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 ability to carry out biometric authentication using generic cameras can be extremely useful for a variety of applications, especially in mobile devices. Palmprint based authentication is appropriate in such cases due to its discrimination power, ease of presentation and scale and size of texture information that is easily captured by commodity cameras. However, the unconstrained nature of pose and lighting in such applications introduces several challenges in 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 one second video, where the natural motion of the palm 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 for practical use. Experimental results on a set of different 100 palms show that the use of multiple frames reduces the error rate from 12.75 *to* 5.79. *We also propose a method for detection of poor quality samples due to specularities and motion blur.*

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 presents the biometric modality in an unrestricted and intuitive manner. 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 [6]. The use of palmprint as a modality 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.

Our recent work [7] in this direction addressed the problem of pose variations with unconstrained palm capture. 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, correcting for pose variations addresses only a part of the problems that is 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 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.

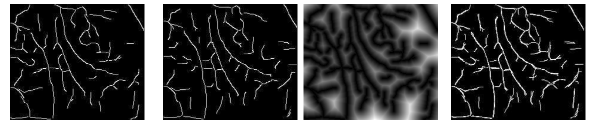
2. Combination of Multiple Frames

The most common way to reconstruct a single image from multiple images is by using super resolution. A detailed study on super resolution of images has been provided by Farsiu *et al.* in [3]. Registration is a pre requisite for super resolution. The error tolerance in registration step for super resolution should not exceed 1 or 2 pixels. But to achieve such high levels of accuracy in registration, the image should have rich textural information. But this is not possible in the case of palm images. Since the only texture present is in the form of weak lines. Also, even if registration is dealt with, the processing time taken to super resolve images is usually of the order of a few minutes[3].

This makes it difficult to directly use super resolution for a recognition based application like biometrics. This led Arandjelović *et al.* to propose implicit super resolution to achieve pose and illumination invariance in low quality videos [1] for Automatic Face Recognition(AFR). They achieved this by the offline learning of a hierarchy of gSIM models, 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, it is more intuitive to reconstruct the image in feature space; where the feature chosen is the line information present in the image. Hence, we propose a novel method to combine the image information in the feature domain. We choose Gabor filter response as our working model. We present a step by step description below:

DATA ENROLLMENT: A high quality video is taken at the time of user enrollment.

MATCHING QUERY VIDEO: The following steps are involved for automatic matching: 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)



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

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.

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 which 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 have been described in the next Section. 5) *Matching:* Since we work with Gabor responses, they can directly be used to generate matching scores as described in Section 4.

3. Registration and Frame Combination

Registration is the process of overlaying two images taken at different times. In [8], Barbará Zitova *et al.* point out that this alignment is achieved by following a 4-step procedure: a) Feature detection, b) Feature matching, c) Transform model 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 [5] have suggested edge based registration using a Hausdorff distance function modified by using a voting scheme as fitting function. Such a method could be used for palm line based registration with some modifications.

In this work, we propose a registration method for palmprints that does not need to find strictly corresponding landmark points. We just need to know 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.

Table 1. EER and FTA (in %) with varying number of frames and different quality thresholds.

	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
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

For each linemap, we compute loose matches using euclidean distance transform ([2]) and the gradient vector of the euclidean distance transform at each point(See Figure 2). This leads us to the line pixel closest to the pixel under consideration in the temporal domain. In this manner, we are able to handle missing points by assigning it another corresponding point present nearby. And 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, as more and more wrongly assigned correspondences get labelled as outliers. Outliers are excluded in the process by using an experimentally determined threshold. 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 [4].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 not possible. We define the corresponding point of a particular pixel to be the pixel closest to it in the frame nearest to it in the temporal domain. *Model Estimation and Image Transformation:* After the loose correspondences are found, a simple pseudo inverse based technique is applied to find out the affine transformation parameters. *Frame Combination:* Once the images have been registered, we combine them by taking an averaging function which ensures that the line patterns repeating the most number of times are given a high weightage and hence retaining them in the threshlding process. This image is then used for score computation using Hamming distance.

4. Experimental Results and Discussion

The Dataset contains videos of the hand taken by a commonly available web camera, in an unconstrained image capturing setup. This is the first video database containing palmprint videos. 100 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 amd I_m obtained by matching the real and imaginery Gabor responses respectively. 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 secs to combine 11 frames into a single image as compared to a few mins taken for super resolution techniques. A total of $3,528$ genuine match scores and $1,75,065$ imposter match scores were recorded. The first row of Table 1 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.

However, on observing the curve in semilog axis(3(b)), we notice a slow rise in the GAR initially. A similar behaviour of the curve was observed in [7]. 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. This is termed as the Failure To Acquire(FTA) rate. 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 independentally. The natural reason for this being that information is being added at each stage, which might make some of the washed out templates accumulate more information. The improved ROC curve is shown in Figure 3(c).

5. Conclusion and Future Work

In this paper, we proposed a novel approach for palmprint recognition in low-quality palmprint videos captured in an unconstrained setup using low-end camera devices. The dataset consists of both right and left hand videos of 100 palms.We address the challenges of using palmprint recognition in a practical scenario, having illumination and pose variations. We used a palm video instead of a single image. We combine meaningful data with considerable reliability. One could improve the accuracy by improving the registration technique for palmprints. We also present a technique to determine wahsed out images at the query stage and systematically study its effect on EER.

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] 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.

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

[5] 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.

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

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

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

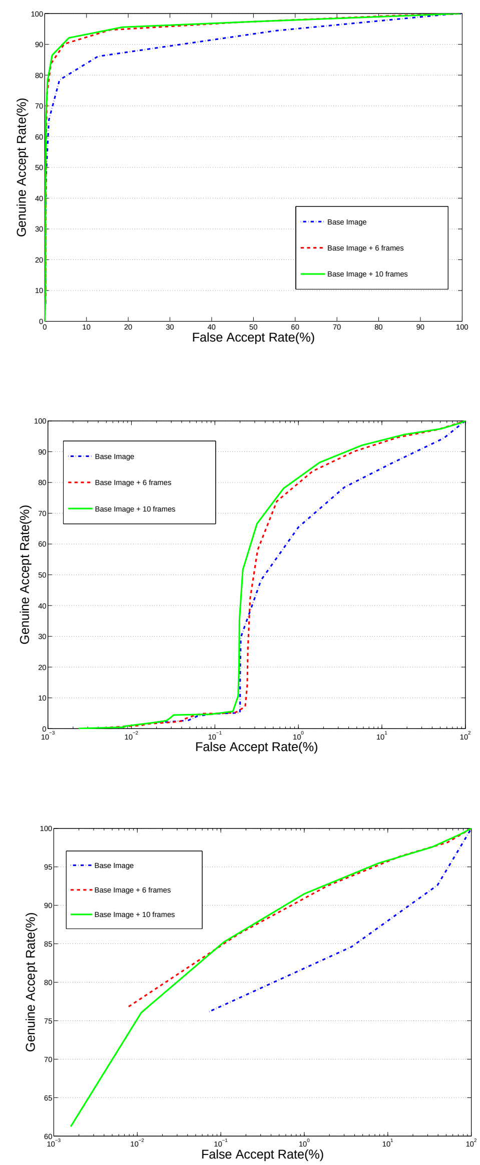


Figure 3. ROC curves on a) ***linear*** **and b)*****logarithmic*** **scales, and the result of removing poor captures.**