

## Loughborough University Institutional Repository

---

# *Colour based semantic image segmentation and classification for unmanned ground operations*

This item was submitted to Loughborough University's Institutional Repository by the/an author.

**Citation:** COOMBES, M., EATON, W.H. and CHEN, W.-H., 2016. Colour based semantic image segmentation and classification for unmanned ground operations. International Conference on Unmanned Aircraft Systems (ICUAS 2016), Arlington, VA USA, 7th-10th June 2016, pp. 858-867.

### **Additional Information:**

- © 2016 IEEE. Personal use of this material is permitted. Permission from IEEE must be obtained for all other uses, in any current or future media, including reprinting/republishing this material for advertising or promotional purposes, creating new collective works, for resale or redistribution to servers or lists, or reuse of any copyrighted component of this work in other works.

**Metadata Record:** <https://dspace.lboro.ac.uk/2134/22729>

**Version:** Accepted for publication

**Publisher:** © IEEE

Please cite the published version.

# Colour Based Semantic Image Segmentation and Classification for Unmanned Ground Operations

Matthew Coombes, William Eaton, Wen-Hua Chen

Department of Automotive and Aeronautical Engineering

Loughborough University, Loughborough, LE11 3TQ UK

Email: {M.J.Coombes@lboro.ac.uk, W.H.Eaton@lboro.ac.uk, W.Chen@lboro.ac.uk}

**Abstract**—To aid an automatic taxiing system for unmanned aircraft, this paper presents a colour based method for semantic segmentation and image classification in an aerodrome environment with the intention to use the classification output to aid navigation and collision avoidance. Based on previous work, this machine vision system uses semantic segmentation to interpret the scene. Following an initial superpixel based segmentation procedure, a colour based Bayesian Network classifier is trained and used to semantically classify each segmented cluster. HSV colourspace is adopted as it is close to the way of human vision perception of the world, and each channel shows significant differentiation between classes. Luminance is used to identify surface lines on the taxiway, which is then fused with colour classification to give improved classification results. The classification performance of the proposed colour based classifier is tested in a real aerodrome, which demonstrates that the proposed method outperforms a previously developed texture only based method.

**Keywords**—Unmanned Ground Operations; Semantic Segmentation; Image Segmentation; Superpixel; Colour Classification; Bayesian Network

## I. INTRODUCTION

Although Unmanned Aerial Systems (UASs) are commonly associated with their roles in military and defence, there are also many potential civil applications for UAS. Unmanned aircraft could bring huge benefits in areas such as search and rescue, border security, law enforcement, aerial photography and environmental monitoring. As such, there is a large drive for UAS to be integrated within the national airspace system [1] [2].

Small scale UAS will likely be easy to integrate, as they will operate with minimal interaction with existing airspace users. However, as civil UAS proliferate, they are likely to grow larger and become more similar to manned aircraft. As current large UAS are predominantly military, they are commonly operated out of segregated aerodromes. However for civil operations, it would be impractical to operate ground facilities solely for unmanned aircraft. Therefore, it is expected that future UAS will operate from existing civil aerodromes, alongside manned aircraft.

Despite these intentions, current UAS are often lacking even basic automated capabilities. Without the need to interact with other users, current UAS generally employ unconventional taxiing methods, such as manual retrieval, which are incompatible with civil operations. This current inability to operate in non-segregated aerodromes represents a large barrier to bringing UAS into the National Airspace System (NAS). In addition to operating in new

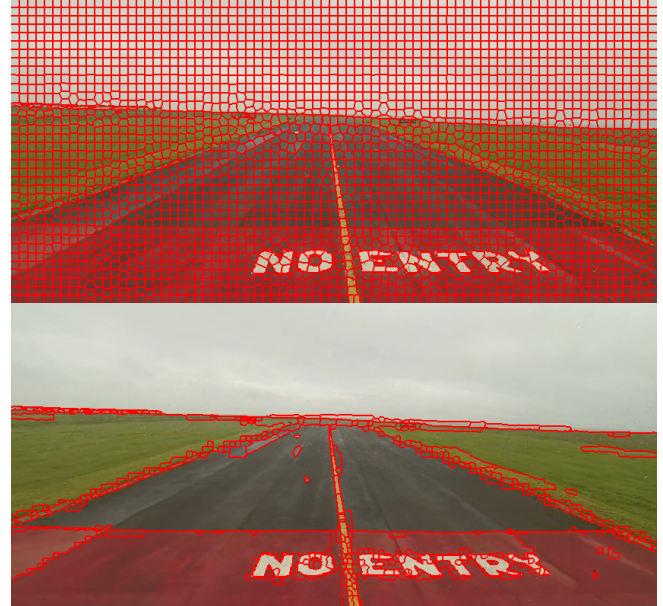


Fig. 1: Upper picture showing superpixel Oversegmentation and the lower image showing Final Untrained segmentation performed on example runway image

environments, future UAS are also likely to operate with less flight crew. It has previously been identified that there is a large economic benefit for a single operator to command multiple UAS, during both ground and air operations [3]. As directly controlling multiple craft is impossible, a shared-autonomy approach (Such as Global Positioning System (GPS) driven taxiing with human collision risk detection) is also insufficient. Therefore, a great number of robust functions are required. For this work, focus has been placed on the essential requirements of accurate ownship localisation and collision risk detection.

This work has been undertaken in conjunction with BAE systems, who have provided the practical test data used for validation. As a result, the system proposed in this paper is constrained by the availability of sensors on the test-platform. Matching the current BAE UAS hardware, this work assumes that direct sensing capabilities are limited to a single forward facing monocular camera. As such, a machine vision approach is the only feasible method of direct sensing.

### A. Machine Vision

For collision risk detection, the forward facing camera must be used to detect and identify any objects in front of the

UAS. Although a library of potential collision risks could be assembled for object detection, there would always remain the possibility of an unexpected object posing a collision risk. Therefore, a generic collision risk detection method is more suitable. As aerodromes are designed to minimise risk to taxiing aircraft, the majority of the environment will have minimal objects with which to collide. As such, the camera image will commonly only consist of terrain features; such as asphalt and grass. If these surfaces types can be reliably identified within an image, any remaining regions will therefore not represent terrain and can therefore be considered a collision risk. Therefore, the purpose of the machine vision system for this work is to robustly classify a small number of expected classes. Any remaining regions within an image which are not classified with a high confidence can be regarded as a collision risk. (It should be noted that whilst this paper shows methods of robust colour-based classification, the techniques used to identify collision risks are detailed in a sister paper [4]).

