A novel multiscale recursive recognition method for flying airplane objects

Tianxu Zhang, Gang Zhou, Ying Yang, Wenjie Weng, Xiangdong Sun
Institute for Pattern Recognition and Artificial Intelligence,Huazhong University of Science and Technology
State Key Laboratory for Multispectral Information Processing Technologies,Wuhan, Hubei,P.R.China

Abstract - *In actual imaging process, the projection or silhouette of a three-dimensional moving object is variable in shape, scale and resolution to make its recognizability also unstable. A remedy is suggested on concepts defined in this paper as the dynamic feature space of the pattern. The necessity of developing feature models of multi-scale characteristic view for three-dimensional moving airplane objects and the rationality of using a usual constraint on airplane object moving characteristics are discussed. Thus, a novel multi-scale intelligent recursive recognizing (MUSIRR) method for image sequences is proposed. An intelligent recognizer composed of hybrid neural networks and logic decision-making modules is constructed with the BP neural networks and RBF networks as the building blocks of the recognizer. The rationality and effectiveness of the new method proposed in this paper have been verified by results of massive simulation and actual experiments on several types of airplane object.*

Keywords: dynamic feature space; moving three-dimensional object recognition; multi-scale feature models; dynamic recognizability; hybrid neural-networks.

1 Introduction

The objective worlds is multidimensional and in dynamic state, while the recognition of three dimensional moving objects is a very important research project. Three dimensional moving bodies may be rigid or nonrigid, and the car, ship and aircraft are typical example of the former, while animal and human bodies are of the latter. This paper refers to only the former. Existent works are mostly focused on the recognition of stationary three dimensional bodies with mature suggestions and applications on feature extraction and classifier design. However, recognition of moving object is more difficult but has more practical and theoretical worthiness than that of stationary object. Yet there are some valuable researches have been reported. Y. LeCun, F, Huang and L. Botton [1] utilized a large amount of images taken from 50 kinds of toys, and built their feature space by Principal Component Analysis (PCA). Seibert and Waxman studied the automatic generation and adaptive recognition problem of two-value characteristic view of unoccluded 3D aircraft object, suggested and implemented the algorithm on analogous VLSI [2].Literature [3] studied digital recognition method for unoccluded 3D aircraft object using silhouette plus range information, in which a large amount of views was taken as training model and only single to three views of test sample objects were used to identify the aircraft object, and recognition based on sequential images, multi-scale problem, imaging condition and the influence of imaging obscurity were not considered. While literatures on this respect are seldom reported especially in a systematic way. This paper gives some valuable suggests and implements an effective new method to recognize moving targets from image sequence.

Remaining part of this paper is organized as follows. The second section defines the concept of dynamic feature space. The third section gives the multi-scale, multi-view and dynamic modeling as the preparation work for the recognition of 3D moving object. The basic theory of Gaussian Extended Mapping is recalled first, followed by our novel proposal of multi-scale modeling with specific steps to carry out the modeling. The moment invariant feature extraction process has been introduced in literature [4,5], therefore not included in this paper. The fourth section introduces the multi-scale intelligent recursive recognition (MUSIRR) method. The fifth section gives the experimental comparison and analysis on the method proposed in this paper with other typical recognition processes based on single scale and that on single frame static recognition concept.

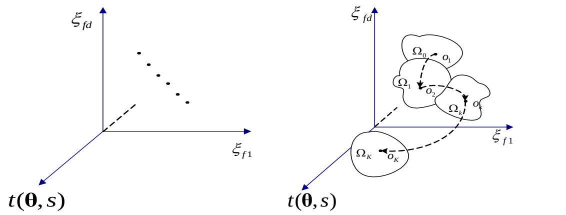
2 Dynamic feature space

According to the classical pattern recognition theory, in a problem domain, the feature space in which the classifier operates is usually stationary. Such a way of defining and representing feature space should be extended and generalized as it is totally unsuitable to solve object recognition problem under complex conditions and in dynamic scene.

