Automatically Generated Interactive Weather Reports based on Webcam Images

Frode Eika Sandnes, Kim Andre Pettersen,
Espen Skaufel and Erling Haugstad

Faculty of Technology, Art and Design
Oslo and Akershus University College of Applied Sciences

Abstract

Most weather reports are either based on data from dedicated weather stations, satellite images, manual measurements or forecasts. In this paper a system that automatically generates weather reports using the contents on webcam images are proposed. There are thousands of openly available webcams on the Internet that provide images in real time. A webcam image can reveal much about the weather conditions at a particular site and this study demonstrates a strategy for automatically classifying a webcam scene into cloudy, partially cloudy, sunny, foggy and night. The system has been run for several months collecting 60 Gb of image data from webcams across the world. The reports are available through an interactive web-based interface. A selection of benchmark images was manually tagged to assess the accuracy of the weather classification which reached a success rate of 67.3%.

1 Introduction

Weather reports are important in many contexts. Hiking in the mountains or other desolate places can be dangerous under certain weather conditions and accurate information about weather condition can be crucial for adequate preparation and survival. Fishermen and others travelling on the seas are also dependent on accurate weather information to avoid danger and loss of life. Even in more protected environments, such as the city, accurate weather reports are important. For instance, the weather affects what clothing we decide to wear.

There are two types of weather reports: real-time weather reports and weather forecasts. Weather forecasting is a complex process whereby historic data combined with the current weather measurements are used to make predictions about future weather conditions. For instance, if the air pressure trend is going from high to low, then one can predict cloudy skies. Air pressure rising from low to high predicts sunny weather. Such simple forecasting is often provided on inexpensive weather stations sold for home use. This study focuses on real-time weather reports, but the strategies presented herein could potentially be combined with forecasting techniques for the purpose of predicting the weather.

This paper was presented at the NIK-2011 conference; see http://www.nik.no/.

Most weather reports are based on data from professional weather stations and satellite images, but underground networks of weather stations initiated by the desire to democratize the weather also exist [1]. Sensor network has also been applied for weather applications [2]. A weather station typically comprises temperature sensors, wind sensors that measure the speed and direction of the wind, humidity sensors and air pressure sensors. Such weather stations are typically regularly dispersed across large geographical areas. When looking at data from several weather stations reliable conclusions can be drawn regarding the weather in a given area. For instance, temperature and wind speed and direction measurements from local weather stations can be averaged, or fitted into a mathematical weather model, to provide a reliable and representative measurement for a local area. Air pressure gives an indication of whether it is cloudy or sunny. High air pressure is often associated with sunny weather and low air pressure is associated with cloudy skies.

Low earth orbit environmental satellites provide more global perspectives of the weather as one can track entire clouds moving across the atmosphere. By combining global satellite data with local weather station data more accurate weather reports and weather forecasts are possible.

There are a myriad of websites that provide aggregated and easy access to weather data from thousands of sources. Such websites often allow visitors to plot time series and observe trends and make comparisons. Recently, several weather sites have started combining weather data with webcams, also known as weather cams, so that visitors also can see the weather conditions in addition to the quantitative instrumental measurements [3]. However, these images need to be interpreted manually in existing systems.

This study attempts to take this one step further by automatically interpreting the weather conditions at the sites monitored by the webcams. There are already thousands of openly available webcams across the globe and these webcams are an immense source of valuable information, including weather. Using this existing infrastructure it is possible to get usable weather information independently of existing sources of weather information.

2 Related work

Webcams have been used for a range of useful applications and some of these are related to environmental and weather related monitoring. For instance, the number of people at beaches has been monitored with webcams and areal images where the level of grey at the white sandy beaches gives an indication of the number of people using the beach [4]. A similar approach was used for the automatic monitoring of beach pollution [5].

Several outdoor camera based monitoring systems have employed compensation strategies to overcome various weather conditions that degrade the monitoring process. One such example is overcoming glare effects that occur at night during fog and haze [6].

Researchers have also proposed strategies for detecting and removing rain from video [7], for instance using Kalman filters [8]. With moving images it is possible to clearly observe the rain. However, it is hard to determine if it is rain or not based on a single still webcam images, because the motion of the raindrops is not visible, the image resolution is often low and the cameras are set up with a wide field of view. However, it has been shown that if the rain falls directly onto the lens, or the optics, then the presence of raindrops can be detected by machine [9].

