Integrated Real-Time Motion Segmentation and 3D Interpretation

DRA Technical Report TR96SMS1

(A slightly different version of this has now been published in ICPR96 [7])

1996

S.M. Smith

Oxford Centre for Functional Magnetic Resonance Imaging of the Brain (FMRIB),
Department of Clinical Neurology, Oxford University, Oxford, UK
(Previously in Computer Vision and Image Processing Group, DRA Chertsey, DERA, UK)
steve@fmrib.ox.ac.uk
www.fmrib.ox.ac.uk/~steve

Abstract

This brief paper describes a real-time integrated motion segmentation and 3D reconstruction/ interpretation system. The motion segmentation system detects and tracks all moving objects, and removes the corresponding features from the original 2D feature data set. The remaining features are passed on to a 3D reconstruction system. The resulting list of 3D feature data, which hopefully contains useful information about the static part of the world, is then interpreted to give some high level understanding of the environment.

Keywords: Obstacle detection, motion segmentation, 3D reconstruction

1 Introduction

By analyzing how image sequences digitized by a camera evolve over time, it is possible to estimate both the motion of moving obstacles and the three dimensional (3D) structure of the world, and to begin to interpret that 3D data. This short paper describes the integration of a real-time motion segmentation system and a 3D reconstruction/interpretation system. The motion segmentation system is used to pre-segment the image sequence into the (probably moving) background and independently moving objects. This enables the structure-from-motion (SFM) system (which, like many SFM algorithms, cannot cope with more than one set of motion parameters in the field of view) to run successfully on the background part of the image sequence. Finally, the output of the SFM system is interpreted to give warning of non-ground-plane obstacles, and the limits of the drivable region are found.

The integrated vision system has been implemented to run in real-time (25 Hz) on a hardware setup using a combination of custom-built hardware feature detection, Transputer-based communications and PowerPC-based fast general purpose processing.

2 ASSET-2 - Moving Obstacle Detection and Tracking

The first part of the integrated system described is ASSET-2 (A Scene Segmenter Establishing Tracking, Version 2), which takes the time-evolving image sequence and analyzes this to detect and track moving objects in the world. In this section a very brief overview of ASSET-2 is given. For more details, see [8, 6, 5].

2.1 Overview

The ASSET-2 system is built around feature-based image motion estimation. The features used primarily are 2D (often referred to as "corners"), and edges can be used to refine the results obtained by using 2D feature motion. All motion estimation, segmentation and tracking takes place in the image plane.

ASSET-2 is fed a stream of digitized video pictures, usually taken from a moving platform. Each frame taken by the video camera is initially processed to find 2D features and edges. The 2D feature list is passed to a feature tracker which uses a 2D motion model to match and track features over as many frames as possible. The resulting vector field is passed to a flow segmenter which clusters the flow vectors using the model of first order flow variation (i.e., using six motion parameters within each cluster).

Next, this cluster list is compared with a temporally filtered list of clusters, and the filtered list is updated using the newly found clusters. (Each cluster has its motion parameters and boundary updated using simplified Kalman filters.) This gives good results as it stands, but the boundaries of the clusters in the filtered list can be further refined by using image edges, if available.

This list of clusters which are spatially and temporally significant is finally used to provide information about the motions of objects viewed by the video camera. The 3D positions and motions of these objects can be estimated by making simple assumptions about the world. In the case of an autonomous vehicle, this information is of practical value; for example it can enable the avoidance or the following of other vehicles.

2.2 Example Result

Figure 1 shows an example output from ASSET-2. There are two moving vehicles in the field of view of the camera, which is moving at roughly the same speed as the Landrover.

