1. Introduction

In the discipline of information science and technology, as one of the information systems frontiers, computer vision enables computers to interpret the content of pictures captured by cameras (Turban et al. 2005). A considerable interest in computer vision has been witnessed in past decades (Li et al. 2007; Xu 1999, 2006; Zhou et al. 2003, 2007). Examples of applications of computer vision systems include systems controlling processes, event detection, information organizing, objects modeling, etc (Bennett et al. 2008; Juang and Chen 2008; Kerner et al. 2004; Wei et al. 2006). Object recognition is one of sub-domains of computer vision. Accurate moving object recognition from a video sequence is an interesting topic with applications in various areas such as video monitoring, intelligent highway, intrusion surveillance, airport safety, etc (Hsu and Wallace 2007). In recent years moving object recognition has been considered as one of the most interesting and challenging areas in computer vision and pattern recognition.

In last few years, researchers have proposed a variety of methodologies for moving object detection and classification, most of them are based on shape and motion features. For example, an end-to-end method for extracting moving targets from a real-time video stream has been presented by Lipton et al. (1998). This method is applicable to human and vehicle classification with shapes that are remarkably different. Synergies between the recognition and tracking processes for autonomous vehicle driving have also been studied (Foresti et al. 1999). The attention on object recognition has been focused on specific parts of the visual signal and assigning them with symbolic meanings. Vehicles are modeled as rectangular patches with certain dynamic behavior (Gupte et al. 2002). The proposed method is based on the establishment of correspondences between regions and vehicles as the vehicles move through the image sequence. An object classification approach that uses parameterized 3-D models has been described by Koller et al. (1993). The system uses a 3-D polyhedral model to classify vehicles in a traffic sequence. Petrovic and Cootes (2004) extract gradient features from reference patches in images of car fronts and recognition is performed in two stages. Gradient-based feature vectors are used to produce a ranked list of possible candidate classes. The result is then refined by using a novel match refinement algorithm. There are many other moving objects classification methods based on multi-feature fusion (Lin and Bhanu 2005; Sullivan et al. 1997; Surendra et al. 2006; Takano et al. 1998; Zanin et al. 2003).

In our captured traffic scenes, moving objects are typically vehicles or pedestrians which can be mainly divided into four categories as trucks, cars, motorcycles, and pedestrians. The
traffic conditions in certain areas may be quite complex and mixed traffic often appears. As a result, a traffic flow detection system may mistakenly recognize a pedestrian for a vehicle. In general, it is considered meaningful to categorize moving objects in realtime from a camera video stream in the intersection. It is expected that the difference of the area, the shape and the velocity can be distinguished among moving objects. However, under certain circumstances, due to possible occlusions, the characteristics of the objects can be more complex. Lighting, weather conditions, occlusions, shadows, camera calibration, and the importance level of the recognition are factors to be taken into consideration.

Artificial Neural Networks and Choquet (fuzzy) Integral (CI) have been used to solve recognition problems in recent years (Li et al. 2003a, b; Sok et al. 1994; Zhu et al. 2008; Zhou and Xu 1999, 2001). Some researchers have integrated the results of several neural network classifiers and/or fuzzy integral to obtain higher quality recognition (Cho and Kim 1995). We use a high-order neural network (HNN) based on Biomimetic Pattern Recognition (BPR) to model the input data and then extract features from the model. BPR was first proposed by Wang (2002). This new model of pattern recognition is based on “matter cognition” instead of “matter classification”; therefore, BPR is rather closer to the functions that human used to than traditional statistical pattern recognition using “optimal separating” as its main principle. The method used by BPR is called High-Dimensional Space Complex Geometrical Body Cover Recognition Method, which studies types of samples’ distribution in terms of feature space, thus samples can be “recognized”. BPR has been used in many fields such as in rigid object recognition, multi-camera face identification, and speech recognition, and the results have shown its superiority (Wang et al. 2003, 2005, 2006). Feature spaces have been studied in existing literature (Li and Xu 2001; Li et al. 2003a, b).

