REAL-TIME VISION-AIDED TRAFFIC MONITORING

E. L. Dagless* and A.T. Ali*

Image processing systems are becoming more suitable for surveillance and traffic monitoring. The paper reports on some techniques for monitoring traffic flow on motorways, using a moving object detector operating in small windows. Traffic monitoring is demonstrated using a relaxed coincidence search technique, and successful results are reported from video of a busy complex junction layout.

1. Introduction

The continuing rise of traffic levels on our roads has been and is a major problem in every day life, not only for transport requirements, but also due to its impact on the economy, road safety, and the damage it can cause to the environment. In order to tackle such rapid traffic growth improved methods of traffic monitoring are needed. Because of the serious limitation of conventional traffic detectors (e.g. inductive loop, humans), attention has shifted towards the use of advanced micro-electronic and computing technology. As well as being suitable for point monitoring these new techniques have potential for wide area surveillance, giving capabilities to track vehicles and pedestrians and to perform global traffic monitoring. Computer vision systems appear to be full of promise and extensive research is being pursued to achieve these aims (Inigo, 1985).

In the Advanced Computing Research Centre at the University of Bristol, a computer vision system has been developed for real-time vision applications (Thomas et al, 1991). Experiments on road traffic monitoring and data collection have been carried out. Traffic and pedestrian flow measurements have been investigated and real-time performance achieved in tracking cars and pedestrians (Ali and Dagless, 1990). The system is capable of counting vehicles, estimating speed, recording statistics on paths through junctions etc. Several field trials have been conducted on roads around the Engineering Laboratories of Bristol University and some sections of the M5 motorway.

2. Image Sensing and Perspective Transformation

The image sensing method, proposed here, is based on a video camera mounted over the road surface on a highway overbridge or a nearby high building (Ali and Dagless, 1990b). The viewing geometry is depicted in Figure 1. When the camera is tilted in the plane of the Z-axis to point to the road the camera co-ordinates change to:

* Advanced Computing Research Centre, University of Bristol, University Walk, Bristol BS8 1TR, UK

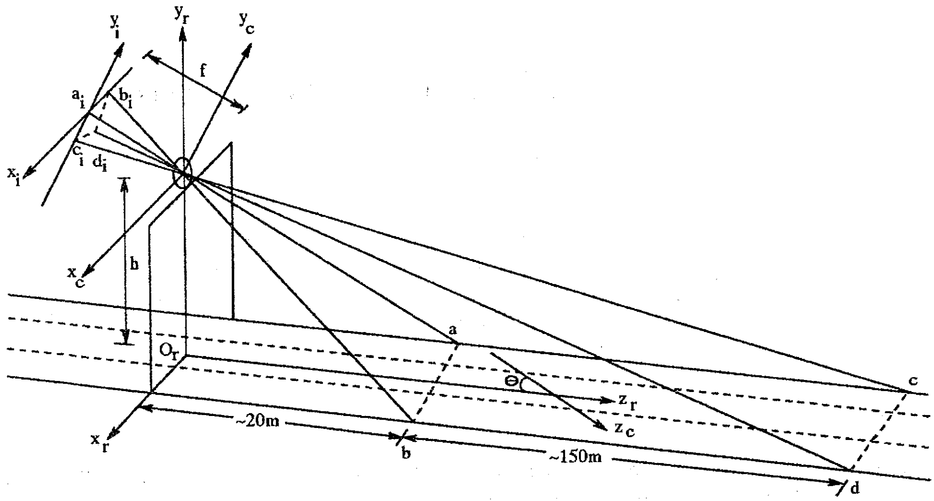

Fig. 1. The viewing geometry of a video camera mounted on a motorway overbridge.

$$\begin{bmatrix} x_c \\ y_c \\ z_c \\ 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & Cos\theta & Sin\theta & 0 \\ 0 & -Sin\theta & Cos\theta & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x_r \\ y_r \\ 1 \\ 1 \end{bmatrix}$$

When the camera is moved vertically a distance h, the following translation matrix applies:

$$\begin{bmatrix} x_c \\ y_c \\ z_c \\ 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & -hCos\theta & Sin\theta & 0 \\ 0 & -Sin\theta & Cos\theta & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x_r \\ y_r \\ 1 \\ 1 \end{bmatrix}$$

These perspective transformation equations which define the relationship between the real-world, camera and image co-ordinates can be used to obtain measurements in the real-world from points in the image plane provided the system has been calibrated correctly.

