Neural Comput & Applic (1997)6:142–147

Neural
Computing
& Applications

Focus of Expansion Estimation by an Error Backpropagation Neural Network

A. Branca, E. Stella, G. Attolico and A. Distante

Istituto Elaborazione Segnali ed Immagini, C.N.R., Bari, Italy

In this work we consider the application context of planar passive navigation in which the visual control of locomotion requires only the direction of translation, and not the full set of motion parameters. If the temporally changing optic array is represented as a vector field of optical velocities, the vectors form a radial pattern emanating from a centre point, called the Focus of Expansion (FOE), representing the heading direction. The FOE position is independent of the distances of world surfaces, and does not require assumptions about surface shape and smoothness. We investigate the performance of an artificial neural network for the computation of the image position of the FOE of an Optical Flow (OF) field induced by an observer translation relative to a static environment. The network is characterized by a feed-forward architecture, and is trained by a standard supervised back-propagation algorithm which receives as input the pattern of points where the lines generated by 2D vectors are projected using the Hough transform. We present results obtained on a test set of synthetic noisy optical flows and on optical flows computed from real image sequences.

Keywords: Focus of expansion; Hough parameter plane; Neural Networks; Optical flow

1. Introduction

In many dynamical applications, like autonomous robot navigation, useful visual information can be obtained from a camera mounted on the mobile vehicle. The vehicle motion in the environment determines spatial and temporal changes in the viewed image, producing a 2D motion field which can be used for recovering both the vehicle motion and the three-dimensional structure of the scene. In this paper, we focus our attention on the information contained in the changing structure of light at an eye due to the movement of an organism relative to a rigid environment, known as Optical Flow (OF). The motion field is a relatively high-level description. Accurate estimates of the motion field from a time-varying sequence of images are usually difficult to obtain [1], since they depend on *a priori* knowledge of many physical properties of the viewed scene. In fact, the use of local, differential techniques, permits to recover a 2D vector field that is only similar to the exact one. Gibson [2] was the first to point out the invariant radial pattern of optical flow emerging from translational motion of an observer in a rigid environment. In fact, the vectors form a radial pattern emanating from a centre point, called the Focus of Expansion (FOE). This global radial pattern is a natural consequence of optical perspective, and it is invariant respect to the 3D structure of the environment. Gibson proposed that this pattern is detected by the observer and used to control locomotion. All points on the image plane seem to move along straight lines whose common intersection is a singular point (where the motion field vanishes), that is a focus at any time. It remains fixed and does not change its qualitative structure. The location of the focus of expansion on the image plane indicates the direction of translation (i.e. the heading of the TV camera).

The problem to be solved is to find the movement parameters of the observer when optic flow at a few points is given. The problem is usually posed as one involving the derivation of the six motion

Correspondence and offprint requests to: A. Branca, Istituto Elaborazione Segnali ed Immagini, C.N.R., Bari, Italy.

parameters characterizing general motion of a rigid body in 3D. Assuming a perspective projection model in which a world point $P = (X,Y,Z)$ projects on the image point $(x,y) = f\left(\frac{x}{z}, \frac{y}{z}\right)$, where f is the focal length, Longuet-Higgins and Prazdny [3] derived the following equations to describe the general rigid motion of an observer moving in a stationary world:

$$u = \frac{T_x - xT_z}{Z(x,y)} - xyR_x + (1 + x^2)R_y - yR_z \tag{1}$$

$$v = \frac{T_y - yT_z}{Z(x,y)} - (1 + y^2)R_x + xyR_y + xR_z \tag{2}$$

with (T_x,T_y,T_z) the 3D translational velocity components, (R_x,R_y,R_z) the 3D rotational velocity components, (u,v) the projected velocity $V(x,y) = (u(x,y), v(x,y))$ of a point (X,Y,Z) on the image plane, and $Z(x,y)$ the depth function.

Though various algebraic approaches have been proposed for the solution of non-linear systems resulting from writing Eqs (1) and (2) in a suitable number of image points, the results obtained are numerically instable due to the large number of equations to be solved and the noise in the (u,v) velocity vector estimates.

