Document Title: Adding Human Expertise to the Quantitative Analysis of Fingerprints, Final Report

Author: Thomas Busey, Chen Yu

Document No.: 230166

Date Received: April 2010

Award Number: 2005-MU-BX-K076

Final Report

NIJ grant #2005-MU-BX-K076

Thomas Busey and Chen Yu, Indiana University

Abstract

The goal of this project is to characterize the nature of human expertise using eye tracking methodologies, and then use these results to develop and refine quantitative metrics of the information contained in friction ridge patterns. Current quantitative approaches to fingerprint matching and analysis are not based on human data and therefore do not take advantage of the full capabilities of the human visual system. Since humans routinely outperform automated fingerprint recognition systems, it is clear that quantitative approaches can be improved by adopting some of the strategies that humans employ. However, humans often have difficulty describing the results of perceptual processing, and may not even know what information they are using. To address this deficit, we used eye tracking to identify what information human experts rely on. We constructed a portable eye tracking system that enables us to collect data from experts and novices while they perform tasks similar to latent print examinations. Once we analyzed the data we obtain a record of the regions visited by the experts as they compared pairs of fingerprints. We then developed a series of computational analyses to identify the nature of the expertise. This took the form of data reduction procedures on pixel crops from the fingerprint images, as well as the development of candidate information metrics that the data from experts helps validate. The results demonstrate clearly that human expertise can be inferred from eye gaze information through a process of carefully designed studies and hypothesis testing of candidate information metrics. Because our candidate metrics take the form of mathematical and computational models, they are readily applicable to machine comparison approaches, and also can be used to identify the diagnosticity and rarity of particular features in novel prints.

Table of Contents

Executive Summary

This Executive Summary is designed to provide an overview of the methods, findings and conclusions of the project. While it provides a summary of the results, many of the mathematical and computational approaches that are described require additional explication in order to gain a complete intuition for the methods. The main body of the report provides such detail, along with supporting figures that demonstrate the conclusions in graphical form. Thus the reader is encouraged to refer to the full text when necessary.

Research Problem

Human expertise can be remarkably difficult to transfer from one individual to another, in part because the tools we have to transfer this knowledge, mainly language, are sparse and incomplete representations of complex perceptual skills. Indeed, psychologists differentiate between knowledge learning and skill learning, as they appear to have different mechanisms supported by different brain substrates. This should come as no surprise to anyone who has ever coached a little league team, taught piano, or mentored a latent print trainee. Describing what we do is very different than actually being able to do it.

The difficulty we have in translating perceptual skills from one person to another also implies that the skill that underlies latent print examinations may be difficult to demonstrate. An examiner can point to a region to indicate to another person what matches, but they can make the person ‘see’ the match in a visual comprehending sense. The MagicEye posters popular a few years back are a good example of how perception is a uniquely personal and private experience: Knowing that a dinosaur is hidden in the image doesn’t help until all of a sudden you can see it completely.

This kind of perceptual experience can be difficult to transfer to others, and this may be why judges have been willing to listen to defense attorneys who have argued that latent print examiners do not have any special expertise, and that fingerprints should be just shown to the jury. At issue here is the fact that much of the process of perception can be difficult to translate into language (Snodgrass, Bernat, & Shevrin, 2004; Vanselst & Merikle, 1993) and examinations may be subject to extra-examination biases (Dror & Charlton, 2006; Dror, Charlton, & Peron, 2006; Dror, Peron, Hind, & Charlton, 2005). As a result, up until now we have been forced to trust the word of the examiner that they rely on the features they say they do.

In this project we have used eye tracking methodologies, combined with computer and mathematical modeling approaches, to infer the nature of the expertise in latent print examiners. This allows us to address questions not only about what information experts rely on, but to create candidate models of this expertise using mathematical and computational approaches that can be validated against the human data. No longer do we need to simply ask examiners what they rely on, which is important because much of perception may reside below the level of consciousness. For example, the picture in Figure 1 illustrates that even though we know at a cognitive level that the two lines are straight, this knowledge does not allow us to perceive them as straight.

The goal, then, of this project is to use eye gaze to identify the features and ridge detail information that experts rely on when conducting latent print examinations. This task is made possible by the fact that the eyes contain a region of high acuity, called the fovea, outside of which the representation is much coarser. Because ridge detail is fairly fine, most information must be acquired by moving the fovea so that it lands on the region of interest. Thus by tracking the position of the eyes we gain a record of what features or ridge detail the examiner considers diagnostic or relevant. Such an approach has been used in question document applications (Dyer, Found, & Rogers, 2008).

This task is complicated somewhat by the fact that the fovea is about 1 degree of visual angle, or about the width of your thumbnail seen with your arm outstretched (go ahead, stick your arm out and see how big this is). This area implies that, depending on the enlargement of the print, it could encompass several different ridges. To gain a final representation of what examiners are evaluating within this window of high acuity, we construct candidate mathematical models that we can test between by comparing data from experts with data from novices. The benefit of these approaches is that we are able to characterize the nature of expertise in computational form that is easy to implement in computer-based matching systems such as AFIS. As part of the development of these models we will need to describe some initial differences between experts and novices that will guide our model development. Note, however, that systems such as AFIS and human examiners have different goals. AFIS is designed to quickly return candidate matching prints from large stored databases. However, human examiners have more information available, since they can use features such as third level detail. In some sense it is desirable to have the two systems use different sources information, because independent sources give more reliability to the final conclusion. Of course, there is a potential built-in correlation if the expert only looks at prints returned by AFIS.

Research Designs and Findings

There are three major research questions in this project:

1) How reliably can eye gaze data be collected from experts?

2) Does this eye gaze data demonstrate differences between experts and novices? If so, what is the nature of the differences?

3) How can data from experts discriminate between different candidate quantitative metrics of the information in latent fingerprints?

These questions take multiple approaches to answer, and as a result we have combined the Research Designs and Findings sections into one, and broken down the individual data analyses into subsections below.

Eye Tracking Data Processing

To gather robust gaze data from examiners in the field, we designed a set of software tools that would allow us to record the video streams from two head-mounted cameras, one that monitored the position of the eye relative to the head, and one that monitored the head relative to the world. By performing calibration procedures and tracking the corners of the computer monitor in the scene video we were able to determine the position of the eye on the fingerprint images to an average error of less than 1 degree of visual angle. Note that this is on par with commercial eye tracking systems and is about the width of the fovea. Therefore, accuracy greater than this resolution probably is diminishing returns.

We developed a number of innovations in order to reliably capture gaze information from videos taken in the field, and our eye model innovation is demonstrated in Figure 4. We had to deal with all kinds of difficulties, such as one recording session when the chandeliers from the hotel lobby were reflected in the pupil of our subjects. Because we can post-process our video we are able to recover much of the data that would be lost if we had tried to do online eye tracking.

