Image and Vision Computing 19 (2001) 119–129

www.elsevier.com/locate/imavis

Road sign detection and recognition using matching pursuit method

S.-H. Hsu, C.-L. Huang*

Electrical Engineering Department, National Tsing-Hua University, Hsin-Chu 30043, Taiwan, ROC

Received 24 January 2000; revised 16 June 2000; accepted 27 June 2000

Abstract

This paper describes an automatic road sign recognition system by using matching pursuit (MP) filters. The system consists of two phases. In the detection phase, it finds the relative position of road sign in the original distant image by using a priori knowledge, shape and color information and captures a closer view image. Then it extracts the road sign image from the closer view image by using conventional template-matching. The recognition phase consists of two processes: training and testing. The training process finds a set of best MP filter bases for each road sign. The testing process projects the input unknown road sign to different set of MP filter bases (corresponding to different road signs) to find the best match.

Keywords: Matching pursuit filters; Template matching

1. Introduction

Recently, many intelligent vision systems have been developed for traffic automation [1–3]. They have many applications such as traffic control and analysis, license plate finding and reading, toll collection, automatic route planning and passive navigation. In this paper, we demonstrate a vision system that can recognize and detect road signs in images of cluttered urban streets as well as country roads. With a camera mounted on a vehicle at a height about 1.7 m, our system can be used to provide the driver with relevant information of the road signs on the scene.

The automatic detection and classification of road signs is clearly an emerging research topic in the field of intelligent vehicle. Different techniques [4–14] have been proposed for road sign detection from a sequence of monochromatic or color images. In Ref. [5], triangular, octagonal and circular contours, which are likely to represent the boundaries of road signs, are selected among the closed edge-contours. This makes the algorithm strongly dependent on the quality of the edge detection process. In Ref. [10] a more complicated strategy is followed, which uses both color and edges clues. Edges are tested at different levels of resolution by using so-called a Hierarchical Structure Code, which allows passing from the signal space of an image into the space of its symbolic representation. It is assumed that closed edge-contours are available at one of these levels of resolution, and failures happen when the outline of the traffic sign merges with the background.

A scheme for shape recognition, based on uncertainty handling and multiple knowledge sources combination and propagation, has been applied to correctly segment the images [4]. Tree structure model based systems indexed by shape, color and pictogram features have been implemented for the recognition of the detected road signs [8]. Ritter et al. [11] also used color segmentation algorithm to find the ROIs, which serve as hypotheses as potential road sign. Yabuki et al. [13] examined the color distribution of the road signs to construct the color similarity map. They incorporated the color similarity shown on the map into image function of an active net model. A road sign is extracted as if it is wrapped up in an active net. Similarly, we use color segmentation and template matching processes to detect the road sign. Our method is tolerant to noise, as the geometrical analysis of edges with the color information does not require that the extracted edges have good quality.

Picciolli et al. [6] developed a road sign detection and recognition scheme using a single monochromatic image, which also subdivides the process into three stages: (1) extraction of a search region; (2) shape detection; and (3) recognition. They applied the Kalman-filter-based temporal integration of the extracted information for further improvement. Escalera et al. [12] proposed a road sign detection and classification system, which has two main parts. The first one uses the color thresholding to segment the image and shape analysis to detect the signs. The second one uses a neural network to classify the road signs. They used a

* Corresponding author. Fax: +886-35715971.
E-mail address: clhuang@ee.nthu.edu.tw (C.-L. Huang).

0262-8856/00/$ - see front matter
PII: S0262-8856(00)00050-0

receptive field neural network with input layer of 32×32 neurons, output layer of nine neurons and four hidden layers. The net was trained to recognize nine road signs. Gravila [14] developed a multi-feature hierarchical algorithm to match N templates with an image using distance transforms. They used the coarse-to-fine search for the translated parameters and grouped the N templates into template hierarchy based on their similarity. This way, they can match the multi-templates simultaneously at the coarse-level matching and then at fine-level matching, they compute the separate distance transform for the features of each sign.

This paper describes a vision system for road sign detection and recognition. The road signs in color images are acquired by a single camera mounted on a moving vehicle. Since for outdoor scene, the illumination conditions can vary considerably, special attention has been devoted to the robustness and flexibility of the system. The system is achieved by two phases: the detection phase and the recognition phase. In detection phase, we use color extraction and template-matching, along with geometrical reasoning based on a priori knowledge to detect the road signs.

The recognition phase is implemented by using robust and flexible MP filters [15]. The MP filters were introduced to represent the signals or images using an over-complete set of bases called MP bases. Phillps [16] modified the MP filters for solving pattern recognition problem. Different from Ref. [16] who used one set of filters for facial recognition, we use different sets of filters for different road signs. In Ref. [16] there are five basis elements for the MP filter which correspond to five different facial features, whereas, we use eight basis elements for each MP filter set. The MP filters are used to decompose a training pattern into a two-dimensional (2D) wavelet expansion. This yields a representation that is explicitly 2D and encodes information locally, unlike template matching that encodes information globally which is easily influenced by the complex environments. In the recognition phase, it finds the coefficient vector as the feature vector of the input road sign image by applying the MP filters. The road sign recognition system has three concerns: (1) how to recognize road signs with different rotation, translation and scale; (2) the resolution and lighting of the road signs can very considerably; and (3) how to present a good discriminative power with a low on-line computational cost.

2. Road sign detection

Similar to Ref. [6], our road sign detection system consists of three stages (see Fig. 1). In the first stage, a region in the captured image where the road sign is more likely to be found is selected. Here we use either the color information or a priori information (such as the possible location of the road signs) to identify the region. Therefore, the road sign location is limited to certain designated region,

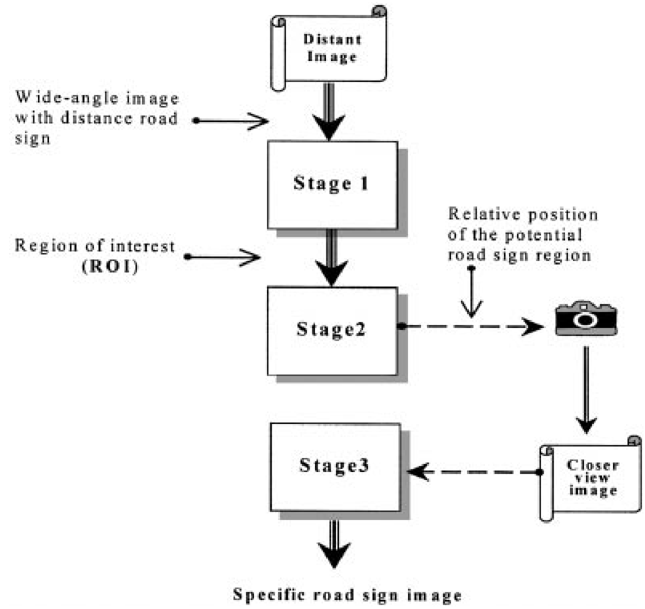

Fig. 1. The road sign detection phase.

called region of interest (ROI). In the second stage, we search the ROI to find the possible location of the triangular or circular shape regions. Then based on the positions of the identified regions, we can capture a closer view image in which the road signs may be detected. In the third stage, we use the template-matching to detect and then extract the specific road sign image.