In accordance with Verri and Poggio [1] and Warren et al. [4], we believe that a more adequate approach to the problem of egomotion perception from optical flow is to restrict the goal to the only computation of FOE associated with the translational component of motion. In fact, in a context like planar passive navigation in which the camera moves forward and the rotational component of motion can be considered negligible, the visual control of locomotion may only require the direction of translation, and not the full set of motion parameters. Moreover, the radial flow pattern is a purely qualitative property of the optical array; it is independent of the distances of world surfaces, and it does not require assumptions about the shape and smoothness of the 3D surface. Several methods in computational vision have been exploited for the computation of FOE from optical flow vectors, the most trivial being the intersection of the set of lines that they determine.

Actually, even a small number of observer rotations or accidental camera vibrations, which are likely to occur in real contexts, can change the radial shape of the flow field and the FOE position. To manage this situation, several methods have been proposed [5–10], most of which use calibrated systems or require a preliminary decomposition of the optical flow field into its translational and rotational components, or propose to compute only a fuzzy region containing the FOE.

In our paper, considering that the FOE is the intersection of lines determined by 2D velocity vector orientations, by using the Hough transform [11], all these lines are projected onto points in the *parameter plane*. Theoretically, those points must lie on a straight line whose slope and intercept represents the FOE coordinates in the image plane. So, once the 2D velocity vectors are projected in the *parameter plane*, the slope and intercept of the line can be estimated by linear regression. Normally, an approximation method based on the LSE (least square error) technique is sensitive to intrinsic noise associated to the optic flow data. Our aim is to investigate the possibility that a distributed neural paradigm may robustly filter out the noise intrinsic to image motion measurements.

We will show how correct heading direction can be computed by a feed-forward neural network trained through a standard supervised backpropagation algorithm.

First, a sparse optical flow field is computed from a pair of successive TV images by matching a few distinct points (extracted by Moravec's interest operator [12]) using the Hopfield neural network [13]. Successively, the FOE location is computed by the proposed neural network.

The input of the network is a binary matrix obtained as a quantisation of a parameter plane where the 2D velocity vectors are coded using the Hough transform. This particular input representation takes into account only the vector orientation, and makes the input pattern independent of the image size and of the number of sparse vectors. The FOE coordinates are represented by the output node values.

The network has been trained to recognise the FOE location ranging in a window of 75×75 pixels centred in the image plane. In the training step, only optical flows relative to some possible FOE locations have been presented. The ability of the network to generalize for new FOE locations and the tolerance to noise makes it potentially useful in robotic vision applications.

In Section 2 the use of mapping onto the parameter plane for optical flow interpretation aims is described. The architecture and dynamics of a neural network are described in Section 3. Finally, in Section 4, all the results obtained both on various test sets adding uniform noise and on optical flow computed from real image sequences are presented.

2. FOE Computation Using the Hough Parameter Plane

Our goal is to recover FOE location when a radial sparse optical flow is given. In the case of a dense map, there is a close correspondence between a FOE location and the associated OF. However, when sparse OFs are used, at each FOE location, several different sparse maps can be associated, depending on locations in the image plane of the features considered for OF estimation. Therefore, different radial patterns (consisting of 2D velocity vectors in different image locations) can be associated at the same FOE location. Patterns independent of vector locations should be extracted from an OF in order to train a network for determining the FOE location.

The optical flow of a planar surface is usually well approximated, over quite large areas of the image plane, by a linear vector field [14]. When an observer translates into a static environment, all those points projected on the image plane move along straight lines determined by the directions of the 2D velocity vectors. The FOE location can be determined as the intersection of all those lines. Since the same line can be determined by vectors placed at different locations in different sparse maps, our idea is to train the network only on straight line orientations, rather than on location and orientation in the image plane of the OF vectors.

In fact, using the Hough transform [11], each sparse vector at point (x,y) in the image plane is represented by a line $b = y - ax$ in the parameter plane, and all vectors having the same orientation of that line in the image plane accumulate on point (a,b) onto a parameter plane, where a is the slope and b the intercept of the line.

From this relationship, each vector of a radial OF (which can be considered as a pencil of straight lines) maps onto points (a,b) in parameter plane, lying on a same straight line (Fig. 1). The slope and intercept of that line is the FOE location referred to by the image plane.

These points (a,b) depend only upon straight lines intersecting in the FOE, and are independent of the number of available sparse vectors. Moreover, when the FOE location changes in the image plane, the OF vectors map onto a different straight line in the parameter plane (Fig. 1).

