Issues in rotational (non-)invariance and image preprocessing

Lalit Jain[1], Michael J. Wilber[1,2], Terrance E. Boult[1,2]
[1]VAST Lab, University of Colorado Colorado Springs [2]Securics Inc
{ljain2 | mwilber | tboult} @vast.uccs.edu *

Abstract

This paper addresses two problems that have been largely overlooked in the literature. First, many systems seek to use, and algorithms claim to provide, rotational in-variance, such as fingerprint minutiae or SIFT/SURF features. We introduce a statistical test for rotational independence, using lossless rotations to show the differences are statistically significant and cannot be attributed to image noise. We use this to experimentally show fingerprint feature extractors fail to be rotation independent. We show the popular "rotation invariant" SURF and SIFT feature extractors, used in both biometric and general vision, also fail the rotation independence test.

We then introduce a match-twist-match (MTM) paradigm and experimentally demonstrate that, by reducing the effective angular difference between probe and gallery, we can improve system matching performance. Our analysis, using FVC2002 and FVC2004 datasets, further shows that differences in extracted features impact the overall system performance of fingerprint matching of both matchers tested. Using the MTM approach, we reduce our secure template system's errors by 10%-20% – helping us to define the current state of the art in the FVC-OnGoing Secure template competition with an EER of 1.698%.

We end by bringing to the forefront the growing danger of sensors over-preprocessing of images. We show examples of the problems that can arise with preprocessing. As our rotation experiments showed, the impact of even modest numbers of feature errors suggest these preprocessing issues are likely very significant. We suggest the need for policy guidelines that require disclosure of preprocessing steps used and the development of standards for testing the impact of preprocessing.

1 Introduction

Throughout computer vision, but particularly in biometrics, people often use image preprocessing/enhancement and rotationally invariant features to address the fact that, in operational settings, data may not always be ideal and/or aligned. This paper examines issues in rotation invariance and preprocessing – challenging some long held beliefs.

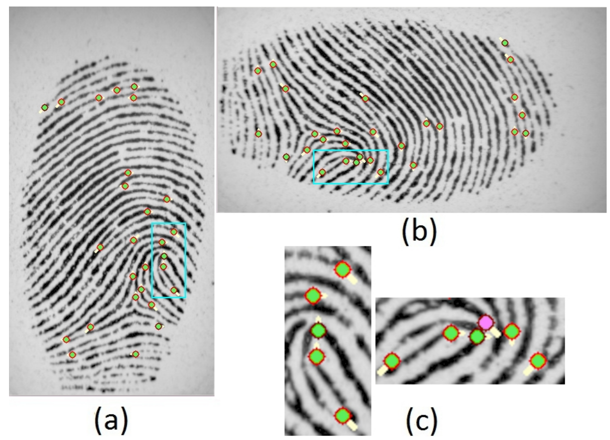

Figure 1. *Original (0°) fingerprint Image(a) with minutiae in green, minutiae orientation in yellow, and a region of interest outlined in cyan. (b) shows the image losslessly rotated 90 degree, then minutiae computed. In (c) we see 250% enlargement of region of interest in 0° and 00°, where it is clear that the 90° a real minutiae, colored magenta, was detected, but was missed in the original. The NIST MINDTCT, with default parameters, was used for extracting minutiae from images at different angles. there are* no *differences in pixels, but there is a difference in the number of minutiae – clearly the detector is not rotation invariant. This paper will show this rotational dependence is common (even for SIFT and SURF), statistically significant, and impacts matching results. We then present an algorithm to mitigate the rotational (non)-invariance.*

Within an automated biometric system, fingerprints are the most widely used modality with applications in criminal investigations, border control, time & attendance and access control systems. After decades of research in this mature subfield, the image processing and feature extraction is generally taken for granted and largely ignored. It is generally claimed (and often strongly believed) that ***fingerprint matching systems are rotation-invariant*** [5, 20], meaning that the fingerprints should match equally well at any angle.

In this paper, we demonstrate that some fingerprint matching systems, including those quoted above, are not com-

*This work was supported by ONR SBIR Award N00014-11-C-0243, ONR MURI Award N00014-08-1-0638 and NIST 60NANB12D0006.

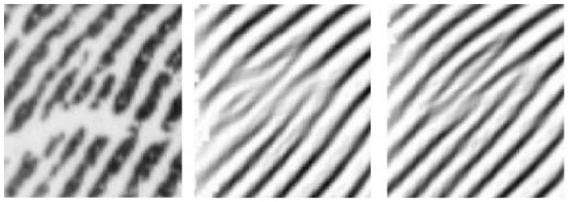

Figure 2. *A second issue explored herein is that of significant image preprocessing. Here we show three images of the same finger. The left image is a cropped region from a fingerprint sensor image with minimal spatial processing. The middle and right images show a similar region from the same finger, acquired with a sensor using aggressive preprocessing. The horizontal cut in the ridges of the left image reflects the real finger condition. The second sensor's preprocessing artifacts act to cover up the cut and produce different artifacts on different measurements. Overall the preprocessed sensor has inflated its NIST Fingerprint Image Quality (NFIQ) score, but preprocessing also induces false minutiae artifacts. The artifacts, which do not exist on the real finger, can present problems for subsequent processing by algorithms or humans, especially because of their inflated quality and unique shape characteristics.*

pletely rotation invariant. The cause of this surprising result is that, while the matching algorithms may be invariant, the underlying feature extraction is not. We show non-invariance also applies to the widely used "rotation invariant features," such as SIFT and SURF.