(1) The first stage. The amount of computation for detecting the road signs in the image can be greatly reduced if the search is limited to a small region of the image where the road sign is likely to be found. There are three pre-processes in the first stage as: (a) down-sampling the input images; (b) using a priori information of the possible location; and (c) extracting color information of road signs in the image. First, we use the wide-angle camera to catch the road signs in the distance, and apply a priori knowledge to localize the road sing by extracting the middle-right part of the image as the ROI. Each image element in the rough search region is classified as 1 (red) or 0 (non-red) according to its hue (which must be between $\pm 30^\circ$ in angular scale that assigns 0° to red, 120° to green and 240° to blue) and saturation (which must be at least 20%).

(2) The second stage. We use either the *coarse search* or the *fine search* to detect the possible road signs location as the ROI. The *coarse search* uses a rectangular block with fixed size, whereas the *fine search* uses the template (triangular or circular) with variable sizes. Both of them use template-matching to find the road signs in the ROI. The second one operates by counting the red image pixels inside the rectangular block or the template, and then selects the rectangular block with red pixel ratio at least 20% or the variable shape-template with red pixel ratio at least 75%. In the second stage, it begins with *coarse search* to extract the possible location of the road signs. If the number of possible

Fig. 2. The variable sized triangular and circular template with fixed thickness belt.

road sign region is more than 10, then the *coarse search* fails and the *fine search* will be activated. When all the possible positions of road signs are found, we compute the relative position and move forward to capture a closer view image and define the ROI.

(3) The third stage. To detect the specific position of road signs in the ROI, we analyze the red component of the ROI to retrieve the circular and triangular contours representing possible road signs. The *triangular/circular road signs localization algorithm*, which looks for triangles/circles of variable size by inspecting the search region, can be outlined as follows:

1. Extract the red component of ROI from the down-sampled closer view image.
2. Find the possible location of the road sign by using the variable-sized triangular/circular templates with fixed thickness belt (see Fig. 2). By moving the templates around the ROI image, we count the number of red pixels inside the belts of the templates. If the number exceeds certain threshold, then they are selected as the triangle/circle candidates.
3. Average the centers of these triangle/circle candidates as the center of the *detection template* and choose the maximal *size* of these candidates as the size of the *detection template*.
4. Extract the regions enclosed by the *detection template*, enhance the contrast between the object and the background, and extract the road sign image by using region-growing technique [17]. After extracting the road sign image, we normalize the triangular/circular road sign region to 45×50-sized region by linear interpolation. The extracted road sign image is the input to the recognition module.

(4) The experiments. Fig. 3(a) shows a scene with a triangular road sign, Fig. 3(b) illustrates the closer view, and Fig. 3(c) demonstrates the red component of the search region. To search for a triangular road sign, we apply template matching by using a set of templates (see Fig. 3(d)). The templates were enclosed by triangular belt with 2 pixels thick and the internal side ranging from 6 to 30 pixel width. The matching was simply performed by counting the number of red pixels inside the triangular belt, and marking the triangular centers and the internal side of the template in which the red pixels found inside the triangular belt occupy 80% of the entire area of the belt. For each center cluster, we compute the centroid and the maximum radius as the location and size of the triangular road sign. Fig. 3(e) shows the contents of the triangular road sign as the input image to the road sign recognition system. Another experiment is also demonstrated. Fig. 4(a) and (b) shows a circular road sign in distant view and in the closer view. Fig. 4(c) shows the red component of the search region, and Fig. 4(d) illustrates the search region of circular road sign obtained by template matching. The templates were

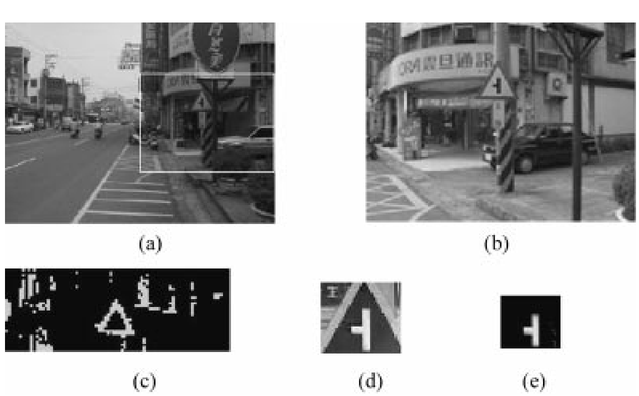

Fig. 3. Detection of the triangular road signs.

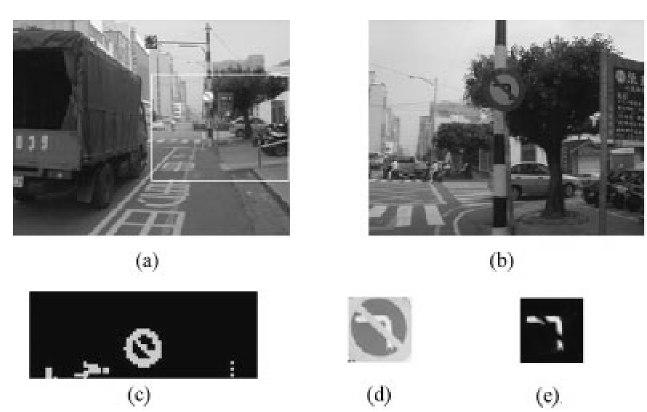

Fig. 4. Detection of the circular road signs.

enclosed by circular with 2 pixels thick and inner radius ranging from 3 to 15 pixels. Fig. 4(e) shows the contents of the circular road sign as the input image to the road sign recognition system.

3. Road signs recognition

Once the road sign has been identified, a recognition process is applied to interpret the road sign. The recognition scheme has to consider the following three concerns. First, the output of the detection algorithm is an image of road sign with nearly constant orientation, but the unknown factors, such as camera viewing direction and their relative position, will complicate the input image. Therefore, we need to use as many rotated, translated, or scaled templates as possible for identification. Second, the resolution and lighting of the road signs can vary considerably. Thus, the system has to cope with a large variety of appearances of each road sign. Third, the on-line recognition procedure needs to present a good discriminative power at a low computational cost.