2.1 Dynamic feature space

Given to a group of patterns ($\omega_i, i = 1, 2, \cdots, I$), every class of pattern vectors $\boldsymbol{\xi}_i = (\xi_{i1}, \xi_{i2}, \cdots \xi_{iD})$ will constitute a unique subspace $\boldsymbol{\Omega}_i$ in the observation space Ω . Assume that the D-dimensional observation space was transformed into a d-dimensional feature space via feature extraction mapping A. Every class of pattern feature vectors will constitute a unique subspace $\boldsymbol{\xi}_{fi} = (\xi_{fi1}, \xi_{fi2}, \cdots \xi_{fid})$ in the d-dimensional feature space $\boldsymbol{\Omega}_f$ .If both $\{\boldsymbol{\xi}_{fi}\}$ and

$\{\Omega_{fi}\}$ vary in time, and then Ω_f was named the dynamic feature space.



(a) feature space with time dimension added (b) shift of the dynamic pattern subspace

Fig. 1. Schematic diagram of the dynamic feature space.

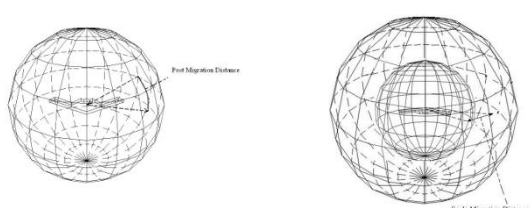
3 Multi-scale feature modelling for 3-D moving airplane objects

3.1 Multi-scale and Multi-view Modeling

The characteristic views of three-dimensional objects obtained by Gaussian extended image mapping are no doubt effective model for two-dimensional representation of a three dimensional object. While the drawback of the model lies in the failure to consider the fact that under actual imaging conditions, blurring of the object image and variation of scale exist constantly, which are induced by the variation of viewing distance due to object movement when the imaging sensor is made of fixed focus length and fixed number of sensitive elements on the focal plane. To overcome these drawbacks, the representation by the conventional characteristic views is extended to multi-scale characteristic view representation in this paper, to reflect the actual imaging conditions.

The extended Gaussian viewing sphere [6] can be divided into a number of characteristic viewing regions in the same way as the longitude and latitude as shown in Fig. 2(a).The multiscale Gaussian viewing sphere was shown in Fig. 2(b).

When a three-dimensional object is in motion, the effect due to variation of its gesture is equivalent to shift the viewpoint on the surface of Gaussian sphere. The effect due to variation of the distance from the three-dimensional object to viewer is equivalent to the scale variation of the object image or the degree of blurring.



(a) single-scale pose migration distance (b) multi-scale migration distance

Fig.2. Comparision of single-scale and multi-scale of the Gaussian viewing sphere.

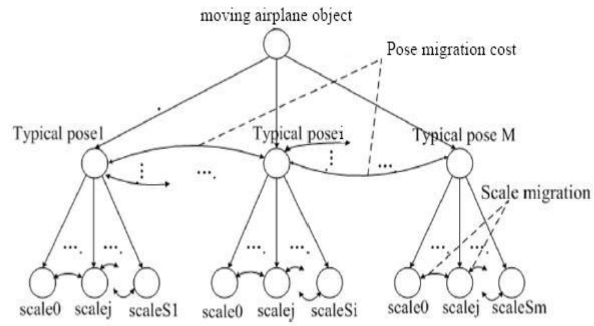


Fig.3. Multi-scale, multi-pose representation of 3-D moving airplane objects

3.2 Multi-scale and Multi-view Modeling Implementation

The 3-D models of airplanes are established using MultiGen Creator with shape data taking from the real object. The projected silhouettes of 3-D objects, especially moving 3-D objects, on to the 2-D plane are considered as simple and robust shape descriptions. Vega was used to get the silhouettes. The observers watching angle around a certain object was set to shift by 20^0 along the elevation from degree 0 to degree 180 and the azimuth from degree 0 to degree 360. In this way, we generated the 162 samples of scale 0 (256x256).

4 Multi-scale intelligent recursive recognition method

In this section an effective multi-scale intelligent recursive recognition method (MUSIRR) is proposed and implemented. A bolck diagram of the recognizing system and algorithm developed are shown in Fig.4, including the three basic procedures of modeling, training, and recognition verification. The system consists mainly of three modules, i.e, the gesture discriminator based on RBF networks, intelligent classifier based on BP networks[7], and fusion decision. The specific method of segmentation is not dealt with in this paper. Suppose an object has been segmented from the sample, but it may have defects, or distortion resulting from the effects of noise, background clutter and imperfect algorithm.

4.1 Single frame recognition based on the multi-scale models and multi-scale BP networks

The most fundamental recognition algorithm is performed by the BP networks based on single-scale model training. Suppose the number of categories to be recognized is K, that is, the object category set is $f = (f_1, f_2, \cdots, f_k)$. The output set of the BP networks is $o = (o_1, o_2, \cdots, o_k)$. The K outputs respectively represent K categories of different objects.

