VEHICLE DETECTION IN TRAFFIC IMAGES

Luigi Di Stefano
DEIS, University of Bologna, Viale Risorgimento 2, 40136 Bologna, Italy - ldistefano@deis.unibo.it

Enrico Viarani
DEIS, University of Bologna, Viale Risorgimento 2, 40136 Bologna, Italy - eviarani@deis.unibo.it

Key words: Applications of Pattern Recognition, Computer Vision, Traffic Monitoring

Abstract: This paper describes a method for detecting vehicles in traffic images which relies on motion estimation provided by a Block Matching Algorithm (BMA). Once motion has been estimated via BMA, the motion field is regularised by means of a Vector Median Filter. Finally, vehicles are detected by grouping together image blocks with similar motion. The main contributions of the paper consist in an effective approach to prevent wrong matches in static areas which yields also a significant computational saving and the use of an adaptive vector median filter aimed at avoiding vehicle erosion.

1. INTRODUCTION

This paper describes the research activity currently going on at DEIS, University of Bologna, on the deployment of computer vision techniques for extraction of traffic information from images. The perspective goal of this activity is the development of a low-cost, vision-based traffic monitoring system capable of measuring fundamental traffic parameters, e.g. queue length at intersections, as well as recovering higher level information such as for example assessment of the traffic situation and congestion detection.

In this framework, one basic problem consists in the development of fast and robust algorithms for vehicle detection and tracking. In this paper we address vehicle detection only. According with related work, we have chosen to rely on the extraction of motion from image sequences so as to segment out vehicles with respect to still objects (e.g. road, buildings...). In particular, our approach is based on the Block Matching Algorithm (Musmann 1994) (abbreviated as BMA), which is used for motion estimation in the MPEG (Baglietto 1997) image compression standard. BMA has been employed for vehicle detection in an application aimed at evaluating turning rates at urban crossroads (Bartolini 1996). Unlike techniques based on spatio-temporal derivatives (Koller 1994) or background subtraction (Michalopulos 1991), BMA provides not only detection of moving pixels but also motion estimation (i.e. a motion vector is associated with image points). This improves the discrimination power in case of partially occluding vehicles and crowded traffic scenes. In addition, although Block Matching is computationally expensive, specialised hardware for BMA-based real-time motion estimation is available nowadays (DeVos 1989). Hence, in a traffic monitoring system the low-level motion estimation task could be carried out by a fast hardware front-end, with the CPU in charge of analysing the extracted motion information.

Once motion has been estimated via BMA, we regularise motion data by means of a Vector Median Filter (abbreviated as VMF) and then detect vehicles by grouping together image blocks having similar motion characteristics.

2. THE BLOCK MATCHING ALGORITHM

BMA consists in partitioning each frame of the sequence into square blocks of fixed size (e.g. 6x6, 8x8,...) and detecting blocks displacement between the actual frame (at time *t*) and the previous one (at time *t-1*), searching inside a given scan area. Each block encloses a part of the image and is represented by a matrix containing the grey tone values of that part of the image. Fig. 1 shows the partitioning of the actual frame and a block, which is going to be found in the previous frame. The displacement vector, *"DV"*, attached to the block will be the result of the search.

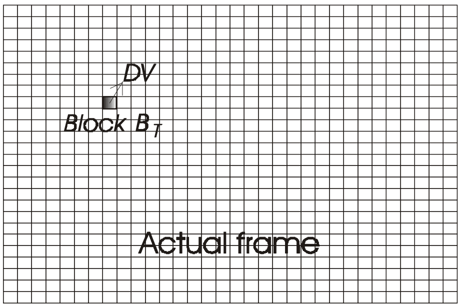

Fig.1 - Frame partitioning and DV

The matching process is described in Fig.2 Each block resulting from the partitioning of the actual frame, B_t, defines a scan-area in the previous frame; the template-block B_t is shifted pixel-by-pixel inside this scan area and compared with the blocks of the previous frame, B_{t-1}, found at each position of the shift. The comparison is aimed at determining the most similar to B_t between the B_{t-1}s .

