Automatic Watermeter Digit Recognition on Mobile Devices

Yunze Gao[1,2], Chaoyang Zhao[1,2], Jinqiao Wang[1,2], and Hanqing Lu[1,2]

[1] National Lab of Pattern Recognition, Institute of Automation, Chinese Academy of Sciences, Beijing, China
[2] University of Chinese Academy of Sciences, Beijing 100190, China
{yunze.gao, chaoyang.zhao, jqwang, luhq}@nlpr.ia.ac.cn

Abstract. Automatic watermeter digit recognition in the wild is a challenging task, which is an application of scene text recognition in the field of computer vision. In this paper, we propose an automatic watermeter digit recognition approach on mobile devices which consists of digit detection and recognition. Specifically, we adopt Adaboost with aggregated channel features (ACF) to detect watermeter digital regions, where the computation is accelerated by the fast feature pyramid technology. Then a small attention bidirectional long short-term memory (BLSTM) is designed for end-to-end digit sequence recognition. Convolutional Neural network (CNN) is exploited to extract discriminative feature and BLSTM is able to capture the rich context in both directions within sequence data. Moreover, an attention mechanism is added to weight the most important part of incoming image features. We validate the performace of our approach on the collected complex dataset. It contains various watermeter images in real scenario which has illumination changes, messy environment, half-digit and blurring. It is observed that the proposed algorithm outperforms existing methods. Our approach runs 10 fps with 96.1% accuracy on HUAWEI Mate 8.

Keywords: watermeter digit recognition, BLSTM, attention model

1 Introduction

It is very time consuming and labor intensive to record the digital number of watermeter manually from house to house. Therefore, if we can take a photo and recognize its digital value automatically, recoding the value of watermeter becomes much convenient and less mistakes caused by meter reader. Traditional approaches for watermeter recognition are mainly based on template matching [1] or BP neural networks [2]. These approaches usually involve several steps including detection, segmentation, binarization and recognition. Each step has several experienced parameters and rigid rules. Therefore, they cannot deal with complex scenes such as various watermeter types, different views and messy environment.

Watermeter digit recognition is an application of scene text recognition. Nowadays, text recognition in the wild has received intensive concerns from

numerous researchers [3, 4], which has a variety of applications, such as automatic car license plate recognition, sign reading in the driveless vehicle, and image retrieval. Traditional approaches [5, 6] focused on the conventional Optical Character Recognition (OCR) method by first segmenting individual characters and then recognizing these characters separately. The diversity of text patterns and blurring, backlight increase the difficulty of character segmentation. So the performance is confined to the inaccuracy of character-level segmentation. Furthermore, recognizing each character individually ignores the relationship between the characters. Recent studies regard scene text recognition as a sequence recognition problem without segmentation. Shi *et al.*[7] proposed a Convolutional Recurrent Neural Network (CRNN) to integrate CNN and RNN for text recognition. Lee *et al.*[8] designed an attention-based RNN approach model for OCR in the wild by weighted sequence modeling.

Different from the scene text, each digit of watermeter is surrounded by a rectangular box, and there exist some partial occlusions since the digits roll in the watermeter. Furthermore, the environment of watermeter is complicated and messy, which also increases the challenge of watermeter digit recognition. Liu *et al.*[1] captured watermeter images through a camera at a fixed angle to obtain the digit region directly, and used template matching to recognize each digit. Rui *et al.* [2] used feature pattern matching method to segment watermeter digits and trained two BP neural networks to classify full digits and half digits, respectively. Conventional watermeter digit recognition methods are inaccurate and inflexible, because the segmentation errors do harm to the recognition accuracy. Besides, the fixed viewpoint and the same feature pattern cannot be generalized to other situations, such as mobile devices and different kinds of watermeters.

In this paper, we propose an automatic watermeter digit recognition approach on mobile devices, which consists of digital region detection and recognition. To extract the feature efficiently in the detection, we adopt the ACF features which are single pixel values in the aggregated channels [9]. Then Adaboost is used to combine decision trees over these features to distinguish digits and background [10]. At the same time, the feature pyramid estimation is applied to accelerate the computation. For the digital recognition, we employ a sequential attention-based model that is specifically designed for sequential recognition. To begin with, a sequence of feature vectors is extracted by the BLSTM [11, 12] which is on top of CNN. According to the feature representation, the watermeter digits are predicted recurrently by an attention decoder [13]. The attention mechanism [14] can perform feature selection and recurrent network can learn the sequential dynamics of digits. Our system performs well in the task of automatic watermeter reading on mobile devices and can achieve higher accuracy than traditional methods.