We currently have data from 72 experts, 10 trainees, and 32 novices, which we have processed through our software. We ask our participants to perform abbreviated latent print examinations that allow 20-30 seconds per pair of images. This allows us to limit the recording sessions to 15 minutes to avoid eye fatigue, and to obtain eye gaze data from a wide variety of images, which is necessary for the data reduction procedures. This relative brief exposure requires the experts to move their gaze to regions that they consider to be most diagnostic.

Figure 8 illustrates the eye gaze data collected from one expert, along with the fixations that are found by grouping the raw gaze data into clusters by computing the velocity of the track. It is this kind of data that provides the foundation for the analyses described in the next section.

Eye Tracking Data Analyses

We have three major types of experiments that we conduct. The first set uses standard latent/inked comparisons from NIST special database 27. We also obtained a set of clean prints from donors at Indiana University, and conducted a set of studies using clean images. More recently we have begun to add texture noise to clean prints to encourage examiners to treat them as latent prints. In all cases we ask our participants to determine whether the two prints come from the same source, although we give a ‘too soon to tell’ option as well. Based on these and the fact that we know the ground truth allows us to compute accuracy. In all cases we find that experts outperform novices in terms of overall measures of sensitivity.

In the summaries below we characterize the nature of the individual analyses and provide a summary of the results. We do not provide the raw statistics for each conclusion, but refer the reader to the full text for such detail.

Fixation statistics

As a first step, we compared our experts and novices in terms of how often and how far them moved their eyes. We found no differences between the two groups in terms of the amount of time they spend in one location, but when experts move their eyes, then tend to make shorter jumps, at least within one print.

This behavior is consistent with the idea that the information used by experts is not individual features to be matched across the prints, but collections of features. Since collections are much less likely to be spuriously matched, they provide much more diagnostic value. Of course the expert must be able to maintain the entire collection in visual memory, and other data from our lab suggests that they have longer-lasting visual memories (Busey & Vanderkolk, 2005).

Consistency among experts and novice

The gaze data provides collections of fixations for a given pair of prints for several different examiners. We can ask whether experts use similar kinds of information by looking to see whether they visit the same locations on the prints. We used a metric called the Earth Mover Distance, which computes the similarity between any two experts based on how easy it would be to move the fixations from one expert onto the fixations of another.

The Earth Mover metric provided some surprising results. When applied to latent/inked pairs, we found that experts were *less* consistent as a group. This suggests that they tended to use different sources of information, perhaps by wandering off the high-visibility regions of the prints into different parts of the noisy areas. The choice of which noisy areas to visit may have depended on their individual expertise and strengths. It will be up to the latent print community to decide whether this variability is a source of concern.

We found the opposite result for clean image pairs. Now experts are showing more consistency as a group, despite the overwhelming wealth of ridge detail available in clean prints. This is consistent with the suggestion that experts have an implicit sense as a community for which features are most diagnostic and tend to fixate them immediately.

Use of minutiae by experts and novices

One reason that experts produced greater consistency as a group with clean prints is that they tended to focus on the core and delta regions, which are rich in minutiae. Figure 13 shows the locations of the fixations for experts and novices for a clean image pair, and the experts tend to send their gaze to regions that have much higher minutiae count. We did not have enough data to look at individual differences between examiners who profess to ignore minutiae and those who explicitly rely on minutiae, but such an analysis would be potentially revealing about the nature of individual differences among examiners.

Machine translation approaches to data analysis

In a separate project we addressed the ability of temporal information to determine whether two features match across an image pair. There have been reports in the eye tracking literature that gaze information may be more diagnostic than over responding, since subjects will often fixate a target one or more times before actually identifying it in a visual search task.

We used algorithms adapted from machine translation approaches to identify corresponding regions in the two prints. Using the data from multiple experts, we found that the two images could be treated as two separate languages, and the pattern of eye gaze switches between the two prints acts as a translation dictionary. This approach builds up correspondences between regions in the prints, and the algorithm was much more successful in finding correspondences in data from experts than data from novices.

This analysis suggests that gaze data alone may be sufficient to determine corresponding regions, and that there may be more information in the eye gaze channel than initially available from the behavioral responses of the examiner.

Feature induction from fixation data

In order to determine the exact nature of the information acquired by experts, we cropped out small, 48x48 pixel patches centered on each fixation. We then performed a data reduction procedure known as Independent Component Analysis, which seeks to find a common alphabet or feature set that best describes the entire dataset. We had over 30,000 image crops from experts, and we distilled these down to just 120 basis functions that, when combined, allow a fairly accurate reconstruction of the original image crops.

This analysis demonstrated fundamental differences in the nature of the information acquired by experts and novices. While novices appeared to focus on fine details in high-signal regions, the experts were more likely to seek out regions that had more visual noise and lower spatial frequency information.

Information theoretic approaches

Our most sophisticated computational approaches rely on metrics derived from Shannon Information Theory (Shannon & Weaver, 1949). We trace the ridge elements and determine the orientation of the ridge at each pixel location. This allows us to collect these orientations, as well as the relative orientation change, which provides ridge curvature information. Figure 19 illustrates this procedure.

The collection of orientations provides a measure of the *entropy* of the patch, and we can construct several different measures of entropy depending on the assumptions made about the frequency distribution.

We computed the entropy of patches extracted at the fixations of experts and novices, and determined that some measures of entropy distinguished between experts and random fixations, while others distinguished between experts and novices. This approach appears promising in part because entropy can be based on sources of information than ridge orientation.

We were surprised to find that some measures of entropy that we though were extremely viable turned out not to discriminate between experts and novices, and others that we thought would not mean much turned out to be very diagnostic. This observation illustrates the value that the eye gaze data from experts and novices can provide in the model generation and testing process.

Conclusions

We had three main questions that we addressed in this project, which we can now answer.

1) How reliably can eye gaze data be collected from experts?

The accuracy of our system, computed by comparing the estimated eye gaze position against where we asked the participants to look, demonstrated that our system performs at the same level of accuracy as commercial systems. In addition, our post-processing procedures and innovative forward eye model allow us to deal with field data collection that would prove problematic for commercial systems. The fact that we built the system ourselves based on research articles describing different calibration procedures means that we understand and trust all aspects of our data. Thus the answer to this question is a definitive yes, and our results demonstrate that eye gaze data can provide strong evidence for the nature of expertise in examiners.

2) Does this eye gaze data demonstrate differences between experts and novices that would support the continued use of human experts as expert witnesses is court cases? If so, what is the nature of the differences?

