Traffic Signals and Autonomous Vehicles: Vision-based or a V2I Approach?

Matthew L. Ginsberg  
Connected Signals, Inc.  
355 Goodpasture Island Road, Suite 200  
Eugene, OR 97401-2119  
ginsberg@connectedsignals.com

Abstract

Autonomous vehicles operating in urban settings will need to know both the current and – ideally – the predicted states of the traffic signals they will encounter. This paper describes the various approaches that can be taken to address this problem: two vision-based approaches and two V2I-based approaches. We discuss the likely strengths and weaknesses of each, and suggest that a hybrid approach will likely be necessary in any relatively near-term deployment of autonomous technology.

Keywords: V2I, Traffic Signals, Connected Vehicles  
Topic Area: Connected/Autonomous Vehicles

Introduction

Drivers need to understand traffic lights, and this is just as true for autonomous drivers as it is for human ones. While there have been some optimistic predictions that traffic lights would vanish when autonomous driving becomes a reality (1, 2, 3), this seems a bit farfetched unless the pedestrians become automated as well. There will also obviously be an extended transition period as human drivers are gradually replaced by digital counterparts.

If autonomous vehicles are to share traffic lights as a fact of life with the rest of us, how are they to actually identify where the lights are, and whether any particular light is red, yellow, or green? At a high level, there seem to be two approaches. In one, autonomous vehicles deal with lights by observing the same traffic lights as do the rest of us, a vision system of some sort. In the other approach, the information is made available to autonomous drivers using an entirely different data stream.

Each of these approaches has at least two subtypes. A vision-based approach might use general vision, picking the lights out from the visual field and then determining their color. Alternatively, a vision system might use some more specialized technique, exploiting additional information to both reduce the complexity of the vision problem somehow and make the solutions more reliable as well.

The “data stream” approach also has two subtypes, depending on whether existing data infrastructure is used (presumably the wireless Internet, as connected cars will surely be a reality long before autonomous drivers are), or whether some new and at least partly
dedicated mechanism is used instead (such as the long-promised V2I connectivity provided by Dedicated Short-Range Communication, or DSRC).

In this paper, we discuss each of these four possibilities. While it is clear that none of them has been fully realized at this point, many partial or related implementations exist and we can draw reasonable conclusions from the successes and failures of that other work. Doing this is our goal in this paper.

Before proceeding, however, we should note that it is not our intention here to predict the future, to say, “Autonomous vehicles will interact with traffic lights in the following way.” In actuality, none of the four approaches seems likely to work in isolation and, at least initially, the makers of autonomous vehicles will need to either restrict their domains of operation or rely on a hybrid approach. Our hope is that by presenting the prospective strengths and weaknesses of each option, we can help to guide the design choices that makers of autonomous vehicles are facing.

**Vision-based Approaches**

Identifying traffic-signal state using some kind of vision system certainly appears to be the most natural approach. Unfortunately, computer vision is in general far more challenging than one would expect. The reason, roughly speaking, is that we (humans) use a tremendous amount of world knowledge in analyzing any particular visual image. As the picture from Denver below indicates, it’s hardly a matter of simply identifying red, yellow, or green circles in the field of view.

![Traffic lights in Denver](image)

**Figure 1: "Traffic lights" in Denver**

To make matters worse, computer vision systems are generally installed in environments where their apparently inevitable (albeit perhaps infrequent) errors have no catastrophic consequences. Traffic lights are not like this: an autonomous vehicle needs to get all of the lights right all of the time.
Using vision to find traffic lights involves solving two separate subproblems. First, the lights must be found in the image. Second, the states of the lights need to be determined based on the information in the image.

**Signal Location**

This is hardly the place to present an introductory course in vision processing, but at least some understanding of the techniques involved is needed if the difficulties are to make sense.

Objects whose structures are known are typically found in images by first creating a “Canny” image, which is basically a black-and-white drawing of the edges of objects in the original frame (4,5). These edges can be found because the color in the image typically changes abruptly from, for example, a yellow balloon to an off-white sky.

Having found the edges, one can look for objects of any particular shape (i.e., the circular shape of a traffic light) by looking for circles in the Canny image. A variety of techniques are available for doing this, the most common of which is known as the Hough transform (6,7).

While this all works well in theory, practice is generally not so benign. A traffic light, for example, may not be circular because its edge is obscured (perhaps by the sunlight deflector at the top of the light itself), or because perspective causes the shape to be somewhat elliptic. Computer vision methods are generally extremely sensitive to small changes such as these, and any of them can cause some particular signal to be overlooked entirely. Consider, for example, the image below, where even to a human it is not obvious at a glance which are the lights and which are not.

![Figure 2: Which are the lights, which are the cars?](image)

If the structure is not known, techniques from machine learning are generally used instead. The state of the art in computer vision is to use deep learning with convolutional neural networks (8) to extract the features of any particular image; the image itself is then classified and analyzed hierarchically as in ImageNet (9). This approach currently has the best classification rates of any system (including humans) on traffic sign recognition (10).

