Towards a Human Tracking System for a Mobile Robot Using Neural-Based Motion Detectors

[1]John A. Perrone, [2]Tony Voyle, [2]Margaret E. Jefferies
[1]Department of Psychology
University of Waikato
[2]Department of Computer Science
University of Waikato
jpnz@waikato.ac.nz, tv6@cs.waikato.ac.nz, mjeff@cs.waikato.ac.nz

Abstract

In this paper we outline a human tracking system for an autonomous mobile robot. Unlike the majority of human tracking methods to date our approach does not use some aspect of the appearance of the human form to recognise the human to be tracked. Rather we use the motion in the scene which is characteristic of human motion to recognise and track people moving in a robot's field of view. Our method also differs from most other object tracking approaches in that we use motion sensors based on the motion sensitive neurones found in the medial temporal area of the primate brain [1]. Networks of these sensors have been developed for detecting and eliminating the background motion generated by the moving robot. Similar networks of motion sensors, comprising a sensor tuned to each component of a walking person's movement, are being used to detect and track people.

Key Words:

Optical flow, motion sensitive neurones, background motion, self motion, human tracking, autonomous mobile robots

1. Introduction

Identifying and tracking humans is an important problem in robotics. Autonomous mobile robots that operate in environments populated with humans, depending on their tasks, interact with these humans in a variety of ways. A robot building its own map of its environment as it explores [2, 3] needs to identify fleeting objects such as humans and ensure that they do not become part of its map. Fig. 1 shows a map that results when humans are not identified. For a robot that engages with humans, for example a tour guide robot, it helps to know the location of the human the robot is interacting with. Surveillance robots not only need to identify humans but also need to track their motion. However motion tracking is not just the domain of surveillance robots. All robots that share environments with humans need to account for its motion if they are to avoid colliding with them. Robots that realistically engage with humans may want to follow the humans. Often the robot will need to distinguish a human that has just entered its field of view from those it currently has in view.

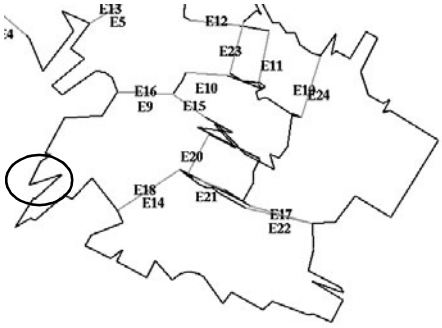


Fig. 1 The human mapping problem. The encircled section of the map encompasses the legs of one of our students.

Fig. 2 illustrates the problem faced by a video-based sensor system as the robot carrying the sensor moves through an environment and a person walks into the scene. The detection of visual motion from image sequences is a difficult problem and has a long history [4, 5]. Recently the detection of human motion has become one of the most active areas of research in computer vision [6, 7]. For an autonomous mobile robot the difficulty is compounded in the person detection task because the motion of the target is heavily masked by the background motion generated by the robot movement. The task is not simply to detect motion, but to identify and separate out target motion from background motion.

Approaches currently employed in computer vision to track human motion can be categorised by whether they firstly detect the humans or the motion in a scene. An example of the latter is the work of Tsukiyama and

Fig. 2 Motion in a scene. The dark arrows show background motion generated from the motion of the camera. The white filled arrows show the motion of the human.

Shirai [8] where all moving objects are first identified; from these the humans are distinguished and are then able to be tracked. By contrast the Pfinder system [9] first creates initial representations for the humans in the scene. The tracking procedure updates these representations. The majority of methods falling into either category recognise some aspect of the "appearance" of the human. Our approach is different in that the human is identified by the characteristics of its motion. Thus the flow of motion in the scene is used to both pinpoint the human and track its movement.

The person identification problem would be greatly simplified if we could somehow subtract out the image motion caused by translation of the robot. Any "anomalous" image motion that deviates from the predicted motion is likely to arise from motion of the person walking in the scene or another moving robot. Systems which estimate self motion, i.e. the robot motion have been developed [10-12].

Our approach has some similarities with each of these systems. In Montemerlo et al.'s [12] system the robot must keep track of its own location in the map it computes of its environment and also track the location of any people in its vicinity. Separate particle filters are used to estimate the locations of the robot and each person, and Brownian motion is used to model the typical motion of a person. Our system is also concerned with keeping track of a human's location within the robot's map. However we are not only concerned with where the robot moves but how the object being tracked moves so we can identify it as "human".