2 The Proposed Approach

As shown in Figure 1, the watermeter image is captured by a mobile phone. The digital region of watermeter is detected by boosting ACF features. Then

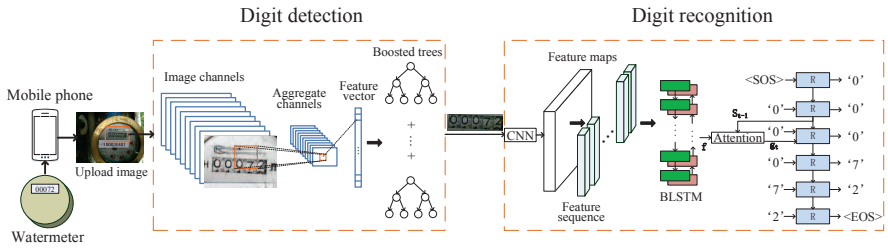

Fig. 1. Overview of mobile watermeter digit recognition.

the features of cropped digital region are extracted by CNN and BLSTM, and input into the attention decoder to obtain the meter reading.

2.1 Digit Detection

We adopt adaboost with ACF features to detect the digital region from a captured watermeter image. Since a cell phone often provides limited computational power in CPU, our detection part has to be realized with small computational complexity. Methods such as DPM or CNN based approaches are far beyond our option. In this paper, we choose ACF features and boosting framework for both fast-to-compute and high accuracy properties. To begin with, given an input image, several channels including normalized gradient magnitude, histogram of oriented gradients and LUV color channels are computed. These channels are divided into 4×4 blocks. Afterwards, the pixels in each block are summed, resulting in aggregate channel features. All pixel values are vectorized to form a pixel lookup table as the feature description of the image. To increase the accuracy, we use adaboost to train and combine multiple decision trees over these features to distinguish digits from background. Based on sliding windows over multiscale feature pyramid, the digital region of watermeter can be detected accurately.

When constructing the feature pyramid, we utilize a pyramid scale estimation [9] to accelerate the computation. Only the features for a sparse set of scales are computed by resampling the image and recomputing the channel features $C_s = \Omega(R(I, s))$, where C_s are the channel features at scale s, I is the original image, R is the sampling function and Ω is the channel computation function. For intermediate scales, C_s is computed by $C_s \approx R(C_{s'}, s/s')(s/s')^{-\lambda_\Omega}$, where s' is the nearest scale and λ_Ω is a channel specific power-law factor. With this method, the cost of feature computation is greatly reduced. The fast feature pyramid construction and scale estimation not only provide good performance but also guarantee the speed.

2.2 Digit Recognition

After digital detection, we adopt an end-to-end sequence recognition network to obtain the digit sequence. First, a feature sequence is generated by a network that

combines convolutional layers and recurrent layers. Next, an attention decoder predicts one digit at each step recurrently according to the feature representation.

Sequence Feature CNN has demonstrated strong ability to learn rich semantic description and robust representation from an input image [15], so we employ the convolutional layers to extract features of digital region. Each column of the feature map is corresponding to a receptive field of original image, which can be seen as the descriptor of the region. But this method ignores the dependence of adjacent regions, so we apply recurrent layer on the top of convolutional layers to obtain the long term context information. Before being fed into the recurrent layer, the feature maps are converted to a feature sequence, by extracting the same column of all feature maps and concatenating these columns into a vector as one element of feature sequence.

Considering the context information in the left and right are both helpful to recognize digits, we use the BLSTM to model the dependencies within the sequence in both directions. Then the BLSTM outputs the feature sequence that contains the latent relationship of these digits. The output sequence is denoted by $f = (f_1, f_2, \cdots, f_n)$, where n is the width of feature maps. The combination of convolutional layer and recurrent layer can effectively generate the discriminative feature for sequence recognition.

Attention Decoder According to the feature sequence, the digits can be predicted recurrently by an attention-based decoder, which is also a recurrent network with attention mechanism like [14]. At each step t, the LSTM cell is applied to decode the weighted feature and predict the digit. First, conditioning on feature sequence and previous recurrent cell state, attention weights are computed by scoring each element in f separately and normalizing the scores:

$$e_{tj} = w^T tanh(W^T s_{t-1} + V^T f_j + b) \tag{1}$$

$$\alpha_{tj} = \frac{exp(e_{tj})}{\sum_{j=1}^{n} exp(e_{tj})} \tag{2}$$

where $\boldsymbol{\alpha_t} \in R^n$ is a vector of attention weights; s_{t-1} is the recurrent cell state of previous frame; f_j is a vector of feature sequence; and W, V are weight matrices; w, b are weight vectors. The attention weight α_{tj} can be regarded as the relative importance of feature vector f_j. Then the input g_t of recurrent cell is computed by weighted sum of feature vectors based on the attention weights:

$$g_t = \sum_{j=1}^{n} \alpha_{tj} f_j \tag{3}$$

Following that, the internal state s_t of the recurrent cell is updated by taking input g_t, previous state s_{t-1} and output y_{t-1} into account. Next, the probability

estimation over the label is computed by:

$$y_t = softmax(U^T s_t) \tag{4}$$

The class with the highest probability is output as the predicted digit. Besides of ten digits, the labels also include "start of sequence"(SOS) which starts the prediction and "end of sequence" (EOS) which ends the prediction procedure. In this approach, at each step, we can focus on the most relevant content to make more accurate prediction. In addition, the recognition network also allows that the input and output sequence have arbitrary length, thus we can recognize various watermeters with different number of digits.

3 Experiment

3.1 Experiment Setting

There is no public watermeter images dataset to evaluate the performance. Therefore, we collect watermeter images by mobile phones in the wild and establish a complex dataset including 8781 watermeter images with five digits. These images are captured in the horizon or vertical angle including various illumination, messy situations, complicated environment, half digit and blurring. We randomly divide 8781 images into two parts: 7781 images are as train set and 1000 images are as test set.

Table 1. Network architecture of the digit recognition network.

attention-decoder 2 layers, 256 units per layer
BLSTM 1 layer, 512 units
Cov7 512, 2×2, stride 1×1, bn
Maxpooling 2×1, stride 2×1
Cov6 512, 3×3, stride 1×1
Cov5 512, 3×3, stride 1×1, bn
Maxpooling 2×1, stride 2×1
Cov4 256, 3×3, stride 1×1
Cov3 256, 3×3, stride 1×1, bn
Maxpooling 2×2, stride 2×2
Cov2 128, 3×3, stride 1×1
Maxpooling 2×2, stride 2×2
Cov1 64, 3×3, stride 1×1

In the experiment, we set 2048 decision trees with depth-two in the detection module. And we show the proposed digit recognition network in Table 1. Convolutions are performed with zero padding and ReLU activation function. During the phrase of training and testing, all the images are resized to 32-pixel height and meanwhile maintains original aspect ratio.

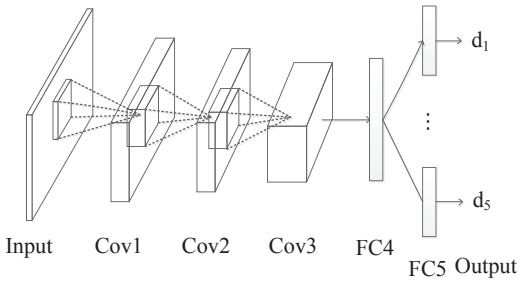

Fig. 2. The structure of Multi-softmax network.

3.2 Experiment Results

To verify the efficiency of the digital detection part, here we compare the ACF detector with boosting methods that combined with several other features, including HOG [16] and LBP [17]. The detection performance is shown in Figure 3. Here we use log-average miss rate for evaluation (lower the better). As shown in Figure 3, the ACF feature shows most promising result. Although boosting with HOG shows comparable result, its computation of 31 gradient orientations suffers more computational complexity compared to ACF. LBP shows the worst result on the digit detection task.

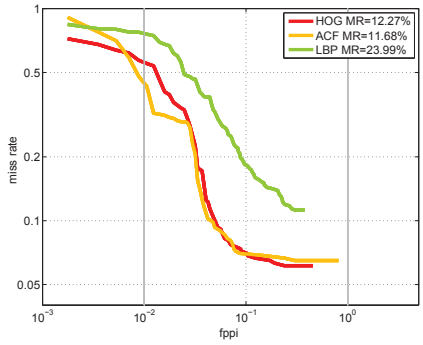

Fig. 3. Comparison of watermeter digits detection by different features.

We also conduct two baseline approaches even segmentation and Multi-softmax for recognition on the same test dataset. Even segmentation is to divide the character region into five parts evenly and each part is classified by an ordinary convolutional neural network. Multi-softmax is a method to recognize digits directly without segmentation, but no recurrent unit and attention mechanism.

As shown in Table 2, our approach outperforms even segmentation. For even segmentation, the poor performance indicates that segmentation mistakes could influence the final recognition result greatly.

Table 2. Comparison results with traditional methods.

Method	Accuracy(%)
Even segmentation	51.2
Multi-softmax	91.6
ours	96.1

For Multi-softmax approach, the network is illustrated in Figure 2. We use the cifar network, which consists of three convolutional layers and two fully connected layers. To predict multiple labels for an input image, the last fully connected layer and softmax are copied according to the number of watermeter digits. As shown in Table 2, our approach outperforms the Multi-softmax approach by 4.5%. This can be explained that more relevant context information is obtained by recurrent network and the attention mechanism.

