41. Analysis of Image Sequences for the Unmanned Aerial Vehicle*

Hung Son Nguyen, Andrzej Skowron, and Marcin S. Szczuka

Institute of Mathematics, Warsaw University
Banacha 2, 02–097, Warsaw, Poland
{son,skowron,szczuka}@mimuw.edu.pl

A method for extracting relevant information from image sequence data is presented. The image sequences, being output of video system of the Unmanned Aerial Vehicle, are analysed with use of EM-clustering techniques and Rough Set based methods. The possibilities of construction of an automated system for recognition/identification of cars on the road, on the basis of colour-related data are discussed.

41.1 Introduction

The issue of constructing and controlling an autonomous Unmanned Aerial Vehicle (UAV) is a multi-fold one. The idea of constructing such a vehicle (helicopter) for the purposes of traffic control drives the WITAS project (see [41.8]). Apart of difficulties in construction of proper hardware the problem of establishing software is a challenging one. The UAV is supposed to recognise the road situation underneath on the basis of sensor readings and make the decision about acts that are to be performed. The issue of constructing adaptive, intelligent and versatile system for identification of situation was addressed in [41.5]. In the paper we focus on one of the subtasks necessary for the entire system to work – the problem of discerning between objects that are visible to the UAV.

The most crucial information for UAV is provided by its video systems. We have to be able to provide UAV control system with information about car colors and so on. Such information may allow for making the identification that is core for operations performed by UAV, such as tracking a single vehicle over some time.

In the paper we address only a part of issues that have to be resolved. The particular task we are dealing with is identification of techniques that may be used for the purpose of discerning and/or classifying objects from image sequence data. Given a series of images gathered by UAV's video system we have to extract the valuable information about cars present in the image. The key is to have compact set of features that at the same time are robust the image data may be heavily distorted. The unwanted effects coming from changes in UAV's position, lighting conditions, scaling, rotation and weather conditions have to be compensated.

*Supported by the Wallenberg Foundation and grant KBN 8T11C02519.

T. Terano et al. (Eds.): JSAI 2001 Workshops, LNAI 2253, pp. 333−338, 2001.

41.2 Data Description

At the current stage we are dealing with two sequences of images consisting of 100 frames each. They represent two situations on the road, each about 4 second long. Every frame is a 24 bit .tiff image with resolution 726×512 pixels. The image sequences have been manually interpreted. Altogether 18 objects representing cars on the road have been identified The object instance (colour blob) is represented with 30 attributes. number (identifier) assigned to an object, two numerical attributes representing X and Y coordinates (within an image) of the center of colour blob (object) and 27 attributes representing coordinates in the RGB colour space for 9 pixels being a 3×3 matrix surrounding the center of colour blob. For each of 18 identified object we have 100 instances, one for each image in sequence (1800 samples in total).

41.3 The Task

The overall problem of situation identification on the basis of image (and possibly some other) data is very compound. In the first stage, described in this paper we would like to find the answers to the following questions:

1. Is the existing amount of information (27 colour-related attributes) sufficient for construction of classification support system that is able to distinguish between 18 pre-identified objects?
2. Is it possible to transform the existing 27 dimensional attribute space to the form better supporting car colour classification tasks?
3. Is it possible to learn the basic concepts allowing for establishment of prototypes rules of classification provided we have part of the sequence, say first 50 images, and then classify objects for the rest of sequence properly?

41.4 The Method

Initially, an attempt to perform car (colour blob) dissemination and/or classification with use of typical methods from the Rough Set armory (see [41.4]) have been made. Unfortunately, it turned out that the data is too vague and distorted for the typical tools like RSES ([41.7]) or Rosetta ([41.6]).

We came to the conclusion that some method for extraction of more relevant features from the raw data is needed. Therefore we turned our attention at unsupervised learning methods that allow for identification of characteristic features of objects in the corpus. The main intention is to eliminate unwanted effects caused by changes in object RGB colours as the object (car) moves between zones of different light. The particular approach we apply uses clustering and simple time series analysis.

First, we perform clustering treating all 1800 measurements as points in 27 dimensional space (9 points$\times$3 RGB coordinates). To do the clustering we utilise Expectation Maximisation (EM) method. EM-clustering is an iterative, unsupervised clustering method aimed at establishment of possibly small number of not intersecting clusters constructed with assumptions about normal distribution of objects. For details about EM clustering see [41.1] and [41.3].

After the clustering have been found we recall the information about sequential character of our data. Namely, we analyse the sequences of cluster assignments for each of 18 cars. Going frame-by-frame we check to which clusters the object belong in scope of this frame. In this way for each of 18 cars we get a vector of 100 cluster assignments. Such vectors may be compared and on the basis of differences between them we may discern one car from the others.

41.5 Results

The clustering was applied to the entire data. As a result of several experiments we got 15 to 18 clusters on the average. For all objects the assignment to cluster was very characteristic. In most cases it was possible to distinguish 2-3 clusters to which the samples corresponding to the single cases were assigned. These 2-3 clusters contained more than 80% of car on the average. Moreover, it was possible to correlate the change of cluster assignments with changes in lighting of car on the road. As the car enters the area of shadow, the visual perception of its colour is changing and so its cluster assignment. This effect is very welcome from our point of view since it makes clear evidence of cluster relevance.

On the basis of clustering new features were constructed for the objects. For each object (car) C_i ($i = 1, ..., 18$) we construct new attributes $na_1, ..., na_c$ where c is the number of the clusters derived. The value of attribute na_j for the car C_i is the number of occurrences of an object representing i-th car in j-th cluster. So, if the value of attribute na_1 for car C_1 is 20 then we know that an object corresponding to this car was assigned to first cluster 20 times out of hundred. This new set of attributes undergone further analysis. By applying Rough Set based techniques it was possible to find out that attributes derived from clusters are sufficient for discernibilty. Namely, it was possible, with the use of RSES software (see [41.7]), to calculate a set of *if..then..* decision rules classifying (discerning) the cars. In this way we got a simple set of rules such that there was exactly one rule for each of 18 cars.