While Google, Baidu, Audi, Mercedes and Volvo have all announced the use of deep learning in their autonomous vehicle projects, the technique is computationally more intensive than the Canny/Hough combination and it is not clear that deep learning is required to recognize objects that have been designed to be as globally uniform as possible.
It is difficult to understand the strengths and weaknesses of any particular approach because the academic literature is quite sparse. Most players in the autonomous car space (automakers, Google, etc.) are incredibly tightlipped about the details of their techniques. John et.al (11) does discuss the use of a deep learning system for these purposes, while the more recent survey article by Jensen et.al (12) echoes our conclusion that these two competing approaches cannot yet be thoroughly compared.

It may also be helpful to examine a non-traffic system solving similar problems and about which a bit more is known. The ArcAngel system developed by Sportstech LLC (13) tracks basketballs in high-definition video in real time in order to determine whether or not a particular shot is likely to go into the hoop.

The goal here is very similar. Basketballs are round. They are a distinctive color. They are often partially obscured by the players. The information needs to be processed quickly.

What Sportstech has found is that a combination of circle detection and color pruning produces the best results, and these lessons may well apply to traffic signals as well. Sportstech has also learned that it is impractical to search entire images in real time, and ArcAngel works by searching entire images as rapidly as possible. Between the frames that are analyzed in their entirety, the system looks for objects (be they basketballs or lights) that are nearby to objects that were successfully found in the last fully analyzed frame. These “complete” and “partial” analyses are interleaved on a multi-core processor to ensure that a timely data stream is constantly available to the user.

For traffic lights, there is an alternative possibility, which is to use information from existing data sources to place the light accurately in 3-dimensional space. That 3-dimensional placement can then be combined with GPS data to determine the light’s expected position in the image. Having identified the expected location of the signal, the light can either be assumed to be at its specified location or (likely better, given the Sportstech results) the light can be assumed simply to be in a relatively small neighborhood of that location. That lights can in fact move somewhat during somewhat normal conditions is made clear in the image in Figure 3. (And note as well the extent to which the light itself no longer presents as a circle in this shot.)

While Google (and presumably others) are taking this approach (14,15), there are some circumstances (Figure 4) that simply cannot be addressed in this fashion, as Gomes has noted (16).

Figure 3: Signals moving in the wind
Figure 4: A signal that will not appear in any database

For this reason, other players appear to be using general vision techniques to ensure that temporary signals are found as reliably as permanent ones. But they must then deal with images such as:

Figure 5: Unanalyzeable?

Signal Phase

Having identified the locations of the signals, it remains to determine their colors. In general, this can be expected to be fairly straightforward. But it can, for example, be too hot or too cold:
The bottom line is that signals break. When this happens, but only when it happens, an autonomous driver needs to realize that the malfunction is in the signal and not in the vision system, and to then respond appropriately cautiously.

In Figure 6, none of the drivers at the intersection will be able to determine the state of the light in question. But in other cases, only the vehicle in question may be having problems:

![Figure 6: Temperature extremes](image)

The differences between Figures 6 and 7 are crucial. In Figure 6, every other driver or pedestrian will be aware of the problem and will be responding similarly cautiously. In Figure 7, however, other actors will not only be aware of the current phase of the light, but will probably have no idea that the vehicle confronted with the image shown is not.

![Figure 7: Lights in an unknown phase](image)

**V2I-based approaches**

Given the difficulties that will need to be addressed by a vision-based approach to this problem, it seems likely that an alternative data source should be found. Indeed, given the current enthusiasm for connecting vehicles to the outside world generally, perhaps this is a more obvious approach to take. Direct provision of information to vehicles will also allow those vehicles to ascertain the states of lights that are currently invisible because they are far
away, around a curve or otherwise obscured, and to obtain additional useful information, such as the fact that a pedestrian has pushed a button and is waiting to cross.

There appear to be two viable approaches to getting traffic light information to vehicles directly. The older one is to have the signals communicate directly with the vehicles, installing a DSRC radio at each light and a receiver on each vehicle. This approach has been in the works for approximately a decade and centers on a standard known as SPaT (Signal Phase and Timing) (17). In the United States in 2014, there were approximately 200 signalized intersections for which SPaT data was available via DSRC (18).

An alternative approach, suggested more recently, is to provide traffic light data to vehicles using existing infrastructure. Many urban traffic lights currently connect to Traffic Management Centers, or TMCs, and provide those TMCs with real-time information regarding signal state, vehicle and pedestrian calls, and so on. The TMCs are in general connected to the Internet and can push this real-time data out to third parties. This author’s company, Connected Signals, has recently gone even further, offering a free hardware device that listens to the network information flowing between the traffic lights and the TMCs and then pushes out traffic signal-related data directly. Connected Signals currently gets real-time signal information from approximately 10,000 signalized intersections worldwide.

