Object Identification in a Bayesian Context*

Timothy Huang, Stuart Russell
Computer Science Division
University of California
Berkeley, CA 94720, USA
{tthuang,russell}@cs.berkeley.edu

Abstract

Object identification—the task of deciding that two observed objects are in fact one and the same object—is a fundamental requirement for any situated agent that reasons about individuals. Object identity, as represented by the equality operator between two terms in predicate calculus, is essentially a first-order concept. Raw sensory observations, on the other hand, are essentially propositional—especially when formulated as evidence in standard probability theory. This paper describes patterns of reasoning that allow identity sentences to be grounded in sensory observations, thereby bridging the gap. We begin by defining a physical event space over which probabilities are defined. We then introduce an *identity criterion*, which selects those events that correspond to identity between observed objects. From this, we are able to compute the probability that any two objects are the same, given a stream of observations of many objects. We show that the *appearance probability*, which defines how an object can be expected to appear at subsequent observations given its current appearance, is a natural model for this type of reasoning. We apply the theory to the task of recognizing cars observed by cameras at widely separated sites in a freeway network, with new heuristics to handle the inevitable complexity of matching large numbers of objects and with online learning of appearance probability models. Despite extremely noisy observations, we are able to achieve high levels of performance.

1 Introduction

Object identification—the task of deciding that two observed objects are in fact one and the same object—is a fundamental requirement for any situated agent that reasons about individuals. Our aim in this paper is to establish the patterns of reasoning involved in object identification. To avoid possibly empty theorizing, we couple this investigation with a real application of economic significance: identification of vehicles in freeway traffic. Each refinement of the theoretical framework is illustrated in the context of this application. We begin with a general introduction to the identification task. Section 2 provides a Bayesian foundation for computing the probability of identity. Section 3 shows how this probability can be expressed in terms of appearance probabilities, and Section 4 describes our implementation. Finally, Section 5 presents experimental results in the application domain.

*This work was sponsored by JPL's New Traffic Sensor Technology program and by California PATH under MOU 152 and 214.

1.1 Conceptual and theoretical issues

The existence of individuals is central to our conceptualization of the world. While *object recognition* deals with assigning objects to categories, such as 1988 Toyota Celicas or adult humans, *object identification* deals with recognizing specific individuals, such as one's car or one's spouse. One can have specific relations to individuals, such as ownership or marriage. Hence, it is often important to be fairly certain about the identity of the particular objects one encounters.

Formally speaking, identity is expressed by the equality operator of first-order logic. Having detected an object C in a parking lot, one might be interested in whether $C = MyCar$. Because mistaken identity is always a possibility, this becomes a question of the *probability of identity*: $P(C = MyCar|$ all available evidence). There has been little work on this question in AI.[1] The approach we will take (Section 2) is the standard Bayesian approach: define an event space, assign a prior, condition on the evidence, and identify the events corresponding to the truth of the identity sentence. The key step is the last, and takes the form of an *identity criterion*. Once we have a formula for the probability of identity, we must find a way to compute it in terms of quantities that are available in the domain model. Section 3 shows that one natural quantity of interest is the *appearance probability*. This quantity, which covers diverse domain-specific phenomena ranging from the effects of motion, pose, and lighting to changes of address of credit applicants, seems to be more natural and usable than the usual division into sensor and motion models, which require calibration against ground truth.

[1] In contrast, reasoning about category membership based on evidence is the canonical task for probabilistic inference. Proposing that *MyCar* is just a very small category misses the point.

1.2 Application

Figure 1: Images from two surveillance cameras roughly two miles apart on Highway 99 in Sacramento, California. The top image is from the upstream camera, and the bottom image is from the downstream camera. The boxed vehicle has been identified at both cameras.

The authors are participants in *Roadwatch*, a major project aimed at the automation of wide-area freeway traffic surveillance and control [Malik and Russell, 1997]. Object identification is required for two purposes: first, to measure *link travel time*—the actual time taken for traffic to travel between two fixed points on the freeway network; and second, to provide *origin/destination* (O/D) counts—the total number of vehicles traveling between any two points on the network in a given time interval. The sensors used in this project are video cameras placed on poles beside the freeway (Figure 1). The video streams are processed at each camera site by vehicle tracking software running on customized parallel hardware. The resulting streams of chronologically ordered vehicle reports are sent to the TMC (Traffic Management Center). The TMC uses these reports to determine when a vehicle detected at one camera has reappeared at another. These matches are used to build up a path for each vehicle as it travels through the freeway network. The set of paths can be queried to compute link travel times and O/D counts as desired. The output of the system is a traffic information display, updated in real time for use by traffic operations managers or by individual drivers.