The calibration can be done automatically on site using objects of known size in the field of view of the camera. These tests need to be done close to the camera and at a distance to ensure that accurate determination of the angle of tilt is obtained since small errors can have a large influence on the perspective effects and measurements at a distance from the camera. Furthermore the camera must be mounted as high as possible so as to reduce occlusion between obstacles which can affect the performance of the system and to give a wide coverage if traffic flow is to be monitored.

3. The System Architecture

To meet the real-time requirements of the application the computer vision system has to have fast data transfer and high speed processing capabilities. The design is based on the idea that raw image data is 'grabbed' directly into the local memory of a transputer or transputers. Some of the most low-level vision tasks are implemented in hardware so as to allow a transputer array to deal with the other vision tasks more efficiently. The development of the vision system has followed this philosophy leading towards the development of a 'heterogeneous' system which is described in more detail below. Figure 2 shows a diagram of a general configuration of the vision system. The hardware structure is described and the major software modules are explained.

Video interface. Using A/D and D/A devices mapped into input and output lookup tables, the video signal is digitised into 256 * 256 * 8 bit grey-level pixels at 50 field/sec, whilst the output can be displayed on RGB video monitor. The digital video signals are conveyed over dedicated video busses between the digitiser and the

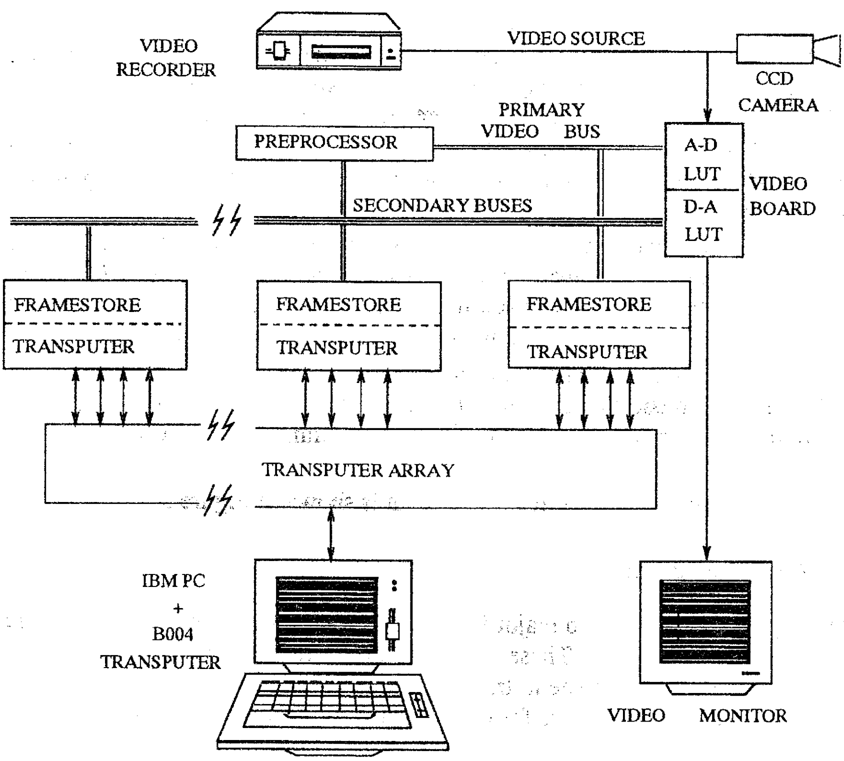

Fig. 2. A schematic diagram of the vision system.

framestores. Video input and video output have separate channels. A new version of this module has been developed which uses MAXBUS compatible digital video signals (Thomas et al, 1991).