Most of these feature extractors are moderately invariant, with only small differences in detected features under rotation. Because image noise results in feature variations as well, the approximate invariance likely leads to the actual non-invariance being overlooked. However, when systems need to approach 95+% accuracy, even small differences can matter. We tested for invariance using lossless transformed (flipped) images, and statically rejected, at very high significance levels, the hypothesis of rotational independence. We tested NIST, Digital Persona, and Neurotec fingerprint minutiae extractors as well as SIFT and SURF, and, for each, rejected the hypothesis of rotational invariance. For fingerprint matching, we show this non-independence has an impact on matching performance on public algorithms. We then show how estimation of inter-image rotation angles, followed by de-rotation (twist) and re-matching, improves fingerprint matching performance. We evaluate on both open source code, for easy reproduction of our experiments, and on state-of-the-art systems. In particular, we show the match-twist-match approach helps in achieving the state of the art in the FVC-onGoing "secure template" competition.

The paper also addresses issues in image preprocessing before feature extraction. Artifacts introduced by preprocessing in fingerprint images are rarely discussed in literature. To highlight/explain the types of issues raised by preprocessing, Figure 2 shows different image chips from different sensors/preprocessing. The first image has minimal processing, but yields no high-quality (NFIQ=1 or 2) minutiae in that region, largely because the cut on the finger limits confidence for the minutiae actually present. A human fingerprint examiner, like the algorithm, can identify minutiae in the region but would also discount the cut region slightly. The other two images highlight the type of highly preprocessed artifacts produced by some sensors – with multiple NFIQ=1 minutiae in each window. However, only one of those "high quality" minutiae matches is shared between them; the other minutiae are just "high-quality" phantasms from the sensor's distortion of reality. If the middle image was the enrollment and the right was used as a basis of comparison, then the high-quality false minutiae could cause problems. Those features might even be considered points of exclusion by a human expert, allowing the person to repute an asserted match, causing legal problems during forensic/criminal investigation or other situations when human matches would be necessary.

We have analyzed false artifacts from multiple sensors, but not in statistically significant volumes. Our intention is not to "point fingers" at any specific sensor vendors, but to highlight potential problems occurring due to excess preprocessing and to call for more work and standards in this area. Basic pixel-level normalizations are to be expected and offer little risk. However, recent systems have been adding spatial image processing, often not described publicly. Such processing may make the image "look better" and enhance feature extraction, e.g. allowing higher NFIQ scores to be touted in company marketing claims. This is an area often taken for granted, and such claims might go unchallenged. However, we suggest such preprocessing can introduce new classes of errors which can cause problems for later stages of processing, or even humans interpretation.

We have four main contributions in this paper:

1. We introduce a testing methodology with necessary conditions for any truly rotationally invariant system, which may be used to test feature extractor designs.
2. We assess fingerprint minutiae feature extractors as well as popular computer vision feature extractors, such as SURF [1] and SIFT [13], and show they fail to be fully rotation invariant. This is novel, as many people incorrectly believe, and many papers claim, these features/systems are rotationally invariant.
3. We introduce the Match-Twist-Match paradigm for mitigating minor rotational dependence and show it improves matching performance in real systems.
4. We introduce and discuss issues of spatial image preprocessing in fingerprint systems, highlighting often overlooked issues and suggesting the need for more research, policy guidelines, and new standards.

2 A Rotational Invariance Statistical Test

There are many feature detectors, from minutiae to general vision features, that claim to be rotationally invariant. Given a feature extractor that claims to be rotationally invariant, how do we test that claim? This section gives a statistically grounded method for testing such claims.

Formally, one might find a single counter example, as we shown in Fig. 1, that contradicts the invariance property. However that does not mean it has any "significant" impact, especially on the overall performance. Thus our proposed "necessary approach" to test such a claim is to define a hypothesis of the impact on match scores, then statistically test that hypothesis.

