Qualitative Motion Understanding

Wilhelm Burger and **Bir Bhanu***

Honeywell Systems & Research Center
3660 Technology Drive, Minneapolis, MN 55418, U.S.A.

ABSTRACT

The vision system of an Autonomous Land Vehicle is required to handle complex dynamic scenes. Vehicle motion and individually moving objects in the field of view contribute to a continuously changing camera image. It is the purpose of motion understanding to find consistent three-dimensional interpretations for these changes in the image sequence. We present a new approach to this problem, which departs from previous work by emphasizing a qualitative nature of reasoning and modeling and maintaining multiple interpretations of the scene at the same time. This approach offers advantages such as robustness and flexibility over "hard" numerical techniques which have been proposed in the motion understanding literature.

1. INTRODUCTION

Visual information about the environment is an indispensable clue for the operation of an autonomous land vehicle (ALV). Due to the vehicle's egomotion, however, the resulting camera image is continuously changing even in a completely stationary environment. Any change in the 2-D image is the result of some change in 3-D space, either induced by vehicle motion or by moving objects in the field of view.

The purpose of our approach is to find consistent interpretations of time-varying images obtained from the camera on a moving vehicle. Specifically we are interested to determine

- how is the camera moving ?
- what is moving in the scene and how does it move ?
- what is the approximate 3-D structure of the scene ?

The original input is a sequence of images, taken at a constant rate. A reliable low-level correspondence technique[4] is assumed to be available, which extracts distinct features from every image and supplies their 2-D image locations in successive frames. The correspondence algorithm labels each feature and tracks it over time by generating tupels (*feature-label, time, x-location, y-location*).

Previous work[2] in motion analysis has concentrated mainly on numerical techniques for computing motion parameters and scene structure from image sequences. While a completely stationary environment has often been assumed for the recovery of the camera motion, the possible presence of moving objects must be accounted for in this scenario. Similarly, we cannot rely on a stationary camera setup to detect those moving objects. Clearly, some kind of common reference is required, against which the movement of the vehicle as well as the movement of objects in the scene can be related. For this purpose, a vehicle-centered *Qualitative Scene Model* (QSM) is constructed and maintained over time, representing the current set of feasible interpretations of the scene.

2. IMAGE OBSERVATIONS

The first step of our approach is to determine the vehicle's motion relative to the stationary environment between each pair of frames. If the vehicle moves along a straight line, all stationary features seem to expand from one single point in the image, the *focus of expansion* (FOE). Given the accurate location of the FOE, the relative range of any (stationary) point P in the image can be determined at time t by the relation

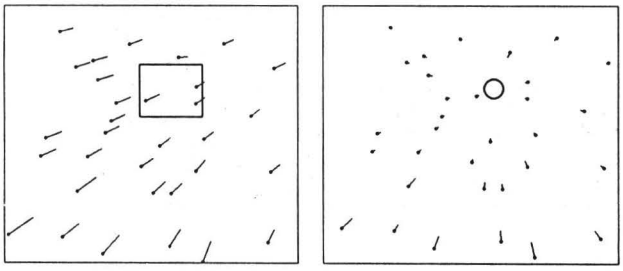
$$Z(t) \propto V(t)\,\frac{r(t)}{v(t)}\,,$$

where $Z(t)$ denotes the actual distance of the point from the image plane in 3-D, $V(t)$ is the velocity dZ/dt of the vehicle perpendicular to the image plane, $r(t)$ is the 2-D distance between the image of P and the FOE, and $v(t)$ is the radial velocity dr/dt. Since $V(t)$ is the same for any stationary point in the scene, $r(t)$ and $v(t)$ can be measured in the image and depth can be computed up to a common scale factor. Furthermore, by knowing the vertical distance of the camera to the ground, absolute values for vehicle velocity and range can be obtained.

While the vehicle is traversing the environment, $Z(t)$ keeps changing for every feature in the field of view. If the depth map itself was used as the scene model, the model would have to be updated continuously. The topology of the stationary part of the scene, however, should remain unchanged.

In reality the ALV does not travel along a straight line but performs small rotations, which induce an additional vector field in the image (Fig. 1(a)). The fact that all the displacement vectors of stationary features must intersect at a common point can be used to "derotate" the image[3] and compute the vehicle's rotation and direction of translation (Fig. 1(b)). Due to inertia, the direction of translation as well as the amount of rotation about either axis cannot change drasti-

*This work was supported by the Defense Advanced Research Projects Agency under contract DACA 76-86-C-0017 and monitored by the U.S. Army Engineer Topographic Laboratories.



(a) (b)

Figure 1. *(a)* A typical displacement field taken from the ALV undergoing translation and rotation. The FOE lies inside the area marked by the rectangle. *(b)* After *derotation* all vectors intersect the FOE-area, marked by a circle, giving the direction of instantaneous heading of the ALV.

cally between two frames. Thus the previous motion parameters can serve as a good initial guess for computing the current vehicle motion. We do not assume that the exact location of the FOE in the image plane can always be determined, but that a region can be specified, which contains the FOE with high certainty. Although this first step is done exclusively in the image plane, knowledge about the stationary features to be used must be available in some form. This information is supplied by the QSM which is described below.