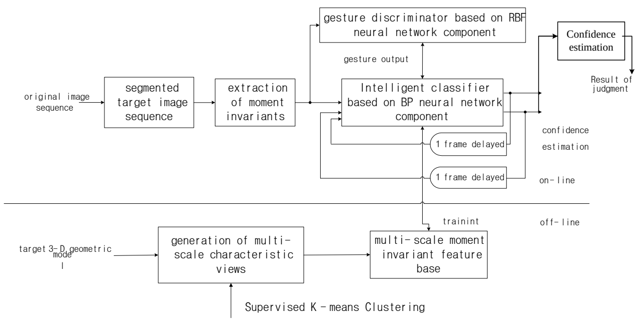
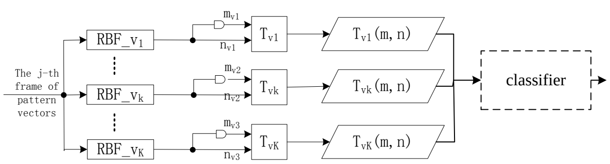
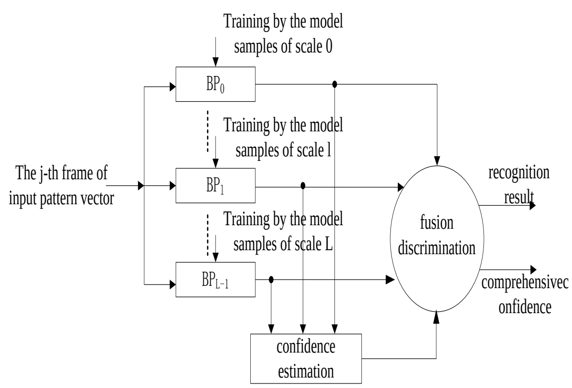


Fig. Fig.4. Block diagram of the recognizing system



(a) Multi-object gesture discriminator based on RBF network components



(b) Multi-scale recognizer based on BP network components

Fig. 5.Schematic diagram of the recognizing system

4.1.1 The relative confidence

With the output results $o=(o_1,o_2,\cdots,o_k)$ arranged from big to small as $o'=(o'_1,o'_2,\cdots,o'_k)$, $0\le o'_i\le 1$, the relative confidence in single frame recognition can rationally be defined as:

$$c=(o'_1-\frac{1}{1\times 2}o'_2-\cdots-\frac{1}{K(K-1)}o'_K)/(o'_1+o'_2+\cdots+o'_K) \qquad (1)$$

This definition is intuitive and rational:

(1) The range of variation of the confidence should be restrict to [0,1];

(2) If and only if the maximum output $o'_1=1$ and $o'_2=\cdots=o'_k=0$ the confidence has the maximum value 1;

(3) If and only if $o'_1=o'_2=\cdots=o'_k$, confidence c is the minimum value 0;

(4) The greater o'_1 is, the greater the confidence c while the smaller from o'_2 to o'_k, the greater the confidence c;

(5) With $\forall$m ≤ K, if $o'_1=o'_2=\cdots=o'_m=1$ and $o'_{m+1}=o'_{m+2}=\cdots=o'_k=0$, m equal maximum values would be obtained, if the corresponding category of o'_1 is chosen as the recognition output result, obviously that the confidence c should only be 1/m.

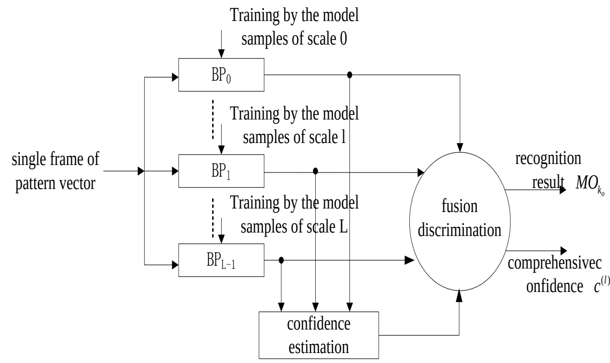


Fig. 6. Block diagram of single frame recognizer based on multi-scale models and multi-scale BP networks

Suppose the input feature vector $\mathbf{f}_j\in(\mathbf{f_1},\mathbf{f}_2,\cdots,\mathbf{f_K})$. Let the maximum output value of each network component BP_l of multi-scale BP networks respectively be $O_k^{(l)}$ ($k=1,2,\cdots K$, $l=0,1,2,\cdots L-1$) , l is the level of scale, then we can get the following results.