Image pre-processor. This unit is intended to perform low-level vision tasks such as filtering, convolution, image comparison, etc. No image pre-processor was used in the experiments reported here although this unit is available in the new system becoming available soon.

Framestores. Images are provided within the address space of the transputers through dual-ported, double-buffered frame store modules connected to the digital video busses. The VME 32-bit bus-backplane is used for data transfer between the frame store and the transputer modules The double buffering and dual ported memory allows the framestore to be loaded from the digitiser while the transputer processes the previous image. Furthermore the display module can receive output frames at the same time. This provides a very responsive and highly effective debugging environment for developing image processing algorithms. Multiple frame-store boards can be used to increase the storage of captured frames. Furthermore, multiple transputer modules can share a single framestore. A newer version of the system (Thomas et al, 1991), which is MAXBUS compatible provides frame store and transputer on a single board.

Transputer boards. The transputer modules that have access to the framestore also have 1Mbytes of local memory. They are responsible for the lowest level of processing and the management of transfers of image patches to the array of transputers for further processing tasks. In the experiments reported here only one transputer had access to the framestore and as well as performing the processing of some image data it was responsible for passing small patches of the image to other transputers so they could help with the low level image processing tasks.

Transputer array. This part of the system is a collection of standard transputer modules (TRAMS) located on motherboards, as appropriate, housed inside PCs. They are connected up as required for the experiments under investigation. Their code is provided by the host and data is provided by the other transputers as required.

Host processor. This is a standard transputer module located inside a PC. It is responsible for all system control functions and provides for access to disc storage and other conventional I/O resources. It loads all the other transputers with code at boot time.

A picture of a small configuration of the system is shown in Figure 3.

4. Software Arrangements

The software is partitioned into two major inter-communicating components: the image analyser and the traffic analyser. These two tasks are pipelined so that while image analysis is being performed on frame n, traffic analysis is being performed on frame n-1. This allows a complete frame time for each of these major tasks to be performed to achieve real-time operation.

Image analyser. This software module which is mapped onto the transputer with the framestore and a few other to which image patches are transferred. The task of the

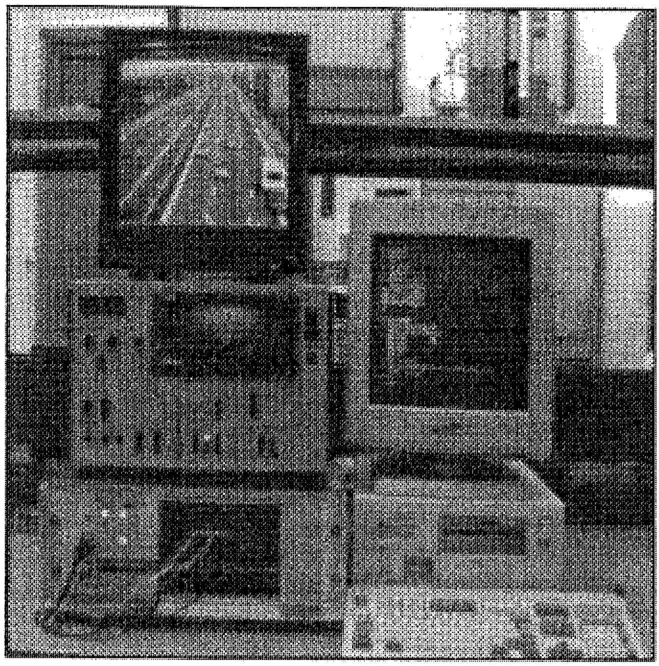

Fig. 3. A prototype vision system in use.

image analyser is to extract relevant data from the image sequences that is then passed on for traffic analysis. There are algorithms for detecting vehicles within regions of interest, for tracking the motion of vehicles, and for extracting motion trajectories.