When real sparse OF maps are considered, because of the noise on the vectors, the FOE location referring to the parameter plane becomes a fit problem. In fact, the intercept and the slope of the line can be evaluated by a linear regression in the parameter plane (Fig. 2). However, a LSE (least square error) approach can be sensitive to an incorrect OF vector that can be found in a sparse OF map, while a neural network may robustly filter out the noise. For each OF to be analysed, the network input is represented by a binary matrix of size $L_a \times L_b$. Each matrix element (i,j) represents a point in the parameter plane relative to the slope and intercept (a_i,b_j). Initially, these cells are set to zero. Then for every point (x_k,y_k) in the image plane, the element (i,j) in the matrix, corresponding to the slope and the intercept of the OF vector at that point, is incremented by 1.

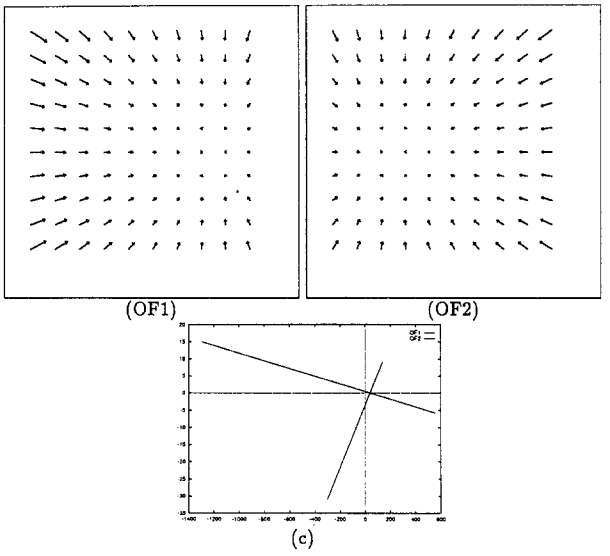

Fig. 1. Different radial OF fields map onto different straight lines in parameter planes.

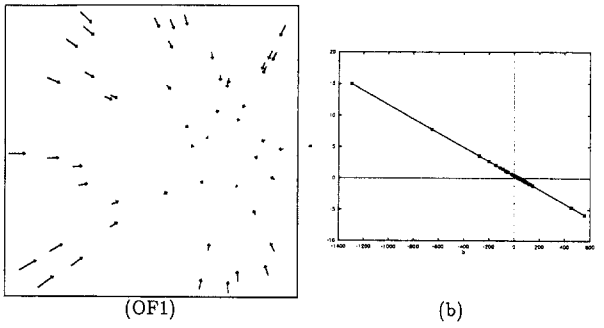

Fig. 2. Vectors in a sparse OF field map onto points in the parameter plane that can be fitted by a straight line.

3. The Neural Network Architecture and Dynamic

The neural network we propose is characterised by a feed-forward architecture consisting of three layers: one input, one hidden and one output layer. A set of input nodes representing the cells in the

parameter plane feed signals through a hidden layer to a layer of output cells representing the direction of heading (*forward step*). The signals are fed between a layer trough fully connected matrix of weighted connections.

The network is trained on a set of parameter plane matrices to map a set of sparse and radial OF fields to their corresponding FOE coordinates. The nodes in the output layer represent the FOE position on the image plane. The net input of each neuron in the hidden and output layers is computed as the weighted sum of the activity of the neurons to which it is connected:

$$inp_i = \sum_{j=1}^{n} act_j w_{ij} + w_0 \tag{3}$$

where n is the number of units to which the ith node is connected, w_{ij} are the connection weights, and w_0 is the bias weight.

The activity of neurons is obtained through the sigmoidal transfer function:

$$act_i = \frac{1}{1 + e^{-\beta inp_i}} \tag{4}$$

The weights of the connections are variable, and are trained through the supervised backpropagation algorithm [15]:

$$\Delta w_{ij}(t + 1) = \eta\, e_j\, act_i + \alpha \Delta w_{ij}(t) \tag{5}$$

The weight change $\Delta w_{ij}(t + 1)$ at instant $t + 1$ is proportional to the error e_j computed for the jth unit, to the activity act_i and to the weight change at the previous step t. The momentum term α is used to help 'smooth out' the weight changes. The error e_j in the output layer is computed as the difference between the desired and actual output. This error, transformed by the derivative of the transfer function, is 'backpropagated' to prior layers where it is accumulated. The backpropagated and transformed error becomes the error term for that prior layer. The backpropagation process continues until the first layer is reached (*backward step*). These two steps (*forward* and *backward*) are iterated until a minimum error is reached.

