Dynamic Scene Understanding for Autonomous Mobile Robots

Wilhelm Burger and Bir Bhanu

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

ABSTRACT

A new approach to the dynamic scene analysis is presented which departs from previous work by emphasizing a qualitative strategy of reasoning and modeling. Instead of refining a single quantitative description of the observed environment over time, multiple qualitative interpretations are maintained simultaneously. This offers superior robustness and flexibility over traditional numerical techniques which are often ill-conditioned and noise-sensitive. The main tasks of our approach are *(a)* to detect and to classify the motion of individual objects in the scene, *(b)* to estimate the robot's egomotion, and *(c)* to derive the 3-D structure of the stationary environment. These three tasks strongly depend on each other. First, the direction of heading (i.e. translation) and rotation of the robot are estimated with respect to stationary locations in the scene. The focus of expansion (FOE) is not determined as particular image location, but as a region of possible FOE-locations called the *Fuzzy FOE*. From this information, a rule-based system constructs and maintains a *Qualitative Scene Model*. Results of this approach from real and synthetic imagery are presented.

1. Introduction

Visual information plays a key role in mobile robot operation. Even with the use of sophisticated inertial navigation systems, the accumulation of position errors requires periodic corrections. Operation in unknown environments or mission tasks involving search, rescue or manipulation critically depend upon visual feedback. Motion understanding becomes vital as soon as moving objects are encountered in some form, e.g. while following a convoy, approaching other vehicles or to detect moving threats. In the given case of a moving camera, image motion can also supply important information about the spatial layout of the environment and the actual movements of the ALV.

Previous work in motion understanding has mainly concentrated on numerical approaches for the reconstruction of 3-D motion and scene structure from 2-D image sequences (see Nagel [7] for a comprehensive review). While a completely stationary environment is frequently assumed for the visual estimation of camera motion, the possible presence of moving objects in the field of view must be accounted for in the given scenario. I.e., the observed scene cannot be treated as a single rigid object. Similarly, due to the vehicle's egomotion, the stationary objects in the scene are not necessarily mapped onto static image locations.

In the classic approach, structure and motion of a rigid object are computed simultaneously from successive perspective views by solving a system of linear or nonlinear equations [4,11]. This technique is known to be noise sensitive even when more than two frames are used [3]. Non-rigid motion or the presence of several moving objects in the field of view would be indicated by a large residual error for the solution to the system of equations. However, in some cases of non-rigid motion an acceptable numerical solution may exist that corresponds to a rigid interpretation. In such a case, the movements of individual entities in the field of view would not be detectable by the classic numerical scheme. Adiv [1] generalized this approach to handle multiple moving objects by using a complex two-stage grouping process to segment the optical flow field.

While it has been a common view to consider scene structure as a *by-product* of rigid motion computation we argue that deriving and modeling the 3-D scene structure is a necessary prerequisite for motion understanding. The approach that we propose is novel in two important aspects. *First*, the scene's 3-D structure serves as a link between motion analysis and other processes that deal with spatial perception, such as shape-from-occlusion, stereo, spatial reasoning, etc. A 3-D interpretation of a moving scene can only be correct if it is acceptable by all the processes involved.*Secondly*, numerical techniques have been complemented by a *qualitative* strategy of reasoning and modeling. The use of qualitative techniques in Computer Vision has received growing interest recently (e.g.[10,12]). Basically, instead of having a system of equations approach a single rigid (but possibly incorrect) numerical solution, we maintain multiple qualitative interpretations of the scene. All the existing interpretations are kept consistent with the observations made in the past. The main advantage of this approach is that a new interpretation can be supplied immediately when the currently favored interpretation turns out to be false.

