BUILDING AN ACTIVE PALMPRINT RECOGNITION SYSTEM

Xuefeng Du[1], *Dexing Zhong*[1], *Huikai Shao*[1]

[1]School of Electronic and Information Engineering, Xi'an Jiaotong University,
28 West Xianning Road, Xi'an, Shaanxi 710049, P. R. China

ABSTRACT

Palmprint recognition allows accurate identity verification to build a security system. Recently, researchers introduce deep learning to this area that largely improves the recognition accuracy. However, as a supervised approach, its performance relies on availability of data and labels for every registered identity. For large-scale security systems, after image acquisition, we need to check the whole dataset and manually assign labels through comparison, which is a time-consuming task. Besides, labelling some redundant training samples contributes little to the recognition result. In this paper, we introduce an active learning framework to select the best sample set for label assignment. We regard the active learning as a binary classification task and attempt to make the labeled and unlabeled set indistinguishable. Experiments on different datasets demonstrate our model can reduce the annotation cost while achieving comparable recognition performance.

Index Terms— Biometrics, Palmprint Recognition, Active Learning, Information Security

1. INTRODUCTION

As one of emerging technologies of biometrics, palmprint recognition uses palmprint pattern and texture to identify a person. A typical palmprint recognition system consists of image acquisition, preprocessing, feature representation and classification modules [1]. Recently, researchers proposed several representative features, such as line direction [2], texture descriptors [3] and phase information [4]. A lot of them are hand-engineered features so the generalization ability is low. After feature extraction, certain classifiers are applied to conduct palmprint verification or identification.

Due to the progress achieved in deep learning, image representation contains more semantic information with higher adaptability to rotations and noise. Therefore, many researchers introduced deep learning models, specifically Convolutional Neural Network (CNN) to extract features from palmprint images [5, 6] and then fully connected layers were stacked to obtain the target labels. Typically, these end-to-end training relies on large amount of supervised training epochs to converge. As a result, we need to annotate many training samples carefully. When we build a large-scale palmprint recognition system, millions of images may be unlabeled after we obtain raw palmprint images. It is important to figure out how to train a promising classifier with as little labelled data as possible because not all samples contribute to the recognition accuracy. Meanwhile, labelling all samples is tedious and time-consuming. Therefore, it is necessary to build a palmprint recognition system where the annotation cost and training complexity are both minimized with fewer supervised training samples.

In this paper, we propose an active learning framework inspired by domain adaptation [7]. Our task is specified as building a labeled palmprint dataset for recognition out of a large unlabeled set. We start from a sparsely labeled palmprint dataset. At each iteration, we query some "salient" images from unannotated set according to query rules and label them for supervised training. To find an appropriate querying rule, we need to model the true data distribution of unannotated set using the right samples. One heuristic is simulating the data distribution using generative models, but they are hard to train and inappropriate for high-dimensional palmprint data. Therefore, we consider for every sample that how certain that it comes from the unannotated set compared to the annotated set. If we can infer with high confidence that it is from the unlabeled set, then adding it into training set will help the recognition model perform better. Therefore, we introduce a discriminator to predict the domain of features extracted by CNN (*i.e.* from unlabeled or labeled set?). Then, we select the samples with the highest confidence to come from unlabeled set and annotate them for further supervised tasks. The schematic diagram is shown in Fig.1.

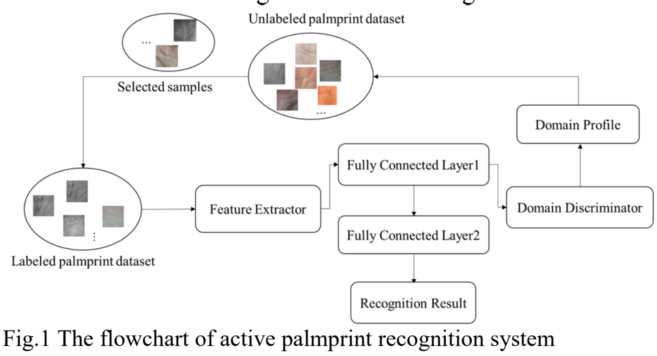


Fig.1 The flowchart of active palmprint recognition system

The rest of the paper is organized as follows: Section 2 reviews related works. Section 3 explains our proposed method. Section 4 describes the experiments and results. Section 5 concludes the paper.