Given as many rotated and translated observations as possible, the MP filter model is trained to generate a set of bases in which those observations are embedded. So the on-line recognition scheme can apply these bases to discriminate the input observation. The bases are also invariant to different illumination level. Our recognition scheme is fast, because the basis elements of MP filters, which are off-line trained, have been used to represent all the possible observations. Each observation is represented by projecting the input image onto a set of basis elements of the MP filters as a coefficient vector. The conventional template matching method must compare input signal with all templates one-by-one to find the best match. MP filter has conformed all possible templates to a set of bases and the dictionary of bases in the system is off-line trained.

Road sign recognition has two components. The first component is a training process, which determines a set of basis elements to represent a road sign. Each road sign is represented by a coefficient vector generated by applying the set of basis elements on the road sign image. The second component is the testing process. Given an unknown sign, it applies every set of per-trained MP basis elements to find a best match.

3.1. An overview of MP filters

The matching pursuit (MP), originated from Ref. [15], uses a greedy heuristic to iteratively construct a best-adapted decomposition of a function *f*on $\boldsymbol{R}$. The algorithm works by choosing, at each iteration i, the wavelet g in the dictionary $\mathbf{D}$ that has maximal projections onto the residue of f. The best-adapted decomposition is selected by the following greedy strategy. Let $R^0 f = f$; then g_i is chosen such that

$$|\langle R^i f, g_i\rangle| = \max_{g \in D} |\langle R^i f, g_i\rangle| \tag{1}$$

where $R^{i+1} f = R^i f - \langle R^i f, g_i\rangle g_i$ for $i \geq 1$.

Each wavelet in the expansion is selected by maximizing the right hand term in Eq. (1). This equation allows for an expansion based on a single function, and minimizes the reconstruction error. The extension from function f on $\mathbf{R}$ to functions (templates) t on $\mathbf{R}^2$ is straightforward. A dictionary of 2D wavelets is used.

(1) Time–frequency atoms. Decompositions of signals over a family of functions that are well localized both in time and frequency have found many applications in signal processing. Such functions are called time–frequency atoms. Scaling, translating and modulating a single window function $g(t)$ can generate a general family of time–frequency atoms. We also impose that $\|g(t)\| = 1$. For any scale $s > 0$, frequency modulating ξ and translation u, we denote $\gamma = (s, u, \xi)$ and define

$$g_\gamma(t) = \frac{1}{\sqrt{s}} g\left(\frac{t-u}{s}\right) \mathrm{e}^{\mathrm{i}\xi t} \tag{2}$$

The factor $1/\sqrt{s}$ normalizes to 1 the norm of $g_\gamma(t)$. If $g(t)$ is even, which is generally the case, $g_\gamma(t)$ is centered at the abscissa u. Its energy is mostly concentrated in a neighborhood of u, whose size is proportional to s. Let $\hat{g}(\omega)$ be the Fourier transform of $g(t)$. Eq. (2) yields

$$\hat{g}_\gamma(\omega) = \sqrt{s}\hat{g}(s(\omega - \xi)) \mathrm{e}^{-\mathrm{i}(\omega-\xi)u} \tag{3}$$

Since $|\hat{g}(\omega)|$ is even, $|\hat{g}_\gamma(\omega)|$ is centered at frequency $\omega = \xi$. Its energy is concentrated in a neighborhood of ξ, whose size is proportional to $1/s$. The family $D = (g_r(t))_{\gamma \in \Gamma}$ is extremely redundant. To represent efficiently any function $f(t)$, we must select an appropriate countable subset of atoms $(g_{r_n}(t))_{n \in N}$, with $\gamma_n = (s_n, u_n, \xi_n)$, so that $f(t)$ can be written as

$$f(t) = \sum_{n=-\infty}^{+\infty} a_n g_{\gamma_n}(t). \tag{4}$$

(2) Matching pursuit. Let $g_{\gamma 0} \in D$. The vector f can be decomposed into

$$f = \langle f, g_{\gamma 0}\rangle g_{\gamma 0} + Rf \tag{5}$$

where Rf is the residual vector after approximate f in the direction of $g_{\gamma 0}$. Clearly, $g_{\gamma 0}$ is orthogonal to Rf, hence

$$\|f\|^2 = |\langle f, g_{\gamma 0}\rangle|^2 + \|Rf\|^2 \tag{6}$$

To minimize $\|Rf\|$, we must choose $g_{\gamma 0} \in D$ such that $|\langle f, g_{\gamma 0}\rangle|$ is maximum. A matching pursuit is an iterative algorithm that decomposes the residue Rf by projecting it on a vector of D that matches Rf at best, as it was done for f. This procedure is repeated each time on the following residue that is obtained. When the dictionary is complete,

$$f = \sum_{n=0}^{+\infty} \langle R^n f, g_{r_n} \rangle g_{r_n}. \tag{7}$$

3.2. MP filters for road sign recognition

A particular road sign is represented as an n-dimensional vector $(a_0, \ldots, a_{n-1})$, called a coefficient vector. One computes the coefficient value a_i by projecting the road sign image onto a set of bases, $\{g_0, \ldots, g_{n-1}\}$, which need not be orthogonal. Because the basis is not necessarily orthogonal, an iterative projection algorithm is applied to calculate the coefficients. If the basis is orthogonal, then the algorithm reduces to the standard projection method. The projection algorithm adjusts for the non-orthogonal property by using residual images. If I is an image (or template), then $R^i I$ is the residual image during iteration i, where $R^0 I = I$. The coefficient a_i is the projection of the residual image R^i onto the basis element g_i, as

$$a_i = \langle R^i I, g_i \rangle, \tag{8}$$

where $\langle \cdot, \cdot \rangle$ is the inner product between two functions. The residual image is updated after each iteration by

$$R^i I = R^{i-1} I - a_{i-1} g_{i-1} \quad \text{for } i \geqq 1. \tag{9}$$

After the nth iteration, an image I is decomposed into a sum of residual images:

$$I = \sum_{i=o}^{n-1} (R^i I - R^{i+1} I) + R^n I. \tag{10}$$

Rearranging Eq. (9) and substituting into Eq. (10) yields

$$I = \sum_{i=o}^{n-1} a_i g_i + R^n I \tag{11}$$

and the approximation of the original image after n iterations is

$$\hat{I} = \sum_{i=o}^{n-1} a_i g_i. \tag{12}$$

The approximation needs not be very accurate, since the encoded information is enough for recognition.

The goal of the recognition algorithm is to determine to which particular class an observed pattern belongs. Therefore, there must be a way of measuring the difference between two patterns. With MP filters, the recognition process compares the coefficient vectors from two patterns, where the coefficient vectors are generated by the same set of bases. The similarity measure between two patterns is the angle between their coefficient vectors. This measure is invariant under linear changes in the contrast of the image [15,16]. Since if the contrast image is changed then only the magnitude of the coefficient vector will be changed. Furthermore, if the basis is composed of wavelets, then the similarity measure is also invariant to the illumination level in the image.