Since clustering have led us to so promising results in terms of ability for object dissemination, we tried to exploit its potential to the limit. Since the clustering process takes some time in case of 1800 objects and 27 numerical attributes we were looking for the way to make it simpler. Reduction of computational effort is in our case very important since major part of recognition process has to performed on-line, during UAV operation. We found out that the clustering-based approach is quite powerful. We performed an experiment using reduced information about colour blobs. Instead of 27 attributes representing three RGB coordinates of 9 points (3×3 matrix) we take only three. These three are averages over 9 points for Red, Green and Blue coordinate values respectively. For this reduced set of features we obtained a clustering and it was still possible to have good discrernibility between objects. Moreover, the time needed for computation was reduced several times.

The results presented above address the question about amount of useful information that can be retrieved from image sequences. The other question on our task list was the one about potential abilities for construction of classification system.

Initial experiments aimed at construction of classification method based on inductive learning of concepts were performed. We wanted to check what are the possibilities to create a system that will be able to classify previously unseen objects as being similar to the prototypes learned during presentation of training sample. For this purpose we first split our set of examples into halves. One half, used for training, contains first 50 samples for each car i.e. frames 1 to 50 from both image sequences. The remaining 50 frames from each sequence form the dataset used for testing. On the basis of training set we establish clustering-based features and decision rules using these features. Then we take a sample from the testing set and label them with the car numbers.

In the experiments we use simplified version of cluster-based attributes presented above. Instead of attributes $na_1, ..., na_c$ for training samples we take binary attributes $ma_1, ..., ma_c$. Attribute ma_1 for a given sample is equal to 1 iff $na_1 > 0$ for this sample, and 0 otherwise.

Since we have to check abilities of classification system we start first with the learning phase. Learning of classification (decision) rules is done on the basis of 18 samples of 50 frames each. So the learning data consists of 18 objects, each object described by c attribute values, where c is the number of clusters.

First attempt was performed for testing samples consisting of entire 50 remaining frames. By matching those examples against previously created clusters, producing cluster-based attributes and the assigning decisions (car numbers) to the samples we got the result for training sample. In this particular experiment we got a perfect accuracy (100%).

Unfortunately, taking 50 frames requires approximately two seconds which is too long for real-time application. Therefore, we would like to be able to reduce the number of frames in testing sample to no more than 15-20 and still retain good classification ratio.

To do that we process our test data and produce testing samples with use of moving window. We set a size of the window to be some integer not greater than 50. Then from 50 frames we produce the testing sample by taking as many sequences of the size of window as possible and calculate cluster-related attributes $ma_1, ..., ma_c$ for them. For instance, if we set the size of the window to be 15 then we will get 35 samples for each car. First of these samples will contain frames from 51 to 66 while the last will consist of frames 86 to 100. So, altogether for 18 cars we will get 830 testing instances.

The key is now to find the size of the window to at the same time small enough to allow on-line classification and big enough to have good quality of this classification. From several attempts we have learned that with the methods of attribute generation and decision rule derivation depicted above, we are able to get perfect accuracy of classification for testing sample if the size of the window exceeding 17. For the window size less than 17 the accuracy decreases, being 89% and 78% for the windows of size 16 and 15, respectively. It is worth mentioning that these experiments are, at the moment of writing, only initially finished. We expect to improve the results by allowing more information to be passed to classifier e.g. by using the original attributes $na_1, ..., na_c$ instead of simplified $ma_1, ..., ma_c$.

41.6 Conclusions

The method for extracting information from image sequences was presented. It is based on combination of unsupervised clustering with Rough Set based approach. From the initial experiment we may see that this approach has a significant potential and may be further developed into complete solution. The proposed method have to be tuned to fit the requirements for co-operation with other components of UAV's control system as well as expectations about robustness, versatility and speed of operation.

The natural next step is the application of developed solutions to other sets of image data. We expect that some further evolution of the methods will be necessary, since many problems may arise. We believe that with more data we will be able to generalise our approach using tools such as more compound time series analysis.

References

41.1 Ambroise C., Dang M., Govaert G. (1997), Clustering of Spatial Data by the EM Algorithm.In: A. Soares et al. (eds), geoENV I-Geostatistics for Environmental Applications, Kluwer, Dordrecht, pp. 493–504.

41.2 Backer E., Kandel A.(1995), Computer-Assisted Reasoning in Cluster Analysis, Prentice Hall, New York.

41.3 Celeux G., Govaert G. (1992) A classification EM algorithm for clustering and two stochastic versions, Computational Statistics and Data Analysis, 14, pp. 315 – 332.

41.4 Komorowski J., Pawlak Z., Polkowski L., Skowron A.(1999). Rough sets: A tutorial. In: S.K. Pal and A. Skowron (eds.), Rough fuzzy hybridization: A new trend in decision-making, Springer-Verlag, Singapore, pp. 3–98.

41.5 Nguyen H.S., Skowron A., Szczuka M. (2000), Situation Identification by UAV, Proceedings of CS&P 2000, Informatik Berichte, Humboldt-Univerität zu Berlin
41.6 The Rosetta Homepage – `www.idi.ntnu.no/~aleks/rosetta`
41.7 The RSES Homepage – `logic.mimuw.edu.pl/~rses`
41.8 The WITAS project Homepage –`www.ida.liu.se/ext/witas/`