We found that experts out-performed novices in all experiments, both in terms of the number of correct identifications and the number of correct exclusions. Experts were in general less likely to use the 'too soon to tell' option.

This behavioral accuracy improvement is supported by a number of differences in the eye gaze data. Experts had shorter saccades between fixations, suggesting that they were looking for clusters of features rather than trying to match individual features. The combinatorial increase in diagnosticity that multiple features in combination provides makes this strategy potentially much more accurate.

Experts and novices differed in their consistency as groups depending on the nature of the stimuli. For latent prints, experts showed *less* consistency overall than the novices, and we speculated that they were venturing into the less visible print regions and choosing different regions to visit depending on their own individual strengths. Novices tended to just focus on the clean portions of the latent prints regardless of whether there was diagnostic information there or not. For the clean prints, however, experts showed more consistency overall despite the fact that there was a wealth of information through the print. This suggests that in this case experts have an implicit set of features that they acknowledge are most diagnostic and tend to fixate these as a group.

Experts also tended to send their gaze to regions with more minutiae. Their fixations tended to land in and around the core and delta, where there is much more activity that creates minutiae.

Our machine translation procedures used the temporal correspondences across subjects to identify regions that might match across a pair of prints. This technique found many more corresponding regions for the experts, suggesting that the eye gaze pattern alone can reveal portions of the prints that the examiner believes matches.

3) How can data from experts discriminate between different candidate *quantitative metrics of the information in latent fingerprints?*

We used two complementary approaches to address the role that eye gaze data can play in discriminating between different candidate computational metrics that characterize the amount of information in fingerprints.

The first used data reduction procedures on pixel crops taken at the locations of fixations. This analysis revealed strong differences between experts and novices in terms of the nature of the information acquired by each group. Experts appear to be acquiring information from regions that are noisier, perhaps because their training allows them to tolerate more noise. They appear to be relying on lower spatial frequencies and coarser features, suggestive of configural processing (Busey & Vanderkolk, 2005).

The second approach defined several candidate entropy metrics using information theory and the angles of the ridge elements within a window. We found that some entropy measures performed better than others at discriminating between experts and novices. Overall this approach is extremely promising because in theory the angles of ridge elements provides a complete description of all of the information contained in fingerprints.

Implications for Practice

Together these results lay to rest the suggesting that experts have nothing more to add to court proceedings than a novice might. While it is true that humans in general have an intuitive sense for what features might be diagnostic, experts outperformed and looked different than novices on virtually all measures that we tested.

From an engineering standpoint, human expert eye gaze data provides an important validation procedure for any proposed quantitative approach to fingerprint matching. Human experts have proven to be superior to computers in the actual matching procedures, in part because they have an ability to alter which kinds of information they rely on. If Level 3 detail is available, they likely rely on it, while they may be able to determine when distortions might have occurred and therefore not trust ridge curvature as much as they might otherwise. This suggests that targeted experiments with these particular scenarios will help engineers collect eye gaze data that would help them train a ‘collection of experts’ model in which separate modules each make a decision and then a voting scheme is performed to make a final decision.

Experts should be cautioned that these differences between groups should not lead to complacency. Indeed, special abilities often come with vulnerabilities and care should be taken to avoid overconfidence.

Introduction

The major goal of this grant is to acquire eye gaze data from expert latent print examiners in order to determine what features or detail they rely on when conducting latent print examinations. This information is important because experts still outperform computers when conducting actual examinations (although machines have the advantage of vast storage space and can provide an initial candidate set for an expert to examine). Thus the human expert appears to have knowledge that is still not incorporated into machine-based approaches.

To determine which features experts rely on when examining latent and inked prints, we take advantage of the fact that the human visual system has a region of high acuity known as the fovea, and detail in regions outside this are (the periphery) are represented in much coarser detail. Since ridge information in fingerprints tends to be very fine detail, humans typically must move the location of their gaze around on a pair of prints, looking for regions that might match or definitively not match. By monitoring the position of the eye relative to the head, and the position of the head relative to the pair of prints, we can determine where their gaze lands on the ridges and therefore what information they choose to acquire.

Up until this project no one had collected eyetracking data from latent print examiners. One reason for this is that experts tend to be located in crime labs, away from research facilities. Eyetracking equipment is typically large not portable, often taking whole rooms in research labs. A major initiative of the current program has been to develop a portable, research-grade eye tracker that could be used to reliably collect data from expert examiners. This hardware/software package, which we have made open-source, is termed ExpertEyes (a play on the word ‘expertise’). This package is a major contribution of the present grant, and we will describe its functionality in section A below. We currently have data from 72 experts, 10 trainees, and 32 novices, which we have processed through our software.

Once the eyetracking data has been processed, we have a record of where the gaze falls on the prints, for each point in time. This data is very simple in form, because it is just the x-y locations in time. However, it is also a very rich dataset, which can be related to the features of the print, or other measures. We discuss the analysis of the eye tracking data in section B below.

Our rational for this research comes from the observation that no systematic study of expertise in latent print examiners has been conducted using eye gaze methods. Thus it is not known whether the experts are using features they say they are using. Indeed, there is evidence from radiology that expert radiologists fixate on regions other than what they say they are looking at (Krupinski, 1996).

Because we have two major sets of accomplishments, one related to data collection, and the other related to data analyses, we discuss these in two separate sections, and within each section we include the methods, results and conclusions from each section. We have also discussed the relevant literature within each section.

A. Eye tracking data processing

Why not just ask examiners what features they rely on? One issue with this approach is the fact that perception is often difficult to express in language and many perceptual processes take place at a level below consciousness (Snodgrass, et al., 2004; Vanselst & Merikle, 1993). An easy demonstration of this is the perceptual illusion in Figure 1. The red lines are actually parallel, but knowing this fact does not allow observers to change their perception of the two lines warping. However, the edge of a sheet of paper reveals that the lines are actually straight. This illustrates that conscious processes do not easily alter the output of the perceptual system, and that we do not have access to the basic levels of perception at the level of awareness. Many experts report having a flash of recognition or insight, and only then go back and document features that they now perceive as matching, a phenomenon known as the "Eureka" effect in the Cognitive Science literature (Ahissar & Hochstein, 1997).

How can we access the information that experts use, when the expert may not be consciously aware of what they rely on? How can we implement a quantitative approach (e.g. a cognitive model and/or a computer program) to extracting features that we don't know? Our solution is to rely on collecting fine-grained behavioral data from experts in combination with tailored experiments and computational modeling to infer the set of features that characterize the information content in friction ridge impressions.