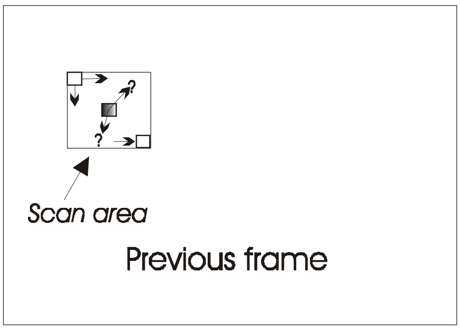

Fig 2 - Matching process.

The size of the scan area should account for the highest possible block displacement between two frames. However, since computation time largely depends on this size, it is convenient to keep it as small as possible. The similarity between blocks can be evaluated on the basis of several matching criteria (Musmann 1994) among which we have adopted the Normalised Cross-Correlation Function:

$$NCCF = \frac{\sum_{i,j}[B_{t-1}(i,j)\cdot B_t(i,j)]}{\sqrt{\left[\sum_{i,j} B_{t-1}{}^2(i,j)\right]\cdot\left[\sum_{i,j} B_t{}^2(i,j)\right]}}$$

where "·" represents the product between the corresponding elements of matrixes B_t and B_{t-1}, and indexes *i,j* rolls all over the matrix size. The matching process yields a displacement vector between the position of B_t and that of the best matching B_{t-1}. This vector is considered to be applied in the centre of B_t and represents an estimation of block's position for the successive frame, at the time $t+1$. The set of displacement vectors found into the actual frame is called "Displacement Vector Field", DVF, of the frame.

Unfortunately, due to noisy grey tone fluctuations, BMA generates many wrong vectors over static blocks located on the background of the road lane. This is typically handled by biasing the selection of the best matching displacement towards the null displacement (ITU-T 1995); in particular, a fixed value is subtracted to the mismatch measure calculated for the same position in the previous frame (i.e. the null displacement case). Hence, for those blocks having not a reasonably good matching position within the scan area, the displacement vector will tend to be null.

Yet, with this approach the matching process is still carried out over the whole scan area even though the displacement vector turns out to be null. Instead, we have addressed the problem so to achieve also a significant computational saving. In fact, we calculate first the correlation of a block with the block in the same position of the previous frame (i.e. null displacement) and compare the result with a fixed threshold: if the correlation is higher than the threshold, then the null displacement vector is chosen for the block; otherwise we proceed by calculating the correlation over the scan area and then searching for its maximum value. By using a relatively low threshold we favour null vectors in low textured static areas; moreover we compute just a single correlation step when the vector turns out to be null.

Fig. 3 - This picture contains less than 5% of non null vectors.

Our experimental results show that this approach is effective in preventing wrong matches within static areas, yielding at the same time a computational saving for the Block Matching Algorithm which is usually of the order of 90%. For example, let's consider a scan-area size of 20x20 pixels and a block size of 8x8 pixels. The matching process for each block requires $(20)^2 = 400$ cross-correlation calculations. In the 512 x 512 pixels picture shown in Fig. 3, there are about 180 non-null displacement vectors out of 4096 ones. Hence, our approach yields

```
4096*400 =          1638400-  (std BMA)
(180*400)+4096-180= 75916=  (our method)
--------------------------
                    1562484,
```

saved cross correlation calculations per frame. Therefore, by calculating just a single correlation when the displacement vector is null, in the case of Fig. 3 we obtain a computational saving as high as the 95.6%.

3. VECTOR MEDIAN FILTERING

Even if we set a very selective cross-correlation threshold in the BMA, we usually obtain a noisy DVF due to the unavoidable grey tone fluctuations of moving objects between successive frames as well as to excessive vehicle speed. Noisy vectors may cause significant over-segmentation of vehicles (i.e. splitting of a single vehicle into different objects) during the grouping process (which will be described in Section 4). Fig. 4 shows the DVF associated with the vehicle into the rectangular black window of Fig. 3 as well as the result of the grouping process, with the vehicle split into two objects due to the noisy vectors.

