Analysis of Image Sequences for the Unmanned Aerial Vehicle

Marcin S. Szczuka and Nguyen Hung Son

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

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 (the Wallenberg laboratory for research on Information Technology and Autonomous Systems) project (see [4], [14]). 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 [10]. In the paper we focus on one of the subtasks necessary for the entire system to work. We cope with the problems related to 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 (description) about car colours 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 solved. 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 we are dealing with may be heavily distorted. There are several problems such us changes in car's (object's) colour characteristics as the moving vehicle passes from light to shadow and so on. Also, the unwanted effects coming from changes in UAV's position, lighting conditions, scaling, rotation and weather conditions have to be compensated.

In the paper we present the results of some experiments with image data aimed at establishment of proper and robust methods for object (colour blob) classification. The proposed approach uses selected techniques from the area of Rough Set Theory as well as probability-related clustering algorithms.

2 Basic notions

2.1 Rough Sets

The structure of data that is subject of our study is represented in the form of *information system* [11] or, more precisely, the special case of information system called *decision table.*

Information system is a pair of the form $\mathbf{A} = (U, A)$ where U is a *universe* of *objects* and $A = (a_1, ..., a_m)$ is a set of *attributes* i.e. mappings of the form $a_i : U \to V_a$, where V_a is called *value set* of the attribute a_i. The decision table is also a pair of the form $\mathbf{A} = (U, A \cup \{d\})$ where the major feature that is different from the information system is the distinguished attribute d. In case of decision table the attributes belonging to A are called *conditional attributes* or simply *conditions* while d is called *decision* (sometimes *decision attribute*). We will further assume that the set of decision values is finite i.e. $V_d = \{d_1, ..., d_n\}$.

The $i-$th *decision class* is a set of objects $C_i = \{o \in U : d(o) = d_i\}$, where d_i is the $i-$th decision value taken from decision value set $V_d = \{d_1, ..., d_n\}$. In our particular case the set of conditional attribute values will be finite and bounded.

The key notion to this study is the *indiscernibility.* For any subset of attributes $B \subset A$ *indiscernibility relation* IND(B) is defined as follows:

$$xIND(B)y \Leftrightarrow \forall_{a \in B} a(x) = a(y) \tag{1}$$

where $x, y \in U$.

$B \subset A$ is a *reduct* of information system if $IND(B) = IND(A)$ and no proper subset of B has this property.

Decision rule is a formula φ of the form

$$(a_{i_1} = v_1) \wedge ... \wedge (a_{i_k} = v_k) \Rightarrow d = v_d \tag{2}$$

where $1 \leq i_1 < ... < i_k \leq m$, $v_i \in V_{a_i}$. Atomic subformulae $(a_{i_1} = v_1)$ are called *conditions.* We say that rule r is *applicable* to object, or alternatively, the object *matches* rule, if its attribute values satisfy the premise of the rule. With the rule we can connect some characteristics. *Support* denoted as $Supp_{\mathbf{A}}(r)$ is equal to the number of objects from $\mathbf{A}$ for which rule r applies correctly i.e. premise of rule is satisfied and the decision given by rule is similar to the one preset in decision table. $Match_{\mathbf{A}}(r)$ is the number of objects in $\mathbf{A}$ for which rule r applies in general. Analogously the notion of matching set for a collection of rules may be introduced. By $Match_{\mathbf{A}}(R, o)$ we denote the subset M of rule set R such that rules in M are applicable to the object $o \in U$. The rule is said to be *optimal* if removal of any of its conditions causes decrease of its support.

2.2 Clustering

In principle, clustering is the process of grouping the data samples into larger clumps that share the same characteristics. The clustering is widely recognised as one of the methods of unsupervised machine learning (see [2], [7]). In previously defined terms we have an information system $\mathbf{A} = (U, A)$ where $A = (a_1, ..., a_m)$ and $U = (o_1, ..., o_n)$. We want to create a family of object groups (clusters) $G_1, ..., G_k$, where $G_i \subseteq U$ for $i = 1, .., k$. The clusters may or may not intersect, depending on the clustering method chosen.

Probably the best known and most intuitive clustering method is the *k-means clustering*. This method decompose the data sample into exactly k disjoint clusters. The clustering is based on pre-determined measure of distance between objects.

This method, however easy to comprehend, has several limitations. Firstly, we need to set the number k of clusters which may be hard to do in case of data we have no deep knowledge of. Secondly, the distance measure must be set (more in [2], [7]).