As any part of the image could represent a collision risk, a whole-image classification approach is required. This has been achieved using a process of semantic segmentation, whereby an image is divided into visible distinct regions (i.e. segmented) and each region is assigned a high level semantic classification (e.g. taxiway, surface marking). This work continues from the initial research conducted in [5]. In this previous work, a semantic segmentation approach was implemented, with image texture used as the main feature for classification. Despite good results, any texture based-approach is computationally intensive, making real-world application more difficult. As such, this paper intends to use the same segmentation method but also seeks to identify if colour data alone is sufficient as the primary feature for classification.

The classification process is to be achieved using a Bayesian Network (BN) approach. BNs provide many advantages compared to the naïve data fusion approach used in [5]. For each type of feature used for classification, the data is original expressed in a function unique to each data type. BN make use of probability so that incompatible sources can be fused together to achieve better results. For this work, Hue-Saturation-Value (HSV) colour data is used alongside Normalised Relative Luminance (NRL). This novel implementation is intended to improve the detection and classification of surface markings.

## B. Paper Contents

The remainder of this paper is organised as follows; section II discusses image segmentation, and how untrained segmentation is performed using methods explored in a previous work [5]. Section III explains the proposed HSV based colour classifier. Section III-B details how luminance is used to provide yellow and white line detection. Section IV introduces the BN data fusion method proposed in this paper, which is followed by a brief introduction to BNs. Section V gives details on the final image classifying BN, including the network's structure, inputs and discretisation. Section VI compares the performance of the current texture-only classification method compared to the proposed BN methodology, using a test set of labelled aerodrome images.

## II. IMAGE SEGMENTATION

Modern semantic segmentation techniques commonly attempt to perform both image segmentation and classification simultaneously. By simply categorising pixels at a small scale (either individually or in small groups), larger regions within an image are formed where many neighbouring pixels share the same class. As the final segmentation is based directly on classification results, this method is commonly referred to as classifier-led-segmentation. Without the need for separate stages, data is extracted from each pixel only once, making such techniques highly efficient. As such, this approach has proven popular, with classifier-led-segmentation commonly used to rapidly categorise images for 'semantic retrieval'. This enables search engines to match images without requiring manual descriptions of image contents [6].

Due to the complexity of fusing many different information types, non-deterministic approaches, such as Artificial Neural Network (ANN), are commonly used. A nearly ideal output for this work was achieved in [7], which made use of an ANN to generate segmented images for urban scene recognition, specifically for Self-Driving Car (SDC). However, regulatory approval of non-deterministic methods (such as ANN) for aerospace use is unlikely. The use of non-deterministic systems has already been identified as a potential barrier to entry for new technology. From [8],

*“Existing adaptive/ non-deterministic algorithms have not been widely applied to safety-critical civil aviation applications in part because of the lack of a mature process for designing, implementing, and testing such algorithms.”*

As such, non-deterministic methods are difficult to propose for use on an UAS when there is no certainty that they could ever be used. Consequently, this work differs from the majority of modern semantic segmentation techniques as it only makes use of using deterministic techniques.

In addition, segmentation and classification are undertaken as separate stages. Prior to ANN becoming common, semantic segmentation often made use of distinct segmentation and classification stages, such as [9]. Rather than use pixel classification to define regions, segmentation is instead achieved using basic low-level image features. As prior knowledge of the image contents has no bearing on the segmentation results, such methods are commonly known as 'untrained segmentation'. Classification can then be applied to each region individually.

This work continues to use the segmentation approach outlined in [5]. This is a two-stage process which begins with an initial 'superpixel' clustering. The term 'superpixel' refers to a small cluster of pixels, grouped together based on their colour and spatial distance. Superpixels are becoming increasingly popular for use in computer vision applications as they can significantly reduce computational load. As only very similar pixels are clustered together, the mean data for a superpixel is highly representative of each pixel it contains. With every pixel in the original image grouped into a superpixel, an image with millions of pixels can be reduced to a meaningful representation of only a few hundred superpixels. Apply machine vision techniques at the superpixel level can then produce similar results but far faster than equivalent pixel-based processes.

As in [5], this work uses Simple Linear Iterative Clustering (SLIC) to generate superpixels [10]. SLIC is currently

the foremost superpixel generation algorithm, capable of generating superpixels of regular shape and size, whilst also minimised colour variance within each superpixel. The efficiency of the algorithm also makes it very practical, with a standard desktop computer providing sufficient computational power for SLIC to process high resolution images in real time [11]. The upper half of Fig. 1 is an example of superpixel segmentation achieved using SLIC.

As superpixels are spatially constrained, the end result is always a significant over-segmentation. This minimises the risk that any superpixel will constrain more than one category of object, whilst also ensuring that superpixel boundaries conform to the boundaries of objects within the image. However, an over-segmentation also introduces many borders which have no presence in the original image. For classification, larger clusters are preferable, as each cluster has more data available for the classifier. In addition, fewer regions to classify increases computational speed. Therefore, the over-segmentation is resolved by a second application of clustering, grouping superpixels into larger, visually similar regions.

This secondary clustering is achieved using the Density-Based Spatial Clustering of Applications with Noise (DBSCAN) algorithm, as outlined in [12]. DBSCAN is a method of clustering that uses density reachability, so that dissimilar superpixels can be merged if connected neighbours between them are similar enough to form a path. This is essential to overcome the effects of distance on large regions such as taxiways, which introduce colour and texture gradients proportional to physical distance.

The metric used to compare superpixels is colour distance within the CIE  $L^*a^*b^*$  (CIELAB) colourspace. As the original superpixel boundaries formed by SLIC are preserved during DBSCAN clustering, the final result retains the sharp resolution required for 3D estimation. Shown in Fig. 1 is an example of DBSCAN clustering of the super pixels from Fig. 1. By varying the threshold used by DBSCAN based on empirically inspected results, each output cluster should ideally represent a single object or material type. This makes the scene easier to interpret, as all data within each cluster can be used during classification.

### III. COLOUR CLASSIFICATION

In the previous work undertaken in [5], image texture was used as the primary feature for region classification. Unlike colour data, where mean values can effectively represent a region, texture data is the variance of colour values over multiple pixels. This necessitates the use of complex feature descriptors to enable comparisons for classification.

As such, texture data is extremely data intensive. Even the most simplified results are far more complex than colour, and are significantly more computationally expensive to extract. Therefore, this paper seeks to identify if colour data alone is sufficient for semantic segmentation. Although [5] did make use of colour data, it was mainly used to increase classification confidence, or differentiate between classes with similar texture data. For example, differentiating between asphalt and painted surface markings is most easily achieved using colour.

A simplified approach to semantic segmentation is required. As the intention is to infer collision risks by identifying terrain types, the number of classes is limited to only

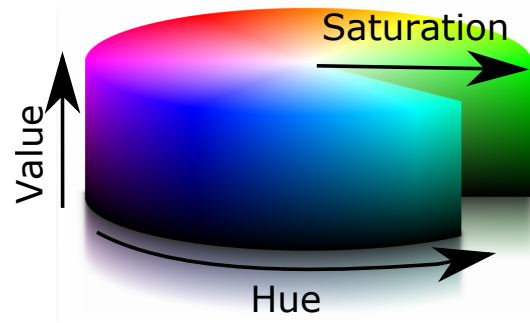


Fig. 2: HSV cylindrical colourspace

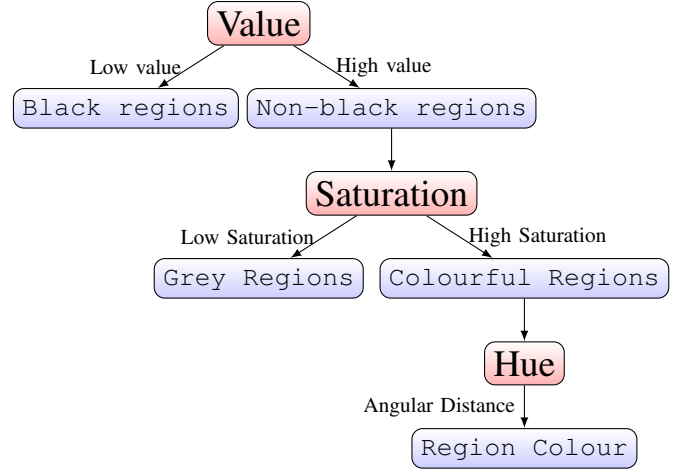


Fig. 3: Colour extraction process within HSV

the most common aerodrome features. These are; asphalt, grass, painted surface markings (red, white and yellow) and sky. As each of these classes have distinct colours, they can be differentiated using colour alone. Similarity to known examples of each class can then provide a measure of confidence which can be passed to the BN. As multiple objects can share the same colour, additional information must be used in the BN to help differentiate objects.

In machine vision, the word ‘colour’ refers to the wide array of different methods of representing pixel values. As different representations of colour can make processing far easier, the rest of this section is divided between conventional segmentation using HSV, and a novel application of NRL, designed to improve the detection of taxiway surface markings.

#### A. Hue, Saturation and Value

For this work, colour classification is primarily achieved within the HSV colourspace, which is commonly used for image classification [13]. Unlike Red-Green-Blue (RGB), HSV is designed to make human interpretation easier, separating colour data into channels reflecting how human vision functions. The main benefit is that the image intensity (*Value*) is separated from the colour information. This separation allows colour information to be more robust to changes in lighting. In addition, as *Hue* is defined within a single channel, colour distance calculations are very simple.

Rather than use Cartesian co-ordinates, HSV uses a cylindrical colour-coordinate system, as shown in Fig. 2. This allows for intuitive rules to be established for classification.

In practice, the BN classifier processes data from all three HSV colour channels simultaneously. However, the logic behind HSV classification is more easily thought of as a sequential process as shown in Fig. 3.

The ‘first’ channel to process is *Value*, which represents distance from black within the cylindrical co-ordinates. Unlike luminance (which is true distance from black), the colour primaries (i.e. red, green and blue) and pure white will all share the same maximum *Value*, despite primary blue only being around 10% as bright as pure white. This is extremely useful, as it provides clear distinction between dark coloured pixels and black, such as for grass and asphalt. As all other classes are typically bright in colour, low *Value* is considered indicative of asphalt.

The *Saturation* colour channel denotes the intensity of colour and is therefore applicable to the classification of every region which is not identified as black. Colourful objects have high *Saturation* values whilst grey objects have low values. Yellow and red surface markings, and grass were all expected to be colourful whilst white surface markings and asphalt were not. To align with testing conditions, sky was assumed to be grey to reflect the overcast skies typical of UK weather.

As both *Value* and *Saturation* are highly influenced by light levels and atmospheric conditions, the colour of a single object will vary dramatically during a single day. Therefore, it is extremely difficult to define typical levels for each class. Instead, the tendency of the data to either have high or low *Value* and *Saturation* has been used. For use in the BN, probability is assigned proportional to the distance from the appropriate end of each colour channel. I.e. regions with very low *Value* have a high probability of being asphalt, while regions with high *Value* have an equal probability of being any other class. The only class which does not use *Value* during classification is ‘Red surface markings’, as the *Value* range has been found to be too great to be meaningful.

The final colour channel is *Hue*, which is used to differentiate regions with high *Saturation* and *Value*. Only surface markings (red and yellow) and grass use *Hue* during classification. As *Hue* represents the angular position within the cylindrical colourspace, confidence is derived using angular distance from known hue values for each class. As human perception of colour varies with only slight changes in *Hue*, upper and lower threshold limits are imposed to judge whether *Hue* criteria had been met, rather than distance from a singular value. This distance is then converted into probability through a Gaussian Probability Density Function (PDF) for use in the BN.

### B. Surface Marking Detection

As relatively few classes have been selected, the majority of classes can be differentiated using just the simplistic HSV approach outlined above. However, low saturation classes (i.e. ‘grey’) are more difficult. If the BN classification were to be performed using the above data alone, it would be difficult to differentiate between asphalt, white surface markings and sky. Therefore, additional information is required to differentiate between them. As sky clusters are constrained in terms of position (i.e. they always appear above the horizon line), they are the simplest to classify. For this work, an horizon detection algorithm is used and high-*Value* low-

*Saturation* regions which appear wholly above the horizon are considered to be sky.

Differentiating between white surface markings and asphalt is more complex, as the perceived intensity of colours is always influenced by other colours within the same scene. In low light, the white surface markings may be less bright than the asphalt in direct sunlight. It is only when they are compared to neighbouring clusters that their brightness becomes apparent. Therefore, comparison between pixel clusters is the easiest way to differentiate between them. Although relative *Value* could be used, as aforementioned, the *Value* channel is not a direct assessment of brightness, but simply the distance of each colour from black. As such, regions of the image which are perceptually very bright will not be well represented, as they will be similar in *Value* to any region of pure colour. As surface markings are often worn, an extremely sensitive measure of brightness is required.

The intensity of light is commonly referred to as *Luminance*. As image data is used, automatic white-balancing and aperture effects prevent the actual luminance value from being used. Instead, relative luminance of all pixels within the image is more common. As clusters of sky pixels will always be brighter than ground pixels [14], the relative luminance within an image is typically based on the sky. As both asphalt and surface markings are found on the ground, this work suggests the approach of using the horizon detection algorithm to redefine the maximum brightness.

The perceived difference between surface markings and asphalt is increased by re-normalising the luminance values within the image, relative to the maximum luminance of any region on the ground. This new image, which we refer to as NRL, then represents a fairly consistent measure of the brightness of pixels on the ground. A benefit of this approach is that as NRL strongly emphasises what appears bright to human eyes, it is also highly effective in detecting yellow surface markings. Relative luminance can be derived from RGB colourspace using Eq. 1, where R, G and B are the respective mean pixel values in each colour channel per cluster, and Y is the relative luminosity of each cluster/

$$Y_i = 0.2126\bar{R}_i + 0.7152\bar{G}_i + 0.0722\bar{B}_i \quad (1)$$

NRL can then be calculated using Eq. 2.

$$NRL_i = \frac{Y_i}{Y_{max}} \quad (2)$$

As relative luminance is highly sensitive to hue, atmospheric effects can have a large influence in the effectiveness of NRL over great distance. This can be mitigated by combining cluster luminance information with cluster distance to camera information, and is explained in greater detail in section V-B.

## IV. BAYESIAN NETWORK AND DATA FUSION

For classification, a technique is required that enables the fusion of metrics which are not directly comparable. From the HSV and surface marking classifiers previously described, discrete class, continuous NRL and realworld-distance estimates must all be combined to produce a final classification. For this work, the decision has been made to use Bayesian Networks (BNs), which are commonly used

to represent knowledge and reasoning under uncertainty. They are built around a probabilistic graphical model, which represents a set of random variables and their conditional dependencies. As such, BNs can provide comparisons using any types of data, provided the result can be expressed in terms of probability.

In addition to a discrete class estimate, the final output from a BN includes a probability for each cluster. As this final value incorporates all previous information, a simple threshold can be applied below which all clusters are simply considered unknowns, and therefore potential collision risks. This provides a simplistic method of tuning the classifier, should a higher degree of confidence be required.

The use of BN for semantic segmentation is well established. However, the majority of approaches use BN within a classifier-led-segmentation effort (such as [15] and [16]), rather than the independent classification stage used here.

#### A. Data Discretisation

As a discrete BN implementation is used, the input variables require discretisation before use. Due to HSV data being stored using 8-bit integers, it is already discretised. However, the 16 million potential combinations are far too numerous to be of use. Instead, each of the colour channels is discretised into far larger bands. As *Hue* represents the angular position within the cylindrical colourspace, it is discretised in to 24 discrete states for 0 - 360° in 15° increments. Both *Saturation* and *Value* are simply assigned 10 states, in increments of 0.1 between 0 and 1. As *Hue* and *Saturation* are difficult to perceive at low *Value*, a further simplification is made that all colours below 10% *Value* are considered black. This produces a colourspace with 2161 discrete colours in total. An image converted to show these discrete colours is shown in Fig. 5

TABLE 1: Relative NRL discrete states

NRL states	
High	$NRL_i > 0.77$
Medium	$0.65 < NRL_i < 0.77$
Low	$NRL_i < 0.65$

NRL values are discretised differently. Rather than use uniform bands, the NRL data is divided in to three states *High*, *Medium* and *Low*. The thresholds for each state have been determined empirically, to best capture the results. Obvious surface markings are captured within the *High* state, worn and faded lines are captured within the *Medium* state and all other clusters are assigned the *Low* state. The values used for this work are shown in Table 1, having been determined empirically from test images. *Dist* is the discretised distance of the cluster from the camera, which can have the states *Close*, *Mid* and *Far*, the use for this will be explained in the next section.

#### V. BAYESIAN NETWORK STRUCTURE

There are three parts to a Bayesian network; a Directed Acyclic Graph (DAG), a Conditional Probability Distribution (CPD) for each node on the DAG, and an inference engine used to solve the network.

The DAG for the proposed BN is shown in the lower part of Fig. 4. The BN consists of two sub-networks; one

for HSV colour classification and another for surface line detection. Each sub-network provides information to the final *Class* estimation node, which fuses the data together. The network is individually applied to each cluster produced during the segmentation phase, with the input variables (*Dist*, *NRL*, *H*, *S* and *V*) extracted directly from each cluster. The final output is a probability of the cluster belonging to each class, with the highest probability indicating the most likely class. This is summarised in Eq. 3, where  $c^i$  is the class assigned to cluster  $i$ .

$$c_i = \arg \max_{Class_i} P(Class_i | H_i, S_i, V_i, Dist_i, NRL_i) \quad (3)$$

In order to complete the network, the CPDs need to be determined. For each node, the CPD states the marginal probability of a each variable with respect to the other variables within that node. Where suitable, parameter estimation techniques have been used to calculate the CPDs. For example, despite the significant decrease from 16 million to 2161 colours for HSV colour classification, there are still far too many discrete colours for manual conversion.

Parameter estimation is common in BN design. Unlike the intuitive network structure, some numerical parameters are harder to elicit from human experts. To this end, in literature a number of methods have been developed to estimate the parameters for both complete and non-complete data. As this BN is not overly complex and has complete data, it is possible to provide manual classification for every cluster in the training set. This allows Maximum Likelihood Estimation (MLE) parameter estimation to be used. MLE selects the set of values of the model parameters that maximizes its likelihood function. MLE parameter estimation has been used previously for image classification, and has been demonstrated to improve performance in skin detection [17]. The application of MLE to Bayesian networks is explained in detail in [18].

In order to minimise complexity, each CPD is trained within it's sub-network, reducing the number of examples required for each training set. The final class estimate CPD has been manually filled out, as it fuses the two sub-networks using logic that can be easily applied by an expert. Table 2 contains the CPD for the final node, *Class*, showing exactly how the fusion is performed.

When the HSV colour classifier indicates the *Colourclass* is a neither white nor yellow surface markings, and *Line* is *F* (false) the probability of the final *Class* being that colour class is 1. However if line is *T* (True) then the probability of it being one of those non-line classes is reduced to 0.1 and there is a probability of it being a white or yellow line of 0.45 each. In a similar way, If the colour class is either a white or yellow line and *Line* is true then the probability is 1. If *Line* is *F* then this probability is reduced to 0.1, and then all other classes will have an equal 0.225 probability. The 0.1 probability for both cases is to say that even though the line classifier contradicts what the colour classifier is saying, there still is a chance the line estimation is incorrect. As it is only 0.1 this shows that there is a great deal of confidence in the line estimate over the colour classifier, which is due to NRL being a very robust identifier.



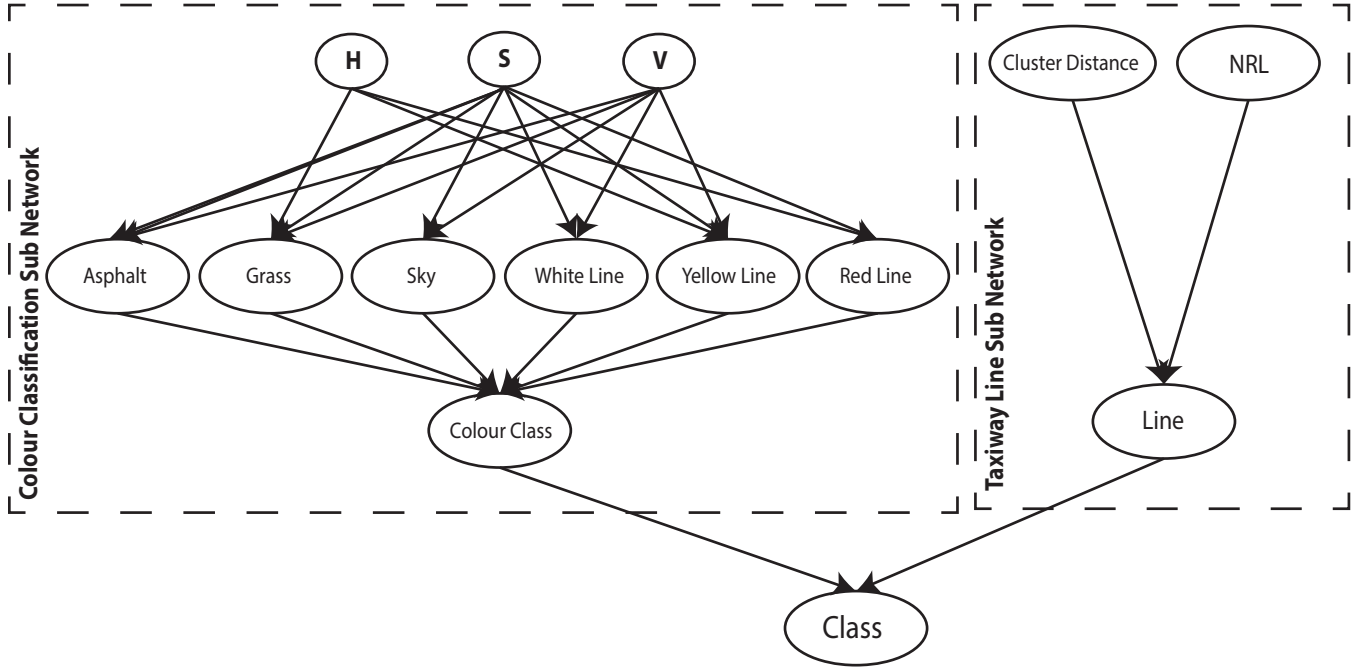


Fig. 4: Full Bayesian Network

TABLE 2: Class CPD  $P(Class|ColourClass, Line)$

Colour Class	Line	Asphalt	Grass	Sky	White	Yellow	Red
Asphalt	T	0.1	0	0	0.45	0.45	0
Grass	T	0	0.1	0	0.45	0.45	0
Sky	T	0	0	0.1	0.45	0.45	0
White	T	0	0	0	1	0	0
Yellow	T	0	0	0	0	1	0
Red	T	0	0	0	0	0	1
Asphalt	F	1	0	0	0	0	0
Grass	F	0	1	0	0	0	0
Sky	F	0	0	1	0	0	0
White	F	0.225	0.225	0.225	0.1	0	0.225
Yellow	F	0.225	0.225	0.225	0	0.1	0.225
Red	F	0	0	0	0	0	1

#### A. Bayesian Colour Classification

The HSV sub-network based on HSV colour is shown on the left side of Fig. 4. The six possible classes are represented by individual nodes, for which MLE has been used to train the CPDs. Each node has two discrete states: true ( $T$ ) and false ( $F$ ). As discussed in Section III-A, every colour channel may not be useful for every class. Therefore, the edges indicate which classes are conditioned using *Hue*, *Saturation* and/or *Value*.

The node *colourclass* is a hidden node which simply combines the individual true/false probabilities into a single node. This simplifies the network, making it easier to observe the output of the colour classifier, as all classes can be compared in a single node. In addition, this also makes the CPD of the final *class* estimate node much simpler, as it will have only a single parent.

An example of the classifier output is shown in Fig. 5. In this figure, the top image shows a typical aerodrome scene which has undergone discretisation in HSV. The class probabilities of each cluster are calculated from  $P(Colourclass|H, S, V)$ . This is the marginal probability distribution of *Colourclass* with  $H, S, V$  entered as evidence. As before

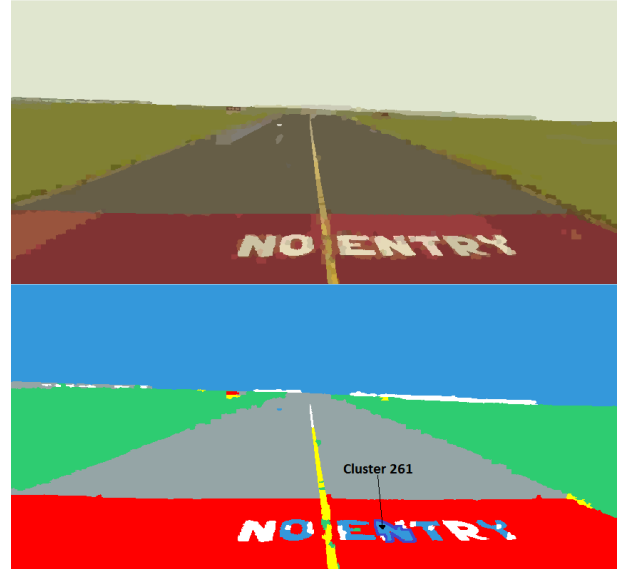


Fig. 5: Discretised HSV colour image and subsequent Bayesian network colour classifier output

$c_i = \arg \max_{ColourClass_i} P(ColourClass_i|H_i, S_i, V_i)$  is the chosen class for that cluster. The lower half of Fig. 5 depicts the classifier output.

Using only the HSV colour classifier, the percentage of correctly classified pixels for the example image is 95.6 %. From Fig. 5 it can be clearly seen that the largest source of error is the misclassification of white surface markings as sky. This is due to the two classes sharing the same discrete colour and performing classification without any additional context. A more detailed discussion on this misclassification is presented in Section VI-A, referring specific to 'cluster 261', as labelled in Fig. 5.

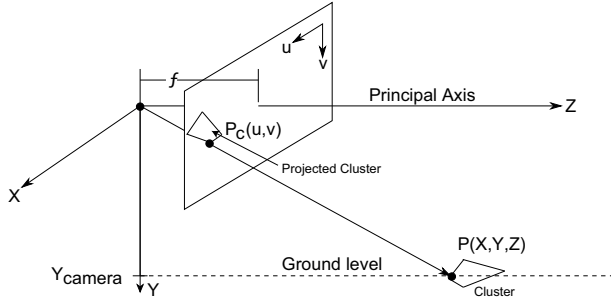


Fig. 6: Pinhole camera model used for depth estimation

### B. Normalised Relative Luminance Line Detection

The second sub-network is specifically intended to provide an additional probability of a cluster being either a white or yellow surface marking. This sub network is shown on the right side of Fig. 4, with the variable *Line* representing this probability. As with the discrete classes within the HSV sub-netowrk, *Line* has two states, true (*T*) and false (*F*).

As mentioned in Section III-B, atmospheric effect can affects NRL over distance. On clear days, Rayleigh scattering scatters blue light more than red, lowering the NRL values of objects in the distance. Conversely, on overcast/rainy days, the presence of water droplets in the air scatters all light wavelengths equally. As such, all colour channels increase, with objects in the distance tending to have a higher NRL value. As the test footage used for this paper was taken on a wet, overcast data, NRL values increase at extreme distance. As a results, classes which would have *Low* NRL states in the foreground gain *Medium* or *High* states when far from the camera.

A different example image in Fig. 7a shows the NRL values for each cluster, clearly indicating the suitability of NRL for surface marking detection. This figure also shows the atmospheric effects created by rain, as the areas highlighted with red circles have high NRL values despite being grass and asphalt. To mitigate these atmospheric effects, clusters in the distance that have a higher NRL values need to be given a much lower probability of being a line. This is easily achieved using the BN structure shown in Fig. 4, where cluster NRL and cluster distance from camera (*Dist*) are combined together.

As only distant clusters are affected significantly, precise distance estimation is not required. Therefore, for simplicity of concept, distance to cluster is approximated using the pinhole camera model. For any point captured by a pinhole camera, similar triangles can be used to map between the 3D position of the point  $P(X, Y, Z)$ , and the position of the point within the image  $P_c(u, v)$  as shown in Fig. 6. Given the focal length  $f$  and the height of the camera above the ground  $Y_{camera}$ . The ground position  $X$  and  $Z$  to the base of the cluster is calculated below. Using simple trigonometry the ground distance  $D_c$  to the cluster can be calculated.

$$\begin{aligned} Z &= \frac{f \cdot Y_{camera}}{v} \\ X &= \frac{u \cdot Y_{camera}}{v} \end{aligned} \quad (4)$$

To be used within the BN, cluster distance must be discretised. This is achieved in a similar way to NRL, with

*Dist* discretised in to three states: *Close*, *Mid* and *Far*, as shown in Table 3.

For the same example image as in Fig. 7a the discrete states for each cluster are displayed in Fig. 7b. In ellipse 1, there are several distant clusters which have high NRL values due to the weather. By introducing distance, these clusters are in the *Far* state, and therefore will no longer have any chance of being misclassified as a line.

Ellipse 2 demonstrates a more difficult result, as the clusters have similar NRL values to actual surface markings at the same distance. Therefore, a *Mid* state simply reduces the probability of being a line. As actual lines generally have far higher NRL values, this is seen as an effective solution. The CPD of the *Line* node is compiled to represent these relationships and is shown in Table 4.

TABLE 3: Dist discrete states

Dist states	
Close	$D_c < 20m$
Mid	$20m < D_c < 55m$
Far	$D_c > 55m$

Fig. 8 depicts the final probabilities of clusters being considered surface markings, based on the sub-network alone. Actual surface markings are clearly well defined. Although some non-surface marking clusters have none-zero probabilities of being a line, the actual probability remains low at 0.25 which will not affect the final classification unduly.

## VI. RESULTS

In order to evaluate the proposed methods, a specific aerodrome data set has been created. Video footage and avionics data has been collected during mock-taxiing performed at Walney Island Airport (WIA), in the UK. For semantic segmentation, several hundred video frames underwent manual segmentation and classification, to produce data for training and validation.