3.3. The training process

The training process is developed to find a set of basis elements for each road sign so that the road sign images can be recursively decomposed and then represented by the coefficient vectors effectively. The main purpose of the training process is to find sets of basis elements so that the coefficient vectors of different road sign classes are as separated as possible, whereas those of the same training class are as close as possible. Therefore, we have the intra-class training for finding a set of basis elements for each road sign, whereas the interclass training process readjusts those basis elements for different road signs.

3.3.1. Intra-class training

Normally, the same road sign should have the same coefficient vector, and all occurrences of this coefficient vector represent that designated road sign. So we must train different examples of the same road sign to find a set of bases from which the cluster of coefficient vectors can be generated to represent the different occurrences of the same road sign. The coefficient vector represents the cluster center is referred to as a *proto-sign* (or, in general, a *proto-object*).

The MP filter for road sign identification is trained based on m different examples of that particular road sign. Let $\{I_l, \ldots, I_m\}$ be m appearance of the designated road sign. These road sign images are aligned with their centers located at the origin. A greedy algorithm is applied to select the basis elements. In iteration i the basis function g_i is selected. How to choose g_i is determined by the residual images $R^i I_l$ and coefficients a_j^l from previous iterations, i.e. $j < i$. Let the coefficient $\alpha_j^l = \langle R^j I_l, g_j \rangle$, that is, the jth coefficient for road sign image I_l. The set of coefficients generated through the ith iteration is denoted by $\Lambda_i = \cup_l(\alpha_0^l, \ldots, \alpha_i^l)$, $i \geq 0$, and $\Lambda_{-1} = 0$.

There are three steps operating iteratively for the training process. In the first step, a new basis function g_i is selected or updated. In the second step, the coefficient vectors for each road sign image I_l are updated. In the third step, the residual images are updated by $R^{i+1} I_l = R^i I_l - \alpha_i^l g_i$. The ith basis function is selected by the following optimization procedure:

$$g_i = \arg \min_{g \in D} C_g(R^i I_1, \ldots, R^i I_m, \Lambda_{i-1}) \tag{13}$$

where C_g measures how well the coefficient vectors cluster when the ith basis function is g_i. It is defined as

$$C_g = 1/m \sum_l \| \Gamma - (\alpha_0^l, \ldots, \alpha_{i-1}^l, \langle R^i I_l, g \rangle) \| \tag{14}$$

where Γ is the centroid of $(\alpha_0^l, \ldots, \alpha_{i-1}^l, \langle R^i I_l, g \rangle)$. The function C_g is a mean distance evaluated for each $g \in D$, and the specific g that minimizes C_g is selected as the basis element g_i. In the current implementation of C_g for a given g, the

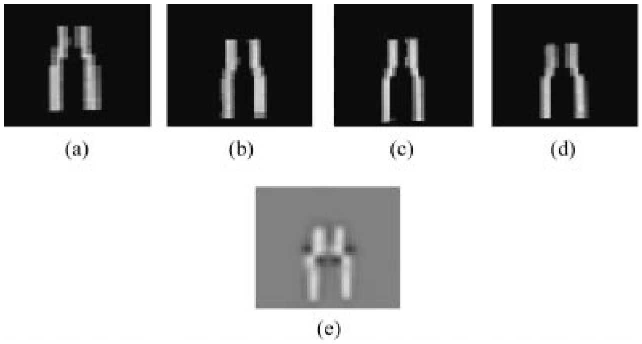

Fig. 5. Design of recognition filters: (a)–(d) are training set; (e) particular sign is reconstructed by Eq. (12) using Gabor function.

cluster vector is the mean of $(\alpha_0^l, \ldots, \alpha_{i-1}^l, \langle R^i I_l, g \rangle)$, $1 \le l \le m$. Once the cluster vector is determined, C_g computes the average distance from the coefficient vectors to the cluster vector. This distance is a measure of variance of the coefficient vectors about the cluster vector. If the dispersion is small, then g_i is a good candidate. On the other hand, if the dispersion is large, then g_i is a poor choice.

The training algorithm generates an ordered list of n basis elements and a list of n coefficients. The training algorithm generates k classes of proto-signs, and the MP filter dictionary consists of k sets of bases and k coefficient lists (coefficient vectors). The location of the basis elements encodes the geometric structure of a road sign. The centers of the basis-elements g_i are usually not aligned. The MP filter can be used to represent a road sign image that is larger than the support of an individual basis element. The projection operation $\langle R^j I_l, g_j \rangle$ used to compute α_j^l is accomplished by translating every basis elements g_j by (u_1,u_2), and projecting the image $R^j I_l$ on the translated basis elements.

Fig. 5 illustrates the training of a recognition filter. Fig. 5(a)–(d) are the training set, and Fig. 5(e) is the reconstructed sign by using Eq. (12). The algorithm is iterated

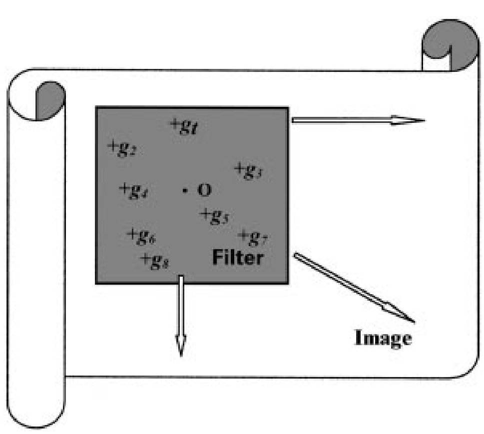

Fig. 6. Matching pursuit filter scanning an image. The center of the filter is O, which moves as the image is scanned. This filter has eight basis elements, g_1, g_2, g_3, g_4, g_5, g_6, g_7, and g_8. The centers of these wavelets g_i relative to O are marked by " + ".

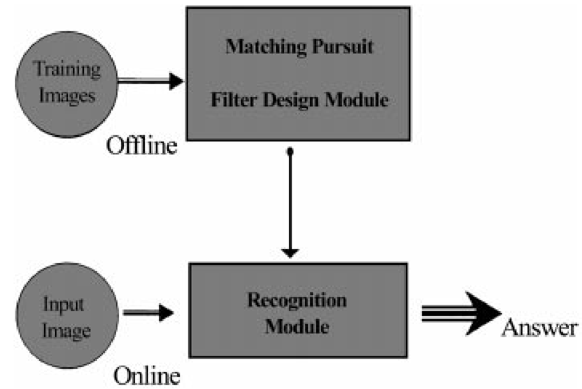

Fig. 7. The road sign recognition system organization.

until n basis elements are selected. The choice of the number of basis elements (i.e. n) is usually determined experimentally. If n is too small, then the false-alarm rate is too high; if n is too large, the filter will not be generalized to represent the signs outside the training set.