A.1. Hardware and recording devices

After a survey of the marketplace we determined that no commercial eye tracker fit our needs of high resolution and portability. We need to collect data at latent print conferences like IAI and the Daubert Las Vegas seminar where we have access to large numbers of potential participants. We needed a portable system that is tolerant to different lighting conditions, as well as eyeglasses since good examiners tend to be older and require glasses. We also required a system that had good spatial resolving power, since the features we thought experts might rely on could be as small as a single ridge.

As no off-the-shelf system met these needs, we designed our own eyetracking system. It consists of two cameras, one of which records the scene and therefore the viewer's location relative to the image, and the other which records the position of the eye relative to the head. Through some calibration procedures we can recover the position of the eye relative to the fingerprint image. The system is portable, lightweight, and quickly adaptable to different subjects for rapid data acquisition at conferences.

Rather than try to analyze eye position in real time, we simply recording the video streams of the eye and scene cameras. We can take advantage of offline processing, which allows data processing that is extremely precise. We spend about 12 seconds processing each frame of a 30,000 frame movie that represents 15 minutes of data collection. Based on data collected during our calibration procedures, we reduce our location estimation error to around 1° of visual angle, which is about the size of the fovea and therefore represents close to what is meaningful since the human is acquiring information from the entire region contained within the fovea. We further improve our spatial resolving power by enlarging the fingerprint image on a 21" LCD monitor, thus making each feature large relative to the fovea.

Our eye tracker consists of a set of cameras mounted on a set of safety glasses (see Figure 2). This set of cameras records the position of the head and eye at 30 frames per second, which is suitable for a task in which most of the fixations are fairly stable.

A.2. Pupil and corneal reflection extraction

To illuminate the eye, we shine an infrared LED on the eye from a position near the eye camera, which is pointed at the eye. This results in a bright spot appearing on the cornea, called the corneal reflection. The relation between this spot and the pupil varies systematically as the eye changes orientation. The first step is to identify the pupil and corneal reflection. While a variety of techniques have been proposed, we created a novel procedure in which a forward eye model is created by drawing dark and light ovals over the eye image and adjusting the position of the two to find the best match. Figure 3 illustrates this procedure, while Figure 4 demonstrates the software that fits the forward eye model.

These adjustment procedures are quite time-consuming, and take 6-8 seconds per frame to compute. Thus it can take several hours on an 8-core computer to fit the model to all 30,000 frames for each subject.

A.3. Calibration with the scene view

To make the link between the scene camera and the eye camera, we ask participants to look at black dots on a white screen. These appear for about 5 seconds and then move to another location. The ExpertEyes software contains a module that allows us to link an eye position with a position in the scene camera. The user selects a frame where the participant is assumed to be looking at the black dot, and then indicates the location of the dot in the scene camera. The computer then finds this dot in several subsequent frames to reduce noise.

We then fit a two-dimensional polynomial function that relates the u-v location of the pupil to an x-y location on the scene camera.

A.4. Monitor corner detection and rectification

The second stage of analysis requires that we identify the location of the monitor in the scene camera. We apply a barrel distortion correction algorithm to each scene camera image and then have the user click on each of the four corners of the monitor in a thresholded scene image. We then use Gabor jets, which are an adaptive template match algorithm, to identify the corners in the rest of the images. Figure 5 demonstrates how we correct for the barrel distortions that all cameras induce, and Figure 6 illustrates the procedures by which we search and identify the corners. The search space is found for each corner by a human coder using a shuttle-jog to quickly mark the search area for the entire movie for each corner.

Once we know where the eye is in the scene view for a particular frame, as well as where the corners of the monitor are in the scene view, we can then interpolate the eye position back into the coordinates of the monitor. This gives us the position of the eye on the images that contain fingerprints.

These procedures provide the ultimate goal of the eye tracker: the position of the eye on the fingerprint ridge detail. We can then tell, with accuracy that is about the size of the fovea, which information the expert is using when performing examinations.

A.5. Trial and event extraction

To obtain enough information about the diagnosticity of different kinds of information, we often conduct experiments with 30-40 images, each shown for 20-30 seconds to encourage the examiners to focus on only the most diagnostic features. The ExpertEyes software contains a module that automatically extracts event information and allows human verification and correction.

A.6. Data cleaning and export

Eyetracking data invariably has some missing data, usually when the participant moves their head enough so that a corner of the monitor moves off the edge of the scene camera. Figure 7 illustrates the procedures that allow the user to identify and mark regions of the data that are bad. We typically throw out less than 5% of our data.

A.7. Fixation finding

The final step in the analysis process is to find fixations. The eye tends to move ballistically from one location to another on static images, with dwell times that average about 300 ms in duration. During this dwell period the eye experiences micro-saccades, tiny movements that prevent the visual world from fading. However, these micro-saccades are typically not meaningful from an analysis standpoint, and there is also jitter from error in the estimation procedures. Thus to smooth the data we rely on fixation finding routines that perform a cluster analysis that groups similar eye gaze locations into fixations and saccades.

We have developed our own algorithm of eye fixation finding which is composed of five step: 1) First, we compute the magnitude of velocity from raw eye movement data (x,y); 2) we next use a pre-defined threshold to segment the whole continuous stream into several big segments that correspond to dramatic eye location changes; 3) we zoom into each big segment and re-segment each into individual segments that correspond to small eye position changes. Those small segments may or may not correspond to fixations; 4) Finally, we took spatial information into account by merging small segments (detected from Step 3) if they are spatially close to each other (e.g. eyes moving around an area with a certain speed). After the above five steps, we typically are able to successfully segment a continue eye movement stream into several eye fixations by integrating both temporal (the speed of eye movement, etc.) and spatial (the overall spatial changes of eye gaze location) information.

Figure 8 shows the raw eye trace along with the results of the fixation finding algorithm. The parameters are tuned by eye, since no objective criterion for fixations and saccades exists in the literature.

A.8. Calibration verification and quantification

An all-important task is to identify whether our calibration procedures accurately measure eye position. To ensure this, we ask our participants to perform an additional calibration procedure at the end of the experiment. We ask them to look at known locations on the monitor and then verify whether we can accurately track their gaze. As Figure 9 illustrates, we typically find that our calibration accuracy is quite high, comparable to commercial systems. In many respects our system is superior to commercial systems because we are able to go back and re-fit the eye model if the parameters are incorrectly specified. If we were recording live from the field we would simply not be able to use that data. Given how valuable the data is from experts, we are fortunate in that we are able to make full use of almost all of the data we gather from examiners.

A.9. System integration

The ExpertEyes software integrates four major modules: eye calibration, trial marking, data cleaning and eye model fitting. Figure 10 illustrates the main screen of the software that integrates these different modules for one subject's data.

