Understanding ACE-V Latent Fingerprint Examination Process via Eye-Gaze Analysis

Aakarsh Malhotra, *Graduate Student Member, IEEE,* Anush Sankaran, *Member, IEEE,* Mayank Vatsa, *Senior Member, IEEE,* Richa Singh, *Senior Member, IEEE,* Keith B. Morris, and Afzel Noore, *Senior Member, IEEE*

Abstract—The latent fingerprint examiners often mark minutiae and perform a comparison of latent fingerprints with exemplar fingerprints of known identity. The interpretation of details in the fingerprints is based on the proficiency of examiners and how they perceive details. However, different examiners discern fingerprint regions and details differently due to an unconscious choice of certain features and details. In this study, we aim to draw inferences from the perceptual behavior by collecting eye gaze of examiners while they mark minutiae and perform comparison. The study shows the patterns observed across different forensic examiners and infers specific heuristics used by examiners to discern features. These practices could be inculcated back into an AFIS system to improve automated comparison and help train novice examiners. To draw inferences, novice and expert examiners perform latent to exemplar fingerprint comparison by following the ACE-V mechanism. During the comparison, examiners provide a value determination, quality score, and minutiae markup and confidences. 29 distinct examiners perform a total of 158 trials, where, the eye gaze is recorded simultaneously. Using the eye gaze fixation, we empirically find Region of Interests (ROI) of examiners on the prints and utilize it towards developing an understanding of the search strategy.

Index Terms—ACE-V, minutiae markup, latent fingerprint, gaze analysis, value determination

1 Introduction

Latent fingerprints are an unintentionally transferred impression of friction ridge detail [1] from the fingers. Examiners compare these latent fingerprints with exemplar fingerprints. The procedure involves assessing the value of the latent fingerprint. The examiner then compares the distinctive details in the unknown latent to the exemplar fingerprint of known source. In this process, examiner notes any and all similarities and differences between the latent and exemplar fingerprint. To determine the value and perform a comparison, there are specific steps that examiners follow. For instance, in the ACE-V procedure, examiners follow a 3-step process of Analysis, Comparison, and Evaluation [2]. These steps are abstract instructions to examiners, which guide them to perform sub-tasks for the two fingerprints in a specific order. These sub-tasks include value assessment, feature markup, and comparison.

Despite the sequential ordering of tasks, the complete process of latent to exemplar fingerprint comparison is not translated into the documentation. The primary reason is variability amongst forensic examiners [3], [4]. Different examiners discern fingerprint regions and details differently. This variability can arise due to an unconscious choice of certain features or examiner's ability to perceive details [5], [6]. Additionally, the way examiners are trained to perform comparison may also vary. While such variability amongst perception and experience exists, it is challenging to document these variations [7]. Ultimately, each examiner looks for discriminative features in latent and exemplar fingerprints that can be used for comparison.

To understand which are the regions in which examiners find distinctive details, one way is to monitor their eye gaze. Eye gaze tracking is an unobtrusive and non-invasive mechanism to observe and understand the perceptual process of an examiner. The locations of eye gaze over the displayed content provide information regarding what region is of importance to the user. Other forensic science studies have also used eye gaze information to draw inferences [8]. These inferences could be used for interpreting examiner's behavior or towards gaze-based forensics.

In the context of latent and exemplar fingerprint comparison, the location visited by the examiner's eye gives an intuition on the relevance of details to the examiner. Fixation locations, where the gaze is accumulated continuously, are regions where examiners acquire information visually. The eye gaze of an examiner, along with their corresponding annotation data, highlights: (i) on how examiners determine suitability in the analysis stage, (ii) on the minutia marking procedure in analysis and comparison stages, (iii) the properties of the location visited by the examiners, and (iv) how the conclusion decision is made by an examiner.

In this research study, we aim to understand the perceptual process of the latent fingerprint examiners while they annotate and compare a latent with an exemplar fingerprint. The understanding of perceptual process of examiners can potentially help in: (i) building better Automated Fingerprint Identification Systems (AFIS) [9], [10], (ii) reducing the effort of examiners and detecting their fatigue level while comparison [11], (iii) informed training of novice

- *A. Malhotra and A. Sanakaran are with IIIT-Delhi, India (email: {aakarshm, anushs}@iiitd.ac.in)*
- *M. Vatsa and R. Singh are with IIT Jodhpur, Rajasthan, India (e-mail: {mvatsa, richa}@iitj.ac.in)*
- *K.B. Morris is with Department of Forensic & Investigative Science, WVU, Morgantown, USA (e-mail: keith.morris@mail.wvu.edu)*
- *A. Noore is with Frank H. Dotterweich College of Engineering, Texas A&M University, Kingsville, USA (e-mail: afzel.noore@tamuk.edu)*

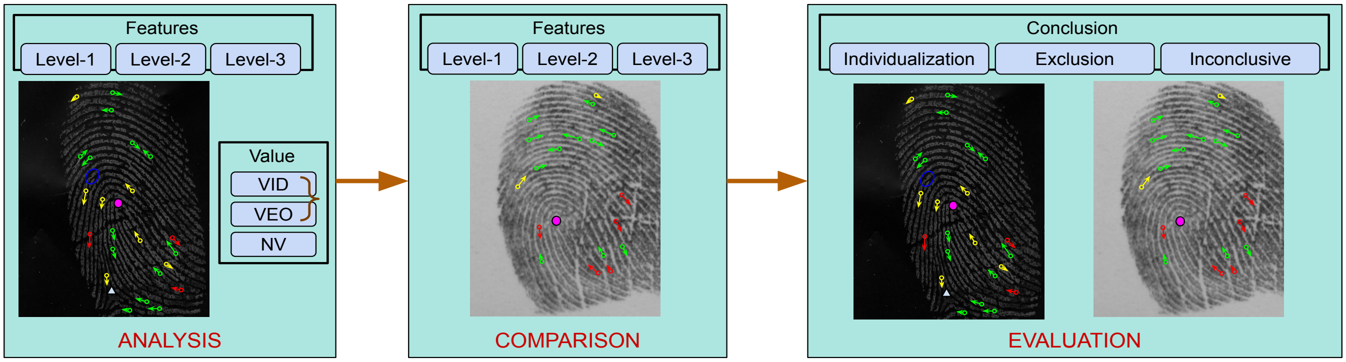


Fig. 1: ACE mechanism for comparing latent fingerprint to exemplar fingerprints

examiners [12], and (iv) act as a testimony for the examiner's conclusion decision in the court of law [5], [6], [13]. The five-fold contribution of this research is as follows:

- **In a real world scenario, data collection from forensic examiners:** Data collection in this study aims to gather the markup information from the examiners on latent fingerprints. Concurrently, eye gaze is tracked in a non-intrusive manner. 29 subjects took part in the study, ranging from certified, experienced examiners to novice examiners who have just completed a course on latent fingerprints.
- **Real-world setup for fingerprint comparison:** In a real-world scenario, there is no time restriction for examiners to perform comparison. Hence, unlike other studies where examiners made the conclusion within a few seconds [3], [14], [15], our study imposes no such restrictions. The data collection follows an ACE procedure, where examiners completes the analysis of the latent fingerprint before looking the exemplar fingerprint" [2].
- **First eye-gaze study on latent fingerprints with minutiae markup, confidence, and value determination:** While researchers have used eye gaze information to understand the thought process of latent fingerprint examiners, ours is the first study to combine non-intrusive eye gaze collection with an explicit marking of the core, delta, and minutia with their confidence. The exact locations of core, delta, and minutia are recorded with the sequence of the ACE-V process.
- **Region of Interest (ROI) algorithm:** Using the eye-gaze information, a novel K-means clustering based algorithm is proposed to find Region of Interest (ROI) of latent prints.
- **Understanding the human process:** The marked minutiae and value assessment by examiners act as annotation data for our experiments. The annotations and the eye gaze locations during the ACE-V procedure helps us find the rationale behind the steps followed by the forensic experts.

2 The ACE-V procedure

The process of ACE-V (Analysis, Comparison, Evaluation, and Verification) aids in sequential assessment of latent fingerprint and comparison with the exemplar fingerprint [2]. The ACE-V method ensures that the examiners reach one of the four conclusions with consensus:

- **No Value:** The latent is not suitable for comparison since it does not have sufficient information.
- **Individualization:** There are sufficient details in agreement to conclude that the latent fingerprint and the exemplar fingerprint belong to the same individual.
- **Exclusion:** There are sufficient details in disagreement to conclude that the latent fingerprint and the exemplar fingerprint are from different sources.
- **Inconclusive:** There is neither sufficient agreement to individualize nor sufficient disagreement to exclude [1] the two fingerprints.

ACE-V is a four-step sequential process, where examiners perform specific tasks. The four stages and the tasks performed are shown in Fig. 1 and are described as follows:

1) **Analysis (A):** Analysis stage begins with assessing latent fingerprint towards the suitability and sufficiency [16] for comparison. The examiner performs a manual markup of the established features (level-1, level-2, and level-3) [17] with personal comments (notes). Examiner also assigns a value determination as either (i) Value for Identification (VID) or (ii) Value for Exclusion Only (VEO) or (iii) No Value (NV). In the case of No Value (NV), the examiner does not proceed to the next stage, and no conclusions (individualization or exclusion) [18] are made.
2) **Comparison (C):** In Comparison [19] stage, examiner marks features in the exemplar fingerprint and compares with the features of the latent fingerprint.
3) **Evaluation (E):** From the inferences of comparison stage, the examiner makes a conclusion decision of either individualization, exclusion, or inconclusive.
4) **Verification (V):** A subsequent examiner performs an independent examination of fingerprints using ACE procedure.

Since Verification (V) step is a repetition of the ACE mechanism by a different examiner, we refer to ACE-V as ACE mechanism throughout the paper. In these stages, human examiners perform tasks such as quality assessment and minutiae markup. In Fig. 2(a)-(c), there are a vast set of details that examiners can use for value determination and minutia marking. However, the details considered and their exact locations on the fingerprint may vary. For instance, in Fig. 2(a), the decision of enclosure depends on the examiner's ability to perceive fine details. Similarly, whether to consider the region in Fig. 2(b) for the task of comparison, depends on how the examiner interprets the quality and clarity of the region [20], [21]. The primary reason is that,

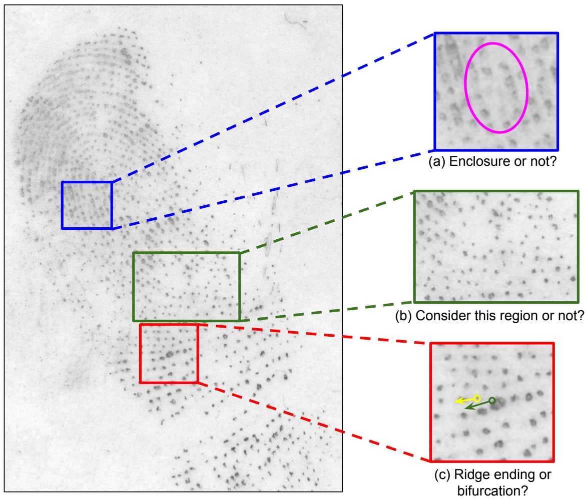


Fig. 2: Understanding the process behind human examiners analyzing a latent fingerprint.

though it is known what features are to be marked, the actual process the examiners follow and the entire cognitive process is not translated into documentation. Also, many studies [4], [16], [22], [23], [24] highlight existence of variability across examiners regarding the number of marked minutiae and the value of the latent print.