3.3.2. Inter-class training

For a particular road sign, the MP filter training procedure selects a basis set in which the coefficient vectors clustered, and it also generates one *coefficient vector* Λ to represent a class of road signs. To distinguish different classes of road signs, the training algorithm generates a coefficient vector for each individual. The lth road sign class is represented by a coefficient vector $\Lambda^l = \{\alpha_0^l, \ldots, \alpha_{n-1}^l\}$. To measure the similarity between an unknown road sign k and the designated road sign class l, we compare coefficient vector Λ^l and the coefficient vector $a^k(u_1,u_2)$. The algorithm computes the image coefficient vector $a^k(u_1,u_2)$ by expanding the image about the pixel (u_1,u_2). This expansion is accomplished by translating every class of basis elements g_i^k by (u_1,u_2), and projecting the image on the translated basis elements (see Fig. 6). Let $a_i^k\ (u_1,u_2)$ be the ith coefficient of $a^k(u_1,u_2)$; then

$$a_i^k(u_1, u_2) = \langle R^i \mathrm{I}, g_i^k(\cdot + u_1, \cdot + u_2) \rangle \tag{15}$$

The unknown road sign k is recognized as the road sign class l if the distance between Λ^l and $a^k(u_1,u_2)$ is the shortest. To decrease the likelihood that signs are misidentified, the MP training algorithm searches for a basis that *separates* the Λ^l coefficient vectors between different classes of road signs. The algorithm for selecting the ith basis element for differentiating the lth class road signs from other road signs is similar to the intra-class training, but with a *modified select basis function*, $\hat{C}_{\hat{g}_l}$. To recognize the lth particular road sign, we select the lth set of base function by using the following algorithm:

Loop
{
1. for $\boldsymbol{l} = 1$ to **N**, (**N** classes of road signs)
g_i^l: the ith basis of the $\boldsymbol{l}$th class road sign.

2. for $\boldsymbol{j} = 1$ **to N**, $\boldsymbol{j \neq l}$
  compute $\hat{g}_i^l = \arg\max_{g \in D} \hat{C}_{\hat{g}_l}(R^i I_1^j, \ldots, R^i I_m^j, \Lambda_{i-1}^j)$;
  Return $g_i^l = \hat{g}_i^l$;
}
If $\sum_l d_\theta(g_i^l, \hat{g}_i^l) < threshold$
  then **stop** and **return** g_i^l
  else **do Loop**

where $\{I_m^j\}$ is the set of training images for the jth class sign, j is the road sign class index, and m is the training image number for the same road sign class.

On the correct implementation of $\hat{C}_{\hat{g}_l}$ for a given $\hat{g}_l$, there are $N-1$ cluster vectors which are means of $(\alpha_0^{k,j}, \ldots \alpha_{i-1}^{k,j}, \langle R^i I_k^j, g^l \rangle)$, $1 \leq k \leq m$, $1 \leq j \leq N$, $j \neq l$, respectively. Let the coefficient be $\alpha_i^{k,j} = \langle R^i I_k^j, g_i \rangle$ that is, the ith coefficient of the kth training image of the jth class. The $\hat{C}_{\hat{g}_l}$ measures the similarity (in terms of angle) between the average of coefficient vectors of the lth set of bases and

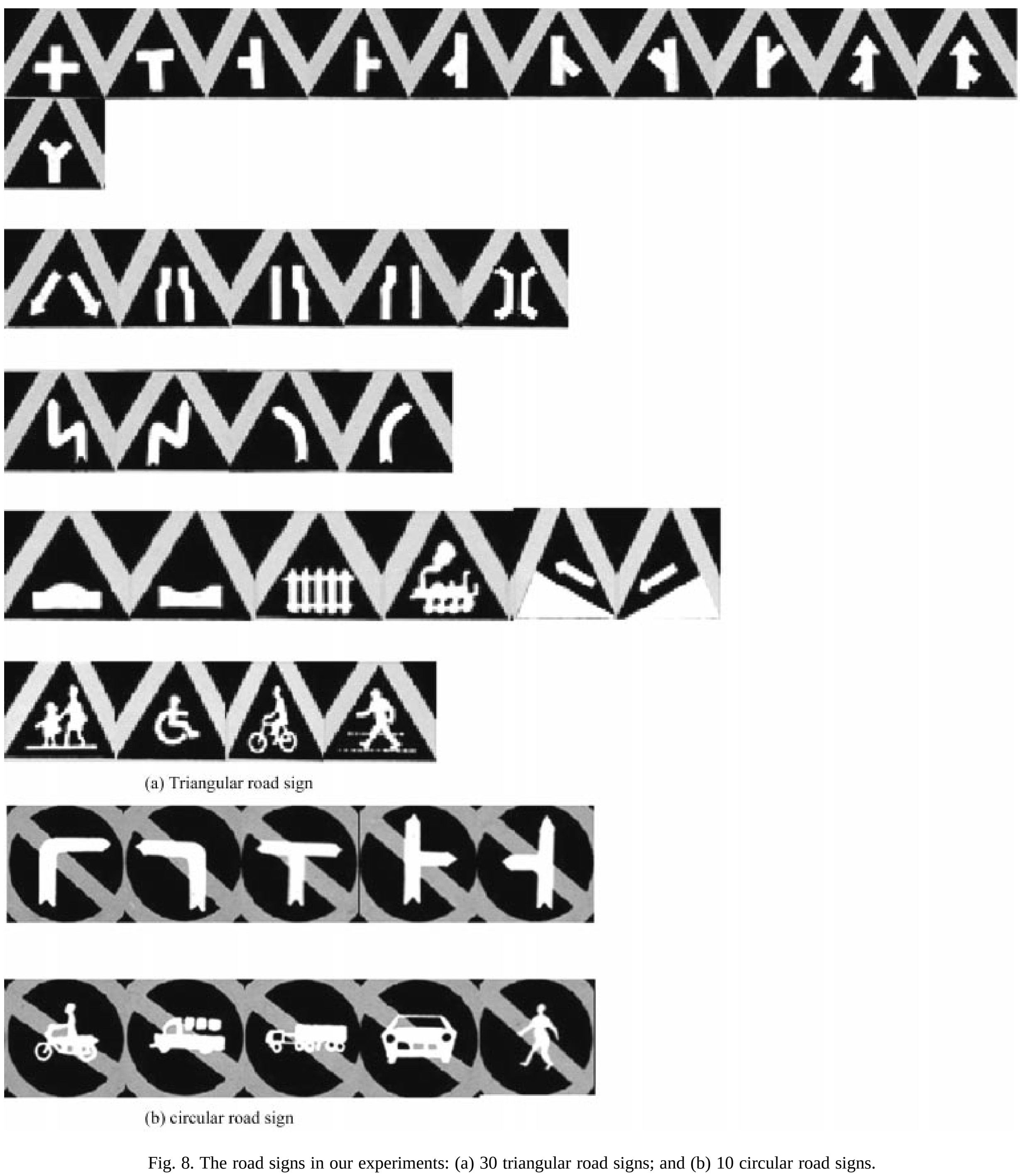

Fig. 8. The road signs in our experiments: (a) 30 triangular road signs; and (b) 10 circular road signs.