Franz [10, 11] models the motion sensitive tangential neurones in the fly brain to detect self motion. His system, unlike ours, uses 2-D motion sensors based on the gradient method [4]. However it has problems obtaining accurate translation estimates. The difficulties with translation estimates could be overcome if the robot could accurately measure its velocity. In practice this is not possible. Overcoming the inaccuracies that accumulate in robot mapping due to wheel slippage has been one of the great challenges in robotics research for some time.

Our system uses self-motion estimation templates based on neurones in the medial superior temporal (MST) area of the primate brain [13, 14]. The Perrone and Stone model is able to extract the relative depth of points in the environment from the 2-D image motion generated during forward translation of an observer. Over a succession of frames, any motion caused by objects not fixed in the environment (e.g. walking people) will stand out in the "3-D space" output of the model's depth detecting system. Until recently, the model has only been able to be tested with theoretical vector flow field inputs. However we have now developed a 2-D motion sensor (see section 2) that can be used as a front-end to the model and which enables the model to be applied to image sequences.

2. 2-D Motion Sensors Based on Biological Principles

We have developed motion sensors based on the properties of motion sensitive neurones in the primate brain [1]. These are more selective to the speed of the movement than standard approaches to image motion measurement [4] and they do not suffer from the correspondence problem associated with feature tracking methods. The sensors are built up from two specially designed spatiotemporal filters (S and T). The S type has low-pass temporal frequency tuning and the T type has band-pass tuning. The spatial frequency tuning of the two types also differs slightly in a specific way. In the spatiotemporal frequency domain (u, ω), these two filters overlap along a line that is oriented in (u, ω) space. The outputs of the two filters are combined using the following equation:

$$WIM(u, w) = \frac{\log(S + T + \alpha)}{|\log T - \log S| + \delta}$$

α and δ are constants which fine tune the properties of the sensor. The sensor is tuned to a particular speed *v*. The resulting sensor has been called the Weighted Intersection Mechanism (WIM) model [1], because it maximises the response of the sensor along the line of intersection of the two (S and T) filters in spatiotemporal frequency space. This line happens to correspond to the location of the spectrum of an edge moving at a particular speed *v* [15]. The sensor is therefore very speed selective and is better at discriminating different edge speeds compared to other systems based just on spatiotemporal filtering (e.g., motion energy models). A number of these basic WIM sensors can be combined to produce an overall 2-D motion sensor that is very se-

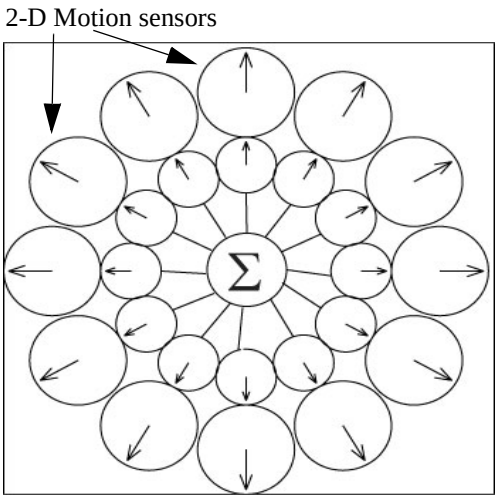


Fig. 3 A heading template showing 2-D motion sensors arranged in a radial pattern.

lective for a particular velocity (speed and direction) of image motion. We have shown that these sensors have properties very similar to motion sensitive neurones in the middle temporal (MT) extra-striate area of primates [1].

3. Recovering 3D layout from optic flow

We have also developed networks of the above sensors for detecting the image motion generated during movement of a camera platform through the environment [13, 14]. An example of one of these self-motion estimating 'templates' (matched filters) is represented in Fig. 3. This template is designed to detect the radial patterns of image motion that occur during forward translation through a scene. A number of radially aligned 2-D motion sensors are coupled together and their output is summed to generate a heading signal in the detector. A non-linear stage is included which selects the maximum output (winner-takes-all) from a number of 2-D sensors at any image location and sends that output to the heading template. By incorporating a number of heading templates tuned to different heading directions and selecting the most active one, it is possible to estimate the heading direction of the moving observer (or camera platform). There is evidence that many animals (including humans) use this type of neural mechanism for determining their heading direction [16]. It is also possible to use similar types of 'full-field' motion detecting networks for determining other aspects of a person's (or robot's) movement through the environment (e.g., eye, head or camera rotation. See Perrone, 1992). In addition, once the heading direction is determined, it is possible to use the activity in the 2-D sensors connected to the winning template to estimate the relative depth of objects in the scene, i.e., to obtain a depth map (see Perrone & Stone, 1994).

