http://ansinet.com/itj

ITJ

ISSN 1812-5638

INFORMATION TECHNOLOGY JOURNAL

ANSInet
Asian Network for Scientific Information
308 Lasani Town, Sargodha Road, Faisalabad - Pakistan

Information Technology Journal 8 (8): 1189-1196, 2009
ISSN 1812-5638

A View-Based Approach to Three Dimensional Object Recognition

Xu Sheng and Peng Qi-Cong
University of Electronic Science and Technology of China,
611731, Chengdu, Sichuan, China

Abstract: To improve the performance of three-dimensional object recognition systems, we propose a view-based method in this study. First we extract wavelet moments, texture features and color moments from the 2D view images of 3D objects. Wavelet moments have the multi-resolution properties in addition to the invariant properties under translation, scaling and rotation. Texture features can distinguish objects which have similar shapes and different appearance. Color moments are robust and insensitive to the size and pose of objects. Support Vector Machine (SVM) is chosen as classifier. Then the feature subset selection and SVM parameters optimization are accomplished automatically and simultaneously using Genetic Algorithm (GA) in an evolutionary way. We assessed our method based on the original and noise corrupted 3D object dataset COIL-100. One hundred percent correct rate of recognition was obtained when the number of presented training views for each object was 36 (10 degrees interval) and 18 (20 degrees interval). When the number of training views was reduced, the correct rate of recognition was also satisfied.

Key words: 3D Object Recognition, wavelet moments, GLCM, color moments, genetic algorithm, support vector machine

INTRODUCTION

In recent years, the view-based (or appearance-based) three-dimensional object recognition (3DOR) has attracted much attention and been widely researched (Bicego *et al*., 2005; Delponte *et al*., 2007; Haizhai *et al*., 2007; Choi *et al*., 2008). In a two-dimensional image, the appearance of a three-dimensional object depends on its shape, reflectance properties, pose and the illumination conditions in the scene. Among the view-based 3DOR approaches, Murase and Nayar (1995) proposed a parametric eigenspace method to recognize 3D object directly from their appearance. They developed a near real-time recognition system to recognize complex objects and got accurate recognition results. Roth *et al*. (2002) proposed a view-based algorithm using a network of linear units, the Sparse Network of Winnows (SNoW) learning architecture. The SNoW method tries to learn the hyperplane using either intensity data or edge information from 32×32 gray-scale images. Nodes in the input layer of the network typically represent relations over the input instance and are being used as the input features. Each linear unit is called a target node and represents a concept of interest over the input. Target nodes could represent an object in terms features extracted from the 2D image input. Early in the 1990s, a new learning algorithm based on statistical learning theory was proposed by Vapnik and coworkers. This algorithm, named Support Vector machine (SVM) (Vapnik, 1995), has produced excellent performances in a number of difficult learning tasks and is also used in 3DOR problems recently. Haizhai *et al*. (2007) decomposed 2D images to sub-images using wavelet transformation, then extracted feature vectors from each sub-images using Singular Value Decomposition (SVD), finally these feature vectors were sent to SVM for classification and got good results (Haizhai *et al*., 2007).

In a view-based recognition system, typically a set of numerical features are extracted from an image rather than using the pixels directly. The most common reason is that feature extraction is used to reduce the dimension of the input data and in turn helps to minimize the training time needed by the classifier. Generally the dimension of feature space is as high as hundreds (Roth *et al*., 2002; Haizhai *et al*., 2007). But in fact only those features that show the main differences among similar views are useful. The selection of discriminative features is a crucial step in the process, since the next stage sees only these features and acts upon them. The filter approach and the wrapper approach are two known general approaches (Frohlich *et al*., 2003; Zexuan *et al*., 2007). The filter methods perform feature selection as a preprocessing step to the following learning algorithm. The criterion for feature selection is independent of the actual generalization performance of the classifier. On the other

Corresponding Author: Xu Sheng, University of Electronic Science and Technology of China, 611731, Chengdu, Sichuan, China Tel: 008613084414490

hand, wrapper methods take into account each feature subset and the learning algorithm at the same time to return the estimated generalization performance which is used to build the classifier. Generally they outperform filter methods in terms of prediction accuracy but more computationally intensive.