3 Literature Review

Various studies have aimed to interpret external factors and examiner's behavior while fingerprint comparison. For instance, in 2015, Ulery et al. [25] studied the changes made by examiners in markup while they proceeded from analysis to the comparison stage. One of their findings is that examiners usually deleted or added minutia when examiners making an individualization conclusion. Ulery et al. [26] further discovered factors such as minutia count and presence of core and delta, which contribute to the exclusion determination of latent fingerprint. Vogelsang et al. [27] explored if examiners holistically process latent fingerprints to be compared against exemplar fingerprints. Recently, Dror and Langenburg [28] studied cognitive behavior during inconclusive conclusions. Additionally, the authors discuss and suggest scenarios where inconclusive conclusions are justified and when they are not.

Such studies develop an understanding of a comparison process of examiners. However, a closer look can be achieved by tracking the exact location an examiner is looking. Eye gaze tracking is an unobtrusive and non-invasive procedure. The examiners perform the ACE procedure on a monitor while a device tracks their exact gaze location on the screen. Studies have integrated eye gaze information to understand the perception process while examiners compare latent with exemplar fingerprint. Yu et al. [14] in 2010 introduced the concept of eye gaze monitoring for understanding latent fingerprint examiner's ROI. The eye gaze is collected with 12 expert and 12 novice examiners performing comparison by spending 20 seconds on a latent and exemplar fingerprint pair. They devised a mechanism to map the corresponding ROI in latent fingerprint and exemplar fingerprint. Using this mechanism, authors compared the eye gaze of novice with expert examiners. One of the takeaways from their study was that the expert examiners find better corresponding ROI than novice examiners. In 2011, Shapoori and Allinson [29] studied search strategy of 10 examiners for a single latent fingerprint using eye gaze data. They utilized clustered eye gaze data to train a neural network for predicting search strategy.

Busey et al. [3] collected eye gaze of latent fingerprint examiners in two experiments. In the first experiment, six novices and six experts performed comparison of latent fingerprints with exemplar fingerprints (duration: 1 minute). With same setup, the second experiment was conducted with 20 seconds duration with participants increasing to twelve novices and twelve experts. They compared the gaze of two examiners using Earth Mover's distance [30]. Authors concluded that: (i) fixation of experts were consistent in the second experiment than the first, (ii) experts had better performance, and (3) experts marked higher number of "too soon to tell". In 2013, Busey et al. [15] performed another experiment where twelve novices and twelve experts took part in a latent and exemplar fingerprint comparison experiment (duration: 20 seconds). On comparing the forward and backward gaze movements, authors concluded that novice examiners made shorter sequences and more "links" are present in the gaze of experts. In 2015, Busey et al. [11] studied the impact of fatigue on latent fingerprint examiners using their eye gaze. They collected eye gaze before and after fatigue experiment. A few outcomes of their experiment were that fatigue induces participants to give up sooner on a fingerprint which was clear. However, there was no significant difference in the mean fixation duration and Earth Mover distance across sessions. Recently, Hicklin et al. [31] aimed to develop a cognitive understanding of region localization. The examiners were given a region as context in the latent fingerprint, which they had to localize in the exemplar. With 117 subjects, 675 trials of 8-10 seconds were performed under three scenarios. Authors conduct experiments which include analysis on gaze (like fixation and saccades), time taken on localizing the target, and the speed of gaze while searching the target. Authors concluded that context affects the gaze behavior of examiners. Additionally, the study claimed that repeated back and forth of gaze between latent and exemplar fingerprints suggests the limited scope of visual memory.

4 Materials and Methods

As a part of analyzing the strategies adopted by examiners for minutiae markup and value assessment, the first step is to collect data which could encode the perception and perspective of examiners during markup. We collect the annotation data and eye gaze data for 29 subjects. These data are obtained when a stimulus is shown to examiners. The stimulus is a pair of latent and exemplar fingerprints, presented to each examiner following the ACE mechanism. Each examiner provides annotation and analysis of 3 to 7 pairs of latent and exemplar fingerprints. The dataset includes 158 trials, out of which eye gaze remained calibrated in 148 samples. The dataset also contains annotation data, which corresponds to the minutiae markup (location and confidence), conclusion decision, and value determination. The details of the experiments are described as follows.

4.1 Participants

In the study, a total of 29 participants took part. Each participant is above the age of 18 years, with age ranging from 19 to 74 years. The age of examiners is grouped into either of: (a) 18-20, (b) 20-25, (c) 25-35, (d) 35-50, (e) 50-60, or (f) above 60 years. As a forensic science examiner, their experience ranged from 6 months to 30+ years. These examiners are categorized as either a novice or an expert examiner. A novice examiner is a participant with basic knowledge of fingerprints obtained on the completion of a course on latent fingerprints or fingerprints. On the contrary, an expert examiner includes practitioners and course instructors with at least two years of experience in fingerprint analysis. The experience is categorized as: (a) less than 1 year, (b) 1-2 years, (c) 2-5 years, or (d) more than 5 years. A summary of the responses is shown in Fig. S.1[1] (supplementary). The novices are students from the WVU Department of Forensics and Investigative Science. Experts are from various forensic science institutions, who volunteered during their visits to Chesapeake Bay Division of the International Association for Identification (CBD-IAI) conference in April 2017. The study was approved by the IRB at WVU.

4.2 Procedure

The latent-exemplar fingerprint pair is shown to each participant on a 21″ LCD monitor at a resolution of 1920 × 1080 pixels, with each fingerprint displayed at a resolution of 700 × 960. A custom tool is built for the examiners to perform comparison. Sample screenshots of the tool in each stage of the ACE mechanism is shown in Fig. S.2(supplementary). As a part of this research, the tool will be released for the research community[2].

While the stimulus is shown in the form of latent and exemplar fingerprint for comparison, participants have to follow a specific protocol. The protocol is motivated by the ACE-V [2] procedure of manual fingerprint comparison and thus, is partitioned into three steps, namely, Analysis, Comparison, and Evaluation. In the Analysis stage, the examiner determines if a latent fingerprint has enough value for further processing or not. The examiner provides the value determination of the fingerprint, where, in case of No Value (NV), the latent fingerprints are discarded from further processing. However, in our experiments, examiners are instructed to proceed with those latent fingerprints as well. It facilitates to collect crucial information in the next stages, like the knowledge on how examiners approach minutiae markup in low-quality regions.

The participant then marks the level-1 features (core and delta points) followed by level-2 features (ridge endings and ridge bifurcations). The level-2 features are marked using a GYRO color coding scheme [32]. The examiners mark minutia using one of the three colors: green, yellow, or red. Minutiae annotated in green are the highest confidence minutiae whereas the ones marked in red are of the lowest confidence. In our study, orange coloring is not included. Orange color in GYRO scheme indicates minutia marking in a later stage (for instance, latent fingerprint minutia marked in Comparison stage). Since we disallow participant to mark or edit minutia in subsequent stages, our experiment does not include the orange component of GYRO scheme. Each examiner is instructed to mark as many minutiae as they would do in a case work. Each minutia is marked by two clicks, the first denoting the location (x, y) and the second click denoting the orientation (θ). Once all minutiae points are marked, the examiner provides the quality score of the latent fingerprint and the lifting technique.

1. Fig. numbers in supplementary material are prefixed with 'S.'.
2. Link for tool: http://iab-rubric.org/resources/ace-tool.html

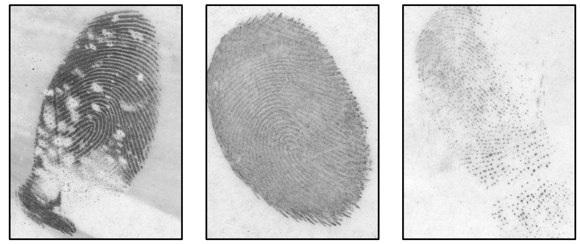


Fig. 3: An illustration of some varied quality latent fingerprints, used as stimuli in the experiments.

In the Comparison stage, an exemplar fingerprint is shown next to the latent fingerprint. The examiner analyzes the exemplar fingerprint in comparison with latent fingerprint and marks level-1 and level-2 features for comparison. The protocol for feature markup is same as the one followed in the latent fingerprint analysis stage. After feature markup in the latent and exemplar fingerprint, the examiner makes one of the three conclusion decisions, "Individualization"(same source), "Exclusion"(different source), or "Inconclusive" in the Evaluation stage.

4.3 Database

The stimuli for the experiment is created from the fingerprints in FBI BioCoP Latent Fingerprint database[3] [33]. These stimuli contained latent fingerprints developed using varied techniques such as black powder, super glue, and ninhydrin. Atmost seven latent and exemplar fingerprint pairs are shown to each examiner, with the flexibility to annotate and compare lesser number of fingerprint pairs. A total of six batches are created for the experiment. Of these six batches, four batches contain seven latent-exemplar fingerprint pairs (batches 1-4), one has three pairs (batch 5), and another contains six latent-exemplar fingerprint pairs (batch 6). In the experiment, the participant is shown one of the six batches to perform the comparison using the ACE procedure. The batch is selected depending on the examiners' time availability. To incorporate the effect of the quality of the latent fingerprints, the pairs shown comprise of latent fingerprints with diverse quality. For batches 1 to 4, a latent-exemplar fingerprint pair is shown twice to check repeatability for an examiner, whereas, a pair of fingerprints is same across all batches to examine reproducibility amongst examiners. The latent fingerprint in the repeated pair across the participants is chosen to be of good quality.

While examiners mark features and compare the fingerprints, multiple information is collected. The data collected is shown in Fig. 4 and is summarized as follows:

3. As per database regulations, any images of the database are not shown in this paper. The lookalikes are created in a lab environment for illustration. A few sample images are illustrated in Fig. 3.