2. RELATED WORKS

Palmprint recognition is popular these years. From literature, most successful recognition models are based on deep learning. They typically require many representative training samples and training iterations to converge. Recently, we used Siamese Network [8] to extract features from palmprints whose similarity were compared by neural networks. Besides, we also encoded palmprint images as hash codes to recognize them quickly [9]. The benchmark dataset we use contains over 6000 images but are already labelled. However, when using our dataset to simulate the practical application scenarios with less image acquisition constraints, the manual labelling work is unavoidable, which takes a lot of time and isn't solved in previous works.

As a popular method to avoid labelling cost, active learning aims to decide which subset of samples, if labelled, will lead to improvements in recognition accuracy with the same annotating budget. Three heuristics of active learning were proposed until now. Uncertainty-based methods scored unlabeled samples according to the classifier's uncertainty about their class, and annotated examples with the highest uncertainty [10]. Instead of scoring samples according to the output probabilities, some researchers selected the examples which were closest to the decision boundary of the model [11]. The second type focuses on the expected model change [12]. Huang *et.al.* [12] chose the examples which will influence the model most when labelled by calculating the expected gradient length (EGL) of the loss. The third type focuses on batch mode active learning which means selecting a set of images to label instead of querying one image at a time [13]. It attempts to choose a batch which ε- covers the unlabeled set closely and meanwhile forces the query to be diverse in order to reflect the multimodal distribution of unlabeled data. The final metric space where ε- cover is optimized is the learned representation of the data.

Theoretically, our method is closest to the third type, which uses the learned representation for judging whether to assign labels. Our method is also similar to domain adaptation [14]. It trains an image classifier and domain discriminator jointly, which attempts to make the source and target representation indistinguishable. However, our method trains the palmprint classifier and discriminator separately by firstly obtaining the learned representations and then optimizing on the domain adaptation-like objective functions. Compared to Generative Adversarial Nets (GAN) [15] which tries to fool a discriminator and changes data distribution in a differentiable way by a neural network, our method incrementally moves unlabeled data to labeled dataset, which greedily maximizes the discriminative loss in a non-differentiable way.

3. METHOD

Suppose we have a training set, which consists of our initial labeled dataset $LD(0)$ and its labels $L(0)$. The unlabeled palmprint images are denoted as $UD(0)$ with labels $U(0)$ (Only seen if it is queried). We first train a regular supervised model by minimizing the cross-entropy loss function on the testing set which has the same label space with the training set. Note that W_1 denotes the parameters of the initial recognition model.

$$\min_{W_1} Loss(W_1, \{LD(0), L(0)\}) \tag{1}$$

The structure of our recognition module is specified in Fig.2.

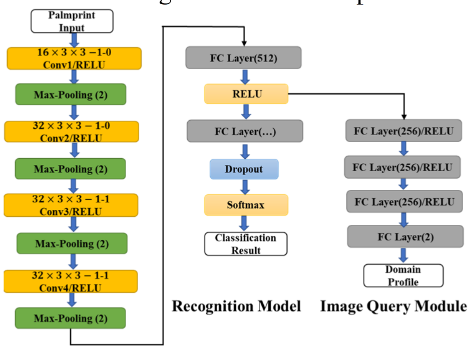


Fig.2 Structure of active palmprint recognition system. The configuration of each layer is shown in the box. For instance, "16×3×3-1-0" means 16 convolutional filters with 3×3 kernel size, 1 stride and 0 padding. "FC layer (256)" means 256 units in fully connected layer. Note that the shape of the final FC layer changes according to the number of classes. Max-Pooling (2) means 2×2 filters are used in the max-pooling layer.

