Combining Rotation-Invariance Images and Neural Networks for Road Scene Understanding

Zhigang Zhu, Haojun Xi, Guangyou Xu

Department of Computer Science and Technology, Tsinghua University

Beijing 100084, China . e-mail: gyx-dcs@mail.tsinghua.edu.cn

ABSTRACT

In this paper we present the results of training and testing backpropagation networks for the outdoor road scene understanding. Both the road orientations used for vehicle heading and the road categories used for vehicle localization are determined by the integrated system. The main features of the work are as follows. (1) The comprehensive image analysis techniques are combined with the adaptive neural networks. (2) An omni-view image sensor is used to extract image samples. The rotation-invariance image features are obtained for the classification network, and the results are used to select the orientation-estimation networks. (3) The internal representation, especially the number of the hidden units, is analyzed. Experimental results with real scene images are given.

1. Introduction

Image analysis and pattern recognition have been important application areas of artificial neural networks (ANNs) since early days. ANN architectures for early vision, especially motion perception, have been proposed based on physiological as well as psychological evidences [1,2]. In a much larger project, a full-sized self-driving van named ALVINN (Autonomous Land Vehicle In a Neural Net) equiped with video camera "eyes" and an onboard "brain" made from four workstations has been developed and built at CMU[3]. Recently neural nets have found potential applications in visual telecommunications[4].

An autonomous vehicle should have three basic functions when it moves safely in an outdoor road environment: road following, obstacle avoidance and self-localization. The ANN is a reasonable choice for these real world problems since it can learn efficiently and there are enough image data to constrain the model. Our work shares the similar goals with the ALVINN of CMU. However, there are three distinguish features of our approach: (1) The system estimates not only the headings of the vehicle but also its locations along the route. (2) Omni-directional view image and more global sensing data is used. In this way the system will not be troubled by the limited view angle of the camera which has brought lot of problems in the past work of road following. Moreover, rotation-invariance and rotation-dependence features are separated which greatly simplifies the heading decision and vehicle localization. (3) We use the pre-processed image data as the inputs of the network. As the result the number of the input variables is reduced.

2. Rotation-Invariance Images

2.1 Omni-view image sensor

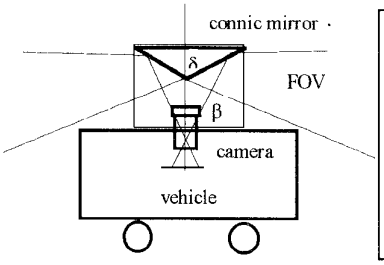


Fig. 1. Omni-view image sensor

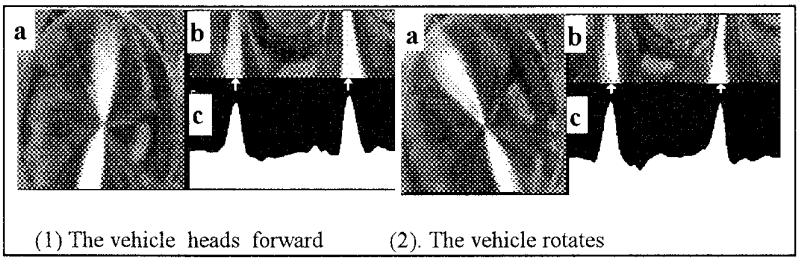


Fig. 2. (a) Omni-view images, (b). polar transform (c) orientation histogram

To capture the omni-directional views of the environment, various imaging methods have been explored, including rotating camera, fish-eye lens, and conic mirror. We adopt a conic projection image sensor similar to the one described in [5]. However we apply it to the outdoor natural scene environment. A prototype of the omni-view image (OVI) sensor is shown in Fig. 1. A conic mirror (with a vertex angle δ=55°) and a TV camera (with a viewing angle β = 20°) are fixed together by vertical thin bars. The OVI sensor mounted on the top of the vehicle and the vertical axis of the sensor aligns with the rotating axis of the vehicle. The image taken by the OVI sensor is a 360° view image ranging from about 5 meters to 20 meters on the ground around the vehicle,(Fig. 2a).

Although the resolution of the OVIs is relatively low, the 360° view image has some distinctive advantages when it is used in the road scene understanding: (1) The vehicle cannot miss the road. (2) The image is of rotation invariance in the sense that the structure of the image is not changed if the vehicle rotates around the optical axis of the camera, no matter what kinds of 3D structures of the environment are. (3) The vehicle can use not only the road information in front of itself but also the information behind and beside it. (4) The low resolution sensor image is suitable for the qualitative recognition (classification) of road categories. Since the lateral offsets of the vehicle on the road do not make great changes in the omni-view image, and the image appearances remain similar if the vehicle moves within the same road segment(category).