In order to solve the feature selection problem, we need to make use of global search and optimization techniques and these techniques should be able to deal with solution spaces which have constrained parameters and a large number of local extrema. In this respect, Genetic Algorithm (GA) imitates the long-term optimization process of biological evolution for solving mathematical optimization problems so that offers a natural way to solve this feature selection problem at hand. And GA is easy to implement and provide good results in terms of selected features and the overall performance of the classifier (Huang and Wang, 2006).

In this study we introduce a view-based 3D object recognition approach. First we extract multiple discriminating features from each 2D view of 3D objects, including wavelet moments, texture characteristics and color moments. Then we propose to use GA to select the optimum set of features and SVM model in an evolutionary way simultaneously for our 3DOR task. This is a wrapper method since an optimal feature subset is dependent on the generalization performance of SVM. The proposed method has been tested on the Columbia Object Image Library (COIL-100) dataset (Murase and Nayar, 1995) and distinguished a large number of 3D objects with similar shapes, colors, texture and poses.

FEATURE EXTRACTION

The purpose of feature extraction technique in image processing is to represent the image in its compact and unique vector form. Wavelet moments have multi-resolution properties and the invariant properties to the translation, scaling and rotation. Texture and color are the intrinsic properties of object. So, we extract wavelet moments, texture and color moments and combine them as features of 3D objects.

Wavelet moments: It is well known that wavelet transform is different from the traditional Fourier transform, because it can provide the local information in time domain and frequency domain at the same time. This property is especially suitable for extracting the feature of local difference. Shen and Ip (1999) proposed a set of wavelet moments which are rotation and scale invariants.

The rotation invariant moments are expressed as follows:

$$F_{pq} = \iint f(r,\theta) g_p(r) r dr d\theta \tag{1}$$

where, F_{pq} is the pq-order moment, $g_p(r)$ is a function of radial variable r, p and q are integer parameters. It is easy to prove that the value of $\|F_{pq}\|$ is rotation invariant. In order to reduce the problem of feature extraction from a 2D image object to extraction from a 1D sequence, expression (1) is rewritten as follows:

$$\begin{aligned} F_{pq} &= \int S_q(r) \cdot g_p(r) r dr \\ S_q(r) &= \int f(r,\theta) e^{jq\theta} d\theta \end{aligned} \tag{2}$$

Then $S_q(r)$ is now a 1D sequence of variable r. It is important to note from (2) that if $g_p(r)$ is defined on the whole domain of variable r and then F_{pq} is a global feature. On the other hand, if the function of $g_p(r)$ is locally defined and then F_{pq} can be used as a local feature.

Replace the basis functions $\{g_p(r)\}$ by wavelet basis functions:

$$\psi^{a,b}(r) = \frac{1}{\sqrt{a}} \psi\left(\frac{r-b}{a}\right)$$

where, a ($a \in R^+$) is a dilation parameter and b ($b \in R$) is a shifting parameter. In practical applications, the values of parameters *a* and *b* are usually discrete. The discretization of the dilation parameter is done by choosing $a = a_0^m$, where *m* is an integer and $a_0 \neq 0$. b is discretized by taking the integer (positive and negative) multiples of $b_0 a_0^m$, where, $b_0 > 0$ is appropriately chosen, so that $\Psi((r-b)/a)$ covers the whole domain at different values of m. Then the wavelet defined along a radial axis in any orientation is denoted by:

$$\psi_{m,n}(r) = 2^{m/2} \psi(2^m r - 0.5n) \tag{3}$$

Now wavelet moment invariants are defined as follows:

$$\left\|F_{m,n,q}\right\| = \left\|\int S_q(r) \psi_{m,n}(r) r dr\right\| \tag{4}$$

where, $\Psi_{m,n}(r)$ is the mother wavelet and m = 0, 1, 2, 3; n = 0, 1,..., 2^{m+1}, q = 0, 1, 2, 3.

In this study, cubic B-spline function is adopted as the mother wavelet because it has the optimal local properties in spatial frequency domain and the multi-resolution characters of wavelet transform. The mother wavelet $\Psi(r)$ of the cubic B-spline in Gaussian approximation form is:

$$\psi(r) = \frac{4a^{n+1}}{\sqrt{2\pi(n+1)}}\sigma_w \cos(2\pi f_0(2r-1))\exp\left(-\frac{(2r-1)^2}{2\sigma_w{}^2(n+1)}\right)$$

where, $n = 3$, $a = 0.697066$, $f_0 = 0.409177$ and $\sigma^2_w = 0.561145$.

