HOW IS THE WEATHER: AUTOMATIC INFERENCE FROM IMAGES

Zichong Chen, Feng Yang, Albrecht Lindner, Guillermo Barrenetxea and Martin Vetterli

{zichong.chen, feng.yang, albrecht.lindner, guillermo.barrenetxea, martin.vetterli}@epfl.ch
School of Computer and Communication Sciences
Ecole Polytechnique Fédérale de Lausanne (EPFL), Lausanne CH-1015, Switzerland

ABSTRACT

Low-cost monitoring cameras/webcams provide unique visual information. To take advantage of the vast image dataset captured by a typical webcam, we consider the problem of retrieving weather information from a database of still images. The task is to automatically label all images with different weather conditions (e.g., sunny, cloudy, and overcast), using limited human assistance. To address the drawbacks in existing weather prediction algorithms, we first apply image segmentation to the raw images to avoid disturbance of the non-sky region. Then, we propose to use multiple kernel learning to gather and select an optimal subset of image features from a certain feature pool. To further increase the recognition performance, we adopt multi-pass active learning for selecting the training set. The experimental results show that our weather recognition system achieves high performance.

Index Terms— Weather recognition, panorama images, image segmentation, multiple kernel learning, active learning

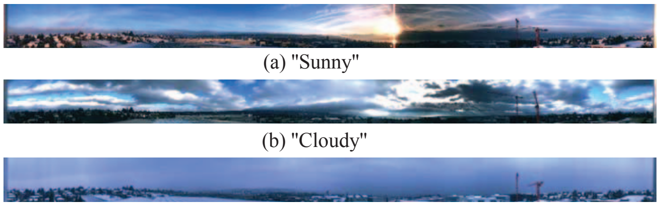
1. INTRODUCTION

Weather report is a traditional way to provide meteorological information. Due to the restricted density of weather stations (e.g. only around 30 major stations across Switzerland), low-cost wireless sensor networks have emerged to collect local environmental information [1] [2]. Among all available sensing capabilities, image sensors provide unique visual information of the target field. In particular, some non-measurable weather information can be obtained from images, such as the cloud amount defined by World Meteorological Organization (WMO Code 2700). However, retrieving such information autonomously remains a challenging problem. Typically, all the images are manually labeled with different weather condition semantics. This procedure requires a lot of labor.

To improve the labor efficiency, the following problem is considered: There are N raw images collected by a static environmental monitoring panorama camera, which is programmed to capture images periodically, and only a portion of J images can be manually labeled with a proper semantic term (e.g., sunny or cloudy). Then, all the other images need to be labeled automatically with a high confidence.

This problem is a specific application of image recognition [3]. There are some related work on weather prediction from images as in [4] [5] and the drawbacks are as follows:

1) The whole image was treated as input, which is inaccurate because not all parts are directly related to the weather condition.
2) Different image features (color, shape, etc.) were combined into a single feature vector for SVM training. This will increase the dimensionality dramatically. As a result, more training samples and computational power are required (*Curse of Dimensionality* [6]).
3) Not all image features are necessary for weather recognition task. However, to the best of our knowledge, there is no methodological approach existing for selecting an optimal subset of features from a certain feature pool.
4) Single pass SVM learning is inefficient for a learning task with training budget.



(a) "Sunny"

(b) "Cloudy"

(c) "Overcast"

Fig. 1. Panorama images taken from the roof of BC building at EPFL.

In this paper, we propose several methods to solve these problems and build a systematic weather inference framework. As the weather information is mainly concentrated in the cloud patterns, in Section 2, we propose a method to extract sky region to eliminate the disturbance of the foreground (e.g., buildings, mountains). In Section 3, we propose to use multiple kernel learning (MKL) to gather and select an optimal subset of image features from a feature pool. We adopt active learning technique for selecting training sets to increase the recognition performance. Section 4 evaluates the overall system using the panorama image dataset collected on the roof of BC building at EPFL[1]. Each image is categorized into three possible weather categories: sunny, cloudy, or overcast as shown in Fig. 1. Experimental results show that our system has high accuracy for the labeling.