2.2 Image transformation

The omni-view image is preprocessed by a pipline image processing machine named PIPE ,hosted by a PC 486. Suppose the origin of the OVI coordinate system xoy is in the center of the image where the conic vertex is projected, we transform the Cartesian coordinate (x,y) into a polar coordinate (r,θ) (Fig. 2b):

$$r = \sqrt{x^2 + y^2}\ , \quad \theta = \tan^{-1}(y/x) \tag{1}$$

where r is the radius and θ is the orientation angle(0 - 2π). For a 256x 256 original sensor image, the resolution of the angle in the polar image is about 1 degree. The polar transformation can be carried out at the frame rate by PIPE.

The 2D polar image I(r, θ) is transformed to a 1D orientation histogram u(θ) by using a projection transformation along a given orientation θ:

$$u(\theta) = \sum_r I(r,\theta) \tag{2}$$

In this paper the rotation-independence image data is obtained by using the Fourier transform of the original data, other than determining the road orientation in the preprocessing stage as in the paper [6], which may bring errors to the samples. The orientation of the road is estimated hereafter. The projection transformation is also implemented by the PIPE at the frame rate.

2.3 Rotation invariance and dependency

First the orientation histogram u(θ) is sampled and then normalized as U = {u(n), n=0,...,N-1} so that the mean is zero , and the standard deviation is 1. The normalized procedure eliminates or reduces the influence of any illumination changes when taking images at different times. If rotation angle of the vehicle is

$$\phi = -\frac{2\pi n_0}{N} \tag{3}$$

where $\phi \in [-2\pi, 0]$, then the resulted orientation histogram v(n) is a circular shift of u(n) by n_0 , denoted as

$$v(n) = u(n - n_0) \tag{4}$$

where u(n) is the orientation histogram when the vehicle heads for the front road. The Fourier transform of u(n) is

$$a(k) = \frac{1}{N}\sum_{n=0}^{N-1} u(n)\exp(-\frac{j2\pi kn}{N})\ , \quad \text{k} = 0,...,\text{N-1} \tag{5}$$

so the Fourier transform of v(n) can be expressed as

$$b(k) = a(k)\exp(-\frac{j2\pi n_0 k}{N}), k = 0,1,\ldots,N-1. \tag{6}$$

By representing a(k) and b(k) in amplitude-phase forms

$$\begin{aligned} a(k) &= a_k \exp(j\psi_k) \\ b(k) &= b_k \exp(j\varphi_k) \end{aligned}, \text{ k=0,...,N-1,} \tag{7}$$

we have the following results:

$$b_k = a_k, k = 0,\ldots,N-1 \tag{8}$$

$$e^{j\psi_k} = e^{j(\varphi_k + k\phi)} \tag{9}$$

$$\psi_k = 2\pi m_k + \varphi_k + k\phi \text{ , k=1,...,N-1} \tag{10}$$

where m_k is a integer which indicates $2\pi m_k$ additive ambiguous in the kth phase value. Equation (8) says that the Fourier amplitudes are invariant to the rotation of the omni-view images. Therefore they are appropriate features for road scene classifications. Equations (9) and (10) give the basic relation to estimate the orientation difference between two omni-view images. For real scene images, the equality can not keep strictly, so the orientation difference should be estimated by searching the minimum value of the following distance

$$d(\phi) = \sum_{k=1}^{N/2} (a_k e^{j\psi_k} - b_k e^{j(\varphi_k + k\phi)})^2 \tag{11}$$

for each $\phi=\phi(n_0)$, $n_0=0,1,\ldots,N-1$. The computing complex of this procedure is $O(N^2)$. Here we give an alternative approach which has only O(N) time complex. From equation (9) we can obtain

$$e^{j\phi} = e^{j(\Delta\psi_k - \Delta\varphi_k)}, k = 1,\ldots,\frac{N}{2}-1 \tag{12}$$

where $\Delta\psi_k = \psi_{k+1} - \psi_k, \Delta\varphi_k = \varphi_{k+1} - \varphi_k$. The final orientation vector can be estimated as

$$e^{j\phi} = \frac{1}{N/2-1} \sum_{k=1}^{N/2-1} w_k e^{j(\Delta\psi_k - \Delta\varphi_k)} \tag{13}$$

with the weight function

$$w_k = (\frac{a_k}{|a_k - b_k|+1}) / \sum_{i=1}^{N/2-1} \frac{a_i}{|a_i - b_i|+1} \tag{14}$$

