Vehicle Detection and Tracking using Wavelet Transforms

K. SUBRAMANIAM, S.S. DLAY and F.C. RIND
Department of Electrical and Electronic Engineering
University of Newcastle-Upon-Tyne
Merz Court, Newcastle-Upon-Tyne NE1 7RU
UNITED KINGDOM

Abstract: - This paper presents, an algorithm using wavelets to detect objects from motion. The algorithm is designed to be part of a Vehicle Tracking System. The main aim is to integrate information from Gabor and Mallat wavelet transforms to improve the accuracy and speed of vehicle detection. The Gabor-wavelet analysis allows a rapid estimation of image flow vectors with low spatial resolution. Furthermore, it enables us to produce a histogram over the image flow field, its local maxima providing motion hypotheses. These then serve to increase the accuracy of the subsequent object detection using the Mallat-wavelet transform, which provides the required high resolution. The reliability of the motion detection is improved by integration over time. Since no object model is used, the system can detect even small, disconnected, and openworked objects of arbitrary numbers, such as dot patterns. The algorithm behaves well even in the complex motion situation.
Key-Words: - Mallat-Wavelet Transform, Gabor-Wavelet, Motion Tracking, Motion Hypotheses.
CSCC'99 Proceedings: - Pages 5011-5016

1. Introduction

The detection algorithm presented here has been motivated in context of a larger effort to build a Vehicle Tracking system. It is an attempt to develop a motion-tracking algorithm, which improves the accuracy and the speed of the system. This is the main reason why Gabor and Mallat wavelet transform are used here. Object detection from motion algorithms can be divided into two broad classes, the filter based methods and the matching methods. A typical example of a filter-based method is the gradient method, where a motion constraint analysis [4] is used to estimate image flow. As this method estimates motion locally in a small region of the image, it suffers from the aperture problem. The matching techniques on the other hand usually suffer from the correspondence problem; i.e. there is an ambiguity as to which feature point in one frame has to be matched to which feature point in the next frame. In this work we integrate these two types of techniques, filters and matching, in order to overcome the aperture problem and reduce the correspondence problem significantly. This achieved by the adaptation of an algorithm developed by [1], to tracking vehicle motion in depth.

2. The System

2.1 mage Flow Estimation

The first stage of the algorithm is the computation of an image flow field. The method used here is based on a convolution with a *Gabor Wavelets transform* [2], which has the shape of localised, plane waves. This Gabor wavelet transform yields complex coefficients with phases usually varying with the main frequency of the corresponding kernels. The distance d between two points and the phase difference Δf of the coefficients at these two points have an approximate relationship $\Delta f = dk$, where k is the wave vector of the kernel's main frequency. Using several kernels with different orientations can provide an accurate estimate of d based

on Δf. Similarly, phase differences between the coefficients in two successive frames (Fig.1) but at the same location can be used to estimate the translation of the underlying grey value distribution at that location. Doing this at each pixel provides an image flow field [5].

The Gabor wavelet transform for an image $I(x)$ is defined as a convolution

$$J_j(x) = \int I(x')y_j(x-x')d^2x' \qquad (1)$$

with a family of Gabor wavelets

$$y_j(x) = \frac{k_j^2}{s^2}\exp\left(-\frac{k_j^2 x^2}{2s^2}\right)\exp(ik_j x) \qquad (2)$$

in the shape of plane waves with wave vector k_j, restricted by a Gaussian envelope function. We employ a discrete set of 5 different frequencies, index $u = 0,..,4$ and of 8 orientations, index $m = 0,\ldots,7$,

$$k_j = \begin{pmatrix} k_{jc} \\ k_{jy} \end{pmatrix} = \begin{pmatrix} k_n \cos j_m \\ k_n \sin j_m \end{pmatrix}, k_n = 2^{-\frac{n+2}{2}}p, j_m = m\frac{p}{8} \quad (3)$$

with index $j = m + 8u$. This sampling evenly covers a band in frequency space. The width s/k of the Gaussian envelope is controlled by the parameter $s = 2p$. One speaks of a wavelet transform since the family of kernels is self-similar, all kernels being generated from one *Mother Wavelet* by dilation and rotation.