The following error function between the actual and the desired output has been used to measure the learning performance, and to stop the training when a suitable convergence threshold is reached:

$$AVGERR_{FOE} = \frac{1}{tot} \sum_{j=1}^{tot} \left[\frac{1}{2} \left(\|tFOE_x^j - oFOE_x^j\| + \|tFOE_y^i - oFOE_y^j\| \right) \right] \tag{6}$$

where *tot* is the total number of examples, *tFOE* is the true FOE, and *oFOE* is the output FOE.

4. Experimental Results

The proposed neural network has been trained on theoretical flow fields and tested both on theoretical and on real optical flow maps. A theoretical pattern consists of flow vectors relative to a random dot plane, with the FOE in a randomly chosen image point ranging into a window of 75×75 pixels centred on the image plane. The real optical flow has been computed with the token-based approach proposed in Branca *et al.* [13].

Various experiments have been performed with the aim of finding the optimal values for the number of units in the first and second layers, and for the parameters α, η, μ and β.

The network has been trained on a theoretical OF of size 100×100, with 50 sparse vectors, mapped into a binary matrix of 20×20 elements, making necessary 400 units in the input layer. In the hidden layer, 30 units have been sufficient. The output layer consists of two neurons to code the FOE coordinates in the image plane.

The network parameters such as α, η, μ and β have been set empirically as follows:

$$\alpha = 0{\cdot}01 \tag{7}$$

$$\eta = 0{\cdot}9 \tag{8}$$

$$\beta_1 = 0{\cdot}05 \tag{9}$$

$$\beta_2 = 0{\cdot}5 \tag{10}$$

where β_1 has been considered in the first layer of weights, and β_2 in the second.

In the testing phase, the performance reached by the network has been measured by the accuracy in estimating the FOE position on new optical flow patterns, using the same error function used in training step 6. The number of sample images used in the training phase and in each test on theoretical maps has been set to 4500 (i.e. 10 OF examples for each of the 450 FOE positions randomly chosen).

4.1. Experimental Results for Theoretical Flow Fields

After the class of experiments performed on theoretical flow fields, the network performance with respect to the tolerance to noise and the ability to generalise to new FOE locations and to new vector distributions have been tested.

In the training phase, the network reached an

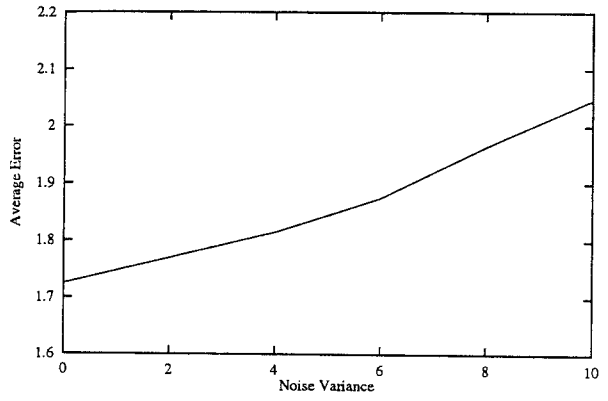

Fig. 3. Results of tests on noise tolerance, on a sample of 100 × 100 OF with 50 sparse points.

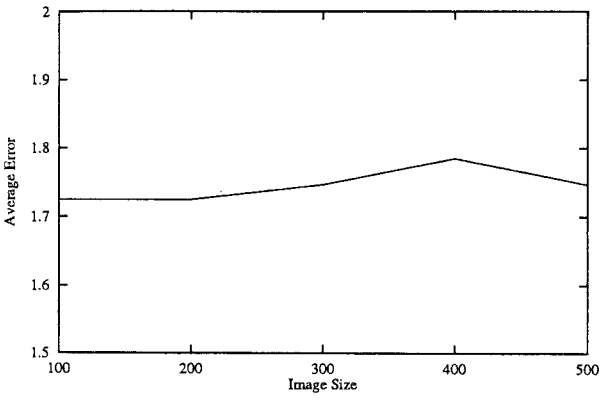

Fig. 5. Results of tests on the independence of image size, on a sample OF with 50 sparse vectors perturbed with uniform noise, with mean 0 and variance 6.