This paper mainly focuses on a recognition and classification method for multiple moving objects on a real-time basis. In performing classification tasks, observations from different sources are combined. Firstly, moments, area, and velocity of an object are extracted through classic background subtraction techniques. Secondly, BPR is used to classify the invariants obtained at the first step. Finally, CI is adopted for multi-features fusion purpose based on the area and velocity of the object extracted during the first step and the moments classified at the second step all together. A multi-stage classification procedure is realized thus the accuracies can be further improved. An experiment has been conducted for a mixed traffic intersection. Experimental results indicate that the learning ability and the accuracy of the proposed method are satisfactory.

The rest of the paper is organized as follows. Moving object detection and feature extraction are presented in Section 2. A multi-stage recognition model is discussed in Section 3. The experimental results are presented and discussed in Section 4. Finally, in Section 5, a conclusion is provided.

2. Moving objects detection and feature extraction

Detection of moving objects from image streams is one of our main concerns in this study. Although numerous studies have been conducted, many problems remain outstanding such as the changes of appearance caused by motion of an object or camera, occlusion of a target, and overlapping of objects. In this study, velocity, area and invariant features are used as features for classification purpose.
2.1 Moving objects detection
Segmentation of moving objects in traffic scenes requires the background estimate to evolve over the weather and time as lighting conditions change. We address the problem of moving object segmentation using background subtraction.

Optical flow is a powerful image processing tool for measuring motion in digital images. The optical flow algorithm provides an estimate of the velocity vector at every pixel from a pair of successive images. The traffic background is estimated through optical flow methods (Horn and Schunck 1981; Ji et al. 2005, 2006), and every frame (image) is analyzed to segment the moving objects from background, and a fusion algorithm is used that is based on background segmentation flow field and edge extracted by Canny’s operator in the image sequences acquired by a fixed camera (Canny 1986).

2.2 Complex Zernike moments
Moments and functions of moments have been utilized as pattern features in a number of applications to achieve invariant recognition of two-dimensional image patterns. One advantage of complex Zernike moments is the ability to easily reconstruct the original signal from the moment values (Teague 1980). We first extract the area, shape and velocity of the moving object. Then we may take the first four order Zernike moments as shape parameters as the input of the HNN classifier based on BPR.

Complex Zernike moments are constructed using a set of complex polynomials which form a complete orthogonal basis set defined on the unit disc $x^2 + y^2 \leq 1$. They are expressed as $A_{pq}$. Two dimensional Zernike moment is,

$$A_{mn} = \frac{m + 1}{\pi} \int x \int y f(x,y)[V_{mn}(x,y)]^*dxdy$$  \hspace{1cm} (1)

where $x^2 + y^2 \leq 1$, $m = 0, 1, 2, \ldots, \infty$, $f(x,y)$ is the function being described and $*$ denotes the complex conjugate. $n$ is an integer (that can be positive or negative) depicting the angular dependence or rotation, subject to the conditions:

$$m - |n| = \text{even}, |n| \leq m$$  \hspace{1cm} (2)

and $A_{mn}^* = A_{m,-n}$ is true. The Zernike polynomial $V_{mn}(x,y)$ (Wang and Lai 2005) expressed in polar coordinates are

$$V_{mn}(r,\theta) = R_{mn}(r) \exp(jn\theta)$$  \hspace{1cm} (3)

where $(r,\theta)$ are defined over the unit disc and $R_{mn}(r)$ is the orthogonal radial polynomial, defined as

$$R_{mn}(r) = \sum_{s=0}^{m-|n|} (-1)^s \frac{(m-s)!}{s!(\frac{m+|n|}{2} - s)!((m-|n|)/2 - s)!} r^{m-2s}$$  \hspace{1cm} (4)