Definition 1 RIST: Rotational Invariance Statistical Test *Let $S(P, G)$, be a match score, e.g. Equal Error Rate, of a system given probe set P and gallery set G. Letting P_X be the probe set with each image rotated by X, we can compare the results when matching a rotated probe set P_X with the gallery G, i.e.,. compare $S(P_X, G)$ for different X. To avoid confounding rotational dependence with image/sensor noise, RIST uses probe sets $P_0, P_{90}, P_{180}, P_{270}$, with the original probe images losslessly rotated images by 90, 180 and 270 degrees respectively. Thus the pixel data in each probe set is identical, only the rotation is different.*

The RIST null hypotheses H_o are that $S(P_0, G) == S(P_X, G)$ for $X = 90, 180, 270$. Given that the experiments use the same probe and gallery pairs, RIST uses a two-sided paired t-test to assess the statistical significance of the result. Note that this hypothesis is necessary, but not sufficient for true rotation invariance.

To analyze the rotation invariance for fingerprint matching, we run experiments using NIST and Digital Persona fingerprint minutiae exaction and the NIST Bozorth matching systems on the complete FVC2002 [14] and FVC2004 [4] database. We also did experiments using the Neurotec fingerprint system and found that even the Neurotec matching system, in addition to NIST and Digital Persona systems, is not rotation invariant. However, since we only had access to the demo version, with limited operations, we did not do full statistical testing on Neurotec.

The FVC2002 and FVC2004 databases contain four datasets: DB1A, DB2A, DB3A, DB4A. Each dataset has 800 images, I_{ij} in datasets F_m $(1 \leq m \leq 4)$ consists of one hundred subjects $(1 \leq i \leq 100)$ and eight samples per subject $(1 \leq j \leq 8)$. The standard FVC2000 protocol [15] is used to analyze the error from matching and non-matching distribution. For matching distribution, each probe image I_{ij} is matched against gallery images J_{ikl} $((j < k \leq 8)$ and $(1 \leq l \leq 4))$, where we extend it to include at rotations at 0, 90, 180 and 270 degrees, as $l = 1..4$ respectively. The total number of genuine recognition attempts are 11200 (((8*7*4)/2)*100). For non-matching distribution, each probe image I_{i1} $(1 \leq i \leq 100)$ is matched against gallery images consisting of first fingerprint image from non-match fingers at 0, 90, 180 and 270 degree J_{k1l} $((i < k \leq 100)$ and

Rotation	90°	180°	270°
02 DB1	9.0E-79	1.9E-80	2.9E-59
02 DB2	1.8E-105	8.9E-76	4.2E-106
02 DB3	4.8E-10	1.3E-16	5.2E-26
02 DB4	2.6E-45	1.1E-21	5.3E-23
04 DB1	1.11E-31	2.15E-24	6.58E-39
04 DB2	2.29E-62	1.17E-41	3.20E-49
04 DB3	5.31E-13	1.92E-65	1.42E-21
04 DB4	2.00E-13	6.42E-15	1.58E-08

Table 1. *P-values for two sided paired T-test for rejecting H_o, the hypothesis that the system is rotation invariant is tested on various FVC databases using MINDTCT minutiae. All results are very significant, so we reject rotational independence and conclude that the rotation dependence of the MINDTCT minutiae extractor's impact on Bozorth matching is statistically significant.*

Rotation	90°	180°	270°
02 DB1	1.1E-31	2.2E-24	6.6E-39
02 DB2	2.3E-62	1.2E-41	3.2E-49
02 DB3	5.3E-13	1.9E-65	1.4E-21
02 DB4	2.0E-13	6.4E-15	1.6E-08
04 DB1	5.2E-43	4.3E-01	1.2E-43
04 DB2	4.4E-05	4.8E-02	9.1E-03
04 DB3	1.1E-182	2.0E-10	2.0E-178
04 DB4	3.8E-07	9.2E-02	6.4E-03

Table 2. *P-values for rejecting hypothesis that rotation does not impact scores using Digital Person Fingerjetfx on various FVC databases. Most results are very significant ($p << .01$), but a few of the 180 degree are only significant ($.01 < p < .05$). Again, we conclude that the rotation dependence of the FIngerjetFX minutiae extractors is statistically significant.*

$(1 \leq l \leq 4))$ and corresponding impostor matching scores are recorded. The total number of impostor recognition attempts are 19800 (((100*99)/2*4)). Our score $S(P_X, G)$ is the equal error rate obtained by the NIST NBIS Bozorth matcher on the features of the images. Testing with other features, such as False NonMatch Rate (FNMR), also called False reject rate) at a fixed False Match Rate (FMR) have similar results. Table 1 and Table 2 show the p-values from a 2-tailed paired T-test with the 7750 scores before and after rotation, using NIST and Digital Persona Fingerjetfx respectively. For every rotation, the null hypothesis H_o is rejected as very significant, so we conclude the rotations impact scores in a way that is statistically significant.

We also performed tests on the Bozorth matcher using minutiae files in which the minutiae coordinates were algebraically rotated and found no significance in matching scores. Because the matching algorithm appears to be rotationally invariant, we can assign the cause for non-invariance to the feature extractors themselves. This is consistent with our finding, as we know the number of features extracted varies under even lossless rotations.