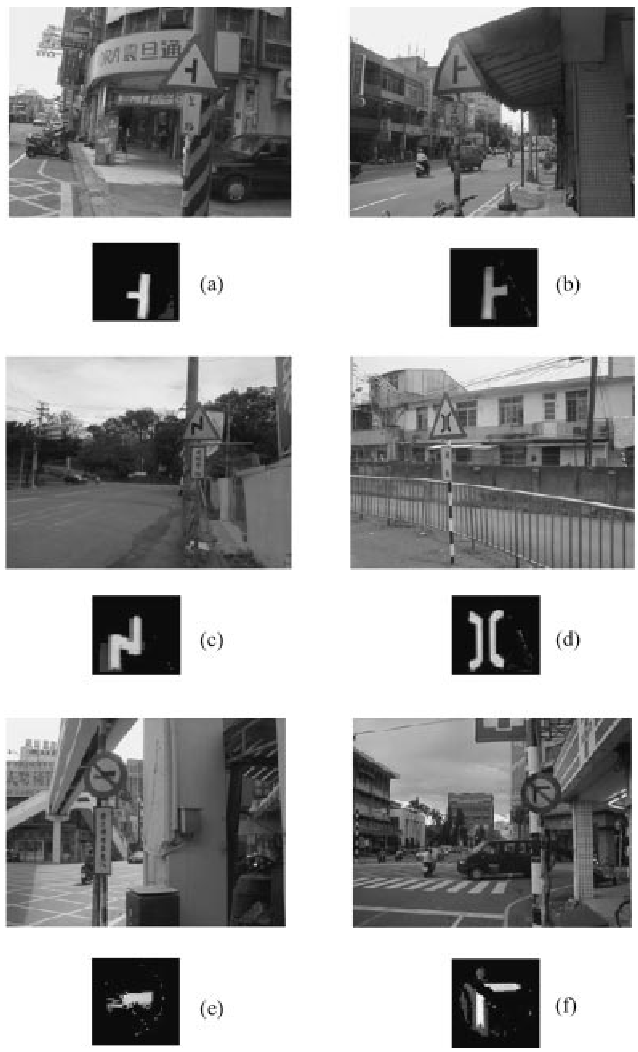

Fig. 9. Examples of road signs with complex environment are detected (a)–(f).

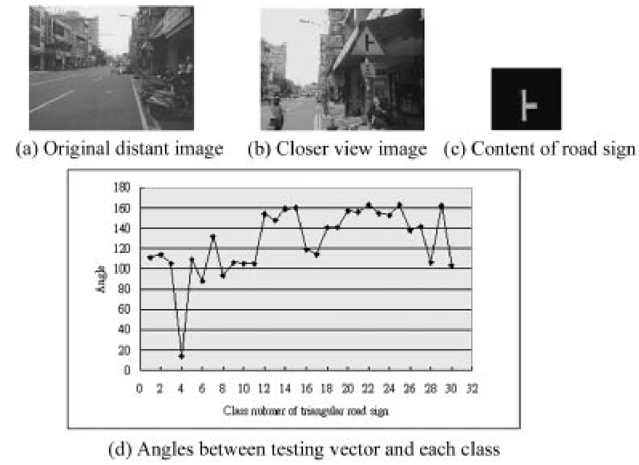

Fig. 10. Recognition of the fourth class of the triangular road sign.

the jth cluster vector, $1 \leq j \leq N$, $j \neq l$, its operation is similar to Eq. (14). If these angles are very large (wide dispersion), then the $\hat{g}_i^l$ is a good candidate to replace g_i^l (obtained in intra-class training stage). The iteration is performed for every class l until the sum of difference between g_i^l and $\hat{g}_i^l$, where $l = 1\ldots N$, is below a very small threshold.

3.4. Road sign identification

Here we apply the MP filters to recognize a particular road sign by scanning every particular MP filters (corresponding to a designated road sign class k) across an image, which results in responses S^k. The response at pixel (u_1,u_2),denoted as $S^k(u_1,u_2)$, measures the similarity between the region centered at location (u_1,u_2) and the designated proto-sign (see Fig. 6). One criterion detects the center of the sign at the maximum response. An alternative method reports all points above a threshold as the center of the sign. In identification phase, for each input image, the algorithm uses the kth MP filter set to compute the coefficient vector $\boldsymbol{a}^k(u_1,u_2)$ at a pixel (u_1,u_2), and then compares every particular proto-sign coefficient vectors $\boldsymbol{\Lambda}^k$ in the database with the image coefficient vector $\boldsymbol{a}^k(u_1,u_2)$. There are K image coefficient vectors, $\boldsymbol{a}^k(u_1,u_2)$, $k = 1\ldots K$, generated for each input image.

$\Lambda^k = (\alpha_0^k,\ldots,\alpha_{n-1}^k)$ be the kth cluster vector that represents the kth particular proto-sign. Then $S^k(u_1,u_2) = d_\theta(\boldsymbol{\Lambda}^k, \boldsymbol{a}^k(u_1,u_2))$, where $d_\theta(\cdot,\cdot)$ is the cosine of the angle between two vectors, i.e. the response is the cosine of the angle between $\boldsymbol{\Lambda}^k$ and $\boldsymbol{a}^k(u_1,u_2)$. The $S^k(u_1,u_2)$ measures the similarity between the input image to the road sign k represented as$\boldsymbol{\Lambda}^k$, a coefficient vector that represents the kth road sign. After all the coefficient vectors of all road sign classes having been determined, the next step is to recognize the

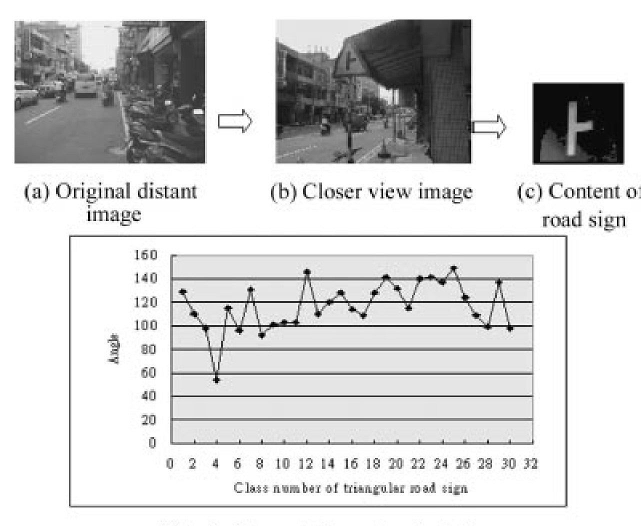

Fig. 11. Recognition of the fourth class of the triangular road sign with translation.