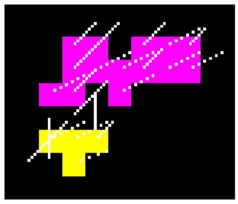

Fig. 4 - Over-segmentation of a vehicle

Since the noise found in the DVF may be modelled as the presence of outliers into a rather regular vector field, we have chosen to rely on vector median filtering to smooth DVFs. This approach has been employed for regularisation of velocity (Astola 1990) and optical flow fields (Bartolini 1993) as well as DVFs (Bartolini 1996). The median value of a set of n scalars is defined as $(n/2)^{th}$ element of the ordered set. Evidently, this definition does not embody any straightforward extension to vector-valued signals. However, Astola et. al. have introduced a "median" operator for vector valued signals derived from a bi-exponential two-dimensional probability density function using the maximum likelihood approach (Astola 1990). This vector operator has properties very similar to those of the standard median for one-dimensional signals (Astola 1990). The vector median operator by Astola et. al. is obtained by taking the element of a set of vectors which minimises the sum of the distances from all the other elements, with distances evaluated on the basis of the L_2-norm. In the $\Re^2$ space, the L_2-norm of a vector $\boldsymbol{v}(x_v,y_v)$ is expressed as:

$$\|\vec{v}\|_{L_2} = \sqrt{x_v^{\ 2} + y_v^{\ 2}}$$

and the distance between two vectors, $\boldsymbol{u}(x_u,y_u)$ and $\boldsymbol{v}(x_v,y_v)$ based on the L_2-norm is given by:

$$\|\vec{u} - \vec{v}\|_{L_2} = \sqrt{(x_u - x_v)^2 + (y_u - y_v)^2}$$

In our application of the vector median operator, each displacement vector is replaced by the vector median of the set given by the vector itself and its eight nearest neighbour's displacement vectors. Fig. 5 shows an example of the regularisation action provided by the filter.

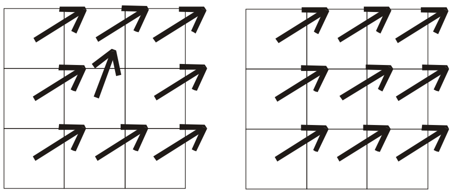

Fig. 5 - Vector median based on an 8-neighbourhood

Regularisation by vector median filtering significantly reduce the noise and produces sets of blocks having more similar displacement vectors.

4. THE BLOCK-GROUPING STEP

After obtaining the restored DVF, we detect vehicles by grouping together connected clusters of blocks having similar (and non-zero) displacement vectors. The grouping process is basically a sequential, iterative connected components labelling algorithm (Haralick 1992) in which the blocks having non-null displacement play the role of "foreground points" and two blocks are considered neighbours if they are 8-connected and the L_2-norm of their displacement vectors is less than a threshold, *Th*:

$$\|\vec{u} - \vec{v}\|_{L_2} = \sqrt{(x_u - x_v)^2 + (y_u - y_v)^2} < Th$$

This process is sometimes referred to as δ-labelling. Moreover, we have chosen to filter out the groups made out of very few blocks in order to increase robustness with respect to noise. The value of the threshold on the group size depends on the size of the block with respect to the size of vehicles.

5. THE ADAPTIVE VECTOR MEDIAN FILTER

We have found that vector median filtering severely erodes most of the connected components present in the DVF. As a result, several vehicles may be split, recovered very poorly or even lost, during the grouping process. Fig. 6 shows the DVF, as well as the result of grouping, for the vehicle of Fig. 3 after the application of vector median filtering. If we compare this with Fig. 4 it is evident that median filtering cause significant erosion of the vehicle shape.

Fig. 6 - Erosion of a vehicle due to median filtering

The problem arises because we deal with vector fields having a large number of null vectors and relatively few non-null vectors, with outliers among the latter ones. Moreover, since BMA gives few displacement vectors over frequent low textured areas like an even colour body (an example in shown Fig. 7), it is important to avoid to miss vectors which shall be useful to grant DVF's connection.

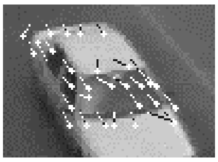

Fig. 7 - A sparse DVF over a low textured vehicle

If we simply median-filter the DVF, most of non-null vectors are changed into null-ones since neighbourhood tends to be dominated by null vectors. Fig. 8 shows an example of this: the vector in the centre of the "8-neighbourhood" is going to be deleted rather than made similar to its non-null neighbours because it has more null neighbours than non-null ones.

