Video Based Palmprint Recognition

Chhaya Methani and Anoop M. Namboodiri
International Institute of Information Technology, Hyderabad, India
chhaya@research.iiit.ac.in, anoop@iiit.ac.in

Abstract

The use of camera as a biometric sensor is desirable due to its ubiquity and low cost, especially for mobile devices. Palmprint is an effective modality in such cases due to its discrimination power, ease of presentation and the scale and size of texture for capture by commodity cameras. However, the unconstrained nature of pose and lighting introduces several challenges in the recognition process. Even minor changes in pose of the palm can induce significant changes in the visibility of the lines. We turn this property to our advantage by capturing a short video, where the natural palm motion induces minor pose variations, providing additional texture information. We propose a method to register multiple frames of the video without requiring correspondence, while being efficient. Experimental results on a set of different 100 *palms show that the use of multiple frames reduces the error rate from* 12.75% *to* 4.7%*. We also propose a method for detection of poor quality samples due to specularities and motion blur, which further reduces the EER to* 1.8%*.*

1. Introduction

The cameras in mobile devices such as cell phones and laptops can effectively double as a biometric sensor, providing security and ease of use for access to the devices as well as other services. The cameras in such devices are fixed and the user should be able to present the biometric modality in an unrestricted and intuitive manner. As a result, the captured images vary considerably due to variations in illumination, background, and pose as well as blur due to motion and incorrect focus. Methods for dealing with variations in illumination and pose have been studied extensively for modalities such as face [1] and gait [7]. Palmprint as a modality [9] has the advantages of ease of presentation and discrimination ability compared to face or gait as well as having a suitable size and scale of texture for capture with a mobile camera.

Figure 1. Palm line variations with change in view.

The problem of pose variations due to unconstrained palm capture was recently addressed in the context of palmprint [8] and hand geometry [10, 3] based authentication. However, the visibility of palm lines may be hindered due to specular reflections from the skin and motion blur, making the problem challenging. Moreover, even minor variation in the view direction causes significant changes in the visibility of palm lines (see Figure 1). The detection and characterization of palm texture is made further difficult by the poor quality of cameras in mobile devices as well as low levels of ambient lighting leading to higher levels of noise and lower contrast. In other words, correction of pose variations as addressed by previous works, solves only a part of the problem encountered in practice.

In this paper, we propose a novel approach for measuring and addressing the degradation of palm images caused by the aforementioned factors. The primary idea is to combine the information from multiple frames of a short video (say 0.5 seconds) to improve the information content in a sample. As most digital cameras are capable of capturing videos, this is quite practical. The natural motion of palm during the video capture provides sufficient variations in view, resulting in significant improvements in the information content. If the information from multiple frames can be integrated effectively, one can expect to see an improvement in performance of the authentication algorithm. However, traditional approaches like super-resolution for the in-

1051-4651/10 $26.00 © 2010 IEEE
DOI 10.1109/ICPR.2010.334

tegration of the images would be too slow to be of any practical use.

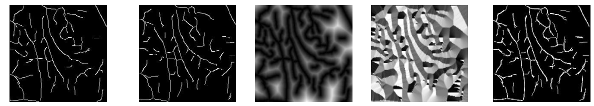
The primary contributions of this paper include: i) A method to register multiple palm images from a video without relying on correspondences, which are difficult to obtain, ii) A method to integrate the information from multiple frames in the feature space, and iii) A method to detect poor quality acquisitions due to specularity and motion blur so that they can be rejected without comparison. We demonstrate the effectiveness of our approach on a dataset of low quality palmprint videos consisting of a 100 users acquired using a webcam.

2. Combination of Multiple Frames

The most common way to reconstruct a single image from multiple images is by using super resolution. Farsiu *et al.* [4] provides a detailed study of super resolution techniques for generic images. Accurate registration is a pre requisite for super resolution, where the error tolerance is less than a pixel. But to achieve such high levels of accuracy in registration, the image should have rich textural information. Unfortunately, this is not possible in the case of palm images, especially when captured using low resolution cameras with uncontrolled lighting. Moreover, even if registration is accurate, the processing time taken to super resolve images is usually of the order of a few minutes[4] in the best case. This makes it difficult to directly use super resolution techniques for biometric authentication.

This led researchers to look for approximate methods that are efficient and practical. Arandjelović *et al.* proposed a method for implicit super resolution to achieve pose and illumination invariance in low quality videos [1] for face recognition. They achieved this by the offline learning of a hierarchical model, sub-sampled at multiple scales. In our case, we are mainly dealing with low-textured images having missing data. We must use an intelligent registration scheme that ignores the missing data when aligning the various frames. Since the discriminative information present in these images is in the form of lines, which also forms the basis for registration, it is both efficient and practical to combine the information from different frames in feature space. We propose a method to carry out the registration using lines detected from each frame, while combining the registered images in the feature space. We choose Gabor filter response as our feature set [9].