To calculate the Zernike moments of an image $f(x,y)$, the image (or region of interest) is firstly mapped to the unit disc using polar coordinates, where the center of the image is the origin of the unit disc. Those pixels falling outside the unit disc are not used in calculation. Translation and scale invariance can be obtained by shifting and scaling the image prior to the computation of the Zernike moments. The first-order regular moments can be used to find the image center and the 0th order central moment gives a size estimate.
2.3 Extraction of area features

The area feature can be obtained after each moving object is segmented. For region $R$, we assume that the area of each pixel is 1, the area of $A$ can be easily obtained. Since the scene to be monitored may be far away from the camera, assuming the world coordinate system and the image coordinate system is the same and the coordinate of the target point is $\{a, \beta, \gamma\}$, then the projection on the image plane is

$$
\alpha_1 = \frac{f\alpha}{\gamma}, \beta_1 = \frac{f\beta}{\gamma}
$$

Generally, a truck has the largest three-dimensional size, and its projected area is the largest also. A car is smaller than a truck, and its projected area is next large. Motorcycles and pedestrians projected area are the smallest ones. However, the above assumption is not always correct since the projected area of an object is not only related to the object's actual area, but also the distance between the object and the camera lens. Therefore, objects of different categories may have similar project areas. Other recognition features besides area and shape must be taken into consideration.

2.4 Extraction of velocity features

The velocity of a moving object is denoted by the velocity in images and it can be obtained by classical block-matching algorithm. The match template is obtained during the moving object extraction process.

In general, the velocities in images are in proportional to the real velocities; therefore, we can differentiate different moving objects according to the velocities in images. The velocity of a car or a motorcycle is faster than the pedestrians. The speed range of various moving objects is different. Thus we can estimate the classes of the moving objects with velocities exceed certain values.

However, as an object moves along the camera optical axis, or in the intersections, or in jams, the above conclusion may not hold. As a result, multiple features are required to recognize a moving object.

3. Multi-stage object classifier

After detecting regions and extracting their features such as area, shape, and velocity, the next step is to determine whether or not the detected region is a vehicle. In order to obtain a reliable conclusion, information from several sources is to be integrated. A multi-stage object classifier system for classifying the detected objects has been developed. We first perform a quick classification by BPR using the first forth order Zernike moments obtained as input. CI is then used to perform a second classification of the results of BPR and the extracted results of velocity and area features.

3.1 Classification based on BPR

BPR intends to find the optimal coverage of the samples of the same type. An HNN which covers the high dimensional geometrical distribution of the sample set in the feature space based on BPR is used as a fast classifier (Wang et al. 2005; Wang and Lai 2005). The input layer has nine neurons ($A = \{A_{00}, A_{11}, A_{20}, A_{22}, A_{31}, A_{33}, A_{40}, A_{42}, A_{44}\}$) and the output linear layer has four neurons.

A moving scene can be seen as made up of regions with different motion parameters. We assume that each frame can be segmented into $L$ subsets, forming moving regions, denoted
The moving objects are considered as compact moving entities, consisting of one or more moving regions. Each moving object is assigned to a class. Each subset $Z_k$ is associated with a nine-dimensional representative vector $\mu_k$, describing the Zernike moment information of a certain moving region.

The architecture of a three-layer HNN is shown in Fig. 1. Each of these input units is fully connected with K hidden layer units. Again, all the hidden layer units are also fully connected with the four units of the output layer. Let $\{Y_1, Y_2, Y_3, Y_4\}$ denote the object recognition category, which represents trucks, cars, motorcycles, and pedestrians, respectively. The number of hidden layer units is determined experimentally. For an input vector, the output of j-th output node produced by an HNN is given by

$$\Phi(W, X) = \sum_{i=1}^{n} \left( \frac{W_{ij}}{|W_{ij}|} \right)^S |W_{ij}(X_i - W'_{ij})|^P - \theta$$