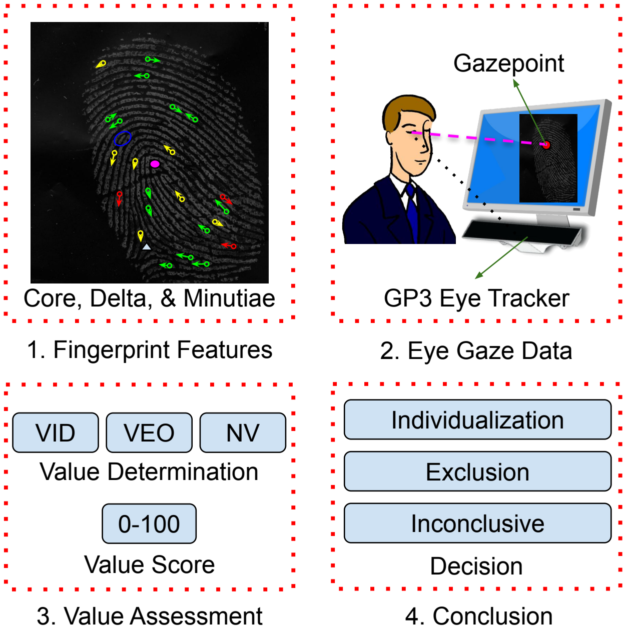


Fig. 4: A Summary of the data collected.

1) **Fingerprint Features:** These include level-1 (core and delta) and level-2 features (ridge endings and ridge bifurcations) marked by examiners for both the latent fingerprint and the exemplar fingerprint.
2) **Latent Fingerprint Value Assessment:** The value assessment includes the value of the latent fingerprint (VID: Value for Identification, VEO: Value for Exclusion Only, and NV: No Value) as determined in the analysis stage. The examiner also provides a quantitative number ranging from (0-100) that best indicates the latent fingerprint quality.
3) **Eye Gaze Data:** The eye gaze is collected through the Gazepoint GP3 eye tracker with Gazepoint Analysis UX edition tool. The sensor operates at 60 Hz and has a ± 15 cm range of depth movement [34]. The device has a 0.5 – 1° of visual angle accuracy and is suitable for screen sizes smaller than 26". The tool provides the following data in a CSV format:
   • **Gaze data** (x, y): Location of the gaze point on the screen along with its recorded time and a valid flag (to check for calibration).
   • **Fixation point**: A fixation point is the center of a small region where the gaze is accumulated continuously for more than t time. The (x, y) location along with the amount of time is recorded.
   • **Pupil details**: The diameter of pupil of both eyes and a scale factor denoting how near the examiner is to the eye tracker is also recorded.
4) **Comparison Data:** It includes the decision made by the examiner: "Individualization"(same source), "Exclusion"(different source), or "Inconclusive".

As a part of this study, we will release the eye gaze distribution and the annotation provided by the examiners.

5 Region of Interest (ROI) Estimation

While searching for relevant features in a fingerprint, the eye gaze of the examiners consists of fixations and saccadic

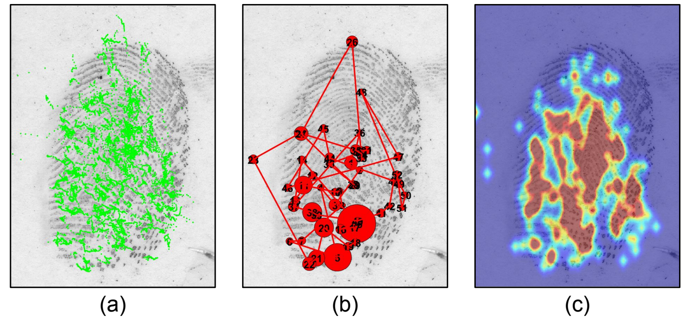


Fig. 5: Examiner's eye gaze during latent fingerprint analysis: (a) Eye gaze locations, (b) Fixation map, and (c) Heat map.

movements. Fixation location highlights the place where examiners found relevant details and looked for features. On the other hand, saccades are rapid eye gaze transitions between a pair of fixations. Saccades convey ridge trace information followed by the examiner. Consequently, in our experiments, each gaze location visited by the examiner is of importance. However, individual points of gaze position are scattered and also increases the computation while inferring. To overcome this challenge, we introduce a clustering-based approach to combine gaze location and find Region of Interests (ROI) in this section.

As shown in Fig. 5, the eye gaze pattern is distributed across the latent fingerprint. However, if an examiner spends more time in a particular region, the region would have more gaze points. Such regions are labeled as high frequency regions. Similarly, regions where gaze points are tightly packed are labeled as high density regions. Thus, regions with high frequency and high density of gaze patterns would mark a ROI for an examiner. If examiners find relevant features or think they might find some detail in a particular area, they tend to look it more often. Such areas hold importance to the examiners for discriminative information and are defined as Region of Interest (ROI). Fixations are capable of representing such regions. They represent a pixel location where an examiner looked for a time greater than an arbitrarily chosen t seconds.

Latent fingerprint examiners perform tasks such as tracing ridges and looking closely at the spacing between ridges to infer the presence of a minutia. Hence, we need to group fixation for broader understanding of surrounding regions. To encode ROI of examiners based on their gaze, the gaze points are grouped using K-means clustering [35]. The gaze points are clustered into pre-defined k-clusters. k is hypothetically chosen as 15 to make circular clusters with nearly 150-pixel diameter. A larger k decreases the diameter size of the region. However, a small-sized region would fail to incorporate the pattern of eye gaze when minutiae are present in nearby areas, and they end up in different clusters. On the contrary, a smaller value of k would increase the cluster size. This can result in the creation of regions with mixed properties. For instance, a larger cluster region may have bad and good quality areas together, or, many minutia points falling inside one region itself. This would affect the study of region specific characteristics. Hence in our experiments, with $k = 15$, the region size is optimal between losing local information (larger k) and mixing characteristics of unrelated areas (smaller k). It would segregate

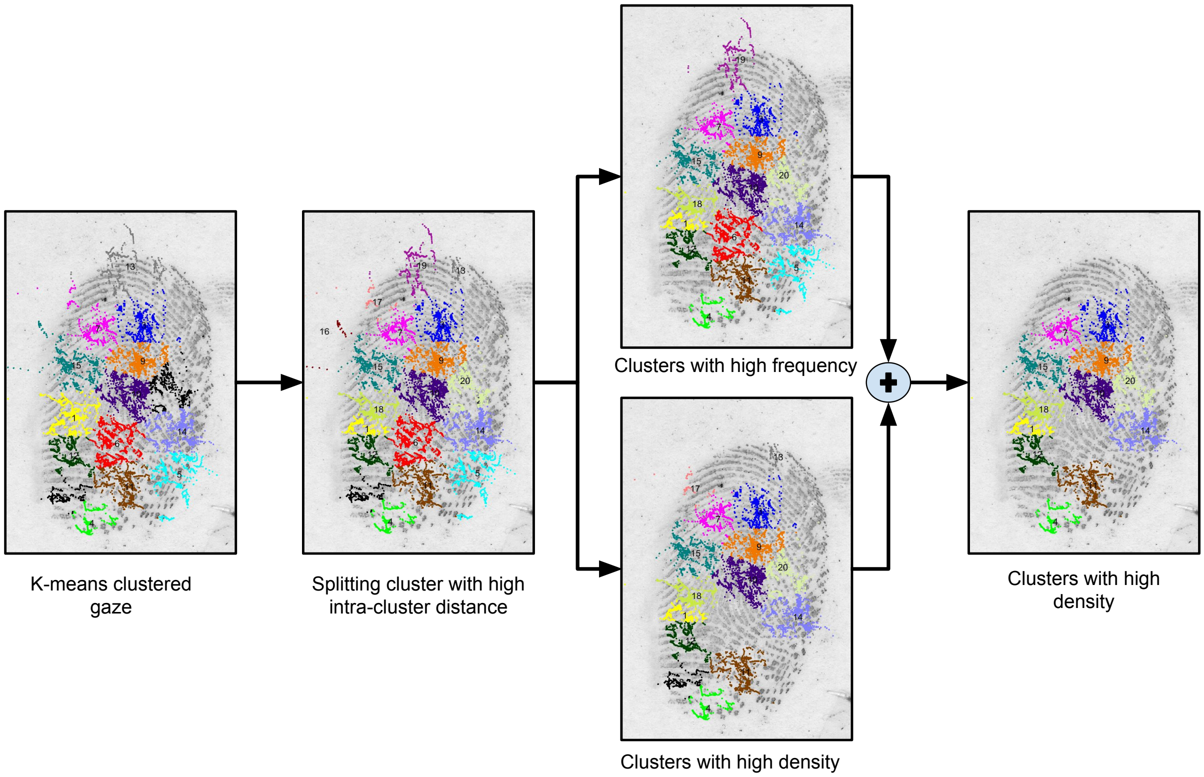


Fig. 6: Analyzing the region of interest (ROI) for examiners based on their gaze.

fingerprints into compact regions of similar properties while keeping contextual information of eye gaze intact.

Fig. 6 illustrates the clusters obtained for an examiner's eye gaze pattern. As seen from the figure, there are a few clusters with high intra-distance between extreme eye gaze points. These are mainly due to the presence of eye gaze outlier points in the cluster. We observed that about one-third of the clusters have gaze outliers. The outlier clusters are generally present near the fingerprint boundary. To resolve the above concern, the aim is to further partition an outlier cluster such that outliers are removed, and the cluster has high density of eye gaze. The partitioning to remove outliers is achieved by splitting the cluster with outliers into two clusters (thus, $k = 2$). The division ensures that one cluster has the outliers, while the other contains the eye gaze clutter. Since approximately one-third clusters had outliers, we chose five clusters ($15/3$) with the highest intra-distance for re-clustering. The value of k is selected as two. The obtained clusters are the final set of regions from which the ROI of examiners is found in the next step.

Of all the clusters after the previous step, a cluster (or region) could be of interest to an examiner if the region has more gaze points (high frequency) and these gaze points are tightly packed (high density). Hence, 15 high-frequency and 15 high-density clusters are selected. Finally, ROI is found by taking into account those clusters which are common amongst the high frequency and high-density clusters. This output and the complete algorithm to obtain the ROI is illustrated in Fig. 6. Utilizing the ROI clusters, we study properties of regions where the examiners look more often.

6 Results

The analysis can be broadly classified into two categories. The first is based on finding inferences from the eye gaze. With eye gaze, we infer the characteristics of regions visited by examiners and the thought process of the examiners. The second analysis revolves around value determination and the minutia count. Thus, the second part of inferences includes all 158 samples, whereas, 148 calibrated trials are considered for the first part.

6.1 Characteristics of Regions Focused by Examiners

Different examiners follow different strategies for markup and comparison. Using gaze and markup data, the results include: (i) Gaze statistics, (ii) Relation of quality and gaze clusters, (iii) ROI and clarity variation, and lastly (iv) Analysis on area with high gaze concentration.

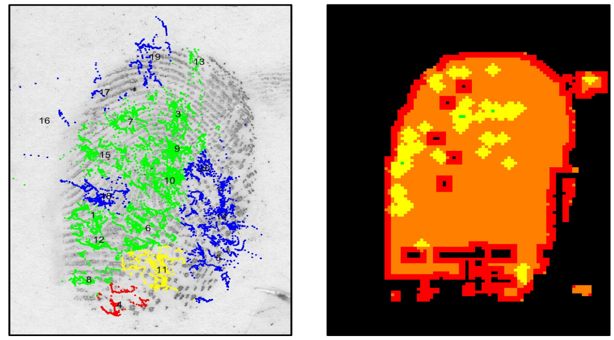
6.1.1 *General Gaze Statistics*

Using the 148 calibrated samples, we first calculate the low-level gaze statistics across the individual stages of analysis, comparison, and evaluation. These statistics have been summarized in Table 1. We present the experimental duration, the fixation details, and the saccadic information of eye gaze.