These interpretations are built in three separate steps (see Fig.1). First, significant features (points, boundaries, corners, etc.) are extracted from the image and the 2-D displacement vectors are computed for this set of features. In the second step, the vehicle's direction of translation, i.e. the focus of expansion (FOE), and the amount of rotation in space are determined. Most of the necessary quantitative computations are performed in this 2-D step which is described in Section 2. The third step (2-D Change Analysis) constructs the internal 3-D *Qualitative Scene Model*, outlined in Section 3. Experiments with our approach on real images taken from the Autonomous Land Vehicle (ALV) are discussed in Section 4.

2. Fuzzy FOE

When a camera performs pure translation along a straight line in space, the images of all stationary features seem to diverge from one particular location which is commonly called the "focus of expansion" (FOE). In reality, however, the vehicle not only translates but also rotates more or less about its three

CH2605-4/88/0000/0736$01.00 © 1988 IEEE

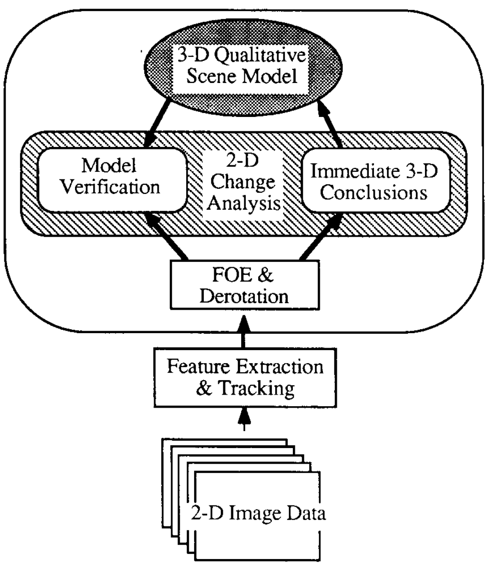

Figure 1. Main steps of the approach. First, discrete features are extracted and tracked between successive images. The resulting field of 2-D displacement vectors serves to compute the FOE and the amount of camera rotation. The 3-D scene model is constructed in a hypothesize-and-test circle.

major axes. The movement **M** of a land vehicle can be sufficiently approximated by a translation **T** followed by rotations about the horizontal axis $\mathbf{R}_\theta$ (pan) and the vertical axis $\mathbf{R}_\varphi$ (tilt). A 3-D point X = (x,y,z) in the camera-centered co-ordinate frame is transferred by the camera movement **M** to a new location X′=(x′,y′,z′)

$$\mathbf{M}:\ X \rightarrow X' = \mathbf{R}_\varphi\, \mathbf{R}_\theta\, \mathbf{T}\ (X).$$

If the observed scene is completely stationary, the effects upon the image caused by the camera movement **M** can be described by a 2-D transformation **d** (for *displacement*), which takes the original image **I** to the following image **I′**. The 3-D rotations $\mathbf{R}_\varphi$ and $\mathbf{R}_\theta$ and translation **T** have their equivalents in **d** as the separate 2-D transformations $\mathbf{r}_\varphi$, $\mathbf{r}_\theta$, and **t** :

$$\mathbf{d}: \mathbf{I} \rightarrow \mathbf{I}' = \mathbf{r}_\varphi\, \mathbf{r}_\theta\, \mathbf{t}\ (\mathbf{I}).$$

Since pure camera rotations do not create new views of the environment, the corresponding 2-D transformations $\mathbf{r}_\varphi$ and $\mathbf{r}_\theta$ are effectively mappings of the image onto itself. Conversely, the image effects **t** of pure camera translation depend upon each 3-D point's actual location in space. At this point we introduce a (hypothetical) intermediate image **I***, which is the result of a pure camera translation **T**:

$$\mathbf{t}: \mathbf{I} \rightarrow \mathbf{I}^*$$

Notice that the image **I*** is never really observed, except in the special case of pure camera translation. However, **I*** has two important properties: *First*, all displacement vectors between corresponding points in **I** and **I*** seem to diverge from a particular image location (xf,yf) (the FOE), unless the camera does not translate at all. We call this property of the displacement field "radial_mapping (**I**,**I***)." *Secondly*, for given pan and tilt angles φ and θ, **I*** can be obtained regardless of the 3-D scene structure by applying the inverse mappings $\mathbf{r}_\varphi^{-1}$ and $\mathbf{r}_\theta^{-1}$ (which always exist) to the observed image **I′**:

$$\mathbf{I}^* = \mathbf{r}_\theta^{-1}\, \mathbf{r}_\varphi^{-1}\, \mathbf{I}'.$$