GLCM based texture analysis: Texture analysis is used in a variety of applications, including remote sensing, automated inspection and medical image processing. Texture is an important feature of objects in an image. When traditional threshold technical can not be used effectively, texture analysis can be helpful when objects in an image are more characterized by their texture than by intensity. In this study we propose to utilize texture analysis method in our 3DOR task.

Gray Level Co-occurrence Matrix (GLCM) (Haralick *et al*., 1973) is one of the effective texture analysis methods, which estimates image properties using second-order statistics. Gray level co-occurrence matrix $P_d(i, j)$ is constructed with each element (i, j) equals to the number of occurrence of gray level i and j pair which are a distance d apart in original images. Haralick proposed 14 statistical features extracted from GLCM to estimate the similarity of GLCM in different distance d and different occurrence of gray level i and j pair. To reduce computational complexity, we only use 4 statistical features for recognition as follows:

- **Correlation:** A measure of how correlated a pixel to its neighbor over the whole image. Here u_x, u_y, σ_x, σ_y are the average and standard variance of P_x, P_y separately. P_x is the sum of each row of $P_d(i, j)$ and P_y is the sum of each column of $P_d(i, j)$

$$\text{Correlation} = \frac{1}{\sigma_x\sigma_y}\sum_{i,j}(i-u_x)(j-u_y)P_d(i,j)$$

- **Energy:** Also known as Angular Second Moment, provides the sum of squared elements in the GLCM

$$\text{Energy} = \sum_{i,j}P_d^2(i,j)$$

- **Contrast:** A measure of the intensity contrast between a pixel and its neighbor over the whole image

$$\text{Contrast} = \sum_{i,j}(i-j)^2P_d(i,j)$$

- **Homogeneity:** Measure the closeness of the distribution of the elements in GLCM to the diagonal

$$\text{Homogeneity} = \sum_{i,j}P_d(i,j)/(1+|i-j|)$$

Color moments: The color of objects is quite robust and insensitive to size and orientation of objects. In this study, we use the method proposed originally by Stricker and Orergo (1995) to store the first three moments of each color channel of an image. For an image of RGB format or HSI format, only 9 moment values are required.

A probability distribution is uniquely characterized by its moments according to probability theory. Color distribution of an image also can be regarded as a probability distribution, so color distribution also can be determined by its moments. The first moment, the second and the third central moment of each color channel can be used. The first moment is the average color of an image. And the second central moment of an image is the variance; the third central moment of an image is the skewness of each color channel. To make the value of the moments somewhat comparable, the standard deviation and the third root of the skewness of each color channel of an image are used, in this way all the values have the same unit. If p_{ij} is the pixel of a digital image f (x, y) of M×N dimension, A is the area of the image and then the first three moments for each HSI color channel can be defined as:

$$\mu = \frac{1}{A}\sum_i\sum_j p_{ij}, \sigma = \left[\frac{1}{A}\sum_i\sum_j(p_{ij}-\mu)^2\right]^{1/2}, s = \left[\frac{1}{A}\sum_i\sum_j(p_{ij}-\mu)^3\right]^{1/3}$$

SVM AND ITS PERFORMANCE

After extracted above-mentioned 3D object features from 2D views, this features based 3DOR task is a typical classification problem from the point of view of pattern recognition. In this section we will briefly describe the basic SVM concepts for typical two-class classification problem. These concepts can also be found by Vapnik (1995), Pontil *et al*. (1998), Pontil and Verri (1998) and Haizhai *et al*. (2007).

Consider the classification task given by a data set $S = \{(x_i, y_i), i = 1, 2, ..., l\}$ in two classes, where each instance x_i belongs to the input space R^N and $y_i \in \{-1, +1\}$. If the training set x are linear separable, there exist separating hyperplane $w \bullet x_i + b=0$ such that:

$$y_i(\mathbf{w}\cdot\boldsymbol{x}_i + b) \geq 1, i = 1, 2, ..., l \quad (5)$$

where, $w \in R^N$. From statistic learning theory, if all the training vectors are separated correctly by the hyperplane and the distance of the nearest sample data from the hyperplane is maximum, this hyperplane is named Optimal Separating Hyperplane (OSH), as shown in Fig. 1.