One useful application of weather related webcam monitoring is to monitor the movement of snow on busy roads [10] as roads covered in snow can be a severe traffic hazard. Typically, a webcam is positioned to monitor an accident ridden stretch of road.

If the road gets covered by snow, or snow slides from the roadside into the road, the snow is detected by detecting differences in images taken before and after the event.

The detection of bad weather conditions such a snow storms is important, but the identification of pleasant weather is also of interest. Sunny days are characterized by directional lighting which results in shadows. Although not explicitly applied to quantifying the degree of sunlight, several approaches have been proposed for detecting shadows in images [11-16].

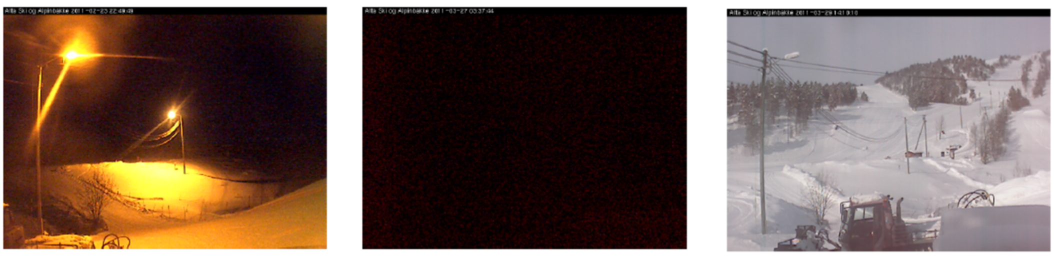
Fog is another visually distinctive weather condition. This work is partially inspired by fog filtering techniques proposed in the literature [17], where Sobel filters are applied to images. Non foggy images will have many edges, while foggy images have few edges. This is illustrated further in subsequent sections. More sophisticated and ambitious algorithms has been proposed for estimating both the wind speed and vapor pressure in webcam images [18] based on principal component analysis, but the reported strategies are in early stages of research. There are also preliminary studies that relate images of the sky to weather [19].

Another avenue of research employs whole sky cameras which captures the entire sky using either fish eye lenses or curved mirrors. Whole sky cameras are used to study cloud cover, measure UV, cloud fractional coverage, sky polarization, computing the cloud base height and wind speed [20, 21].

3 Method -Weather classification and webcams

The proposed system comprises a webcam monitoring module, web interface and a weather classification engine. We will first describe the core of the system, namely the weather classification engine followed by a description of the web interface.

In order for the system to adapt to the conditions at each site a representative set of images are first collected over time so that there are good chances that one has captured the most representative weather conditions, that is, the learning phase. Then, k-means clustering is used to classify the training images according to the various features described in the following sections. The centeroids of each cluster are then used as reference values for the various weather conditions at each site.



a) Evening floodlight b) Night c) Daytime

Figure 1: Day versus night classification.

Day-night classification

Each image is first classified into day and night images on the assumption that it is hard to determine the weather from a dark night image. Night images are thus discarded. This is not to say that it is impossible to extract weather information from night images. Polar regions have midnight sun during the summer months and other sites are often artificially lit.

There are several ways to determine whether an image is a night image or day image. Images captured with high end cameras are encapsulated with EXIF information such that the exposure value can be computed and used directly [22]. Inexpensive webcam images often do not contain EXIF information and content based strategies are needed. Previous strategies include averaging the pixel intensity and classifying images according to the overall intensity [23]. However, this study proposes an even more efficient and reliable measure. First the image is converted to grayscale. Then dark pixels with a value of 35 or less are counted. This value was found through experimentation. During training K-means with a k-value of 2 are used to cluster the training images into day and night. This procedure reveals the threshold for day and night in terms of number of dark pixels. Figure 1 illustrates day-night classification.