2. SKY EXTRACTION

In our algorithm, we infer the weather information from the sky parts of the images. Thus the first step is to detect the sky parts in the images. Our sky extraction algorithm is based on two observations. First, clouds in the sky are dynamic, while camera and buildings are static. Secondly, the sky is at the top of the images, and buildings are at the bottom.

[1] `http://panorama.epfl.ch` provides high resolution (13200 × 900) panorama images from 2005 till now, recording at every 10 minutes during daytime.

978-1-4673-2533-2/12/$26.00 ©2012 IEEE

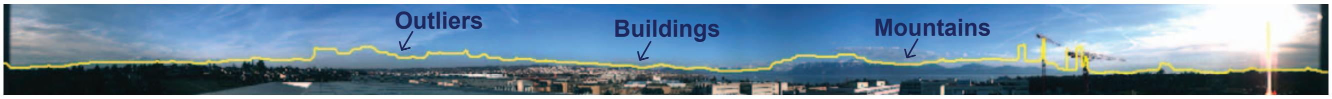


Fig. 2. Sky region extracted using Algorithm 1: the sky and the foreground buildings/mountains are separated by the yellow line.

The details of sky extraction is described in Algorithm 1. The main idea is to calculate the accumulative residual image from successive image frames, and then to apply morphology operations and thresholding in order to obtain a sky region mask. As the sky is more dynamic compared to the foreground, it has a higher residual value. We can discriminate the sky from buildings and mountains through the residual value. Some reflective facets of buildings can also vary greatly as light conditions change. To distinguish this with the sky region, we explore the second observation, i.e., weighting the residual map according to the vertical height in a squared manner.

Fig. 2 shows the sky region extraction result obtained from a sample image sequence of 30 frames (3 days). It can be seen that over 98% of the sky is correctly classified as the sky region, and all foreground buildings and mountains are classified as the non-sky region. The outlier (as denoted in Fig. 2) is attributed to the fact that there is some part of the sky comparatively static in a certain period. Nevertheless, as the camera is static in our setup, such errors will be eventually averaged out in a long run.

Algorithm 1 Pseudocode of sky region extraction algorithm

```
initialize WEIGHT: scaled in a squared manner w.r.t height
LAST=1st image frame, RESIDUAL=0, COUNT=0
threshold THRE= 40, refreshing period PERIOD=20
while a new image CURRENT is loaded do
    RESIDUAL = RESIDUAL + abs(CURRENT-LAST)×WEIGHT
    MASK = normalize(RESIDUAL) > THRE
    image erode and dilate applied to MASK to remove small fragments
    if COUNT%PERIOD==1 then
        RESIDUAL = RESIDUAL×MASK to cleanup accumulative errors in the foreground
    end if
    COUNT++
    LAST=CURRENT
end while
output MASK as the sky region
```

3. RECOGNITION

After the sky region is properly segmented from the raw images, our weather recognition system includes two main stages: feature extraction and learning. To overcome the feature gathering and selection problem as mentioned in Section 1, we propose to use multiple kernel learning (MKL) to select an optimal linear combination of image features. Then, we adopt active learning technique as a multi-pass recognition framework to further improve the recognition accuracy.

3.1. Feature gathering and selection

We use the "bag of words" method [3] to extract features from the raw image. This approach generates spatially uncorrelated features, and thus is suited for our problem because cloud patterns are also randomly distributed. The details of the feature extraction algorithm for each feature are explained in Algorithm 2.

Algorithm 2 Pseudocode of the feature extraction algorithm

```
1: Build a 2-D feature map (the size of image $3966 \times 270$) from the raw image, for a certain feature, e.g., HSV color.
2: Mask the feature map with the sky region, and divide the masked part into small tiles (e.g., 600 $32 \times 32$ tiles for each image).
3: Compute the local histogram of each tile (e.g., 600 $128 \times 1$ histograms for each image).
4: Aggregate histograms from all images, and cluster them into $K$ clusters using K-means clustering algorithm. Each tile is assigned an id in the range $1 \sim K$.
5: For each image, calculate the distribution of tiles' id (a $K \times 1$ histogram). This is the final feature vector extracted.
```

After different features are extracted (e.g., HSV, gradient, etc), we need to solve the problem on how to gather all these features for recognition. MKL [7] has been recently proposed for similar problem. The main idea behind this technique is to learn an optimal linear combination of feature kernels. In this way, the dimensionality of the problem is only increased with a few weighting coefficients, while the recognition accuracy is improved by gaining higher discriminative power.