Traffic analyser. This software module receives the image parameters from the image analyser and computes the traffic information. It translates the vehicle trajectories into real world co-ordinates. From this information it is then able to compute traffic parameters like traffic volume, speed, headway, and occupancy.

The rest of the paper describes these tasks in more detail and reports on some results obtained using the architecture described.

5. Vehicle Detection

The method of vehicle detection uses small regions of interest in which salient events are sensed (Ali and Dagless, 1991). This avoids the need to process a whole image, which would seriously impair real-time performance. However, the placing of these windows is critical if successful detection of events is to be achieved. The size of the windows is also an important parameter: too large and it becomes harder to capture a single vehicle, (and the method then fails), and performance is impaired with the need

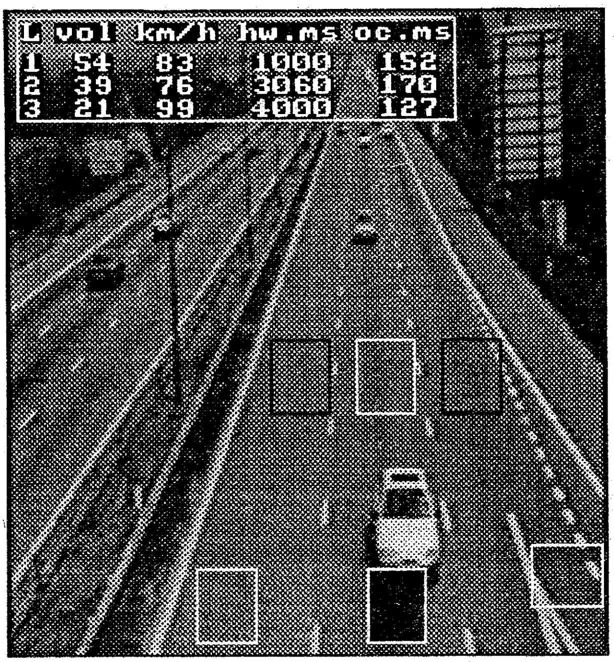

Fig. 4. A hard copy of the system monitor showing real-time data collection of motorway traffic from a motorway overbridge.

to process many pixels; too small and accuracy is compromised because it is more likely to not see the front (or back) of the vehicle for accurate location and timing, the field of view is reduced, and reliability is reduced because detection can be more unpredictable. In some cases fixed windows are used for interval timing, (Fig. 4), and these must be far enough apart to give accurate timing, but close enough to ensure only a single vehicle is within the active region. For vehicle tracking the window is located on the point of entry, and the vehicle is then tracked dynamically until it leaves the scene. There will be 4 or more windows in a typical scene at a junction, or 6 or more in multi-lane monitoring applications, e.g. Figure 4.

The detection of a vehicle is the first task. The segmentation is based on two frame differencing methods; reference frame differencing and interframe differencing. In reference frame differencing a background reference frame is subtracted from each input frame. The resulting difference image is thresholded to produce a binary image that is then filtered using a dilation/erosion process, so that small, random binary blobs are eliminated. Bright parts of the final image are objects moving against the background scene. The reference frame is captured and updated using robust techniques (Ali and Dagless, 1992). To capture the reference a 'MODs' filters (see the next paragraph) search sub-divisions of the dynamic scene and transfers any empty areas into the reference image. Thus initialisation of the reference is automatic and quick provided all parts of the scene are static for at least a small part of the time.