Once suitable mappings $\mathbf{r}_\theta^{-1}\, \mathbf{r}_\varphi^{-1}$ have been found, the FOE can be located for the pair of images **I** and **I***. However, it is not trivial to determine how close a given displacement field is to a radial mapping without knowing the location of the FOE. In most of the proposed schemes for testing this property the displacement vectors are extended as straight lines to somehow measure the spread of their intersections [5,8]. Unfortunately, the resulting error functions are noise-sensitive and not well behaved for varying values of φ and θ, i.e., they require expensive global search.

However, for a given FOE-location, the optimal rotations angles can be found analytically by minimizing second order functions [2] and the deviation of the "derotated" displacement field from the ideal radial pattern is easily measured. The resulting error function is usually smooth and monotonic within a large area around the actual FOE, i.e., even from a poor initial guess the global optimum can be found by local search methods. This technique is fairly robust in the presence of noise and under small camera translation. However, the 2-D error function flattens out in those cases and the location of minimum error may be considerably off the actual FOE. The local shape of the error function is therefore an important indicator for the accuracy of the result.

This raises the question whether it is necessary to locate the FOE as one particular point in the image. After all, even humans seem to have difficulties in estimating the direction of heading under similar conditions [9]. As we demonstrate in the following section many conclusions about the 3-D properties of the scene can be drawn even if only the approximate location of the FOE is known. The following algorithm searches for a connected region of possible FOE-locations which we call the *Fuzzy FOE*. The final size of this region depends upon the local shape of the 2-D error function. A large Fuzzy FOE reflects a flat error function, i.e., little accuracy in the location of the FOE, whereas a small region indicates a distinct local optimum.

Fuzzy_FOE (I,I*):

(1) Guess initial FOE (x_0,y_0) (e.g. the FOE obtained from the previous frame pair) and compute the corresponding optimal rotations φ_0,θ_0 and the deviation from a radial flow field (error) e_0.

(2) From (x_0,y_0) start a local search for an FOE-location (x_C,y_C) that results in a minimum error e_C.

(3) Create the set FUZZY-FOE={ $(x_C,y_C,\varphi_C,\theta_C,e_C)$ }.

(4) Grow the set FUZZY-FOE by including adjacent FOE-locations until either *(a)* a certain error ratio e_{max}/e_C within the FOE-region is reached or *(b)* the region exceeds a predefined size (to stop when the error function is flat). •

After computing the Fuzzy FOE and the angles of horizontal and vertical rotation, a good estimate for the motion parameters of the vehicle is available. Notice that this is possible without knowing the 3-D structure of the observed scene. To measure the camera motion with respect to the stationary world, however, none of the displacement vectors used for this

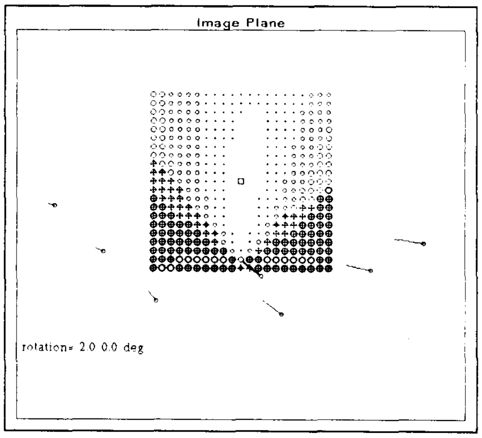

Figure 2. Fuzzy FOE for a simulated displacement field for a vehicle translating and rotating to the right by 2°. The small square in the center is the location of the actual FOE. The error values for surrounding (i.e., hypothesized) FOE locations are shown with circles of proportional size. Notice the elongated shape of the FOE region which is due to the particular distribution of displacement vectors (typical for road scenes).

computation may belong to another moving object. This information is supplied by the internal scene model (as described in the following), which, among other things, tells us what features are currently believed to be stationary. Fig. 2 shows the results of applying this algorithm to a simulated sparse displacement field. The shape of the error function around the actual FOE is plotted with circles of size proportional to the error. The blank area in the center of Fig.2 marks the resulting Fuzzy FOE.