To reduce the annotation cost, we query the optimal subset of palmprints $Sub(t)$ at every iteration t and assign labels to them. In order to model the true data distribution of the whole palmprint dataset using only the labeled set, an intuitive method inspired by domain adaptation is to train a domain discriminator D to distinguish whether the learned representation is from labeled or unlabeled set. Since we train a recognition model on the labeled training set, we want our model to work well on the unlabeled dataset or the combination of both. Therefore, our goal is to minimize the distribution difference between labeled and unlabeled images or their representations. Let $LR(t)$ and $UR(t)$ denote their respective outputs from the first fully connected layer. According to [14], the $\mathcal{H}$ divergence between source distribution D_S and target distribution D_T is formulated as follows:

$$d_{\mathcal{H}}(D_S, D_T) = 2 \sup_{f \in \mathcal{H}} |\mathbb{P}_{x \in D_S}[f(x) = 1] - \mathbb{P}_{x \in D_T}[f(x) = 1]| \tag{2}$$

Where $\mathcal{H}$ is a hypothetical class in a certain domain. $d_{\mathcal{H}}(D_S, D_T)$ is the value of $\mathcal{H}$ divergence. f is a characteristic function such that if x belongs to class f, then $f(x) = 1$. As a result, Equation (2) demonstrates for every hypothesis in $\mathcal{H}$, the maximum difference of the probabilities for data which belong to same class but from difference domains.

Therefore, we formulate the distribution difference similarly in our scenario as follows:

$$d_{\mathcal{H}}(\mathrm{LR(t)}, UR(t)) = 2 \sup_{f \in \mathcal{H}} \left| \frac{1}{|L(t)|} \sum_{x \in LR(t)} f(x) - \frac{1}{|U(t)|} \sum_{x \in UR(t)} f(x) \right| \tag{3}$$

$\frac{1}{|L(t)|} \sum_{x \in LR(t)} f(x)$ is the average probability distribution of the labeled dataset. Minimizing Equation (3) is equivalent to training a robust binary classifier D which can discriminate the labeled and unlabeled data and gradually can't distinguish which domain the data comes from. The relationship between its input and output is shown in Equation (4). The label is in the form of the regular one-hot encoding.

$$D(x) = \begin{cases} [0,1], x \in LR(t) \\ [1,0], x \in UR(t) \end{cases} \quad (4)$$

In each iteration, we first train the recognition model to find the maximum value of $d_{\mathcal{H}}(\mathrm{LR(t)}, UR(t))$, then we annotate the samples which contribute most to the difference. These images are intuitively predicted by the discriminator to come from unlabeled dataset with the highest confidence, which satisfy $arg \max_{x \in UR(t)} P(y \in U(t)|D)$. Since there is a huge representation difference between these unlabeled samples and labeled palmprints, labelling them should be informative for supervised classification on the updated labeled dataset. In general, we train this active recognition model greedily. At each iteration, it queries the most appropriate samples to label and we manually annotate and move them into the labeled dataset, which reduces annotation cost while simultaneously ignoring the redundant samples for higher efficiency. In that the number of training samples reduces, the neural network is easier to train with fewer epochs to converge.

4. EXPERIMENTS AND RESULTS

4.1 Datasets

4.1.1 PolyU Multispectral Database

PolyU multispectral dataset contains palmprint images under four spectral bands, *i.e.* blue (PB), red (PR), green (PG) and NIR (PN). 6000 grayscale images from 500 palms were collected for each band [16], including 195 males and 55 females. For each palm, 12 images were collected. The images are cropped to form Region of Interest (ROI) which are 128×128 pixels.

4.1.2 XJTU Uncontrolled Dataset

To simulate the practical image acquisition, we set up four palmprint datasets where the users can freely move their hands. The XJTU-A (XA) image dataset was built in the lab with proper illumination and low noise level. We used CMOS camera to collect 57 people's grayscale images of both palms. For each palm, we collect 10 images. XJTU-B (XB) was constructed outdoors so the quality of the images is lower. We collected 1960 RGB images from 196 different people. The images in dataset XJTU-C (XC) and XJTU-D (XD) were collected by iPhone6s with and without flash light in the RGB form. The number of images is 1960 and 1970 from 98 people respectively.

4.1.3 Tongji Palmprint Dataset

In Tongji (T) palmprint dataset, images were collected from 300 volunteers, including 192 males and 108 females in two sessions. In total, the database contains 12000 images from 600 different palms [17]. The exemplary images and their ROIs are demonstrated in Fig.3.