Obviously, a license-plate reader would render the vehicle identification task trivial, but for political and technical reasons, this is not feasible. In fact, because of very restricted communication bandwidth, the vehicle reports sent to the TMC can contain only about one hundred bytes of information. In addition, the measurements contained in the reports are extremely noisy, especially in rainy, foggy, and night-time conditions. Thus, with thousands of vehicles passing each camera every hour, there may be many possible matches for each vehicle. This leads to a combinatorial problem—finding most likely consistent assignments between two large sets of vehicles—that is very similar to that faced in *data association*, a form of the object identification problem arising in radar and sonar tracking. Section 4 explores this connection in more detail. We adopt a solution from the data association literature, but also introduce a new "leave-one-out" heuristic for selecting reliable matches. This, together with a scheme for online learning of appearance models to handle changing viewing and traffic conditions, yields a system with performance good enough for practical deployment (Section 5).

2 Inferring identity from observations

This section shows how the probability of identity can be defined in terms of physical observations and events. We begin with the formal framework and then illustrate it in the traffic domain.

2.1 Formal Bayesian framework

Let $\mathbf{O}$ be a random variable whose domain is the set of complete observation histories. That is, any particular value of $\mathbf{O}$ might correspond to the complete set of observations of objects made by some agent: $\mathbf{O} = \{o_1, \ldots, o_n\}$. Let o_a and o_b be two specific observations made, which we may think of as having been caused by objects a and b.[2] Informally, we may write the probability of identity of a and b as $P(a = b | \mathbf{O} = \{o_1, \ldots, o_n\})$.[3]

To make this probability evaluable, we define an event space $\mathbf{S} = \langle H_1, \ldots, H_N \rangle$, where each H_k is a random variable denoting the "life history" of the kth object in the universe, and each event is an N-tuple specifying the life history of all N objects. We can think of the index k as an invisible "object identification number." We impose a prior distribution $P(\mathbf{S})$, with the restriction that the prior is exchangeable, i.e., invariant under any permutation of object indices. Also, we will use the notation $o_i \in H_k$ to mean that observation o_i was generated by the kth object.

[2] For the purposes of this paper, we will assume that each observation corresponds to exactly one physical object. This assumption can be relaxed, at the cost of introducing into the theory the mechanism whereby objects generate observations.

[3] One is tempted to write this as $P(a = b | o_1, \ldots, o_n)$, i.e., to condition on a conjunction of the "positive" observations. However, conditioning on the positive observations is not the same as conditioning on both positive observations *and negative observations*—that is, observations of no vehicles at a given time and place. The temptation therefore reflects a natural "semi-closed-world" assumption: one assumes that the stated positive observations are all that have been made *in the past*, and that all other observations were negative. Obviously, one does not make this assumption regarding the future.

Now the key step is to provide an *identity criterion* to select those events corresponding to a and b being the same object. We write this as

$$a = b \iff \bigvee_k [(o_a \in H_k) \wedge (o_b \in H_k)]$$

That is, the two observed objects are the same if each observation was generated by the life history of the same object. This is the basic step in relating propositional observations to identity sentences.

Since the propositions in this disjunction are mutually exclusive, we have

$$\begin{aligned} P(a = b|\mathbf{O}) &= \frac{P(a = b, \mathbf{O})}{P(\mathbf{O})} \\ &= \frac{1}{P(\mathbf{O})} \sum_k P((o_a \in H_k) \wedge (o_b \in H_k), \mathbf{O}) \end{aligned}$$

Conditioning on the event space $\mathbf{S}$ yields

$$P(a = b|\mathbf{O}) = \frac{1}{P(\mathbf{O})} \int_{\mathbf{s} \in \mathbf{S}} \big(\sum_k P(o_a, o_b \in H_k, \mathbf{O}|\mathbf{s}) \big) P(\mathbf{s}) d\mathbf{s}$$

where $o_a, o_b \in H_k$ is shorthand for $(o_a \in H_k) \wedge (o_b \in H_k)$. Finally, we expand $P(o_a, o_b \in H_k, \mathbf{O}|\mathbf{s})$ to obtain

$$\frac{1}{P(\mathbf{O})} \int_{\mathbf{s} \in \mathbf{S}} \big(\sum_k P(\mathbf{O}|o_a, o_b \in H_k, \mathbf{s}) P(o_a, o_b \in H_k|\mathbf{s}) \big) P(\mathbf{s}) d\mathbf{s} \tag{1}$$

In this way, we express the probability of identity in terms of the probability of observations given events. We now make this framework more concrete in the context of the traffic domain.