It is illustrative to consider the "rotation invariance" of

Rotation	90°	180°	270°
Bedroom	3.74E-08	2.2E-04	0.608
CALsuburb	0.029	0.008	0.113
Industrial	2.38E-17	3.99E-14	0.645
Kitchen	0.0617	0.0072	0.1905
Living room	4.80E-12	2.15E-08	0.272
Overall	7.73E-30	4.22E-25	3.42E-02

Table 3. *P-values for rejecting the hypothesis H_o, that rotation does not impact scores on various Scene Categories from a database with five objects: namely bedroom, CALsuburb, industrial, kitchen and living room having 200 different images each. Most individual and the overall results are significant ($p < .05$) and most are very significant ($p < .01$), so one can conclude that the rotation dependence of the SURF feature extractor are very statistically significant.*

Rotation	90°	180°	270°
Bedroom	0.456	0.091	0.885
CALsuburb	0.011	7.0E-4	0.116
Industrial	0.084	0.134	0.147
Kitchen	0.705	0.002	0.276
Living room	0.391	0.060	0.465
Overall	.207	2.15E-07	.032

Table 4. *P-values for rejecting a hypothesis that rotation does not impact scores on various Scene Categories from a database with five objects: bedroom, CALsuburb, industrial, kitchen and living room having 200 different images each. For sift only about 1/3 of the individual results are significant, and overall only 180 and 270 show significance. Thus one can conclude that the rotation dependence of the SIFT feature extractor are statistically significant at some rotations. The SIFT significance was clearly lower than SURF, so also applied a paired t-test on the difference between SIFT and SURF, and strongly reject the hypothesis that SURF is as rotational invariant as SIFT.*

other computer vision feature extractors, such as the well known SURF [1] and SIFT [13]. While SIFT/SURF are used in face recognition [2, 6], fingerprints [17], and other biometrics , we evaluated on scene classification [12] datasets, just to show the issue is very general. We used five objects: bedroom, CALsuburb, industrial, kitchen and living room, having 200 different images per category, i.e. a total of 1000 different images in our experiments. For feature extraction, we use the OpenCV implementation of both extractors. As for fingerprints, each probe image is rotated by 0, 90, 180 and 270 degree using lossless transformation, and separate features are produced. For simplicity, we let $S(P_x, G) = |\#F(P_x) - \#F(G)\|$, i.e. we simply compare the number of features extracted in the probe and gallery. If the feature extractor was invariant, that number would not change under rotation. We consider the same null hypothesis H_o, and again we apply a 2-tailed paired t-test in testing. Tables 3 and 4 show the the p-values from a 2-tailed paired T-test with SURF and SIFT feature extractor respectively. The tests use 200 images per class and 1000 images overall. For each lossless rotation angle, the null hypothesis H_o is rejected at very significant to significant levels, so we conclude the detectors are not rotation invariant, and it impacts the scores in a way that is statistically significant. We also considered the relative difference in scores for SIFT and SURF. The magnitude of errors was smaller for SIFT and the pair-wise t test reject the null hypothesis that SURF has equal or better invariance as SIFT with p-values of 3.00E-14, 3.03E-07 and 9.85E-06 for $90, 180$, and 270 respectively. Thus we conclude **SIFT is statistically significantly better, in terms of these rotational invariance tests, than SURF.**

As we mentioned, the RIST approach with hypothesis H_o is necessary but not sufficient for rotational invariance. True rotational invariance requires invariance at all angles; however, testing that requires modeling sampling effects. The proposed RIST method for evaluation is easy to implement/use and could/should be used by designers of rotationally invariant feature extractors to evaluate the algorithms. Future work may explore the root cause of the non-invariance, e.g. is it inherent or just an implementation flaw. For now, we address what we can do about it at a systems level.

3 A Twist To Improve Matching

Given that we have shown that minutiae extraction differences lead to matching systems that are not rotation invariant, it is natural to explore if it can be corrected. We designed an approach, Match-Twist-Match (MTM), to reduce the difference in rotation between the probe and gallery images. We test this approach using two matching algorithms.

Even with the variations in extracted minutiae, matching prints will likely have a partial match, so we propose the MTM process of coarse matching, rotation estimation, twist (derotation of one image by the negative of the estimated rotation), reprocessing to extract minutiae, and then rematching. This concept, pre-processing with a twist, can be applied to any matching system.

For our fingerprint matching, we modified the NIST BOZORTH3 minutiae matcher to output an estimated rotation angle between the original probe and gallery images. The Securics/UCCS Bipartite Biotope algorithm, described below, already provides estimated rotation. It may be worth noting that, while there are many papers on ridge orientation field estimation, that is not the problem being examined here. While other papers may also do overall fingerprint orientation estimation, any algorithm for estimating orientation between the probe and gallery would do for this application; we don't consider our Bozorth modifications significant. The contribution here is the MTM, the idea of using the estimated inter-print orientation to twist (derotate) the pair and match again as a way of addressing feature extractors that are only approximately rotationally independent.