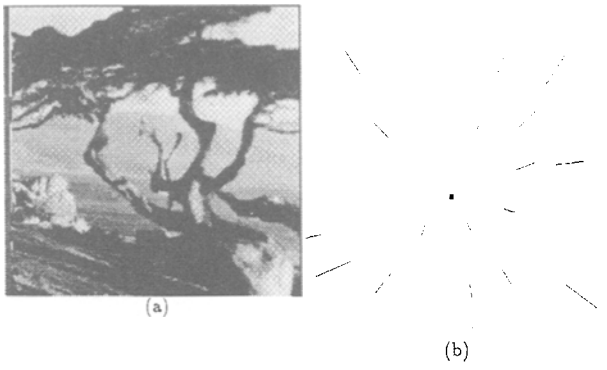

Fig. 6. Diverging tree sequence created by Fleet. The camera translates along its line of sight and the FOE is at the centre of the image plane, whose size is 150 × 150. The FOE is estimated at (74,77).

average error of 0·75 pixels after 730 interations. In Fig. 3, the results of the experiments performed on an OF of size 100 × 100 pixels with 50 sparse 2D vector points perturbed with uniform noise, with a mean 0 and specific variances, are reported. We observe how the average error increases more slowly and is always less than two pixels.

In Fig. 4, the relationship between the number of random dots and the test performances is shown. The OFs used are of a size of 100 × 100 pixels, and the vectors are perturbed with uniform noise with a mean 0 and variance 6·0. Since the network has been trained on an OF of 50 sparse vectors, the average error increases for a lower number to 4·5 pixels.

In Fig. 5, the independence of image size is shown by plotting the average error on an OF with 50 sparse vectors perturbed with uniform noise, with a mean 0 and variance 6·0: the error is constant.

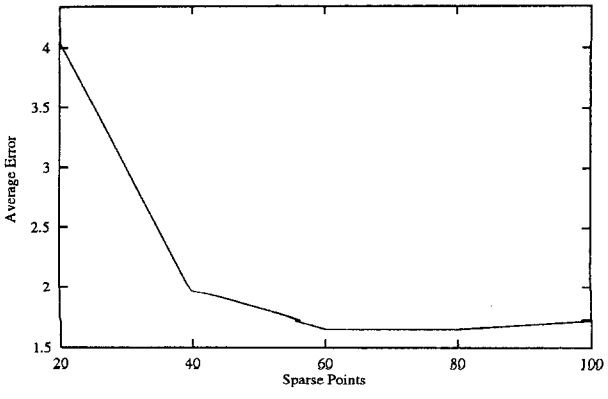

Fig. 4. Results of tests on the independence of the sparse point number, on a sample of a 100 × 100 OF perturbed with uniform noise, with mean 0 and variance 6.

4.2. Experimental Results for a Realistic Optical Flow Field

A second class of experiments has been performed to test the performance of the network in real computational contexts, in which the input data are intrinsically noisy.

In Figs 6–9, results obtained on four distinct real

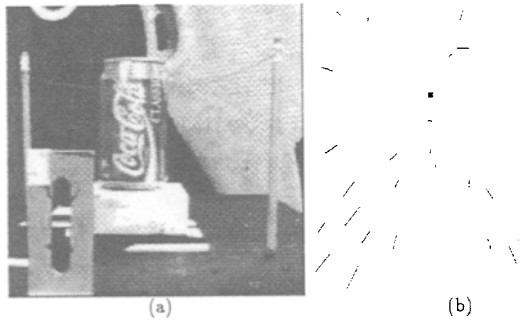

Fig. 7. Sequence collected at NASA Ames Research Center. The scene is static and the initial distance camera scene is of 600 mm. Image size 150 × 150 pixels. The FOE is estimated at (62,47).