2.2 Identity in the traffic domain

The basic vehicle identification task involves two cameras on a freeway, one upstream and one downstream. The TMC monitors a vehicle's progress by matching the appropriate upstream observation of the vehicle with the appropriate downstream observation. It is natural to keep the observation history $\mathbf{O}$ separated in two parts, $\mathbf{U}$ (observations by upstream camera u) and $\mathbf{D}$ (observations by downstream camera d). Each observation $o_i^u \in \mathbf{U}$ of vehicle i is a pair (r_i^u, f_i^u), where r_i^u includes the location of the vehicle and the time of observation, while f_i^u corresponds to observed values of intrinsic features of the vehicle, such as color and size. The same holds for observation $o_j^d \in \mathbf{D}$ of vehicle j. It is reasonable to assume that each r is unique, since a vehicle cannot be in two places at once, nor can two vehicles be in the same place at once.

H_k can be thought of as a *trajectory* for the kth vehicle, specifying its position as a function of time. The identity criterion for observed vehicles a and b now becomes

$$a = b \iff \bigvee_k \big[(r_a^u \in H_k) \wedge (r_b^d \in H_k)\big]$$

To illustrate Eq. (1), consider the simplified case where the universe contains exactly two vehicles of similar appearance, each moving at constant velocity along the same road. Two reliable cameras, located at y_u and y_d, make observations whenever vehicles pass by. From time $t = 0$ to $t = T$, the observation history $\mathbf{O}$ consists of $\mathbf{U} = \{o_a^u\}$ recorded at t_a and $\mathbf{D} = \{o_b^d\}$ recorded at t_b. The event space $\mathbf{S} = \langle H_1, H_2\rangle$ therefore ranges over all possible trajectories for two vehicles; these can, in turn, be defined by the random variables $\langle y_1, v_1\rangle$ and $\langle y_2, v_2\rangle$, where y_1 and y_2 are the positions of the vehicles at time t_a and v_1 and v_2 are the velocities. Figure 2 shows two possible events consistent with the observations—one where $a = b$ and one where $a \neq b$.

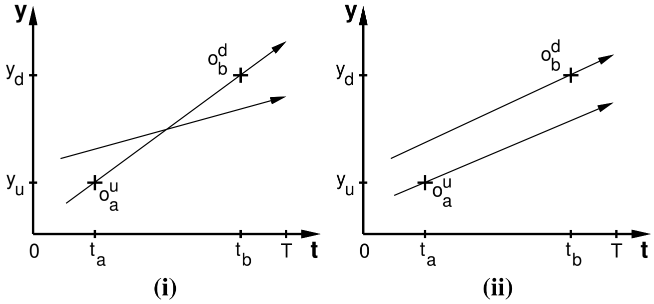

Figure 2: Time space diagrams showing two possible events in the two-car universe, given observations o_a^u and o_b^d (and no other observations). In (i), we have $a = b$, while in (ii), we have $a \neq b$.

It turns out to be easiest to compute the probability of identity from the ratio

$$\frac{\int_{\mathbf{s} \in \mathbf{S}} P(a = b, \mathbf{O}|\mathbf{s}) P(\mathbf{s}) d\mathbf{s}}{\int_{\mathbf{s} \in \mathbf{S}} P(a \neq b, \mathbf{O}|\mathbf{s}) P(\mathbf{s}) d\mathbf{s}}$$

since the $P(\mathbf{O})$ term in Eq. (1) cancels. Suppose now that $y_d - y_u$ is 2000 meters, and that $t_b - t_a$ is 100 seconds. Furthermore, suppose $P(\mathbf{S})$ is such that y_1 and y_2 are independently and uniformly distributed in the range $[y_u, y_d]$, and v_1 and v_2 are independently and uniformly distributed in the range $[10m/s, 40m/s]$. Then $P(\mathbf{s})$ is constant over the range of integration and the above ratio is given by

$$\frac{\int_{y_u}^{y_d - 10(t_b - t_a)} \int_{10}^{\frac{y_d - y_2}{t_b - t_a}} dv_2 dy_2}{\int_{10}^{\frac{y_d - y_2}{t_b - t_a}} dv_1 \int_{y_u}^{y_d - 10(t_b - t_a)} dy_2} = \frac{1}{2}$$

hence $P(a = b|\mathbf{O}) = 1/3$.