Any reference will quickly become incorrect, except in artificially lit scenes, and it must be updated if reliable performance is to be achieved. The reference is refreshed by updating individual points within small patches on each cycle. By adjusting the update rate the refresh can be made to suit the traffic conditions and the rate of brightness changes. However this can consume significant processing power if the update rate is high. Therefore, a second correction is performed to provide a more immediate compensation for when rapid changes in lighting occur. The change in brightness in known 'static zones' is monitored and the information adjusts the threshold used to convert the difference image into a binary image. In time, the refreshing of the reference frame will adjust for the change that occurred and the correction effect will gradually reduce.

In some situations the reference frame may be difficult to obtain, or simply not work due to rapid scene changes. In which the second method can be used. Usually both techniques will be used together to improve reliability. The second method is interframe differencing and is achieved by subtracting the present image from the previous frame. Any object that has moved will be accentuated at the edges where there has been a change in the pixel intensity. Thus the front and back of moving vehicles will be seen; often the central part of a vehicle will not be visible. Thus the method complements the first method because, while it will always find moving objects the image will often not allow the whole of the object to be identified.

In order to recover the intermediate pixels of the vehicle difference image which are weakened during the differencing process, a combination of gradient edge detection and coincidence operation has been used. The order in which these operations are applied to consecutive frames produces different results. These different constructions are called 'moving object detectors' or 'MODs'. An example of a 'MOD' is:

[Thresholded Sobel (current frame)] * [Thresholded Sobel (current frame - previous frame)].

Full details of the construction and the performance of a range of MODs and the reference frame differencing technique can be found in (Ali and Dagless, 1992). A suitable 'MOD' filter is used for capturing the static parts of the scene to obtain the reference frame (see above).

6. Traffic Tracking

Once a vehicle has been located within a window it can be tracked by moving the window region and relocating the vehicle in the new window. Prediction of the new window position is based on the vehicle trajectory and its offset within the window. The front, back or centroid of the vehicle may be used (Ali and Dagless, 1990). The centroid can only be used when the whole of the vehicle is within the window. However, all three position indicators are prone to error when trying to relocate a vehicle carrying out a complex manoeuvre. When tracking, the window size is adjusted to a sub-image which surrounds the object. This improves the system performance because only the pixels within the window are used in the computation. Tracking proceeds as long as the

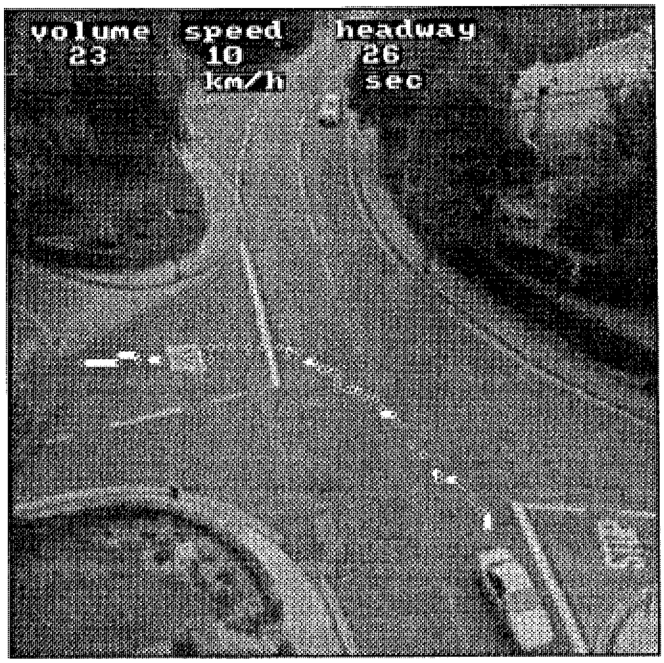

Fig. 5. A hard copy of the system monitor showing a turning manoeuvre at a complex junction being tracked in real-time.

vehicle is within the camera's field of view. At the end of each algorithm cycle, (usually a frame time), vehicle trajectories are sent to the host transputer to be converted into various traffic parameters (Fig. 5).