We intend using these heading detector networks for registering and removing the image motion caused by robot movement so that the motion of a person walking into the path of the robot can be more easily detected.

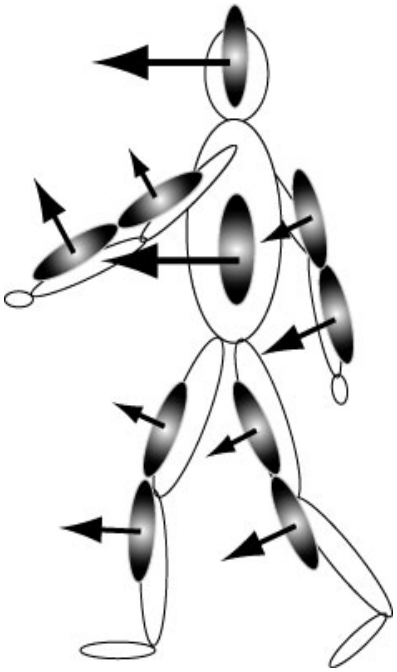


Fig. 4 Representation of a walking person with motion sensors arranged to selectively pick out the walk-movement.

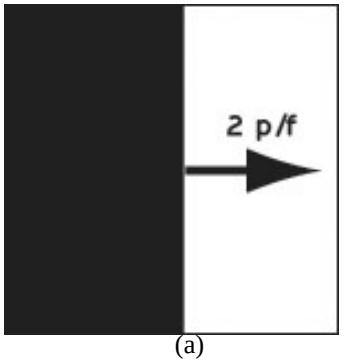
4. Detecting human movement in a scene

Having developed the computational framework for people detection and tracking by an autonomous mobile robot we are currently focusing on the problem of detecting a "human" moving within the robot's field of view. The robot's motion sensing apparatus comprises a network of sensors which collectively detect the human motion. The individual sensors are tuned to a particular aspect of human motion, e.g. one sensor could be tuned to the characteristic motion of the upper arm when a person is walking. The proposed scheme is shown in Fig. 4.

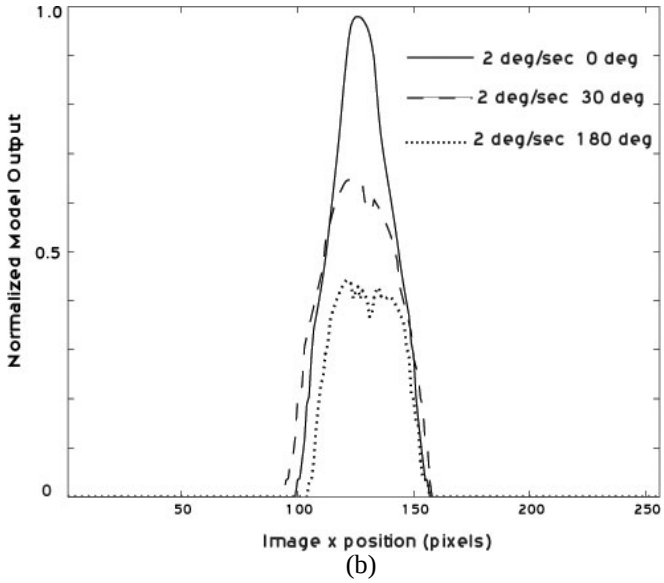
Fig. 5 shows an example of tests we have carried out on the basic 2-D motion sensors. The sensor in this example would detect the moving edge of a human body part. It responds selectivity to one particular speed and direction as shown in Fig. 5 (b) and (c). By using collections of such sensors we can cover the appropriate tunings needed for each moving body part (see Fig. 4). The outputs of the different sensors are summed. Collectively the sensors can be viewed as a template which responds selectively to walking motion.