3 Appearance models

The previous section showed how to express the probability of identity in terms of the probability of observations given events. Some domains, including traffic surveillance, involve observation sets that contain initial observations of objects as well as subsequent observations of objects. In these situations, appearance probabilities, which define how objects observed at some point in the past can be expected to appear at some point in the future, seem to provide a more usable model than standard motion and sensor models. In this section, we show how to express Eq. (1) in terms of appearance probabilities and describe the specific appearance probabilities used in the vehicle identification domain.

3.1 Identity in terms of appearance probabilities

Suppose we are given observation history $\mathbf{O} = \mathbf{U} \cup \mathbf{D}$, where $\mathbf{U}$ consists of initial observations of objects, and $\mathbf{D}$ consists of subsequent observations of objects. Keeping these sets separate in Eq. (1) gives the following for $P(a=b|\mathbf{U},\mathbf{D})$:

$$\frac{1}{P(\mathbf{U},\mathbf{D})}\int_{\mathbf{s}\in\mathbf{S}}\big(\sum_k P(\mathbf{U},\mathbf{D}|o_a,o_b\in H_k,\mathbf{s}) \\ P(o_a,o_b\in H_k|\mathbf{s})\big)P(\mathbf{s})d\mathbf{s}$$

Expanding $P(\mathbf{U},\mathbf{D}|o_a,o_b\in H_k,\mathbf{s})$ yields

$$\frac{1}{P(\mathbf{U},\mathbf{D})}\int_{\mathbf{s}\in\mathbf{S}}\big(\sum_k P(\mathbf{D}|\mathbf{U},o_a,o_b\in H_k,\mathbf{s})P(\mathbf{U}|o_a,o_b\in H_k,\mathbf{s}) \\ P(o_a,o_b\in H_k|\mathbf{s})\big)P(\mathbf{s})d\mathbf{s}$$

At this point, we define a new event space Ω, which is a coarsening of $\mathbf{S}$. First, we need some new terms: A *matching* is a simply a pairing indicating that two observations were generated by the same object. For example, the matching (a,b) indicates that the same object generated o_a and o_b. Given a set of initial observations and a set of subsequent observations, an *assignment* is a set of matchings for every observation in each set. The *assignment space* Ω is an event space that ranges over the space of possible assignments. It thus divides the space $\mathbf{S}$ into subsets of events, such that each subset is relatively *homogeneous* if we condition on the observations—because each subset then effectively specifies that starting point and ending point of each vehicle's trajectory.

Summing over only those assignments ω that satisfy the identity criterion for a and b gives

$$P(a=b|\mathbf{U},\mathbf{D}) = \frac{1}{P(\mathbf{U},\mathbf{D})}\sum_{\omega\in\Omega:(a,b)\in\omega} P(\mathbf{D}|\mathbf{U},\omega)P(\mathbf{U}|\omega)P(\omega)$$

Since the principle of exchangeability requires a uniform prior over Ω, and since $P(\mathbf{U}|\omega)$ is constant given no information about the observations to which the observations in $\mathbf{U}$ correspond, these constant terms can be grouped outside of the summation along with the normalization constant $P(\mathbf{U},\mathbf{D})^{-1}$ so that we have

$$P(a=b|\mathbf{U},\mathbf{D}) = c\big(\sum_{\omega\in\Omega:(a,b)\in\omega} P(\mathbf{D}|\mathbf{U},\omega)\big)$$

Finally, we assume that the probability of a specific subsequent observation, given a specific initial observation and a matching between the two observations, is independent of the other observations and matchings. Hence, we can factor $P(\mathbf{D}|\mathbf{U},\omega)$ into the product of the individual probabilities so that the identity equation becomes

$$P(a=b|\mathbf{U},\mathbf{D}) = c\big(\sum_{\omega\in\Omega:(a,b)\in\omega}\big[\prod_{(i,j)\in\omega} P(o^d_j|o^u_i,i=j)\big]\big) \quad (2)$$

In this expression, $P(o^d_j|o^u_i,i=j)$ is an *appearance probability*, the probability that an object that initially generated observation o^u_i subsequently generated observation o^d_j. We will write this as $P(o^d|o^u)$ where no confusion is possible. It is important to note that the appearance probability is *not* the probability that $i=j$. This probability can only be computed by taking into account all other observations.

To ground this discussion, we will now discuss the specific observed features and appearance probability models used in the traffic domain.