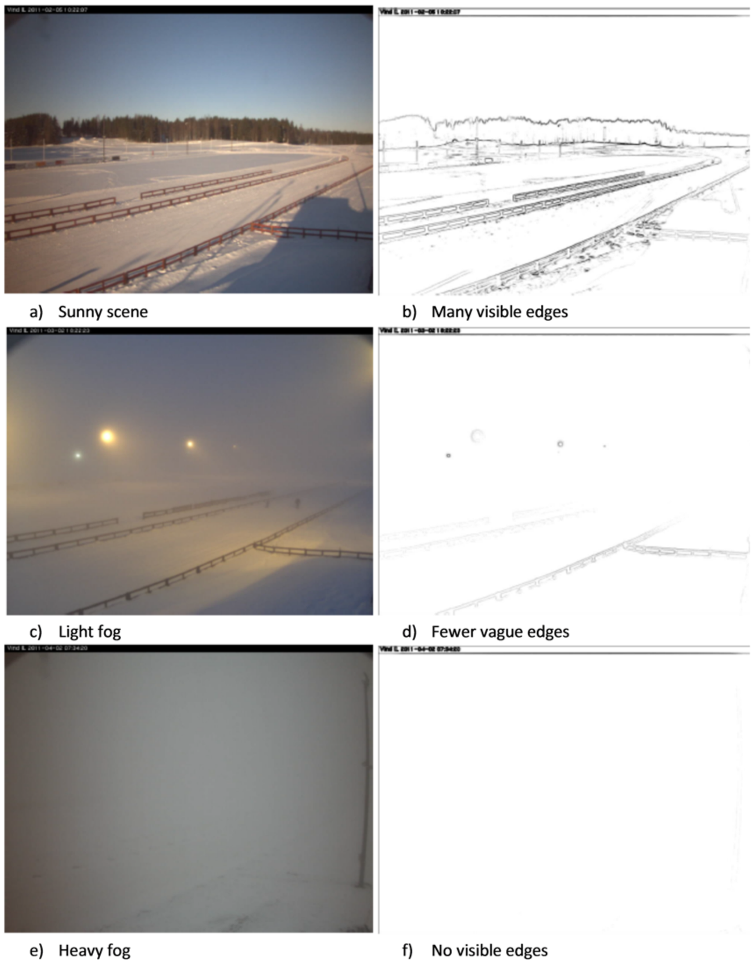


Figure 2: Sobel fog detection.

Fog detection

The next step is to apply fog detection to all day images. Fog detection is achieved using a Sobel operator for edge detection as proposed in the literature [17], namely:

$$g(x,y) = \sum_{k=-1}^{1} \sum_{j=-1}^{1} h_{k,j} f(x-j, y-k) \tag{1}$$

where the horizontal and vertical kernels h are defined as

$$\begin{pmatrix} -1 & 0 & 1 \\ -2 & 0 & 2 \\ -1 & 0 & 1 \end{pmatrix}$$

and

$$\begin{pmatrix} -1 & -2 & 1 \\ 0 & 0 & 0 \\ 1 & 2 & 1 \end{pmatrix}$$

Then the Sobel pixel value is given by

$$sobel(x,y) = 255 - \sqrt{g_{horiontal}(x,y)^2 \times g_{vertical}(x,y)^2} \tag{2}$$

Note that the Sobel image is inverted so that edges appear black and non-edges appear white when visually inspected. Next, the edge pixels are counted. A high edge count indicates non-fog, and a low edge count suggests that the image contains fog. To find the appropriate thresholds, k-means with k-value of 2 was applied to the training set. Fog detection with the Sobel operator is illustrated in Figure 2.

Sun-cloud classification

A very important aspect of the proposed system is the classification of weather into sunny versus cloudy days and the degree to which it is sunny or cloudy. In addition to temperature, the degree of sun or clouds is a key weather characteristic that users often are interested in. In this study several strategies for analyzing the degree of sun and clouds in the skies were explored.

A manual inspection of a large number of images revealed that the degree of sunny clear skies is strongly correlated with saturated colors. That is, images of sunny scenes contain more saturated colors than images of cloudy scenes that appear less saturated as cloudy scenes contain more white components. Thus, the overall level of saturation can thus be used to quantify the level of sun. We therefore converted the image from RGB into HSB (hue, saturation and brightness) and then summed the saturation of all the pixels. The correlation between saturation and sun is illustrated in Figure 3.

Another characteristic of webcam images are that they often contain large portions of sky, that is, typically, sky separated by the horizon and then the ground. On a sunny day the sky is blue, while on a cloudy day the sky is grey. We therefore devised a blue measure and a grey measure. The blue measure is obtained by counting the number of blue pixels, and summing the blue value of these pixels. We defined these pixels to have a hue in the range of 170-240 degrees on the color wheel with saturation over 0.05 to avoid white pixels. The blue-measure is thus:

$$blue = 1.05 - \frac{n_{blue}+1}{X \cdot Y} \times \frac{S_{blue}}{1.5} \tag{3}$$

where n_{blue} is the number of blue pixels, S_{blue} is the sum of all blue pixels and X and Y are the image dimensions. The grey measure was computed by counting all the