Data points which satisfy (5) with equality are called support vectors since they define the orientation of the

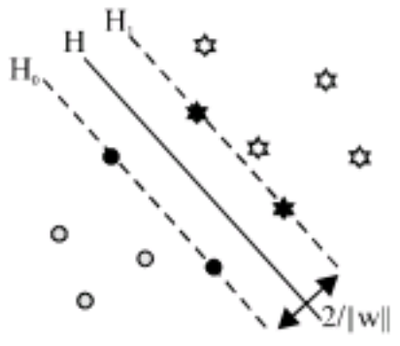

Fig. 1: OSH and Support Vectors (filled marks)

resulting OSH. We require the separating hyperplane to separate the data set in two classes correctly with maximum distance. Because the distance equals $2/\|w\|$, the maximum distance problem equals minimum $\|w\|^2/2$ and thus the solution to the OSH can be written as the quadratic programming problem:

$$\min_{w,b}\left(\frac{1}{2}\|w\|^2\right), \quad \text{s.t.: } y_i\ (w.x_i + b) \geq 1,\ i = 1, 2, ..., 1 \tag{6}$$

Non-linear separable problem: The 3DOR problem studied in this paper belongs to non-linear separable problem. When the training set x is non-linear separable, the training vectors can be mapped into a (usually high dimensional) linear feature space through a nonlinear mapping function $\phi(\cdot)$. In this high dimensional linear feature space, the OSH $w\cdot\phi(x_i)+b = 0$ is constructed and the OSH problem (6)is then regarded as Eq. 7:

$$\min_{w,b,\xi}\left(\frac{1}{2}\|w\|^2 + C\sum_{i=1}^{l}\xi_i\right), \quad \text{s.t.: } y_i(w\cdot\phi(x_i)+b) \geq 1-\xi_i, i = 1,2,...,l, \xi_i \geq 0 \tag{7}$$

where, ξ_i, i = 1, 2, …, *l* is non-negative slack variables and C is the penalty coefficients that allows one to trade off training error vs. model complexity. Instead of solving Eq. 7 directly, it is easier to solve its dual problem:

$$\min_{\alpha}\left(W(\alpha) = -\sum_{i=1}^{l}\alpha_i + \frac{1}{2}\sum_{i=1}^{l}\sum_{j=1}^{l}y_i y_j \alpha_i \alpha_j K(x_i, x_j)\right), \quad \text{s.t.: } \sum_{i=1}^{l}\alpha_i y_i = 0,\ 0 \leq \alpha_i \leq C \tag{8}$$

where, α_i denotes the Lagrange multipliers. From the solution α^* of Eq. 8, the decision function can be computed as:

$$y(x) = \text{sgn}\left\{\sum_{i=1}^{l}\alpha_i^* y_i K(x_i, x) + b^*\right\} \tag{9}$$

where, $K(x_i, x) = \phi(x_i)\phi(x)$ is the kernel function. Though the training set is mapped into high dimensional feature space by non-linear function$\phi(\cdot)$, this non-linear function need not be calculated explicitly. We only need to calculate the kernel function $K(x_i, x)$ in the initial feature space and thus avoid the curse of dimensionality problem of feature space. Different SVM can be designed if we select different inner product kernel function which meets Mercer condition. In this paper, we choose the Gaussian Radial Basis Function (RBF) with kernel width equals to σ: $K(xi, x) = \exp(-\|xi-x\|^2/2\sigma^2)$ since RBF kernel function can analysis higher-dimensional data and require only two parameters.

Performance of SVM: Traditionally k-fold cross-validation or Leave-One-Out is used as an estimator of the generalization error of SVM (Jun *et al.*, 2004; Huang and Wang, 2006). For most practical problems, these methods are computational intensive and not practical for SVM. In this 3DOR problem, we propose to use the theoretical bound on the generalization error for SVM instead of performing cross-validation. It can be computed at no extra cost immediately after training a single SVM instead of retrain k times on each feature subset. The goal of SVM is to minimize the expected generalization error (or risk) over all possible patterns drawn from the unknown distribution P(x, y). Since, we do not know P(x, y), the generalization performance can be estimated. Several theoretical bounds on the leave-one-out error exist for SVM, the better one mentioned in (Frohlich *et al.*, 2003) is cited here:

Let ρ be the size of the maximal margin and $\phi(x_1)$, ..., $\phi(x_n)$ the images of the training patterns in feature space which are lying within a sphere of radius R. Let α^* be the tuple of Lagrange multipliers which are obtained by maximizing function (8). If the images of the training data of size n belonging to a sphere of radius R are separable with the corresponding margin ρ, then the expectation of the test error probability P^{n-1} has the bound:

$$EP_{err}^{n-1} \leq \frac{1}{n}E\left\{\frac{R^2}{\rho^2}\right\} = \frac{1}{n}E\left\{R^2W(\alpha^*)\right\} \tag{10}$$