When compared to vision-based approaches, this sort of direct data acquisition has both strengths and weaknesses. Most importantly, of course, the current state of computer vision means that data acquired directly from the signals should simply be more accurate and more robust than data acquired via a vision system. But that may not be an unmixed blessing, because it is possible for a signal to fail on the street but for the TMC to be completely unaware of the problem. Faced with the situation of Figure 6, for example, drivers on the street will realize that the signals have failed. Drivers getting their data automatically may not.

There are also important differences between the DSRC and Internet-based approaches:

1. Coverage. While any signal or vehicle can be equipped with a DSRC radio, only signals that are already connected to TMCs can be expected to provide their state over the Internet. Currently only some 100,000 of the 300,000 signalized intersections in the United States (19) are connected to TMCs; roughly speaking, it only makes sense to do so when the signals need to coordinate in some way with neighboring signals. The TMC handles synchronization among the signals on its network even though the clocks on individual signals may drift. For many rural lights, it seems unlikely that they will ever be connected to the Internet in this fashion. (The 100,000 number is based on the fact that approximately half of the 300,000 signalized intersections are in urban locations, and our experience with individual municipal agencies suggests that approximately 70% of urban lights are connected.)

2. Cost. While the primary argument against an Internet-based approach to this problem is coverage, the primary argument against the DSRC approach is simply cost. Estimates of the cost required to equip a single intersection range from $17,000 to $18,000 per intersection, plus backhaul costs of up to $40,000, and ongoing operations and maintenance costs of $2,000-3,000 annually (18); a recent Econolite FAQ (20) similarly suggested a cost of approximately $17,500 per intersection. While federal agencies have indicated a great interest in the DSRC approach generally, they have not shown a willingness to provide the approximately $5 billion in funding that would be
required to equip all of the nation’s intersections with this functionality, and local or state agencies are less likely still to have access to funds of this magnitude.

3. Protocol. One additional argument against the DSRC approach is that the SPaT protocol is relatively impoverished, including information only about the current phase of the signal in question. As autonomous drivers become more sophisticated, they will surely want access to additional information. Perhaps the simplest example is the one already mentioned involving pedestrians. If a pedestrian has pushed a button indicating a desire to cross in front of an autonomous vehicle, that vehicle should at least attempt to find that pedestrian in its field of view before proceeding through the intersection. This is true even if the autonomous vehicle has the right of way, since pedestrians are not always patient.

These latter two points, together with a variety of additional technical concerns, led to a congressional peer review of the USDOT’s report on DSRC to take a relatively pessimistic view of the technology (21).

All told, however, we do not view the DSRC and Internet-based approaches as competitive. For the 100,000 signalized intersections that are already connected, there is surely no reason to spend $1.75 billion installing DSRC radios. But for the 200,000 intersections that remain, DSRC may well turn out to be the method of choice. We expect the data-driven (as opposed to vision-based) approaches to use a hybrid of these two technologies.

Summary

In this paper, we have described a variety of techniques for getting traffic-light information into autonomous vehicles. Two of those techniques have been vision-based (depending on whether the light locations were found visually or those locations were found in a database) and two have been data-driven (DSRC and Internet). We have also presented a variety of situations in which these techniques can be expected to fail, roughly corresponding to the figures that have appeared throughout the paper. If we summarize the abilities of the four methods to deal with the various failures, it looks something like this:

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Vision</th>
<th>Data-driven</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Location visually</td>
<td>Location from data</td>
</tr>
<tr>
<td>1</td>
<td>Visual features resemble lights</td>
<td>✔</td>
<td>✔</td>
</tr>
<tr>
<td>2</td>
<td>Too many lights</td>
<td>✔</td>
<td>✔</td>
</tr>
<tr>
<td>3</td>
<td>Lights move</td>
<td>✔</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>Temporary lights</td>
<td>✔</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>Lights broken</td>
<td>✔</td>
<td>✔</td>
</tr>
</tbody>
</table>
It should be obvious that no single method is a panacea, and a safe autonomous vehicle will almost inevitably need to use a combination of approaches. To the extent that specific conclusions can be drawn, we expect them to be the following:

1. Although the two data-driven approaches appear to be competitors, it is probably most appropriate to view them as likely providing similar functionality in different locations. We would expect the Internet-based approach to reach the most signalized intersections the most quickly, and the DSRC-based approach to provide supplementary coverage as the standard becomes more general and funds become available.
2. The “vision-based, location from data” approach strikes us as the least likely to be part of deployed systems, because this approach is incapable of dealing with many of the issues addressed by a purely vision-based approach, and also because this approach appears to be (mostly) dominated by the data-driven approach in locations where data is available.

Of course, there is no particular reason to believe that the examples we have chosen to present will be representative, or that all of the problems will fit naturally into one of the groups in the above table. And here, perhaps, is the single greatest lesson that can be learned from the considerations we have presented: Autonomous driving is hard. It requires advancements in both engineering and in science. An intellectual environment in which all of the players guard progress so closely is unlikely to be in the interests of any of us.

**Acknowledgement**

I would like to thank all of the members of the Connected Signals team for creating the stimulating environment in which we have jointly worked on these problems, and would especially like to thank David Etherington, Pamela Kinion and Tim Stirling for their help.