3. Qualitative Scene Model

The choice of a suitable scheme for the internal representation of the scene is of great importance. The *Qualitative Scene Model* (QSM) is a 3-D camera-centered interpretation of the scene, which is built incrementally from visual information gathered over time. The nature of this model, however, is a *qualitative* rather than a precise geometric description of the scene. The basic building blocks of the QSM are *entities*, which are the 3-D counterparts of the 2-D *features* observed in the image. For example, the point feature A located in the image at x,y at time t (Point_Feature $A\ t\ x\ y$) has its 3-D counterpart in the model as (Point_Entity A).

Since the model is camera-centered the image locations and 2-D movements of features are implicitly part (i.e., known facts) of the model. Additional entries are the properties of entities (e.g., "stationary" or "mobile") and relationships between entities (e.g. "closer"), which are not given facts but the outcome of some interpretation step (i.e., hypotheses). This is expressed in the model as either

(Stationary *entity*) or (Mobile *entity*) .

It is one of the key features of the QSM that it generally contains not only one interpretation of the scene but a (possibly empty) *set* of interpretations which are all pursued simultaneously. At any point in time, a hypothesis is said to be "feasible" if it exists in the QSM and is not in conflict with some observation made since it was established.

Interpretations are structured as an inheritance network of partial hypotheses. Individual scene interpretations are treated as "closed worlds", i.e., a new conclusion only holds within an interpretation where all the required premises are true. Interpretations are also checked for internal consistency, e.g., entities cannot be both stationary and mobile within the same interpretation. The QSM is maintained through a generate-and-test process as the core of a rule-based blackboard system. The two major groups of rules are: Generation Rules and Verification Rules.

Generation Rules examine the (derotated) image sequence for significant changes and modify each interpretation in the QSM if applicable. Some of these observations have unconditional effects upon the model. E.g., if an image feature is found to be moving *towards* the Fuzzy FOE (instead of diverging away from it), then it belongs to a moving entity in 3-D space. The actual rule contains only one premise and asserts (MOBILE ?x) as a globally known fact (i.e., one that is true in every interpretation):

```
(defrule DEFINITE_MOTION
    (MOVING_TOWARDS_FOE ?x ?t)    < observation at time t >
=>
    (at ROOT (assert (MOBILE ?x)))).          < a global fact >
```

The directive "at ROOT" in the above rule places the new fact at the root of the interpretation graph, i.e., it is inherited by all existing interpretations.

Other observations depend upon the facts that are currently true in a "world" and therefore may have only local consequences inside particular interpretations. For example, if two image features A and B lie on opposite sides of the Fuzzy FOE and they are getting closer to each other, then they must be in relative motion in 3-D space. If an interpretation exists that considers at least one of the two entities (x,y) stationary, then the other entity cannot be stationary (i.e., it must be mobile). The following rule "fires within" each interpretation that considers the first entity (x) stationary:

```
(defrule RELATIVE_MOTION
    (OPPOSITE_FOE ?x ?y ?t)              < observation 1 >
    (CONVERGING ?x ?y ?t)                < observation 2 >
    (STATIONARY ?x           < true inside an interpretation >
 =>
    (assert (MOBILE ?y))).   < new fact local to this interpretation >
```