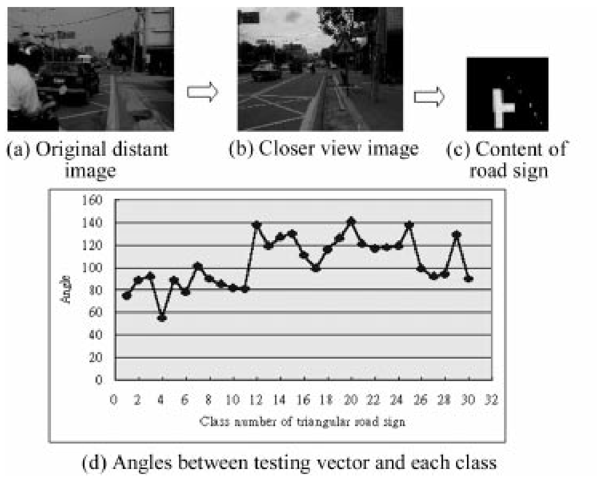

Fig. 12. Recognition of the fourth class of the triangular road sign at twilight.

road sign image, which is found by searching for the maximum response over all the S^k

$$\hat{S}^{\hat{k}}(\hat{u}_1, \hat{u}_2) = \max_{k,(u_1,u_2)} S^k(u_1, u_2) \tag{16}$$

where $(\hat{u}_1, \hat{u}_2)$ is the estimated center of the sign in the image. The algorithm reports that the identity of the input image is the road sign $\hat{k}$. Using the above algorithm, we develop the recognition module for road sign recognition (see Fig. 7).

3.5. Implementation and comparison

The complexity of our method is determined by the off-line training process, which generates a number of basis elements. For the on-line recognition, we just move the MP filter to scan the image to find the best match. The computation complexity of recognition process is determined by the following three scenarios: (a) the number of selected basis elements, i.e. $g_0,\ldots,g_{n-1}$, which indicates different rotation, translation and scaling version of the time-frequency atoms; (b) the number N of road signs to be identified which is the number of sets of the bases elements; and (c) the number of locations that we apply the matching pursuit basis elements (or the dimension of coefficient vector). The choice of the number of basis elements (i.e. n) is usually determined experimentally. If n is too small, then the false-alarm rate is too high; if n is too large, the filter will not be generalized to represent the signs outside the training set.

In our experiments, we have thirty different time–frequency atoms or basis element g_i (i.e. $n = 30$). In the ith iteration, the matching pursuit operation is operated by translating the g_i to a relative location (u_1,u_2) to match the ith time–frequency atoms with the residue images R^iI for the best match and the coefficient (see Eq. (14)). We may find eight best locations (see Fig. 6) relative to the center of O to

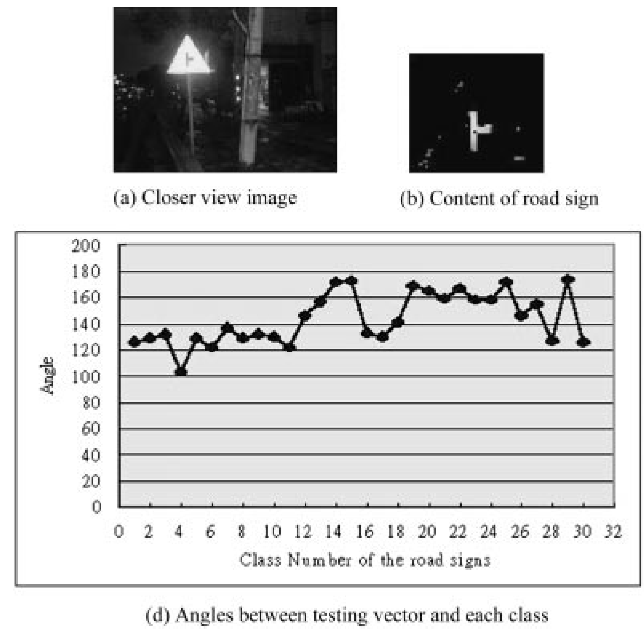

Fig. 13. Recognition of the fourth class of the triangular road sign at night.

recursively apply the projection operations. Therefore, the dimension of coefficient vector is 8. The matching pursuit operation is illustrated as follows:

(1) Scan the internal region (30×40) of the road sign image, for each location, try all basis elements $(n = 30)$ and find the best basis element that provides the best projection (the one generates the largest projection coefficient).
(2) Compare all the projection operations on all locations and select the best one for current scan.
(3) Update the residual image (Eq. (9)) for the next time scan.
(4) Repeat the matching pursuit operations (steps (1)–(3)) eight times and generate eight projection coefficients.

The computation complexity of our on-line recognition module is determined by the order of $8 \times 30 \times (30 \times 40) \times N$. Actually, in real implementation, to speed our matching pursuit operation, we may reduce the density of scanning by a factor of 1/4 or 1/9. The MP filter consists of a set of best-adapted bases, which are embedded with all the possible information of the target. The recognition scheme can apply these bases to discriminate the input with different conditions. However, as the variation of the road signs becomes larger, we need to increase the dimension of the coefficient vector to represent the different appearance of the road sign effectively.

Template-matching is a straightforward and effective method which recognizes the target by matching the input image with all possible appearance of target called templates. However, this algorithm becomes complicate once the appearance of target changes dramatically. The complexity of

Table 1
Detection phase

Detection phase	Total testing road signs	Correct detection (%)	False detection
Triangular road signs	356 individuals	93	25
Circular road signs	125 individuals	95	7

template-matching increases with the number of templates and the complexity of the input image in the order of $O(N(TSR))$ under the assumptions that object appears with different translation (T), Scale (S), and rotation (R).

4. Experimental results

The experiments are tested for 30 triangular road signs and 10 circular road signs (see Fig. 8). To demonstrate the capability of our system, we perform two different experiments. The first one is the road sign detection, the second one is the road sign identification, it includes recognizing the road sign under different viewing and lightening conditions. Our system is tested mostly in the urban streets in the City of Hsin-Chu, Taiwan. The algorithm has been implemented on Pentium II PC 300 MHz with Oculus F-64 Frame grabber and SONY XC-7500 CCD sensor. The computation time of the detection phase is 100 ms for color information processing and then extracting the road sign image from a 512×480 image, whereas the recognition operation using matching pursuit method requires about 250 ms.