3.2 Observed features for traffic

When a certain camera c observes some vehicle i, it generates a vehicle report consisting of various *features*. Thus, the observation o^c_i in our system is a vector of features. Currently, we use the features in this table:

Name	Description
t	time of observation
x	lane position (1, 2, 3, etc.)
y	forward position in lane
$\dot{x}$	lateral velocity
$\dot{y}$	forward velocity
w	vehicle width
l	sum of vehicle length and height
h	mean vehicle color hue
s	mean vehicle color saturation
v	mean vehicle color value
C	histogram of color distribution over vehicle pixels

The matching algorithm is designed to be independent of the specific features used; new features of arbitrary complexity, informativeness, and noise level can be added without changing the algorithm.

3.3 Appearance probabilities for traffic

The appearance probability is currently treated as the product of the following independent models:

- lane (x): discrete distribution $P(x^d|x^u)$
- size (w, l): multivariate Gaussian
  $$P(w^d,l^d|w^u,l^u) = N_{\mu_{w,l},\Sigma_{w,l}}(w^d - w^u, l^d - l^u)$$
- color (h, s, v): multivariate Gaussian
  $$P(h^d,s^d,v^d|h^u,s^u,v^u) = N_{\mu_{h,s,v},\Sigma_{h,s,v}}(h^d-h^u, s^d-s^u, v^d-v^u)$$
- arrival time (t): univariate Gaussians conditioned on upstream and downstream lane
  $$P(t^d|t^u,x^d,x^u) = N_{\mu_t^{x^d,x^u},\sigma_t^{x^d,x^u}}(t^d - t^u)$$

The arrival time model is particularly important, since it drastically reduces the number of vehicle pairs that are considered to be plausible matches. The parameters $\mu_t^{x^d,x^u}$ and $\sigma_t^{x^d,x^u}$ represent the mean and standard deviation of the predicted link travel time for cars that start upstream in lane x^u and end up downstream in lane x^d. This allows the system to accurately model, for example, the fact that cars in the car pool lane travel faster than cars in other lanes.

In examining the empirical distributions for the appearance probability, we were surprised by the level of noise and lack of correlation in measurements of the same feature at two different cameras. Some features, such as saturation, appear virtually uncorrelated. In all, we estimate that the size and color features provide only about 3 to 4 bits of information.

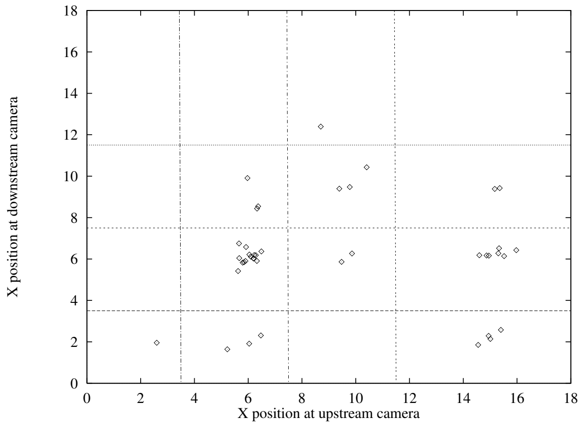

Figure 3: Diagram showing observed upstream and downstream x-position data for a sample of 41 matched vehicles from the Mack Road and Florin Road cameras. The horizontal axis corresponds to upstream x-position and the vertical axis corresponds to downstream x-position. Each marked point corresponds to a single matched vehicle. Lane dividers are shown as horizontal and vertical lines. For example, 13 vehicles are observed in lane 4 (onramp, highest x values) upstream, of which 7 end up in lane 2 (middle lane) downstream, indicating that $P(x^d = 2|x^u = 4) \approx 0.54$.

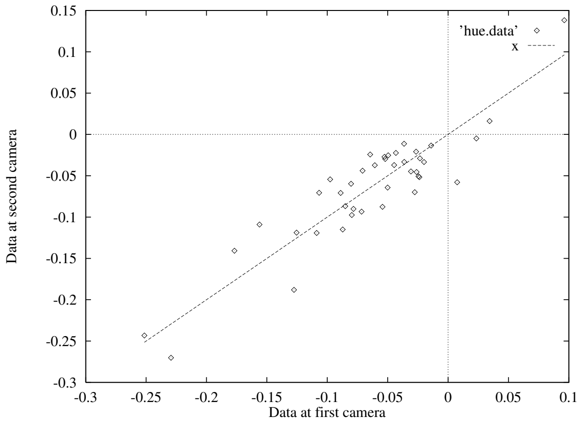

Figure 4: Diagram showing observed upstream and downstream hue data for a sample of 25 matched vehicles from the Mack Road and Florin Road cameras. The horizontal axis corresponds to upstream hue, and the vertical axis corresponds to downstream hue. The appearance probability for color, which includes hue, saturation, and value components, is modeled as a multivariate Gaussian.