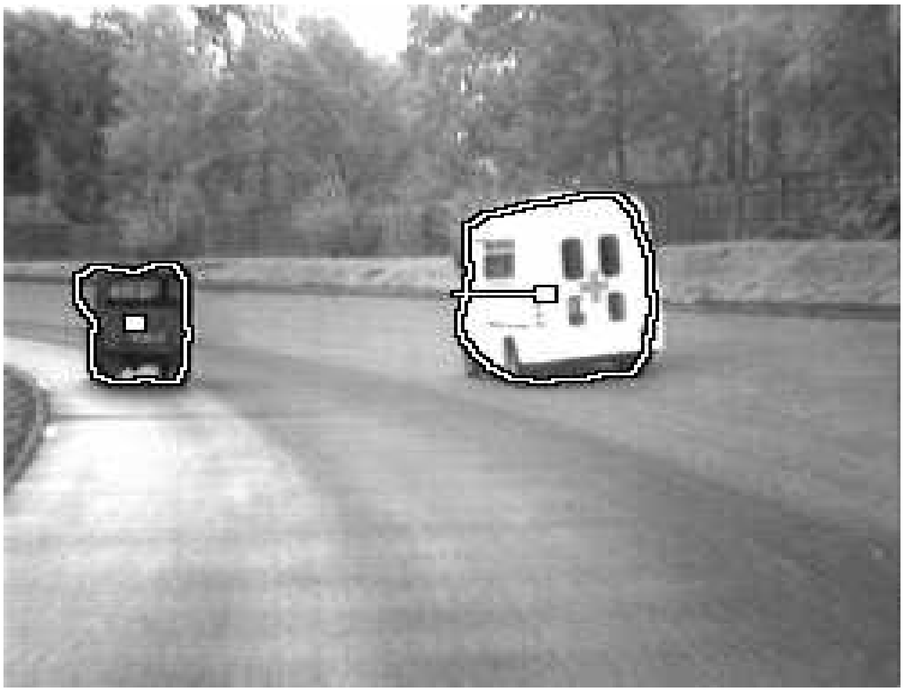

Figure 1: An example output from ASSET-2.

3 ALTRUISM - 3D Road Scene Interpretation

The second part of the integrated system is known as ALTRUISM - Automatic Local Three-dimensional Representation Using Image-based Scene Measurements. ALTRUISM attempts to provide useful automatic interpretation of sets of sparse 3D world measurements. In this section a very brief overview of ALTRUISM is given. For more details, see [3, 4]

3.1 Overview

The image processing system which provides the 3D scene information for this research is DROID, developed mainly at Roke Manor Research Limited and implemented as a real-time system at DRA Chertsey. DROID provides a list of 3D scene measurements, made at well localized places in the image, as the camera is moved through the environment. DROID also produces an estimate of the camera's own motion in the world, given only an initial space scaling factor (enabling the resolution of the scale/speed ambiguity). It works by tracking 2D features over time, using a Kalman filter for the 3D tracking of each point. For further details on DROID, see [1] and [2].

ALTRUISM takes in the resulting 3D data and camera egomotion and attempts to fit a 3D road model to an appropriate subsection of this input data in order to provide 3D data segmentation, 3D road model parameters including its boundary, scale drift correction (for DROID) and estimated steering direction control. The main elements of ALTRUISM are:

- A recursive least squares fit of a planar-quadratic "ribbon" to the three dimensional tracked features is computed, ignoring outliers. This gives an estimate of the three dimensional "road" parameters. These four parameters are temporally filtered.
- Those three dimensional features (tracked by DROID) which lie "on" the computed drivable surface are distinguished from those which do not.
- This information is used to compute a radial map of the limits of the drivable surface in the central-tangential plane of that surface. The resulting list of distances may extend further than objects which lie on it but are not perceived as forming its boundary.
- The radial map is filtered within each new estimation to reduce noise and to interpolate missing data, and also filtered temporally.
- The computed planar-quadratic ground surface is used to provide a measure of scale drift within DROID, a serious problem over large sequences unless camera calibration is completely perfect. The estimated height of the camera above the ground, assumed to be constant in reality, is used to drive a feedback loop to force this height measurement to stay constant.
- The scale drift correction allows good estimates of the camera's total journey (in three dimensions) to be made.
- An estimated steering direction command is produced from the drivable region's outline; this can be output to the ALV.

3.2 Example Results

The sequence used to produce the example outputs given in this section was taken on a seven minute circuit around a rough track at DRA Chertsey. (A map of this track forms part of Figure 2.) As explained earlier, the DROID-ALTRUISM system ran at full frame rate. For every fifth processed frame, details of the estimated camera position and of the road model parameters were output to file.