In 2011, Busey et al. [3] collected approximately 38,400 seconds of eye gaze data of latent fingerprint examiners in 2 different experiments. In our research, eye gaze data is collected while examiners compare latent fingerprint with an exemplar fingerprint using ACE mechanism. From 29 subjects, 39,550.28 seconds of calibrated eye gaze data is collected. This is the first of a kind research where examiners mark minutiae and compare fingerprints while following the ACE methodology. The eye gaze collected in our study

	Overall ACE	Analysis (A)	Comparison (C)	Evaluation (E)
Total gaze time (sec)	39550	22046	16451	1053
Avg. time per sample (sec)	267.23	148.96	111.16	7.11
Avg. fixation duration (sec)	0.62	0.62	0.63	0.40
Avg. fixation time (sec)	218.68	119.80	93.22	5.65
Avg. number of fixations	350.47	189.43	148.44	12.60
Avg. saccade count per min	103.59	96.35	125.82	215.24
Avg saccade length (pixels)	143	128	160	143

TABLE 1: Gaze statistics from the calibrated samples.



(a) (b)

Fig. 7: (a) Cluster colors using minutiae marked inside them. (b) Quality maps using NBIS Mindtct for 8×8 blocks (black: 0 (worst), red: 1, orange: 2, yellow: 3, green: 4 (best)).

depicts a real world comparison scenario. In terms of the duration of eye gaze recording, it is highest amongst all the research efforts in comparison of latent fingerprints with exemplar fingerprints. Further, participating examiners are not restricted in terms of time for making a decision. On an average, an examiner took 267.23 seconds for making a decision, unlike earlier studies, where examiners were given fixed viewing duration of maximum 20 seconds [3], [14], [15], [31] or one minute [3].

In our experiments, it is observed that examiners tend to spend more time in the analysis phase compared to comparison and evaluation phase. The higher time can be affirmed by noting the total time, average time, overall fixation duration, and the number of fixation, which are all highest in the analysis phase. It highlights that the *examiners mark features conservatively in the analysis phase*, since latent fingerprints are of lower quality compared to the good quality exemplar fingerprint image. The least time is spent in the evaluation stage, with an average duration of 7.11 seconds and least number of fixations of 12.60. This indicates that *the conclusion decision based on the minutiae markup is largely determined in the comparison phase itself and evaluation phase is utilized towards the confirmation*.

The average fixation duration of the evaluation stage is also quite small for evaluation phase (5.65 seconds). We observe that the total time, average time, overall fixation duration, and the number of fixation are less in comparison phase compared to the analysis phase. We can infer that *examiners are able to mark minutiae quickly in exemplar fingerprints due to the clear ridge valley details*. Regarding average fixation duration, there is no significant difference between the analysis and comparison stages (0.62 seconds vs. 0.63 seconds: $p = 0.55$). However, since no minutia needs to be marked, the average fixation duration is only 0.40 seconds in the evaluation stage, which is significantly higher in analysis and comparison stage ($p = 4.1 \times 10^{-29}$ and $p = 7.2 \times 10^{-27}$ respectively). *Due to marking of minutiae, examiners transition slowly and fixate longer to mark minutiae precisely*.

Saccades are the rapid eye gaze transitions between a pair of fixations. The number of saccades and average saccade length can convey if an examiner is looking in surrounding areas while marking a minutia. A lower saccadic length signifies examiners integrate information from nearby areas. At a fingerprint display resolution of 700×960, ridge width ranges from 7.58 to 13.97 pixels ($\mu = 9.93$ pixels). Hence, a saccadic length of 10 pixels approximates to width of one ridge.

As seen from Table 1, the average number of saccades made by examiners increases from 96.35 to 125.82 ($p = 7.7 \times 10^{-4}$) as they move from analysis to comparison stage. It further increases to 215.24 in the evaluation stage ($p = 1.6 \times 10^{-8}$ with respect to analysis and $p = 1.8 \times 10^{-5}$ with respect to comparison). These values are also similar to average saccade length, where the lowest is 128 pixels for analysis phase compared to 160 pixels in comparison and 143 pixels in evaluation ($p = 3.4 \times 10^{-143}$ for analysis-comparison, $p = 2.9 \times 10^{-8}$ for analysis-evaluation, and $p = 1.9 \times 10^{-8}$ for comparison-evaluation). This clearly indicates that while marking the features in the analysis stage, examiners move fluently and integrate visual information from nearby areas than erratic movements of gaze. In both comparison and evaluation phase, eye movements cover longer distances due to which higher number and longer saccades exist. This is a clear indication showing that *examiners tend to trace ridges when the fingerprint is of lower quality (latent fingerprint in analysis phase)*.

6.1.2 *Relation of Quality and Gaze Clusters*

To analyze the clusters made by K-means and validate the relation between the ROI and region quality, the minutiae markup confidence and quality map (as returned by mindtct for each 8 × 8 blocks [36]) are used. The clusters made by the algorithm described in the above section are given a color (green, yellow, and red) based on the average minutia confidence of minutiae marked inside the clusters. For example, if there are three green and one yellow confidence minutia present in a cluster, the cluster is assigned as a green color. Similarly, if there are one red and one green confidence minutia in a cluster, the cluster is assigned as a yellow color. There would be some clusters where there are no minutia marked by an examiner. These clusters are assigned the color blue. Fig. 7(a) represents an illustration of cluster coloring.

Ideally, there should not be a red cluster with many green confidence minutiae. Similarly, there should not be green clusters with many red confidence minutiae. To understand average cluster colors, the count of each cluster color with their constituent minutia confidences and count

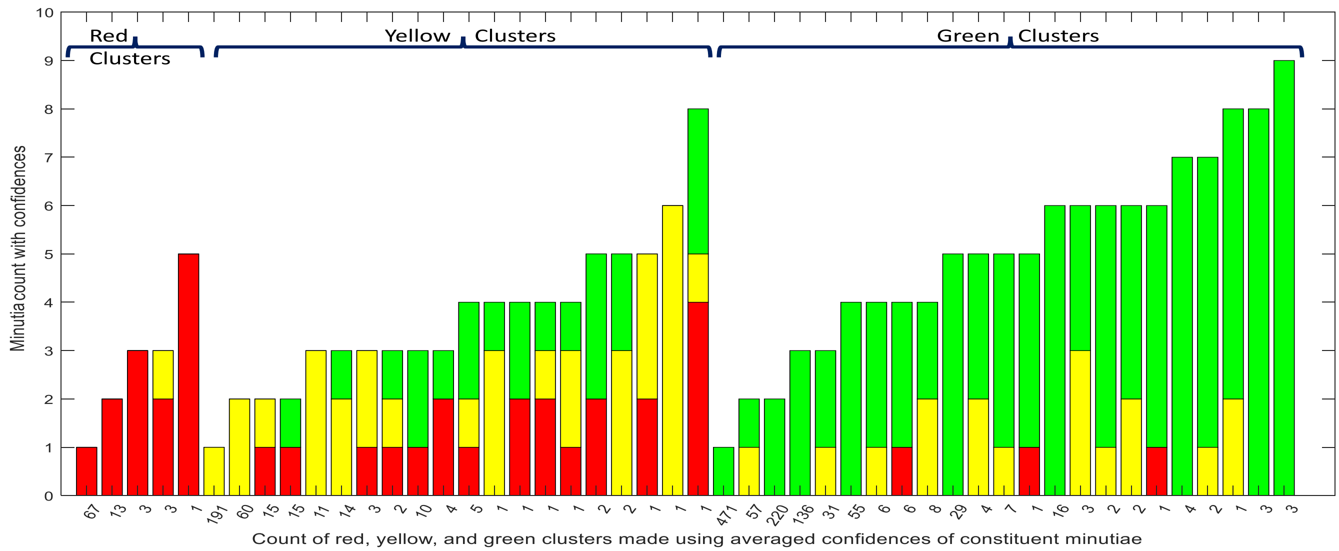


Fig. 8: Analyzing count of red, yellow, and green clusters. The color coded clusters represent the average confidence of the minutiae present in the cluster.

is shown in Fig. 8. It can be seen that red confidence clusters do not contain any green minutia. Similarly, the number of high (green) confidence regions (clusters) with red minutiae is also quite low. However, in regions (or clusters) where examiners labeled both red and green minutiae, the regions are typically assigned yellow confidences.

Next, blocks of size 32×32 are generated from the four colored clusters. To find the relation of cluster confidences to the fingerprint quality, the quality score of latent fingerprint is extracted using quality map generated from mindtct of NBIS. While NFIQ tool is known to work for fingerprints, it mostly returns the worst quality score for latent fingerprints. However, the quality map score for each 8×8 block gives diverse score ranges between 0 to 4 for latent fingerprints. A score of 0 represents the lowest quality, whereas, a score of 4 signifies the best quality. A quality map is illustrated in Fig. 7(b). Mindtct gives a quality map for each 8×8 region, and average quality for each 32×32 region is extracted from averaging quality across 16 regions of size 8×8.

The mean quality of 32×32 blocks generated from the clusters using the marking done by examiners are as follows: red cluster blocks: 1.41, yellow cluster blocks: 1.83, green cluster blocks: 2.32, and blue cluster blocks: 1.24. As seen from these values, mean quality of green blocks are highest, followed by yellow blocks, and blocks from the red clusters have the least quality. *It confirms that there is a direct relationship between the minutia confidence with the region quality*. Thus, if minutia colors are marked correctly, minutia confidence is a useful metric to determine the region quality. The quality of blocks from blue clusters are even lower than quality of red cluster blocks. Blue clusters have no minutia marked due to the low quality or a background,resulting in quality map score of 0.

Considering clusters for being a part of ROI, we observe that *examiners tend to select better quality regions as their ROI*. For 32×32 blocks generated from red clusters, the mean quality is 1.58 for ROI blocks and 0.96 for non-ROI blocks ($p = 2.6 \times 10^{-11}$). Similarly, mean quality of ROI blocks

Cluster Color	Is ROI?	Avg. quality (0-4)	
		Individual	Combined
Red	✓	1.58	1.41
Red	×	0.96	
Yellow	✓	1.94	1.83
Yellow	×	1.53	
Green	✓	2.36	2.31
Green	×	2.23	

TABLE 2: Mean quality of 32×32 blocks from ROI and non-ROI clusters, which shows that examiners focus on better quality regions over regions with unclear ridge valley details.

from yellow cluster is 1.94 compared to 1.53 for non-ROI blocks ($p = 2.2 \times 10^{-21}$). For blocks of green ROI clusters, the average quality of 2.36 is observed compared to 2.23 for non-ROI blocks ($p = 1.3 \times 10^{-9}$). The details of number of blocks derived from the clusters is shown in Table 2. Despite having an average quality of 2.23, fingerprint regions end up in non-ROI category. The rationale behind this observation is that examiners preferred some latent fingerprint regions for comparison over another despite having similar good quality. One such instance, which is discussed in the following sections, is that examiners preferred regions closer to the core and delta. Thus, in all three scenarios, examiners prefer to search minutiae in better quality regions as ROI have higher quality map score compared to non-ROI regions.