To overcome the known shortcomings of k-means clustering several other clustering methods have been developed.

One of adaptations of k-means iterative idea to more probabilistic fashion is the *EM-clustering* (EM - Expectation Maximisation) algorithm. It is an iterative algorithm that in subsequent steps performs two operations:

1. Calculation of cluster probabilities for each data instance (expectation).
2. Estimation of data distribution on the basis of cluster probabilities.

The procedure repeats until the logarithmic measure of cluster quality known as likelihood stops to increase.

The EM-clustering algorithm uses the estimation based on multidimensional, normal distribution based mixture model. It works well for the data represented as tuples of real numbers (points in $\mathbb{R}^s$), especially if we can make an assumption about independence of attributes. The number of clusters may be induced in different way. We may either define it or let the algorithm to adaptively find the right number of clusters starting with a random choice. For more information about EM-clustering consult [1], [2], [3], [7].

3 The WITAS project

The WITAS project is long term (years 1997-2003) research project aimed at establishment and implementation of methods allowing for construction of Command and Control Software Architecture (CSSA) for Vertical Take-off and Landing (VTOL) Unmanned Aerial Vehicle (UAV). Although the research is goal-oriented, the idea is to construct an architecture that will be possibly universal and adaptable to the case of controlling unmanned system other than UAV e.g. autonomous ground vehicles.

The WITAS CSSA may be characterised as a hybrid, multi-layered deliberative-reactive architecture. In its setting it shares some similarities with the systems proposed in [6] and [8]. Three main layers in this architecture are the deliberative layer, the reactive layer and the process layer. The system is also equipped with the Knowledge and Data Repository which stores all the information necessary for the systems' operation.

The issues of object classification addressed in the paper are related to one more element of the architecture, the Dynamic Object Repository (DOR). The DOR is a database-like structure that allows for storing, updating and retrieving with the use of queries the information about objects identified by UAV's sensors (see [5]) .

The actual WITAS hardware platform is constructed on board of the Yamaha mini-helicopter. The information to the CSSA is provided by:

- The vision system equipped with video camera that is mounted in specially designed stabilised housing (turret).
- Inertial navigation system providing information about changes in UAV's position.
- Differential Global Positioning System (GPS) allowing for positioning UAV in universal world coordinates.
- Geographical Information System (GIS) storing digital maps of the territory underneath.
- Internal system reading such as throttle, pitch angles etc.

For more information about WITAS project refer to [4] and [14].

4 Data description

At the current stage we are dealing with two sets of data being sequences of images (video feeds) consisting of 100 frames each. They represent two situations on the road, each about 4 seconds 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. 8 of them appear in the first sequence and remaining 10, in the second sequence (see figure 1). The object instance (colour blob) is represented with 30 attributes. These are:

- 1 attribute representing the number (identifier) assigned to an object.
- 2 numerical attributes representing X and Y coordinates (within scope of a single 726×512 image) of the point being taken as the centre of colour blob (object).
- 27 attributes representing coordinates in the RGB colour space for 9 pixels being a 3×3 matrix surrounding the centre of colour blob.

For each of 18 identified object we have 100 instances, one for each image in sequence (1800 samples in total).