These methods work well for tracking vehicles that are well separated. However, for situations like motorway traffic in the distance, or traffic on a roundabout or complex junction where vehicles travel closer together or are occluded by others, the tracking algorithm can be confused and can often pick up the wrong vehicle. To overcome these problems a more robust method for vehicle re-identification is used. A suitable technique is A Relaxed Coincidence Search Technique, (ARCST). The method uses a variant of the correlation method using a template of the vehicle and the current image (Ali and Dagless, 1992).

The experiments show that the main sources of error are due to a change in the scene brightness, or unwanted matches between the template and similar image areas from the background or other objects in the scene. The technique is modified to reduce such errors by excluding the background from the correlation process. This can be done if the raw image of a current frame is 'ANDed' with a segmented version of the current frame containing the object features only. The segmentation can be done using reference frame differencing or MODs.

The resultant ANDed image, which is a raw object image or raw object edge image, is correlated with the template created from the object in the previous frame, using either the search area or sub-search area coincidence technique as explained above. Since the edge pixels of an image have a high contrast and are less affected by a change in brightness, the method is less sensitive to lighting changes. Performing the correlation within a well defined search window located in the region of the predicted position of the object reduces the likelihood of an unwanted match with such other objects.

7. Concluding Remarks

Two techniques of visually monitoring road traffic are presented in this paper. The first monitors the movement of vehicles between two static regions of interest (ROI) pre-located on the road image. Using the reported image processing techniques employing 6 vehicle detection windows the system can monitor traffic flow on a multi-lane motorway. From timing information and knowledge of the site geometry useful traffic data is extracted. Real-time traffic detection is performed using up to nine transputers to monitoring three lanes of a motorway, (Fig. 4). Using nine parallel processors, the system operates at video rates, algorithm speed-up is 5.5, and processor efficiency is about 65%. Initial experiments achieve an accuracy of 90 to 95% detecting vehicles moving at speeds up to 80mph.

The second technique uses a tracking algorithm to follow moving vehicles. Then their trajectories are transformed from the image domain into real-world measurements for conversion into traffic data. Real-time vehicle tracking has been achieved by successful prediction of the vehicle position within a well defined search area. The recent results of applying the ARCST technique, which is intended for object re-identification, have shown that good matching can be achieved between a template object image and a corresponding object image even when there are changes in the object orientation. These techniques can cope with partially occluded objects and those with different scales and resolutions, but more experiments are required to improve the stability of the technique when faced with changes in the brightness.

References

Ali A.T. and Dagless E.L. (1990): *Vehicle and pedestrian detection and tracking*.- IEE Colloquium on Image Analysis for Transport Applications, IEE, Savoy Place, London, February, Digest 035, pp.5/1-7.

Ali A.T. and Dagless E.L. (1990b): *Computer vision for security surveillance and movement control*.- Proc. IEE Colloquium Electronic Images and Image Processing in Security and Forensic Science, IEE, Savoy Place, London, May, Digest 087, pp.6/1-7.

Ali A.T. and Dagless E.L. (1991): *Computer vision-aided road traffic monitoring*.- Proc. ISATA 24th. Int. Conf. Road Transport Informatics (RTI) & Intelligent Vehicle-Highway Systems (IVHS), Florence, Italy, May, pp.55-61.

Ali A.T. and Dagless E.L. (1992): *Alternative practical methods for moving object detection.-* Proc. IEE 4th. Int. Conf. Image Processing and its Applications, Maastricht, Holland, April, pp.77-80.

Inigo R.M. (1985): *Traffic monitoring and control using machine vision: a survey.-* IEEE Trans. on Industrial Electronics, v. IE-32, No.3, August, pp.177-185.

Thomas B.T., Dagless E.L., Milford D.J. and Morgan A.D. (1991): *Real-time vision guided navigation.-* Engng Applic. Artif. Intell. v.4, No.4, pp.287-300.