The term R^2 of bound (10) can be estimated by the expression:

$$\frac{1}{n}\sum_{i=1}^{n}k(x_i, x_i) - \frac{1}{n^2}\sum_{i,j=1}^{n}k(x_i, x_j)$$

GA BASED FEATURE SELECTION AND SVM MODEL OPTIMIZATION

Based upon Darwin's principle of the survival of the fittest, Genetic Algorithm (GA) is well known for its ability to efficiently search large space about which little is known (Huang and Wang, 2006). Problem solutions are abstract as individuals in a population. An

initial population of chromosomes is generated randomly. Then each individual is evaluated by a fitness function. A fitness function (known as the objective function in standard optimization algorithms) assesses the quality of a solution in the evaluation step. It has great impact on performance and guides the evolutionary learning process determining the probability that an individual can hand down genetic information to the subsequent population via evolutionary mechanisms. In this paper we propose that the fitness of each individual is evaluated by the bound (10), which can be computed no extra cost once after training a single SVM. To derive a new population from an old one at each generation, pairs of parents are selected. Each pair of parents gives rise to one offspring via a recombination process which consists of two separate processes: crossover and mutation. In addition, the best individual from the previous generation is copied over unchanged to the new population. This procedure is known as elitism. The evolutionary algorithm continues until a pre-specified number of generations have passed.

Feature selection: After getting the 3 types of features discussed above, we can combine them and obtain a feature set $F = \{f_1, f_2, ..., f_n\}$ of length N. The number of features to be selected in our 3DOR problem is not known beforehand. In this case we have a search space of size 2^N. We take the binary encoding to represent this search space, that is:

$$B = \{b_1, b_2, ..., b_N\}, \quad b_i \in \{0,1\}, \quad i = 1, 2, ..., N$$

where, each bit 1 or 0 in B represents whether the related location in F is selected. Let G be the performance of SVM, then the optimization of feature selection can be represented as:

$$\text{Problem 1: } \max_{s} G(B),$$
$$\text{s.t. } B = \{b_1, b_2, ..., b_N\}, b_i \in \{0,1\}, i = 1, ..., N$$

This can be regarded as a kind of wrapper methods, since candidate feature subsets is obtained by the estimating of generalization performance of classifier.

SVM model selection using GA: Before the training of SVM, the parameters of SVM including kernel width σ and penalty coefficients *C* should be assigned. This process is formally known as model selection for SVM and plays an important role to obtain accurate classification results for a given task. Similar to the selection of a feature subset, we can optimize kernel parameter and penalty coefficients of the SVM by means of GA. We can represent this optimization problem of SVM parameters as following Problem:

$$\text{Problem 2: } \max_{M} G(M),$$
$$\text{s.t.: } M = \{\sigma, C\}, \sigma, C \geq 0$$

Combination of Feature Subset and SVM Model Selection: In fact, when using SVM, we always confront aforementioned Problem 1 and 2, that is: 1) choose the optimal input feature subset for SVM; 2) choose the best parameters for SVM. Naturally, the two problems should be considered at the same time:

$$\text{Problem 3: } \max_{S,M} G(B,M),$$
$$\text{s.t.: } B = \{b_1, b_2, ..., b_N\}, b_i \in \{0,1\}, i = 1, ..., N;$$
$$M = \{\sigma, C\}, \sigma, C \geq 0$$

where (B, M) is the hybrid vector, that is we can select an optimal feature subset and an optimal SVM training model simultaneously. This strategy is reasonable, because the choices of SVM parameter C and σ are influenced by the feature subset taken into account and vice versa. In Problem 3, individuals become the hybrid vector (B, M), their fitness function G can be evaluated by the bound (10) in a more computational efficient way.

EXPERIMENTS

The Columbia Object Image Library (Murase and Nayar, 1995) (COIL-100) is a dataset consisting of 7,200 size normalized true color images of 128×128 resolution of 100 objects (72 images per object). The objects have a wide variety of geometric and reflectance characteristics. Each object was placed on a motorized turntable against a black background. The turntable was rotated through 360 degrees to vary object poses and acquire images at pose intervals of 5 degrees, i.e., 72 poses per object. The images of each object with view angel at 30° are shown in Fig. 2 and some binary views are shown in Fig. 3. In this study, given the use of COIL 3D objects dataset, it is assumed that the illumination conditions remains constant, thus we only consider object poses as variable. We evaluated the recognition rate of our method based on both original and noise corrupted COIL dataset. Our implementation was performed on Matlab2008b simulation environment and Intel Core2 CPU running at 1.83 GHz and 1 GB RAM.