The feature-property *stationary/mobile* and the *closer* relationship between features are the basic building blocks of the QSM. (*Mobile* **a**) means, that **a** has been found moving once and is not considered part of the stationary environment. A point **a** is said to be *closer* than **b** (*closer* **a b**), if **a** is closer to the camera plane than **b** in space. The current set of image features together with assigned properties and mutual relationships constitute an interpretation of the scene. An interpretation is *feasible*, as long as it is free of internal conflicts, e.g. (*closer* **a b**) *and* (*closer* **b a**). The QSM comprises a set of feasible interpretations of the current scene. As the vehicle proceeds, new observations in the image are incorporated into the model, while existing hypotheses inside the QSM are verified.

Since the interpretation of image motion is inherently ambiguous, different forms of visual information are employed for the construction of the QSM: *geometric*, *spatial*, and *semantic* information. Each of these three types of knowledge is formulated as a group of rules in a Blackboard environment. The following examples shall illustrate the basic ideas:

Rule GEOMETRY-1:
if image-point **a** is moving toward the FOE-region
then assert (*mobile* **a**).

Rule GEOMETRY-2: An image point **a** is said to be *inside* a point **b**, if **a**,**b** are in some neighborhood and **a** is at a smaller distance from the FOE than **b**. $D_{a,b}(t)$ denotes the Euclidean distance from **a** to **b** at time t.
if (*stationary* **a b**) & (*inside* **a b**) & ($D_{a,b}(t) > D_{a,b}(t+1)$)
then hypothesize (*closer* **a b**).

Rule GEOMETRY-3: This constraint-rule verifies that a hypothesis generated in the previous rule can be maintained over time.
if (*stationary* **a b**) & ($D_{a,b}(t) > D_{a,b}(t+1)$)
then verify ($D_{a,b}(t+1) > D_{a,b}(t+2)$).

While the results from geometry are valid for arbitrary configurations, certain assumptions can be made about the spatial layout of the scene which is encountered by the ALV. For example, the fact that (for an upright camera) the *lower* features in the image are generally *closer* to the vehicle can be expressed in the following heuristic rule.

Rule SPATIAL-1: For any pair of image points **a**,**b**:
if (*lower* **a b**)
then hypothesize (*closer* **a b**)

As a consequence, it is very unlikely that a feature is farther away than all its surrounding neighbors.

Occlusion is another important source of spatial information, which is applicable for more complex features such as lines and regions. A feature occluding another is certainly closer to the viewer than the occluded feature.

Semantic information becomes an important factor as soon as partial interpretations of the scene are available. For instance, if the horizon has been identified, any object above it must be in the sky and is probably not stationary. Similarly, the features of an object recognized as a building would not be considered moving in an ambiguous situation.

3. INTERPRETATION AND CONFLICT RESOLUTION

Using a set of rules like the ones described in the previous section, the *Qualitative Scene Model* is constructed. Here we describe how the model develops over time, how new interpretations of the scene are generated, and how conflicts are resolved.

An example with a scene containing three feature points **a**,**b**,**c** is shown in Figure 2. Initially (at $t = t_0$) nothing is known about the spatial relationships between these points and whether they are stationary or not. The default assumption is that any point is stationary unless there is an indication that this is not true. The initial interpretation of the scene thus contains only

Interpretation $A(t_0)$:
(*stationary* **a b c**).

Suppose that between t_0 and t_1 all three points show some amount of expansion away from the FOE, giving rise to the conclusion (e.g. by rule GEOMETRY-2) that **a** is closer (to the vehicle) than **b**, **a** is closer than **c**, and **c** is closer than **b**. From the information gathered up to this point, the interpretation of the scene at time t_1 looks like this:

Interpretation $A(t_1)$:
(*stationary* **a b c**),
(*closer* **a b**), (*closer* **a c**), (*closer* **c b**).

At time t_2 one of the rules claims that **c** is closer than **a** and tries to assert this fact into the current interpretation. Clearly, the new interpretation would contain the conflicting facts
(*closer* **a c**) *and* (*closer* **c a**),
which would not be a feasible interpretation. The conflict is