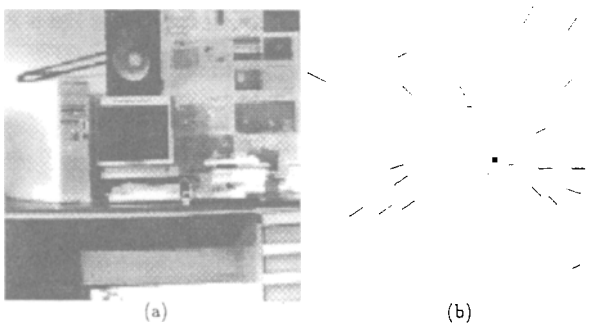

Fig. 8. Image of a laboratory (128 × 128 pixels). The CCD camera is mounted on the mobile platform LABMATE by TRC. The average distance camera-scene is 3000 mm. The FOE is estimated at (81,67).

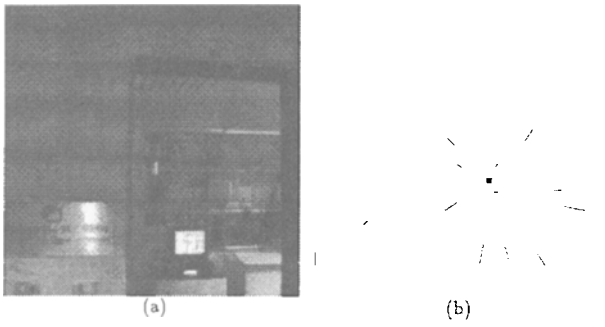

Fig. 9. Image of a laboratory (143 × 143 pixels). The CCD camera is mounted on the mobile platform LABMATE by TRC. The FOE is estimated at (88,86).

image sequences are reported. The recovered FOE is indicated by a black square. We observe that the network computes the correct heading direction, even with a minimum number of sparse vectors (15–20) and noisy vector estimates.

5. Conclusions

We have proposed to map, using the Hough transform, the optical vector field arising in passive navigation in a parameter plane. A supervised feed-forward neural network model has been applied to the computation of the FOE position from radial optical flow fields obtained by an observer translational motion relative to a static scene. Tests on both exact and noisy 2D vector fields obtained as the projection on the image plane of the 3D scene velocities have been performed. The network is able to recover the true FOE position with a mean error of three pixels. Moreover, it appears resistant to noise, and also gives good results when tested on intrinsically noisy optical flow fields computed with a token-based approach.

References

1. Verri A, Poggio T. Motion field and optical flow: qualitative properties. IEEE Trans PAMI 1990; 11(5): 490–498
2. Gibson JJ. The Perception of the Visual World. Houghton Mifflin, Boston, 1950
3. Longuet-Higgins HC, Prazdny K. The interpretation of a moving retinal image. Proc Roy Soc Lond, Ser B 1980; 208: 385–397
4. Warren WH, Morris J, Kalish M. Perception of translational heading from optical flow. J Experimental Psychology: Human Perception and Performance 1988; 14(4)
5. Nelson RC, Aloimonos J. Obstacle avoidance using flow field divergence. IEEE Trans Patt Anal Mach Intell 1989; 11(10)
6. Burger W, Bhanu B. Estimating 3D egomotion from perspective image sequences. IEEE Trans PAMI 1990; 12(18): 1040–1058
7. De Micheli E, Torre V, Uras S. The accuracy of the computation of optical flow and of the recovery of motion parameters. IEEE Trans Patt Anal Mach Intell 1993; 15(5).
8. Hummel R, Sundareswaran V. Motion parameter estimation from global flow field data. IEEE Trans Patt Anal Mach Intell 1993; 15(5).
9. Meyer FG. Time to collision from first-order models of the motion field. IEEE Trans Rob Autom 1994; 10(6)
10. Prazdny K. Determining the instantaneous direction of motion from optical flow generated by a curvilinearly moving observer. CGIP 1981; 17: 238–248
11. Ballard DH, Brown CM. Computer Vision. Prentice-Hall, Englewood Cliffs, NJ
12. Moravec HP. The Stanford cart and the CMU rover. Proc IEEE 1983; 71(7): 872–878
13. Branca A, Convertino G, Distante A. Hopfield neural network for correspondence problems in dynamic image analysis. Int Conf Artificial Neural Networks. October 1995; Paris
14. Campani M, Verri A. Motion analysis from first-order properties of optical flow. CVGIP: Image Understanding 1992; 56(1): 90–107
15. Rumelhart DE, McClelland JL. PDP Research Group. Parallel Distributed Processing: Explorations in the Micro-structure of Cognition. Vol. 1: Foundations. MIT Press, Cambridge, MA 1987