B. Data analyses

The analyses of eyetracking data have subsumed much of our research efforts for the latter portion of the grant. Perfecting our eye tracker to enable research-grade data required an iterative procedure of testing and development, and the result is that we are much further along with the software tools to gather eye tracking data than we are with the data processing. The data analysis is complicated by the fact that there are several major classes of approaches that can be used to identify the nature of expertise in latent print examiners.

Below we describe the six different types of analyses that we have conducted. Note that each of these approaches relies on sophisticated mathematical modeling and machine learning techniques, which are quite time consuming to implement and test. Thus the results below represent in many cases completed analysis, but also signposts to future analyses as is often the case in the hypothesis development and testing loop of science. The first three sections below are described in Busey et al. (Submitted Manuscript) and portions of the text below also appear in the submitted manuscript. The methods of data collection are presented in the appendix of this Final Report.

We compared the behavioral results found a mean d' for experts of 2.47 and a mean d' for novices of 1.30. Thus experts seem to be able to reliably out-perform novices and have fairly high d' values even with short viewing times. The experts had a false alarm rate of 0.007 (or .7%) while the novices had a false alarm rate of .25. These numbers are based on relatively small number of participants, but it does indicate the rather large accuracy difference between the two groups.

B.1. Fixation statistics

The first step in analyzing eyetracking data is to compute various summary statistics that are likely to be related to the strategies used by experts and novices. These are essential to determine whether the eyetracking data contains the kind of structure that will allow us to extract information about the quantitative information content as revealed by eyetracking scan paths.

At each point during the experiment, the eyes are at a location on the latent/inked print pair. Eyetracking researchers distinguish between two main states of eye position: fixations and saccades. Fixations are periods where the eye is relatively stable on the image, and tends to last between 200 and 500 milliseconds. Saccades are periods where the eyes move rapidly from one region to another for a period of around 20-170 ms. Note that these two categories characterize a dataset that is often more continuous, since the eyes can move slowly and smoothly at times as in smooth pursuit of moving objects or when experts follow a ridge. These, however, tend to be second-order effects and there are a number of accepted techniques that are used to decide whether the eye is currently in a fixation or saccade state.

Once the fixations and saccades have been identified, we can compute statistics about the expert and novice groups. We had no prior expectations for these data, since this kind of data has not been collected before. Perhaps somewhat surprisingly, experts and novices spend the same proportion of time looking at latent and inked prints (about 67% on latent) and the mean fixation duration is about the same on the latent image, for about 400 ms. Both groups made longer fixation durations on the inked print, by a factor of 50% more. Thus for these two groups, the statistics are quite similar, and may reflect low-level eye movement programming behavior that is shared by all human observers.

We found that experts and novices have equal numbers of saccades ($F(1,11)<1$) but the experts had much shorter saccades than novices both on the latent print side (66.3 vs. 100.7 pixels, $F(1,11) = 9.0$, $p<0.05$) and on the inked print side (57.5 vs. 99.8 pixels, $F(1,11) = 12.3$, $p<0.01$). These results are consistent with experts making smaller eye movements to regions that are close together.

In addition, experts make an average of 58.9 saccades within a fingerprint vs. 24.7 for novices. So while experts make more saccades than novices, they still find a way to suppress saccade length with a mean length of 171.8 vs. 191 pixels for novices. One possibility is that experts might know the layout of fingerprints, allowing them to encode information at highly diagnostic fixations and reduce extraneous saccade. Also, experts are able to make more contiguous fixations on one side before switching (mean 5.46 vs. 4.32 for novices). All of this converting evidence suggests that they approach the task of print matching in a different way than novices and may intuitively understand which features are most diagnostic.

To validate these candidate hypotheses, however, we have to look at the locations visited by experts, and we rely on a comparison across observers to look for consistency within groups. If experts rely on a common feature set and quickly identify the most informative regions in a print pair, we should see common regions visited by all experts. This is discussed in the following section.

B.2. Consistency among experts and novices

To visualize the regions visited by experts and novices, we create a visualization called a heat map by accumulating the duration spent at each location and representing this information as color applied transparently over the print pair. Figure 11 shows two example print pairs from two participants, one expert and one novice.

These heat maps represent regions that were visited for longer durations as hotter colors. As can be seen in Figure 11, the expert tends to focus their attention on a few regions, while the novice tends to distribute their gaze across disparate locations.

We noticed that this was a general trend. Not only did experts tend to focus on fewer regions, they tended to be the *same* regions across experts for a print pair.

To confirm this, we computed a consistency analyses. This relies on a statistical algorithm known as the Earth Movers Distance. To gain an intuitive sense of this comparison technique, imagine that one expert's heat maps represent piles of dirt, with hotter colors representing taller piles. A second subject's heat map represents holes in which dirt can be added. The Earth Mover's Distance computes the shortest path that a dump truck could take to move the dirt from the piles to the holes. If two experts have very similar maps then this is a very short distance since the piles and the holes align. However, if two subjects (say two novices) have very different heat maps, the distance to move the 'dirt' from one subject to another will be quite large. In essence this quantifies in a formal way the intuitive procedure of printing out both heatmaps on transparencies and holding them up to see if they align across subjects.

We performed this analysis with both latent and inked prints. With latent prints, of the 90 trials in the experiment, experts were more similar to other experts on only 29 trials, while on the remaining 61 trials the novices were more consistent with each other. We computed the mean of the inter-expert distances, which was 48.4. The mean of the novices was smaller at 45.04. A paired t-test computed across the trials was significant ($t(89) = 2.19$; $p < 0.031$).

With inked prints, we found the opposite results. We again computed the earth mover distance for each trial (including each of the individual presentations within each trial) between each expert and every other expert, as well as between each novice and all the other novices. However, contrary to the latent/inked prints, we find that now the experts show much more consistency than the novices. Experts had smaller distances to other experts on 78 trials and novices had smaller distances to other novices on only 12 trials. Experts had an average distance to other experts of 133.2, while the average distance of novices to other novices was 196.4 ($F(1,178) = 97.9$; $p<0.001$). This suggests that experts have an intuition for which areas are likely to be the most diagnostic, and given high quality prints they are able to quickly move to those areas and begin to acquire information.

Implications for training and practice

The consequences of greater or less variability is something that we believe the community should discuss. It is entirely possible that a consistent conclusion could be reached by examiners using different sources of information. However, it may be troubling from a legal or policy standpoint to have different examiners relying on different sources of information. Until we have an independent information metric (discussed later), we do not have direct comments on this issue. We do believe that issues of consistency should be discussed by professional organizations. It is possible that it did not appear as an issue until now due to the lack of eyetracking data.