In the above equations, the final orientation vector is the average of (N/2-1) orientation estimates. If the errors of these estimates are independent random noises with zero mean and variance σ, the variance of the average error reduces to $\sigma/(N/2-1)$. Moreover, each orientation vector is weighted by the corresponding Fourier amplitude and divided by the amplitude difference. The reason for this operation is that the stability of the phase shift is proportional to the absolute amplitude and inversely proportional to the amplitude difference of the two sequences. This method is used in the data collection for the networks since the behavior of the vehicle can be controlled in the training stage. In the real operations, the orientation is estimated using the neural networks.

3. The Architecture

The basic model for Road Understanding Neural Networks(RUNN) is a adaptive combination of a image processing module (IPM) and several fully connected two- or three-layered backpropagation networks --- a single Road Classification Network (RCN), a Road Orientation Network (RON) for each road category . The inputs of the RCN are P (<=N) rotation-invariance image data (e.g. Fourier amplitudes), and the outputs are M road categories. The inputs of each RON are Q (<=N) rotation-dependence image data (e.g. Fourier phases), and the outputs are L road orientations. The data representation of the inputs and the number of hidden units for each network is decided by the experiments of training and testing.

The RCN and RONs are built up using *Nworks* tool[7], while the IPM is composed of a PC486 and the pipeline image processing system PIPE (Fig. 3). The processing of the RUNN is as follows. (1). The omni-view image is captured and transformed by the IPM and the rotation-invariance Fourier amplitude $A=\{a_k, k=0,...,N-1\}$ and rotation-dependence Fourier phases $\Psi=\{\psi_k, k=0,...,N-1\}$ are obtained. (2).The rotation-invariance data A is used to decide the road category by the RCN. (3).The road category estimation is used to select the correct RON, and the rotation-dependence data Ψ or the original orientation histogram U is feed into the RON to estimate the road orientation.

Advantages of the separation of the road classification and orientation estimations are intuitive. First, As the rotation invariance data are used as the input of the RCN instead of the original image, the distinctiveness of the input units is increased, and therefore the complexity of the network is reduced. If the Fourier amplitude A is used as the input of the RCN, the number of the input units can be reduced (P<=N/2). Second, since a separate RON is used to estimate the road orientation for each road category and the classification result is used to select the right RON, the efficiency of the network will improved.

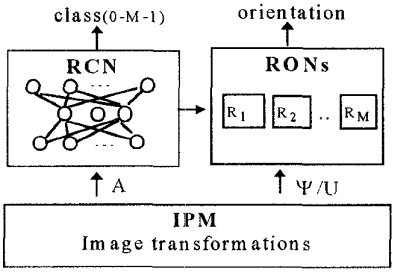


Fig 3. The Architecture of RUNN

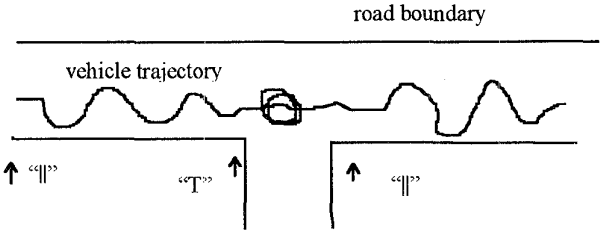


Fig. 4. Collecting the Data

The basic model of the processing elements(PEs) are basically determined by the summation function and the transfer function. The summation function is

$$I_i = \sum_j w_{ij} x_j + \beta \tag{15}$$

where i is the current PE, j is a PE that i is connected to, x_j is the output of PE j, w_{ij} is the weight of the connection of i and j, and β is the bias value. The transfer function is the hyperbolic tangent (TanH), whose range is from -1 to +1.