6.1.3 *Region of Interest and Clarity Variation*

Using ROI from the analysis stage, we study the relation of ridge clarity [37] and ROI. Clarity is the discernibility of fingerprint features, specifically, whether or not fingerprint features are perceivable or not. As highlighted earlier, experts mark regions of the latent fingerprint as good, bad, and ugly. Using the average of region clarity provided by different experts, we establish the ground truth clarity for each pixel in the latent fingerprint.

From 148 calibrated trials, we obtain the average clarity of each location where the examiner gazed by considering the annotated region clarity. Next, the average clarity is found for each location which is a part of the examiner's ROI. On comparing the average clarity of all gazed locations with the average clarity of gaze locations which are part of ROI, it is found that examiners prefer to concentrate more on clear regions. This is because out of a total of 148 trials, the average clarity of ROI locations is higher than all gazed locations for 130 instances. On the other 18 occasions, where the average clarity of ROI locations is lower, the standard deviation of gaze locations increased on 11 cases compared to all gazed locations. *It indicates that while searching for features in a higher clarity region, examiners extended their search in lower clarity areas to find relevant details*.

6.1.4 *Areas with High Gaze Concentration*

Next, we visualize the eye gaze data on a latent fingerprint in the form of heat maps and fixation map.

- **Heat map:** Heat map of eye gaze presents a distribution of regions which are viewed for the maximum duration.
- **Fixation map:** Fixation map is a sequential graph-like representation, which shows the order in which the image is gazed. The radius of each circle in the sequential graph corresponds to the duration of gaze.

Visualizing both heat maps and fixation maps provide an intuition about which regions of the latent fingerprint helped in making the decision. Fig. 5(b) and Fig. 5(c) shows the fixation map and heat map of an examiner in the analysis stage. The heat map shows that during analysis, most of the gaze is concentrated around the singular points. Fixation map representation provides the sequence flow of gaze, which help us validate corresponding areas between latent fingerprint and exemplar fingerprint.

Building upon heat maps, we analyze if the regions around core and delta are of interest to examiners or not. For each gaze point in ROI and non-ROI fingerprint regions, the nearness of these gaze points to the core and delta points is calculated based on Euclidean distance as follows:

1) Depending on the type of latent fingerprint, either one or two deltas are marked manually.
2) In some cases, the region around core or delta is not clear. In such situations, core and delta are marked with approximation. In our experiments, the core is always present in all latent fingerprint except one. In this scenario, both core and delta are not marked.
3) For each gaze point inside the ROI, we calculate the Euclidean distance from the gaze point to the core and the delta(s). Considering the minimum of the two distances, the mean distance is calculated for all gaze points inside the ROI.
4) The above step is repeated for non-ROI gaze points.

On average, the distance of gaze points, lying in ROI, from the singular points is 186 pixels compared to 224 pixels for gaze points which are not a part of the ROI ($p = 3.2 \times 10^{-6}$). The median distance for gaze points lying in ROI is 181 pixels compared to 206 for gaze points which are not part of ROI. It highlights that ROI areas, where gaze concentration is high, are closer to the singular points. Individually, out of a total of 148 calibrated samples, the gaze points present in ROI are closer to the core and the delta in 109 cases (73.65%).

The above results show that *examiners use core and delta points as reference while searching for minutiae*. The relative positioning of minutia points around the core or delta act as a distinct trait during the comparison, and ergo, high concentration is observed near core and delta points. This inference can be utilized towards developing better AFIS.

	Earth mover distance (pixels)		*p-value*
Experience	Expert	Novice	6.1×10^{-6}
	27.37	32.20	
Correctness of Conclusion	Correct	Incorrect	2.2×10^{-7}
	29.58	38.07	

TABLE 3: Earth mover distance amongst the gaze distribution across two different experiments.

6.1.5 *Inferring fixations using Earth-mover distance*

In this experiment, gaze distribution is compared using the Earth mover metric [30]. Earth mover metric finds a dynamic solution to convert one set of fixation distribution to the other, with the least effort. The distance corresponds to effort made in converting one set of fixation to the other. In an ideal scenario, two examiners analyzing the same latent fingerprint regions would mean the same fixation distribution. Hence, it would imply a close to zero Earth mover distance.

In the first experiment, the Earth mover distance between all expert instance pairs (expert-to-expert) is compared against the distance between all pairs of novice instances (novice-to-novice). While making pairs, we ensure that the same latent fingerprint is being analyzed. The results are reported in Table 3. We observe that experts as a group are more consistent than novices, with lower earth mover distance amongst their gaze distributions (27.37 and 32.20 pixels, respectively). It can be attributed to expert examiners implicitly agreeing on a particular set of fingerprint features, worthy enough for a confident conclusion.

In the second experiment, we compute Earth mover distance when examiners make an incorrect conclusion (incorrect-to-incorrect) vs. examiners making a correct conclusion (correct-to-correct). While making pairs, we ensure that the same latent fingerprint is being analyzed. We observe that examiners are more consistent during correct conclusions with a lower distance of 29.58 pixels in comparison to 38.07 pixels for incorrect conclusions. A possible explanation could be that looking at lesser-visited indiscriminative (or unclear) regions results in higher Earth mover distance. Eventually, the selection of indiscriminative or unclear regions during analysis results in the incorrect conclusion.

6.1.6 *Effect of the Choice of Clustering Algorithm*

The proposed algorithm uses K-means to group and understand eye gaze patterns. To understand the generalizability of previous observations across different clustering algorithms, we performed experiments with Mean Shift and EM based clustering approaches as well. The key observations are summarized below and the detailed results are available in Section 2 of the supplementary material.

- The observations are consistent across the three clustering algorithms. For all three clustering algorithms, the average quality of ROI is better than non-ROI clusters.

- The combined region quality increases from red to yellow to green cluster regions, validating the direct relationship between the minutiae confidence and the region quality.
- From the 148 calibrated instances, we count the instances where examiners incline towards better clarity regions. The results of mean shift and EM clustering are broadly in agreement with the K-means clustering algorithm, with the majority of cases showing an inclination for better clarity regions.
- For all three, the ROI clusters have higher mean fixation duration compared to the average fixation duration of all the fixations. With the K-means algorithm, we had observed that the mean and median distances from ROI points were smaller. Mean shift and EM clustering algorithms also demonstrated similar results, showing the proximity of ROI points towards the singularities.
- K-means clustering is optimized with the Euclidean distance with respect to the center. Similarly, mean shift is optimized towards packing density around the cluster center. Therefore, K-means and mean shift clustering tends to create denser clusters. However, EM clustering provides relatively sparse ROI clusters compared to the other two algorithms. The primary reason is that EM clusters are modelled with a Gaussian distribution. It results in elliptical clusters instead of circular clusters, due to which the points can be sparsely distributed along the major axis.

6.2 Value Determinations and Minutia Count

During the ACE procedure, examiners are instructed to annotate their findings. These include the minutiae markup with their confidences, the value label (VID, VEO, and NV), and a quality score of the latent fingerprint on a scale of (0-100). While eye gaze is useful to analyze the regions where an examiner is looking at, these annotations help in understanding the sufficiency and suitability [1] of the latent fingerprint. In this section, the annotation data provided by the examiners is utilized to understand the relation between quality of the fingerprint, minutia count, and the conclusion decision. To do so, all the 158 samples of the annotated data are taken into consideration.

6.2.1 Repeatability and Reproducibility of Value Determination

The value determination of latent fingerprint plays a crucial part in determining the suitability of latent fingerprint for examination. An examiner labels the value of the latent fingerprint into one of the three categories: VID, VEO, or NV. Typically, latent fingerprint with VID and VEO labels are known to be “of value”. For the latents that are “of value”, examiners are shown latent fingerprint and exemplar fingerprint together for further comparison and evaluation [38]. However, for NV latent fingerprints, the examiners do not proceed further to comparison and evaluation phase and the fingerprints are kept secure in the case files.

Value determination plays an important role in illustrating the quality of a fingerprint. If a latent fingerprint is erroneously determined of “No Value”, the latent fingerprint might skip the comparison and evaluation stages. On the other hand, if a latent fingerprint of “No Value” is labeled as VID or VEO, the chances of an erroneous conclusion increases [39]. In our research, to study the consensus amongst examiners regarding latent fingerprint value determination, we calculate the repeatability and reproducibility of value determination. These values are highlighted in Table 4. Repeatability in value determination can be defined as a scenario where the same latent and exemplar fingerprint pair is shown again to the same examiner, to collect the value information. Repeatability checks the agreement of an examiner with his/her previous annotation [40], [41].

	Agreement			Disagreement		
	VID	VEO	NV	VID-VEO	VID-NV	VEO-NV
Repeatability (28 pairs)	20	4	0	2	2	0
	Total: 24			Total: 4		
Reproducibility (757 pairs)	498	7	59	114	47	32
	Total: 564			Total: 193		

TABLE 4: Repeatability & reproducibility of value determinations. The disagreement for value increases when two different examiners review the latent fingerprint.

In this study, there are a total of 28 instances of repeatability. The examiners stuck to their previous decision on 24 instances (85.71%). Of these 24 instances, 20 are labeled as VID while 4 are labeled as VEO. Fig. S.3(a)(supplementary) illustrates heat map of eye gaze where the examiner’s previous value determination agreed with his second attempt. On the other hand, the 4 cases where examiners did not repeat their value determination, 2 of them changed amongst VID and VEO while 2 of them changed amongst VID and NV. A change from VID to NV can impact casework hugely since a NV examination will not proceed whereas a VID may result in a conclusion decision.

Observing the pattern, it can be said that latent fingerprints with borderline quality are sometimes labeled as VID and sometimes VEO. However, saying the same for VID-NV does not hold true. When an examiner says that a latent fingerprint is of “No Value”, the latent fingerprint should lack distinctive information. However, the same examiner labeling the fingerprint as VID is highly unlikely. In both of these cases, examiner made an inconclusive conclusion. On observing the latent fingerprint, in one case the fingerprint is partial while the other latent fingerprint is developed using ninhydrin. Heat map of disagreement under repeatability conditions are shown in Fig. S.3(b)(supplementary). Though the latent fingerprints looked good in the first place to the examiner, the examiner ended up making an inconclusive decision. Despite performing ACE mechanism at a later time, the examiner might have remembered their conclusion decision from the last trial and thus changed the value of fingerprint to NV in the second trial.