2.2 Motion Hypotheses

The second stage of the algorithm is concerned with the extraction of motion hypotheses from the image flow field. This can be simply extracted by detecting local maxima in the image flow histogram, a histogram over the flow field vectors. In order to avoid detecting too many irrelevant maxima, a low-pass filter is applied to the histogram and maxima of a certain minimal height are accepted. The result of this stage is usually a small number of displacement vectors $\vec{u}_n$ representing frequently occurring flow vectors. A displacement $\vec{d}$ can be compared with a motion hypothesis $\vec{u}_n$ by the displacement similarity function

$$S_d\left(\vec{u}_n, \vec{d}\right) = \max\left(1 - \frac{\left(\vec{u}_n - \vec{d}\right)^2}{r^2}, 0\right) \qquad (4)$$

with a parameter r to set its sensitivity.

The advantage of the motion hypotheses is that they drastically reduces the correspondence problem for an image flow estimation algorithm based on more localised features, such as edges. Instead of testing all possible displacements within a certain distance range, only a few of them suggested by the motion hypotheses need to be taken into consideration. This greatly improves the accuracy and speeds up the processing. See in Fig.2, Illustration of the reduced correspondence problem: Without motion hypotheses, a whole region within a certain diameter has to be tested for possibly corresponding pixel locations with a similar grey value gradient. With motion hypotheses, the regions to consider reduce to a few small spots.

2.3 Edge Valuation

The third stage of the algorithm uses the *Mallat Wavelet transform* which can be thought of as the grey-value gradient at different levels of resolution. This stage is similar to a matching algorithm and is therefore faced with the correspondence problem, which is particularly severe for such a simple feature as the local gradient. We employ two methods to bypass the correspondence problem. Firstly, the evaluation of the Mallat-wavelet transform is restricted to edges, as defined by the modulus maxima [3]. Edges have a particularly high

information content and are less ambiguous than gradients in general. Secondly, we match the edges between two frames, using the image flow field obtained in the first stage. This reduces the correspondence problem significantly, but leaves it partially unresolved at the same time. It is important from a conceptional point of view that the modulus maxima of the wavelet transform provide sufficient information to reconstruct images.

Given a gradient similarity function $S_g\left(\vec{g}, \vec{g}\right)$ comparing direction and magnitude of two gradients $\vec{g}$, the accordance of a gradient $\vec{g}\left(\vec{x}, t\right)$ with a motion hypothesis $\vec{u}_n$ is defined as the maximum over

$$A\left(\vec{x}, t, \vec{u}_n\right) = \max_{\vec{d}} \left\{ S_d\left(\vec{u}_n, \vec{d}\right) S_g\left(\vec{g}\left(\vec{x}, t\right), \vec{g}\left(\vec{x} - \vec{d}\right) t - 1\right) \right\}, \quad (5)$$

found by varying $\vec{d}$. Notice that only edge pixels are evaluated and taken into consideration. This accordance function defines an accordance map for each motion hypotheses indicating whether an edge might have moved by the respective flow vector $\vec{u}_n$ or not. It is important to note that the accordance maps may still contain ambiguities as to which edge belongs to which motion hypothesis. This would be indicated by high accordance values in more than one accordance map for a single edge pixel.

2.4 Integration over a Sequence of Frames

The quality of the accordance maps can be significantly improved by integrating them over a sequence of frames. The main difficulty is to determine which motion hypothesis in one pair of frames corresponds to which motion hypothesis in the next pair of frames. In order to avoid assumptions about the type of movement objects may perform, We associate the motion hypothesis via the spatial overlap in the intermediate frame, i.e. the one that is common to both pairs. This is possible because the first pair of frames already provides a crude segmentation result. See in Fig.3, Accordance maps for Frame 15 with respect to the two relevant motion hypotheses. This accordance map was integrated over time and takes into account the results preceding frames. The integrated accordance map is the geometric mean of the previous integrated accordance map and the current instantaneous accordance map.

2.5 Segmentation

The last stage of the algorithm performs the segmentation. Each edge pixel is classified as belonging to that motion hypothesis for which it has the highest accordance value. One can think of as a pixel-wise winner-take-all competition between the accordance maps. See in Fig.4, Motion Detection based on the accordance maps.