In Figure 2 the X co-ordinate of the estimated camera position is plotted against Z (i.e., this is a plan view of the camera's trajectory, in the initially established global 3D ground plane). This trajectory is shown as a thick dashed line, superimposed upon a line drawing map of the area, produced photographically for the site.

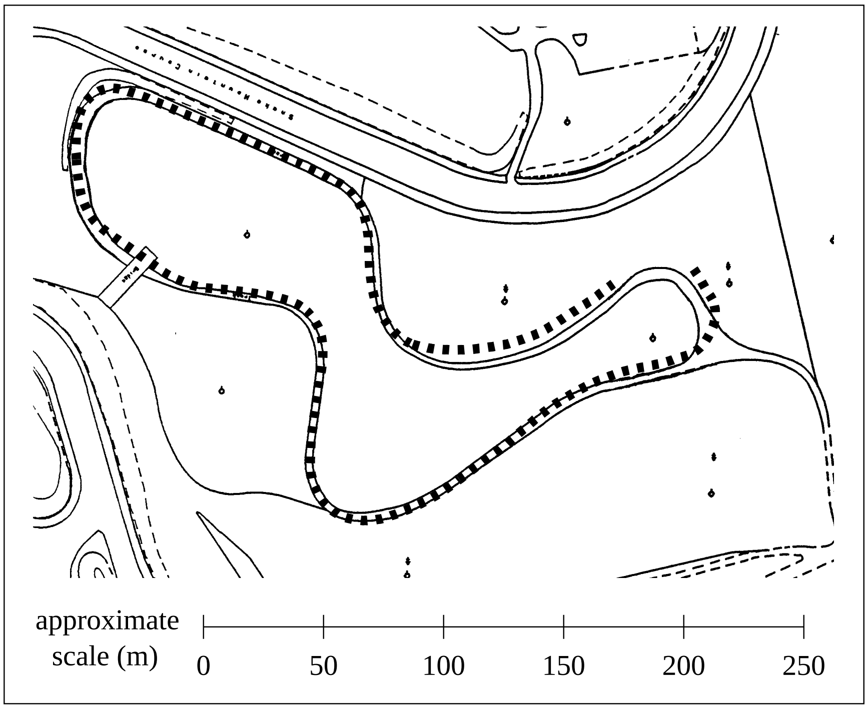

Figure 2: Estimated route taken by the camera with ALTRUISM providing scale drift correction, overlaid on a map of the area.

Figure 3 shows an example of the feature segmentation stage of ALTRUISM, on one of the images in the rough track sequence. The model fitting has provided information about which 3D features are "on road" (dark features with light borders) and which are "off road" (light features with dark borders). The discrimination is working well.

Figure 4 shows an example of the drivable region radial map estimation stage of ALTRUISM. The map is projected back into the image.

Figure 5 shows the same radial map from plan view, as well as steering angle estimation. The bottom vertex of the radial map corresponds to the camera's position, and the angle subtended by the the first non-zero radial estimates corresponds to the field of view. The vertical line indicates the local Z axis, i.e., "straight ahead". The non-vertical line indicates the predicted necessary steering direction (in the plan view frame).