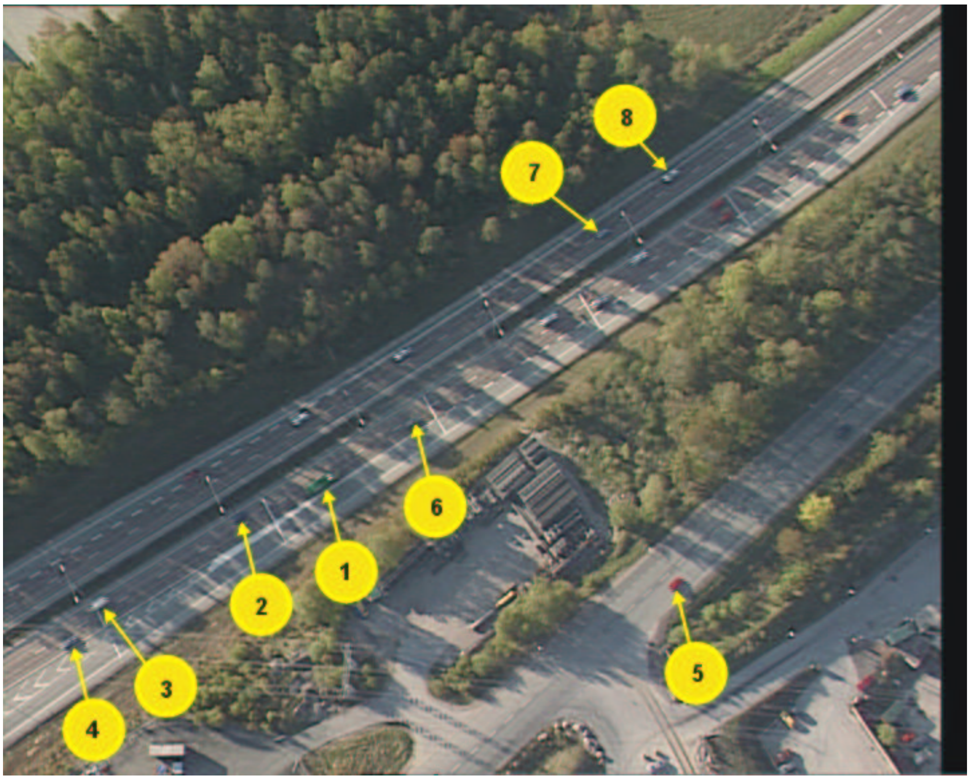

Fig. 1. One frame from first image sequence.

Some statistical tests were performed to check the quality of data. It turned out that the samples are distorted quite heavily. The standard deviation was very high for most of the cases. This information was a clue for considering the methods that are robust and invariant to the distortion.

In fact, the particular set of images we are considering in this study is a bit more complicated than we expect the average operational data to be. The pictures are taken from a helicopter flying at high altitude (see 1) while the WITAS UAV is supposed to fly closer to the ground.

5 Tasks

The overall problem of situation identification on the basis of image 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 (features) 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?

6 Results

Initially, an attempt to perform car (colour blob) dissemination and/or classification with use of typical methods from the Rough Set armoury (see [9]) have been made. Unfortunately, it turned out that the data is too vague and distorted for the typical tools like RSES ([13]) or Rosetta ([12]).

It also turned out from experiments that conversion of existing RGB representation into different colour space (e.g. YUV, HSV, CMYK) do not make much difference or even makes things worse. We cannot tell whether some other method for colour space conversion would or would not be effective. It is possible that some non-linear conversion method may improve the validity of data but, we were unable to find such a conversion in the easy way.

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 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 (see [1] and [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. The examples of such cluster assignment sequences are shown in figure 2.

The clustering was applied to the entire data. As a result of several experiments we got 16 to 18 clusters on the average. One of the objects (Car 5) was very characteristic due to its intensive red colour. For this car we got perfect clustering i.e. for every attempt there was one cluster that contained only the objects corresponding to that car. For the rest of objects the situation was less comfortable. However, 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

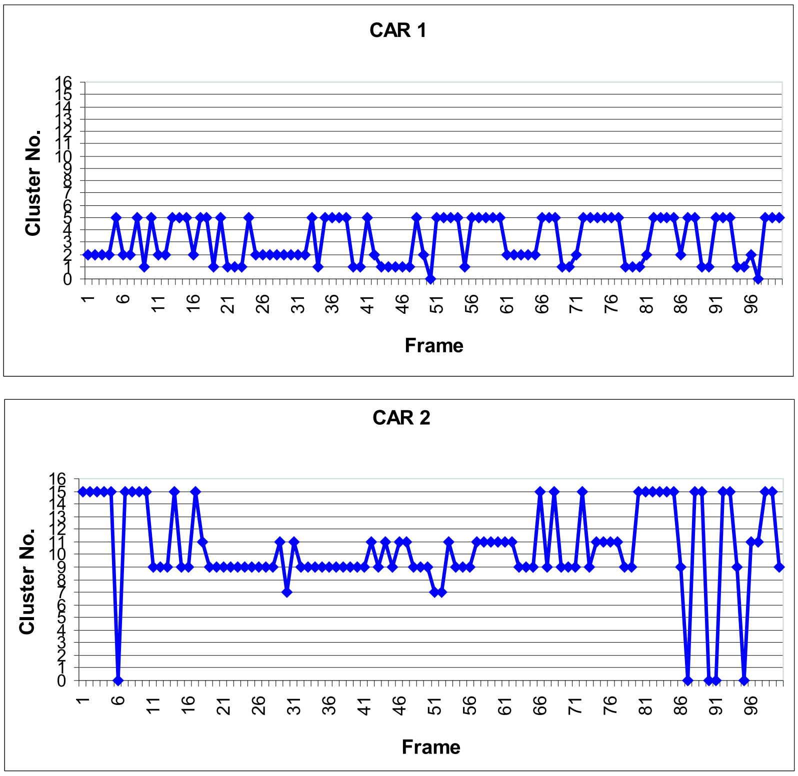

Fig. 2. Cluster assignments for two vehicles from the first sequence.