In the detection phase, the original distant images are acquired by a camera mounted on a vehicle at the height of 1.70 m (above the ground). The optical axis of the camera is parallel to the moving direction of the vehicle. The vehicle can be assumed to move forward and travel along the direction of the road. In the first and second stages of the detection phase, we compute the possible position of the road sign, and move the camera directly to that position to capture the closer view image as the input to the third stage process extracting the road sign image.

In the recognition phase, the MP filters, which are obtained in the training process, are used to provide the most effective representation of the road sign. They are composed of the modulated Gabor functions, whose locations and variances along the time and frequency axes are determined by the parameters (s_n, u_n, ξ_n). The MP filters are constructed for each road sign, and each coefficient vector has eight components. We show the results of recognizing the same road sign images captured under the different illumination conditions, such as daytime, twilight, or at night.

Finally, we test the performance of our overall system. Fig. 9 shows the road sign images extracted from the closer view images. From Figs. 10–13, we illustrate the results of our system recognizing the same road sign captured in different viewing directions and lightening conditions. It shows that our methods recognize the road signs effectively. From the last sub-figure of Figs. 10–13, we can find that the angle between the coefficient vector of the test image and the cluster vector of the fourth road sign (Λ^4) is the smallest. It indicates that the input image is identified as the fourth road sign (see Fig. 8). We may also find that the input image is much more similar to the first eleven triangular road signs (see Fig. 8) than the other 19 triangular road signs, and their angles are much smaller.

The detection phase is tested on 356 individual triangular toad signs and 125 individual circular road signs. Table 1 shows the percentage of road signs which are correctly detected. In the training process, we use 200 individual road sign images, each class consist of five training road sign images. In the testing process, we test our system using 331 triangular road signs and 118 circular road signs. Table 2 shows the percentage of the road signs correctly recognized. Table 3 shows the performance of our overall system.

Our algorithm is based on the results of color segmentation method followed by the template matching. The

Table 2
Recognition phase

Recognition phase	Total testing road signs	Correct recognition (%)	False recognition
Triangular road signs	331 individuals	94	19
Circular road signs	118 individuals	91	10

Table 3
Overall system

Overall system	Total testing road signs	Correct recognition (%)	False recognition
Triangular road signs	356 individuals	88	44
Circular road signs	125 individuals	86	17

detection algorithm may fail when the background of the input image has too much red color information or the red component of the road sign is not obvious. For twilight or night scene, the road sign may not be detected, even if it is identified, the road sign images may not be accurately extracted, and that makes the recognition process fail. For some similar road signs (i.e. the 13th, 14th, and 15th road signs shown in Fig. 8), the corresponding MP filters basis set may not be well separated in the training phase. As the varieties of the road signs increase, the clusters of coefficient vectors of these road sign images may overlap, and the recognition algorithm may miss-identify the road sign based on the similarity measurement (i.e. Eq. (15)). One way to increase the recognition rate is increasing the number of basis elements and the dimension of coefficient vector. The other method is using the Bayesian approach [17]. Both of them require more complex computation in both the training and recognition phases.

5. Conclusion

In this paper, we propose a road sign detection and recognition system. In the detection phase, we have used color features effectively to detect the road signs under noisy and complex environment. In the recognition phase, we use the MP filter to recognize the road signs effectively. The purpose of using the MP filters is to produce the image representation. Our approach can be applied for the development of an automatic pilot system.

References

[1] E.D. Dickmanns, B.D. Mysliwetz, Recursive 3-D road and relative ego-state recognition, IEEE Transactions on PAMI 14 (1992) 199–213.

[2] D. Koller, N. Heonze, H.-H. Nagel, Algorithmic characterization of vehicle trajectory from image sequences by motion verbs, IEEE CVPR, Lahaina, Hawaii, 1991, pp. 90–95.

[3] K. Sakurai, H. Oka, Y. Usshioda, S. Ozawa, Analysis of a road image as seen from a vehicle, ICCV (1987) 651–656.

[4] B. Besserer, S. Estable, B. Ulmer, Multiple knowledge sources and evidential reasoning for shape recognition, Proceedings of the Fourth ICCV, Berlin, Germany, 1993, pp. 624–631.

[5] M. De Saint Blancard, Road sign recognition: a study of vision-based decision making for road environment recognition, in: I. Masaki (Ed.), Vision-based Vehicle Guidance, Springer, Berlin, 1992, pp. 162–172.

[6] G. Piccioli, E. DeMicheli, P. Parodi, M. Campani, Robust method for road sign detection and recognition, Image and Vision Computing 14 (1996) 209–223.

[7] R. Janssen, W. Ritter, F. Stein, S, Ott, Hybrid approach for traffic sign recognition, Proceedings of the Intelligent Vehicle Symposium, Tokyo, Japan, 1993, pp. 390–395.

[8] L. Priese, V. Rehrmann, R. Schian, R. Lakmann, Traffic sign recognition based on color image evaluation, Proceedings of the Intelligent Vehicles Symposium, Tokyo, Japan, 1993, pp. 95–100.

[9] B. Mertsching, H. Austermeier, U. Buker, G. Hartmann, Interpretation of traffic scenes using a hierarchical data structure, Proceedings of the Intelligent Vehicle Symposium, 1993, pp. 163–168.

[10] S. Masciangelo, 3-D cues from a single view: detection of elliptical arcs and model based perspective back-projection, Proceedings of the British Machine Vision Conference'90, Oxford, UK, 1990, pp. 223–228.

[11] W. Ritter, F. Stein, R. Janssen, Traffic sign recognition using color information, Mathematics and Computational Modeling 22 (4–7) (1995) 149–151.

[12] A. de la Escalera, L. Moreno, M.A. Salichs, J. Armingol, Road traffic sign detection and classification, IEEE Transactions on Industrial Electronics 44 (6) (1997).

[13] N. Yabuki, Y. Matsuda, Y. Fukui, S. Miki, Region detection using color similarity, Proceedings of IEEE ISCAS'99, 4, Monterey, USA, 1999, pp. 98–101.

[14] D.M. Gavrila, Multi-feature hierarchical template matching using distance transform, Proceedings of the 14th ICPR, vol. 1, Brisbane, Australia, 1998, pp. 439–444.

[15] S. Mallat, Z. Zhang, Matching pursuit with time-frequency dictionaries, IEEE Transactions on Signal Processing 41 (1993) 3397–3415.

[16] P.J. Phillips, Matching pursuit filers applied to face identification, IEEE Transactions on IP 7 (8) (1998).

[17] R.M. Haralick, L.G. Shapiro, Computer and Robot Vision, vols. 1 and 2, Addison-Wesley, Reading, MA, 1992.