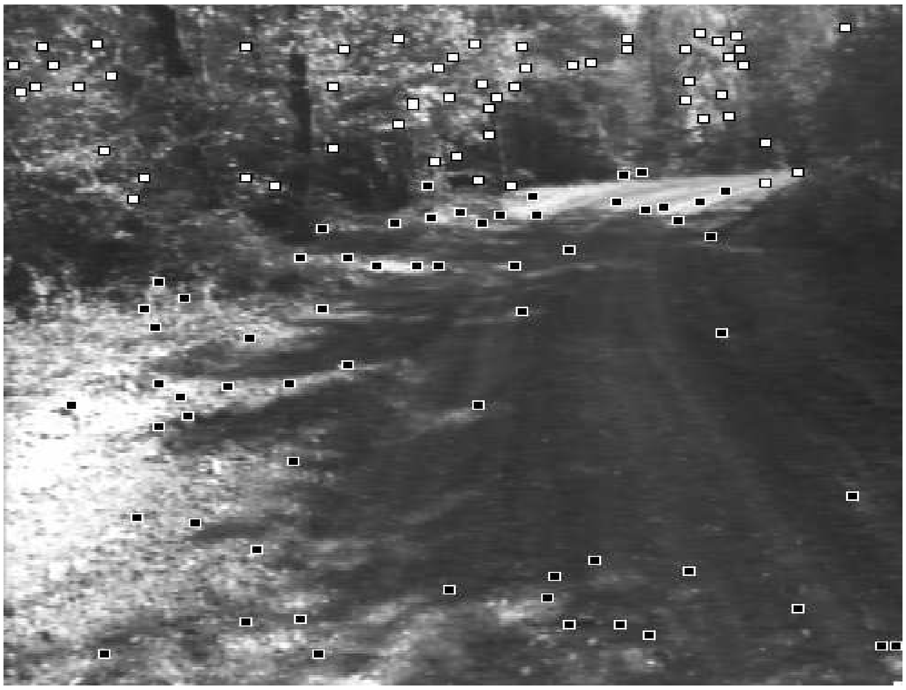

Figure 3: Tracked 3D features marked as “on” and “off” road.

Figure 4: Estimated drivable region outline.

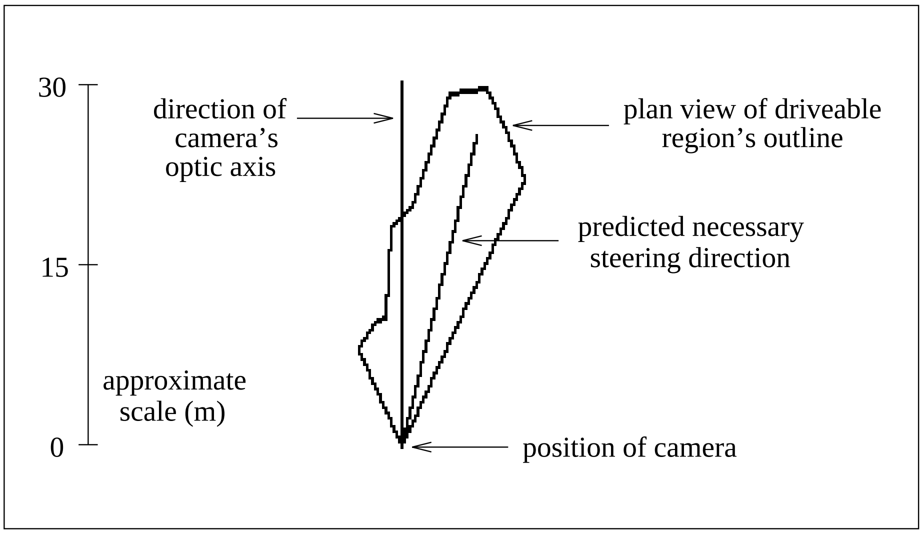

Figure 5: Plan view of the drivable region's outline, together with predicted necessary steering direction.

4 Integration of ASSET-2 and DROID-ALTRUISM

4.1 Introduction

As it stands DROID cannot cope with moving objects in the world - only one motion is allowed, i.e., the relative motion of the camera to the world. In order to be able to run the DROID-ALTRUISM system in more general situations, i.e., when moving objects are present, it is necessary to first pre-segment other moving objects out of the data which DROID will process. This can be achieved using ASSET-2, which has been adapted to track all moving objects, determine which one corresponds to the world (i.e., the background) and report to DROID all 2D image features ("corners") which lie on the world. DROID can then process these features.

4.2 ASSET-2

The first stage in this integrated system is to run ASSET-2. In this combined system ASSET-2 tracks all of the moving objects and then subtracts from the original list of 2D features those features which correspond to these objects, leaving a list of features which have been produced by the static part of the scene, i.e. the rest of the world. This list is output to DROID, which runs concurrently.

In Figure 6 an example image is shown, with the complete (original) list of 2D features superimposed on it.