Reproducibility in value determinations can be defined as a scenario where the same latent fingerprint is shown twice to 2 different examiners to provide its value determination. In our study, there are a total of 757 cases of reproducibility. Overall, reproducibility of 74.50% is observed (compared to 85.71% in repeatability), with examiner repeating value determination in 564 instances. The majority of consensus is in VID and NV with 498 and 59 cases

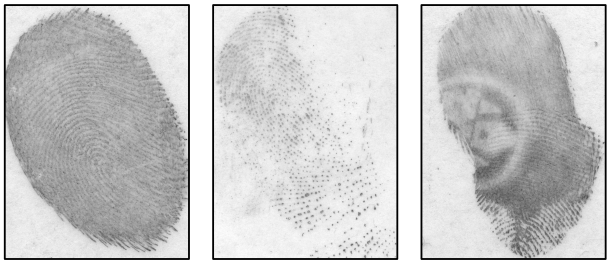


Fig. 9: Sample latent fingerprints where examiners failed to have consensus on value determination.

respectively. Fig. S.3(c)(supplementary) illustrates heat map of eye gaze of two different examiners where both of them agreed on the value determination of the latent fingerprint.

Table 4 highlights the individual scenarios of disagreement where an examiner labeled the value of the latent fingerprint differently compared to another examiner. While cases of disagreements do exist on borderline instances of fingerprint value such as VID-VEO or VEO-NV, there are also cases where external factors influence [26], [42]. The difficulty, as perceived by the examiner, in determining the ridge flow, the exact location of minutiae, and overall quality may vary from examiner to examiner. One such scenario is varying experience of the examiner. For instance, heat maps of eye gaze are shown in Fig. S.3(d)(supplementary). It has a scenario where an experienced examiner labeled as VID, while a novice labeled as NV. Overall, as shown in stimuli in Fig. S.3(supplementary) and Fig. 9, *we infer that the variations in value determinations arise on complex latents fingerprints*.

6.2.2 *Impact of Quality on Minutia Confidence*

For markup, examiners are instructed to color code the minutia markings to encode their level of confidence. These colored markups determine the region properties of the latent fingerprint. Additionally, colored markups also show how well examiners follow GYRO standards. To estimate impact of region quality on markup, we found the relation of color codes of markup with both region quality (as described by the quality map by mindtct) and region clarity (as defined by manual markup). It is achieved as follows:

1) For each marked minutia, a score of 0, 1, or 2 is assigned based on the color of minutia (red, yellow, and green respectively).
2) A 40×40 pixel region is defined around each marked minutia. For each region, average region quality is calculated by averaging the quality map score of twenty-five 8×8 blocks. The average region clarity is also calculated for each 40×40 region based on the manual ground truth annotation.
3) A correlation of minutia score (0, 1, or 2) and the average quality and clarity of the 40×40 region around the minutiae is found.

Using the above methodology, the Pearson correlation coefficient of 0.32 is observed between region quality and minutiae confidence. A similar correlation coefficient of 0.38 is found between the correlation of neighborhood clarity and minutia confidence. The above results indicate that not only region clarity and quality are interrelated, but there is a positive correlation between the quality of the neighborhood and minutiae confidence. *It means that as the neighborhood quality improves, the tendency to mark a green minutia increases and with unclear ridge valley details, examiners prefer to mark a red minutia*.

However, despite the intuitive results and positive correlation, the observed correlation coefficient is quite low. To understand the rationale, we study standard deviation of region clarity and quality. For areas around red minutiae, the average quality has standard deviation of 1.21 (mean=1.73) compared to 0.82 (mean=2.67) for regions around green minutiae. *It indicates that marking of red minutia is not directly dependent on region quality, as sometimes examiner mark red minutia in better quality areas as well (as shown by a higher standard deviation, and yellow and green clusters in Fig. 8). Thus, it can be concluded that examiners are not very particular about the markup of red minutiae*. Yet, their confidence is clearly reflected when they mark minutiae as green or yellow.

6.2.3 *Minutiae and Latent Fingerprint Value*

The results presented in the previous section showed us the impact of local clarity and quality with the minutiae confidence. That leads us to further analyze the impact of minutiae count and minutiae confidence on the overall value of the fingerprint. The value of the fingerprint is classified into three categories, namely: VID, VEO, and NV. Fig. S.4 (supplementary) shows a bar plot of all 158 samples, sorted based on the number of minutia marked. It illustrates the number of minutia, showing the red, yellow, and green confidence minutiae separately for each value label.

The value of the latent fingerprint is closely related to the latent fingerprint quality. We study the impact of the value, clarity, and quality of the latent fingerprint with minutia count. Similar to the score of 0, 1, or 2 assigned to minutiae based on the color, we assign 0, 1, or 2 to NV, VEO, and VID markup respectively. These values are assigned to each of 158 trials. It would result in a vector of size 158×1 for value determination (V). Similarly, vector of size 158×1 can be generated for quality map (Q_M) and clarity scores (C_S).

Next, Pearson correlation is found for value determination (V), quality map score (Q_M), and clarity score (C_S) with minutia count (X) and percentage (P) of green, yellow, and red minutiae. Each of these vectors is of length 158, and are shown in the equations below. Here, C_i, G_i, Y_i, and R_i corresponds to total minutia count, green minutia count, yellow minutia count, and red minutia count respectively for the i^{th} trial.

$$X_c = [C_1, C_2, C_3, ..., C_{158}] \tag{1}$$

$$X_g = [G_1, G_2, G_3, ..., G_{158}] \tag{2}$$

$$X_y = [Y_1, Y_2, Y_3, ..., Y_{158}] \tag{3}$$

$$X_r = [R_1, R_2, R_3, ..., R_{158}] \tag{4}$$

$$P_g = [G_1/C_1, G_2/C_2, G_3/C_3, ..., G_{158}/C_{158}] * 100 \tag{5}$$

$$P_y = [Y_1/C_1, Y_2/C_2, Y_3/C_3, ..., Y_{158}/C_{158}] * 100 \tag{6}$$

$$P_r = [R_1/C_1, R_2/C_2, R_3/C_3, ..., R_{158}/C_{158}] * 100 \tag{7}$$

The correlation results are shown in Table 5. It can be seen from the results that the *quality and clarity correlates*

R	(c) Total minutiae	(g) Green minutiae	(y) Yellow minutiae	(r) Red minutiae
$R(X,V)$	$R(X_c,V)$	$R(X_g,V)$	$R(X_y,V)$	$R(X_r,V)$
	0.46	0.50	-0.02	-0.23
$R(P,V)$	-	$R(P_g,V)$	$R(P_y,V)$	$R(P_r,V)$
	-	0.71	-0.24	-0.50
$R(X,Q_M)$	$R(X_c,Q_M)$	$R(X_g,Q_M)$	$R(X_y,Q_M)$	$R(X_r,Q_M)$
	0.61	0.60	0.05	-0.09
$R(P,Q_M)$	-	$R(P_g,Q_M)$	$R(P_y,Q_M)$	$R(P_r,Q_M)$
	-	0.56	-0.22	-0.36
$R(X,C_S)$	$R(X_c,C_S)$	$R(X_g,C_S)$	$R(X_y,C_S)$	$R(X_r,C_S)$
	0.66	0.65	0.07	-0.15
$R(P,C_S)$	-	$R(P_g,C_S)$	$R(P_y,C_S)$	$R(P_r,C_S)$
	-	0.51	-0.17	-0.36

TABLE 5: Pearson correlation (R) of minutia count (X) & percentage of minutia (P) with value determination (V), quality map score (Q_M), & clarity score (C_S). It shows that minutia count, quality, & ridge-valley clarity are all directly proportional to each other.

		Value Label		
		VID	VEO	NV
Conclusion Decision	Individualization	39	0	0
	Exclusion	63	12	4
	Inconclusive	12	9	19

TABLE 6: Value determination with respect to the conclusion decision for all the 158 trials by 29 participants.

positively with the minutia count, green minutia count, and the percentage of green minutia marked. It indicates that as the quality of latent fingerprint improves, the tendency to mark minutia and their confidence increases. Also, the total minutia count, green minutia count, and the percentage of green minutia show a positive Pearson correlation coefficient with the quality and value of the fingerprint. Regarding the red minutia, it is observed that red minutia count is not a clear indicator of the overall latent fingerprint clarity, quality, and value since a strong correlation is not observed (-0.15, -0.09, and -0.23 respectively). However, the percentage of red minutia is a good indicator of the value of the fingerprint, since it signifies that the value and proportion of red minutia are negatively correlated (-0.50). These results are notable since red minutia (least confidence) is not strongly correlated with local or global quality and clarity. However, a significant negative correlation of the presence of red minutia is found with the value of the latent fingerprint.

In the Database section, we mentioned that the data collection included a quantitative score for value assessment between 0 and 100. The quality score is capable of translating the value contained in the latent fingerprint as per the examiner. Fig. 10a shows a scatter plot and distribution on how expert and novice examiners determine the value score, along with the minutia count on Y-axis. The illustration is for latent fingerprint during the analysis stage. The scatter plot is segregated based on the examiner's experience. Fig. 10a indicates that for expert examiners, the value score is higher compared to novice examiners. It can be inferred that *experience accounts for better value score by the examiners, since expert examiners may be able to perceive the ridge detail precisely*.

6.2.4 *Value Determination (VID, VEO, or NV) vs Actual Decision (Individualization, Exclusion, or Inconclusive)*

During comparison, examiners decide if the latent and exemplar fingerprint are from the same source or not. Comparing the value determination with the conclusion can reveal insight into the behavior of the examiner. It has a potential to tell if an examiner is risk-averse or too ambitious while taking a conclusion decision. For 158 trials, all the scenarios are tabulated in Table 6.

Comparing value label and actual conclusion, examiners remained consistent on 44.30% occasions. Since the exemplar fingerprint is not shown in the analysis stage, the examiner might decide to opt for an exclusion conclusion despite saying VID in the analysis stage. In such a scenario, examiner feels that latent fingerprint has value for identification (VID). However, on observing exemplar and performing a comparison, the examiner might conclude that the latent and exemplar fingerprints are not from the same source. Instances with conclusion decision contradicting the previously marked value determination need to be studied in detail. From Table 6, it can be seen that there are 12 cases where examiners made an inconclusive conclusion despite labeling latent fingerprint as VID. There are additional nine instances where examiners made an inconclusive conclusion after labeling the latent fingerprint as VEO in the analysis stage. *Both of these scenarios indicate the risk-averse nature of examiners while taking a decision*. While examiners initially thought that fingerprint had value, they faced difficulties in comparing the dotted Ninhydrin or partial latent fingerprints with exemplar fingerprints in the comparison stage. Thus, they resorted to a risk-averse decision of inconclusive. These latent fingerprints are shown in Fig. 9.