4.1.2 Single frame recognizing measure of the multi-scale BP networks recognizer

The single frame recognizing measure of the multi-scale BP networks recognizer can rationally be specified as:

$$MO_k=\frac{1}{L}\sum_{l=0}^{L-1}(O_k^{(l)}\times c^{(l)})\ ,k=1,2,\cdots,K \qquad (2)$$

Where $O_k^{(l)}$ is the maximum output value of BP_l and $c^{(l)}$ is the relative confidence as weighting factor.

4.1.3 The criterion for single frame recognition of multi-scale BP networks recognizer

If $MO_{k0}=\max\{MO_k\}$, $k=1,2,\cdots K$,then the input feature vector f_j is said to belong in the k0-th category $k_0\in\{1,2,\cdots,K\}$.

4.2 Multi-frame recursive recognition based on the multi-scale models

In contrast to the single frame image, the sequential images generated by the three-dimensional objects in motion contain more information about the objects. It is very valuable to study how to make full use of such information to increase the correct rate of recognition. See Fig.7 for the recursive multi-scale recognizer extended from the single frame recognizer in Fig.6, where the gesture changing information of inter-frame are used.

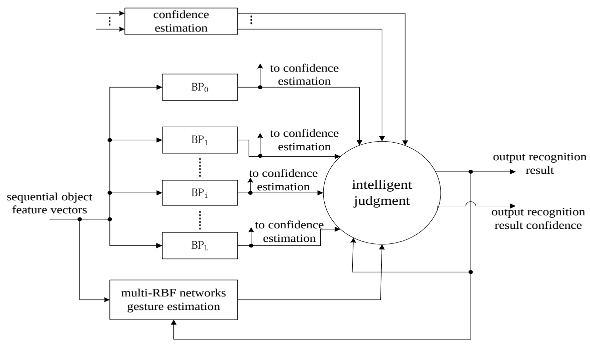


Fig. 7. Schematic diagram of the recognizing system

Suppose the feature vector $f(j)$ of the j-th frame of image of a certain object is input. Let $MO_k(j)$ be the response of the k-th node at the output end of the multi-scale recognizer (MSR) to the current single frame while the response at the preceding interval at the k-th out-put end of the multi-scale recursive recognizer (MSRR) is $MMO_k(j-1)$, then the response $MMO_k(j)$ of MSRR at the k-th output end at the current interval can be obtained by merging the information found by the current frame and that found earlier, the recurrence formula being as follows:

$$\begin{aligned} MMO_k(j) &= \frac{1}{j}[(j-1)MMO_k(j-1) + w_k(j)\cdot MO_k(j)] \\ &= \frac{1}{j}[\sum_{j_0=1}^{j-1} w_k(j_0)\cdot MO_k(j_0) + w_k(j)\cdot MO_k(j)] \end{aligned} \quad (3)$$

Where $w_k(j-1)$ is the (j-1)-th frame weight coefficient and $w_k(j)$ is the weight oefficient of the j-th frame. The value of $w_k(j)$ can control the magnitude of contribution of the new information obtained by the current frame to $MMO_k(j)$ on which making the current judgment relies. Below is an analysis of the rational value taken for the weight coefficient.

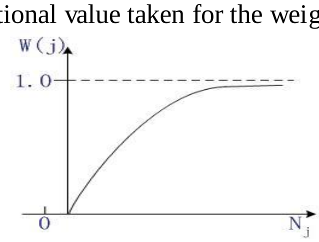


Fig. 8. The relationship in variation of weight coefficients with the number of object pixels

4.2.1 The criterion of multi-frame recursive recognition

If $MMO_{k0}(j) = MAX\{MMO_k(j), k = 1,2,\cdots,K\}$ (4)

We can get the input sample sequences $\mathbf{f} = \{\mathbf{f}(1), \mathbf{f}(2), \cdots, \mathbf{f}(j)\} \in \mathbf{f}_{\mathbf{k_0}}$.

The complete intelligent recognizer (MUSIRR) composed of mixed neural networks and logic decision modules for recognizing multiple categories of moving airplane objects is shown in Fig.9.