Figure 7 shows the estimate of the object's outline made by ASSET-2. The number is the object number of the Landrover, as labelled internally by ASSET-2. The square near the centre of the Landrover is the tracked 2D object centre. The line coming out of this

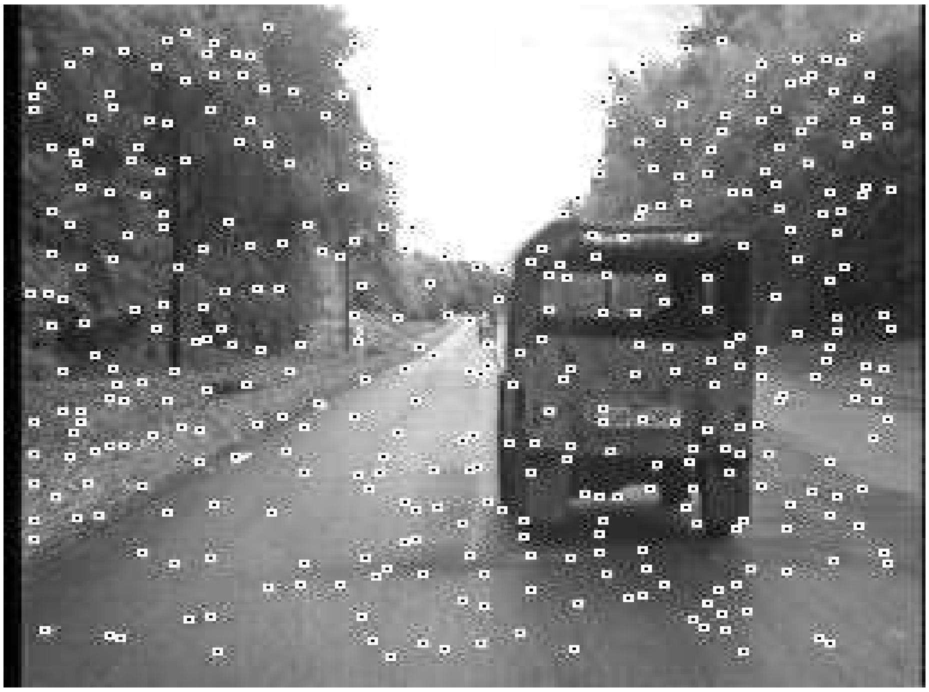

Figure 6: An example image with a large moving object travelling in front of the vehicle carrying the camera. The original set of 2D image features ("corners") is shown superimposed.

square is the overall image motion of the object, scaled by a factor of two.

Figure 8 shows the final list of 2D features output by ASSET-2; this list is then fed into DROID.

4.3 3D Processing

DROID now processes the list of 2D features which correspond to the static part of the world. It tracks these features in 3D, resulting in a list of 3D world points. Finally, ALTRUISM is fed this list of 3D points and determines which correspond to "drivable" or "road" points and which correspond to "obstacle" or "off-road" points. Note that to plan local navigation on an autonomous vehicle it is still necessary to combine the information from ASSET-2 (about the moving objects) with that from ALTRUISM.

Example outputs of the 3D processing and segmentation can be seen in Figures 9 and 10, where successfully tracked 3D points are split into "road" and "off-road" groups. The moving objects have been ignored and 3D interpretation is successful.

4.4 Real-Time Implementation

The integrated system has been implemented at DRA Chertsey to run live on the ROVA (ROad Vehicle Autonomous, the ALV at DRA Chertsey) PowerPC-based image processing

Figure 7: The outline of the moving object as estimated by ASSET-2.