One might think this would approximately double the

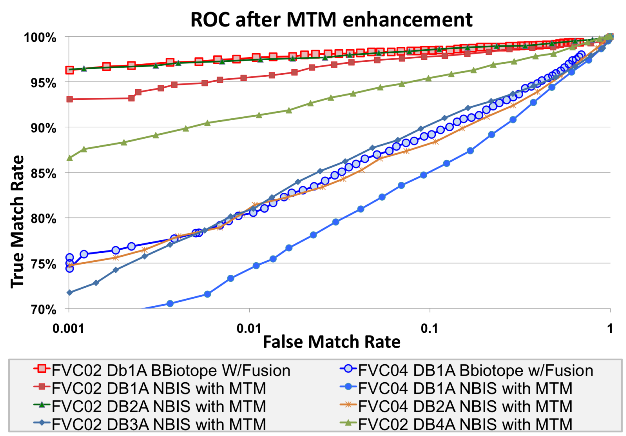

Figure 3. *ROC curve show for both NBIS enhanced with MTM and the Bipartite Biotope (BBiotope) fusion of two components enhanced with MTM. The BBiotope on DB1A of both FVC2002 and FVC2004, shown as open symbols, does significantly better than NBIS (filled symbols of same shape). The BBiotope with MTM and fusion also does better than the original Biotope algorithm, see (Fig.3 in [3]), especially at lower FMR.*

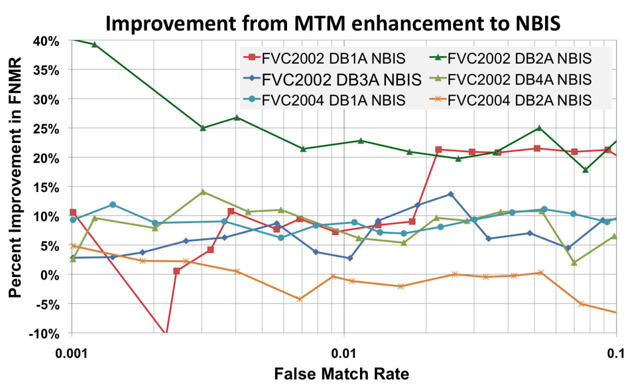

Figure 4. *Improvement plots for MTM enhanced NBIS on FVC2002 and FVC2004 datasets. Percent improvement in False Non-Match Rate (FNMR) at a fixed False Match Rate over the practical region of FMR. Above 10% FMR has little security, and below .01%, there is insufficient data for meaningful analysis. From the plots, it clear that MTM generally provides moderate improvement, with the exception of FVC04 DB2.*

matching cost. However, the derotate-rematch step is needed only for probes with scores near the decision boundary – if the coarse level match score is sufficiently high (clear match), or extremely low (not even close to a match), then derotation and rematching will not likely change the outcome and can be skipped.

Once again, the standard FVC protocol [15] is used to analyze the error from matching and non-matching images in FVC2002 and FVC2004 dataset. We analyzed two algorithms. The first was the NBIS Bozorth Matcher with MINDTCT extractor already discussed, both native and with MTM enhancements. Our second algorithm was to apply these enhancements to the Securics/UCCS Bipartite Biotope

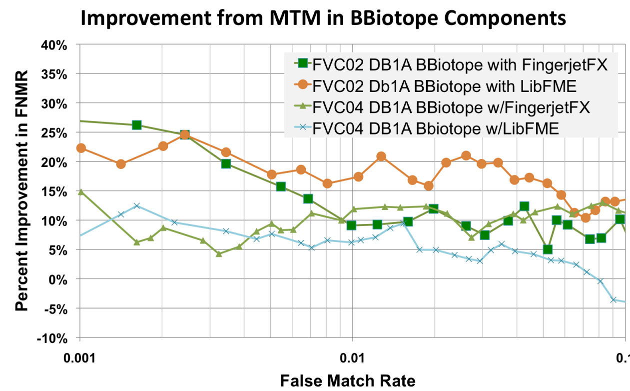

Figure 5. *Percentage improvements using MTM approach on the components used in the Bipartite Biotope secure-template matching. The core BBiotope approach was run separately with different minutiae feature extractors (libFME and FingerJetFX), each of which estimated matched, twisted and matched again. There is generally a significant improvement for each approach, except at high FMR for libFME on FVC04 DB1a.*

(BBiotope) algorithm[3] as extended in [18] and to study choices of minutiae extractors and the effect of rotation. The BBiotope algorithm is a leading secure template matching approach to which we have access, including its extractor libFME which was derived from MINDTCT but with numerous modifications. BBiotope, like NBIS, is also supposed to be rotationally independent – the matcher is but the extractors are not.

