EVALUATION FRAMEWORK ON TRANSLATION-INVARIANT REPRESENTATION FOR CUMULATIVE FOOT PRESSURE IMAGE

Shuai Zheng, Kaiqi Huang, Tieniu Tan

National Laboratory of Pattern Recognition,
Institute of Automation, Chinese Academy of Sciences
{szheng,kqhuang,tnt}nlpr.ia.ac.cn

ABSTRACT

Ground reaction force can be used to distinguish different human gait like limb movement. Cumulative foot pressure image is a 2-D data that recorded the spatial and temporal change of ground reaction force during one gait cycle. However, when putting it into practice as a new biometric for gait recognition, it suffers from the problem of large translation variation within class caused by wearing different shoes and walking in different speed. In this paper, an evaluation framework is proposed to address the problem. The framework consists of a database containing cumulative foot pressure images and a well designed benchmark. The data are collected from 118 subjects. A locality-constrained sparse coding scheme is developed to be compared with the benchmark PCA approach. Experimental results show the potential of evaluation framework for evaluating the translation-invariant power of image representation algorithms.

Index Terms— Ground Reaction Force, Foot Pressure, Translation Invariant Representation, Gait Recognition

1. INTRODUCTION

Ground reaction force and its variation footprint can be used as evidence like limb movement to identify the unique gait of pedestrian [1]. This attribute allows us to identify a person in many scenarios including the some scenes without monitor, jailhouse security system, entrance of Japanese house, biomedical system and shoes designing purpose application. Previous works [2] have been developed based on the unique limb movement pattern, while few works have been proposed based on the ground reaction force. The ground reaction force and footprint are not as popular as the limb movement does in pattern recognition and image process communities. Ground reaction force is easily affected by the noise since it is a 1-D signal which records the temporal information of gait. Recognizing footprint is intractable because it only contains the rough foot pressure distribution.

In these cases, we consider that recognizing cumulative foot pressure image instead of ground reaction force or footprint. Cumulative foot pressure image is a 2-D cumulative ground reaction force during one gait cycle. Different from previous foot pressure data[3, 4, 5] like ground reaction force or footprint, it provides both temporal and spatial distribution information of the gait. Different individual has his/her distinct gait, hence they have their own unique cumulative foot pressure image. However, recognizing cumulative foot pressure image as a biometric is still challenge since wearing different shoes bring the large inner-class appearance variation. In this paper, we address the problem of developing a foot pressure image representation to achieve tradeoff between discriminative power and translation-invariant ability for foot pressure image recognition.

Finding translation-invariant and discriminative representation for object recognition is also one of the key problems in computer vision and pattern recognition. Numerous descriptor-based approaches [6, 7, 8] are proposed in many other representative challenge recognition dataset such as MINIST and Caltech 101. Some low-level feature descriptors such as SIFT and HOG are designed to be invariant to minor translations of input images. However, only utilizing these low-level features as image representation leads to lose the image structure information, which is in terms of spatial distribution of object components. This information is crucial for both object recognition in life images and shoes-invariant cumulative foot pressure image recognition. Inspired by the evidences from neuroscience and cognition, some mid-level feature representation is proved to be potential in preserving both discriminative ability and translation-invariant ability for effective object recognition. Recent state-of-the-art mid-level feature representation approaches (e.g. bag of words, sparse coding and locality-constrained linear coding) in most challenge PASCAL VOC object recognition dataset show that sparsity and locality constrained are both essential for obtaining a discriminative and translation-invariant feature representation [9, 10, 11]. Hence, we consider integrating sparsity and locality constrained in the mid-level feature representation learning for cumulative foot pressure image recognition.

In this paper, we proposed an evaluation framework, benchmark algorithm. The evaluation framework collects the
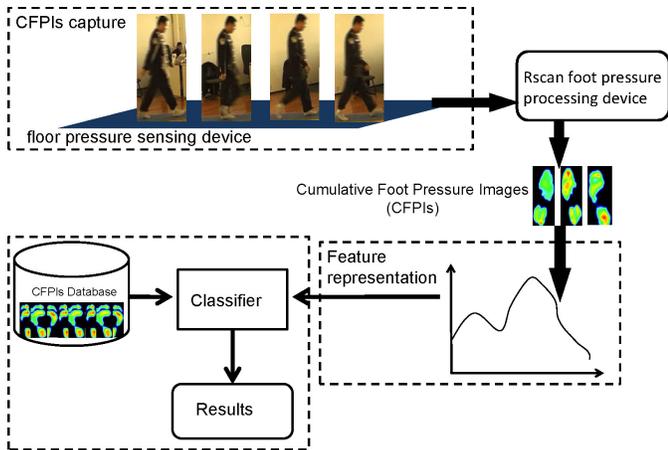
Fig. 1. Setup for cumulative foot pressure image collection.