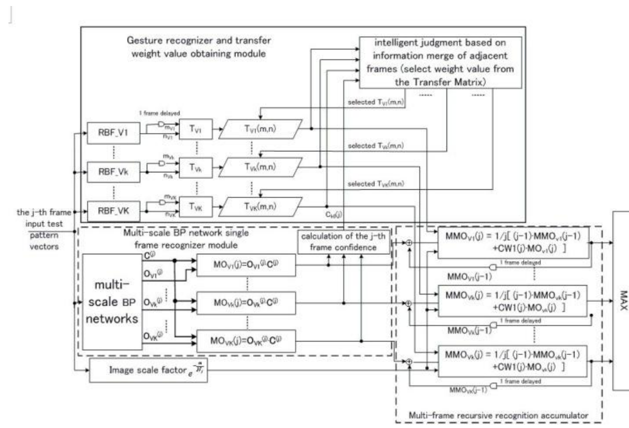


Fig. 9. Block diagram of a multi-category object recognizer (MUSSIRR)

4.3 Multi-frame recursive recognition based on the multi-scale models

An issue may arise that whether the correct recognition rate of MUSIRR will be convergent, e.g., to 1.0, as the recursive process is progressing. It is difficult to theoretically rigorously prove the convergency of the algorithm. However, experiments have proved that the classifiability of 3D moving objects has been significantly improved by using MUSIRR. The following logical results can be inferred:

(1)Within a relatively big variation of scale and shape distortion, the multi-frame recognition rate of MUSIRR will approach a higher value than that of single-frame recognition, especially for larger object images (smaller scale), the correct rate could be in the range of 0.95-1.0, when the number of image frames increases.

(2)The multi-frame correct recognition rate of MUSIRR using the weighted measure as Eq.(2) will approach a value higher than that using non-weighted simple accumulation measure as follows:

$$\begin{aligned} MMO^n{}_k(j) &= \frac{1}{j}[(j-1)MMO_k^n(j-1) + MO_k(j)] \\ &= \frac{1}{j}[\sum_{j_0=1}^{j} MO_k(j)]\ ,\ (k = 1,2,...,K),(j = 1,2,3,...,J). \end{aligned} \quad (5)$$

(3)For very small object image (very large scale) due to big loss of shape information, the multi-frame recognition rate of MUSIRR cannot converge to a very high value even though the number of frames has been largely increased, only when object images gradually go into smaller scale space.

5 Experiment results

5.1 Generation of training samples

The 3-D models of five airplane categories, including B2, F117, Mirage2000, F22 and Su27, were created see the appendix.For each model, 162 silhouette images were respectively generated from 162 different viewpoints based on the Gaussian View Sphere, and a small quantity of characteristic silhouettes obtained with a clustering process was used as the 2D representation of the models. Based on the original scale of characteristic silhouettes, those of other 6 scale levels were generated to get 7 scales of characteristic silhouette sets for each object. The stated 7 levels of scale included 256×256, 192×192, 128×128, 96×96, 64×64, 48×48 and 32×32. On each scale level, the five classes of airplane objects owned characteristic views in 38, 39, 44, 41, 42, respectively. So that , the five categories of objects had 7×38, 7×39, 7×44, 7×41, and 7×42 characteristic silhouettes respectively to reflect the different effects of different viewing points, distances, degrees of fuzziness and scales. From each characteristic silhouettes, an eight dimensional moment invariant feature vector could be obtained as invariant feature[8].

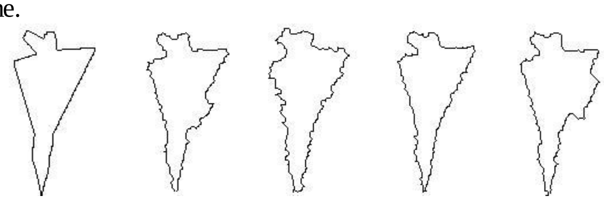
5.2 Generation of recognition testing samples

5.2.1 Generation of testing samples of stationary objects

For each object class, 450 frames of images were randomly generated by viewing on the Gaussian sphere. This kind of samples included 13 scale levels, with 7 of them being the same with training samples, and 6 new intermediate scale levels added, which size were 224×224, 160×160, 112×112, 80×80, 56×56, 40×40. In this way, there were 5×13×450 frames of static check samples without distortion for all 5 object classes. The segmentation algorithm is beyond the scope of this paper and the objects were assumed to have been segmented from the images already. Coordinate noise could be superimposed on the coordinates at the points of their contours causing the contour distorted to simulate the segmentation errors of object images due to the image noise and defects in segmentation. Examples of distortion intensity are shown in Fig. 10.