The Figure 3 shows the ROC curve for fingerprint matching of probe and gallery image using MTM enhanced NBIS and the Bipartite Biotope approach on FVC datasets. It can be noted that fingerprint matching is enhanced when probe and gallery images are aligned before matching, especially at the generally more important lower FMR rates. The Figure 4 gives improvement results of MTM enhanced NBIS method on FVC datasets. The improvements were consistent on FVC202 datasets. We found similar but slightly smaller improvements on DB1A and DB3A of FVC2004 datasets, and, for DB2A and DB4A, the MTM performs the same as non-rotated images. In the collection of FVC2004 [4], the subjects were instructed to distort their prints, pressing hard and/or rotating while collecting with the goal to increase difficulty in recognition. These distortions increase the chances of error in estimating the angle between the probe and gallery images in the first stage of the twist algorithm and also increase the overall match accuracy. Either of these could have impaired or masked improvements of the the Match-Twist-Match approach. We do note that while the “twist” did not help, it did not degrade performance either.

The amount of improvement for the components used in the BBiotope are shown in figure 5. We see that for each feature extractor, the MTM approach improves performance. Using this approach, we reduce the errors of the components for our secure template system’s errors by 10%-20%. The final system use for the FVC-OnGoing competition results, does a max score-level fusion of the two MTM enhanced matchers, improving results further. The improvement was important to use because various changes since 2009, including dealing with the Bipartite storage itself as well as a potential attack on the template, had slightly reduced accuracy at a fixed template size. By using MTM and fusing two different features, it helped us to define the current state of the art in the FVC-OnGoing Secure template competition, with an EER of 1.698%.

4 Spatial Image Processing of Fingerprints

In fingerprint verification/authentication, the ridge structure and/or minutiae are the most widely used features [10]. The extraction of these features is impacted by the quality of a fingerprint image. In operational environments, fingerprint quality may be reduced because of poor contact with the fingerprint sensor, physical distortions on the print (e.g. cuts/abrasion), noise at the sensor (e.g. surface or sensor noise), and many other reasons. To overcome these limitations, the fingerprint sensor manufacturer includes preprocessing on the raw sensor data, generally implying that the preprocessing improves the recognition performance. Early sensors were direct imaging, or images of Frustrated Total Internal Reflection (FTIR) images. The raw sensor data in such systems is a direct image, and processing was largely per-pixel brightness/contrast adjustments, and overall geometric distortion correction.

A significant amount of literature is available for fingerprint enhancement techniques that do image processing on "raw fingerprint images" [11, 21, 9, 7, 19] to improve their use in fingerprint recognition. Since the raw image was already formed, such processing is a type of post-processing. When carefully examining the raw image and processed image in [11, 21, 9, 7, 19], it is clear that occasional artifacts can be observed on the processed image, especially in low quality regions. It is important to assess the impact of image quality, including artifacts, especially in applications where enrollment and verification may use different sensors. Quality testing and inter-operability testing has long been a part of fingerprint research.

Sensor properties plus preprocessing done by fingerprint sensors can create artifacts. The author in [8] used six different sensors for capturing fingerprint images and used NIST and Neurotec fingerprint matching systems. EER estimated using NIST and Neurotec algorithms shown in [8] were statistically different for all six of the different sensors. The reason why errors exist when the fingerprint image is captured from different sensors is not explored.

For multiple reasons, organizations developed measures of the quality of fingerprint image, such as the NIST NFIQ measure. Such image quality measure scores were originally intended to be predictive of relative performance of a minutiae-based fingerprint matching system and served a very useful purpose. However, they are now being used to get certificates such as PIV and for general marketing as a measure of sensor "quality." Such uses are at odds with the goal of maintaining fidelity to the original image and the goal of predictive power of the quality measure – one could trivially produce an image with perfect NFIQ by returning the same "perfect" artificial image every time.

Figure 6. *Images captured from the same finger using a sensor with advanced spatial image processing. The image output by the sensor looks great, but has hidden dangers. The minutiae in red colored region represent different minutiae information for the same finger.*

With the introduction of capacitive and other solid-state sensors, and more recently multi-camera and multi-spectral sensors, combined with the increasing computational ability that can be embedded in a sensor, the level of advanced image pre-processing before the fingerprint algorithms even see the images has increased dramatically. In particular, the systems are now doing extensive spatial image processing, and, in this section, we explore some of the issues on spatial image processing on fingerprint images which we believe need to be addressed at the level of policy and standards.

Figure 6 shows the full prints of the same finger shown in Fig.2 from a sensor that includes advanced spatial image processing. Let us briefly examine the differences between these images. The local loop + ridge ending in the left of the circular region is a double bifurcation with a smaller local loop + ridge ending in the circle on the right. In the rectangle on the left, there is another double bifurcation structure while, on the right, there is no minutiae at all. In both cases, the image processing has made good contrast and smooth ridging with good connectivity and flow, i.e. good quality minutiae and ridges. In the pentagon on the left, there is yet another double bifurcation (lower contrast this time, in the upper part of the pentagon), and on the right, there is a bifurcation - ridge - ending in a bifurcation triple. In the diamond, there are again significant structural differences. The structural changes show that the spatial preprocessing performed by the sensor creates spurious minutiae or removes genuine minutiae. Not only are there differences, but a number of these minutiae have very uncommon structures, structures our algorithm decided were ideal for matching, especially if trying to minimize the template size. This caused us significant problems before we understood the cause of reduced performance on this particular sensor.