Fig. 4. Examples of watermeter digits detection and recognition results.

Some examples of watermeter detection and recognition results are shown in Figure 4. By analyzing the mistake results, we find most of recognition errors appear in the case of half digits. As mentioned before, half-digit recognition is a difficult issue because of the various patterns.

Table 3. The time of each section.

	Detection	Recognition	Total
Time(ms)	50	70	120

Furthermore, the computation time of the watermeter digit recognition on HUAWEI Mate 8 is shown in Table 3. The digit detection spends 50 ms and the

digit recognition spends 70 ms. We can get the recognition result of an image within 120 ms.

4 Conclusions

In this paper, we propose an automatic watermeter digit recognition method on mobile devices which is composed of digit detection and recognition. For the detection part, the ACF detector with boosting method can extract the digital region accurately and efficiently. During digit recognition, BLSTM is utilized for the context information and attention mechanism is also employed for weighting image features. With this method, the digits can be recognized without segmentation, which improves recognition accuracy greatly. Some comparative experimental results show that our approach performs well on validity and practicability. Next, we will add Spatial Transform Network to recognize tilted and rotated watermeter images.

References

1. Liu, Ying and Han, Yan-bin and Zhang, Yu-lin: Image Type Water Meter Character Recognition Based on Embedded DSP. arXiv preprint arXiv:1508.06725 (2015)
2. Xiao-ping, Rui and Xian-feng, Song: A character recognition algorithm adapt to a specific kind of water meter. Computer Science and Information Engineering, 2009 WRI World Congress on.5, 632–636 (2009)
3. Neumann, Lukáš and Matas, Jiří: Real-time scene text localization and recognition. In Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on. IEEE, 3538–3545
4. Jaderberg, Max and Simonyan, Karen and Vedaldi, Andrea and Zisserman, Andrew: Synthetic data and artificial neural networks for natural scene text recognition. arXiv preprint arXiv:1406.2227 (2014)
5. Wang, Tao and Wu, David J and Coates, Adam and Ng, Andrew Y: End-to-end text recognition with convolutional neural networks. Pattern Recognition (ICPR), 2012 21st International Conference on, 3304–3308 (2012)
6. Bissacco, Alessandro and Cummins, Mark and Netzer, Yuval and Neven, Hartmut: Photoocr: Reading text in uncontrolled conditions. Proceedings of the IEEE International Conference on Computer Vision, 785–792 (2013)
7. Shi, Baoguang and Bai, Xiang and Yao, Cong: An end-to-end trainable neural network for image-based sequence recognition and its application to scene text recognition. IEEE Transactions on Pattern Analysis and Machine Intelligence, (2016)
8. Lee, Chen-Yu and Osindero, Simon: Recursive Recurrent Nets with Attention Modeling for OCR in the Wild. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2231–2239 (2016)
9. Dollár, Piotr and Appel, Ron and Belongie, Serge and Perona, Pietro: Fast feature pyramids for object detection. IEEE Transactions on Pattern Analysis and Machine Intelligence.36(8), 1532–1545 (2014)
10. Friedman, Jerome and Hastie, Trevor and Tibshirani, Robert and others: Additive logistic regression: a statistical view of boosting. The annals of statistics.28(2), 337–407 (2000)

11. Graves, Alex and Mohamed, Abdel-rahman and Hinton, Geoffrey: Speech recognition with deep recurrent neural networks. Acoustics, speech and signal processing (icassp), 2013 ieee international conference on, 6645–6649 (2013)
12. Hochreiter, Sepp and Schmidhuber, Jürgen: Long short-term memory. Neural computation.9(8), 1735–1780 (1997)
13. Shi, Baoguang and Wang, Xinggang and Lyu, Pengyuan and Yao, Cong and Bai, Xiang: Robust scene text recognition with automatic rectification. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 4168–4176 (2016)
14. Chorowski, Jan K and Bahdanau, Dzmitry and Serdyuk, Dmitriy and Cho, Kyunghyun and Bengio, Yoshua: Attention-based models for speech recognition. Advances in Neural Information Processing Systems, 577–585 (2015)
15. Krizhevsky, Alex and Sutskever, Ilya and Hinton, Geoffrey E: Imagenet classification with deep convolutional neural networks. Advances in neural information processing systems, 1097–1105 (2012)
16. Dalal, Navneet and Triggs, Bill: Histograms of oriented gradients for human detection. Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on.1, 886–893 (2005)
17. Ahonen, Timo and Hadid, Abdenour and Pietikäinen, Matti: Face recognition with local binary patterns. European conference on computer vision, 469–481 (2004)