The authentication process proceeds as follows. During the enrollment phase, multiple samples (short videos) of the users palm are captured. The resulting frames from each video are combined to create a feature representation of the palm as in traditional single image approaches. The following steps are involved in the enrollment step. 1) *Frame Extraction from the video*: First of all, we need to find all the valid frames from the videos. A valid frame is one which has a clear unobstructed view of the palm. 2) *Palm Extraction*: The relevant part is extracted from all the frames. This is the first step of registration. This corrects for the in plane rotations, hand orientation is set to a pre fixed direction. 3) *Registration:* This step employs a specially tuned registration method that corrects for scale and pose variations in the absence of a rich texture. This results in the images being registered within 3-4 pixel range. 4) *Combination across multiple frames in the feature space:* In this step, we combine the information on a pixel by pixel basis. Registration and frame combination steps are described in the next Section.



(a) (b) (c) (d) (e)

Figure 2. Registration Process: a,b) Line maps of two palms, c) Distance Transform of a, d) Gradient Transform of c, and e) overlapped image.

The *matching* step is similar to traditional palmprint authentication, where the binary feature representations of two palms are compared using hamming distance [9].

3. Registration and Frame Combination

Registration is the process of overlaying two images taken at different times. Zitova and Flusser [11] note that this alignment usually follows a 4-step procedure: a) Feature detection, b) Feature matching, c) Transformation estimation, and d) Image re-sampling and transformation. Generally, robust landmark points are treated as features. In palmprint images, it is difficult to find robust landmark points due to the weak textural information. Cases of missing and erroneous correspondence computation occur easily; both because there is similarity in the texture and also due to the differences in the computed line map. For such a case, Gut *et al.* in [6] have suggested edge based registration using a Hausdorff distance function modified by using a voting scheme as fitting function. We utilize a similar approach, where the distance between closest matching line points is used as the criterion of registration.

As noted before, we need a registration method for palmprints that does not require accurate correspon-

Table 1. EER and FTA (in %) with combination of varying number of frames and different quality thresholds (τ_i). Results for score level fusion of frames is given by SF.

	Base Image (BI)		BI+2 frames		BI+6 frames		BI+8 frames		BI+10 frames	
	EER	FTA	EER	FTA	EER	FTA	EER	FTA	EER	FTA
SF	12.75	0	24.30	0	19.25	0	17.48	0	36.19	0
τ_0	12.75	0	13.99	0	4.70	0	5.15	0	5.79	0
τ_1	7.64	6.57	8.07	8.09	5.35	7.58	4.42	7.5	4.50	8.09
τ_2	14.36	9.44	5.82	13.15	4.64	10.9	4.46	11.12	3.62	11.8
τ_3	3.69	11.8	3.40	16.86	1.90	13.99	1.80	13.99	7.75	14.5

dences between landmark points. We note that we just need the sets of pixels constituting the lines in the two images to be matched. Assuming the underlying transformation to be affine, parameters are estimated iteratively. For each linemap, we compute loose matches using euclidean distance transform (DT) [2] at each point. The gradient vector of the DT at each point (see Figure 2) is used to determine the direction of motion of each point for alignment. This leads us to the line pixel closest to the pixel under consideration in corresponding frame. In this manner, we are able to handle missing points by assigning it another corresponding point present nearby. Since we use all the line pixels for determining the image transformation instead of a few landmark points, the collective group of points iterates to the most consistent match. At each step, points with correspondences beyond a threshold are labelled as outliers and are excluded in the following iteration. This method is more flexible as opposed to using strict point correspondences. This implicit adaptivity is extremely helpful in taking account of missing and erroneous data. The details of the line detection method and the method to find point matches has been described below:

Line Combination: Major lines are extracted using a Sobel filter [5]. The image is filtered in four directions viz. $0°$, $45°$, $90°$ & $135°$. The prominent lines are then thresholded out to form the prominent features after rejecting the unstable wrinkles. *Finding Correspondence:* As mentioned earlier, finding exact correspondence for a line image is difficult. We define the corresponding point of a line pixel to be the closest line pixel in the second image. *Model Estimation and Image Transformation:* In each iteration, a direct pseudo inverse based technique is used to find out the affine transformation parameters [11]. *Frame Combination:* Once the images have been registered, we combine them by taking an average of the responses, which ensures that the line patterns repeating the most number of times are given a high weightage and hence retaining them in the thresholding process. This image is then used for score computation using Hamming distance.

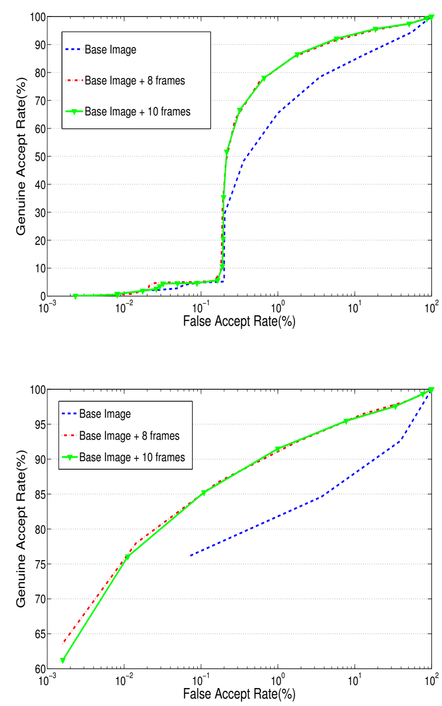


Figure 3. ROC curves on a)*logarithmic* **scales, and the result of removing poor captures.**

