in [9]. Bayesian networks are also widely used [10] [11]. Non statistics based approaches have also been used to recognize activities. [12], proposes an approach based on declarative models of activities and defines scenarios for *Vandalism*, *Access forbidden and Holdup* and uses a hierarchy of facts ranging from abstract to concrete to recognize these situations. [13] investigates the use of qualitative spatio-temporal representations and abduction in an architecture for Cognitive Vision. [14] employs a context representation scheme for surveillance systems. [15] considers an activity to be composed of action threads and recognizes activities by propagating constraints and likelihood of event threads in a temporal logic network. [16] uses *scenarios* to declare spatio-temporal knowledge in vision applications. [17] uses qualitative representation of uncertainty.

The primary motivation for employing a logic programming based identity maintenance and activity recognition approach is the expressive power it bestows upon our system that permits us to encode not only complex propositions but also functions and quantification. Use of well studied logic programming languages provides us with an efficient and ready-to-use mechanism for searching and backward chaining.

### 4 Reasoning Framework

Logic programming systems employ formulae that are either facts or rules to arrive at inferences. In visual surveillance, rules can be used to define various activities of interest as well as intermediate inferences such as that of equality of individuals. Rules are of the form “ $A \leftarrow A_0, A_1, \dots, A_m$ ” where each  $A_i$  is called an atom and ‘,’ represents logical conjunction. Each atom is of the form  $p(t_1, t_2, \dots, t_n)$ , where  $t_i$  is a term, and  $p$  is a predicate symbol of arity  $n$ . Terms could either be variables (denoted by upper case alphabets) or constant symbols (denoted by lower case alphabets). The left hand side of the rule is referred to as the head and the right hand side is the body. Rules are interpreted as “if body then head”. Facts are logical rules of the form “ $A \leftarrow$ ” (henceforth denoted by just “ $A$ ”) and correspond to the input to the inference process. These facts are the output of the computer vision algorithms, and include “atomic” events detected in video (entering/exiting a door, picking up a bag) and data from background subtraction and tracking. Finally, ‘ $\neg$ ’ represents negation such that  $A = \neg\neg A$ .

#### 4.1 Default Logic

Logic programming based visual surveillance systems apply a set of predefined logical rules defining each activity to logical facts generated in real time from events transpiring in video to recognize activities [18]. Traditional logic programs are based on deduction,

which is a method of exact inference. If the body of a rule evaluates to true, then the head always evaluates to true; in classical logic, there exists no provision of changing the truth value of the head over time. Deduction therefore requires information to be complete, precise and consistent. By contrast, in real world surveillance scenarios, one has to deal with incomplete, imprecise and potentially inconsistent information. Humans possess the ability to reason effectively under such circumstances using what is termed “common sense reasoning”. Default logic [19] is an attempt to formalize common sense reasoning using default rules. Default logic expresses rules that are “true by default” or “generally true” but could be proven false upon acquisition of new information in the future. This property of default logic, where the truth value of a proposition can change if new information is added to the system, is called nonmonotonicity.

**Definition 1 (Default Theory).** A default theory  $\Delta$  is of the form  $\langle W, D \rangle$ , where  $W$  is a set of traditional first order logical formulae (rules and facts) also known as the definite rules and  $D$  is a set of default rules of the form  $\frac{\alpha:\beta}{\gamma}$ , where  $\alpha$  is known as the precondition,  $\beta$  is known as the justification and  $\gamma$  is known as the inference or conclusion.

A default rule of this form expresses that if the precondition  $\alpha$  is known to be true, and the justification  $\beta$  is consistent with what is currently in the knowledge base, then it is possible to conclude  $\gamma$ . Such a rule can be also written as  $\gamma \leftarrow \alpha, \text{not}(\neg\beta)$ . ‘not’ represents the negation by “failure to prove” operator and the consistency check for  $\beta$  is done by failure to prove its negation.

**Example 2** Assume the following set of rules and facts:

$$\begin{aligned} \neg\text{equal}(P_1, P_2) &\leftarrow \text{distinct}(P_1, P_2). \in W \\ \text{equal}(P_1, P_2) &\leftarrow \text{appear\_similar}(P_1, P_2), \text{not}(\neg\text{equal}(P_1, P_2)) \in D \\ &\quad \{\text{appear\_similar}(a, b)\}_t \\ &\quad \{\text{appear\_similar}(a, b), \text{distinct}(a, b)\}_{t+1} \end{aligned}$$

where  $\{\dots\}_t$  indicates the set of facts at time  $t$  and  $\text{distinct}(a, b)$  indicates that  $a$  and  $b$  appear as two separate and distinct individuals at some point of time.

In this example, at time  $t$ , given the rules and the set of facts, the system concludes that since it cannot prove  $\neg\text{equal}(a, b)$  and  $\text{appear\_similar}(a, b)$  is true, therefore  $\text{equal}(a, b)$  is true. However, at time  $t+1$ , it is now possible to prove  $\neg\text{equal}(a, b)$  because  $\text{distinct}(a, b)$  is true and therefore the system now can no longer conclude  $\text{equal}(a, b)$  (the default rule is blocked by the definite rule) and concludes  $\neg\text{equal}(a, b)$  instead.

While the property of a conclusion blocking another default rule is desirable since it bestows nonmonotonicity upon the system, it can also create a problem.