Another probable reason could be the lack of corresponding areas or poor quality exemplar. However, a significant overlap in the fingerprint ridge information is present between the exemplar and the latent fingerprints on the majority of these instances. Also, as shown with NFIQ and NFIQ 2.0 scores in Table 7, exemplars are of above-average quality. Amply marked minutiae in exemplar fingerprint by the examiners also re-assures its quality. Thus, in most of the above mentioned 21 cases of our study, we observe that the examiners are risk-averse. Detailed analysis on examiners' inconclusive conclusion despite VID or VEO is shown in Section 3 of the supplementary material.

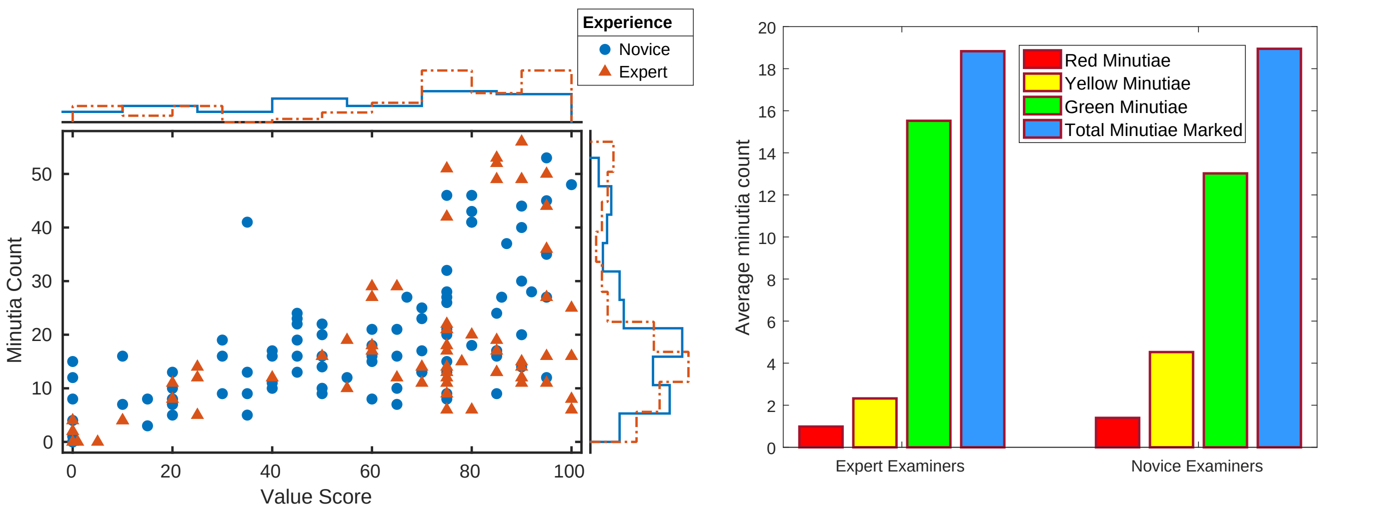
There are also four instances where examiners took an exclusion decision in spite of saying that the latent fingerprint is of No Value (NV). Making a conclusion in such a scenario might be erroneous, given the latent fingerprint is No Value. On evaluating these four exclusion conclusion with the ground truth, it is observed that all four cases were an erroneous exclusion. Though in real situation, examiners do not proceed with the comparison of the fingerprint if the fingerprint is of "No Value", it is to be understood that examiners should not be ambitious in taking a conclusion decision if the latent fingerprint lacks clarity in ridge valley.

6.2.5 *Time Spent per Minutia*

As highlighted in the gaze statistics, examiners spend most of their time in analysis stage of the ACE mechanism. Examiners tend to find distinctive details in the latent fingerprint which could assist in the later stages. Thus, the time spent per minutia in the analysis phase could potentially give an insight of the value of the latent fingerprint and the conclusion decision of the examiner. To calculate the average time spent in minutiae markup, we consider trials where at

	No. of Instances	NFIQ score for exemplar fingerprint (1-5; 1 best)	NFIQ 2.0 score for exemplar fingerprint (1-100; 100 best)	Latent minutiae count (manually marked)	Exemplar minutiae count (manually marked)
VID $\rightarrow$ Inconclusive	12	2.08	62.42	17.17	29.00
VEO $\rightarrow$ Inconclusive	9	2.44	55.11	9.78	23.44

TABLE 7: Minutiae count and exemplar fingerprint quality for instances where examiners concluded as Inconclusive despite labelling latent fingerprint as VID or VEO.



(a) Relation of value scores and minutiae count with experience **(b)** Bar chart showing effect of experience on minutia count & confidence

Fig. 10: Understanding the impact of experience on value & minutiae counts. We see that expert examiners are confident to mark green minutiae and accounts higher value score. On the contrary, novice examiners are uncertain and mark more yellow minutiae.

least one minutia has been marked in the latent fingerprint. Thus, we consider 153 samples out of a total of 158 samples.

Based on the overall quality of the latent fingerprint, an examiner defines the value as either VID, VEO, or NV. In the experiments, the average time spent in minutiae markup is lowest for VID (7.98 seconds), highest for NV (21.95 seconds), and intermediate for VEO (13.32 seconds). *These results indicate that examiners spent more time searching for features in the lower quality fingerprints*.

6.2.6 *Relation of Minutia Count and Confidences with Conclusions*

Erroneous conclusions, especially the false positives, have been a highlight in recent studies. Studies discuss and evaluate false positive rates (FPR) incurred in latent to exemplar fingerprint comparisons [40], [43]. On similar lines, we assess the relationship between minutia count and minutia confidence with conclusions. Minutia counts and their confidences might help in checking the reliability of a conclusion decision. Thus, we study relations between minutia counts and confidences with the correct decisions. Out of a total of 158 samples, examiners made 96 correct conclusions (39 individualizations and 57 exclusions). Another 40 instances are inconclusive and there are 22 erroneous conclusions. Thus, examiners achieved an accuracy of 81.46% (inconclusive conclusions excluded).

On average, examiners marked approximately 22 minutia in correct conclusions. However, while making a correct individualization conclusion, a higher average count of 30 is observed. Similarly, during correct exclusion, average

Conclusion Decision	**Ground-truth**			
	Experts		Novice	
	Mated	Non Mated	Mated	Non Mated
Individualization	20	0	19	0
Exclusion	*5*	25	*17*	32
Inconclusive	7	8	9	16
Total	65		93	

TABLE 8: Conclusion decision for expert and novice examiners with respect to the ground-truth.

minutia count is 17. Consequently, it can be said that for making an individualization conclusion, examiners require higher confidence (based on the number of minutia) to make the conclusion. Still, the lowest minutia count in correct exclusion conclusion is 0 and true individualization conclusion is 6. The 0 count is observed in a scenario where pattern type of latent is different from the exemplar fingerprint.

In case of incorrect conclusions, there is a lower average minutia count of 19. However, the least minutia count for an erroneous exclusion is three, which is higher than the lowest minutia count of correct exclusion. In the false conclusions, there are no incorrect individualizations. *It shows that examiners are risk-averse and need to be confident to make an individualization conclusion*.

6.2.7 *Effect of Examiner's Experience*

In 2016, Ulery et al. [4], [44] highlighted that difference in minutiae markup exists among latent fingerprint examiners. They established that factors such as region clarity and lack of standardization in minutia markup contribute towards inter-examiner variations. In this study, we evaluate the effect of experience of examiners. As highlighted earlier, the participants of our study had a diverse amount of experience. The expertise of examiners ranged from 6 months to more than 40 years. Each examiner is labeled as novice or expert based on experience. An examiner with less than two years of expertise is termed as a novice, while, examiners with more than two years are called expert examiners. Using markup data, we perform experiments that analyze the difference in minutia counts. We also use conclusion decisions made by expert and novice examiners to understand their success rate of correct comparison.

The first set of inferences incur from minutia counts. Fig. 10b shows the average number of minutia marked by expert and novice examiners. These minutiae marked are segregated as red (least confidence), yellow (intermediate confidence), and green (highly confident) minutiae as well. It can be observed that expert and novice examiners mark a similar number of minutia, averaging to 18.83 and 18.95 respectively ($p = 0.9578$). While the total number of marked minutia for latent fingerprints are comparable, the difference arises in confidences. Due to the lack of expertise, novice examiners are uncertain during minutiae markup. Hence, compared to expert examiners, they mark more yellow minutiae ($p = 7.1 \times 10^{-5}$). Though, expert examiners are confident enough to annotate a higher proportion of green minutiae, the improvement is not significant ($p = 0.25$). Similarly, the markup of low confidence minutia is similar for both expert and novice examiners ($p = 0.21$).

Table 8 shows the conclusion decision for expert and novice examiners with respect to the ground-truth. The mated pairs arise from the same fingerprints of the individual, while non-mated fingerprints arise from different fingers. Due to the risk-averse nature of examiners, examiners make an individualization conclusion only when they are confident enough. Hence, there is no erroneous individualization. However, the role of expertise can be observed with incorrect exclusion. Proportionally, novices make higher erroneous exclusion compared to expert examiners. Also, novice examiners make a slightly higher amount of inconclusive conclusion. Overall, from the results, *we can conclude that experience plays a pivotal role during latent to exemplar fingerprint comparison*. Experience helps the examiners to make minutiae markup confidently, interpret details in latent fingerprint precisely with fewer inconclusive and erroneous conclusions.

7 Conclusion

Latent fingerprint comparison involves the intervention of forensic examiners for markup of features and verification of results. Using their knowledge, examiners determine the value of the latent fingerprint and take an individualization, exclusion, or inconclusive decision. However, the process of perceiving details from latent and exemplar fingerprint is unknown. This study collects and studies the eye gaze patterns of examiners while they mark features and compare a latent with an exemplar fingerprint using ACE procedure. The eye gaze pattern of a forensic examiner provides insights into the process and employed heuristics.

In this work, we empirically find the Region of Interest (ROI) using the eye gaze patterns of examiners. The ROI indicates essential aspects and search strategies, such as examiners having a preference of region near singularities. Gaze based results are accompanied by assessing minutiae markup and value determinations of latent fingerprints and its relevance with the local clarity. Some of the key findings of the research are: (i) examiners are risk-averse in minutiae markup during the analysis phase, (ii) examiners tend to trace ridges in low-quality regions, (iii) examiners are careful enough to never make an incorrect individualization, however, they do make an incorrect exclusion, and (iv) the experience plays an essential role during the ACE-V procedure, resulting in inter-examiner variability in markup confidences and erroneous conclusions (FP). The results highlight the patterns observed across different forensic examiners. The inferences of this study could be inculcated back into an AFIS system to improve automated comparison and help train novice examiners.

Acknowledgments

Authors thank K. Ayers, R. Wood, C. Venter, and the participants for making the data collection possible. A. Malhotra is partially supported by the Visvesvaraya Ph.D. Scheme and ORF. The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of the U.S. Govt. or the Govt. of India.

References

[1] " Standard Terminology of Friction Ridge Examination (Latent/Tenprint)," http://www.clpex.com/swgfast/documents/terminology/121124_Standard-Terminology_4.0.pdf, 2012, [Online; 09-Aug-2019].