4. Experimental Results and Discussion

We collected a dataset containing videos of the hand taken by a web camera, in an unconstrained image capturing setup, as existing datasets provide single images of each palm. 50 subjects were asked to pose for a fixed camera in a manner intuitive to them. 6 videos each were recorded for both the left and the right hand for each subject. The matching between two palms, $palm_1$ and $palm_2$ is performed by computing the two hamming distances R_m and I_m between the real and imaginary Gabor responses. The final distance score is taken to be the maximum of these two dissimilarity scores. We performed experiments on a set of 100 different palms having an average of 6 videos each. The proposed algorithm takes only 1.4 seconds to combine 11 frames for a MATLAB implementation as compared to a few minutes taken for the fastest and most optimized super resolution techniques for these images. A total of $3,528$ genuine match scores and $1,75,065$ imposter match scores were recorded. The first row in Table 1 shows the result for fusion of scores obtained by matching frames directly without combination. The max rule was used to obtain these results. The second row of Table shows the improvement in the EER(Equal Error Rate) as we add more frames from the video to the Base Image. The corresponding ROC curve has been shown in Figure 3. We note that the EER first drops from 12.75% to 4.7% with the addition of 6 frames to the base image. However, the addition of further frames decreases the accuracy. This could be due to the noise present in the frames, as well as the blurring of lines arising from misalignments during the registration phase.

On observing the ROC curve in semilog axis(3(a)), we notice a slow rise in the GAR initially. A similar behaviour of the curve was observed in [8]. The drop indicates the presence of few imposter scores having better matching scores than genuine scores. This happens due to the texture appearance being partially washed out due to specularity. This is characteristic of an unconstrained imaging system. This effect can however be eliminated by automatically detecting and discarding these washed out samples at the time of query itself, and are modelled as Failure To Acquire(FTA). Our second experiment consists of studying the effect of variation of the FTA rate on the performance of the recognition system. We determine the washed out samples by measuring the average response of the image to Gabor Filter. We used three different thresholds to study the effect of removing bad samples from the dataset incrementally. The results for these three different parameters namely, τ_1, τ_2 and τ_3 have been provided in the Table 1. The FTA rate is different at various levels of frame addition on account of rejecting samples at each stage independently. Figure 3(b) shows the resulting ROC curve, where the EER drops to 1.8% on combining 9 frames. We note that the imposters with high matching score have been completely removed in the process.

5. Conclusion and Future Work

We address the problem of palmprint recognition from low-resolution videos of the palm captured in an unconstrained setup using low-end cameras. We proposed an efficient and robust method for feature level integration of information from multiple frames. We show that the EER of an authentication system reduces from 12.75% to 4.7% by integrating information from just 9 frames, on a dataset of 600 samples from 100 palms. We also propose a method to detect and remove low quality captures, where the texture information is washed out, which further enhances the EER to 1.8% albeit with rejection of some samples. The resulting system is both efficient and robust to be practical. One could further improve the performance of the system by improving the registration accuracy, enabling the combination of larger number of frames.

References

[1] O. Arandjelović and R. Cipolla. A manifold approach to face recognition from low quality video across illumination and pose using implicit super-resolution. *ICCV*, 2007.

[2] G. Borgefors. Distance transformations in digital images. *Comput. Vision Graph. Image Process.*, 34(3):344–371, 1986.

[3] J. Doublet, O. Lepetit, and M. Revenu. Contact less hand recognition using shape and texture features. 3, 2006.

[4] S. Farsiu, D. Robinson, M. Elad, and P. Milanfar. Advances and challenges in super-resolution. *International Journal of Imaging Systems and Technology*, 14:47–57, 2004.

[5] S. Garg, J. Sivaswamy, and S. Chandra. Unsupervised curvature-based retinal vessel segmentation. pages 344–347, 2007.

[6] P. Gut, L. Chmielewski, P. Kukolowicz, and A. Dabrowski. Edge-based robust image registration for incomplete and partly erroneous data. *CAIP*, pages 309–316, 2001.

[7] A. Kale and A. R. Chowdhury. Towards a view invariant gait recognition algorithm. *IEEE Conference on AVSS*, pages 143–150, 2003.

[8] C. Methani and A. M. Namboodiri. Pose invariant palmprint recognition. *IEEE/IAPR Proceedings of International Conference On Biometrics*, 3:577–587, 2009.

[9] D. Zhang, W.-K. Kong, J. You, and M. Wong. Online palmprint identification. *IEEE Trans. Pattern Anal. Mach. Intell.*, 25(9):1041–1050, 2003.

[10] G. Zheng, C.-J. Wang, and T. E. Boult. Application of projective invariants in hand geometry biometrics. *IEEE Transactions on Information Forensics and Security*, 2(4):758–768, 2007.

[11] B. Zitová and J. Flusser. Image registration methods: a survey. *IVC*, 21:977–1000, 2003.