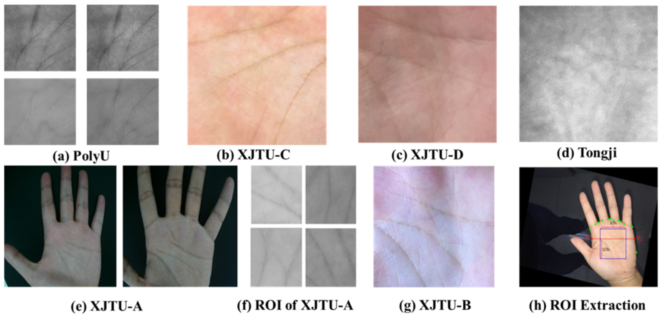


Fig.3 Different datasets used in our experiments. (h) denotes the ROI extraction process. We select two points on the edge of palm (the red point) as the reference points. A square area, whose side length was three-fifths of the length between the two points, was extracted as ROI.

4.2 Baselines

In order to validate the effectiveness of our proposed active palmprint recognition system, we compare our method with state-of-the-art active learning methods for neural networks. (1) **Supervised Training**: we train a fully supervised models on the whole training set. (2) **Random Query**: At each iteration, the selected batch is randomly selected from the unlabeled palmprint images. (3) **Uncertainty based Query**: The query batches are selected according to the classifier's uncertainty about their class. The examples with highest uncertainty are annotated [10]. (4) **Core-set Selection:** The diverse batch which ε- covers the unlabeled set closely is selected [13]. (5) **Bayesian based Selection**: Bayesian convolutional neural networks [18] is integrated with active learning for performing subset selection.

4.3 Active Palmprint Recognition

We test two settings for validating our active palmprint recognition model. The first is that only one querying method is used to select the unlabeled samples. The second is that we split the querying set into two parts and apply two querying methods separately. We finally add their respective queried images together to avoid bias caused by one method.

4.3.1 One Querying Method

First, we specify the experiment settings. We resize all the images to 128×128 pixels. For PolyU datasets, we use 4000 images from 500 classes as the training set, with remaining 2000 testing images. We randomly assign 100 labels to form the initial labeled dataset. Then, at each iteration out of 20 querying iterations, we add 100 samples to labeled palmprint dataset. In order to generate a diverse queried image set, we split 100 samples to 10 sub-batches. We then apply the domain discriminator to select for 10 times, with 10 samples each time being the most possible samples from unlabeled dataset. Therefore, we only label 2000 samples which reduces the annotation cost by a half. We run 70 supervised training iterations on the labeled set with the batch size to be 32. For the binary classier, we regard the representations from the training set as input. We train the classifier for 500 epochs with the batch size to be 128. We also apply early stopping when the accuracy reaches 98% to prevent overfitting. Adam

Optimizer [19] is used with the learning rate to be 1e-4. The hyperparameters for other 8 datasets are listed in Table.1.

Table 1 Experiment setting

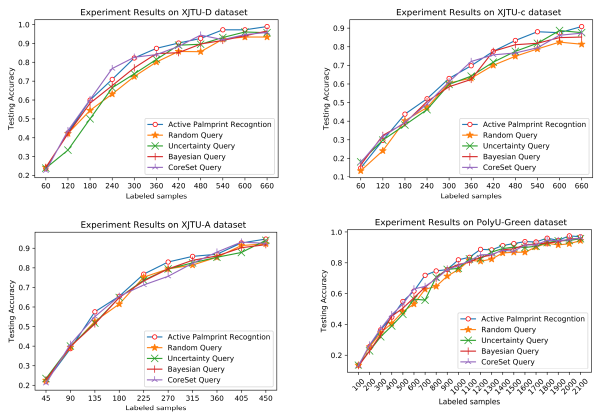
Dataset	Training set	Testing set	Initial dataset	Subset size number	Iteration
PolyU(4)	4000	2000	100	10/10	20
XA	678	452	45	9/5	9
XB	600	400	40	10/4	10
XC	1176	784	60	12/5	10
XD	1182	788	60	12/5	10
T	7200	4800	100	10/10	20

We test the recognition performance on the testing set in 9 datasets. The recognition accuracy at the last iteration is reported in Table.2. To illustrate, we replace the baselines by their initial character. For example, 'R' denotes Random query. 'U' denotes uncertainty-based methods. 'S' denotes supervised training on the whole training set.