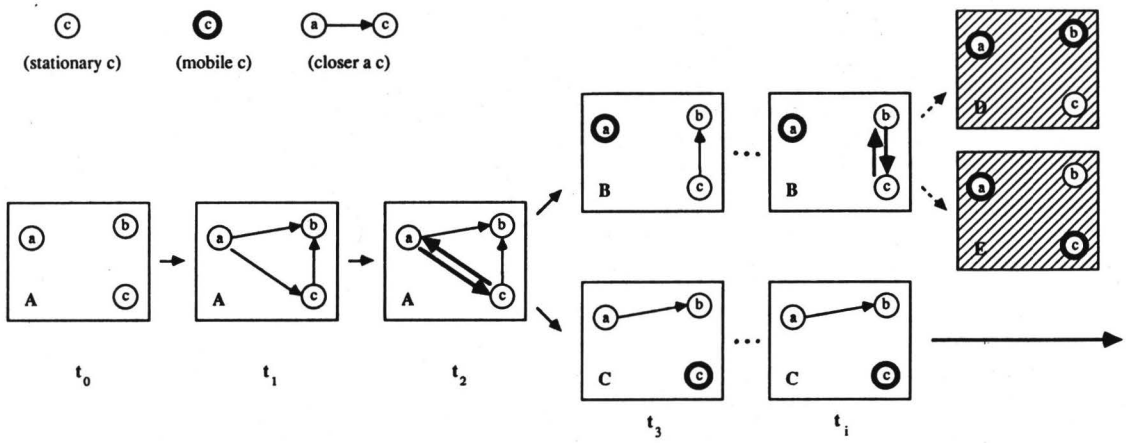


Figure 2. Development of the *Qualitative Scene Model* (QSM) over time: At time t_0 three features **a**, **b**, **c** are given, which are initially assumed to be stationary. At time t_1 three *closer*-relationships have been established between **a**, **b** and **c**. At time t_2 a conflict occurs in interpretation **A** by the contradictory facts (*closer* **a c**) and (*closer* **c a**). Two new interpretations (**B** and **C**) are created, each containing one feature considered *mobile* (**a**, **c** respectively). At time t_i a new conflict occurs in interpretation **B** from the additional fact (*closer* **b c**). Since another interpretation (**C**) exists at the same time which could absorb this fact (**c** is *mobile* in **C**), **B** is not branched out but discontinued. **C** remains as the only feasible interpretation.

resolved by creating two disjunct hypotheses **B** and **C**, with either **a** or **c** as mobile:

Interpretation **B**(t_2):
(*mobile* **a**), (*stationary* **b c**), (*closer* **c b**).

Interpretation **C**(t_2):
(*mobile* **c**), (*stationary* **a b**), (*closer* **a b**).

Notice, that when a feature is hypothesized to be *mobile*, all its *closer*-relationships are removed from the interpretation. At this point in time (t_2), *two* feasible interpretations of the scene are active simultaneously. All active interpretations are pursued until they enter a conflicting state, in which case they are either branched into new interpretations or removed from the QSM.

In our example we assume, that both interpretations **B** and **C** are still alive at time t_i. At this point some rule claims, that if **b**, **c** are both stationary, then **b** is closer than **c**. This creates a conflict in interpretation **B**, because **B** contains the contradictory fact (*closer* **c b**)! Again we could branch interpretation **B** into two new interpretations (**D**,**E**), with either (*mobile* **b**) or (*mobile* **c**). At this time, however, there exists another active interpretation (**C**), which could absorb (*closer* **b c**) without causing an internal conflict (**c** is *mobile* in **C**). Therefore **B** is not branched out but removed altogether from the model, and only interpretation **C** survives.

In summary, the development of the QSM is controlled by the following *meta*-rules:

- After a hypothesis has been created by one of the analyzing rules, try to integrate this hypothesis as a fact into every active interpretation.
- If the new fact is consistent with the interpretation, make it a part of this interpretation.
- If the new fact is *not* consistent with the interpretation, and there are currently no other interpretations active, then create a new set of interpretations containing this fact without conflict.
- Otherwise prune the search tree by deleting the conflicting interpretation from the model.

4. CONCLUSIONS

In this paper we presented a new approach for the problem of motion interpretation in the scenario of an Autonomous Land Vehicle, following a qualitative line of reasoning and modeling. Different forms of visual information are combined in a rule-based framework to construct and maintain a three-dimensional qualitative model of the environment. Instead of refining a single numerical model a set of disjunct interpretations of the dynamic scene are pursued simultaneously.

The work reported here shows the conceptual outline of our approach. While we have considered only point-features so far, the integration of lines and regions will be a natural extension. An implementation is currently under way using actual ALV imagery.[1]

REFERENCES

1. B. Bhanu and W. Burger, "DRIVE - Dynamic Reasoning from Integrated Visual Evidence," Proc. DARPA Image Understanding Workshop pp. 581-588, Morgan Kaufmann Publishers (February 1987).

2. H.-H. Nagel, "Image Sequences - Ten (octal) Years - From Phenomenology towards a Theoretical Foundation," Proc. Intern. Conf. on Pattern Recognition pp. 1174-1185, Paris (1986).

3. K. Prazdny, "On the Information in Optical Flows," *Computer Vision, Graphics, and Image Processing* **22** pp. 239-259 (1983).

4. S. Ullman, *The Interpretation of Visual Motion*, MIT Press, Cambridge, Mass. (1979).