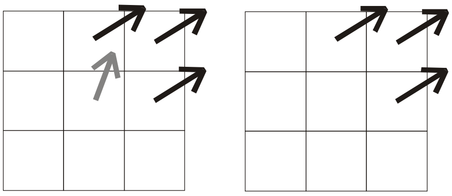

Fig. 8 - Action of the median on a sparse DVF

Therefore, we have devised an "adaptive" filter in which only non-null vectors are taken into account during the smoothing process. That is, each non-null vector is turned into the vector median computed on the basis of its non-null 8-neighbours. This operator may be viewed as an adaptive statistical order (or rank) filter (Shalkoff 1989) in which the order changes as a function of the local characteristics of the signal.

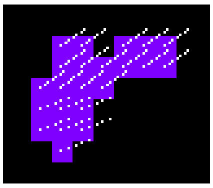

Fig. 9 - The vehicle of Fig. 3 after adaptive filtering and grouping

Fig. 9 shows the vehicle of Fig. 3 after adaptive median filtering and grouping. The DFV has now been regularised without causing erosion of the vehicle, which is now detected as a single moving object.

6. CONCLUDING REMARKS

We have done experiments using video sequences available from the Internet (http://i21www.ira.uka.de/image_sequences/) as well as sequences obtained from several traffic monitoring and surveillance cameras installed throughout the city of Bologna, and kindly provided to us by the Traffic Bureau. Our experimental results shows that the developed procedure (i.e. BMA with the correlation threshold, adaptive filtering and grouping) is effective in extracting vehicles. However, there can be still some over-segmentation effect when vehicles are too big with respect to block size. Examples are buses or, when images are significantly affected by perspective, vehicles very close to the camera (see Fig. 10).

Fig. 10 - A vehicle close to the camera split into 3 objects

To deal with this problem we plan to develop a region-merging algorithm to be run after the grouping step. This algorithm might work differently for near and far regions so as to deal with perspective distortion. In addition, future work on vehicle detection will address night images and the vehicle tracking problem.

REFERENCES

Astola, J., Haavisto, P., Neuvo, Y, 1990, *Vector Median Filters*, Proceedings of IEEE, vol. 78, No. 4, April.

Bartolini, F., Cappellini, V., Colombo, C., Mecocci, A., 1993, *Enhancement of Local Optic Flow Techniques*, Proceedings 4th International Workshop on Time Varying Image Processing and Moving Object Recognition, June 10-11, Florence, Italy, pp. 359-366.

Bartolini, F., Cappellini, V., Giani, 1996, C., *Motion Estimation and Tracking for Urban Traffic Monitoring*, Proceedings 3rd IEEE International Conference on Image Processing ICIP'96,Lausanne, Switzerland, September 16-19, pp. 787-790.

Baglietto, P., Maresca, M., Migliardi, M., Zingirian, N., 1997, *Customizing MPEG Video Compression Algorithm to Specific Application Domains: the Case of Highway Monitoring*, Lectures Notes in Computer Science, Vol. 1311, pp. 46-53: Springer Verlag.

De Vos and Stegherr, M., 1989, *Parametrizable VLSI Architectures For Full-search Block-matching Algorithms*, IEEE Transactions on Circuits and Systems, vol. 36, October, pp. 1309-1316.

Haralick, R.M., .Shapiro, L.G., 1992, *Computer and Robot Vision, Vol.1,* pp.28-33:Addison-Wesley.

ITU-T, 1995, *sl Video Codec Test Model*, TMN5, ITU-TSS LBC-95, Study Group 15, Working Party 15/1, Expert's Group on Very Low Bitrate Visual Telephony, Telenor Research, January, http://web.fou.telenor.no/fou/DVC/tmn5/tmn5.html

Koller, D., Weber, J., Huang, T., Malik, J., Ogasawara, G., Rao, B., Russel, S., 1994, *Towards Robust Automatic Traffic Scene Analysis in Real-Time*, Proc. Int'l Conf. Pattern Recognition, pp. 126-131.

Michalopulos, P. G., 1991, *Vehicle Detection Video Through Image Processing: The Autoscope System*, IEEE Transactions on vehicular technology, Vol. 40, No. 1,February, pp. 21-29.

Musmann, H.G., Pirsch, P., Grallert, H.J., 1994, *Advances in Picture Coding*, Proceedings of the IEEE, Vol. 73, No. 4, April, pp. 523-546.

Shalkoff, R. J, 1989, *Digital Image Processing and Computer Vision*, pp.137-144: Wiley.