Fig. 2. Cumulative foot pressure image under different wearing conditions.

data as Fig.1 illustrates. When a person walks through foot pressure measurement floor, cumulative foot pressure images are acquired. There are total 3000 images collected from 121 subjects. The representation algorithm for cumulative foot pressure image can be conducted and compared based on the dataset. Inspired by the progress in sparse coding[12] and convolutional deep belief network[7] ,we proposed locality-constrained sparse coding model as a comparison algorithm with traditional benchmark algorithm PCA plus LDA.

The remainder of this paper is organized as follows. In Section 2, an onverview introduction of cumulative foot pressure image dataset is introduced. In Section 3, we present a translation-invariant representation benchmark algorithm for cumulative foot pressure image. In Section 4, exmperiments are reported. Conclusions and future works are given in Section 5.

2. FOOT PRESSURE IMAGE DATABASE

There are many reports proposed on foot pressure image or footprint, but there is no open database for evaluating translation-invariant representation of cumulative foot pressure image. As far as we know, our cumulative foot pressure image data is the first open database.

As the framework focuses on the robustness of recognition algorithm, the images in cumulative foot pressure image database are all well selected and aligned. We set up a access scenario specially for the data acquisition as illustrated in Fig. 1. When a subject walks through foot pressure measurement floor, cumulative foot pressure images are acquired. One RScan[13] USB pressure sensor setup on the floor as a carpet. The size of the pressure sensor is 3mx1m. Hence, there are enough cumulative foot pressure images would at least be collected for one individual at one time. Fig.2 shows some sample images from the cumulative foot pressure image database.

We evaluate robustness of recognition algorithm in two aspects: walking speed effects and different shoes wearing conditions.

When a subject walked in the scene, he/she was asked to walk naturally along a straight line through the pressure sensor 5 times first and 5x88 normal walking cumulative foot pressure images were captured for each subject. There are 20 female subjects, 66 male subjects. All subjects are Asian aged between 20 and 60. The data are representative and esnure the future experiments to be not limited in any specific group of people. After normal walking, the subjects were asked to walk fast through the pressure sensor 5 times. In that case, there are totally 10x88=880 cumulative foot pressure records, three cumulative foot pressure images for each record, so there are totally 2640 images for considering the effect of walking speed. These subjects were all asked to walk without their shoes.

To evaluate the algorithms' robustness to translation, we also collect the data from the walking subjects with different shoes. To compare fairly, 30 subjects were asked to walk through pressure sensor without shoes wearing 2 times, wearing shoes 4 times. There are 6 cumulative foot pressure images were captured for each subject. 30 subjects were asked to walk with running shoes, Chinese cloth shoes, and leather shoes. Totally, there are $30 \times 6 \times 3$ cumulative foot pressure images in database. There are 6 female subjects, 24 male subjects. Most subjects are young people and they aged between 20 and 40.

All the images were stored as image files encoded with bmp codec. The image size is normalized as 40x90. It is a large gait database in terms of subjects number. The NLPR cumulative foot pressure image dataset is provided at web site[14].

3. TRANSLATION-INVARIANT REPRESENTATION

In this section, we present a traditional benchmark algorithm PCA plus LDA and a locality-constrained sparse coding

	PCA+LDA+NN	LSC+LDA+NN
correct rates	84%	90%

Table 1. Experiments are conducted under the condition that probe data and gallery data are collected from normal walking subjects. The data is collected from 88 subjects.

for representing cumulative foot pressure image, inspired by the convolutional deep belief network[7] and supervised translation-invariant sparse coding[12, 15].