5.2.2 Generation of image sets of moving airplane

With randomly picked start points and directions, serial view point loci were produced on the Gaussian Spheres of different radiuses. For each locus, we could find the corresponding projected image from each view point on the locus, and make them to form image sequences as our recognition samples of moving objects. Some sequences described pose transient on one scale,some of them describe the case when pose transient and scale migration(i.e. on the Gaussian Spheres of different radiuses) occurred at the same time.



(a) (b) (c) (d) (e)

Fig.10.Examples of superimposing 4 different intensity distortions on object contour images

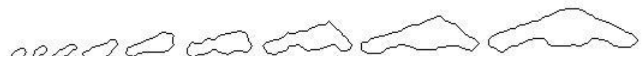
(a) original image (b) distortion tensity level1

(c) distortion tensity level2 (d) distortion tensity level3

(e) distortion tensity level4

5.2.3 Performance of MUSIRR on the image sequences

To testify the capability of MUSIRR on recognition, we did mass experiments based on the recognition test samples mentioned above. Experiment results show that with MUSIRR, moving airplane objects in the image sequences can be correctly recognized within about 10 frames, and video with 50 frames per second can be handled without problem in real time. Fig.11 shows the curves of recognition rate under 4 distortion intensity.

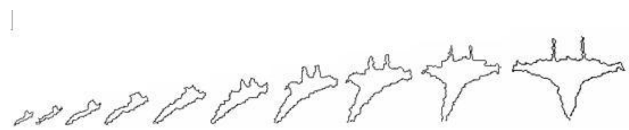


(a)sequential images of the airplane B2

(b) sequential images of the airplane F117

(c)sequential images of the airplane Mirage2000

(d)sequential images of the airplane F22



(e)sequential images of the airplane Su27

Fig.11 Distorted sequential object images with the scale and gesture varing at the same time

Fig. 11 gives such an example of image sequences in which the scale and gesture of object vary at the same time.Here the object image size in the sequence is increasingly becoming larger and larger. Fig.12 shows the curves of recognition rate under 4 distortion intensity.

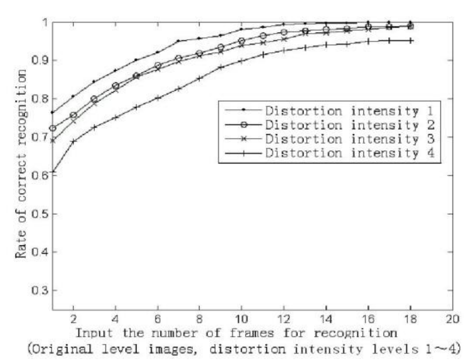
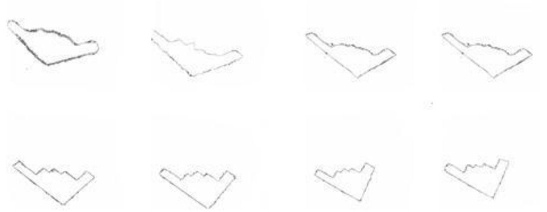
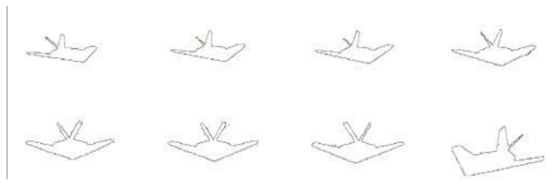


Fig. 12. Curves of recognition rate of multi-frame sequential mages with the gesture and scale varying at the same time

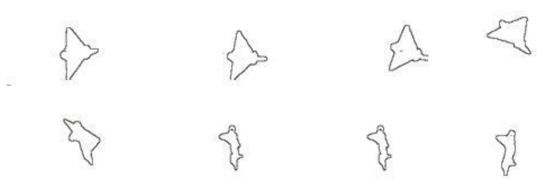
Part of the image sequences of five objects from the video disc are demonstrated in Fig.13 and Fig.14. For all these real image sequences, the single frame recognition rate by MUSIRR reaches 90.2%, and recognition rate of the multi-frame by MUSIRR reaches 100% within no more than 7 frames.