While some image observations allow direct conclusions about motion in the scene, other observations hold cues about the stationary 3-D structure. If the *exact* location of the FOE is known then the depth of each stationary point (i.e., its 3-D distance from the camera) is proportional to the rate of divergence (from the FOE) of its image [8]. Applied to the Fuzzy FOE, where a set of potential FOE locations is given, the distance $Z(A)$ of a stationary point A is determined as an *interval* instead of one single number:

$$Z^{min}(A) \leq Z(A) \leq Z^{max}(A).$$

Therefore, a point A is closer in 3-D than another point B, if the corresponding ranges of depth do not overlap, i.e.,

$$Z^{max}(A) < Z^{min}(B) \Rightarrow (\text{ CLOSER } A\ B\ ).$$

Since this conclusion only holds if both features are actually stationary, the following rule fires only within a suitable interpretation (if it exists):

```
(defrule CLOSER_FROM_DIVERGENCE
   (STATIONARY ?x)                 < interpretation where both x
   (STATIONARY ?y)                   and y are stationary >
   (< (Zmax ?x) (Zmin ?y))         < no overlap in range >
 =>
   (assert (CLOSER ?x ?y))).
```

To compare the ranges of 3-D points, another criterion can be used which does not measure the individual rate of divergence. Instead, the change of distances *between* certain pairs of features is observed. If two stationary points lie on the same side of the FOE and the distance between them is becoming smaller, then the *inner* feature (i.e., the one which is nearer to the FOE) is closer in 3-D space. This is a valuable test for features that are relatively close to each other. It can be employed even if the image is not (or incorrectly) derotated and the location of the FOE is either only known very roughly or is completely outside the field of view (i.e., for a side-looking camera):

```
(defrule CLOSER_FROM_CHANGING_DISPARITY
   (STATIONARY ?x)                 < interpretation where both x
   (STATIONARY ?y)                   and y are stationary >
   (SAME_SIDE_OF_FOE ?x ?y)        < e.g. right of the FOE >
   (CONVERGING ?x ?y)              < disparity is decreasing >
   (INSIDE ?x ?y)             < x is nearer to the FOE than y >
 =>
   (CLOSER ?x ?y).
```

While the purpose of the generation rules is to establish new hypotheses and conclusions the purpose of verification rules is to review interpretations after they have been created and, if possible, prove that they are false. When a hypothesis is found to be inconsistent with some new observation it is usually removed from the QSM. Any interpretation that is based on that hypothesis is removed simultaneously. Since we are always trying to come up with a single (and hopefully correct) scene interpretation this mechanism is important for pruning the search tree.

Verification rules are typically based on image observations that, used as generators, would produce a large number of unnecessary conclusions. For example, the general layout of the scene seen from the top of a land vehicle suggests the rule of thumb that things which are *lower* in the image are generally *closer* to the camera. Although this rule is not strong enough to draw direct conclusions, it may be used to verify existing hypotheses:

```
(defrule LOWER_IS_CLOSER_HEURISTIC
   (CLOSER ?x ?y)                      < existing hypothesis >
   (BELOW_THE_HORIZON ?x ?t)           < rule does not apply to
   (BELOW_THE_HORIZON ?y ?t)             things in the air etc. >
   (BELOW ?y ?x ?t)                 < actually x should be below y >
 =>
   (assert (CONFLICT LOWER/CLOSER ?x ?y))).
                  < mark this interpretation as conflicting>
```

Whenever an existing hypothesis (CLOSER ?x ?y) violates the above rule of thumb, this rule fires and marks the interpretation as conflicting. How the conflict is eventually resolved depends upon the global state of the QSM. E.g., simply removing the afflicted interpretation would create an empty model if this interpretation presently is the only one. This task is handled by a set of dedicated *conflict resolution rules* (see [2]).

The kind of rules described up to this point are mainly based upon the geometry of the imaging process, i.e., perspective projection. Other important visual clues are available from occlusion analysis, perceptual grouping, and semantic interpretation. *Occlusion* becomes an interesting phenomenon when features of higher dimensionality than points are employed, such as lines and regions. Similarities in form and motion found by *perceptual grouping* allow us to assemble simple features into more complex aggregates. Finally, as an outcome of the recognition process, *semantic* information may help to disambiguate the scene interpretation. If an object has been recognized as a building, for example, it makes every interpretation obsolete that considers this object mobile.