The locality-constrained sparse coding model the local feature descriptors $X \in R^m$ from cumulative foot pressure image using a linear combination of dictionary matrix $D \in R^{m \times n}$ and coefficients $z \in R^n (n > m)$. The optimization process follows the following convex formulation:

$$L = \min \frac{1}{2} \|x - Dz\|_2^2 + \lambda \sum_i |w_i z_i|. \quad (1)$$

Where w_i is the Euclidean distance among z_i and its nearest Gaussian distribution center. There are two terms in equation 1, one is optimizing the reconstruction error penalty, another one is minimizing the regularization penalty. These constrains ensure the sparsity constrain in feature space, while the w_i ensure the locality constrain in spatial image space.

4. EXPERIMENTS

We designed two sets of experiments (Experiment Set A and B) for evaluate the translation-invariant representation for cumulative foot pressure image in recognition system. Experiment Set A is for investigating how walking speed affect the recognition performance and an algorithm's robustness to different walking speed. For each individual, there are 5 normal walking sets, 5 fast walking sets. Experimental results on benchmark algorithm are shown in Table 1. Experiment set B is for investigating how shoes wearing condition affect performance and an algorithm's robustness to different shoes wearing condition. In Experiment set B, we consider the two scenarios: collecting probe data from barefoot, gallery data from shoes wearing foot; collecting probe data from shoes wearing foot, gallery data from shoes wearing foot.

In this section, we report experimental results based on the two proposed cumulative foot pressure dataset: barefoot dataset and shoes-invariant dataset. One dataset contains all barefoot cumulative foot pressure images and its corresponding gender and identity label. It is collected from 88 sub-

	PCA+LDA+NN	LSC+LDA+NN
correct rates	77%	84%

Table 2. Experiments are conducted under the condition that probe data are collected from fast walking individual while gallery data are collected from normal walking individual.The data is collected from 88 subjects.

jects, including 68 male and 20 female. Another dataset contains barefoot cumulative foot pressure images, wearing cloth shoes one, wearing running shoes one and wearing leather shoes one. It is collected from 30 male subjects. These subjects came from different areas and they are constituted with different age groups from 22 to 50. Hence the experimental results are not limited to a specific group of people.

There are three criteria to evaluate the performance of our recognition system. First, we consider recognition correct rates, since it is widely used. Besides, we also consider the other two metrics false acceptance rate (FAR) and false rejection rate (FRR) as performance evaluation criteria. FAR consider the probability that the system incorrectly matches the probe data to a non-matching gallery data. It measures the percent of invalid probe data which are incorrectly accepted. FRR consider the probability that the system fails to detect a match between the probe data and a matching gallery data. It measures the percent of valid probe data which are incorrectly rejected. We also consider the FAR and FRR as the evaluation criteria to evaluate the performance in translation-invariant representation for cumulative foot pressure image.

Walking in each time, each subject leaves three cumulative foot pressure, we only considered the first left foot pressure image to use in experiments (You can also consider the first right foot pressure image.). For experiment set A, the experiment dataset are randomly participated into training and testing data set, where training set contains the data from 44 subjects and the testing set contains the remaining 44 data. For experiment set B, the experiment dataset are also randomly participated into training and testing dataset, where training set contains the data from 10 subjects and the testing set contains the data from 20 subjects.

In the following, recognition algorithms are evaluated as an example by the metrics proposed in this paper. Two components in recognition system play great effect on the performance. One is representation method, while another is classifier. We also choose the Dense SIFT descriptor as our low-level local feature descriptor. In out setup, the SIFT descriptor are extracted from patches densely located by every 4 pixels on the image, under the scale, 4x4 respectively. The dimension of each SIFT descriptor is 128. Here, we focus on the representation method. Hence, to simplify the classifier design, we choose to use Fisher Linear Discriminant Analysis (LDA) and Nearest Neighborhood (NN) as our classifier.

We compare Locality-Constrained Sparse Coding (LSC) and well known Pricipal Components Analysis (PCA) algorithms. PCA is our benchmark considering the similar method described in [16, 1]. We set reduced dimensionality as 200. Dictionary size is set as 128 and regularization penalty term is set as 0.15.

Table 1 and Table 2 show the experimental results of comparison between different methods. Experimental results are reported based on the optimal results from 10 times randomly separate the 30 subject datasets into training set (10 subjects)

	PCA+LDA+NN	LSC+LDA+NN
FAR	1.5%	1.0%
FRR	29%	20%

Table 3. Experiments are conducted under the condition that probe data are collected from shoes wearing foot while gallery data are collected from bare foot.The data is collected from 30 subjects.