It will also be important for the community to continue to monitor the state of eyetracking research. The present results are based on a relatively few numbers of subjects, and more data may refine the conclusions about which group has more variability.

B.3. Use of minutiae by experts and novices

Minutiae provide the foundation for many strategies of latent print identification, and may be used implicitly by examiners even if they do not explicitly rely on these sources. Experts might be more or less likely to visit regions that include minutiae. If the novices all tended to gravitate toward a cluster of high-quality inked prints and the experts tended to focus instead on non-minutiae information such as ridge orientation and inflection, we might find differences in terms of the number of minutiae each group visits.

To assess this computationally, we defined a circle of arbitrary size (we chose 50 pixels but the conclusions do not depend on this value) and centered this circle on each fixation. We then counted the number of minutiae inside this circle, as illustrated by Figure 12.

We found that experts tended to have many more minutiae near each fixation. Experts had an average of 1.50 minutiae near each fixation, while novices had only an average of 1.16, which is a significant difference. The 95% confidence interval based on the null hypothesis of no difference between the two groups is [-0.24, 0.26] which does not include the actual difference of 0.344.

Figure 13 demonstrates why experts have so many minutiae near each fixation. The expert fixations are in red plus signs, and the novices are in magenta asterisks. On this particular trial the experts had 1.42 minutiae near each fixation, while the novices had only 1.05, which is a statistically significant difference. The novices tended to look at the top portion of the print, which has relatively few minutiae. The experts tended to look at the core and lower-right regions, which tend to have much more movement and activity in the ridges, leading to more minutiae. This practice will lead to not only looking at more minutiae, but also greater consistency among experts if they all follow a similar pattern of looking in these high-density regions.

B.4. Machine translation approaches to data analysis

The major goal in this type of data analysis was to integrate data from human experts with statistical machine learning algorithms to provide a quantitative analysis of the information that was available in inked and latent prints. We argue that the quantitative evaluation of the information contained in latent and inked prints can be vastly improved by using elements of human expertise to assist the statistical modeling. The eye tracking technique allowed us to monitor the expert's momentary gaze moments when s/he is engaged in examining a fingerprint. We argue that eye movement data indicate what participants visually attend to when they examine a fingerprint image. Thus, participants may not know explicitly their matching algorithms or strategies, and they may not be able to formally describe what they are doing. However, their eyes are moving rapidly in our task and actively collecting the information for the brain. Every single eye movement indicates what information is required in the internal computation.

We have recruited two groups of participants in our study – an expert group (n = 6) and a novice group (n = 6). Each participant in each group was asked to examine 30 ink and latent print pairs display simultaneously on a computer screen. In each trial, they were asked to make a judgment whether the latent print in the current trial matches with the ink print. Besides this task, there was no constraint how long they may take to make this decision, and what visual features they should pay attention to. We used a Tobii eye tracker to monitor their gaze at a frequency of 50Hz. In total of 30 trials, we've collected 45,000 data points per participant. The goal of our data analysis was to demonstrate potential differences between novices and experts. We intended to find the differences between experts and novices based on their eye movements. This aim was particularly challenging given the following observations: 1) Different experts may visually attend to different subsets of features and used them to perform the matching task; 2) Novices may visually attend to most salient areas in both inked prints and latent prints, which may contain important features; and 3) Different prints may pose different challenges to individual participants.

Our solution was based on advanced techniques in machine translation. Briefly speaking, given two sets of text in parallel (for example, one in English and one in French), current machine translation techniques can compute which words in English maps to which words in French. Similarly here, if we treated the areas of interests in the ink and latent prints as two languages, then we can compute which region in one image is mapped to which area in the other image. Our hypothesis was that experts might be able to identify more such pairings while novices might not be able to do so.

Given a fingerprint-matching task, subjects (either fingerprint experts or novices) generate multiple eye movements on both the ink print and the latent print displayed simultaneously on the computer screen. We are interested in visual features that are used to make a decision. To do so, we need to find which eye movements on the ink print correspond to which ones on the latent print. The correspondences can be potentially estimated based on two streams of eye movements that jointly determine which ink-latent patch pairing is relevant. As shown in Figure 14, we can conceptualize patch-to-patch mapping as a translation problem - how to cognitively translate image patches in an ink print into image patches in a latent print. Thus, the learning mechanism we propose rests on advances in machine translation. Briefly, machines 'learn' word correspondences (which word in one language corresponds to which word in another language) by finding statistical regularities across large parallel corpora in two languages.

Here, we use a similar computational approach with eye movements in latent prints as one language, and eye movements in latent prints as the other. This conceptualization provides a unique way to discover patch correspondences. More specifically, the learning process can be formalized as an expectation-maximization algorithm (EM) (Dempster, Laird, & Rubin, 1977). The idea of EM is that there is a way of representing the data as the sum of component probability distributions. More specifically, the probability of an ink patch is expressed as a weighted mixture of the conditional probabilities of the latent patches. The goal of our data analysis algorithm is to find those reliable correspondences that maximize the likelihood function of observing the whole data set. The method computes association probabilities of all the possible patch pairs simultaneously.

This technique not only reveals the corresponding regions between latent and inked prints, it also simultaneously estimates the features used to make these correspondences. The first step in this process is to identify clusters of fixations and use the temporal sequences of fixations to determine correspondences between the latent and inked prints. Figure 15 shows an example from an expert. The white lines link and highlight corresponding areas in two images identified by two experts, which is very similar to each other.

Figure 16 shows two major results that come out of this analysis. First, experts identify far more corresponding areas between latent and ink prints (m=12.3) compared with novices. This confirmed our hypothesis that fingerprint expertise can be demonstrated by their fine-grained behavioral data in real-time examination. More interestingly, we also found a reliable consistence of the area pairs identified across by different experts. Moreover, Figure 16 showed individual differences with novices. We found an outlier subject #4 in this group. It turned out that subject #4 was a graduate student in the PI's lab and has extensive experiences with fingerprint and interactions with fingerprint experts. Interestingly, even that subject didn't get a formal training to be a fingerprint expert. Through daily research experience with fingerprints, this individual also acquired some expertise that is comparably significant compared with other novices.

B.5. Feature induction from fixation data

The machine translation approach takes advantage of contextual information when matching regions or clusters across observers. If two or more observers tend to jointly fixate two regions across a print pair, the algorithm will identify these correspondences, much like hearing an unknown word in the context of known foreign words help establish the meaning in terms of know English words.

These analyses do not have access to the underling feature information; only the fixations from experts and novices along with the x and y coordinates of the fixations. To take full advantage of the feature information available when linking the fixations back to the ridge detail, we have extracted out small windows of features centered on each fixation, like those shown in Figure 17.