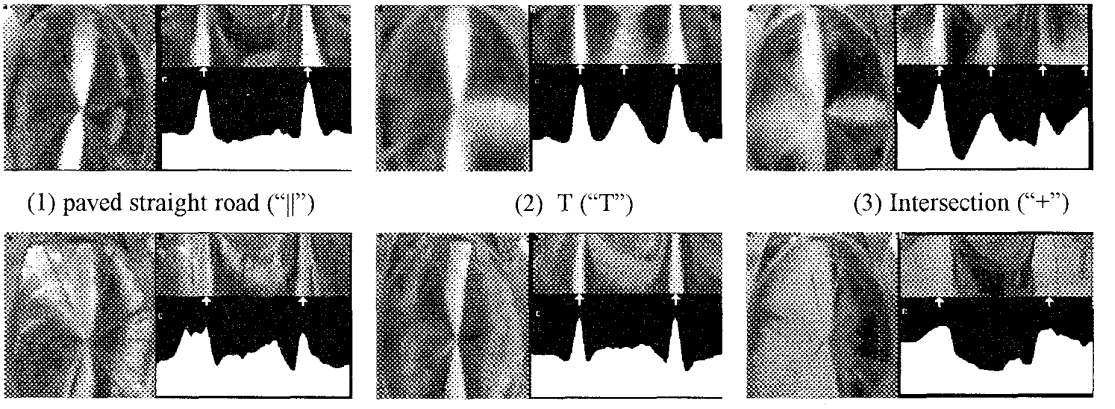
4. Collecting the Data

4.1. Collecting the data for the RUNN

The data for the RUNN is collected while the vehicle is moving on the road. In the current implementation, the vehicle moved along the route around the Main Building at the campus of Tsinghua University. The omni-view image sequences are recorded by a video camera record. At the laboratory, the image sequences are played back and processed by the PIPE machine and the PC486 to determine the road orientation and then extract the rotation-invariance and rotation-dependence features along the road.

At the beginning of each category of road segment, the vehicle heads for the front road (i.e. the road orientation angle is 0), and the desired outputs of the RCN, representing the road categories, are given by human supervisor. For the current experiments, each orientation histogram has 32 elements (N=32) and the road images are classified as 6 categories (M=6). They are the paved straight road surrounded by bushes and trees (denoted as "||"), three road junction("T"), intersection("+"), dirt road surrounded by grass and trees("D"), paved curved road passing through the garden in front of the building("C") and the square in front the Main Building("S"). In order to cover most of the situations, the vehicle moves on the road along a zigzag trajectory rather than a straight one, so that images are captured when the vehicle heads for different possible directions and locates on the road with different lateral offsets (Fig. 4). The orientation difference is calculated for the successive image frames within the same road category using equations (13). The absolute road orientation is obtained by accumulating the orientation

differences. In order to cover most of the cases, the sampled orientation histogram is shifted by software to simulate different road orientations. Both the inputs to the network as well as the desired outputs are mapped into numbers. Fig. 5 shows one typical sample for each of the six categories.



(1) paved straight road ("||") (2) T ("T") (3) Intersection ("+")

(4) Dirt road("D") (5) Curved road ("C") (6) Square ("S")

Fig 5. The sample images of the six road categories

4.2. Selecting and dividing the data

As part of collecting and preparing the data, it is important to make sure that examples selected for training the network do not have any dubious data fields (e.g. outliners). To this end, we calculate the mean Fourier amplitude (FA) vector of each category, and the distance between any FA vector of this category and the mean is used to judge whether it is a "good" example. For best results, the selection of training and testing set is based on the following rules: (1) The data is evenly divided among the various categories. (2) It is reasonably representative of the entire universe. (3) It is best to make the testing and training sets completely separate. The actual selection and division are listed in Table 1.

Category	||	*T*	+	*D*	*C*	*S*	*Total*
Original	1073	371	182	557	441	175	2799
Selected	930	312	160	445	374	148	2342
Training	133	133	133	134	133	133	799
Testing	797	179	26	311	241	15	1569

Table 1. Selecting and Dividing the data for the RCN

h	0	3	4	5	6	7	10	12	16
C	53634	51137	86752	73906	65581	75280	71064	86923	50939
e	0.30	0.25	0.20	0.20	0.15	0.15	0.12	0.10	0.14

Table 2. The learning process of the RCN

5. Training and Testing

The back-propagation learning strategy is used to training the network. In our implementation, the normalized cumulative delta learning rule was used for the RUNN. Examples in the training set were presented to the network randomly during the training to avoid the "learn one thing but forget others" problem. During the training and testing process, we studied the following four issues: (1) the suitable representation of input data, (2) the number of the hidden units, (3) the internal representation of the networks; and (4) the learning problem, for example, the epoch size, the converging speed, etc. .