[2] "Standard for the Documentation of Analysis, Comparison, Evaluation, and Verification (ACE-V) (Latent)," http://clpex.com/swgfast/documents/documentation/121124_Standard-Documentation-ACE-V-tenprint_2.0.pdf, 2012, [Online; 10-Aug-2019].

[3] T. Busey, C. Yu, D. Wyatte, J. Vanderkolk, F. Parada, and R. Akavipat, "Consistency and Variability Among Latent Print Examiners as Revealed by Eye Tracking Methodologies," *Journal of Forensic Identification*, vol. 61, no. 1, pp. 60–91, 2011.

[4] B. T. Ulery, R. A. Hicklin, M. A. Roberts, and J. Buscaglia, "Interexaminer variation of minutia markup on latent fingerprints," *Forensic Science International*, vol. 264, pp. 89–99, 2016.

[5] M. Snodgrass, E. Bernat, and H. Shevrin, "Unconscious perception at the objective detection threshold exists," *Perception & Psychophysics*, vol. 66, no. 5, pp. 888–895, 2004.

[6] M. Van Selst and P. M. Merikle, "Perception below the Objective Threshold?" *Consciousness and Cognition*, vol. 2, no. 3, pp. 194–203, 1993.

[7] "DEPARTMENT OF JUSTICE APPROVED UNIFORM LANGUAGE FOR TESTIMONY AND REPORTS FOR THE FORENSIC LATENT PRINT DISCIPLINE," https://www.justice.gov/file/1037171/download, 2016, [Online; 10-Aug-2019].

[8] V. Cantoni, M. Musci, N. Nugrahaningsih, and M. Porta, "Gaze-based biometrics: An introduction to forensic applications," *Pattern Recognition Letters*, vol. 113, pp. 54–57, 2018.

[9] A. Sankaran, M. Vatsa, and R. Singh, "Latent Fingerprint Matching: A Survey," *IEEE Access*, vol. 2, pp. 982–1004, 2014.

[10] A. Malhotra, A. Sankaran, M. Vatsa, and R. Singh, "Learning Representations for Unconstrained Fingerprint Recognition," *Deep Learning in Biometrics*, pp. 197–221, 2018.

[11] T. Busey, H. J. Swofford, J. Vanderkolk, and B. Emerick, "The impact of fatigue on latent print examinations as revealed by behavioral and eye gaze testing," *Forensic Science International*, vol. 251, pp. 202–208, 2015.

[12] B. Roads, M. C. Mozer, and T. A. Busey, "Using Highlighting to Train Attentional Expertise," *PloS one*, vol. 11, no. 1, 2016.

[13] National Research Council, "Committee on Identifying the Needs of the Forensic Science Community," *Strengthening Forensic Science in the United States: A Path Forward*, 2009.

[14] C. Yu, T. Busey, and J. Vanderkolk, "Discovering Correspondences between Fingerprints based on the Temporal Dynamics of Eye Movements from Experts," in *IAPR International Workshop on Computational Forensics*, 2010, pp. 160–172.

[15] T. Busey, C. Yu, D. Wyatte, and J. Vanderkolk, "Temporal Sequences Quantify the Contributions of Individual Fixations in Complex Perceptual Matching Tasks," *Cognitive Science*, vol. 37, no. 4, pp. 731–756, 2013.

[16] B. T. Ulery, R. A. Hicklin, M. A. Roberts, and J. Buscaglia, "Measuring What Latent Fingerprint Examiners Consider Sufficient Information for Individualization Determinations," *PloS one*, vol. 9, no. 11, p. e110179, 2014.

[17] "Ansi/nist standard," https://www.nist.gov/programs-projects/ansinist-itl-standard/, 2018, [Online; 10-Aug-2019].

[18] "Document #103 individualization/identification position statement (latent/tenprint)," http://clpex.com/swgfast/Comments-Positions/130106-Individualization\-ID-Position-Statement.pdf, 2013, [Online; 08-Sep-2019].

[19] "Standards for Examining Friction Ridge Impressions and Resulting Conclusions," http://clpex.com/swgfast/documents/examinations-conclusions/130427_Examinations-Conclusions_2.0.pdf, 2013, [Online; 10-Aug-2019].

[20] A. Hicklin and C. Reedy, "Implications of the IDENT/IAFIS Image Quality Study for Visa Fingerprint Processing," *Mitretek Systems*, 2002.

[21] C. W. Elham Tabassi and C. I. Watson, "Fingerprint Image Quality," https://www.nist.gov/publications/fingerprint-image-qualitiy, 2017, [Online; 10-Aug-2019].

[22] B. Schiffer and C. Champod, "The potential (negative) influence of observational biases at the analysis stage of fingermark individualisation," *Forensic Science International*, vol. 167, no. 2-3, pp. 116–120, 2007.

[23] S. V. Stevenage and C. Pitfield, "Fact or friction: Examination of the transparency, reliability and sufficiency of the ACE-V method of fingerprint analysis," *Forensic Science International*, vol. 267, pp. 145–156, 2016.

[24] A. Rairden, B. L. Garrett, S. Kelley, D. Murrie, and A. Castillo, "Resolving latent conflict: What happens when latent print examiners enter the cage?" *Forensic Science International*, vol. 289, pp. 215–222, 2018.

[25] B. Ulery, R. Hicklin, M. Roberts, and J. Buscaglia, "Changes in Latent Fingerprint Examiners' Markup Between Analysis & Comparison," *Forensic Science International*, vol. 247, pp. 54–61, 2015.

[26] B. T. Ulery, R. A. Hicklin, M. A. Roberts, and J. Buscaglia, "Factors associated with latent fingerprint exclusion determinations," *Forensic Science International*, vol. 275, pp. 65–75, 2017.

[27] M. D. Vogelsang, T. J. Palmeri, and T. A. Busey, "Holistic processing of fingerprints by expert forensic examiners," *Cognitive Research: Principles and Implications*, vol. 2, no. 1, p. 15, 2017.

[28] I. E. Dror and G. Langenburg, ""Cannot Decide": The Fine Line Between Appropriate Inconclusive Determinations Versus Unjustifiably Deciding Not To Decide," *Journal of Forensic Sciences*, vol. 64, no. 1, pp. 10–15, 2019.

[29] S. Shapoori and N. Allinson, "GA-Neural Approach for Latent Finger Print Matching," in *IEEE International Conference on Intelligent Systems, Modelling and Simulation*, 2011, pp. 49–52.

[30] Y. Rubner, C. Tomasi, and L. J. Guibas, "The Earth Mover's Distance as a Metric for Image Retrieval," *International Journal of Computer Vision*, vol. 40, no. 2, pp. 99–121, 2000.

[31] R. A. Hicklin, B. T. Ulery, T. A. Busey, M. A. Roberts, and J. Buscaglia, "Gaze behavior and cognitive states during fingerprint target group localization," *Cognitive Research: Principles and Implications*, vol. 4, no. 1, p. 12, 2019.

[32] G. Langenburg and C. Champod, "The GYRO system-a recommended approach to more transparent documentation," *Journal of Forensic Identification*, vol. 61, no. 4, p. 373, 2011.

[33] FBI, "Biometric Collection of People (BioCoP) Database," https://biic.wvu.edu/, 2008-2013, [Online].

[34] "GP3 Eye Tracker," https://www.gazept.com/product/gazepoint-gp3-eye-tracker/, [Online; 10-July-2020].

[35] J. A. Hartigan and M. A. Wong, "Algorithm AS 136: A K-Means Clustering Algorithm," *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, vol. 28, no. 1, pp. 100–108, 1979.

[36] " NIST Biometric Image Software," https://www.nist.gov/services-resources/software/nist-biometric-image-software-nbis, 2016, [Online; 09-Aug-2019].

[37] R. A. Hicklin, J. Buscaglia, and M. A. Roberts, "Assessing the clarity of friction ridge impressions," *Forensic Science International*, vol. 226, no. 1-3, pp. 106–117, 2013.

[38] "Statistical Friction Ridge Analysis (SFRA)," https://www.nist.gov/programs-projects/statistical-friction-ridge-analysis-sfra, 2014, [Online; 1-Aug-2019].

[39] S. Gutowski, "Error rates in fingerprint examination: The view in 2006," *Forensic Bulletin*, vol. 2006, pp. 18–19, 2006.

[40] B. T. Ulery, R. A. Hicklin, J. Buscaglia, and M. A. Roberts, "Accuracy and reliability of forensic latent fingerprint decisions," *Proceedings of the National Academy of Sciences*, vol. 108, no. 19, pp. 7733–7738, 2011.

[41] B. T. Ulery, R. A. Hicklin, J. Buscaglia, and M. A. Roberts, "Repeatability and Reproducibility of Decisions by Latent Fingerprint Examiners," *PloS one*, vol. 7, no. 3, p. e32800, 2012.

[42] P. A. Fraser-Mackenzie, I. E. Dror, and K. Wertheim, "Cognitive and contextual influences in determination of latent fingerprint suitability for identification judgments," *Science and Justice*, vol. 53, no. 2, pp. 144–153, 2013.

[43] I. Pacheco, B. Cerchiai, and S. Stoiloff, "Miami-Dade Research Study for the Reliability of the ACE-V Process: Accuracy & Precision in Latent Fingerprint Examinations," *National Institute of Justice*, 2014.

[44] B. T. Ulery, R. A. Hicklin, M. A. Roberts, and J. Buscaglia, "Data on the interexaminer variation of minutia markup on latent fingerprints," *Data in Brief*, vol. 8, pp. 158–190, 2016.

A. Malhotra received his B.Tech. degree (2015) in CSE from IIIT-Delhi and currently, he is pursuing his doctoral degree at the IIIT-Delhi, India.

A. Sankaran received the B.Tech. degree (2010) in CSE from CIT, India and Ph.D. degree (2017) at IIIT-Delhi, India. He worked as a Research Scientist at IBM Research, India and is currently working as Senior Research Scientist at Deeplite, Canada.

M. Vatsa received the M.S. (2005) & Ph.D. (2008) degrees from WVU, USA. He is currently a Professor at IIT Jodhpur, India, and the Project Director of the TIH on Computer Vision and Augmented Reality. He is also an Adjunct Professor with IIIT-Delhi and WVU, USA.

R. Singh received the M.S. (2005) & Ph.D. (2008) degree in CSE from WVU, USA. She is currently a Professor at IIT Jodhpur, India, and an Adjunct Professor with IIIT-Delhi and WVU, USA.

K. B. Morris received B.Sc. (1985), B.Sc. (Hons) (1986), & Ph.D. (1990) in Chemistry from Univ. of Port Elizabeth. He is currently a Professor in the Dept. of Forensic & Investigative Science at WVU, USA.

A. Noore received the Ph.D. degree in ECE from WVU, USA. He currently serves as a Professor and Associate Dean in the Frank H. Dotterweich College of Engineering at Texas A&M University-Kingsville.