Table 2 Comparative Recognition Accuracy

Dataset	R	U	C	B	S	Ours
PG	94.3%	95.6%	96.1%	94.2%	96.1%	**96.3%**
PR	92.5%	93.9%	94.8%	96.9%	96.3%	**96.8%**
PB	93.8%	95.1%	94.4%	91.8%	**96.4%**	95.7%
PN	91.1%	93.7%	93.7%	92.0%	**96.0%**	94.1%
XA	92.2%	94.0%	92.5%	91.8%	**94.6%**	94.5%
XB	77.7%	77.7%	79.4%	75.9%	**85.6%**	83.2%
XC	81.4%	85.7%	85.9%	88.8%	**93.4%**	90.9%
XD	93.3%	95.6%	95.7%	96.5%	97.8%	**97.9%**
T	94.0%	96.4%	95.5%	95.9%	**98.7%**	96.9%

In Table2, our active recognition framework performs close to or even greater than fully supervised models when tested on XD, PG and PR dataset. Meanwhile, it only requires a half of labeled samples so the labelling work to build the palmprint recognition system is reduced. Compared with baselines, our method performs better in both the controlled and uncontrolled datasets under the same annotation cost. Besides, all querying methods outperform random query while no obvious difference among the uncertainty-based, core-set based and Bayesian methods is observed. Generally, experiment results on uncontrolled datasets are worse than that on controlled dataset because of high intra-class variance caused by illumination and noise. The accuracy plot of different methods is shown in Fig.4.



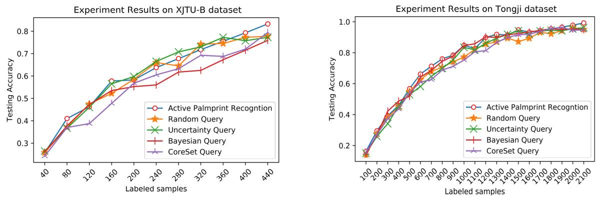


Fig.4 Comparative recognition results

In Fig.4, we demonstrate the process of labeled set expansion. In the initial expansion stage, adding new images makes a huge difference so the testing accuracy increases quickly. In the later stage, testing performance changes slowly. Testing accuracy of some baselines even drops down and fluctuated while our method keeps improving its performance because it queries the most discriminative samples to represent the whole dataset with the fewest labelling cost.

4.3.2 Querying Method Fusion

In general, our active recognition system uses a domain discriminator to distinguish representations. Sometimes, it is hard for the domain discriminator to distinguish samples which are nearby in the latent space but are from different datasets. In order to avoid the potential bias, we combine our method with other heuristic querying methods, which are (1) Random query, (2) Uncertainty-based query, (3) Bayesian query and (4) Core-set selection to improve the performance. We first use our proposed method to query the first half the query batches and then use another querying method to select the remaining half. The testing result is reported in Table.3.

Table 3 Comparative Recognition Results by querying fusion

Dataset	1	2	3	4
PG	94.6%	96.5%	**96.8%**	96.4%
PR	93.8%	96.1%	**97.4%**	97.1%
PB	95.6%	**96.4%**	96.3%	95.5%
PN	92.8%	**96.0%**	95.7%	94.1%
XA	93.3%	92.3%	**94.7%**	93.1%
XB	79.9%	82.3%	**83.7%**	77.0%
XC	88.1%	88.7%	**95.7%**	92.1%
XD	95.8%	95.8%	**97.9%**	97.7%
T	95.4%	**98.1%**	97.0%	97.0%

From Table.3, the combination with Bayesian models reaches better performance because Bayesian CNN provides our model with the uncertainty of the current image which validates that the combination with uncertainty-based methods performs better in dataset PR, PB and T. The combination with random query has the lowest accuracy.

5. CONCLUSIONS

In this paper, we propose a simple way to build a palmprint recognition system that largely reduces the annotation cost by half. We introduce a new batch-mode active learning framework to automatically query the necessary images for representing the whole dataset. To model the distribution of unlabeled set, we aim to make the distribution of labeled and unlabeled set indistinguishable by training a domain discriminator. Experiments on 9 different datasets validate the effectiveness of our model using single or combined querying methods. Further research will be improving the querying method which further reduces the annotation cost.