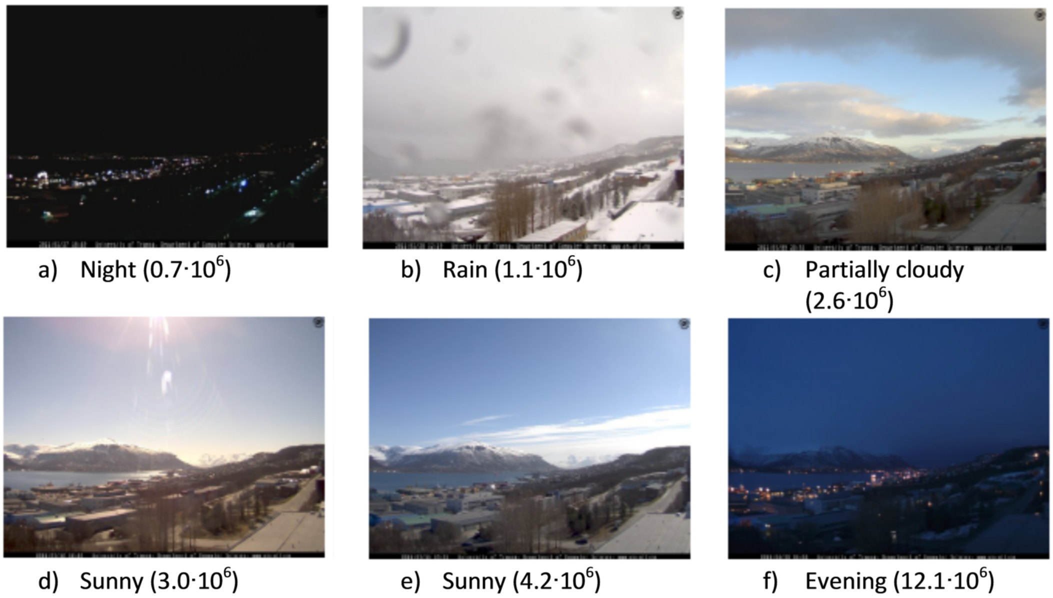


Figure 3: Saturation analysis based on images from a webcam in Tromsø, Norway. The sum of pixel saturations is given in parentheses.

pixels with saturation below 0.2 and brightness in the range 0.70-0.99. The blue and grey measurements are illustrated in Figure 4.

A few noteworthy exceptions to these observations were made. Just after sunrise and just before sunset the captured images are highly saturated with a bluish light irrespective of the weather. One way to distinguish sunset and sunrise images from sunny daytime images is to either look at the time of day of the particular location or consider the boundary of day and night. However, we discovered a more robust and simple strategy. Daytime images have a large range in intensity values from bright to dark, while the range of intensity values is narrower during sunrise and sunset as the overall light intensity is lower. Therefore, an image is tagged as sunny if it is highly saturated and the variance in intensity values is large. For each image the variance of all the pixel intensities are calculated to obtain a measure of spread in intensity. A large spread in intensity indicates daytime image and a low spread in intensity signals night, sunrise or sunset images. The importance of considering spread in intensity is illustrated in Figure 5.

These four parameters, that is, overall saturation, blue measure, grey measure and intensity spread are used with the k-means algorithm to classify images into cloudy, partially cloudy and sunny images.

User interface

The system was written in Phyton using the Django framework with JQuery for the user interface and javascript support, Matplotlib for plotting, SciPy and NumPy for mathematical computations, the Pyton Imaging Library (PIL) and OpenLayers for map functionality. The data for the map was obtained from OpenStreetMap.

Visitors browse the weather using the web application. By registering an account users can also register additional cameras by providing a URL to the webcam and indicate on the map where in the world the camera is located. Users can manage their list of webcams using their accounts.