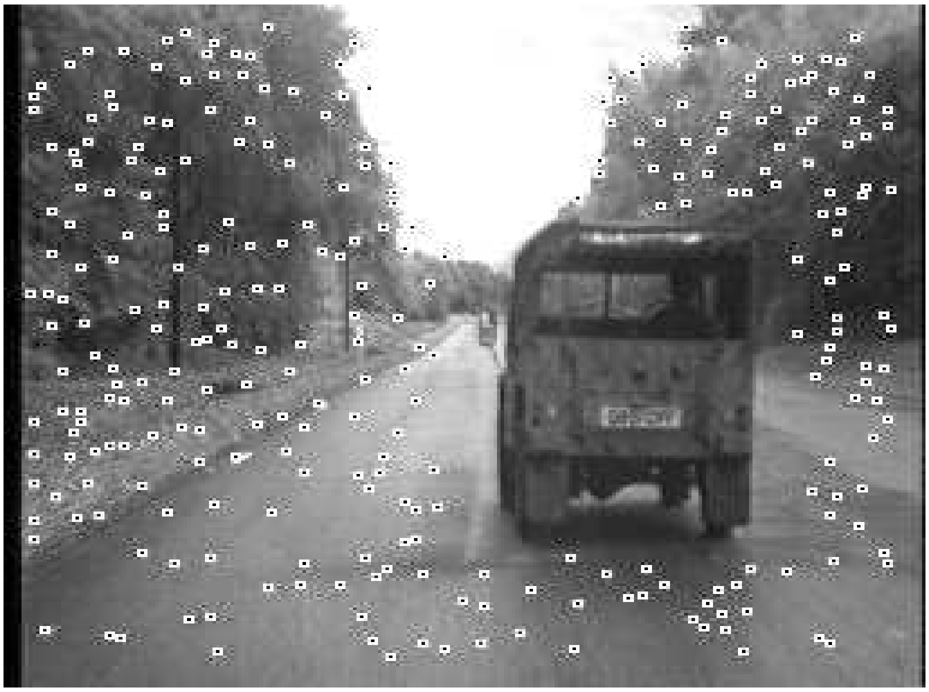

Figure 8: The final list of 2D features output by ASSET-2 to DROID.

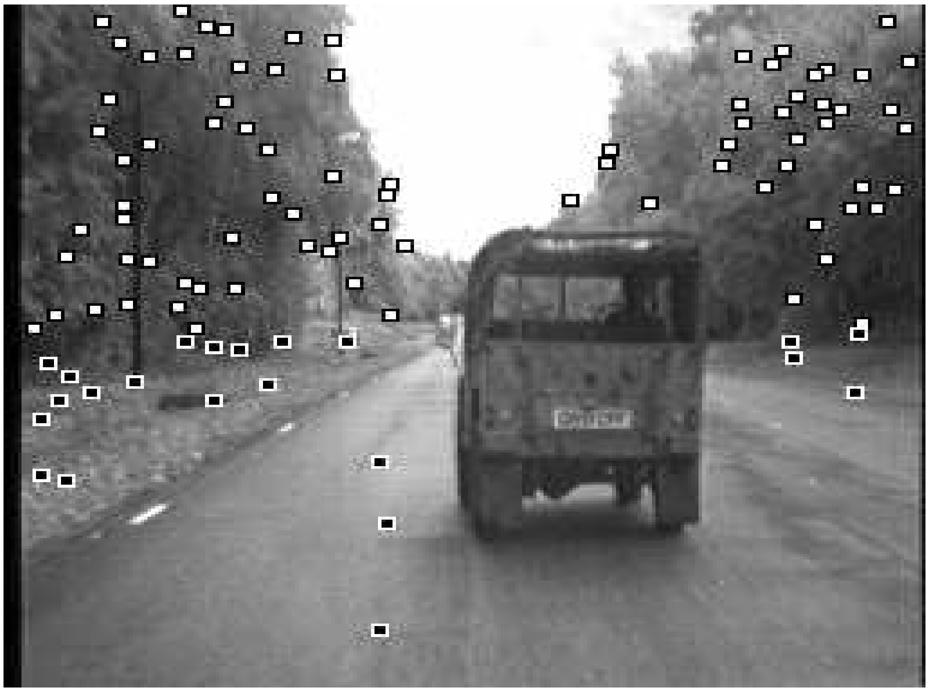

Figure 9: An example result of 3D processing followed by 3D plane segmentation.