On the positive side, some level of spatial image processing is almost a necessity for multi-image integration and can be useful for managing low quality regions. The ability to fill in gaps and deal with low quality can be very useful for low-quality prints, but is fraught with difficulty. The spatial processing are not, however, limited to issues with cuts or low-quality regions. In many of our experiments, we had issues with clean fingers that produce excellent images in regular sensors. For example, Figure 7 shows three captures from a sensor with significant image processing as well as a standard sensor. The circle in Figure 7 shows the finger region in the standard sensor in the the lower left sub-image. That image has a downward bifurcation in the lower left within the circle and a ridge ending in the right side of the circle. In the circles in the preprocessed images, we have 2 bifurcations and a bridge structure (upper left image), a correct image (upper right) and an almost indiscernible image structure (lower right). In the square, we see the normal print has a single upward bifurcation. The square region for the upper left advanced processed image has one upward bifurcation, an added downward bifurcation and another inserted bridge ridge. In the upper right processed image, there is an upward bifurcation as well as an added downward bifurcation. The lower right processed image has the ridge ending and is missing the upward bifurcation. This finger produce far more false-rejects than many other fingers.

It is quite clear that the extensive preprocessing within the sensor can create artifacts and changes the genuine biometrics features recorded in the database. Likely for competitive reasons, the images are made to look nearly ideal, and hide the flaws. When the spurious biometrics features are added and genuine biometrics features are removed due to extensive preprocessing, all while providing excellent apparent image quality, it's a problem. *We propose that, at least for evaluation/enrollment purposes, all sensors should be required to provide a minimally processed image(s), with only per-pixel operations and global geometric corrections. A weaker but useful operational alternative would be for the sensors to provide an* adjustment map*, an image with per-pixel or per-region value that dictate how much processing changed the pixels or the region.* Of course, the sensors should be allowed to provide their enhanced images by default, but without either the raw data or an adjustment map, there is too much risk of artifacts that cannot be detected in those enhanced images. The advantage of the raw image, which is why we propose that approach, is that human examiners can assess differences and researchers can better assess the impact and develop new measures and standards in this area.

Figure 7. *Clean print on a normal fingerprint sensor (lower left) images, and 3 examples from an image with significant spatial image processing. Compare the structures in the circled and squared regions. We really don't know what causes the very distorted images like the lower right, where even the overall flow is wrong.*

The PIV certification performed by NIST just specifies that the artifacts, anomalies, false detail or cosmetic image restoration effects detected on fingerprint images due to a device or image processing shall not significantly adversely impact the supporting intended application [16]. There can be no automatic procedure that checks and gives the score for the amount of preprocessing done unless the raw images are also available. Given the raw images, there could be several tests which might quantify the level of preprocessing in a fingerprint image. For example, one might propose to use any of the many tests used to test the quality of image compression algorithms. However, since we do not have access to the raw images from these sensors, any proposed approach would be purely speculative at this time. Therefore, the only approach we can propose at this time is the need for more research, regulation, and standards in this area.

5 Conclusion

In this paper, we have highlighted two neglected but important problems in fingerprint matching – the claimed rotational invariance and the impact of image processing on fingerprint images, and we have presented a novel solution. Our solution, a Match-Twist-Match paradigm, improves performance when features are approximately, but not fully, rotationally invariant. The rotational non-invariance issue, and the MTM paradigm, transcend fingerprint matching, with broader applications in biometric and vision systems.

We have challenged the standard claims and long-held belief of many " rotationally invariant" features, showing experimentally that, while they are roughly invariant, the rotational dependencies are statistically significant, resulting in statistically significant rotational non-invariance of the system's performance. In particular, because of their image processing, the three minutiae extractors tested are not rotationally invariant, leading to measurable non-invariance in system patching performance. We also analyzed popular "invariant" computer vision feature extractors, such as SURF and SIFT feature extractors, showing that they too are not rotation invariant.

Given that we can measure a difference in fingerprint matching performance as the relative angle between the probe and gallery samples, the next question was what do do about it. To overcome the limitations of the rotation invariant characteristic and enhance fingerprint matching performance, we propose a new paradigm – match-twist-match. This is basically a two stage matching process, with the first coarse level match producing an estimated rotation between the probe and gallery, followed by probe twisting (derotation at that angle), and finally a rematch. This MTM approach is shown, on FVC datasets, to improve fingerprint matching of both the open source NIST NBIS Bozorth Matcher, and the Securics/UCCS Bipartite Biotope Secure template matching algorithm. The latter, when fusing two features with MTM enhancements, is the best performing secure template matching algorithm in the FVC-OnGoing competition, i.e. the MTM approach presented herein helps produce a state of the art algorithm.