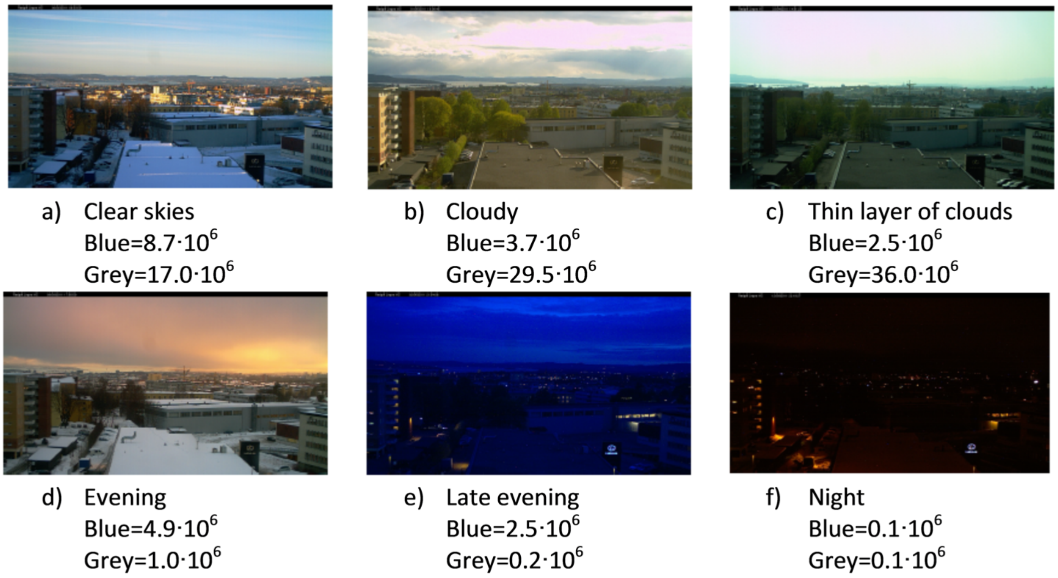


Figure 4: Blue and gray measures for a webcam in Oslo, Norway.

The user is presented with a world map with icons indicating the weather according to the analysis of the webcams using recognizable icons. Such weather icons have become culture neutral conventions for presenting weather. The user can zoom and pan to select a specific region on the map to see more detail. Detailed information is acquired by selecting a specific camera.

Often there are more than one webcam in a specific area. In such cases a voting algorithm is used to determine what weather to report for that region. That is, the weather with most votes is reported to the user. In the world view there can be large number of webcams in small areas and the webcam and the associated weather are also listed on the side of the map to help the user.

The web interface also contains experimental analysis functionality that was used during the development to tune the algorithms. This allows easy access to webcam images with certain features including visualization of the clustering analysis.

4 Experimental evaluation

Towards the end of the testing phase more than 70 cameras across the world were registered and the system collected approximately 7,000 images from these cameras each day, that is, approximately 100 images from each camera on a daily basis. An image was thus downloaded from each camera every 15 minutes. During the entire test-phase a total of 450,000 images were captured and analyzed.

To assess the correctness of the automatic weather reports two types of tests were conducted, a smaller test were the authors manually checked 100 random images and a larger test where 86 volunteers manually checked 860 random images.

Authors' evaluation

The test performed by the authors comprised selecting 20 different webcams and then manually classifying the weather shown on five random images for that camera. The following classes were used: cloudy, partially cloudy, sunny, foggy and night. Cameras