where $W_{ij}$ denotes the direction weight which connects the j-th input and the neuron, and it determines the direction of the neuron. $W'_{ij}$ is the core weight which connects the j-th input and the neuron, it can determine the center position of the neuron; $x_i$ is the j-th input. $\theta$ is the threshold, $S$ and $P$ are the power exponents. As parameter $S$ and $P$ change, the hyper-surface of the high-order neuron changes accordingly. If $W'_{ij} = 0$, $S = 1$, $P = 1$, in Eq. 6 holds, Eq. 6 is transformed into a classic neural network. If $W_{ij} = 1$, $S = 0$, $P = 2$ in Eq. 6 holds, Eq. 6 is transformed into a radial basis function neural network. Here three weight neurons are used to construct the high dimensional space geometry region.

The network is trained with HNN and 840 samples collected under various weather conditions used as training set. According to the principle of BPR, determining the subspace of a certain type of samples is based on the type of samples itself. If we are able to locate a set of multiweights neurons that covering all training samples, the subspace of the neural networks will represent the sample subspace. When an unknown sample is in the subspace, it will be determined if it belongs to the same type of training samples. Moreover, if a new type of samples is added, it is not necessary to retrain any trained types of samples. The training of a certain type of samples has nothing to do with the other ones.

---

Fig. 1. A three layer high-order neural network
4. Recognition based on CI

After the first stage classification, CI is used as the second stage classifier. Fuzzy integral is an effective means to solve complicated pattern recognition and image segmentation (Xu 1988). Recently, since CI has been found useful in data fusion, it has been enjoying successes in image sequence analysis (Cho and Kim 1995; Li et al. 2002; Murofushi and Sugeno 1991; Tahani and Keller 1999). The differences of the shape features of different moving objects is not distinguished under certain circumstances as occlusions happen, the recognition accuracy may not satisfy the requirements. Therefore, combining the area and the velocity to perform further fusion by CI is intended. With this approach, the information sources are given grades of compatibility, and the evidence is weighted and combined accordingly. From three features including area, shape, and velocity, we compute the CI. The definition is as follows,

Fuzzy measure over \( T \) \( \rightarrow \) \( 0, 1 \) is defined as follows:

\[
g^1(X_1) = g^1
\]

\[
g(X_i) = g(X_{i-1}) + g^i + \lambda g(X_{i-1})g^i
\]

CI (Murofushi and Sugeno 1991) \( C^g(h) \) is defined as follows:

\[
C^g(h) = \int h(x) \circ g = \int_0^1 g(h_\theta)d_\theta
\]

where \( h_\theta = \{ x : h(x) \geq \theta \} \).

For \( h(x_1) \leq h(x_2) \leq \cdots \leq h(x_n) \),

\[
C^g(h) = \sum_{i=1}^{n} g(X_i)[h(x_i) - h(x_{i+1})] = \sum_{i=1}^{n} h(x_i)[g(X_i) - g(X_{i-1})]
\]

where \( h(x_{n+1}) = 0, g(x_0) = 0 \)

Let \( T = \{ t_1, t_2, t_3, t_4 \} \) be the object recognition category, which represents trucks, cars, motorcycles and pedestrians, respectively. \( A \) is the object to be recognized, and \( X = \{ x_1, x_2, x_3 \} \) be a set of elements, which represents the recognition result of BPR, the area and the velocity of the object, respectively. Let \( h_k : X \rightarrow [0, 1] \) denote the confidence value of the target \( A \) belonging to class \( t_k \) The flow chart of target recognition algorithm based on CI is shown in Fig. 2

![Flow chart of target recognition algorithm based on CI](image)

Fig. 2. Flow chart of target recognition algorithm based on CI
In our approach, we let \( g_1^k = 0.7, g_2^k = 0.3, g_3^k = 0.3, 1 \leq k \leq 4 \) (Cho and Kim 1995). We first treat each feature (property) as fuzzy variables and then assign fuzzy density to all possible values. \( h_k(x_i) \) represents the degree. \( A \) belongs to class \( t_k \) according to \( x_i \) Figure 2 shows the flow chart of target recognition algorithm based on CI, and Fig. 3 shows the curves of the functions for confidence obtained by experiments.