3.4 Online learning of appearance models

Because traffic and lighting conditions change throughout the day, our system uses online (recursive) estimation for the appearance probability model parameters. As new matches are identified by the vehicle matcher, the parameters are updated based on the observed feature values at the upstream and downstream sites. Figure 3 shows a sample set of x values for matched vehicles, from which $P(x^d|x^u)$ can be estimated. Similarly, Figure 4 shows a sample set of hue values for matched vehicles, from which $P(h^d|h^u)$ can be estimated. To adapt to changing conditions, our system uses online exponential forgetting. For example, if a new match is found for a vehicle in lane x^u upstream and lane x^d downstream, with link travel time t, then the mean travel time is updated as follows:

$$\mu_t^{x^u,x^d} \leftarrow \gamma\mu_t^{x^u,x^d} + (1-\gamma)t$$

The γ parameter, which ranges from 0.0 to 1.0, controls the effective "window size" over which previous readings are given significant weight.

The above assumes that the match found is in fact correct. In practice, we can never be certain of this. A better motivated approach to model updates would be to use EM; this can be approximated by weighting each update by the probability that the match is correct.

4 Matching algorithm

We begin by describing the simplest case, where all vehicles detected at the upstream camera are also detected downstream, and there are no onramps or offramps. In this case, the aim is to find pairs of vehicles a and b such that $P(a=b|\mathbf{U},\mathbf{D}) > 1-\epsilon$ for some small ϵ. This problem is intractable in the worst case (it relates to computing the permanent of a matrix), so we adopt a heuristic approach.

The core of the approach is the observation, from [Cox and Hingorani, 1994], that a most probable assignment (pairing all n vehicles) can be found in time $O(n^3)$ by formulating the problem as a weighted bipartite matching problem and using any of several well-known algorithms for this task. To do this, we construct an *association matrix M* of appearance probabilities, where each entry $M_{ij} = -\log P(o_j^d|o_i^u)$, so that the assignment with least total weight in the matrix corresponds to the most probable assignment, according to Eq. (2).

For our purposes, knowing the most likely assignment is not enough. It can easily happen that some c of the n vehicles are all very similar and fairly close to each other on the freeway. In this case, there will be $c!$ assignments all having roughly the same probability as the most probable assignment. Since matches within this group may be very unreliable, we employ a *leave-one-out* heuristic that "forbids," in turn, each match contained in the best assignment. For each forbidden match, we measure the reduction in likelihood for the new best assignment. Matches whose forbidding results in a significant reduction are deemed reliable, since this corresponds to a situation where there appears to be no other reasonable assignment for the upstream vehicle in question.

For example, suppose we have the following matrix:

	Downstream		
Upstream	x	y	z
a	3.2	2.5	12.7
b	8.5	4.5	4.4
c	7.3	5.0	5.0

Here the best assignment is $\{a=x, b=z, c=y\}$, with a total weight of 12.6. (Notice that $a=y$ is the "closest" match for a, but leaves no good match for the others.) If we forbid $a=x$, the best assignment is $\{a=y, b=z, c=x\}$ with weight 14.2. If the difference between these two weights is greater than some reliability threshold t, [4] i.e., if $14.2 - 12.6 > t$, then

[4] The value compared with the reliability threshold is the negative logarithm of the relative likelihood of the observations given the best

we accept the $a=x$ match, since no other reasonable choice seems to exist. On the other hand, if we forbid $b=z$, the best assignment has weight 12.7. If $12.7 - 12.6 <= t$, then we reject $b=z$, since there is another match for b that yields a good overall assignment. By increasing the threshold t, we obtain more reliable matches, i.e., the error rate ϵ is reduced; however, this reduces the overall number of accepted matches.

4.1 New and missing vehicles

In the general case, vehicles can appear from onramps between the cameras or can disappear onto offramps. (Equivalently, they can fail to be detected at the upstream or downstream camera.) To handle this, we add extra rows and columns to the association matrix. With m upstream and n downstream vehicles, the matrix now has $m+n$ rows and columns to allow for all possibilities. Figure 5 illustrates the structure of the matrix for $m=n=2$. Here, α is the probability that a vehicle exits the freeway, β is the number of vehicles entering the freeway between the cameras per unit time, and $P(o_i)$ refers to the prior probability of seeing a vehicle with features o_i.[5] The formulas in the table explain the interesting fact that human observers feel far more confident matching unusual vehicles than typical vehicles: not only is the probability of confusion with other vehicles lower, but the probability that the upstream vehicle exited, only to be replaced by another vehicle of the same unusual appearance, can be discounted because the extra multiplicative $P(o_i)$ factor for an unusual vehicle would be tiny.