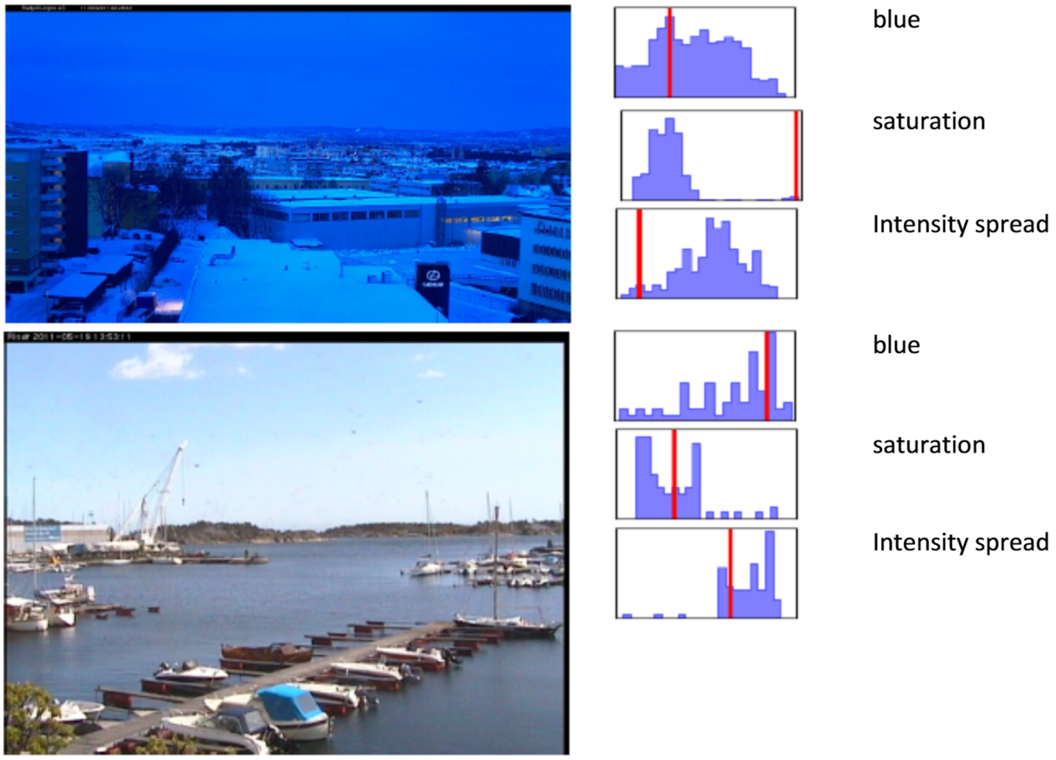


Figure 5: A night image and sunny image together with Blue-test, saturation and intensity spread distributions (blue histograms) for the images captured. The red lines indicate the cluster center for the sunny category. Note that standard deviation is used as the intensity spread measure shown.

not applicable to the tests were replaced. That is, cameras out of service, indoor cameras or cameras of low quality or with incorrect exposure settings that makes it difficult to manually determine the weather (see Figure 7). A total of 100 images were manually classified. The results of the manual versus automatic classification are shown in Table 1.

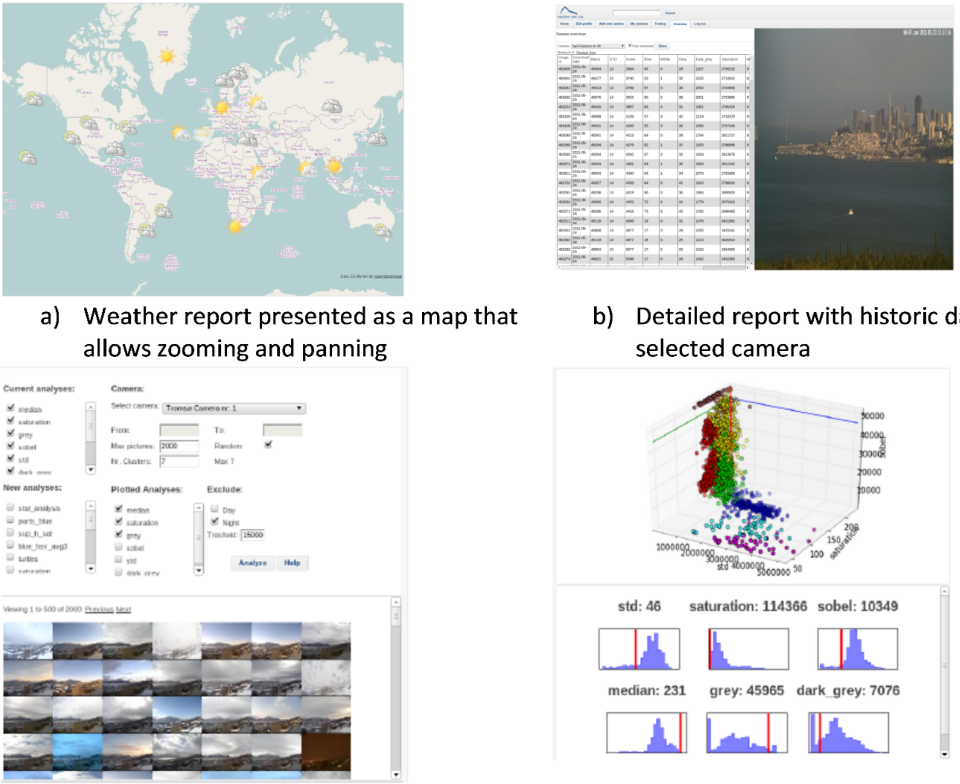
The experiment revealed an overall success rate of 67.3%, that is, 67.3% of the cases were correctly classified by the algorithm. This experiment included five classes and a result of 67.3% is much better than random, as a totally random guess would give a theoretical success rate of about 20%.

The performance metrics in Table 2 shows that classification of night yields the best performance, followed by the classification of cloudy weather and partially cloudy weather. The results for sunny weather are the worst. Table 1 shows that most images with sunny weather are correctly classified, but partially cloudy weather is often misinterpreted as sunny weather. However, note that the test set only comprise seven sunny images as the majority of images (23 images) depicted partially cloudy weather.