In summary, the construction of the QSM and the search for the most plausible scene interpretation are guided by the following *meta rules*:

- Always tend towards the "most stationary" (i.e. most conservative) solution. By default all new entities (entering the field of view) are considered stationary.
- Assume that an interpretation is feasible unless it can be proved to be false (the principle of "lack of conflict").
- If a new conclusion causes a conflict in one but not in another current interpretation then remove the conflicting interpretation.
- If a new conclusion cannot be accommodated by any current interpretation then create a new, feasible interpretation and remove the conflicting ones.

4. Experimental Results

In the following, the operation of the QSM and the surrounding rule base is demonstrated for two instances of an image sequence taken from the moving ALV. Point features were tracked by hand between successive frames on the binary edge images (Fig.3a) to simulate the conditions of automatic feature tracking (e.g., see [6]). The scene contains a number of stationary points and one moving point (24) which belongs to another vehicle that is moving away from the camera. Figure 3.b shows the original displacement vectors (solid lines) between frames 182 and 183, the Fuzzy FOE (shaded area), and the "derotated" displacement vectors (dotted lines). The rotation scale indicates a horizontal rotation angle of almost 1° to the left between the two frames. Vertical rotation is insignificant.

Figures 4a and 4b visualize two separate, feasible scene interpretations for the situation in frame 183. Entities which are considered stationary are marked with circles or plain labels. Arcs from a small circle (or plain label) to a larger circle indicate that a CLOSER-relationship has been established between the two entities (the entity with the larger circle is closer to the camera in 3-D). Mobile entities are marked with squares, or with arrows if the direction of their current movement has been identified.

The existence of two interpretations is due to the movement of the approaching car (point 24). This movement was detected as 2-D motion "across the FOE" (see rule RELATIVE_MOTION in Section 3) between point 24 on one side of the FOE and points 8,11, 19, 20, 22, 23 on the opposite

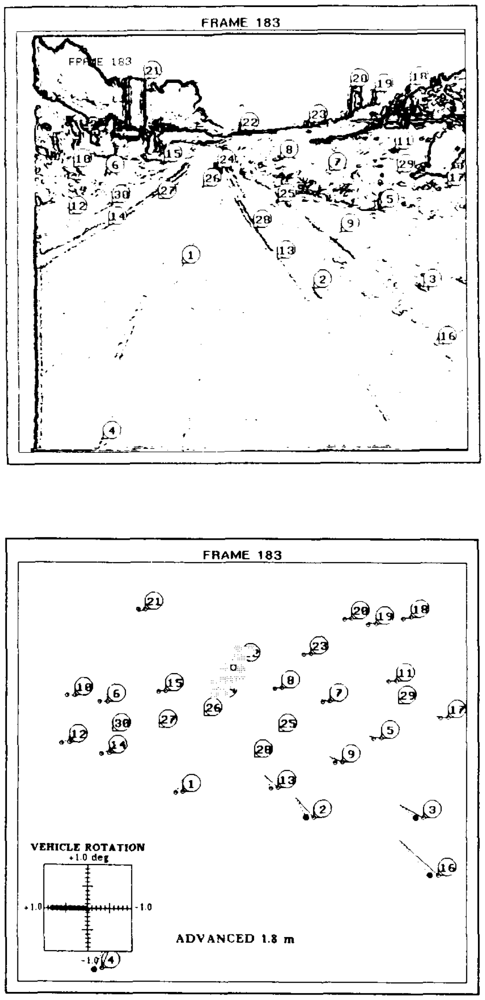

Figure 3. *Top:* Edge image from ALV sequence with point features marked. *Bottom:* Original displacement vectors (solid lines), Fuzzy FOE (shaded area), and derotated displacement vectors (dotted lines). The rotation scale indicates about 1° of horizontal vehicle rotation to the left.