Table 1. Features extracted from cloud patterns

Name	Description	Type
H,S,V	hue, saturation, and brightness of HSV color space	color
PHOW	SIFT on a dense gird at a fixed scale [8]	shape
LBP	local binary patterns (17 bins) in a texture [9]	texture
Gradient	gradient magnitude computed by Sobel operators	texture
Motion	residual computed from reference image	dynamics

Table 1 lists several features that we used in the experiments, which represent various aspects of cloud patterns. Note that for PHOW [8], the feature is not extracted by using Algorithm 2. PHOW itself computes SIFT at a given grid size (tile size), which can be directly clustered to form a "bag of words" feature. The Motion feature represents the dynamics of clouds, which utilizes redundant adjacent images (the original datasets contains six images per hour, while we only label one image per hour). It is based on the intuition that cloudy images may have higher motion than sunny and overcast ones.

To select a good subset of features from such a feature pool, we first use the MKL to learn an optimal weights for all features, and sort their weights in descending order. As the corresponding weighting coefficient of each feature represents its discriminative contribu-

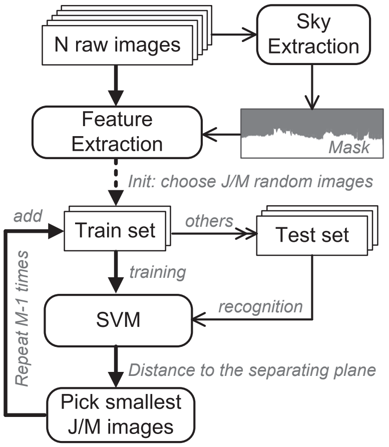


Fig. 3. Recognition routine via active learning: starting with J/M randomly chosen images, J/M additional images are appended to the training set through smart selection at each learning pass. After $M-1$ iterative passes, totally J training images are labeled by human (N unlabeled images in the beginning). The rest $N-J$ images are then labeled autonomously through recognition. At any pass, the training set and the test set constitute the whole image corpus.

tion to the overall recognition performance, thus it can be used as a measure for feature selection. We start from the most discriminative feature. Then, each feature is progressively added for testing according to the ranking, until the recognition performance stops to increase. These selected features provide the optimal choice.

3.2. Recognition via active learning

Traditional image recognition [3] assumes that the training set is fixed. In our case, however, the training set is not given in the beginning and needs to be labeled manually. Thus, we want a small but efficient training set. In the experiment, if the training set is drawn randomly from an image corpus, the recognition precision varies greatly with every iteration. Such phenomena is due to the fact that a random training set can not represent the whole image feature space well.

To improve the recognition performance, the training set is chosen through an iterative procedure, where SVM can query an oracle (human) to label some images during the process of learning. Such methodology is called *active learning* [10]. The basic principle is that in the recognition stage, the SVM returns the distance w_k between an unlabeled image I_k and the separating hyperplane. As a SVM finds the maximum-margin hyperplanes during the training stage, w_k can be treated as a natural measure of the recognition uncertainty of I_k. By sorting w_k of all the unlabeled images, we can select those with small values as the new training set in the next pass. Based on this idea, our weather recognition system is depicted in Fig. 3.

4. EXPERIMENTS

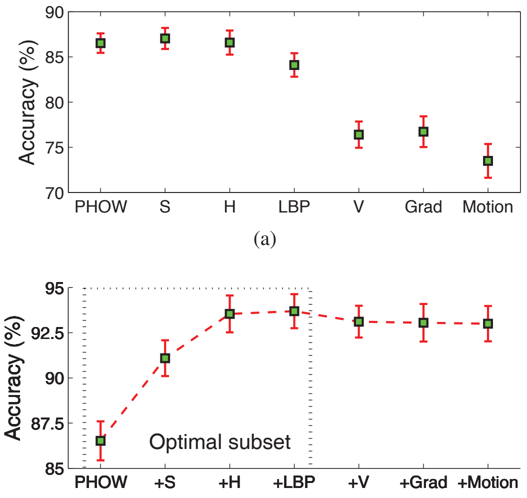
We evaluate our algorithm using 1000 images from our panorama image dataset (one image per hour in 2010, and downsampled to a resolution of 3966 × 270). Each image is categorized into three possible weather categories (as specified in Table 2). All images are manually labeled to serve as the ground truth, from which J images are chosen as the training set and the rest as test set (which is assumed to have no labels in recognition). The implementation of algorithms are based on VLFeat libraries [11].