3. Examples

The strength of the system is that it generates motion hypotheses on a coarse level, but segments on a single pixel level. This allows the system to segment even small, disconnected, or openworked objects. A second strength is that the motions need not be continuous in order to do the integration over a sequence of frames. Objects may jump back and forth, and the algorithm will still be able to track them. This is demonstrated for dot-patterns in Fig.5, fifth frame of the dot-pattern sequence, Showing a stationary circle with eight points and a dotted background moving left. Each frame in the sequence is additionally shifted randomly up to (±2, ±2) pixels, resulting in an additional relative displacement of up to ±4 pixels in each dimension where random shifts of the frames against each other have been introduced artificially.

When tracking several vehicles on the road, the ability of the algorithm to track each vehicle separately is lost if the vehicles travel closely to each other with nearly the same

speed. Altough the algorithm respresents these two vehicles as one object (motion), the segmentation result shows these object as two vehicles. This is shown in Fig.6, three moving vehicles on the road taken from stationary camera. Two vehicle moving towards left closely with nearly same speed and one vehicle just coming into view which is moving toward right.

4. Conclusion

The system presented performs segmentation from motion on a sequence of frames. It is unconventional in several respects. It integrates two different techniques based on Gabor- and Mallat-wavelets to improve the accuracy of detection and it integrates over time for an additional improvement of the segmentation results. Segmentation is only performed on edges and it is argued that edges are the most appropriate representations for segmentation, because they provide enough motion evidence and the complete grey-value distribution could be reconstructed from edges.

Since no object model is used, the system can even segment small and openworked objects. We propose to do the segmentation locally and integrate the local fragments on a higher level. Due to the closeness and nearly same velocity of the motion objects, the two different objects can be attached together as one motion object. This attachment which can be improved by regularisation constraints and explicit object knowledge, while keeping local segmentation decisions.

References

[1]Wiskott L, 'Segmentation from Motion. Object recognition', Tech. Rep. IR-INI 96-10, Institute fur Neuroinformatik, Ruhr-Universitat Bochum, D-44780 Bochum, Germany, 1996.

[2]Daugman J.G, 'Complete discrete 2-D Gabor transform by neural networks for image analysis and compression', IEEE Transactions on Acoustics, Speech and Signal Processing, 36-7, July 1988, 1169-1179.

[3]Mallat S and Zhong S, 'Characterisation of signal from multiscale edges', IEEE Transactions on Pattern Analysis and Machine Intelligence, 14-7, 1992, 710-732.

[4]Horn B.K.P and Schunck B.G, 'Determining optical flow', Artificial Intelligence, 17, 1981, 185-203.

[5]Theimer W.M and Mallot H.A, 'Phase-based binocular vergence control and depth recontruction using active vision', CVGIP: Image Understanding , 60- 3, Nov. 1994, 343-358.

(a) (b)

Fig.1: (a) Frame 14 , (b) Frame 15 is the car moving sequence on the road taken from moving camera. The images have a resolution of 128x128 pixels with 256 grey values.

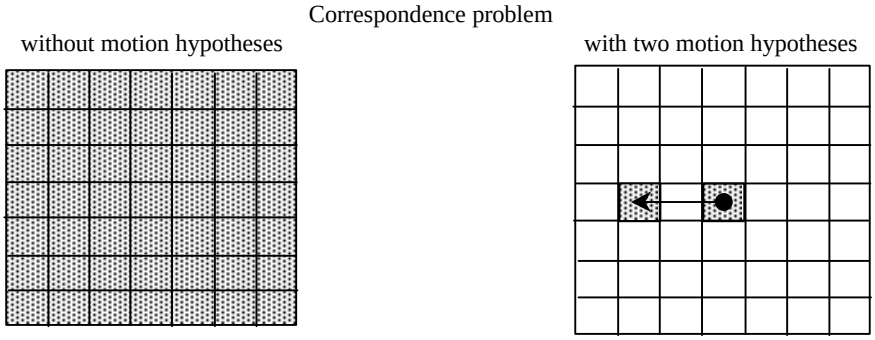
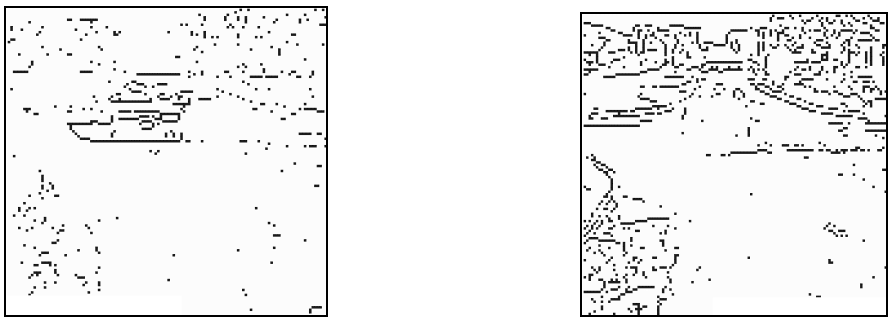
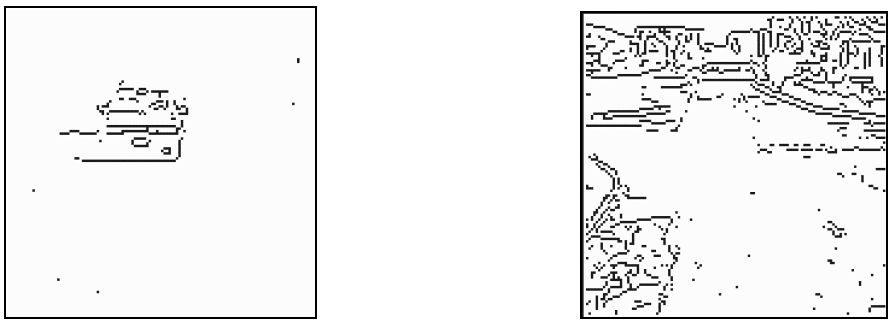