Fig. 2: All objects in COIL-100 dataset with view angle = 30°

Fig. 3: The binary images of preprocessed Tank objects

Experiments based on original views: All of the 3D objects, totaled 7,200 images were used in our experiment. For learning of SVM, 36 views per object (10° intervals) were chose as the training set, amounted to 3600 training views and the remaining views have been used for testing. We extracted the wavelet moments, textures and color moments for the training view images according the methods mentioned in Section 2. To fully use the spatial information of images, we calculated the GLCM in 4 directions of horizontal, vertical and two diagonal of the training views. Then we use GA to select the appropriate feature subset and optimize the SVM model simultaneously.

In all experiments, the parameters setting for GA are: Population size: 25; crossover rate: 0.6; roulette wheel selection; single point crossover; an elitist strategy of passing two fittest population members to the next generations was used to guarantee that the fitness is never declined from one generation to the next. GA performed selection operator by using roulette wheel method to select two mating individuals and then performed single point crossover operator and mutation operation randomly for them to generate individuals of the next population. And the next population was considered as the current population in the next iteration.

Fig. 4: The views making up the training set for a Tank object

Finally, provided with the GA selected feature subset and optimized parameters, the SVM was trained. We used the one-against-one method in which 100*(100-1)/2 classifiers were constructed on each pair of class labels and voting method of maximum win strategy was used for the learned SVM to recognize the testing images of unseen views in COIL.

In the first experiment, the correct rate of recognition of the trained SVM achieved 100% for all the 3600 testing views of 100 objects in COIL, as shown in Table 1. This result indicated the good performance of our method in high-dimensional patterns recognition.

In order to evaluate our method in more realistic situation, we ran several experiments when fewer numbers of views of the 3D objects were presented during training. For each experiment the number of training views was step-wise reduced from 36 views per object (10° intervals) to 2 views per object (180° intervals). When the training views of one object were 4, we could only get information for recognition from front, back, left-side and right-side of the object as demonstrated in Fig. 4. The recognition performance was tested on the remaining views in the database (which were 36 to 70 views per object). We repeated our experiments by decreasing training views and reported the recognition results also in Table 1.

Under these more challenging experimental setups, although it is not surprising to see from Table 1 that the correct rate of recognition decreased as the number of available views decreased during training, it is worth noticing that when the number of training views per object were reduced to 18 (20° interval), our method also achieved 100% correct rate of recognition. At the most hardness setup, we reduced the number of training view to 2, the correct rate of recognition of our method achieved 88.50%. From Table 1, compared with the results obtained on COIL reported in previous literatures (Roth *et al*., 2002; Xiuwen and Srivastava, 2002; Marée *et al*., 2004; Vasconcelos *et al*., 2004), our method outperformed or was comparable to these results when fewer training views were used.

To inspect the performance of each kind of feature, we fixed the number of training views for each object to 18 and conducted another three experiments. From the results shown in Table 2, we could find the correct rate of recognition of wavelet moments was the worst, since

Table 1: Correct rate of recognition with varying view angles

	The No. of training views for each object				
Views	36	18	8	4	2
(Training/Testing) view number	3600/3600	1800/5400	800/6400	400/6800	200/7000
Original views					
Our method	100%	100%	99.19%	95.62%	88.50%
Sub-windows (Marée *et al*., 2004)	99.94%	99.61%	98.47%	95.06%	88%
Spectral (Xiuwen and Srivastava, 2002)	100%	99.50%	96.33%	N/A.	N/A.
K-L SVM (Vasconcelos *et al*., 2004)	99.78%	98.89%	95.22%	84.32%	N/A
SNoW / Edges (Roth *et al*., 2002)	96.25%	94.13%	89.23%	88.28%	N/A.
SNoW / Intensity (Roth *et al*., 2002)	95.81%	92.31%	85.13%	81.46%	N/A.
NNC (Roth *et al*., 2002)	98.50%	87.54%	79.52%	74.63%	N/A.
Corrupted views					
variance = 0.05	99.86%	99.20%	96.58%	88.75%	77.97%
variance = 0.1	99.42%	98.56%	94.64%	85.25%	77.09%