5. Conclusion

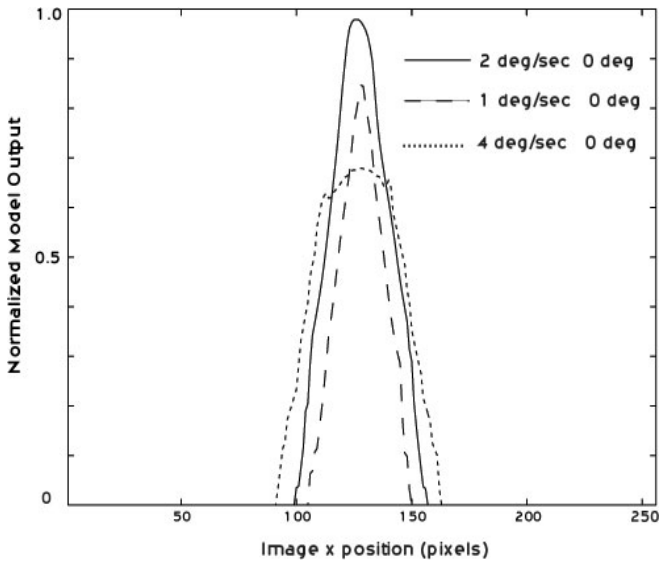
We have outlined a system for a human tracking system for an autonomous mobile robot. Our system is different from most human tracking systems in that it is a



(a)



(b)



(c)

Fig. 5 Test of the basic motion sensor tuned to 2 p/f (pixels per frame) and 0^o direction. (a) one frame of an 8 frame sequence of an edge moving at 2 p/f in a 0^o direction. (b) Outputs of multiple sensors across the image for edges moving at 2 p/f at 30^o and 180^o directions. (c) Outputs for edges moving at 1, 2, and 4 p/f. Maximum response occurs for the edge velocity that matches the sensor tuning.

one step process. The motion detected identifies the human. Most current systems require a separate step to identify the human and then find its motion.

One aspect that we have not addressed yet is the need for different sized templates for different scales, for example, near and far objects. Also to be considered is at what point the robot would be interested in tracking the human object.

Computational complexity is an issue for any motion detection method. our method has the advantage that the sensors can be precomputed. The tradeoff is that they require considerable memory resources but this is a manageable constraint. We are considering an hierarchical approach to reduce the amount of processing.

References

1. Perrone, J.A. and Thiele, A., "A model of speed tuning in MT neurons". *Vision Research*, 42(8): 1035-1051 (2002).
2. Jefferies, M.E., Baker, J., and Weng, W. "Robot cognitive mapping: A role for a global metric map in a cognitive mapping process". in *Workshop on Robot and Cognitive Approaches to Spatial Mapping*, (2003).
3. Yeap, W.K. and Jefferies, M.E., "Computing a representation of the local environment". *Artificial Intelligence*, 107: 265-301 (1999).
4. Horn, B. and Schunck, B., "Determining optical flow". *Artificial Intelligence*, 17: 185-203 (1981).
5. Tomasi, C. and Kanade, T., "Shape and motion from image streams under orthography". *International Journal of Computer Vision*, 9(2): 137-154 (1992).
6. Moeslund, T.B. and Granum, E., "A survey of computer vision-based human motion capture". *Computer Vision and Image Understanding*, 81: 231-268 (2001).
7. Gavrila, D.M., "The visual analysis of human movement". *Computer Vision and Understanding*, 73(1): 82-98 (1999).
8. Tsukiyama, T. and Shirai, Y., "Detection of the movements of persons from a sparse sequence of TV images". *Pattern Recognition*, 18: 207-213 (1985).
9. Wren, C.R., et al., "Pfinder: real-time tracking of the human body". *Transactions on Pattern Analysis and Machine Intelligence*, 19(7): 780-785 (1997).
10. Franz, M.O. and Chahl, J.S., "Linear combination of optic flow vectors for estimating self motion - a real world test of a neural model", in *Advances in Neural Information Processing* (2003).
11. Franz, M.O. and Krapp, H.G., "Wide-field motion-sensitive neurons and matched filters for optic flow fields". *Biological Cybernetics*, 83: 185-197 (2000).
12. Montemerlo, M., Thrun, S., and Whittaker, W. "Conditional particle filters for simultaneous mobile robot localization and people tracking". in *IEEE International Conference on Robotics and Automation*, (2002).
13. Perrone, J.A., "Model for the computation of self-motion in biological systems". *Journal of Optical Society of America A*, 9(2): 177-194 (1992).
14. Perrone, J.A. and Stone, L., S., "A model of self-motion estimation within primate extrastriate visual cortex". *Vision Research*, 34: 2917-2938 (1994).
15. Watson, A.B. and Ahumada, A.J., "A look at motion in the frequency domain", in *Motion: Perception and representation*, J.K. Tsotsos, Editor. New York: Association for Computing Machinery. 1-10 (1983).
16. Duffy, C.J. and Wurtz, R.H., "Sensitivity of MST Neurons to Optic Flow Stimuli. I. A continuum of response selectivity to large-field stimuli". *Journal of Neurophysiology*, 65: 1329-1345 (1991).