Table 2. Weather categories and number of images per category.

Weather label	Description	Number of images
`sunny`	less than 50% of clouds	276
`cloudy`	between `sunny` and `overcast`	251
`overcast`	no visible blue sky	473

Table 3. Weighting coefficients given by MKL (discriminative power in descending order).

`PHOW`	`S`	`H`	`LBP`	`V`	`Grad`	`Motion`
0.308	0.302	0.300	0.282	0.278	0.275	0.268



(a)

(b)

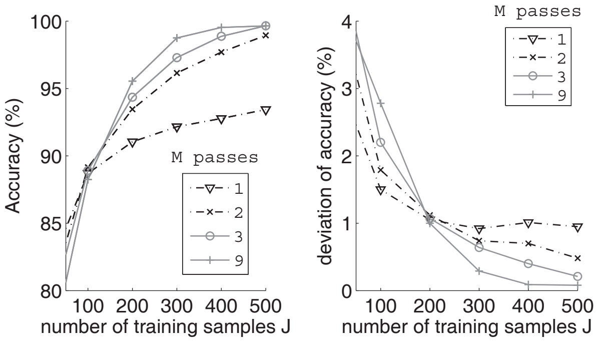
Fig. 4. Recognition accuracy of various features. 500 images are randomly chosen as the training set, and the recognition accuracy is recorded by testing the other 500 images ($J = 500$, no active learning). The error bars show the standard deviations obtained from 100 repetitions of each experiment. (a) Performance of each single feature. (b) Performance of feature combination with new features progressively added to MKL. The first four features provide the optimal choice.

The following parameters are chosen by cross validation and fixed throughout all evaluations: the local histogram bin number is 128 except for LBP, the number of clusters is 200, the tile size of Algorithm 2 is 32, the soft margin of SVM is 10; and the scale of the `PHOW` feature is 24.

4.1. Feature selection

To select an optimal subset of features from the feature pool as listed in Table 1, we first use the state-of-the-art algorithm of MKL [12] with Chi-Square kernel to learn the optimal weights for all features. Table. 3 shows the weights obtained by MKL. The features are sorted according to their weights.

We also evaluate the recognition accuracy of each single feature. For each test, 500 images are randomly chosen as the training set,



(a) (b)

Fig. 5. Recognition performance for different number of passes M (as defined in Fig. 3). PHOW+HS+LBP are used as features and learnt under the Chi-Square MKL. For each test, the recognition accuracy is evaluated from the remaining $1000 - J$ images. (a) Recognition accuracy versus the number of training samples J curves. (b) Corresponding standard deviation of recognition accuracy.

and 100 repetitions are carried out to obtain the mean and standard deviation of recognition accuracy. It is shown in Fig. 4a that the feature with higher weight has better discriminative power (recognition accuracy), as we mentioned in Section 3.1.

Knowing the relative discriminative power of features, we evaluate the recognition accuracy with the PHOW feature first, and then progressively add one more feature to MKL according to their rank in Table 3. As shown in Fig. 4b, the combination of the first four features outperforms other combinations. With these features, the shape, color and texture of images are conveyed respectively. In the following experiments, we choose these PHOW+S+H+LBP as the optimal feature selection for our task. Such procedure shows great advantage in practice, because it reduces complexity by avoiding unnecessary feature extractions, i.e., computation for V+Gradient+Motion can be skipped.