5. Experiment

The videos used in our work are from the website (http://i21www.ira.uka.de) maintained by KOGS-IAKS Universitaet Karlsruhe. Traffic sequence showing the intersection Karl-Wilhelm-/Berthold-Straβe in Karlsruhe that is recorded by a stationary camera under various weather conditions. In Fig. 4, we give a typical example of what we might obtain by moving object detection using the above method. After analyzing the traffic image such as shown in Fig. 4a, the current background is obtained as Fig. 4b shows. Figure 4c shows the detected moving object region, and Fig. 4d shows the foreground objects. We extract the moving objects and obtain the first fourthorder moments and set up the training image set by the algorithm proposed in Section 2.1. Four groups of experiments are conducted for normal conditions, heavy fog, heavy snowfall, and snow on lanes, respectively. Under each condition, we select 60 traffic frames to train the HNN, and then classify the other 30 moving objects. The results show that there exists misclassification between cars and trucks, and pedestrians and motorcycles. If we combine the first classification results by BPR, the area and the velocity as the input of CI classifier, the results can be improved. The comparisons are shown in Table 1 and Table 2. The proposed algorithm was able to classify trucks, cars, motorcycles and pedestrians. It can be concluded that the approach proposed in this paper has better recognition ability.
Fig. 4. a A sample frame in video sequence in normal condition; b current background; c detected moving object region by the background subtraction; d foreground

<table>
<thead>
<tr>
<th>Recognition result (%)</th>
<th>Truck</th>
<th>Car</th>
<th>Motorcycle</th>
<th>Pedestrian</th>
<th>Average</th>
</tr>
</thead>
<tbody>
<tr>
<td>High-order Neural network</td>
<td>90.2</td>
<td>91.8</td>
<td>88.4</td>
<td>82</td>
<td>88.6</td>
</tr>
<tr>
<td>Combined with CI</td>
<td>94.3</td>
<td>94.1</td>
<td>92.9</td>
<td>92.8</td>
<td>93.5</td>
</tr>
</tbody>
</table>

Table 1. Recognition results in normal conditions

In this paper, the image size is $768 \times 576$ and we have conducted our experiments on a personal computer with an INTEL Celeron 2.4G CPU using Microsoft Visual c++ 6.0. The computation time per frame is around 0.1 s, depending on the image quality and the number of moving objects. For most applications, this can be considered as real-time.

<table>
<thead>
<tr>
<th>Recognition result (%)</th>
<th>Truck</th>
<th>Car</th>
<th>Motorcycle</th>
<th>Pedestrian</th>
<th>Average</th>
</tr>
</thead>
<tbody>
<tr>
<td>High-order Neural network</td>
<td>90.8</td>
<td>91.2</td>
<td>88.5</td>
<td>82.3</td>
<td>88.7</td>
</tr>
<tr>
<td>Combined with CI</td>
<td>93.7</td>
<td>94.2</td>
<td>92.5</td>
<td>91.3</td>
<td>92.9</td>
</tr>
</tbody>
</table>

Table 2. Recognition results in normal conditions

6. Conclusion

In this paper, a multi-stage moving objects recognition approach is presented and applied to mix traffic environment. Area, velocity and invariant feature of the moving object were extracted firstly and BPR and CI techniques have been integrated to perform the multi-stage classification. The experimental results show that the proposed approach is effective.

7. References


Vision-based object recognition tasks are very familiar in our everyday activities, such as driving our car in the correct lane. We do these tasks effortlessly in real-time. In the last decades, with the advancement of computer technology, researchers and application developers are trying to mimic the human’s capability of visually recognising. Such capability will allow machine to free human from boring or dangerous jobs.

How to reference
In order to correctly reference this scholarly work, feel free to copy and paste the following: