Per-Pixel Classification of Clouds from Whole Sky HDR Images

Pinar Satilmis[a], Thomas Bashford-Rogers[b], Alan Chalmers[c], Kurt Debattista[c]

[a] *Hacettepe University, Beytepe Kampusu, Cankaya/ANKARA, 06800, Turkey*

[b] *University of the West of England, UK*

[c] *University of Warwick, UK*

Abstract

Accurately identifying cloud types in images has multiple uses from meteorological science to computer graphics, especially as clouds are a major factor influencing atmospheric radiative transport. Understanding which cloud types are present in an image is typically performed on a coarse scale, where cloud types are identified per image, but do not permit a finer, per-pixel granularity of labelling cloud types. This paper presents a novel approach which solves this problem via a per-pixel classification method for identifying cloud types based on High Dynamic Range imagery of skies. The proposed method requires minimal labelling of the training data, and utilises a hierarchical patch-based feature extraction technique which describes the statistical and structural features about regions of the image. This enables the extraction of representative feature vectors which are used for subsequent labelling. This approach is the first to produce a per-pixel classification of cloud types from a single image, with an accuracy of 84%. Additionally, when applied to whole sky cloud classification, our results produce a 98.3% accuracy, which is competitive with the state-of-the-art.

Keywords: Cloud-type classification, per-pixel classification

Email addresses: pinarsatilmis@hacettepe.edu.tr (Pinar Satilmis), thomas.bashford.rogers@gmail.com (Thomas Bashford-Rogers), alan.chalmers@warwick.ac.uk (Alan Chalmers), k.debattista@warwick.ac.uk (Kurt Debattista)

1. Introduction

Clouds play a significant role in the radiation of energy from the atmosphere to the surface of the earth, and studying their properties is important across a diverse range of fields ranging from hydrological studies and meteorological science to computer graphics and imaging. A common requirement in these fields is to acquire, classify and process images of clouds in order to understand their structure and composition. This work focuses on the classification stage which labels cloud types in an image. Due to the enormous variety in cloud shapes combined with a limited range of color values, the classification problem for clouds is non-trivial.

Existing approaches for cloud classification summarise images of the whole sky with a single label for instance "Cumulus", "Cirrus", or "Mixed", see Cheng and Yu [1], Li et al. [2]. This does not provide information about the common situation of multiple cloud types present in an image, or the spatial distribution of clouds in an image. This is insufficient for detailed studies on clouds, where ideally a label is assigned per pixel. Per-pixel classification presents a challenge for initial labelling of the training dataset; different cloud types may appear similar in an image leading to human classification errors, and significant effort would be required to assign a label per pixel for the entire dataset.

Images used in cloud classification are typically gathered from two sources; satellite and ground based measurement systems. While satellite images capture a large region of clouds, the resultant images typically have low resolution which precludes investigation of finer cloud details. Ground based measurements are therefore typically investigated in the meteorological literature, e.g. Tapakis and Charalambides [3]. These measurements consist of views of the sky captured with either specialized equipment, see Tapakis and Charalambides [3], Jarraud [4], or with a conventional DSLR camera equipped with a fisheye lens. Our work uses High Dynamic Range (HDR) imagery which stores real-world lighting as floating point values at higher precision than conventional images, see Chalmers and Debattista [5]. This increased precision permits more accurate classification, but is incompatible with existing approaches for cloud classification such as Heinle et al. [6], Kazantzidis et

al. [7], Liu et al. [8], Cheng and Yu [1] further motivating the proposed novel approach.

This work solves the above challenges, and presents a weakly supervised method for per-pixel classification of clouds from HDR images. We propose a method where down-sampling of the input HDR images in the training set produces a hierarchy which is able to represent cloud features at a variety of scales. At each level of the hierarchy, Feature Vectors (FVs) corresponding to patches in the image are extracted and clustered into Representative Feature Vectors (RFVs). Based on an initial coarse labelling of the input HDR images, a histogram of cloud types is associated with each cluster. At run-time, features are extracted from patches in each level of an input image hierarchy, and are compared to the clustered feature vectors. Labelling per pixel is then assigned from merging the histograms associated with the pixel from each level of the hierarchy. The main contributions of the presented method are as follows:

- A method to compute per-pixel labelling of cloud types from input HDR images. This is based on minimal user input, and uses a novel hierarchical histogram merging method to efficiently compute per-pixel labels.
- Results showing a high classification accuracy for per-pixel cloud classification, and state-of-the-art results for whole sky cloud classification.

2. Background and Related Work

There has been a wide range of research on image classification, object detection, object segmentation etc. (Gehler and Nowozin [9], Wang et al. [10], Woźniak and Połap [11], Li et al. [12]). However, direct application of this research to cloud type classification would be impractical because of the unpredictable structure of clouds. Therefore, specific methods have been proposed in literature to classify and detect clouds in images. These methods mainly depend on colour channels (Red (R), Green (G) and Blue (B)), statistical and structural information obtained from ground level or satellite images. Mean, Standard Deviation, Skewness, Smoothness

and Contrast are commonly used statistical methods to extract features from images (Heinle et al. [6], Cheng and Yu [1], Li et al. [2], Calbo and Sabburg [13], Gan et al. [14] etc.). For the structural/textural features, the ratio between the colour channels, Fourier Transform, Discrete Cosine Transform (DCT), Grey Level Co-occurence Matrix ($GLCM$) and Local Binary Patterns are often used (Heinle et al. [6], Liu et al. [8], Calbo and Sabburg [13]).

It is common to use the R/B channel ratio to find the clear sky pixels, since blue light is scattered more in the atmosphere leading to the blue appearance of the sky. For a clear sky pixel, the R/B ratio will be small compared to a cloud's pixel. Classification of sky images based on this ratio was first implemented by Johnson and Hering [15] and commonly used in the literature (e.g. Koehler et al. [16], Pfister et al. [17]). Heinle et al. [6], Li et al. [18], Yang et al. [19], Chauvin et al. [20] used variants of this ratio such as $R-B$ or normalized B/R which they found to improve classification accuracy. However, thresholding based on this ratio is not guaranteed to succeed as the colour of the atmosphere can change significantly based on atmospheric haze (also known as turbidity), ground reflectance, and solar position, see Hosek et al [21]. Therefore, additional classification techniques have been investigated in the literature. In recent papers by Yang et al. [22] and Dev et al. [23] cloud pixels are identified in images of the sky. The former makes use of differences between the green channels of clear and cloudy-sky images, and the latter uses a supervised method based on Partial Least Squares regression. Hsu-Yung and Chih-Lung [24] merged the results of several classification methods by a voting scheme to detect cloud pixels. They used also different patch sizes to exploit image information at multiple scales. These recent approaches segment cloud pixels from the sky by labelling a pixel as cloud or non-cloud, whereas in our method, it is labelled as being a specific cloud type.

Calbo and Sabburg [13] used thresholding for clear sky pixels, statistical features (standard deviation, smoothness, third moment, uniformity) extracted from the image and features obtained from the Fourier Transform of the image to classify clouds. They achieved classification results of 62% when 8 sky conditions were investigated

and up to 76% for 5 sky conditions, including clear skies from a total 395 images.

Heinle et al. [6] classified whole sky images with a k-NN method into seven classes with an accuracy of 75% using a training set of 200 and test set of 275 images including: Cumulus, Cirrus & Cirrostratus, Cirrocumulus & Altocumulus, clear sky, Stratocumulus, Stratus & Altostratus, Cumulominbus & Nimbostratus. They calculated the spectral features in RGB colour space (mean, standard deviation, skewness, difference of colour components), the structural features from the $GLCM$ of the image and the cloud cover percentage of the image. Kazantzidis et al. [7] improved the method of Heinle et al. [6] by updating the feature vector used in their method and using subclasses of each cloud class. This obtained an accuracy of 79% from a training set of 1050 and test set of 1500 images including 6 cloud types and clear sky. Liu et al. [8] used the Local Binary Pattern (LBP) method to classify clouds and was found to give high accuracy results of 90% from a dataset of 1500 images. Wang et al. [25] also used LBPs for feature extraction based on a histogram approach to get more stable features.

Cheng and Yu [1] used a block based technique to classify an all-sky image according to its cloud type. For each block, statistical features such as mean, standard deviation, skewness, differences of colour components and LBP texture features were extracted. The technique starts by dividing the image into equally sized blocks which helps to classify mixed-cloud type images. To assess the impact of the classifier on the results three techniques were used and compared: k-NN classifier, Bayesian classifier and Support Vector Machines (SVMs). The cloud type of an input whole sky image was found by using the block based classification and a voting scheme based on counting the blocks in the image of similar type.

Hussain and Sayed [26] used a bag of words and minimum distance classification technique to classify sky images as clear, cloudy and sunset sky images. They did not deal with the cloud type in the image. Li et al. [2] also classified a whole sky ground image by using the bag of features technique which gave 90.9% accuracy using a dataset of 5000 images. The images in the dataset were divided into patches and then features of each patch, called micro-structures in this case, were found.

Mean, standard deviation, skewness and contrast were used to extract statistical features. Each image was represented by a feature vector which was generated by calculating weighted frequencies of micro-structures observed in that image. An input image was classified according to its representative feature vector by using SVMs, a one-against-all technique. Zhang et al. [27] proposed a cross-domain classification method where the features of an image were obtained by integrating different feature extraction methods considering the maximum response on different sized regions of the image. Gan et al. [14] also worked on different sized regions of the image where they used a method based on sparse coding, which is efficient in terms of both locality and computation. As in the work by Li et al. [2] they also used mean, standard deviation, skewness and contrast to extract statistical features.

Other approaches have used additional hardware to improve the classification results. Liu and Li [28] used multimodal information obtained from weather station networks and Huertas-Tato et al. [29] used the information collected from a ceilometer. Roman et al. [30] used HDR images combined with a ceilometer to estimate cloud cover from HDR images. HDR imagery was used to estimate a set of ratios, largely based on the commonly used red-blue ratio, and these were then combined with a metric based on sky symmetry around the solar principle plane and information from a ceilometer. Finally, this information was used to classify pixels as cloud or sky based on thresholds.

There is a wide variety of methods which can be used to learn labels for pixels in the literature e.g. Guzella and Caminhas [31], Zhang et al. [32], Boiman et al. [33], Bosch et al. [34]. Supervised methods such as Convolutional Neural Networks have been used to classify objects with high accuracy Krizhevsky et al. [35], Kim et al. [36], however these typically require very large amounts of labelled data, which as discussed in Section 1 is infeasible for images of clouds.

There are also methods proposed to achieve pixel level classification by using weakly supervised learning such as Verbeek and Triggs [37], Vezhnevets et al. [38], Vezhnevets et al. [39], Pathak et al. [40], Pinheiro and Collobert [41] and Papandreou et al. [42]. In these papers, weakly supervised learning is used to learn labels based

on a coarse initial labelling, potentially one label for a whole region or several labels for an image. In these methods training is implemented on pixel-wise labelled data whereas our method infers the output automatically. Also, our work is different as it is robust to incorrect coarse labelling, requires one label for whole image and is designed to infer more labels than were initially assigned to an image. While weakly-supervised deep learning methods exist [43], these also require a large volume of training data. Our method requires a small number of coarsely labelled images and is robust to small changes in cloud appearance. Unsupervised methods learn information from the input data, and can cope with smaller amounts of training data, but the results may not clearly correspond to labels as required in this work. Therefore, we develop a weakly supervised approach which combines the labelling accuracy of supervised methods with the learning style of unsupervised methods.

3. Overview of the Classification Method

The classification method proposed in this work consists of two stages: building a classifier using weak-supervision and run-time classification. The first stage involves creation of coarsely labelled training data and then training the hierarchical classifier. The second step is the classification of an input whole-sky HDR image per-pixel. See Figure 1 for an overview of this process, and Table 1 for a list of symbols used in the following sections.

Table 1: Symbols used in this work

$\wp_p^n$	Patch of pixel p with length n
H	Histogram
N_p	Neighbourhood of pixel p
$\mathcal{V}$	Feature vector (FV)
$\tilde{\mathcal{V}}$	Representative feature vector (RFV)
L	Down-sampling level
ζ_L	Cluster in level L
K_L	Number of clusters in level L

Creation of coarsely labelled training data (visualized in Figure 1, top row, left column) involves manually labelling training images with only the dominant cloud

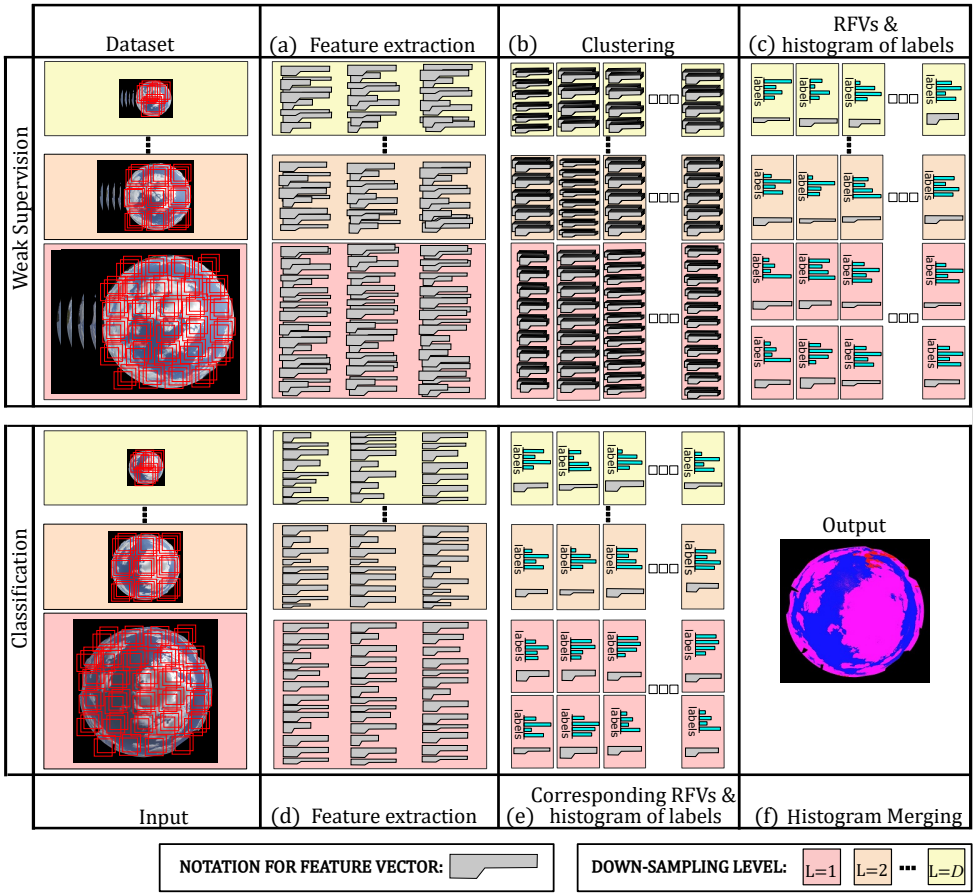

Figure 1: An overview of the two stage method.
Top row: FVs are extracted (a) from images in the dataset, and are clustered (b) ignoring their labels. Since the clusters can include FVs with different labels, an RFV without a label and a histogram of potential labels is generated (c) for each cluster.
Bottom Row: When classification is required for a new input image, the FVs are extracted (d) for each pixel and are matched (e) with the closest RFV and corresponding histogram to find the cloud type associated with the pixel (f).

type observed in the image, which is then used in the learning stage to more precisely classify cloud pixels. This approach of defining the dominant cloud type rather than individually labelling pixels has two advantages. Firstly, it is more robust to mistakes made by subjective per-pixel labelling by users, and secondly it would require an infeasible amount of time to manually label millions of pixels individually in the entire training set. The dominant cloud type is used because typically skies include multiple cloud types, and capturing only one type is highly dependent

on weather conditions. However, in this case, labelling all the cloud pixels with the dominant cloud type causes mislabelling of pixels which belong to a different cloud type. This paper overcomes this problem by using a histogram based approach (described in more detail in Section 4.3) which links the extracted features of a patch to the correct per-pixel cloud type.

Specifically, in the training phase (Figure 1, top row), a down-sampled hierarchy of images is created for each input image. At each level of the hierarchy, patch-based features augmented with the initial labelling are extracted (Figure 1.a, Section 4.1) and then clustered (Figure 1.b) by using K-Means Clustering to generated RFVs (Section 4.2). Rather than assigning one label to each RFV, we propose to use a histogram where each bin corresponds to a label associated with the FVs used to create the RFV (Figure 1.c) which considers the distribution of features over the dataset (Section 4.3). This hierarchical approach allows the method to exploit both local and global structures of the clouds, which has a significant impact on classification performance.

During the classification phase (Figure 1, bottom), downsampling is used to create the hierarchy of patch based features for an input HDR image (Figure 1.d). Then the closest RFV for each pixel at each level in the hierarchy is located (Figure 1.e). Finally, the histogram associated with each RFV is used to assign a cloud type to each pixel by using a histogram merging technique (Figure 1.f) to determine the dominant label in the set of hierarchical FVs forming the cluster (Section 4.3).

4. Weakly-supervised cloud classifier

This stage of the method involves extraction of feature vectors based on statistical and textural methods and then clustering these vectors into RFVs and histograms. The following sections present a detailed explanation of these steps.

4.1. Feature Vector Extraction

This work uses patch-based features due to their ability to exploit a wider range of local image information such as structure which enables a richer set of features to

be extracted. A patch is defined as a square subset of an image and denoted as $\wp^n$ where n is the edge size in pixels, and is always odd. A patch $\wp^n$ is assigned centered at each pixel p in each level of the hierarchy.

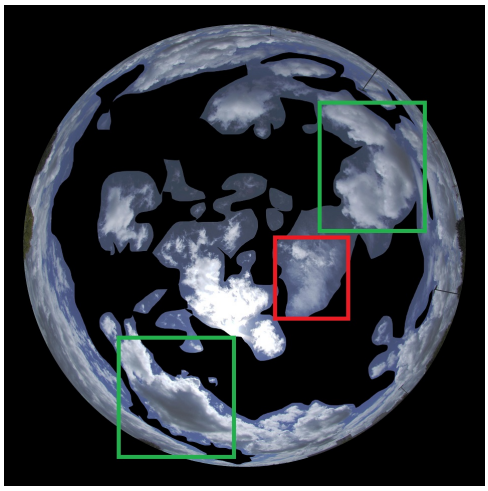

Figure 2: This figure shows a Cumulus image from the dataset. The most common cloud type is Cumulus (green box), therefore all the features extracted from this image are labelled with Cumulus other than the masked clear sky pixels (masked black for visualisation purposes) which are kept with the label of clear sky. However, this image still includes clear sky pixels because of imperfect masking and also Cirrus cloud (red box). Unmasked Clear sky and Cirrus pixels in this image are initially purposefully incorrectly labelled as Cumulus due to the weak labelling. This is subsequently resolved using RFVs and the histogram approach described in Section 4.2.

The features in each patch are described with a d-dimensional FV using information extracted from the patch $\wp_p^n$. If the pixel is associated with a label of clear sky, then clear sky is assigned to the FV, otherwise the label of the dominant cloud type in the image is used, see Figure 2. These labels are used to decide the cloud type of the clusters, described in Section 4.2.

The statistical and structural features were chosen by considering the common features used in literature. Features generated using only one of the colour channels $\{R, G, B\}$ reduces the learning efficiency, since the dynamic range of the skies cause the sky and different cloud types to have similar colour properties depending on the sky condition. An example of this is desaturation around the horizon due to inscattering of light from the sky, leading to larger R and G values which makes it difficult to distinguish clear sky from clouds with only one colour channel. Therefore, lu-

minance (l) and the R/B (r) colour ratio of the pixels are used in the calculation of features. Luminance obtained from HDR images can provide more information about the image compared with previous methods and is calculated by weighting the colour channels: $l = 0.2126 * R + 0.7152 * G + 0.0722 * B$, and the ratio between R and B is a standard feature in the cloud classification literature (Johnson and Hering [15], Koehler et al. [16], Pfister et al. [17], Heinle et al. [6], Li et al. [18], Yang et al. [19], Chauvin et al. [20]).

When considering structural features, methods such as *GLCM*, Fourier Transform, Discrete Cosine Transform can be used. However, structural features such as the ones based on *GLCM* and the Fourier Transform were not used in this method since *GLCM* is only applicable to integer values, and therefore cannot work with HDR images, and the Fourier Transform was found to give low classification accuracy, by Calbo and Sabburg [13]. It should be noted that the *GLCM* could be modified to use quantized HDR values, but this would require further apriori choices of parameters such as number of bins, which would make the method impractical due to significant manual adjustments to account for the range of values from HDR captures of skies. Therefore, this work uses the Discrete Cosine Transform due to the compact description of frequency content and computational speed.

The following equations are used to define the statistical and structural features in the classification:

- **Mean** : $\mathscr{M}_c = \frac{1}{n^2} \sum_{i=0}^{n-1} \sum_{j=0}^{n-1} p_c^{ij}$
- **Standard Deviation** : $\mathscr{SD}_c = \frac{1}{n} \sqrt{\sum_{i=0}^{n-1} \sum_{j=0}^{n-1} (p_c^{ij} - \mathscr{M}_c)^2}$
- **Skewness** : $\mathscr{SK}_c = \frac{1}{n^2} \sum_{i=0}^{n-1} \sum_{j=0}^{n-1} (\frac{p_c^{ij} - \mathscr{M}_c}{\mathscr{SD}_c})^3$
- **Contrast** : $\mathscr{C}_{RB} = \frac{\mathscr{M}_B - \mathscr{M}_R}{\mathscr{M}_B + \mathscr{M}_R}$

•**Discrete Cosine Transform** :

$$\mathscr{D}_c^{q_1 q_2} = \alpha_{q_1}\alpha_{q_2}\sum_{i=0}^{n-1}\sum_{j=0}^{n-1} p_c^{ij}\cos\Big(\frac{\pi(2i+1)q_1}{2n}\Big)\cos\Big(\frac{\pi(2j+1)q_2}{2n}\Big)$$

$$\alpha_q = \begin{cases} 1/\sqrt{n}, & q = 0 \\ \sqrt{2/n}, & 1 \leq q \leq n-1 \end{cases}$$

where q_1 and q_2 denotes the basis functions with values in $\{0,..,n-1\}$.

Due to the typically low frequency nature of clouds, the higher frequency coefficients of $\mathscr{D}$ contain very small values, so we do not use these for classification. Our initial experiments showed that the patches are sparse in the $\mathscr{D}$ domain, and most of the energy is contained in the lower frequency components. Therefore, in this work, the basis functions $\{00,01,10\}$ which correspond to the first horizontal and vertical frequency components, and the DC component of the discrete cosine transform, are used. As the basis (00), $\mathscr{D}_c$ is equal to $\mathscr{M}_c$, we do not include this term. When implementing this method, there is no need to perform a full DCT per patch as only the horizontal and vertical gradients are required, but the features used are motivated by our analysis from the DCT. Hence the following FV is used in this work (the r subscript corresponds to the ratio of red and blue color channels, and luminance is denoted l):

$$\mathscr{V} = \{r, \mathscr{SD}_r, \mathscr{SK}_r, \mathscr{C}_{RB}, \mathscr{D}_r^{00}, \mathscr{D}_r^{10}, \mathscr{D}_r^{01}, \mathscr{D}_l^{00}, \mathscr{D}_l^{10}, \mathscr{D}_l^{01}\}$$

The extraction of the FVs is implemented for the same patch size at each level of the down-sampled image hierarchy and the features belonging to the same down-sampling level are clustered together.

To summarise, the features chosen to construct the FVs were designed based on a balance of features identified from prior work (see Section 2), and computational performance. We have used a combination of color based statistics, designed to

separate clear sky from clouds, and separate different cloud types where color based statistics are effective, and structural information which is useful where color statistics provide insufficient information. While this FV is sufficient for our work, a more comprehensive search for FVs could be performed to identify if there is a set of features which can provide higher classification accuracy.

4.2. Representative Feature Vector

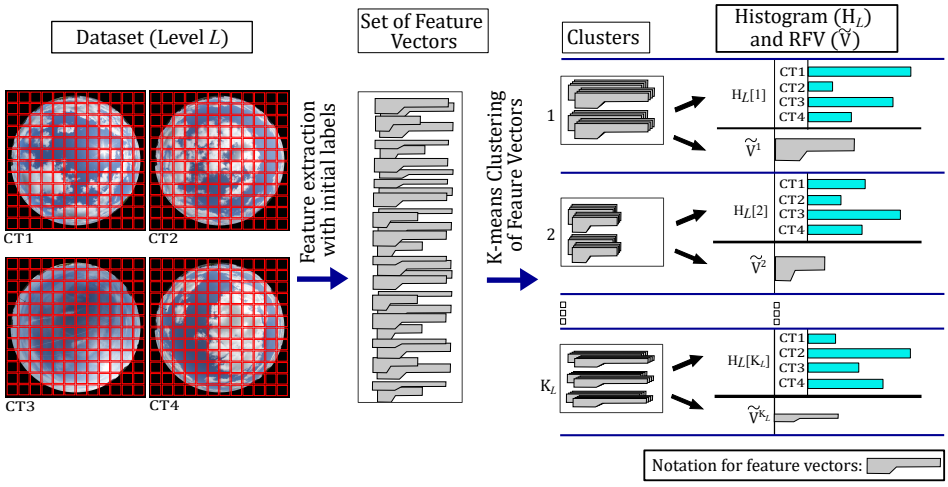

Figure 3: An illustration of labelling of the RFVs for down-sampled level L. As the dataset includes images from several cloud types (labelled CT in the diagram), FVs are labelled with the predominant cloud type of the image they are extracted from, and during clustering their labels are not considered. Therefore, the clusters include FVs with different labels. For each cluster a RFV and a histogram based on the number of labels of FVs are calculated for use when classifying new images.

The features extracted in Section 4.1 produce a large amount of weakly labelled training data with labels assigned as the dominant cloud type in the original image, see Figure 2. This leads to some pixels being mislabelled if multiple cloud types are present in the sky (see Figure 2) To resolve this, these features are clustered at each down-sampling level L and the overall features corresponding to each cluster are represented by a RFV.

During clustering, the initial labels are not considered. Instead, since the clusters include weakly labelled data, a histogram of labels is assigned to each RFV. These labels correspond to the FVs used to create the cluster, see Figure 3. These histograms and associated RFVs will later be used to label the input images. As this clustering

process operates at each level of the down-sampled image, this generates multiple levels of RFVs and associated histograms of labels. These are used in the classification stage as outlined in Section 5.

In this work, the extracted FVs are clustered using k-Means Clustering. While there are several other options for clustering, such as Expectation-Maximization [44] and DBSCAN [45], we use k-means as it is fast, robust, and produces results which work well for construction of the RFVs.

Prior to clustering, the same number of initial FVs are sampled from each image in the dataset using stratified sampling. Then these FVs are clustered into K_L clusters using k-Means Clustering, where each cluster has an associated RFV ($\tilde{V}$). These vectors are the mean of all the features in each cluster and represent the FVs showing similar properties.

4.3. Histogram of Labels

Due to the unsupervised clustering process explained at Section 4.2, the RFVs have no knowledge of the labels associated with the individual FVs used to create the cluster. As it is likely that these FVs were associated with different labels, and in order to use the RFVs for classification, labels need to be re-associated with each RFV at each level of the hierarchy.

We achieve this by computing a normalized histogram of labels with the RFV. This is constructed by storing all the labels associated with the FVs used to build the cluster, creating a histogram with the number of bins equal to the number of unique labels, and finally creating the histogram. However, the number of FVs used to create each cluster can vary, and is also dependent on the number of input images. Therefore, we normalize the histogram with the total number of FVs observed of that cloud type to prevent one cloud type from dominating the others during classification.

We also cluster clear sky pixels from the coarse labelling (see Figure 2), but ensure that these are not clustered with the other cloud types which helps to disambiguate clear sky and clouds during classification. The histogram associated with these clusters contains one bin with the label of clear sky.

At level L of the hierarchy, the total number of FVs labelled with a cloud type (t) is defined as N_t and the number of features labelled with t in the $i'th$ cluster as N_t^i. The histogram is calculated as:

$$H_L[i][t] = \frac{N_t^i}{N_t}, \forall\, i \in \{1, ..., K_L\}. \tag{1}$$

This computes the histogram of ratios of cloud types of features within a cluster given the total number of features with corresponding cloud types in the dataset.

5. Runtime Classification

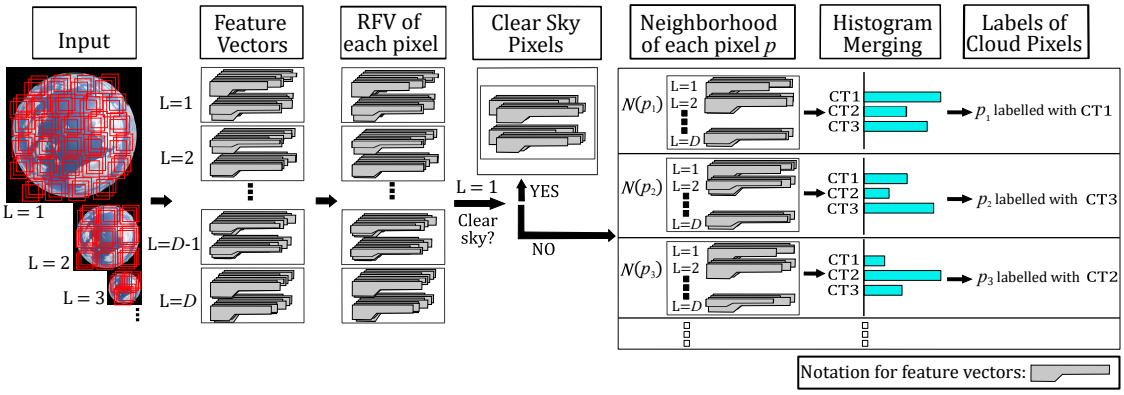

Figure 4: An illustration of labelling the pixels. The input image is downsampled D times and a FV of each pixel at each level of downsampling is calculated. The closest matching RFV of the FV associated with the input is found. If the labelling of a RFV of a pixel (p) is clear sky at the highest level $L = 1$, then the pixel is labelled as clear sky and not investigated further. For the other pixels a histogram is calculated considering its neighborhood at each level and it is labelled according to the label corresponding to the largest value in the histogram.

Classifying pixels at runtime is implemented in two steps. Firstly, at each level of the hierarchy the FV of each pixel is generated from the corresponding patch, see Figure 4. For each FV ($\mathcal{V}$), a RFV ($\tilde{\mathcal{V}}^i$) is found which minimizes the Euclidean distance to the FV:

$$\|\mathcal{V} - \tilde{\mathcal{V}}^i\|^2 \leq \|\mathcal{V} - \tilde{\mathcal{V}}^j\|^2, \forall\, j \in \{1, .., K_L\}. \tag{2}$$

The label of a pixel is found by merging the histograms of RFVs found at each down-sampling level. This provides increased accuracy in labelling as labels lower in the hierarchy are able to accurately represent larger regions of the image, while higher levels provide finer, but potentially less accurate labelling at the pixel level. This motivates our weighting scheme that gives a higher weight to histograms lower in the hierarchy, and a smaller weight to those higher in the hierarchy.

This leads to a set of possible labels and associated weights for each pixel. In our method, we assign the label with the highest weight in the merged histogram. Initially, for each pixel q at level L the matching RFV (see Equation 2) and the associated histogram $H_L[i_q]$ is found (see Figure 3). Then the merging is implemented by summing the weighted histograms corresponding to the pixel p under consideration.

$$H_{merge}[p][t] = \sum_{L} w_L \sum_{q \in N_{\hat{p}}} H_L[i_q][t] \tag{3}$$

where $w_L = \frac{2^L}{\sum_{i=0}^{i=D} 2^i}$ is the weight of level L for a total of D down-sampling levels. We use a factor of two in the weights as this leads to a good balance between histograms at lower and higher levels. Larger values cause the histograms at lower levels to have too much influence leading to mis-classifications with mixed cloud types, and smaller values result in noise in the labelling. $N_{\hat{p}}$ corresponds to a patch around the pixel at the associated level of the hierarchy $\hat{p} = \left\lfloor \frac{p}{2^L} \right\rfloor$.

The size of the patch is computed as $\lfloor n\Delta^L \rfloor$, where $\Delta < 1$ is a user defined value which controls the number of pixels to be considered around $\hat{p}$, and $\lfloor \cdot \rfloor$ is the floor function. As the number of pixels decreases at each level of the down-sampling hierarchy, the size of the neighbourhood is likewise decreased. In this work n and $\Delta = 0.85$ were chosen to get $\lfloor n\Delta^L \rfloor = 1$ at level $L = D$.

Because of the down-sampling, images with low resolution are typically dominated by cloud pixels and clear sky pixels can be lost. Therefore, labelling of clear sky pixels is implemented on the full resolution image where they still have distinctive features compared to the cloud pixels. If in a histogram, the clearsky cloud type has the high-

est value, then its RFV is labelled with clearsky (CS). An overview of the method is given in Algorithm 1.

Algorithm 1 Classification of an image I

for $L \leftarrow 1$ to D **do**
  Calculate down-sampled image I_{ds}
  for $for\ each$ pixel $p \in I_{ds}$ **do**
    Find $\tilde{\mathcal{V}}^{i_p}$
    if $L == 1$ and $\tilde{\mathcal{V}}^{i_p}$ labelled with Cs **then**
      label Cs
    end if
    Calculate $H_L[i_p]$
  end for
end for
for $each$ pixel $p \in I$ **do**
  if not **labelled** Cs **then**
    Calculate $H_{merge}[p]$
    return label associated with largest value in $H_{merge}[p]$
  end if
end for

6. Results

6.1. Dataset

The training set used in this work is composed of whole sky HDR images captured with a Canon EOS 5D Mark 3 with a Sigma 8mm EX DG Fisheye Lens, an aperture of f3.5 and 7 bracketed exposures equally spaced from -8 to 8. The HDR images included surrounding buildings which were eliminated with a fixed size circular mask. However, the tops of some incongruent objects were still visible, these were masked manually during the generation of outputs. All the images had a resolution of 3,840×3,840 pixels. Images were captured in the United Kingdom at various times of the year.

The dataset consisted of three types of cloud fisheye images: 21 Cirrus (Ci), 29 Cumulus (Cu) and 15 stratocumulus (Sc) images. These clouds are local clouds frequently observed in the data collection area. The training images are tagged with the dominant cloud type in the image. The dataset does not include clear sky (Cs)

images because the features extracted from clear sky images and clear sky pixels in a cloudy image show different characteristics and can contain different color information. Therefore, the clear sky pixels in cloudy images are pre-labelled manually and used as clear sky data during the extraction of features. It is not necessary to have 100% correct labelling since the histogram technique used in clustering improves the classification of mislabelled pixels. For each high resolution image in the dataset, approximately 98,000 patches are sampled.

The images used in the results were masked into a square shape of 3,840 × 3,840 pixels and the whole sky image has a radius of 1,766 pixels. An initial experiment with different patch sizes showed that using larger patch sizes lowers the accuracy in classification. We chose 0.15% of the total image size to be used as the smallest patch size, which corresponded to $n = 5$. From initial experiments, this was found to provide a good balance between the calculation time and accuracy, and is the smallest patch size from which enough structural information was able to be extracted.

We also found that increasing the number of clusters to more than 750 adds only 0.2% more accuracy, therefore by considering the computational efficiency $K_1 = 750$ is used in the experiments. Since the total number of features extracted at each level decreases, the total number of clusters should likewise decrease. We empirically determined the values 750, 745, 740, 735, 730, 725, 720, 250 for $L \in [1..8]$. However, our approach is not sensitive to these values. Additionally, we determined that L should be capped at 8, corresponding to a resolution of 15×15 pixels at which further decreases in resolution did not improve performance.

6.2. Classification Performance

The performance of the method was validated using Cross Validation. From the dataset, 50 subsets were generated by randomly skipping one image from each cloud type. The proposed method was trained on images from each subset, and the skipped image associated with the subset was classified per pixel. When calculating results, a total 3,071,297 pixels were sampled from the labelled pixels by using concentric sampling over the disk, see Figure 5. The overall accuracy of the

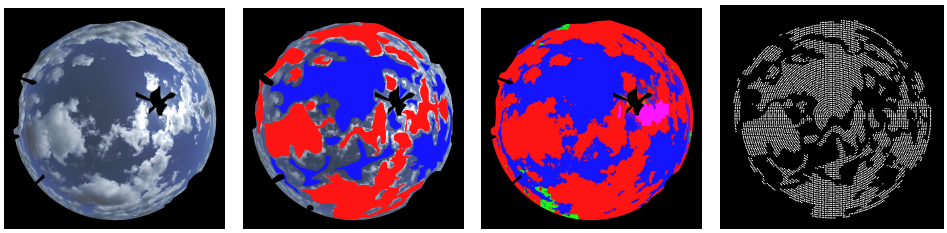

Figure 5: This figure shows how the accuracy of a classification is computed. The image on the left is the input image with the sun and non-cloud structures masked, the center left is the manually labelled image, the center right is the classified image with our method and the one on the right is concentric sampling on the image. Only the pixels where there are specific cloud types exist are tagged. Since it is hard to classify pixels manually on the edges, they are not tagged. The accuracy of classification is checked between the tagged image and the classified image.

method in classifying pixels with the correct cloud type was 84.050%, see Table 2. For each expected cloud type the percentages were calculated as $\frac{NP_t}{NP_{total}}$ where NP_t is the number of pixels classified as cloud type t and NP_{total} is the total number of expected pixels. Our un-optimised method takes on average 203 seconds to label all pixels on an Intel i7-3612QM, 8 GB RAM, using multithreading with 8 threads.

Table 2: Interpretation of the cross validation results with a Confusion Matrix. Expected values are the input cloud type of the pixels. Predicted values are classification results of the inputs. Numbers in the parenthesis indicate the number of pixels.

		PREDICTED				
		Cu	**Ci**	**Sc**	**Cs**	**Overall Accuracy**
EXPECTED	**Cu**	0.80635 (903393)	0.00324 (3634)	0.17718 (198504)	0.01322 (14811)	
	Ci	0.02636 (92099)	0.88292 (3084850)	0.00001 (29)	0.09071 (316942)	
	Sc	0.12725 (228286)	0.00044 (782)	0.87065 (1561982)	0.00166 (2983)	
	Cs	0.01354 (38807)	0.15790 (452434)	0.02649 (75911)	0.80206 (2298080)	0.84050

A visualisation of the per-pixel classification of clouds can be seen in Figure 6 and mixed cloud types in Figure 7, which also includes a comparison to manually assigned labels. This shows that most pixels are labelled correctly by our method, and the error in the classification is like due to strong similarities in appearance between Cumulus and Stratocumulus cloud types. The sun, lens flare and other world objects such as lamps and buildings are masked in the images.

In previous methods, the results are reported based on the dominant cloud type in the image. Our method can also classify the dominant cloud type from the per-

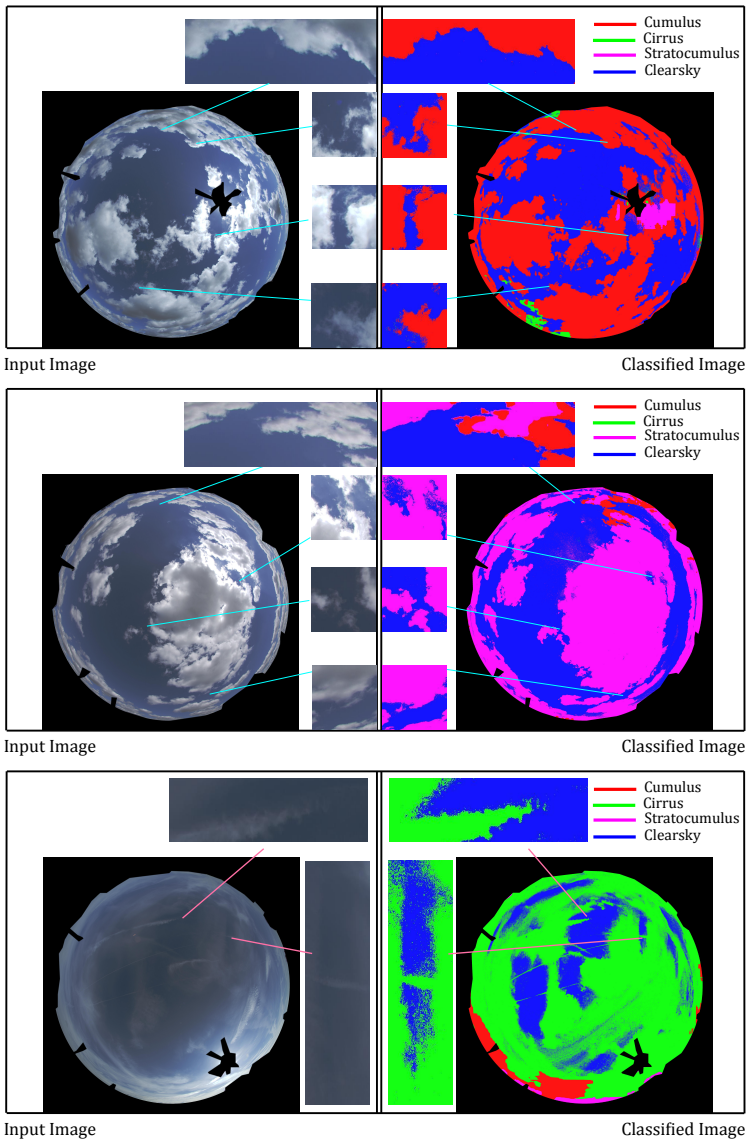

Figure 6: Classification result for Cu (top), Sc (middle) and Ci (bottom) cloud image.

pixel labelling that our method produces by returning the label associated with the largest number of pixels. A 98% accuracy, including results for clear sky images, was obtained by the aforementioned cross validation technique when using this method. Table 3 shows a comparison between our method and the average results reported by related work on cloud type classification of whole sky images. We com-

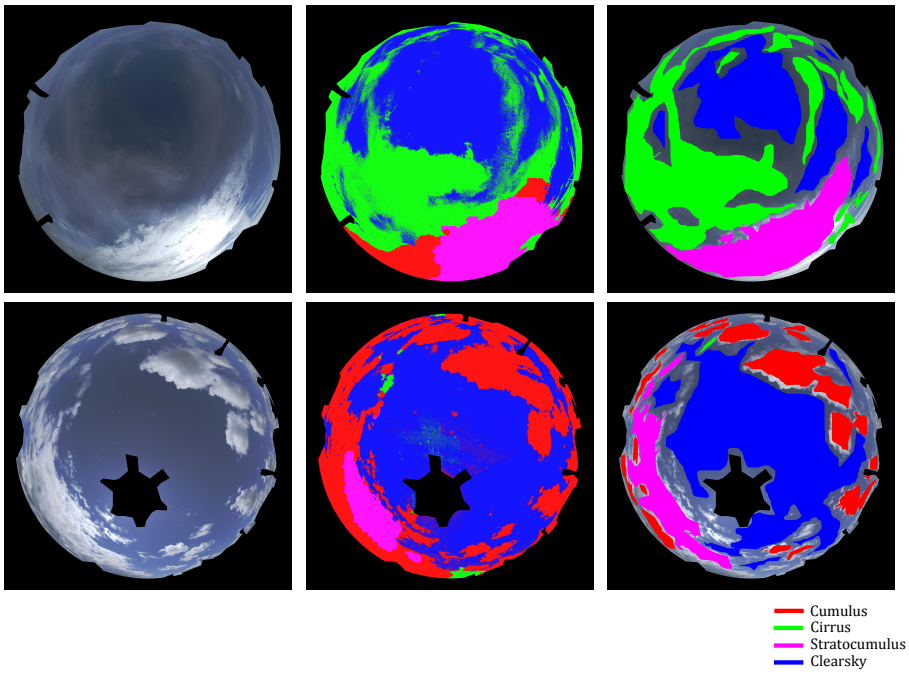

Figure 7: Results for the classification when the input includes different cloud types. The image on top was taken when Stratocumulus clouds have started to form below Cirrus clouds, and the bottom image includes a mixture of Stratocumulus, Cumulus and Cirrus clouds. Left: Input image, Middle: Classified image with our method, Right: Manual labelling of the input image. Most of the pixels have been labeled correctly, and the remaining error in these images is likely due to the similarity in appearance between Cumulus and Stratocumulus clouds.

pare with the following methods: Heinle et al. [6] which classifies the clouds into 7 cloud types using a training set of 200 images. The method by Cheng and Yu [1] uses 3,000 images for training and classifies images into 6 cloud types using block based classification. Li et al. [2] and Gan et al. [14] used 5,000 and 1,000 images respectively and classified the clouds into cirriform, cumuliform, stratiform, clear-sky and mixed cloudiness. For a more detailed discussion of these methods, please see Section 2. The results from our work show a competitive performance with the state-of-the-art. However, these results are focused on whole sky cloud type classification, whereas our approach also performs per-pixel cloud type classification. We additionally show further results in Figure 8.

Table 3: A comparison of whole image classification results based on authors reported results.

Heinle et al. [6]	Li et al. [2]	Cheng and Yu [1]	Gan et al. [14]	Our Method
75%	90.9%	90%	97%	98%

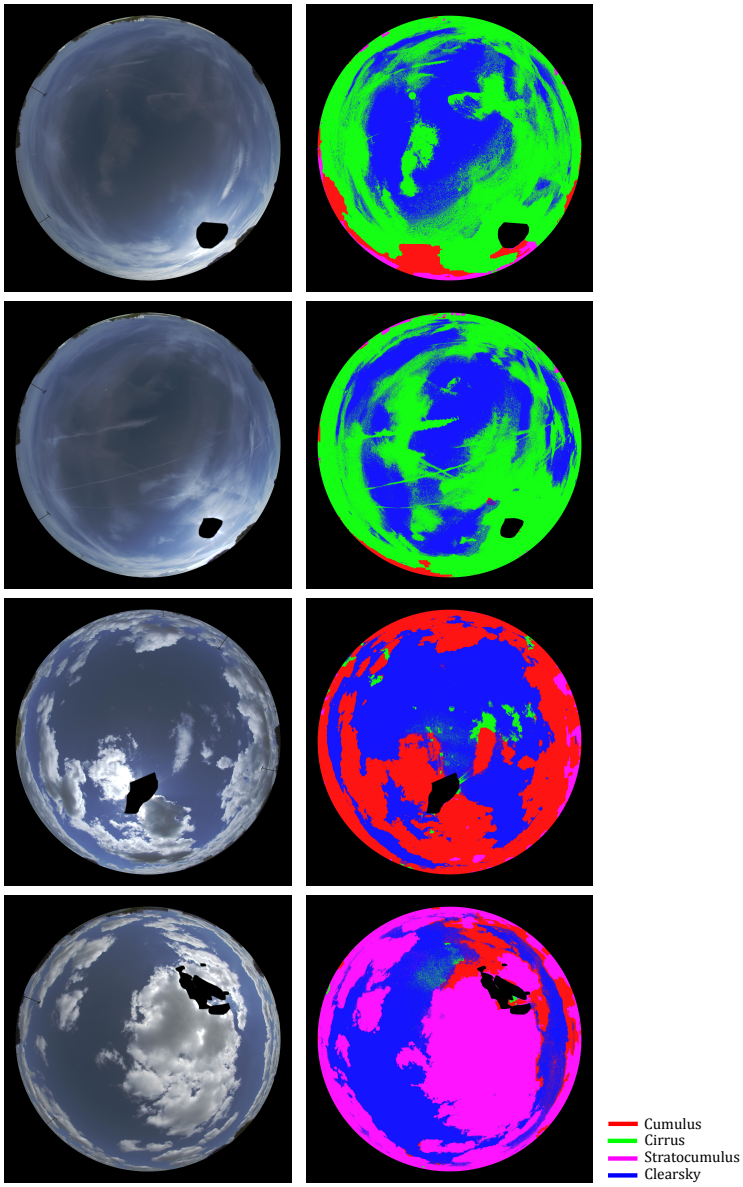

Figure 8: Additional classification results for a variety of cloud types.

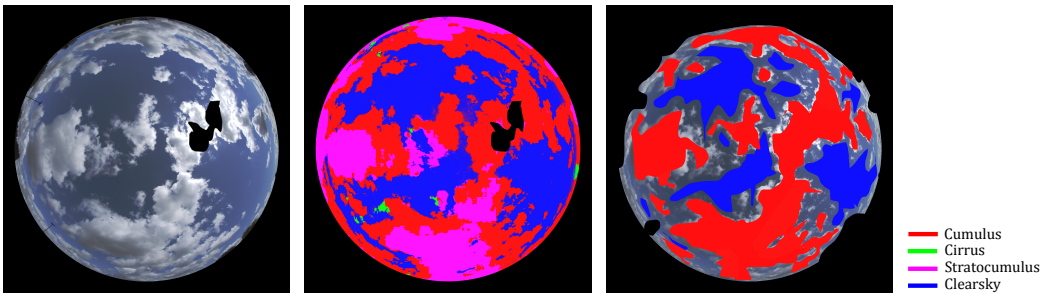

Figure 9: A failure case for the classification. In this case, the dominant cloud type should be Cumulus, however the classifier has labelled regions as being Stratocumulus (middle image). Manual labelling shown on the right illustrates that the error in classification is mostly distributed towards the center the clouds. This is partially due to Cumulus and Stratocumulus clouds having similar appearance, especially when they subtend large regions of the image.

We investigated the performance of our method when using pre-labelled clear sky pixels versus automatically extracting features from predominately clear sky images. The results of this show a significantly lower accuracy of 0.67 for classification when using features extracted from clear sky images, compared to 0.84 with our method. This is due to the features in predominately clear sky images being significantly different to the features in regions of clear sky in a cloudy image.

Figure 9 shows an example of a failure case of our method. In this image, the clouds are largely Cumulus, whereas the classifier has assigned a mixture of the labels of Cumulus and Stratocumulus to this image, as can be seen when compared to the manually labelled image on the right. This is because of the similarity of appearance of Cumulus and Stratocumulus, especially when these cloud types subtend a large region. This leads to similarities in the RFVs, and although this is captured in the histogram, there is not enough information to disambiguate the cloud types. However, the classifier does produce correct labels in regions of this image where there is a clearer match between the features captured and the RFVs. This issue will be resolved in future work through capturing larger datasets, and fine tuning cluster sizes during clustering. However, this is a rare occurrence in our dataset, and as can be seen in Figure 6 and Table 2, our method produces the correct result in the vast majority of cases.

7. Conclusion and Future Work

This paper presents the first per-pixel classification of cloud types from HDR images. We propose a weakly supervised approach, requiring a minimal amount of user input to train the system. Our method extracts statistical and structural features from a hierarchy of patches, and clusters them into RFVs which are linked to the cloud types from the initial labelling. When classifying images, our approach efficiently uses the hierarchy of RFVs to label pixels with high accuracy for both per-pixel labelling (84%) and also achieves state-of-the-art accuracy for whole sky cloud cover classification.

In future, we intend to expand the number of cloud types used for training to cover a wider range of cloud formations. We also intend to collect a larger database of HDR whole sky images, and investigate other machine learning approaches for classification. With a larger HDR sky image database, it could be possible to bootstrap a CNN classifier based on our initial classification.

8. Acknowledgement

Pinar Satilmis is supported by the Council of Higher Education and Hacettepe University, Turkey.

References

[1] H.-Y. Cheng, C.-C. Yu, Block-based cloud classification with statistical features and distribution of local texture features, Atmospheric Measurement Techniques 8 (3) (2015) 1173–1182.

[2] Q. Li, Z. Zhang, W. Lu, J. Yang, Y. Ma, W. Yao, From pixels to patches: a cloud classification method based on bag of micro-structures., Atmospheric Measurement Techniques Discussions 8 (10) (2015) .

[3] R. Tapakis, A. Charalambides, Equipment and methodologies for cloud detection and classification: A review, Solar Energy 95 (2013) 392–430.

[4] M. Jarraud, Guide to meteorological instruments and methods of observation (wmo-no. 8), World Meteorological Organisation: Geneva, Switzerland (2008) .

[5] A. Chalmers, K. Debattista, Hdr video past, present and future: A perspective, Signal Processing: Image Communication 54 (2017) 49–55.

[6] A. Heinle, A. Macke, A. Srivastav, Automatic cloud classification of whole sky images, Atmospheric Measurement Techniques 3 (3) (2010) 557–567.

[7] A. Kazantzidis, P. Tzoumanikas, A. Bais, S. Fotopoulos, G. Economou, Cloud detection and classification with the use of whole-sky ground-based images, Atmospheric Research 113 (2012) 80–88.

[8] S. Liu, C. Wang, B. Xiao, Z. Zhang, Y. Shao, Salient local binary pattern for ground-based cloud classification, Acta Meteorologica Sinica 27 (2013) 211–220.

[9] P. Gehler, S. Nowozin, On feature combination for multiclass object classification, in: Computer Vision, 2009 IEEE 12th International Conference on, IEEE, 2009, pp. 221–228.

[10] R. Wang, W. Li, R. Li, L. Zhang, Automatic blur type classification via ensemble svm, Signal Processing: Image Communication (2018) .

[11] M. Woźniak, D. Połap, Object detection and recognition via clustered features, Neurocomputing (2018) .

[12] Y. Li, F. Cui, X. Xue, J. C.-W. Chan, Coarse-to-fine salient object detection based on deep convolutional neural networks, Signal Processing: Image Communication 64 (2018) 21–32.

[13] J. Calbo, J. Sabburg, Feature extraction from whole-sky ground-based images for cloud-type recognition, Journal of Atmospheric and Oceanic Technology 25 (1) (2008) 3–14.

[14] J. Gan, W. Lu, Q. Li, Z. Zhang, J. Yang, Y. Ma, W. Yao, Cloud type classification of total-sky images using duplex norm-bounded sparse coding, IEEE Journal

of Selected Topics in Applied Earth Observations and Remote Sensing 10 (7) (2017) 3360–3372.

[15] R. Johnson, W. S. Hering, Automated cloud cover measurements with a solid-state imaging system, in: Proceedings of the Cloud Impacts on DOD Operations and Systems-1987, Workshop, 1987, pp. 59–69.

[16] T. Koehler, R. Johnson, J. Shields, Status of the whole sky imager database, in: Proceedings of the Cloud Impacts on DOD Operations and Systems, 1991 Conference, 1991, pp. 77–80.

[17] G. Pfister, R. McKenzie, J. Liley, A. Thomas, B. Forgan, C. N. Long, Cloud coverage based on all-sky imaging and its impact on surface solar irradiance, Journal of Applied Meteorology 42 (10) (2003) 1421–1434.

[18] Q. Li, W. Lu, J. Yang, A hybrid thresholding algorithm for cloud detection on ground-based color images, Journal of atmospheric and oceanic technology 28 (10) (2011) 1286–1296.

[19] J. Yang, W. Lu, Y. Ma, W. Yao, An automated cirrus cloud detection method for a ground-based cloud image, Journal of Atmospheric and Oceanic Technology 29 (4) (2012) 527–537.

[20] R. Chauvin, J. Nou, S. Thil, A. Traore, S. Grieu, Cloud detection methodology based on a sky-imaging system, Energy Procedia 69 (2015) 1970–1980.

[21] L. Hosek, A. Wilkie, An analytic model for full spectral sky-dome radiance, ACM Transactions on Graphics (TOG) 31 (4) (2012) 95.

[22] J. Yang, Q. Min, W. Lu, Y. Ma, W. Yao, T. Lu, J. Du, G. Liu, A total sky cloud detection method using real clear sky background, Atmospheric Measurement Techniques 9 (2) (2016) 587–597.

[23] S. Dev, Y. H. Lee, S. Winkler, Color-based segmentation of sky/cloud images from ground-based cameras, IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing 10 (1) (2017) 231–242.

[24] C. Hsu-Yung, L. Chih-Lung, Cloud detection in all-sky images via multi-scale neighborhood features and multiple supervised learning techniques, Atmospheric Measurement Techniques 10 (1) (2017) 199.

[25] Y. Wang, C. Shi, C. Wang, B. Xiao, Ground-based cloud classification by learning stable local binary patterns, Atmospheric Research 207 (2018) 74–89.

[26] K. F. Hussain, H. A. Sayed, Enhancement of sky and cloud type classification (2013).

[27] Z. Zhang, D. Li, S. Liu, B. Xiao, X. Cao, Cross-domain ground-based cloud classification based on transfer of local features and discriminative metric learning, Remote Sensing 10 (1) (2017) 8.

[28] S. Liu, M. Li, Deep multimodal fusion for ground-based cloud classification in weather station networks, EURASIP Journal on Wireless Communications and Networking 2018 (1) (2018) 48.

[29] J. Huertas-Tato, F. Rodríguez-Benítez, C. Arbizu-Barrena, R. Aler-Mur, I. Galvan-Leon, D. Pozo-Vázquez, Automatic cloud-type classification based on the combined use of a sky camera and a ceilometer, Journal of Geophysical Research: Atmospheres 122 (20) (2017) .

[30] R. Román, A. Cazorla, C. Toledano, F. Olmo, V. Cachorro, A. de Frutos, L. Alados-Arboledas, Cloud cover detection combining high dynamic range sky images and ceilometer measurements, Atmospheric Research 196 (2017) 224–236.

[31] T. S. Guzella, W. M. Caminhas, A review of machine learning approaches to spam filtering, Expert Systems with Applications 36 (7) (2009) 10206–10222.

[32] H. Zhang, A. C. Berg, M. Maire, J. Malik, Svm-knn: Discriminative nearest neighbor classification for visual category recognition, in: Computer Vision and Pattern Recognition, 2006 IEEE Computer Society Conference on, Vol. 2, IEEE, 2006, pp. 2126–2136.

[33] O. Boiman, E. Shechtman, M. Irani, In defense of nearest-neighbor based image classification, in: Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on, IEEE, 2008, pp. 1–8.

[34] A. Bosch, A. Zisserman, X. Munoz, Representing shape with a spatial pyramid kernel, in: Proceedings of the 6th ACM international conference on Image and video retrieval, ACM, 2007, pp. 401–408.

[35] A. Krizhevsky, I. Sutskever, G. E. Hinton, Imagenet classification with deep convolutional neural networks, in: Advances in neural information processing systems, 2012, pp. 1097–1105.

[36] Y. Kim, Convolutional neural networks for sentence classification, arXiv preprint arXiv:1408.5882 (2014) .

[37] J. Verbeek, B. Triggs, Region classification with markov field aspect models, in: Computer Vision and Pattern Recognition, 2007. CVPR'07. IEEE Conference on, IEEE, 2007, pp. 1–8.

[38] A. Vezhnevets, V. Ferrari, J. M. Buhmann, Weakly supervised semantic segmentation with a multi-image model (2011) .

[39] A. Vezhnevets, V. Ferrari, J. M. Buhmann, Weakly supervised structured output learning for semantic segmentation, in: 2012 IEEE Conference on Computer Vision and Pattern Recognition, IEEE, 2012, pp. 845–852.

[40] D. Pathak, E. Shelhamer, J. Long, T. Darrell, Fully convolutional multi-class multiple instance learning, arXiv preprint arXiv:1412.7144 (2014) .

[41] P. O. Pinheiro, R. Collobert, From image-level to pixel-level labeling with convolutional networks, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2015, pp. 1713–1721.

[42] G. Papandreou, L.-C. Chen, K. P. Murphy, A. L. Yuille, Weakly-and semi-supervised learning of a deep convolutional network for semantic image segmentation, in: Proceedings of the IEEE international conference on computer vision, 2015, pp. 1742–1750.

[43] J. Rony, S. Belharbi, J. Dolz, I. Ben Ayed, L. McCaffrey, E. Granger, Deep weakly-supervised learning methods for classification and localization in histology images: a survey, arXiv preprint arXiv:1909.03354 (2019).

[44] A. P. Dempster, N. M. Laird, D. B. Rubin, Maximum likelihood from incomplete data via the em algorithm, Journal of the Royal Statistical Society: Series B (Methodological) 39 (1) (1977) 1–22.

[45] M. Ester, H.-P. Kriegel, J. Sander, X. Xu, et al., A density-based algorithm for discovering clusters in large spatial databases with noise., in: Kdd, Vol. 96, 1996, pp. 226–231.