	PCA+LDA+NN	LSC+LDA+NN
FAR	3.5%	4.2%
FRR	1%	1.1%

Table 4. Experiments are conducted under the condition that probe data are collected from shoes wearing foot while gallery data are collected from shoes wearing foot.The data is collected from 30 subjects.

and testing set (20 subjects). LSC and PCA are all conducted under the same setting. Based on the results, LSC+LDA+NN method outperforms the PCA+LDA+NN methods in terms of recognition correct rates, since the spartsity and locality penalty term is used to ensure representation to preserve the salient structure information from cumulative foot pressure images. The shoes-invariant cumulative foot pressure image recognition is still a difficult problem considering the results in Table 3, and Table 4.

5. CONCLUSIONS AND FUTURE WORK

In this paper, we proposed an evaluation framework on translation-invariant representation for cumulative foot pressure image. There are a large cumulative foot pressure image dataset in the proposed framework. There are total 118 subjects in the dataset. The suitable age and gender distributions in data ensure the representative conclusions made based on it. In addition, two experiment sets are established to evaluate the representation model's robustness to translations caused by different shoes wearing condition and different walking speed.

To achieve the trade-off between discriminative power and translation-invariant ability, this paper presents that it will bring some benefits that integrating sparsity constrained and locality constrained on the latent embedding of feature representation. The sparsity constrain brings the advantages of lower reconstruction error while the locality constrain considers the spatial distribution information of local features in image. These two benefits lead a better performance, comparing with original PCA benchmark. In future work, more image spatial distribution information of the local feature descriptors should be considered.

The framework provides a platform to evaluate the performance of different recognition algorithms using cumulative foot pressure image. In the future, we will enlarge the database to include the corresponding data such gait video

etc. A more systematic evaluation for cumulative foot pressure image representation in recognition system as LFW in face recognition is also our goal.

Acknowledgement

6. REFERENCES

[1] P.C.Cattin and D.Zlatnik, "Biometrics system using human gait," in *Mechatronics and Machine Vision in Practice*, 2001.

[2] D.Tan, S.Yu, K.Huang, and T.Tan, "Walker recognition without gait cycle estimation," in *ICB*, 2007.

[3] K.Nakajima and Y.Mizukami, "Footprint-based person recognition," *IEEE Trans. on Biomedical Engineering*, 2000.

[4] R.B.Kennedy, "Uniqueness of barefeet and its use as a possible means of identification," *Forensic Science International Journal*, 1996.

[5] S.P.Moustakidis and J.B.Thecharis, "Subject recognition based on ground reaction force measurements of gait signals," *IEEE Trans. on SMC Part B*, 2008.

[6] D.G.Lowe, "Distinctive image features from scale-invariant key points," *International Journal of Computer Vision*, 2010.

[7] H.Lee, R.Grosse, and A.Ng, "Convolutional deep belief networks for scalable unsupervised learning of hierarchical representation," in *ICML*, 2009.

[8] K.Kavukcuoglu and M.A.Ranzato, "Learning invariant features through topographic filter maps," in *CVPR*, 2009.

[9] M.A.Ranzato and C.Poultney, "Efficient learning of sparse representation with an energy-based model," in *NIPS*, 2006.

[10] M.ranzato, F.Huang, and et al., "Unsupervised learning of invariant feature hierarchies with application to object recognition," in *CVPR*, 2007.

[11] J.Wang, J.Yang, K.Yu, F.Lv, T.Huang, and Y.Gong, "Locality-constrained linear coding for image classification," in *CVPR*, 2010.

[12] J.Yang, K.Yu, and T.Huang, "Supervised translation-invariant sparse coding," in *CVPR*, 2010.

[13] R.S.Scan Lab, "an floor pressure sensing devices product introduction," http://www.rsscan.co.uk, 2010.

[14] CASIA NLPR, "Casia foot pressure image database," http://www.cbsr.ia.ac.cn/english/, 2011.

[15] Y.L.Boureau, F.Bach, and Y.LeCun, "Learning mid-level features for recognition," in *CVPR*, 2010.

[16] M.A.Turk and A.P.Pentland, "Face recognition using eigenfaces," in *CVPR*, 1991.