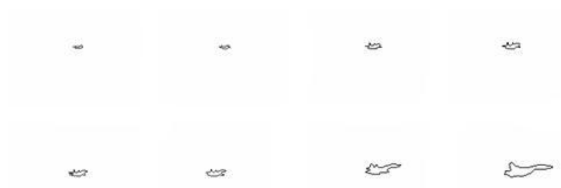
(a)8 out of the total 59 frames of B2 image sequence sample



(b)8 out of the total 120 frames of F117 image sequence



(c)8 out of the total 95 frames of Mirage2000 image sequence



(d)8 out of the total 800 frames of Su27 image sequence

Fig. 13. samples from real image sequences

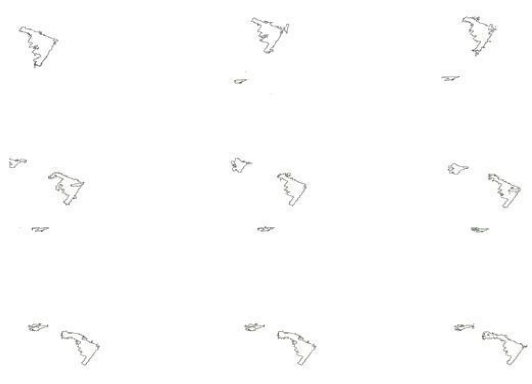


Fig. 13. silhouette samples from real image sequences with three kinds of airplane objects.

The recursive recognition algorithm disinters the sequential information of image sequences of moving object, and makes full use of them. It is an essential factor to secure the effectiveness of the recognition system, and an ultimate distinction from the algorithms to recognize only static object images.

Along with the pose transient and the scale migration brought by moving object, the pose and scale information reflected in the corresponding 2-D images is also changing, so does the recognizability of each frame in one image sequence. Single frame method merely use the information in each frame in isolation from others to put up the recognition, so the recognition result for each frame in image sequence may turn out conflictive, and lead to an unstable recognition rate.

Correspondingly, the confidence of the multi-frame recursive recognition process tends to increase with the increase of frame number, which is very different from the single frame method. Curves of variation of confidence in multi-frame recursive recognition are shown in Fig. 14.

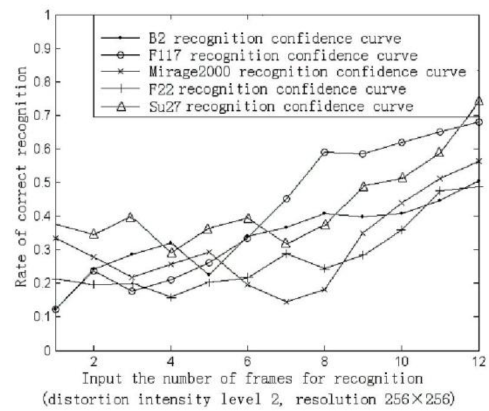


Fig.14.Curves of variation of confidence in multi-frame sequential recognition.

6 Conclusions

This work tries to study in a systematic way to solve the difficult issue of recognizing moving three-dimensional airplane objects. The ideas in this paper are different from those in available literature in the following respects:

(1)The modeling and recognition of the three-dimensional moving objects should be examined with dynamic feature space of patterns as the recognizability is dynamically varying;

(2)Since there exist such constraints as the limited resolution and fuzziness of the imaging sensor under actual imaging conditions, an object model based on the multi-scale characteristic views, or the effective representation of the three-dimensional moving object, should be developed;

(3)A novel multi-scale intelligent recursive method (MUSIRR) for recognizing the sequential images of three-dimensional moving airplane objects is proposed, which utilizes the general kinetic constraint of moving airplane object to get a correct recognition rate higher than that of conventional methods based on single scale models and stationary images.

(4)As few as possible on multi-scale characteristic silhouette contours are used as representation of moving airplane objects, and the moment invariant feature extracted from these contours are used as the training samples input into the new recognizer. So the recognition process would not be interfered by the irrelevant information of how the object appears in the image, like illumination or grey level to reduce the data and computing complexity.

(5)An intelligent recognizer has been constructed with hybrid neural networks, featuring a simple training process, real-time processing ability and higher rate of correct recognition.

(6) Unlike the common 3D target recogniton method, not only the type of the target was identified but also the pose was recognized at the same time with the proposed method, which gives a more precise representation of the target.