The data-set scenario provides a realistic and visually challenging scene, due to the effects of weather and the aerodrome itself. The data set was created over several days, during a period of inclement weather. This directly influences image contents, with many puddles and rain present within the test images. The overcast skies also affect the images as a whole, with poor lighting conditions reducing the *Saturation* and making classification more difficult. In addition, the taxiway surface at WIA is aged and worn, with inconsistent surface colours where repairs have been made. Both asphalt and surface markings vary dramatically throughout the test footage.

The overall classification results have been created using 20 images captured from WIA but not used for training.

TABLE 4: Line CPD  $P(\text{Line}|\text{NRL}, \text{Dist})$

NRL state	Dist state	True	False
Low	Close	0	1
Medium	Close	0.75	0.25
High	Close	0.9	0.1
Low	Mid	0	1
Medium	Mid	0.55	0.45
High	Mid	0.8	0.2
Low	Far	0	1
Medium	Far	0.2	0.8
High	Far	0.6	0.4





TABLE 5: Marginal posterior distribution for *Class* and *ColourClass* for cluster 261

Class	Colour Class	Class
Asphalt	0.0159	0.0051
Grass	0.0159	0.0051
Sky	0.567	0.125
White Line	0.3694	0.542
Yellow Line	0.0159	0.3177
Red Line	0.0159	0.0051

TABLE 6: Percentage error for Bayesian network classifier and colour classifier compared to texture only classifiers, using the test set

Classifier	Mean % pixel misclassification	Mean % cluster misclassification
BN	2.18 %	15.341 %
Colour	2.76 %	24.34 %
LBP	6.29 %	52.81 %
MR8	5.12 %	57.67 %

the correct winning class. The marginal posterior distribution for *Class* is shown in Table 5.

### B. Classification Accuracy

For comparison, the previous texture classification methods used in [5] have been applied to the same data set, using both Maximum Response Filters (MR8) and Local Binary Pattern (LBP) texture descriptors. As this previous method has been shown to give performance adequate for navigation or obstacle detection, this will be used as a comparison for performance when using colour data alone.

The average percentage error for the test set is presented in Table 6, comparing the previous texture based approach against the BN classifier at both a pixel and cluster level. An obvious issue with the texture based classification method is indicated by the large difference between pixel-level error and cluster-level error. This is to be expected, as the size of feature descriptors used in texture based classification often result in the misclassification of small clusters. As colour information is entirely constrained by the boundaries of each cluster, this same issue is not present in the BN classifier. In fact, despite using less data, the BN approach shows a performance increase over the texture classifier. As expected, this is especially true for small clusters, as can be seen from a 4% improvement at the pixel level, but a more significant 40% improvement in cluster classification.

The benefits of including NRL are also apparent. Despite the colour classifier sub-network (*ColourClass*) performing better than the final result of the texture based method, further improvement is made when NRL data is included. Through data fusion, the performance of the full BN is increased by 0.6% per pixel and by 9% per cluster.

It is also useful to compare the classification for individual classes. The breakdown for the LBP, MR8 and BN classifiers are shown in Tables. 7, 8 and 9 respectively. Each row represents the percentage breakdown of the original manually classified class, in terms of the automated segmentation and classification results. The highlighted diagonals represent correct classifications.

Distinguishing between surface markings on texture alone is shown to be significantly difficult, with White and Yellow surface markings often miss-classifying as each other. In addition, red surface marking classification is also very poor for both MR8 and LBP due to the asphalt and red surface texture being very similar.

TABLE 7: Percentage breakdown of LBP Texture only classifier of test set

Man	Auto	Asphalt	Grass	Sky	White	Yellow	Red
Asphalt		95.5	2.7	1.3	0.4	0.1	.003
Grass		3.5	94.5	0	1.8	0.3	0
Sky		0.4	0.6	98.9	0.01	0.01	0
White		4.2	1.7	0	86.5	7.6	0
Yellow		5.4	1.7	0	17.0	76.0	0
Red		86.5	11.2	0	2.0	0.1	0.1

TABLE 8: Percentage breakdown of MR8 Texture only classifier of test set

Man	Auto	Asphalt	Grass	Sky	White	Yellow	Red
Asphalt		99.0	0.5	0	0.2	0.3	.01
Grass		10.1	87.9	0	1.0	0.9	0
Sky		0	0	98.9	1.1	0	0
White		3.5	.03	0	77.3	19.2	0
Yellow		2.5	0	0	2.1	95.3	0
Red		86.5	0	0	0.8	0.9	9.4

Looking at Table 9 it can be seen that using HSV colour and NRL dramatically improves the classification accuracy; increasing correct classification for white, yellow and red surface markings to 99.06%, 92.93% and 97.98% respectively. The biggest remaining error is yellow surface markings misclassified as asphalt for 5.19 % of the pixels. This is a result of worn yellow surface markings having NRL values to low to correct the misclassification from HSV alone.

The Texture only classifiers distinguishes between asphalt and other classes well, with MR8 performing better at 98.9 % compared to 95.5 % for LBP. As the colour classifier only uses H and V to classify asphalt, combining them together in the BN has yielded similar performance at 98.13 %.

As sky typically represents the largest region in each image, it usually has the most information available and therefore has the best correct classification at around 98.9%, for both MR8 and LBP. However small patches of sky, especially those close the horizon, are still affected by the effects of large texture descriptors. As sky is easily distinguishable by colour, these incorrect sky classifications have been completely removed as the BN classifies 100 % of them correctly in Table 9.

The only results which are possibly misleading are related to the detection of grass. In terms of both colour and texture, grass is less consistent than the other classes, with varying blade length and different grass breeds. As such, the texture only performance is not as favourable, at 94.4% for LBP and 87.9% for MR8, compared to the results of the BN classifier at 98.2%. However, as the moisture and chloroform levels within grass determine it's colour, seasonal colour variation is likely to be greatly than texture change. This is not reflected in these results, as both training and testing days were sampled over the course of a few days.