**Example 3** Assume the following set of rules and facts:

$$\begin{aligned} \neg\text{equal}(P_1, P_2) &\leftarrow \text{distinct}(P_1, P_2), \text{not}(\text{equal}(P_1, P_2)). \in D \\ \text{equal}(P_1, P_2) &\leftarrow \text{appear\_similar}(P_1, P_2), \text{not}(\neg\text{equal}(P_1, P_2)) \in D \\ &\quad \{\text{appear\_similar}(a, b), \text{distinct}(a, b)\}_t \end{aligned}$$

In Example 3, the rule for inferring that two individuals are not equal if they appear distinct is now made a default rule<sup>1</sup>. In this case, given the set of facts, at time  $t$ , depending on the order in which the default rules are applied, different sets of conclusions can be produced. If the first default is applied first, it blocks the second default and we conclude  $\neg equal(a, b)$ ; but if the second default is applied first, it blocks the first and we conclude  $equal(a, b)$ .

**Definition 2 (Extensions).** *The different sets of conclusions that can be derived by applying defaults in different orders are called extensions.*

A default theory can have multiple extensions, each capturing a possible outcome of the definite and default rules. While multiple extensions of a default theory list its possible outcomes, they are of not much use if a single solution is needed. There are several different approaches in the literature to obtain a single solution from the space of extensions of the default theory, including specificity [20], prioritized defaults [21] and multi-valued belief states [22]. Our system adopts the latter.

In the multivalued belief states approach, various rules in the system are regarded as different sources of information concerning the truth value<sup>2</sup> of a given proposition. These sources contribute different amounts of information to the decision making process and consequently our degree of belief in these propositions should mirror the information content. For example, default rules are not always correct and could be proven wrong by definite rules. Therefore, in this approach, definite rules provide more information than default rules. We seek a representation that combines truth value of these belief states with the information content of the sources.

## 4.2 Bilattice Theory

Bilattices [22] provide an elegant and convenient formal framework in which the information content from different sources can be viewed in a truth functional manner. Truth values assigned to a given proposition are taken from a set structured as a bilattice.

**Definition 3 (Lattice).** *A lattice is a set  $L$  equipped with a partial ordering  $\leq$  over the elements of  $L$ , a greatest lower bound (glb) and a lowest upper bound (lub) and is denoted by the triple  $(L, glb, lub)$  where  $glb$  and  $lub$  are operations from  $L \times L \rightarrow L$  that are idempotent, commutative and associative*

Informally a bilattice is a set,  $B$ , of truth values composed of two lattices  $(B, \wedge, \vee)$  and  $(B, \cdot, +)$  each of which is associated with a partial order  $\leq_t$  and  $\leq_k$  respectively. The  $\leq_t$  partial order indicates how true or false a particular value is, with  $f$  being the minimal and  $t$  being the maximal. The  $\leq_k$  partial order indicates how much is known about a particular sentence. The minimal element here is  $u$  (completely unknown) while the maximal element is  $\perp$  (representing a contradictory state of knowledge where a sentence is both true and false). The glb and the lub operators on the  $\leq_t$  partial order are  $\wedge$

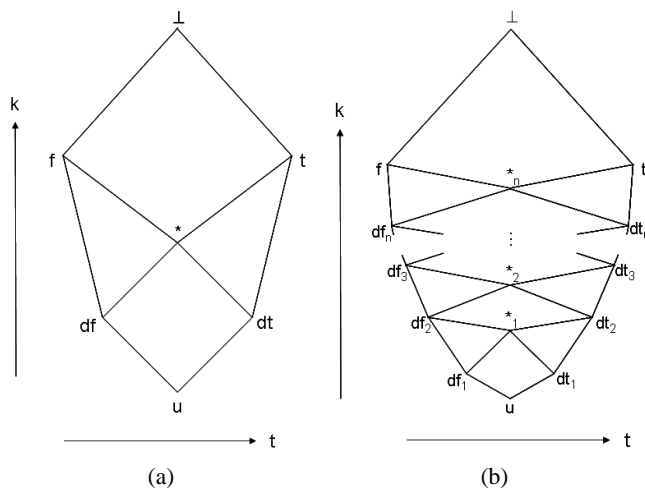
<sup>1</sup> This default rule captures the fact that if there exists a mirror in the world, it could be possible for a single person to appear as two distinct individuals

<sup>2</sup> It is important to note that by truth value we mean our degree of belief in the veracity or falsity of a given proposition. This is different from the actual truth value of the proposition in the real world.

and  $\vee$  and correspond to the usual logical notions of conjunction and distinction, respectively. The glb and the lub operators on the  $\leq_k$  partial order are  $\cdot$  and  $+$ , respectively, where  $+$  corresponds to the combination of evidence from different sources or lines of reasoning while  $\cdot$  corresponds to the consensus operator. A bilattice is also equipped with a negation operator  $\neg$  that inverts the sense of the  $\leq_t$  partial order while leaving the  $\leq_k$  partial order intact.

**Definition 4 (Bilattice [22]).** A bilattice is a sextuple  $(B, \wedge, \vee, \cdot, +, \neg)$  such that

- $(B, \wedge, \vee)$  and  $(B, \cdot, +)$  are both lattices and
- $\neg$  is a mapping such that
  - $\neg^2 = 1$  and
  - $\neg$  is a homomorphism from  $(B, \wedge, \vee)$  to  $(B, \vee, \wedge)$  and from  $(B, \cdot, +)$  to itself.