Second, we explored the impact of sensors' spatial image preprocessing, showing they can create non-obvious artifacts wherein genuine minutiae can be removed and spurious, often uncommonly shaped, minutiae added. In such sensors, the processing attempts to enhance the apparent quality so the artifacts end up with inflated quality measures and clean looking images making errors in the image invisible to a human. At a minimum, such characteristics are something of which algorithm designers should be aware. Furthermore, such processing can introduce a "point of exclusion," with an apparent high-quality but false minutiae that could/should limit the legal use of such images. This renders them generally unfit for use in higher security, such as forensic aid to help criminal investigation, banking, UIDIA or other applications where human validation of prints may be needed. There is, however, no requirement for sensor manufacturers to provide the underlying raw images nor to disclose the type of preprocessing done. We argue that regulations/standards are needed in the area of spatial image preprocessing in fingerprint sensors.

References

[1] H. Bay, A. Ess, T. Tuytelaars, and L. Van Gool. Speeded-up robust features (surf). *Comput. Vis. Image Underst.*, June, 2008. 2, 4

[2] M. Bicego, A. Lagorio, E. Grosso, and M. Tistarelli. On the use of sift features for face authentication. In *CVPR Biometrics Workshop, CVPRW*. IEEE, 2006. 4

[3] T. E. Boult, W. J. Scheirer, and R. Woodworth. Revocable fingerprint biotokens: Accuracy and security analysis. In *CVPR*. IEEE, 2007. 5

[4] M. M. Cappelli, D. Maio, D. Maltoni, J. L. Wayman, and A. K. Jain. Fvc2004: Third fingerprint verification competition. In *Proc. Int. Conf. on Biometric Auth.*, 2004. 3, 5

[5] DigitalPersona. *White Paper Guide to Fingerprint Recognition*. DigitalPersona, September 2011. 1

[6] P. Dreuw, P. Steingrube, H. Hanselmann, H. Ney, and G. Aachen. Surf-face: Face recognition under viewpoint consistency constraints. In *BMVC*, 2009. 4

[7] I. Emiroglu and M. Akhan. Pre-processing of fingerprint images. In *EU Conf. on Security and Detection*, pages 147–151. IET, 1997. 6

[8] D. Gafurov, P. Bours, B. Yang, and C. Busch. Guc100 multisensor fingerprint database for in-house (semipublic) performance test. *EURASIP J. Inf. Secur.*, 2010. 6

[9] S. Greenberg, M. Aladjem, D. Kogan, and I. Dimitrov. Fingerprint image enhancement using filtering techniques. In *Int. Conf. Pattern Recognition*, 2000. 6

[10] R. G. H.C. Lee. *Advances in Fingerprint Technology*. Elsevier, 1991. 6

[11] L. Hong, Y. Wan, and A. Jain. Fingerprint image enhancement: Algorithm and performance evaluation. *IEEE Trans. Pattern Anal. Mach. Intell.*, 1998. 6

[12] S. Lazebnik, C. Schmid, and J. Ponce. Beyond bags of features: Spatial pyramid matching for recognizing natural scene categories. In *Computer Vision and Pattern Recognition*, CVPR '06, 2006. 4

[13] D. G. Lowe. Distinctive image features from scale-invariant keypoints. *Int. J. Comput. Vision*, 2004. 2, 4

[14] D. Maio, D. Maltoni, R. Cappelli, J. Wayman, and A. Jain. Fvc2002: Second fingerprint verification competition. In *Int. Conf. Pattern Recognition*, 2002. 3

[15] D. Maio, D. Maltoni, J. L. Wayman, and A. K. Jain. Fvc2000: Fingerprint verification competition. *IEEE TPAMI*, 2002. 3, 5

[16] N. B. Nill. Test procedures for verifying iafis image quality requirements for fingerprintscanners and printers. Technical report, MITRE Corp., 2005. 7

[17] U. Park, S. Pankanti, and A. Jain. Fingerprint verification using sift features. In *SPIE*, volume 6944(1), 2008. 4

[18] W. J. Scheirer and T. E. Boult. Bipartite biotokens: Definition, implementation, and analysis. In *Advances in Biometrics*, pages 775–785. Springer, 2009. 5

[19] B. G. Sherlock, D. M. Monro, and K. Millard. Fingerprint enhancement by directional Fourier filtering. *IEE Proc. Vision, Image and Signal Processing,*, 1994. 6

[20] C. I. Watson, M. D. Garris, E. Tabassi, C. L. Wilson, R. M. Mccabe, S. Janet, and K. Ko. User's guide to nist biometric image software (nbis), 2007. 1

[21] E.-K. Yun and S.-B. Cho. Adaptive fingerprint image enhancement with fingerprint image quality analysis. *Image Vision Comput.*, 2006. 6