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 discernibility. Namely, it was possible, with the use of RSES software (see [13]), to calculate a set of $if..then..$ decision rules classifying (discerning) the cars. These rules are very simple and there is exactly one rule for each car.

Since clustering have led us to so promising results in terms of classification ability, we tried to exploit its potential to the limit. Unfortunately, the clustering process takes some time in case of 1800 objects and 27 numerical attributes so, 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 be performed on-line, during UAV operation. 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 which is not as nice as in case of all 27 attributes, but still it was possible to have good discernibility between objects. Moreover, the time needed for computation was reduced several times since the EM-clustering method strongly depends on number of features (attributes).

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 (not considered during training) 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 the we got a perfect accuracy (100%).

Unfortunately, taking 50 frames requires approximately two seconds which is too long for real 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. Then, the method could be regarded as applicable for real-time processing.

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 630 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 $ma_1, ..., ma_c$.

7 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. Several questions have still to be answered since the research on the topic is still at early stage. 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. In the current data set we experienced only a fraction of possible problems. So far, the supply of real image data is limited due to technical reasons. We believe that with more data we will be able to generalise our approach using tools such as more compound time series analysis.

Acknowledgement

Authors wish to thank Professor Andrzej Skowron who's remarks and advices helped in many ways to improve this study.

This work has been supported by the Wallenberg Foundation – WITAS Project and by the grant 8T11C 025 19 from the Polish State Committee for Scientific Research (KBN).

References

1. Ambroise C., Dang M., Govaert G. , Clustering of Spatial Data by the EM Algorithm. In geoENV I-Geostatistics for Environmental Applications, A. Soares et al. (eds), Kluwer Academic Publishers, Dordrecht, 1997 pp. 493–504.
2. Backer E., Kandel A., Computer-Assisted Reasoning in Cluster Analysis, Prentice Hall, New York, 1995.
3. Celeux G., Govaert G. A classification EM algorithm for clustering and two stochastic versions, Computational Statistics and Data Analysis, 14, 1992, pp. 315 – 332.
4. Doherty P., The WITAS Integrated System Architecture, Linköping Electronic Articles in Computer and Information Science, Vol. 4 No. 17, Linköping University Electronic Press, Linköping, 1999 `http://www.ep.liu.se/ea/cis/1999/017/`
5. Doherty P., Kachniarz J., Szałas A., Meta-Queries on Deductive Databases, Fundamenta Informaticae, Vol. 40 No. 1, 1999, pp. 17–30
6. Firby J.R., Prokopowicz P.N., Swain M.J., The Animate Agent Architecture, In: Kortenkamp D., Bonasso R.P., Murphy R.(eds.), Artificial Intelligence and Mobile Robots: Case Studies of Successful Robot Systems, chapter 8, AAAI Press/The MIT Press, Cambridge MA, 1998
7. Frank E., Witten I.H., Data Mining: Practical Machine Learning Tools and Techniques with Java Implementations Morgan Kaufmann, San Francisco, 1999
8. Three-layer Architectures, In: Kortenkamp D., Bonasso R.P., Murphy R.(eds.), Artificial Intelligence and Mobile Robots: Case Studies of Successful Robot Systems, AAAI Press/The MIT Press, Cambridge MA, 1998, pp. 195-210
9. Komorowski J., Pawlak Z., Polkowski L., Skowron A., Rough sets: A tutorial. In: S.K. Pal and A. Skowron (eds.), Rough fuzzy hybridization: A new trend in decision-making, Springer-Verlag, Singapore, 1999, pp. 3–98.
10. Nguyen H.S., Skowron A., Szczuka M., Situation Identification by Unmanned Aerial Vehicle, Proceedings of CS&P 2000, Informatik Berichte, Humboldt-Univerität zu Berlin, Berlin, 2000
11. Pawlak Z., Rough Sets: Theoretical Aspects of Reasoning about Data, Kluwer, Dordrecht, 1991
12. The Rosetta Homepage at `http://www.idi.ntnu.no/~aleks/rosetta`
13. The RSES Homepage at `http://logic.mimuw.edu.pl/~rses`
14. The WITAS project Homepage at `http://www.ida.liu.se/ext/witas/`