	Downstream			
Upstream	x	y	off	off
a	$(1-\alpha)P(o_x\|o_a)$	$(1-\alpha)P(o_y\|o_a)$	α	α
b	$(1-\alpha)P(o_x\|o_b)$	$(1-\alpha)P(o_y\|o_b)$	α	α
new	$\beta P(o_x)$	$\beta P(o_y)$	1	1
new	$\beta P(o_x)$	$\beta P(o_y)$	1	1

Figure 5: Extended association matrix for two upstream and downstream observations, showing additional rows and columns to account for entering and exiting vehicles. Each entry will be replaced by its negative logarithmic value before computing the minimum weight assignment.

4.2 Relationship to data association

The vehicle matching problem is closely related to the traditional "data association" problem from the tracking literature, in which new "observations" (from the downstream camera) must be associated with already-established "tracks" (from the upstream camera). Radar surveillance for air traffic control is a typical application: the radar dish determines an approximate position for each aircraft every few seconds, and each new set of positions must be associated with the set of existing tracks. There is a large literature on data association—typically over 100 papers per year. The standard text is by Bar-Shalom and Fortmann [Bar-Shalom and Fortmann, 1988], and recent developments appear in [Bar-Shalom, 1992]. Ingemar Cox [Cox, 1993] surveys and integrates various developments, deriving formulas very similar to those in Figure 5. Cox's aim in his review paper is to present the ideas from the data association field to the computer vision and robotics community, where they might be used to resolve problems of identifying visual features seen in temporally separated images.

Major differences between our work and "standard" data association include the following:

1. Sensor noise and bias are large, unknown, time-varying, site-dependent, and camera-dependent, and sensor observations are high-dimensional.
2. Successive observations of a vehicle are widely separated in time and space, and all observations are asynchronous.
3. Vehicle trajectories are highly correlated.

5 Results

We tested the vehicle matcher with data from a region-based vehicle tracker on video sequences from the sites in Figure 1.

On any given run, the number of matches proposed by the vehicle matcher depends on the reliability threshold selected for that run. In the results discussed below, *coverage* refers to the percent of vehicles observed by both cameras for which matches were proposed, and *accuracy* refers to the percent of proposed matches that were in fact correct. In general, the coverage goes down as the reliability threshold is increased, but the accuracy goes up.

To verify the accuracy of the matcher, the ground-truth matches were determined by a human viewing the digitized sequences with the aid of a frame-based movie viewer. Since this method required about 3 hours of viewing to match each minute of video, it was used only during the early stages of testing. In subsequent testing, we first ran the matcher on the vehicle report data and then used the frame-based movie viewer to verify whether the suggested matches were correct.

Testing our system involved a start-up phase during which it estimated the appearance probability models online. For the results shown in Figure 6, we trained our system on a pair of 60-second video sequences and then ran it on the immediately following 60-second sequences. The sequences contained 29 vehicles upstream, along with numerous vehicles that either entered or exited the freeway in between the cameras. The resulting accuracy/coverage curve in figure 6 shows that despite very noisy sensors, the system achieved 100% accuracy with a coverage of 14%, and 50% accuracy with a coverage of 80%. Incorporating additional features would probably improve system performance. The boxed vehicles in Figure 1 show a pair of vehicles correctly matched by our system.

Link travel times between each camera pair are currently calculated by averaging the observed travel times for matched vehicles. These times were accurate to within 1% over a distance of two miles, over a wide range of coverage/accuracy tradeoff points. This suggests that matched vehicles are representative of the traffic flow—that is, the matching process does not select vehicles with a biased distribution of speeds.

assignment and the observations given the best assignment with a forbidden match.

[5] In our implementation, each of these models is learned online; α and β are also specific to individual lanes.

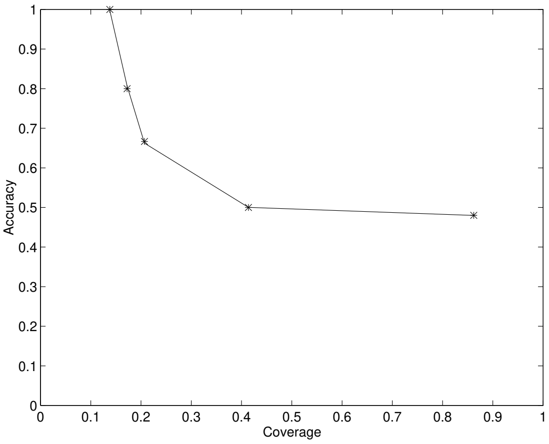

Figure 6: Sample matching results: the graph shows accuracy versus coverage for a range of reliability threshold values. A low threshold implies high coverage and low accuracy, while a high threshold implies low coverage and high accuracy.