Table 2: Correct rate of recognition with varying feature types

Methods	Percentage	Training view No. for each object	Total testing view No.
Wavelet moments	96.02	18	5400
GLCM based texture features	97.89		
Color moments	98.26		

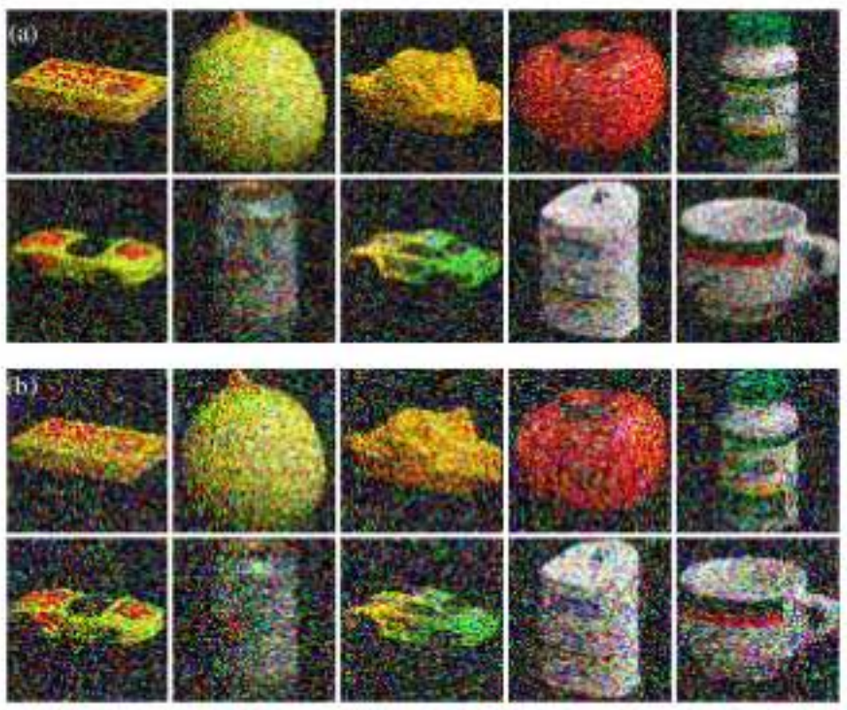

Fig. 5: First 10 noise corrupted objects with view angle = 30°. (a) Views added AWGN of mean = 0 and Variance = 0.05 and (b) Views added AWGN of mean = 0 and Variance = 0.01

wavelet moments only used the mathematic properties of training views. Texture features and color moments achieved better correct rate of recognition. But none of the features could achieve 100% correct rate alone.

Experiments based on noise corrupted views: In order to assess the robustness of our method under noise environment, Additive White Gaussian Noise (AWGN) of zero-mean with variance 0.05 and 0.1 respectively were added to the original images of COIL, showing in Fig. 5a and b. Due to the space constrains, here only the first 10 objects are shown. Under each noise variance condition, also we repeated our experiments using all the 100 objects by decreasing training views.

From the obtained results in the last two rows of Table 1, when noise corrupted training views were presented at 10° intervals, in the case of images were more seriously corrupted by AWGN of variance 0.1, our method achieved 99.42% correct rate of recognition, i.e., only 21 poses was not recognized in the total 3600 testing views. This result can be inferred that our method is robustness in the presence of large number of image noise.

CONCLUSIONS

This study introduced a full view-based 3D object recognition approach including discriminating feature extraction, feature selection, SVM model optimization, SVM training and recognition. The feature extraction method is simple and nature, since it mainly utilizes wavelet moments, texture and color information of 3D objects derived from image views. In this 3DOR task, we proposed a GA-based strategy to optimize feature subsets and SVM parameters simultaneously. To evaluate the fitness function we used a much more efficient method that is to compute the theoretical bound of the generalization error for SVM on a candidate feature subset instead of the traditional way of performing Cross-Validation or Leave-One-Out.

The experiments based on the public 3D object dataset COIL-100 were conducted with different numbers of view angles as training set. The 100% correct rate of recognition could be obtained when the training sets were presented with view angle of every 10° and 20°. And when the number of training views was reduced, the correct rate of recognition was also satisfied. The theoretical analysis and experiments show that our method could effectively select out useful feature subset and good values for SVM model and achieved good recognition performance. This paper addresses the view-based 3D object recognition, that is, we mainly focus on the problem of recognizing 3D object in images, the framework of our method also can be used for general object recognition or image classification, such as face recognition, rock or wood texture classification and handwriting recognition, etc.