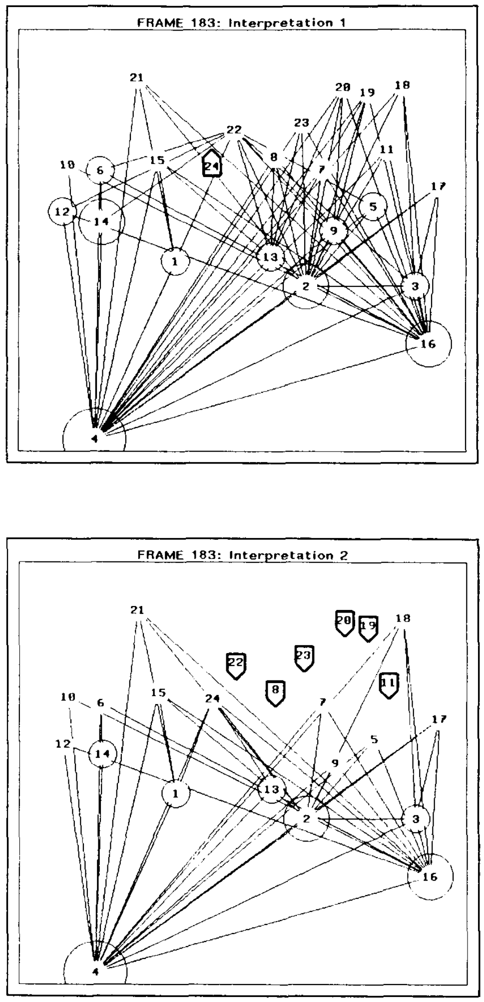

Figure 4. Two separate scene interpretations for frame 183. *Top:* Interpretation 1 considers all entities stationary except point 24 (the moving car) which moves upwards. This movement was detected by relative motion "across the FOE." *Bottom:* Consequently, Interpretation 2 sees points 8,11,...23 moving downwards.

side. Since there was no other indication for the movement of point 24, two interpretations were created.

Interpretation 1 (Fig.4a) considers all entities stationary, except point 24 which is moving upwards (in the 3-D coordinate frame). Since point 24 is located below the horizon, the system could now hypothesize that 24 is also *receding* in space. This is (as we know) the correct solution. However, interpretation 2 (Fig.4b) is also feasible, taking 24 as stationary and points 8,11,...23 as moving downwards. Notice that CLOSER-relationships are only formed between stationary entities.

Interpretation 2 does not "survive" the verification after the following frame. If entities 8,11,...23 were really moving downwards, then they should not exhibit any divergence away from the FOE. In this case at least one of those points undergoes significant divergence which is sufficient to prove interpretation 2 false. Consequently, for frame 184 only one interpretation remains in the QSM (Fig.5).

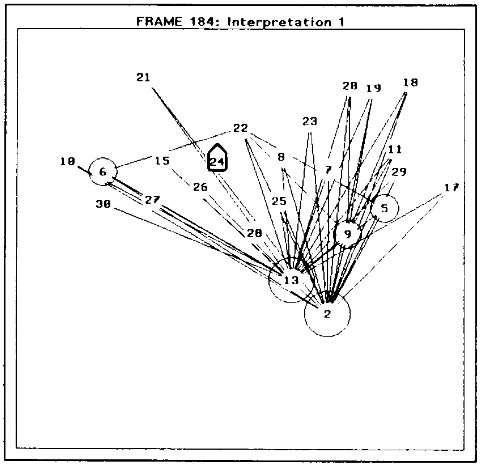

Figure 5. After frame 184 only a single interpretation "survives" with point 24 moving upwards (i.e. receding from the camera). Interpretation 2 (for frame 183) was eliminated because some of the points 8,11,...23 showed inconsistent divergence away from the FOE. Points 4 and 16 have meanwhile left the field of view.