We then collect all such image windows from all 6 of our experts on the clean images and perform several different kinds of feature induction analyses. We have included the data from 25 experts across 4 different experiments, giving a total of over 38,000 image windows from 270 unique fingerprints. We also included some 32,000 fixations from 18 novices on the same images. We analyzed the data separately for the two groups in order to make comparisons across the groups in terms of the features they rely on.

Our first data analysis relies on a technique called independent components analysis (ICA). This approach relies on looking for spatial correlations, much like principal components analysis, but with the orthogonality constraint relaxed. We extract out a set of features that forms the fundamental alphabet for image representation.

The left panel of Figure 18 illustrates the results of one such analysis. It shows a large set of features derived from ICA that represent the basic features used by examiners. The image patches were rotated vertically prior to data reduction, which account for the fact that the majority of energy is in the vertical plane. The individual features are grouped together by similarity to demonstrate how the algorithm can handle spatial uncertainty.

The right panel of Figure 18 illustrates the feature set derived from the novices. Despite an approximately equal number of image patches in the two datasets, there are striking differences between the two sets of recovered features. The experts appear to rely on features that are of coarser scale, which would be consistent with the use of integration over greater spatial extents (see (Busey & Vanderkolk, 2005) for more on configural processing in experts). The novices appear to rely on finer regions that are more noise-free. These regions may not be as diagnostic as more noisy areas, since the identification task sometimes requires obtaining information from regions with poor signal to noise ratios.

We can use these derived filters in several different ways. The feature set can be used as a basis set similar to that used in image compression (Portilla, Strela, Wainwright, & Simoncelli, 2003). We have begun to ask whether the feature set for experts is easier to compress using the fixed 120-element feature set than the novice set. If it were, this would suggest that experts have an implicit agreed-upon set of features that they naturally gravitate to.

The basis set also allows us to characterize the high-dimensional space of each patch (each patch has 48*48 or 2304 dimensions) in a much lower, 120 dimensional space. This space simply indicates the contributions of each of the 120 basis features to each patch. The reconstruction will have some error, but the redundancy seen with the images in Figure 18 suggests it will be minimal. This low-dimensional representation then allows us to compute the relative base rates of different types of features, expressed in terms of their position in the 120 dimensional space. If we have a feature that falls in a relatively sparse region in this space (acknowledging that most things are in sparse regions in a 120 dimensional space) this might suggest that this may be particularly diagnostic or rare.

B.6. Information theoretic approaches

We can use the eye tracking data derived from experts to help test candidate information quantification approaches. The idea is to create several candidate measures of information and use the data from the experts to determine whether they tend to fixate on regions that have more information by a particular metric. If so, we can then use this to begin to explore extensions to the metric to determine how it might be extended and improved. Once we have a revised metric, we can then use to suggest changes that the experts might make to their procedures, or provide base-rate information on particular types of features. Thus the refinement of a set of metrics and testing against human data is a feedback loop that can be used to improve both human performance and measures of information in fingerprints.

All measures of the information content in friction ridge impressions must first create a feature set. For example, many current approaches rely on the locations and orientation of minutiae (Egli, Champod, & Margot, 2007; Neumann, et al., 2007; Su & Srihari, 2008) or ridge elements (Su & Srihari, 2008). These approaches have real strengths, in that they are easily identified and face validity in that experts report using minutiae and relative locations of ridges. They also allow statistical approaches that model the variability within different impressions of the same finger as compared with other prints (Neumann, et al., 2007) and provide generative models that can be used to characterize the probability of any one particular print relative to the entire database (Srihari & Su, 2008). However, minutiae and ridge endings may represent only a subset of the information used by experts, even if the print is too noisy for third level detail. Indeed, there is a published report of a small friction ridge impression that was individualized despite the lack of *any* ridge ending or bifurcation minutiae (Reneau, 2003). Thus, the feature set used by experts may contain many different sources of information, only some of which is captured in current quantitative approaches.

Engineers have tried many different techniques to quantify the information available in fingerprints, including minutiae-based systems, ridge flow and even level 3 detail. However, the development of each of these systems tends to ignore the expertise that resides in the latent print examiners' visual systems, which have proven to be much better than computer recognition systems. We are using our eye tracking data to validate candidate information metrics, under the assumption that experts will move their eyes to regions they deem most diagnostic or contain the most information. Below we describe one set of metrics based on information theory, and then describe the validation process.

One possible metric for measuring characteristics of locations being looked at by finger print examiners is the amount of information contained in the locations. In this analysis, we focus on the ridge edge direction information being taken in by the examiners through their eye fixations on a print. Here, we use Shannon information theory (Shannon & Weaver, 1949) as the measure of the information. The detail of how we perform a measurement is as follows.

As shown in Figure 19, we extract fingerprint ridge orientation information and create the enhanced fingerprint image using Peter Kovesi's implementation[1] of the technique proposed by (Hong, Wan, & Jain, 1998). Then ridge edges are detected from the enhance fingerprint image using boundaries detection function (bwboundaries) from MATLAB Image Processing Toolbox. We then work along the individual ridges in the print and encode the orientation of each ridge edge element into strings. In order to use Shannon theory, the orientation is rounded off to a full degree (so we have 181 symbols representing 0 to 180 degrees to use for encoding a string). The following equation is then used to compute information amount presented:

$$I = -\sum_{i=1}^{n} p(x_i)\log_2(p(x_i)) \qquad (1)$$

where n is the total number of edge elements. $p(x_i)$ is the probability that the orientation of the edge element i appear in the observed area.

Using above metric, we investigate the amount of information taken in by expert and novice fingerprint examiners. Given a fixation point on the fingerprint, we approximate the area observed by taking a square patch around the fixation, a 50x50 pixel area in a 1050 x 1680 pixel fingerprint image in our experiment.

In order to visualize the possible successes and limitations of each candidate information metric, we have developed a color-based visualization that illustrates the amount of information contained at each location as determined by a particular algorithm. These Figures are found at the end of this report due to the large size of the figures.

Figure 20 and Figure 21 illustrate two information metrics, where the amount of red in the visualization display implies a greater amount of potential information available. The relative information metric in Figure 20 seems to correspond to what intuitively seems informative, mainly regions of minutiae, as well as large changes in curvature. Figure 21 also shows these regions, with the addition that this metric may not discriminate as well between high and low information regions. However, whether these metrics are actually useful depends on whether they correspond to whether they are fixated by experts. If so, we can begin to refine our information metric. However, if we don't see differences between

[1]http://www.csse.uwa.edu.au/~pk/Research/MatlabFns/index.html

experts and novices, or between experts and random fixations, we will need to abandon this particular set of information metrics and look to other possibilities.