6. REFERENCES

[1] A. Kong, D. Zhang, and M. Kamel, "A survey of palmprint recognition," *Pattern Recognition,* vol. 42, no. 7, pp. 1408-1418, 2009.

[2] L. Lu, F. Gao, C. Qi, and C. Kim, "Palmprint recognition system on mobile devices with double-line-single-point assistance," *Personal & Ubiquitous Computing,* vol. 22, no. 1, pp. 93-104, 2018.

[3] W. Li, J. You, and D. Zhang, "Texture-based palmprint retrieval using a layered search scheme for personal identification," *IEEE Transactions on Multimedia,* vol. 7, no. 5, pp. 891-898, 2005.

[4] S. Iitsuka, K. Ito, and T. Aoki, "A practical palmprint recognition algorithm using phase information," in *International Conference on Pattern Recognition*, 2008, pp. 1-4.

[5] D. Zhong, Y. Yang, and X. Du, "Palmprint Recognition Using Siamese Network," in *Chinese Conference on Biometric Recognition*, 2018, pp. 48-55.

[6] D. Zhong, X. Du, and K. Zhong, "Decade progress of palmprint recognition: A brief survey," *Neurocomputing,* 2018/08/20/ 2018.

[7] S. Ben-David, J. Blitzer, K. Crammer, A. Kulesza, F. Pereira, and J. W. Vaughan, "A theory of learning from different domains," *Machine Learning,* vol. 79, no. 1-2, pp. 151-175, 2010.

[8] D. Zhong, Y. Yang, and X. Du, "Palmprint Recognition Using Siamese Network," in *13th Chinese Conference on Biometric Recognition, CCBR 2018.*, Urumqi, China, 2018, pp. 48-55.

[9] D. Zhong, M. Li, H. Shao, and S. Liu, "Palmprint and Dorsal Hand Vein Dualmodal Biometrics," in *2018 IEEE International Conference on Multimedia & Expo Workshops (ICMEW)*, San Diego, USA, 2018, pp. 1-6.

[10] Y. Yang and M. Loog, "A benchmark and comparison of active learning for logistic regression," *Pattern Recognition,* vol. 83, pp. 401-415, 2018.

[11] M. Ducoffe and F. Precioso, "Adversarial Active Learning for Deep Networks: a Margin Based Approach," *arXiv preprint arXiv:1802.09841*, 2018.

[12] J. Huang, R. Child, V. Rao, H. Liu, S. Satheesh, and A. Coates, "Active Learning for Speech Recognition: the Power of Gradients," *arXiv preprint arXiv:1612.03226*, 2016.

[13] O. Sener and S. Savarese, "Active Learning for Convolutional Neural Networks: A Core-Set Approach," in *International Conference on Learning Representations*, 2018.

[14] Y. Ganin and V. S. Lempitsky, "Unsupervised Domain Adaptation by Backpropagation," in *International Conference on Machine Learning*, 2015, pp. 1180-1189.

[15] I. J. Goodfellow *et al.*, "Generative Adversarial Networks," *arXiv preprint arXiv: 1406.2661*, 2014.

[16] D. Zhang, Z. Guo, G. Lu, L. Zhang, and W. Zuo, "An Online System of Multispectral Palmprint Verification," *IEEE Transactions on Instrumentation & Measurement,* vol. 59, no. 2, pp. 480-490, 2010.

[17] Z. Lin, L. Li, A. Yang, S. Ying, and Y. Meng, "Towards contactless palmprint recognition: A novel device, a new benchmark, and a collaborative representation based identification approach," *Pattern Recognition,* vol. 69, pp. 199-212, 2017.

[18] Y. Gal, R. Islam, and Z. Ghahramani, "Deep Bayesian Active Learning with Image Data," in *International Conference on Machine Learning*, 2017, pp. 1183-1192.

[19] D. P. Kingma and J. Ba, "Adam: A Method for Stochastic Optimization," *arXiv preprint arXiv: 1412.6980*, 2014.