Our results are comparable with those obtained by Petty *et al.* [Petty *et al.*, to appear], who use an "ensemble" matching approach that detects downstream propagation of distinct arrival time patterns. The two approaches are complementary, in that our approach allows identification of individual vehicles (and hence the generation of O/D counts), whereas the Petty *et al.* approach provides a robust bootstrap method for initializing our arrival time models.

6 Conclusions and further work

This paper has described the patterns of reasoning involved in establishing identity from observations of objects. We proposed a formal foundation based on a prior over the space of physical events, together with an identity criterion defining those events that correspond to observations of the same object. In the case of vehicle matching, the events are the different sets of trajectories of vehicles in a given freeway network. When a single trajectory passes through two vehicle observations, that implies that the observations correspond to the same object. This general approach makes it possible to define the probability of identity and to integrate the necessary patterns of reasoning into an intelligent agent.

This research can be seen as another step in the Carnapian tradition that views a rational agent as beginning with uninformative prior beliefs and then applying Bayesian updating throughout its lifetime. The general relationship between perception and the formation of internal models is a subject that needs much more investigation [Bacchus *et al.*, 1995].

We showed that the abstract probability of identity can be expressed in terms of measurable appearance probabilities, which define how, when, or where objects that were observed at some point in the past are expected to appear at some point in the future. These appearance probabilities can be learned online to adapt to changing conditions in the environment—such as changing weather, lighting, and traffic patterns.

We have implemented and tested a system for vehicle matching using an efficient algorithm based on bipartite matching combined with a leave-one-out heuristic. Despite very noisy feature measurements from the cameras, our system achieved a high level of accuracy in matching individual vehicles, enabling us to build the first reliable system for measuring link travel times. Although experimental camera data were not available for the system to do so, it is already capable of tracking the path of a vehicle over a sequence of camera sites. Thus, O/D counts for a time period can be computed by examining the complete set of recorded paths during that time period. For successful O/D measurement over a long sequence of cameras, however, we need to improve both matching coverage and the detection rate of the tracking subsystem.

The patterns of reasoning described here have broad applicability to other domains. For example, the object identification problem occurs in database management, where it is possible that two different records could correspond to the same entity. For example, US credit reporting agencies record over 500 million credit-using Americans, of whom only about 100–120 million are actually distinct individuals. Applying our approach to this problem could help with maintaining database consistency and with consolidating multiple databases containing overlapping information.

References

[Bacchus *et al.*, 1995] Fahiem Bacchus, Joseph Y. Halpern, and Hector Levesque. Reasoning about noisy sensors in the situation calculus. In *Proceedings of the Fourteenth International Joint Conference on Artificial Intelligence (IJCAI-95)*, pages 1933–1940, Montreal, Canada, August 1995. Morgan Kaufmann.

[Bar-Shalom and Fortmann, 1988] Yaakov Bar-Shalom and Thomas E. Fortmann. *Tracking and Data Association*. Academic Press, New York, 1988.

[Bar-Shalom, 1992] Yaakov Bar-Shalom, editor. *Multitarget multisensor tracking: Advanced applications*. Artech House, Norwood, Massachusetts, 1992.

[Cox and Hingorani, 1994] I.J. Cox and S.L. Hingorani. An efficient implementation and evaluation of Reid's multiple hypothesis tracking algorithm for visual tracking. In *Proceedings of the 12th IAPR International Conference on Pattern Recognition*, volume 1, pages 437–442, Jerusalem, Israel, October 1994.

[Cox, 1993] I.J. Cox. A review of statistical data association techniques for motion correspondence. *International Journal of Computer Vision*, 10:53–66, 1993.

[Malik and Russell, 1997] Jitendra Malik and Stuart Russell. Traffic surveillance and detection technology development: New sensor technology final report. Research Report UCB-ITS-PRR-97-6, California PATH Program, 1997.

[Petty *et al.*, to appear] K. F. Petty, P. Bickel, J. Jiang, M. Ostland, J. Rice, Y. Ritov, and F. Schoenberg. Accurate estimation of travel times from single-loop detectors. *Transportation Research, Part B*, to appear.