5.1. Road classification

First the rotation invariance Fourier amplitudes are used to train the RCN. In this case the RCN has P=N/2=16 inputs (a_1,...,a_{16}) and 6 outputs. Experiments indicated that 16 or 32 is the proper size of the update epoch. The number of the hidden units (h) is decided by experiments in order to find the minimum number and best number for the problem. Table 2 shows the training results for different number of hidden units. "C" is the number of training iterations when the network becomes stable. The RMS errors "e" are also listed in the table.

Each realization of the RCN was tested using the training set, testing set and the original raw data set. When the value of one (e.g. kth) of the six network outputs $Y=(y_1, \ldots, y_6)$ is greater than 0.5 and the other five values are less than 0.5, then the input road image is classified as kth category. Table 3 lists the correct recognition rate(%) for the three data sets under every realization of the RCN.

The training and testing process indicates that 4 is the minimum number of the hidden units for proper classification. Comparing with the learning process of the networks using rotation independent orientation histogram as inputs[6] , the network RCN converges much slower and the recognition rate is slightly lower. The reason may be that the Fourier amplitudes lose the phase information of the orientation histogram and the Fourier transform compacts the energy in the first several terms. However the rotation-invariance of the Fourier amplitude vector makes it a good choice for the input of the RCN.

h	*0*	*3*	***4***	*5*	*6*	*7*	*10*	*12*	*16*
train	86.9	91.8	**96.3**	94.6	96.7	96.6	97.6	98.5	98.5
test	83.5	86.5	**91.6**	88.9	92.5	91.6	92.9	94.4	94.2
origin	77.2	79.5	**86.1**	84.1	85.9	85.3	87.6	89.0	88.6

Table 3. Performance of the RCN using FA

category	||	*T*	+	*D*	*C*	*S*
epochsize	4	1	1	16	8	8
train	100.	99.9	100.	83.9	96.2	89.4
test	99.0	100.	100.	83.1	95.7	86.7
original	96.8	98.1	87.9	70.4	90.8	76.5

Table 4. Performance of the RON with input U

5.2. Road orientation estimation

After the road category is determined , the corresponding RON is activated for this category. We compare the results of the network using original orientation histogram U and the Fourier phases Ψ as inputs. The outputs of the RON are L(=32) sampled orientations. The 3 layered RONs with vary number of hidden units do not converge when the input is the original phase data. So we use the original orientation histogram U as the inputs of the RCN. Experiments indicate that the RONs with no hidden units perform best for all the road categories. Table 4 shows the orientation estimation accuracy, measured by percentage of errorless orientation estimation for each road category. The epoch size of learning process , which is different for each road category, is also presented in Table 4. Analysis of the RONs reveals that the operations of the 2 layered orientation networks are quite similar to correlation functions. The estimation accuracy decreases when the input data become noisy for a certain category (e.g. "D"). This is just the reason of classifying the roads before orientation estimations. The orientation estimation could be improved by using the orientation difference of the temporal sequences (eqn(13)).

6. Conclusion and Discussion

In this paper we present the results of training and testing the backpropagation network for the outdoor road scene understanding. Both the road orientations used for vehicle heading and the road categories used for vehicle localization are determined by the integrated system. Experiments with real scene images are promising. Here we list some of our further works briefly. (1). More comprehensive image, e.g. 2D image patterns, would be used to recognize more complex road scene. (2). It is straight forward to apply spatio-temporal pattern recognition (SPR) network for recognizing image sequence of outdoor road scene. Since the same road category will last for a period of time, SPR network should not be sensitive to the occasional image events and could give a robust recognition.

References

[1]. Zhou , Y. T. , Chellappa, .R.A., " A network for motion perception,", Proc. IJCNN , 1990: pp II841-851.

[2].Atsumi,E.,et al, "Internal representation of a neural network that detects local motion ," Proc. IJCNN, 1993: 198-201.

[3]. Pomerleau,D.A., " Neural network based autonomous navigation," In Vision and Navigation : the CMU Navlab, Kluwer Academic Publishers, Boston, 1990.

[4]. Linggard, R. et al (ed.), Neural Networks for Vision, Speech and Natural Language, Chapman & Hall ,1st Ed., 1992.

[5]. Yagi., Y., et al, " Collision avoidance using omnidirectional image sensor(COPIS)," Proc ICRA, 1991 :910-915

[6]. Zhu, Z.G., Xu. G.Y.,Traning and testing neural networks for outdoor road scene understanding , Proc. Int. Conf. Neural Information Processing, Beijing,1995.

[7]. Neuralware Inc., Neural Computing / Reference Guide/ Using Nworks of Professional II/Plus, 1991.