4.2. Active learning

We evaluate now if active learning can improve the recognition performance. Fig. 5a shows the recognition accuracy versus the number of training samples J curves, for different number of passes M, as defined in Fig. 3. When $J > 100$, multi-pass learning outperforms conventional single pass learning ($M = 1$) substantially. Fig. 5b shows the corresponding standard deviation of recognition accuracy. The stability of the recognition system is also improved using active learning method, especially when $J > 200$. These results suggest that with the help of active learning, our weather recognition system can reliably label most of the images. With 20% of images manually labeled, the system achieves 95% of accuracy. These results are substantially better than the reported performance in [4] [5], because we leverage the latest developments in computer vision, namely, multiple kernel learning and active learning, which are both missing in previous literatures.

It is worth mentioning that in Fig. 5, active learning has lower accuracy as compared to conventional method when J is smaller than a certain bound. This is due to the fact that when the number of training samples is severely insufficient, the multi-pass active learning system cannot learn well in the beginning (the initial number of training samples in active learning is just J/M).

5. CONCLUSIONS

We consider the problem of assigning weather labels, i.e., sunny, cloudy and overcast to panorama images. Given a certain human input constraint, our proposed system can automatically label the remaining images with a high confidence. We first propose a robust sky region extraction algorithm to filter out foreground interference. Then we use the state-of-the-art multiple kernel learning framework to gather and select a combination of image features for optimal discriminative power and low computational complexity. To get a smarter choice of training set, we use active learning to build a multi-pass learning/recognition system. The experimental results show that this system achieves a high confidence.

6. ACKNOWLEDGEMENTS

This research was supported by the National Competence Center in Research on Mobile Information and Communication Systems (NCCR-MICS, http://www.mics.org), and the ERC Advanced Investigators Grant of European Union.

The authors also would like to thank Weijia Gan and Prof. Sabine Süsstrunk for their helpful comments.

7. REFERENCES

[1] G. Barrenetxea, F. Ingelrest, G. Schaefer, M. Vetterli, O. Couach, and M. Parlange, "Sensorscope: Out-of-the-box environmental monitoring," in *Proc. IPSN '08*, 2008, pp. 332–343.

[2] Zichong Chen, Guillermo Barrenetxea, and Martin Vetterli, "Share Risk and Energy: Sampling and Communication Strategies for Multi-Camera Wireless Monitoring Networks," in *Proceedings of the 31st Annual IEEE International Conference on Computer Communications (INFOCOM 2012)*, 2012.

[3] G. Csurka, C. Dance, L. Fan, J. Willamowski, and C. Bray, "Visual categorization with bags of keypoints," in *Workshop on statistical learning in computer vision, European Conference on Computer Vision*, 2004, vol. 1, p. 22.

[4] M. Roser and F. Moosmann, "Classification of weather situations on single color images," in *IEEE Intelligent Vehicles Symposium '08*, june 2008, pp. 798 –803.

[5] Xunshi Yan, Yupin Luo, and Xiaoming Zheng, "Weather recognition based on images captured by vision system in vehicle," in *Proceedings of the 6th International Symposium on Neural Networks: Advances in Neural Networks - Part III*, 2009.

[6] A.K. Jain and B. Chandrasekaran, "Dimensionality and sample size considerations," in *Pattern Recognition in Practice*, P.R. Krishnaiah and L.N. Kanal, Eds., pp. 835–855. 1982.

[7] M. Varma and D. Ray, "Learning the discriminative power-invariance trade-off," in *Proceedings of the 11th International Conference on Computer Vision (ICCV)*, 2007.

[8] A. Bosch, A. Zisserman, and X. Muoz, "Image classification using random forests and ferns," in *Proceedings of the 11th International Conference on Computer Vision (ICCV)*, 2007, pp. 1–8.

[9] Timo Ojala, "A comparative study of texture measures with classification based on featured distributions," *Pattern Recognition*, vol. 29, no. l, pp. 51–59, 1996.

[10] Burr Settles, "Active learning literature survey," Computer Sciences Technical Report 1648, University of Wisconsin–Madison, 2009.

[11] A. Vedaldi and B. Fulkerson, "VLFeat: An open and portable library of computer vision algorithms," in *Proceedings of the international conference on Multimedia*. ACM, 2010, pp. 1469–1472.

[12] Francesco Orabona, Jie Luo, and Barbara Caputo, "Online-batch strongly convex multi kernel learning," in *Proceedings of the 23rd IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2010.