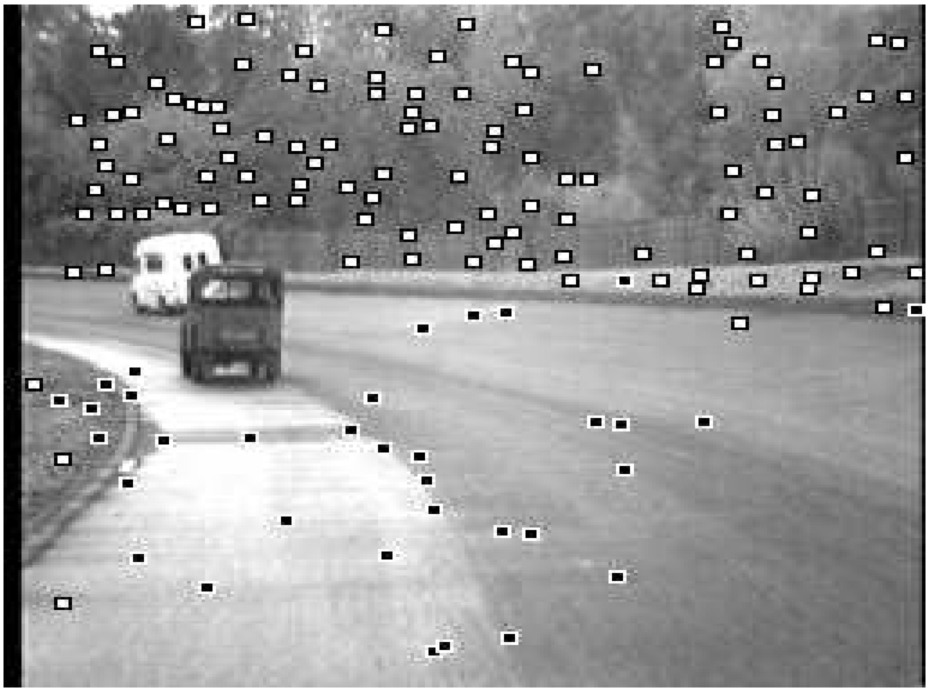

Figure 10: A second example result of 3D processing followed by 3D plane segmentation.

system (see Figure 11). Feature detection is performed by custom hardware supplied by Roke Manor Research Limited, which provides a list of 2D features at full frame rate (25 Hz), using the Harris corner finding algorithm. (Digitization and display are performed using a Datacube Digimax/Framestore VME board pair.) The feature list is sent down a Transputer link to the first of four Parsytec Transputer/PowerPC processing boards, straight into PowerPC memory.

The feature tracking stage of ASSET-2 runs on the first PowerPC board at full frame rate. The resulting vector flow field is sent down a Transputer link to the second PowerPC board, where the remainder of ASSET-2 is run, also at full frame rate. Results can be graphically overlaid onto the original images for real-time display.

The 2D features corresponding to the background in the image sequence are then sent to the third PowerPC board, where DROID runs at full frame rate. DROID then sends the list of 3D points and the 3D camera position to ALTRUISM, which runs on the fourth PowerPC board, also at full frame rate. Again, results can be graphically overlaid onto the original images from this board.

5 Conclusions

The integration of a motion segmentation system with a 3D interpretation system has successfully dealt with scenes which contain moving objects, providing high level interpretation of the 3D world structure in real-time. This is a step forward and upward in the overall problem of image sequence interpretation.

References

[1] C.G. Harris and J.M. Pike. 3D positional integration from image sequences. *Image and Vision Computing*, 6(2):87–90, 1988.

[2] S.M. Smith. *Feature Based Image Sequence Understanding*. D.Phil. thesis, Robotics Research Group, Department of Engineering Science, Oxford University, 1992.

[3] S.M. Smith. ALTRUISM: Interpretation of three-dimensional information for autonomous vehicle control. *Engineering Applications of Artificial Intelligence*, 8(3):271–280, 1995.

[4] S.M. Smith. ALTRUISM: Interpretation of three-dimensional information for autonomous vehicle control. Internal Technical Report TR95SMS3, Defence Research Agency, Chobham Lane, Chertsey, Surrey, UK, 1995. Available at www.fmrib.ox.ac.uk/~steve for downloading.

[5] S.M. Smith. ASSET-2: Real-time motion segmentation and object tracking. Internal Technical Report TR95SMS2, Defence Research Agency, Chobham Lane, Chertsey, Surrey, UK, 1995. Available at www.fmrib.ox.ac.uk/~steve for downloading.