ACKNOWLEDGMENTS

This study was supported by the 2008 Texas Instruments (TI) Innovation Funds.

REFERENCES

Bicego, M., U. Castellani and V. Murino, 2005. A hidden *Markov model* approach for appearance-based 3D object recognition. Pattern Recogn. Lett., 26: 2588-2599.

Choi, J., Y.I. Cho, T. Han and H.S. Yang, 2008. A view-based real-time human action recognition system as an interface for human computer interaction. Virtual Syst. Multimedia, 4820: 112-120.

Delponte, E., N. Noceti, F. Odone and A. Verri, 2007. Appearance-based 3D object recognition with time-invariant features Proceedings of the14th International Conference on Image Analysis and Processing, May 22-22, IEEE Press, Modena, Italy, pp: 467-474.

Frohlich, H., O. Chapelle and B. Scholkopf, 2003. Feature selection for support vector machines by means of genetic algorithm. Proceedings of the 15th IEEE International Conference on Tools with Artificial Intelligence, Nov. 3-5, IEEE Press, pp: 142-148.

Haizhai, J., X.Z. Wang, S.F. Zhang and J. Li, 2007. View-based 3D object recognition using wavelet multiscale singular-value decomposition and support vector machine. Proceedings of the International Conference on Wavelet Analysis and Pattern Recognition, Nov. 2-4, IEEE Press, Beijing, China, pp: 1428-1432.

Haralick, R.M., K. Shanmugam and I.H. Dinstein, 1973. Textural features for image classification. IEEE Trans. Syst. Man Cybernet., 3: 610-621.

Huang, C.L. and C.J. Wang, 2006. A GA-based feature selection and parameters optimization for support vector machines. Exp. Syst. Appl., 31: 231-240.

Jun, F., Y. Yang, H. Wang and X.M. Wang, 2004. Feature selection based on genetic algorithms and support vector machines for handwritten similar Chinese characters recognition. Proceedings of the 3rd International Conference on Machine Learning and Cybernetics, Shanghai, Aug. 26-29, IEEE Press, pp: 3600-3605.

Marée, R., P. Geurts, J. Piater and L. Wehenkel, 2004. A generic approach for image classification based on decision tree ensembles and local sub-windows. Proceedings of the 6th Asian Conference on Computer Vision, (ACCV'2004), Asian Federation of Computer Vision Societies pp: 860-865.

Murase, H. and S.K. Nayar, 1995. Visual learning and recognition of 3-d objects from appearance. Int. J. Comput. Vision, 14: 5-24.

Pontil, M. and A. Verri, 1998. Support vector machines for 3D object recognition. IEEE Trans. Pattern Anal., 20: 637-646.

Pontil, M., S. Rogai and A. Verri, 1998. Recognizing 3-D objects with linear support vector machines. Proceedings of the 5th European Conference on Computer Vision-Volume II, Jun. 2-6, Springer Verlag, London, UK., pp: 469-483.

Roth, D., M.H. Yang and N. Ahuja, 2002. Learning to recognize three-dimensional objects. Neural Comput., 14: 1071-1103.

Shen, D. and H.H.S. Ip, 1999. Discriminative wavelet shape descriptors for recognition of 2-D patterns. Pattern Recogn., 32: 151-165.

Stricker, M.A. and M. Orengo, 1995. Similarity of color images. Proceedings of the Storage and Retrieval for Image and Video Databases, Feb. 9-9, San Jose, CA., USA., pp: 381-392.

Vapnik, V.N., 1995. The Nature of Statistical Learning Theory. Springer, New York, ISBN: 0-387-94559-8.

Vasconcelos, N., P. Ho and P. Moreno, 2004. The kullback-leibler kernel as a framework for discriminant and localized representations for visual recognition. Proceedings of the 8th European Conference on Computer Vision, May 11-14, Prague, Czech Republic, pp: 430-441.

Xiuwen, L. and A. Srivastava, 2002. A spectral representation for appearance-based classification and recognition. Proceedings of the 16th International Conference on Pattern Recognition, Aug. 11-15, IEEE Press, USA., pp: 37-40.

Zexuan, Z., Y.S. Ong and M. Dash, 2007. Wrapper-filter feature selection algorithm using a memetic framework. IEEE Trans. Syst. Man Cybernet. B, 37: 70-76.