**Fig. 2.** (a) Bilattice for default logic (b) Bilattice for prioritized default logic.

**Properties of Bilattices** Figure 2(a) shows a bilattice corresponding to classical default logic. The set  $B$  of truth values contains, in addition to the usual definite truth values of  $t$  and  $f$ ,  $dt$  and  $df$  corresponding to true-by-default (also called “decided-true”) and false-by-default (also called “decided-false”),  $u$  corresponding to “unknown”,  $*$  corresponding to “undecided” (indicating contradiction between  $dt$  and  $df$ ) and  $\perp$  corresponding to “contradiction” (between  $t$  and  $f$ ). The  $t$ -axis reflects the partial ordering on the truth values while the  $k$ -axis reflects that over the information content. This bilattice provides us with a correlation between the amount of information and our degree of belief in a source’s output. Procuring more information about a proposition, indicated by rising up along the  $k$ -axis, causes us to move away from the center of the  $t$ -axis towards more definitive truth values. The only exception to this being in case of a contradiction, we move back to the center of the  $t$ -axis. Negation corresponds to reflection of the bilattice

about the  $\perp -u$  axis. It is also important to note the this bilattice is distributive with respect to each of the four operators. Based on this framework, we can define the truth tables for each of the four operators as defined in figure 3.

$\wedge$	$\perp$	T	F	*	DT	DF	U	$\vee$	$\perp$	T	F	*	DT	DF	U
$\perp$	$\perp$	$\perp$	F	U	$\perp$	DF	U	$\perp$	$\perp$	T	$\perp$	T	DT	$\perp$	T
T	$\perp$	T	F	*	DT	DF	U	T	T	T	T	T	T	T	T
F	F	F	F	F	F	F	F	F	$\perp$	T	F	*	DT	DF	U
*	U	*	F	*	*	DF	U	*	T	T	*	*	DT	*	DT
DT	$\perp$	DT	F	*	DT	DF	U	DT	DT	T	DT	DT	DT	DT	DT
DF	DF	DF	F	DF	DF	DF	DF	DF	$\perp$	T	DF	*	DT	DF	U
U	U	U	F	U	U	DF	U	U	T	T	U	DT	DT	U	U

$\cdot$	$\perp$	T	F	*	DT	DF	U	+	$\perp$	T	F	*	DT	DF	U
$\perp$	$\perp$	T	F	*	DT	DF	U	$\perp$	$\perp$	$\perp$	$\perp$	$\perp$	$\perp$	$\perp$	$\perp$
T	T	T	F	*	DT	DF	U	T	$\perp$	T	$\perp$	T	T	T	T
F	F	F	F	*	DT	DF	U	F	$\perp$	$\perp$	F	F	F	F	F
*	*	*	*	*	DT	DF	U	*	$\perp$	T	F	*	*	*	*
DT	DT	DT	DT	DT	DT	DF	U	DT	$\perp$	T	F	*	DT	*	DT
DF	DF	DF	DF	DF	DF	DF	U	DF	$\perp$	T	F	*	*	DF	DF
U	U	U	U	U	U	U	U	U	$\perp$	T	F	*	DT	DF	U

Fig. 3. Truth table for glb and lub operators for t and the k axis of the bilattice for default logic.

### 4.3 Inference

**Definition 5 (Truth Assignment).** Given a declarative language  $L$ , a truth assignment is a function  $\phi : L \rightarrow B$  where  $B$  is a bilattice on truth values.

The semantics of a bilattice system is given by a definition of closure. If  $\mathcal{K}$  is the knowledge base and  $\phi$  is a truth assignment labelling each sentence  $k \in \mathcal{K}$  with a truth value then the closure of  $\phi$ , denoted  $cl(\phi)$ , is the truth assignment that labels information entailed by  $\mathcal{K}$ . For example, if  $\phi$  labels sentences  $\{p, q \leftarrow p\} \in \mathcal{K}$  as true; i.e.  $\phi(p) = T$  and  $\phi(q \leftarrow p) = T$ , then  $cl(\phi)$  should also label  $q$  as true as it is information entailed by  $\mathcal{K}$ . Entailment is denoted by the symbol ' $\models$ ' ( $\mathcal{K} \models q$ ).

If  $S \subset L$  is a set of sentences entailing  $q$ , then the truth value to be assigned to the conjunction of elements of  $S$  is

$$\bigwedge_{p \in S} cl(\phi)(p) \quad (1)$$

This term represents the conjunction of the closure of the elements of  $S$ . It is important to note that this term is merely a contribution to the truth value of  $q$  and not the actual truth value itself. The reason it is merely a contribution is because there could be other sets of sentences  $S$  that entail  $q$  representing different lines of reasoning (or, in our case, different rules). The contributions of these sets of sentences need to be combined using the  $+$  operator. Also, if the expression in 1 evaluates to false, then its contribution

to the value of  $q$  should be “unknown” and not “false”. These arguments suggest that the closure over  $\phi$  of  $q$  is

$$cl(\phi)(q) = \sum_{S \models q} u \vee [\bigwedge_{p \in S} cl(\phi)(p)] \quad (2)$$

We also need to take into account the set of sentences entailing  $\neg q$ . Since  $\phi(\neg q) = \neg\phi(q)$ , aggregating this information yields the following expression

$$cl(\phi)(q) = \sum_{S \models q} u \vee [\bigwedge_{p \in S} cl(\phi)(p)] + \neg \sum_{S \models \neg q} u \vee [\bigwedge_{p \in S} cl(\phi)(p)] \quad (3)$$

For more information on the properties and logical inference based on bilattice theory see [22].

#### Example 4 (Inference example)

$$\begin{aligned} \phi[\neg equal(P_1, P_2) \leftarrow distinct(P_1, P_2)] &= DT \\ \phi[equal(P_1, P_2) \leftarrow appear\_similar(P_1, P_2)] &= DT \\ \phi[appear\_similar(a, b)] &= T \\ \phi[distinct(a, b)] &= T \end{aligned}$$

$$\begin{aligned} cl(\phi)(equal(a, b)) &= [U \vee (T \wedge DT)] + \neg[U \vee (T \wedge DT)] \\ &= [U \vee DT] + \neg[U \vee DT] = DT + DF = * \end{aligned}$$

In Example 4, we encode our belief that the two rules are only true in general and do not always hold by assigning a truth value of  $DT$  to them. We record our belief in the facts as  $T$  and apply equation 3 to compute the truth value of  $equal(a, b)$ . Note in Example 3, we obtained two extensions with  $equal(a, b)$  being true in one and  $\neg equal(a, b)$  being true in another. Using the multivalued logic approach we collapse these extensions and combine the two conclusions to obtain  $DT + DF = *$  or “undecided”.

#### 4.4 Belief Revision

In classical AI, belief revision is the process of revising a proposition’s belief state upon acquisition of new data. In the bilattice framework presented above, these revisions should only occur if the new data source promises more information than that which triggered the current truth value assignment. Note that the belief combination operator,  $+$  is a lub operator on the  $k$ -axis, meaning it will only choose a sentence with maximum information.

However, this poses a problem for our current theory. Since default rules could be contradicted by other default rules, it is possible that many propositions will suffer from a  $DT, DF$  contradiction and will settle in the  $*$  or undecided state. According to our current theory, only a rule with more information, the definite rules, can release it from this state. Unfortunately in visual surveillance, most rules are default rules and therefore it might be the case that there may be no definite rules to rescue a proposition once it gets labelled “undecided”.

**Example 5** Assume that an individual enters a room we believe to be empty and closed (no other exit). Assume also that after some time, another individual emerges from the room who appears dissimilar from the first individual

$$\begin{aligned} \phi[\neg \text{equal}(P_1, P_2) \leftarrow \neg \text{appear\_similar}(P_1, P_2)] &= DT \\ \phi[\text{equal}(P_1, P_2) \leftarrow \text{enterclosedworld}(P_1, X, T_1), \\ &\quad \text{exitclosedworld}(P_2, X, T_2), T_2 > T_1, \\ &\quad \text{emptybefore}(X, T_1), \text{emptyafter}(X, T_2), \\ &\quad \text{not}(\text{enter\_or\_exit\_between}(P_3, T_1, T_2)).] &= DT \end{aligned}$$

$$\begin{aligned} \phi[\neg \text{appear\_similar}(a, b)] &= T \\ \phi[\text{enterclosedworld}(a, \text{office}, 400)] &= T \\ \phi[\text{exitclosedworld}(b, \text{office}, 523)] &= T \\ \phi[\text{emptybefore}(\text{office}, 400)] &= DT \\ \phi[\text{emptyafter}(\text{office}, 523)] &= DT \\ \phi[\text{not}(\text{enter\_or\_exit\_between}(P_3, 400, 523))] &= T \\ cl(\phi)(\text{equal}(a, b)) &= \dots = DT + DF = * \end{aligned}$$

In Example 5, the first rule states that if two individuals do not appear similar, then they are not equal. The second rule, states that if there exists a closed world that we believe to be empty and we observe an individual enter it and at a subsequent time exit it and no one else has entered or exited the closed world in between, then we can conclude that the two individuals are equal. The set of facts captures the activity of an individual entering a closed empty world and later reappearing and looking dissimilar from the individual who entered. In this case, too, we have contradicting defaults and on applying equation 3,  $\text{equal}(a, b)$  gets labelled \*.

#### 4.5 Prioritized Defaults

This problem arises because thus far we are assuming that all the default rules provide us the same amount of information, causing them to contradict each other and force a proposition into the \* state. However, suppose, instead we assume that different defaults could provide different amounts of information and consequently could alter our belief state by different degrees. It turns out that the bilattice structure very elegantly generalizes to accommodate this assumption. We could modify the previous example and state that inferring equality based on appearance matching is a weaker default than inferring equality based on the fact that the person entered and exited an empty closed world. Therefore, if we then assign a label  $DT_1$  to default 1 and label  $DT_2$  to default 2 and state that  $DT_2$  is a stronger default and has more information than  $DT_1$  we can conclude

$$cl(\phi)(\text{equal}(a, b)) = DT_2 + \neg DT_1 = DT_2 + DF_1 = DT_2$$

Figure 2(b) shows a general bilattice for a prioritized default theory with n priorities. Formally a prioritized default theory  $\Delta_{<}$  is of the form  $\langle W, D, < \rangle$  [21] where  $W$  and

$D$  are as defined in Definition 1 and  $<$  is a strict partial ordering on  $D$ . The semantics of the bilattice on the new set of truth values stays the same as before.

## 5 Reasoning about Identities

Our system primarily employs four identifying cues or traits for reasoning about identities. These cues are based on the individuals possessions, closed world activity, knowledge and appearance. In addition to these cues, we also employ equality axioms of reflexivity, transitivity, and symmetry.

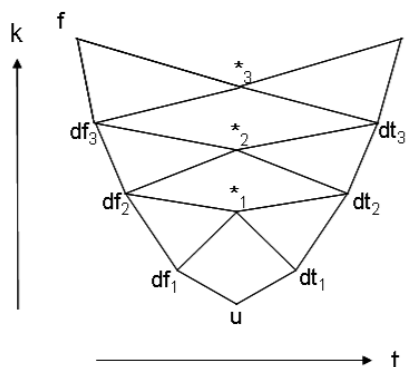
Identity can be verified on basis of a person possessing something that only he can possess. For example, if we know that a vehicle belongs to an individual and later we observe another individual entering that vehicle using a key, we can conclude that they must be equal. An individual can be identified on the basis of certain closed world activities, examples of which we have seen earlier (see Example 5). One can also verify identity on the basis of the knowledge we think an individual possesses. For example, if there is a combination lock on a door controlling access to a office and we observe an individual successfully entering the code and opening the door to enter the room, we can conclude that he must be the owner of that office. Finally appearance based cues help identify individuals based on appearance. We employ a color histogram based appearance matching algorithm.

It should be noted that any rule based on these cues can almost never be definitive and most of them will be default rules. Also, different cues provide us with different amounts of information as they deal with varying degrees of uncertainty. Without loss of generality, we assume three levels or priorities of defaults<sup>3</sup>. Also, we assume that the definite rules are never incorrect and therefore there will never occur a contradiction between  $T$  and  $F$ . Figure 4 shows the resultant bilattice employed in our system.

### 5.1 Rules of identity

In this section we will give English descriptions of various rules employed in our system, and note their priority levels.

**Priority Level 1** Appearance based identification states that if two individuals appear similar to each other then they are equal to each other. On the other hand, if two individuals do not appear similar to each other, then they are not equal. These set of rules



**Fig. 4.** Prioritized bilattice employed in our system

<sup>3</sup> The number of levels depend on the number and type of default rules. In our system and for the environment we are observing as we shall see in subsequent sections, there is no justification to have more than three levels.

are required in situations where we are forced to compare individuals in the absence of any contextual information. Assume an individual disappears from view into an open world (a world with no constraints on the movements of that individual or others) and another person reappears. Since the person reappearing could potentially be anyone in the world, there is significant uncertainty associated with making an identity decision. Therefore, these rules provide us with least information compared to any approach that augments appearance matching with context. We therefore assign to it priority level 1<sup>4</sup>

**Priority Level 2** If a number of individuals are observed entering a closed world and later reappearing, the uncertainty associated with performing appearance matching as before on that limited group of people is significantly lesser than in the previous case. Therefore, this rule, which reduces the space of possible matches via a closed world assumption, provides more information than pure appearance matching and we assign to it priority level 2.

**Priority Level 3** Most of the rules based on possession and knowledge fall in this category as they cause us to depart from comparing groups of individuals to comparing just two individuals. For example, if we observe an individual arrive in the scene in a vehicle, disappear from view and subsequently another individual appears in the scene and uses a key<sup>5</sup> to enter the vehicle, we can conclude, provided they appear similar, that they must be equal. Here we are comparing just two individuals the one who arrived in the vehicle and the one departing in it. Similar reasoning can be applied to offices which require a key or a combination number to enter<sup>6</sup>. Since the comparisons here involve an even more reduced set than the previous case, we assign to this set of rules priority level 3

Another set of rules that fall in this prioritization are purely closed world based rules such as an individual entering a closed world that we believe to be empty and subsequently exiting it such that no other individual is observed entering or exiting the closed world in between. Here, since there exists the possibility of the individual changing their attire inside the closed world (taking off a jacket), appearance matching is not a strong cue. Other rules in this category are rules that state that if we observe an individual enter a closed world and if, while we believe he is still inside, we observe another individual elsewhere in the scene, then they cannot be equal to each other. Closed world rules such as these clearly have more information than rules with priority levels 1 and 2; however it isn't clear that they have more or less information than the knowledge and possession based rules mentioned above. Therefore we assign to these set of rules priority level 3.

---

<sup>4</sup> Note, in our system we employ color histogram based appearance matching which by itself is a poor biometric, however if one were to employ a more powerful biometric system such as fingerprint recognition or even high resolution face recognition, then such a cue would possibly figure higher up in the bilattice.

<sup>5</sup> At present we do not directly recognize an action like using a key. Also, many vehicles have remote door locks which do not require a physical key. The fact that the individual uses a key is a default assumption. We assume that if the individual purposefully walks to the vehicle and enters it, he probably has a key. This is in contrast to loitering around the car for a while or moving from car to car, and then entering one.

<sup>6</sup> provided we have reason to believe that the office usually has only one occupant

**Definite rules** It is very hard to state that two individuals are definitely equal based on visual observation alone. Irrespective of how much information one packs in such rules, it is always possible to find ways to defeat them. Therefore, in our system we do not have a single rule that definitely infers equality. However, it is possible to state that two individuals are not equal. We do that when we observe them as two distinct individuals at the same instant of time<sup>7</sup>. We also consider the equality axioms of reflexivity, transitivity and symmetry to be definite in nature.

## 6 Activity Recognition

We can now use inferences made regarding equality of individuals to reason about the occurrence of various activities in the input video. Moreover we can propagate our degree of belief in the identity statement to the activities that it contributes to. We define three such activities and list some sample rules<sup>8</sup>.

**Theft:** We define theft as the activity of an individual possessing a package that does not belong to him. A package does not belong to an individual  $P_1$  at time  $T_1$  if it belonged to another individual  $P_2$  at some time  $T_2 < T_1$  such that  $\neg equal(P_1, P_2)$ . Formally,

$$\begin{aligned} theft(P_1, B, T_1) &\leftarrow human(P_1), bag(B), possess(P_1, B, T_1), \neg belongs(B, P_1, T_1). \\ \neg theft(P_1, B, T_1) &\leftarrow human(P_1), bag(B), possess(P_1, B, T_1), belongs(B, P_1, T_1). \end{aligned} \quad (4)$$

A package does not belong to an individual  $P_1$  at time  $T_1$  if it was originally possessed by individual  $P_2$  at some time  $T_2 < T_1$  such that  $\neg equal(P_1, P_2)$ .

$$\begin{aligned} \neg belongs(B, P_1, T_1) &\leftarrow original\_possessor(P_2, B, T_2), T_2 < T_1, \neg equal(P_1, P_2). \\ belongs(B, P_1, T_1) &\leftarrow original\_possessor(P_2, B, T_2), T_2 < T_1, equal(P_1, P_2). \end{aligned}$$

**Entry Violation:** Assuming an identity card reader controls access to a building entrance, we define entry violation as the activity of an individual entering the building without scanning his card. Formally,

$$\begin{aligned} \neg entry\_violation(P_1) &\leftarrow enter(P_1, T_1), scancard(P_2, T_2), T_2 < T_1, equal(P_1, P_2). \\ entry\_violation(P_1) &\leftarrow not(\neg entry\_violation(P_1)). \end{aligned} \quad (5)$$

**Unattended Package:** We define a package to be unattended if we observe an individual drop off a package and then cease to be in its vicinity. This is captured by the following rules

<sup>7</sup> The assumption is there are no mirrors in our world. Reflective surfaces such as glass windows never act like true mirrors, thereby giving the individual's reflection a different appearance

<sup>8</sup> Note, due to space constraints, rules listed in this paper are only those pertinent to the scenarios described in the next section and represent a small (modified for ease of understanding) subset of the rules encoded in the system. Typically for any predicate  $p$ , there exist multiple rules deriving  $p$  and/or  $\neg p$  depending on how we want the system to behave under various scenarios.

$$\begin{aligned} unattended(B, T_1) &\leftarrow not(\neg unattended(B, T_1)). \\ \neg unattended(B, T_1) &\leftarrow in\_vicinity(P_1, B, T_1), dropoff(P_2, B, T_2), equal(P_1, P_2). \end{aligned} \quad (6)$$

Propagation of belief states from equality statements to these activities is done using equation 3.

## 7 Experiments

Our system has been implemented as a multi-threaded, C++ application capable of handling multiple cameras. A Prolog reasoning engine has been embedded within this C++ application. Multivalued default reasoning is implemented using meta-predicates provided by Prolog. As currently implemented, this application runs at frame rate while taking input from up to three cameras.

The application consists of two kinds of threads: the (possibly multiple) camera thread(s) which take input from the camera(s) and detect “atomic” events (like entering a door or picking up a bag) and a single reasoning thread responsible for the high level multivalued default reasoning. For each camera connected to the system, we create a camera thread that first performs background subtraction and tracking on the video. It then detects “atomic” events and syntactically structures them as Prolog facts. The reasoning thread, when first created, starts the Prolog engine and initializes it by inserting into its knowledge base all the predefined rules from the default theory. The reasoning thread is subsequently evoked every few seconds. Every time it runs, it assimilates Prolog facts generated by the camera threads and inserts them into the Prolog engine’s knowledge base. Also, for every human observed in the video, it reasons about their identity by applying all applicable equality rules. Finally, equality statements, along with their qualitative confidence values, are used to reason about the occurrence of predefined activities using the rules listed in section 6. If any of the activities can be proven with belief states of  $DT_1$ ,  $DT_2$ ,  $DT_3$  or  $T$  then the reasoning thread generates an alert.

The tool we have built also allows the user to manually click on the image, while setting up the system, to mark and label regions (as ‘closed world’, ‘hand-off region’, ‘card reader’ etc.), in the scene. These regions, as seen in Fig 5 and 6 provide the system with information about the scene structure and properties and also helps the system to recognize a richer set of “atomic” events that log the interactions of individuals with the environment.

### 7.1 Scenarios

We demonstrate our system in action on a multi-camera surveillance setup. We employ cameras that have disjoint fields of view and label certain regions within the scene as hand-off regions. Hand-off regions are areas within an image where individuals disappear and reappear between cameras. We encode simple rules that state that if an individual disappears from the hand-off region in one camera and within a certain time interval appears within a specific hand-off region of another camera and the two individuals appear similar, then they must be equal. These rules as well as the belief states assigned to them are clearly setup specific.

We now describe a few scenarios that were used to test the system and describe how the system performed.

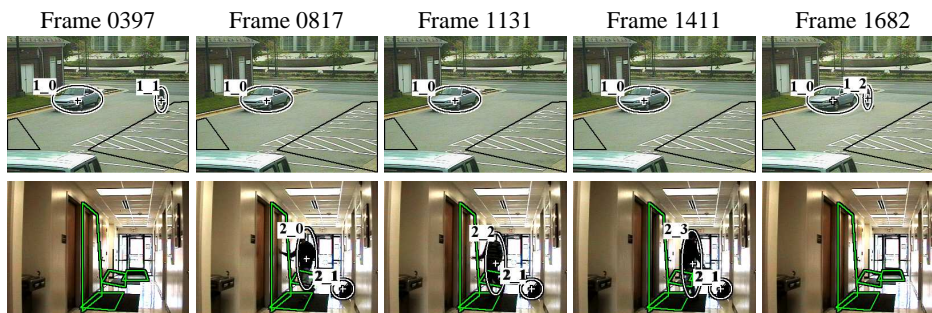


Fig. 5. Figure depicting scenario 1. Top row Camera 1 and bottom row Camera 2

**Scenario 1 (Theft-See Figure 5 and supplemental video)** Vehicle 1.0 enters the scene and individual 1.1 appears from it and disappears from the view of camera 1 from the right hand-off region. He appears in view of camera 2 from its hand-off region as 2.0, drops a bag, 2.1, in the corridor and enters a room (closed world). He is followed by another individual 2.2 (who appears from around the corner) into the room. Subsequently an individual 2.3 exits the room, picks up the bag and exits the view of camera 2 through the hand-off region. He appears in the hand-off region of camera 1 as 1.2 and enters the vehicle using a key and drives away.

In this scenario, the system correctly identifies human 2.0 as being equal to 1.1 due to the hand-off rules encoded for this camera setup. When human 2.3 exits the room, the system attempts to apply the closed world and appearance matching (default priority 2) set of rules mentioned in section 5. However, it turns out 2.3 appears similar to both 2.0 and 2.2, and therefore the system derives both  $\phi[\text{equal}(2.3, 2.0)] = DT_2$  and  $\phi[\text{equal}(2.3, 2.2)] = DT_2$ . Note the system can also prove  $\phi[\text{equal}(2.0, 2.2)] = DF_3$  which is inconsistent if we attempt to establish the transitivity relation. The system therefore is forced to assign  $\phi[\text{equal}(2.3, 2.0)] = *_2$  and  $\phi[\text{equal}(2.3, 2.2)] = *_2$  which represents the undecided state. When 2.3 picks up the bag left behind by 2.0, the system tries to prove whether or not a theft has taken place, however, it can only prove  $\phi[\text{theft}(2.3, 2.1, 1415)] = *_2$  due to the uncertainty involved in the equality statement that contributes to it. The system continues on to correctly conclude that human 2.3 is equal to human 1.2. However, when 1.2 uses a key and enters the vehicle, it can now prove  $\phi[\text{equal}(1.1, 1.2)] = DT_3$ . By transitivity, the system is then able to revise its belief of  $\phi[\text{equal}(2.3, 2.0)]$  from  $*_2$  to  $DT_3$  and consequently revise its belief of  $\phi[\text{theft}(2.3, 2.1, 1415)]$  from  $*_2$  to  $DF_3$ , i.e. no theft has occurred with high confidence.

In the next scenario, we assume there exists a card reader controlling access to a building.

**Scenario 2 (Entry Violation)** Individual 2 approaches the card reader and swipes her card while 1 is at the phone. Individuals 1 and 2 momentarily occlude each other causing the tracker to lose track of the individuals. Subsequently when the two individuals separate out again, tracking is resumed and human 3 enters the building.

In this scenario, after tracking is lost and resumed, the system needs to ascertain whether the person who entered the building is the one who swiped the card. However due to a lack of any context based cues, it is forced to resort to appearance matching (priority level 1) rules. Based on those rules, the system concludes  $\phi[equal(2, 3)] = DT_1$  and  $\phi[entry\_violation(3)] = DF_1$ , i.e. no entry violation has taken place with low confidence.



**Fig. 6.** Figure depicting scenario 2.

**Scenario 3 (Unattended Package)** *Human 2\_16 drops a bag 2\_17 in the corridor and enters an empty room (closed world). Subsequently 2\_18 exits the room.*

In this scenario, the event of 2\_16 entering the room, triggers the unattended package alert as the bag's owner is no longer in its vicinity. However, when 2\_18 appears, based on the closed world (priority level 3) rules, the system is able to conclude  $\phi[equal(2_{16}, 2_{18})] = DT_3$  and therefore it also concludes  $\phi[unattended(2_{17}, 1783)] = DF_3$ , i.e. the bag is not unattended with high confidence.

## 7.2 Complexity

Traditional default logics are computationally intractable. In traditional default logic, inferences can only be made if they are consistent with the current knowledge base. Consistency checks in default logics are a primary source of intractability and are required because the traditional theory does not permit inconsistent information to persist. In our framework, however, since the truth values are really only an agent's belief state about the world, we relax the consistency condition and allow for seemingly contradictory information to persist. Our framework therefore avoids explicit consistency checks.

Another source of intractability for traditional default logics is the method of choosing a consistent set of propositions entailed by the default theory from the set of all its extensions. Regardless of what technique is adopted to achieve this, enforcing the consistency constraint requires one to generate and inspect all possible extensions of the default theory. Note, given  $n$  defaults, there are potentially  $n!$  extensions that need to be examined. We avoid this source of intractability, again, by relaxing the consistency constraint and believing everything our theory tells us (albeit with different degrees of belief). The effect this achieves is that different extensions of our default theory are collapsed into a single solution. This makes sense because we treat our default rules as different sources of information, none of which can be completely discarded, and combine them in an information centric manner. A more disciplined and formal analysis of the complexity of the proposed theory is part of our future work.

## 8 Summary

The problem of identity maintenance is a very important problem in visual surveillance. Many activities that we wish to recognize in surveillance video depend, in some ways, upon the identities of the individuals involved, and therefore have to account for the uncertainty in reasoning about them. Traditionally, identity maintenance has relied solely on appearance matching, however it is extremely important to take into account context and cues provided by certain self-identifying actions to augment reasoning. This work is an attempt to provide a framework to do just that. The development of this framework has been heavily influenced by human reasoning. We believe human reasoning is characterized, among other things, by nonmonotonicity, qualitative belief gradation and prioritization. We have attempted to capture these traits in the proposed theory.

## References

1. Nakajima, C., Pontil, M., Heisele, B., Poggio, T.: Full-body person recognition system. *Pattern Recognition* **36** (2003) 1997–2006
2. BenAbdelkader, C., Cutler, R., Davis, L.: Motion-based recognition of people in eigengait space. In: *Proc of Intl. Conf. on Auto Face and Gesture Recogn.* (2002) 267
3. Zhou, S., Krueger, V., Chellappa, R.: Probabilistic recognition of human faces from video. *Comput. Vis. Image Underst.* **91** (2003) 214–245
4. McCarthy, J.: *Artificial intelligence, logic and formalizing common sense.* *Philosophical Logic and Artificial Intelligence* (1989)
5. Krumm, J., Harris, S., Meyers, B., Brumitt, B., Hale, M., Shafer, S.: Multi-camera multi-person tracking for easy living. *Proc. 3rd IEEE Intl Workshop on Visual Surveillance* (2000)
6. Wei, G., Petrushin, V., Gershman, A.: Multiple-camera people localization in a cluttered environment. *The 5th International Workshop on Multimedia Data Mining* (2004)
7. Starner, T., Pentland, A.: Real-time american sign language recognition from video using hidden markov models. In: *Proc of the Intl Symposium on Computer Vision.* (1995)
8. Wilson, A.D., Bobick, A.F.: Recognition and interpretation of parametric gesture. In: *ICCV.* (1998) 329–336
9. Ivanov, Y.A., Bobick, A.F.: Recognition of visual activities and interactions by stochastic parsing. *IEEE Trans. Pattern Anal. Mach. Intell.* **22** (2000) 852–872
10. Buxton, H., Gong, S.: Advanced Visual Surveillance using Bayesian Networks. In: *International Conference on Computer Vision, Cambridge, Massachusetts* (1995)
11. Intille, S.S., Bobick, A.F.: A framework for recognizing multi-agent action from visual evidence. In: *Proceedings of the sixteenth NCAIIAAI.* (1999) 518–525
12. Rota, N.A., Thonnat, M.: Activity recognition from video sequences using declarative models. *14th ECAI 2000 Berlin Germany* (2000)
13. Cohn, A.G., Magee, D., Galata, A., Hogg, D., Hazarika, S.M.: Towards an architecture for cognitive vision using qualitative spatio-temporal representations and abduction. In: *Spatial Cognition III.* (2002)
14. Bremond, F., Thonnat, M.: A context representation for surveillance systems. In: *ECCV Workshop on Conceptual Descriptions from Images.* (1996)
15. Hongeng, S., Nevatia, R., Bremond, F.: Video-based event recognition: activity representation and probabilistic recognition methods. *CVIU* **96** (2004) 129–162
16. Vu, V., Bremond, F., Thonnat, M.: Automatic video interpretation: A novel algorithm for temporal scenario recognition. *The Eighteenth IJCAI '03* (2003)
17. Fernyhough, J., Cohn, A., Hogg, D.: Building qualitative event models automatically from visual input. *Proc. ICCV* (1998) 350–355

18. Shet, V., Harwood, D., Davis, L.: VidMAP: Video Monitoring of Activity with Prolog. IEEE International Conference on Advanced Video and Signal based Surveillance (2005) 224–229
19. Reiter, R.: A logic for default reasoning. *Readings in nonmonotonic reasoning* (1987) 68–93
20. Horty, J.: Skepticism and floating conclusions. *Artificial Intelligence* **135** (2002) 55–72
21. Brewka, G.: Adding priorities and specificity to default logic. In: JELIA '94: Proceedings of the European Workshop on Logics in Artificial Intelligence, Springer-Verlag (1994) 247–260
22. Ginsberg, M.L.: Multi-valued logics: a uniform approach to reasoning in artificial intelligence. *Computational Intelligence* **4** (1988) 265–316