It has been pointed out in relevant literature that increase of feature dimensionality may not always lead to the improvement of the correct rate of recognition or class separability but may even result in its deterioration. That is because of increasing only the feature dimensionality without dynamically selecting the optimum feature vector. On the contrary, it is proposed in this paper that the dimensionality of the feature space should be increased without increasing the dimensionality of the feature vector. An increase in the dimensionality of the feature space will lead to an increase in the selectivity of subspace, that is, in certain subspace, the rate of recognition may be better than that in other subspaces. Increasing the dimensionality of the feature space is significantly different from increasing the dimensionality of the feature vectors.

7 Acknowledgment

This work was supported by the Key Project of the National Natural Science Foundation of China under grant (No.60736010).

8 Reference

[1] J.J.Koenderink and A.J.V.Doorn. "The internal representation of solid shape with respect to vision". Byo, Cyb. vol.32, no.4, pp.211-216, 1979.

[2] M.Seibert and A.M.Waxman. "Adaptive 3-d Object Recognition from Multiple Views". IEEE Trans. PAMI, vol.14, no.2, pp.107-124, 1992.

[3] A.P.Reeves and R.W.Taylor. "Identification of Three-dimentional Objects Using Range Information". IEEE Trans PAMI, vol.11, no.4, pp.403-410, 1989.

[4] M.K.Hu. "Vision Pattern Recognition with Scale and Rotation Invariance". IRE Trans. Inf. Theory, vol.8,pp.179-187, 1962.

[5] Liu Jin and Zhang Tianxu. "Fast Algorithm for Generation of Moments Invariants". Pattern recognition. , vol.37, no.8, pp. 1745-1756, August 2004.

[6] B.K.P. Horn. "Extended Guassian Images". Proc.IEEE, vol.72, no.12, pp. 1671-1686, 1984.

[7] Jing Guolin, Du Wenting. "Studies on prediction of separation percent in electro dialysis process via BP neural networks and improved BP algorithms". DESALINATION, APR 2012,vol291,p78-93.

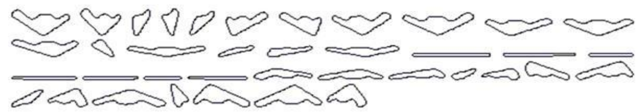
[8] Tianxu Zhang, Jin Liu. "An Investigation on the Stability of Targets". Journal of IR and Millimeter Wave".vol.23, no.3, pp. 197-204, 2004.

9 Appendix

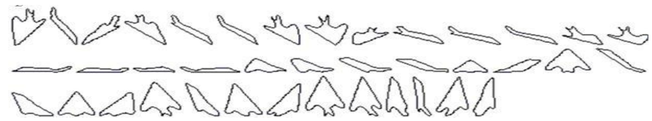
Two scale levels of characteristic silhouette contour models from 7 scale levels of characteristic silhouette contour models (clustering with moment feature vectors).

The 3-rd level scale (96x96)

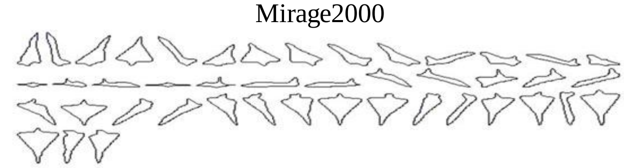
B2



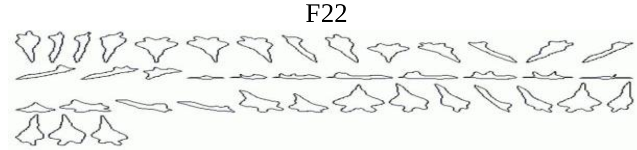
F117



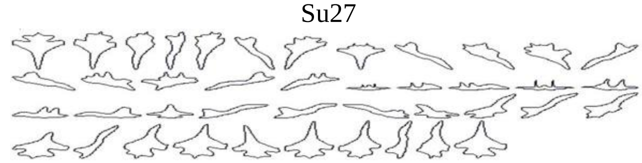
Mirage2000



F22

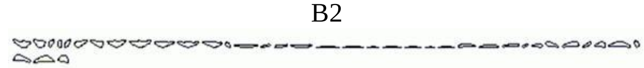


Su27

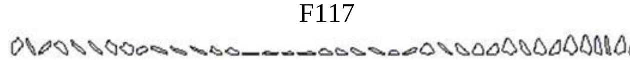


The 6-th level scale (32x32)

B2



F117



Mirage2000

F22

Su27