## VII. CONCLUSIONS

In this paper we have presented a method for semantically segmenting images, primarily using HSV colour and a novel implementation of NRL. This is intended to enable automated taxiing of UAS at non-segregated aerodromes, where the segmented images will enable both visual navigation and collision risk detection. A probabilistic BN was used

TABLE 9: Percentage breakdown of Bayesian network classifier of test set

Man \ Auto	Asphalt	Grass	Sky	White	Yellow	Red
Asphalt	98.13	1.32	0.02	0.50	0.00	0.02
Grass	1.74	98.20	0.00	0.01	0.04	0.01
Sky	0.00	0.00	100.00	0.00	0.00	0.00
White	0.42	0.00	0.01	99.06	0.11	0.41
Yellow	0.17	1.46	0.25	5.19	92.93	0.00
Red	1.90	0.11	0.00	0.01	0.00	97.98

to fuse both HSV and NRL to give an accurate estimate of each clusters class. This probabilistic approach enables fusing data which are normally are incompatible, as well as enabling parameter estimation from training data.

By using HSV colourspace for colour classification, different channels can be used on individual classes to better represent their particular colour features. This makes the classification more robust to changes in light or weather conditions. By training the parameters of some of the CPDs of the BN, compiling the network in not only much faster but also includes the ability to tailor the classifier to the particular imaging device, or even the specific aerodrome.

The novel idea of using NRL for surface marking detection is shown to effectively differentiate both yellow and white lines from other aerodrome features. By combining this results with the HSV classification, this method has been shown to improve the classification performance from the texture only classifiers by 5 % per pixel and by a large 40 % per cluster.

Going forwards, there are many extensions that could be performed on this work. As the BN graph structure is intuitive and easily extensible, the texture classifiers could be added to the BN as another source of information. Lighting conditions could be considered by adding time of day variable to the network, and trained with a larger data set with time information.

#### ACKNOWLEDGEMENT

This work was supported by the U.K. Engineering and Physical Sciences Research Council (EPSRC) Autonomous and Intelligent Systems programme under the grant number EP/J011525/1 with BAE Systems as the leading industrial partner. The work greatly benefits from the data set collected from an airfield provided by BAe Systems and technical advice provided by the technical officer Rob Buchanan.

#### REFERENCES

- [1] FAA, "Integration of civil unmanned aircraft systems (uas) in the national airspace system (nas) roadmap," Federal Aviation Administration, Tech. Rep., 2013.
- [2] European RPAS Steering Group, "Roadmap for the integration of civil remotely-piloted aircraft systems into the european aviation system," European RPAS Steering Group, Tech. Rep., 2013.
- [3] Unmanned Aircraft System Roadmap, "Unmanned aircraft system roadmap: 2005-2030," DOD, Tech. Rep., 2005.
- [4] W.-H. C. Matthew Coombes, William Eaton, "Environment recognition for autonomous taxiing through bayesian network based data fusion," in *Submitted to International Conference on Unmanned Aircraft Systems 2016*, 2016.
- [5] W. Eaton. and W.-H. Chen, "Image segmentation for automated taxiing of unmanned aircraft," in *Unmanned Aircraft Systems (ICUAS), 2015 International Conference on*, June 2015, pp. 1–8.

- [6] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, A. Berg, and L. Fei-Fei, "Imagenet large scale visual recognition challenge," *International Journal of Computer Vision*, vol. 115, no. 3, pp. 211–252, 2015. [Online]. Available: <http://dx.doi.org/10.1007/s11263-015-0816-y>
- [7] J. Alvarez, Y. LeCun, T. Gevers, and A. Lopez, "Semantic road segmentation via multi-scale ensembles of learned features," in *Computer Vision ECCV 2012. Workshops and Demonstrations*, ser. Lecture Notes in Computer Science, A. Fusiello, V. Murino, and R. Cucchiara, Eds. Springer Berlin Heidelberg, 2012, vol. 7584, pp. 586–595. [Online]. Available: [http://dx.doi.org/10.1007/978-3-642-33868-7\\_58](http://dx.doi.org/10.1007/978-3-642-33868-7_58)
- [8] T. Warchocki, *Autonomy research for civil aviation : toward a new era of flight*. Washington, D.C: The National Academies Press, 2014.
- [9] J. Shotton, J. Winn, C. Rother, and A. Criminisi, "Textonboost for image understanding: Multi-class object recognition and segmentation by jointly modeling texture, layout, and context," *Int. Journal of Computer Vision (IJCV)*, January 2009.
- [10] R. Achanta, A. Shaji, K. Smith, A. Lucchi, P. Fua, and S. Ssstrunk, "SLIC Superpixels," EPFL, Tech. Rep., 2010.
- [11] C. Y. Ren and I. Reid, "gslic: a real-time implementation of slic superpixel segmentation," *University of Oxford, Department of Engineering, Technical Report*, 2011.
- [12] P. D. Kovesi, "MATLAB and Octave functions for computer vision and image processing," Centre for Exploration Targeting, School of Earth and Environment, The University of Western Australia, available from: <http://www.csse.uwa.edu.au/~pk/research/matlabfns/>.
- [13] C.-J. Du and D.-W. Sun, "Comparison of three methods for classification of pizza topping using different colour space transformations," *Journal of Food Engineering*, vol. 68, no. 3, pp. 277 – 287, 2005.
- [14] K. He, J. Sun, and X. Tang, "Single image haze removal using dark channel prior," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 33, no. 12, pp. 2341–2353, 2011.
- [15] F. Liu, D. Xu, C. Yuan, and W. Kerwin, "Image segmentation based on bayesian network-markov random field model and its application to in vivo plaque composition," in *Biomedical Imaging: Nano to Macro, 2006. 3rd IEEE International Symposium on*, April 2006, pp. 141–144.
- [16] C. Bouman and M. Shapiro, "A multiscale random field model for bayesian image segmentation," *Image Processing, IEEE Transactions on*, vol. 3, no. 2, pp. 162–177, Mar 1994.
- [17] N. Sebe, I. Cohen, T. Huang, and T. Gevers, "Skin detection: a bayesian network approach," in *Pattern Recognition, 2004. ICPR 2004. Proceedings of the 17th International Conference on*, vol. 2, Aug 2004, pp. 903–906 Vol.2.
- [18] D. Koller and N. Friedman, *Probabilistic graphical models: principles and techniques*. MIT press, 2009.