The classification of fog has a high recall, but low precision. That is, only one foggy image is incorrectly classified as a cloudy image, while several cloudy, partially cloudy and sunny images are classified as foggy images.

Volunteers' evaluation

The second test involved recruiting participants via Facebook to voluntarily participate in a manual classification task. Within two days 86 participants from all over the world



a) Weather report presented as a map that allows zooming and panning

b) Detailed report with historic data for a selected camera

c) Analysis for a webcam

d) Classsification analysis for a webcam

Figure 6: User interface screenshots

volunteered and manually classified 10 images each using a special purpose webpage. The images were randomly selected from the database. Therefore, it was not possible to successfully classify some of the images for the same reasons as explained earlier. The main difference being that these individuals did not report back and got replacement images or could mark images as unclassifiable. Since these images were not replaced it was natural to expect a lower success rate. Note also that these people were anonymous and located in different countries. We therefore cannot be completely certain that all these participants performed the task as instructed.

Table 3 lists the results of the volunteers' evaluation. The table lists the frequencies of datasets with various degrees of correct classifications. Two datasets were correctly classified, while two other datasets only had two correct classifications. A majority of the datasets had 6 correct classifications, and the overall success rate was 60.7%. These results confirm that the system is capable of detecting the correct weather condition in a majority of the cases.

5 Conclusions

This study proposes a strategy for reporting weather based on the existing infrastructure of openly available webcams. Although not as accurate as weather stations it provides an alternative source of high level weather information that can be used to corroborate other sources. The authors' evaluation achieves a success rate of nearly 67.3% and the volunteers' evaluation 60.7%. The current strategy performs best on night and cloudy images, but the strategy is also able to detect partially cloudy days, foggy weather and

Table 1: Results of manual classification.

		automatic				
		cloudy	partically cloudy	sunny	foggy	night
manual	cloudy	20	1	1	1	1
	partially cloudy	4	23	14	2	1
	sunny	2	3	7	2	
	foggy	1			5	
	night		1			15

Table 2: Weather prediction performance.

condition	no. images	precision	recall
cloudy	20	74.1 %	83.3 %
partially cloudy	23	82.1 %	52.3 %
sunny	7	31.8 %	50.0 %
foggy	5	50.0 %	83.3 %
night	15	88.2 %	93.8 %

clear skies. Future work includes improving the weather classification strategy, especially the detection of clear skies using shadow detection.

References

[1] T. Geller. Envisioning the Wind: Meteorology Graphics at Weather Underground. *IEEE Computer Graphics and Applications* 27(5), pp. 92-97 (2007).

[2] K. Sathita, H. Ochiai, and H. Esaki. RainWatch Project: Location-Awared Realtime Detection and Notification of Rain on Internet-Based Sensor Network. *in proceedings Ninth Annual International Symposium on Applications and the Internet*, IEEE Press, pp. 259-262 (2009).

[3] E. Bradley, D. Roberts, and C. Still. Design of an image analysis website for phenological and meteorological monitoring. *Environmental Modelling & Software* 25, pp. 107-115 (2010).

[4] M. Kammler and G. Schernewski. Spatial and temporal analysis of beach tourism using webcam and aerial photographs. *in Managing the Baltic Sea. Coastline Reports 2, G. Schernewski and N. Löser, Eds.*, pp. 121-128 (2004).

[5] S. i. Kako, A. Isobe, and S. Magome. Sequential monitoring of beach litter using webcams. *Marine Pollution Bulletin* 60, pp. 775-779 (2010).

[6] S. G. Narasimhan and S. K. Nayar. Shedding Light on the Weather. *in Proceedings of Computer Vision and Pattern Recognition*, IEEE Press (2003).

[7] K. Garg and S. K. Nayar. Detection and Removal of Rain from Videos. *in proceedings Computer Vision and Pattern Recognition*, IEEE Press (2004).