Fig.2: Illustration of the reduced correspondence problem



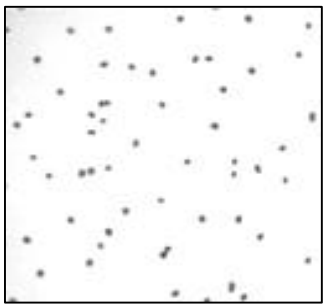
(a) (b)

Fig.3: Accordance maps for Frame 15 with respect to the two relevant motion hypotheses. (a) corresponds to the car, (b) corresponds to the background.

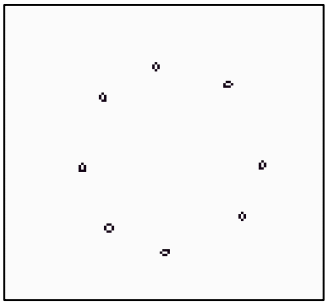


(a) (b)

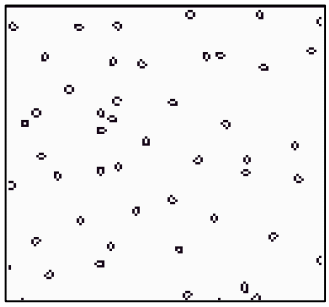
Fig.4: (a) Motion Detection based on the accordance maps. with (b) Segmentation of the background



(a)



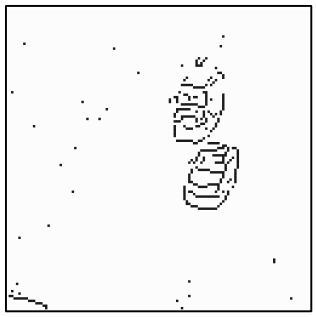
(b)



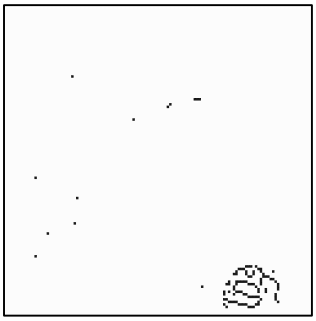
(c)

Fig.5: a) Fifth frame of the dot-pattern sequence. b) Segmentation result for the circle and c) the background.

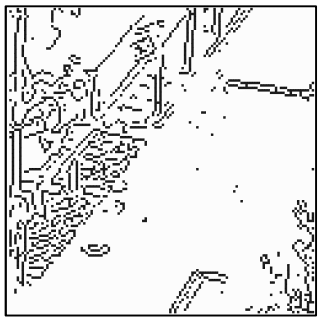
(a)



(b)



(c)



(d)

Fig.6: a) 20th frame of three moving vehicles on the road. b) Detected two vehicle which moving closely as first object. c) second object is the one vehicle moving right. d) is the background.