5. Conclusion

In this paper we presented the conceptual outline of a new approach to scene understanding for mobile robots in dynamic environments. The challenge of understanding such image sequences is that stationary objects do not appear as stationary in the image and mobile objects do not necessarily appear to be in motion.

The approach taken here departs from related work by following a strategy of qualitative rather than quantitative reasoning and modeling. All the numerical efforts are packed into the computation of the focus of expansion (FOE), which is is accomplished entirely in 2-D. To cope with the problems of noise and errors in the displacement field we determine a region of possible FOE-locations instead of a single FOE. Termed the *Fuzzy FOE*, it is probably one of the most robust techniques available today. It is shown that, even without knowing the exact location of the FOE, many powerful conclusions about motion and 3-D scene structure are possible.

An internal 3-D representation, termed the *Qualitative Scene Model*, is constructed and maintained in a generate-and-test cycle over extended image sequences. To overcome the ambiguities inherent to dynamic scene analysis, multiple interpretations of the scene are pursued simultaneously. This model serves as a central pool of accumulated knowledge about the observed scene and allows the merging of various independent categories of visual clues.

Due to limited space, only one example could be given in this paper to show just the most basic operation of our approach on a real image sequence. The results of processing an extended image sequence and additional details can be found in [2]. There we also demonstrate that some apparently simple situations require relatively complex paths of reasoning, especialy in the context of indirect motion detection. Of course, a critical (and still unsolved) point is the problem of computing reliable displacement vectors or optical flow fields. Especially point features appear to be highly unreliable in noisy images while they deliver only sparse displacement fields. To exploit the full potential of our approach we are trying to employ more complex 2-D features, such as line segments and regions, for motion understanding.

REFERENCES

[1] G.Adiv, "Determining Three-Dimensional Motion and Structure from Optical Flow Generated by Several Moving Objects," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. **PAMI-7**, no. 4, pp.384-401 (July 1985).

[2] B.Bhanu, W.Burger, "DRIVE - Dynamic Reasoning from Integrated Visual Evidence," Scene Dynamics Report, DARPA Contract DACA76-86-C-0017, Honeywell S&RC, Minneapolis MN (June 1987).

[3] S.Bharwani, E.Riseman, and A.Hanson, "Refinement Of Environmental Depth Maps Over Multiple Frames," Proc.IEEE Workshop on Motion, Kiawah Island Resort, pp.73-80 (May 1986).

[4] O.D.Faugeras, F.Lustman, G.Toscani, "Motion and Structure from Point and Line Matches," Proc. 1st ICCV, London, pp. 25-34 (June 1987).

[5] C.Jerian, R.Jain, "Determining Motion Parameters for Scenes with Translation and Rotation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. **PAMI-6**, no. 4, pp.523-530 (July 1984).

[6] H.P.Moravec, "Towards Automatic Visual Obstacle Avoidance," Proc. 5th IJCAI, p. 584 (August 1977).

[7] H.-H.Nagel, "Image Sequences- Ten (octal) Years- From Phenomenology towards a Theoretical Foundation," Proc. Int. Conference on Pattern Recognition, Paris, pp.1174-1185 (1986).

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

[9] J.H.Rieger, D.T.Lawton, "Processing Differential Image Motion," J.Opt.Soc.Am. A, vol. **2**, no. 2, pp. 354-360 (1985).

[10] W.B.Thompson, J.K.Kearney, "Inexact Vision," Proc. IEEE Workshop on Motion, Kiawah Island Resort, pp. 15-21 (May 1986).

[11] R.Y.Tsai, T.S.Huang, "Uniqueness and Estimation of Three-Dimensional Motion Parameters of Rigid Objects with Curved Surfaces," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. **PAMI-6**, no. 1, pp. 13-27 (January 1984).

[12] A.Verri, T.Poggio, "Qualitative Information in the Optical Flow," Proc. DARPA Image Understanding Workshop, Los Angeles, pp. 825-834 (February 1987).

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.