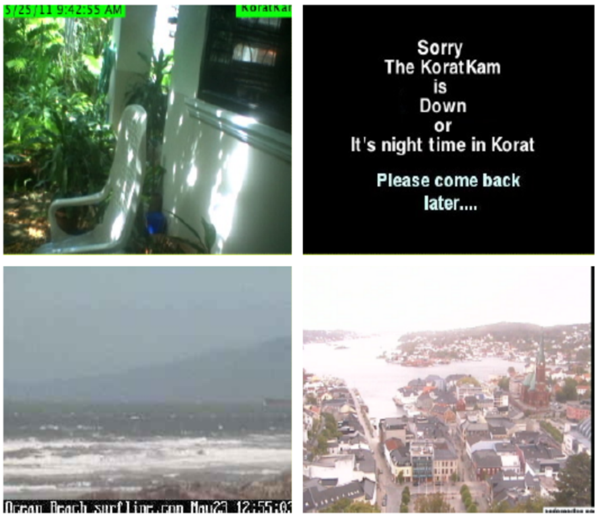


Figure 7: Difficulties encountered determining the weather manually. a) sky not visible, b) out of service, c) low image quality, d) over exposure.

Table 3: Frequency of correct classification in the volunteers' evaluation.

no. correct	Frequency
2	2
3	5
4	8
5	11
6	25
7	21
8	9
9	3
10	2

[8] W.-J. Park and K.-H. Lee. Rain Removal Using Kalman Filter in Video. *in proceedings International Conference on Smart Manufacturing Application*, pp. 494-497 (2008).

[9] F. Nashashibi, R. de Charrette, and A. Lia. Detection of unfocused raindrops on a windscreen using low level image processing,. *in proceedings 11th International Conference on Control Automation Robotics & Vision*, IEEE Press, pp. 1410-1415 (2010).

[10] J. Cai, M. Pervez, M. Shehata, R. Johannesson, W. Badawy, and A. Radmanesh. On The Identification of Snow Movements on Roads. *in proceedings SiPS'2006*, IEEE Press, pp. 357-361 (2006).

[11] Y. Wang, M. Tang, and G. Zhu. An Improved Cast Shadow Detection Method with Edge Refinement. *in proceedings Sixth International Conference on Intelligent Systems Design and Applications*, IEEE Press, pp. 794-799 (2006).

[12] F. E. Sandnes. Determining the Geographical Location of Image Scenes based on Object Shadow Lengths. *Journal of Signal Processing Systems for Signal, Image, and Video Technology* 65(1), pp. 35-47 (2011).

[13] T. Horprasert, D. Harwood, and L. S. Davis. A Statistical Approach for Real-time Robust Background Subtraction and Shadow Detection. *in proceedings IEEE ICCV*, IEEE Press (1999).

[14] P. KaewTraKulPong and R. Bowden. An Improved Adaptive Background Mixture Model for Realtime Tracking with Shadow Detection. *in proceedings 2nd European Workshop on Advanced Video Based Surveillance Systems, AVBS01*, Kluwer Adademic (2001).

[15] Y. Li, T. Sasagawa, and P. Gong. A system of the shadow detection and shadow removal for high resolution city aerial photo. *in proceedings XXth ISPRS Congress*, (2004).

[16] J. M. Wang, Y. C. Chung, C. L. Chang, and S. W. Chen. Shadow detection and removal for traffic images. *in proceedings IEEE International Conference on Networking, Sensing and Control*, IEEE Press, pp. 21-32 (2004).

[17] S. Bronte, L. M. Bergasa, and P. F. Alcantarilla. Fog Detection System Based on Computer Vision Techniques. *in proceedings 12th International IEEE Conference on Intelligent Transportation Systems*, IEEE Press, pp. 30-35 (2009).

[18] N. Jacobs, R. Souvenir, and R. Pless. Passive vision: The global webcam imaging network. *in proceedings Applied Imagery Pattern Recognition Workshop*, IEEE Press (2009).

[19] J.-F. N. Lalonde, S. G. Efros, A. A. What does the sky tell us about the camera?. *in proceedings European Conference on Computer Vision*, (2008).

[20] A. Heinle, A. Macke, and A. Srivastav. Automatic cloud classification of whole sky images. *Atmos. Meas. Tech.* 3, pp. 557-567 (2010).

[21] M. C. Allmen and W. P. Kegelmeyer. The computation of cloud base height from paired whole-sky imaging cameras. *Machine Vision and Applications*, 9(4), pp. 160-165 (1997).

[22] F. E. Sandnes. Where was that photo taken? Deriving geographical information from image collections based on temporal exposure attributes. *Multimedia Systems* 16(4-5), pp. 309-318 (2010).

[23] F. E. Sandnes. Towards calibration-free geo-localization of stationary outdoor webcams. *in proceedings of the Norwegian Computer Conference NIK 2010*, Tapir Academic Publishers, pp. 190-201 (2010).