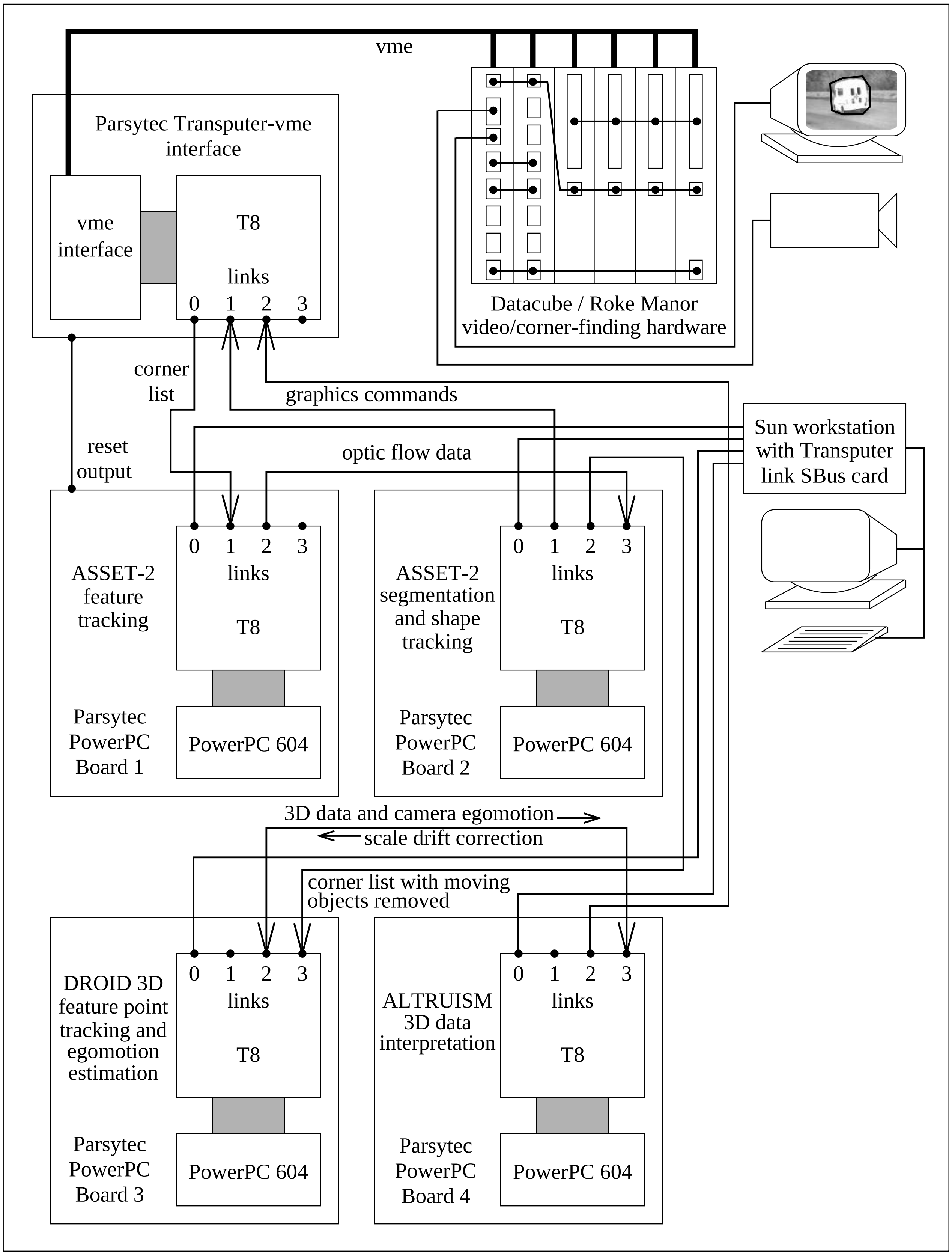

Figure 11: The real-time implementation of the integrated ASSET-DROID-ALTRUISM system on the ROVA PowerPC-based image processing system.

[6] S.M. Smith. ASSET-2: Real-time motion segmentation and shape tracking. In *Proc. 5th Int. Conf. on Computer Vision*, pages 237–244, 1995.

[7] S.M. Smith. Integrated real-time motion segmentation and 3D interpretation. In *Proc. 13th Int. Conf. on Pattern Recognition*, volume 3, pages 49–55, 1996.

[8] S.M. Smith and J.M. Brady. ASSET-2: Real-time motion segmentation and shape tracking. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 17(8):814–820, 1995.