The validation procedures with experts and novices are described next. This is an important step and highlights the value of our expert data. Rather than simply coming up with a metric that we think might work, we verify that experts might intuitively use such a metric, even if they do not explicitly calculate the probabilities. Instead, we are modeling processes in the visual system that may approximate these calculations.

Information Use by Experts and Novices

A reasonable approach to testing various metrics is to determine whether experts rely on the regions determined to be most informative by a particular metric. One could simply ask experts what information they use, but often they have difficulty explaining what regions or information they actually use. This results from the fact that much of perception resides below the level of consciousness, and can be difficult to explain in the sparse code of human language.

We compute the information contained at each location for each subject, and then average these across trials for each subject. We are interested in whether experts tend to visit regions with different information values than novices do.

Figure 22 and Figure 23 show the fixations from two experts, superimposed over a relative information metric based on the difference between two consecutive angles. The fixations appear to fall in regions that have higher information content. Figure 24 and Figure 25 show the same information metric applied to data from two novices. Here, the fixations seem to fall in regions that have lower information content.

To test these observations, we compared the average information value visited by experts and compared this with that visited by novices. We found that three metrics reliably differentiated between experts and novices ($t(11) = 2.88$; $p = 0.015$). We computed the information at each fixation of the experts and averaged these across the fixations for all trials in which individual images were shown. Experts reliably visited regions that had lower values of entropy, meaning that these were regions that tended to have more variable ridge detail. This was true whether entropy was calculated along a ridge line, or just averaged over all pixels in a region around the fixation.

We also found that experts rely on much more informative regions that one might expect on average. We computed the entropy values at random locations, taking X and Y values from different fixations to ensure that we drew entropy values from valid locations. Experts performed reliably better on many measures, including the standard entropy measure described above ($t(11) = 2.81$; $p = 0.017$). In fact, the novice data looks very similar to the random data, suggesting that novices may not have a clear idea of where to look in the image.

These results validate our entropy measure as an approximation to the kinds of information that experts are using when matching latent to inked prints. This is an important step because this metric quantifies the total information available in fingerprints, given our particular definition of information.

Not all metrics proved diagnostic. We had high hopes that the relative angle measure, computed by taking the difference of the angles along a ridge, might also be diagnostic. However it did not reliably differentiate between experts and novices, suggesting that experts may not rely on this information.

Additional Entropy Analyses

In addition to the analyses described above, we have to other analyses that we have begun to explore. The analyses of these metrics are ongoing.

If experts can reliably match a region on the latent with a region on the inked print, they may do so because the two patches have *similar entropy values*. This suggests that we can compute the combined entropy for the two patches across a saccade that bridges the two prints. We should find that these entropy values are reliably different than ones computed from novices across saccade pairs.

This analysis will help determine what strategies and information sources experts use when making saccades across image pairs.

Implications for training and practice

It is likely too soon to draw firm conclusions about possible changes to training procedures given that we have only begun to explore various information theoretic approaches. However, we anticipate being able to make such conclusions in the coming years.

B.7. Linguistic identifiers assigned to extracted features

In addition to the visual-based approaches described in the previous sections, we collected linguistic labels from examiners in an attempt to use these to provide a source of data compression.

To obtain linguistic labels, we presented examiners with 48x48 pixel crops from latent and inked prints that were centered on the fixations from expert examiners. These were not otherwise standardized or rotated. We asked the examiners to provide a description of the features in such a way that this might enable individualization.

This study produced mixed results. The examiners expressed some frustration with this task, since we asked them to provide just 2-4 words or phrases per crop. They discovered that this format was not nearly rich enough to capture the detail of the crop that would enable individualization. They used phrases such as: "core, short ridge, incipient ridge, bifurcation up, recurve, ending ridge, bend down, short near core, and incipient spur". While these were accurate, they were analogous to describing the Mona Lisa as 'white female with dark hair'.

Despite the relative sparse nature of language for the description of uniqueness in fingerprints, there may be situations in which linguistic identifiers may play a role. For example, mentors may use linguistic terms for perceptual features during training procedures with novice examiners. This may enable the trainee to comprehend the classes of different sources of detail in prints.

B.8. Summary of data analyses

Patterns of eye movements reveal whether experts move their gaze to regions that are most informative as defined by different information metrics. The collection of such data allows us to validate candidate metrics and improve them where they tend to deviate from expert eye gaze. In essence we are reverse-engineering the human visual system in order to better engineer computer-based approaches to fingerprint quantification. Linking the gaze back to the physical features of friction ridge patches will illustrate the nature of expertise, and help improve machine-based analyses of latent prints.

C. Conclusions and implications for policy and practice.

Prior to this project, virtually nothing was known about the differences between experts and novices in terms of the information they choose to acquire, since no one had collected the eye tracking data. Once we established a reliable software toolkit that enabled robust data collection, we were able to use expert/novice comparisons to demonstrate a number of important conclusions:

First, while the basic timing of eye fixations and saccades does not differ between the two groups, experts make smaller saccades, which suggest that they are looking at clusters of nearby features or combinations of features. Novices may be attempting to match individual features across the prints, which are not nearly as diagnostic since the combinatorial explosion of clusters makes these much more diagnostics.

Second, experts acquire data from resource-rich regions of fingerprints, which are associated with more minutiae. However, they are also more willing to wander into regions that have poorer signal-to-noise ratios in an attempt to find information from a partial print that might prove diagnostic. We can infer this from the structure of the feature set found thorough the ICA data reduction procedures.

Third, we documented the conditions under which experts may be more or less consistent with each other as a group, depending on image quality and the content of the fingerprints.

Fourth, we adapted techniques from machine translation to demonstrate how these procedures could be used to identify corresponding regions of the fingerprints.

Finally, we demonstrated how we could create several candidate information theoretic metrics that describe in quantitative terms the information contained in fingerprints. We used the expert/novice comparisons to demonstrate how one can discriminate between these candidate measures.

These conclusions have a number of applied applications. First, they go a long way toward addressing the criticism that experts do not bring anything of value when they testify. It is clear now from a number of studies that experts have special abilities that exceed those of novices, and therefore should be allowed to share that expertise in court.

The results also suggest that engineers who are refining fingerprint-matching algorithms should look to data from human experts to validate their approaches. The advantage of approaches such as the data reduction and information theoretic metrics described in Section B is that the models are built on the same math and computer code that computer-based fingerprint identification systems use. Therefore the results should be easy to adopt by those working on computer-based matching.

All of our studies were done using abbreviated testing conditions in order to obtain the amount of data necessary to test candidate models, and therefore care should be taken when generalizing these results to full-blown latent print examinations. Future studies should address the changes that might occur when experts are given more time with the images.

Finally, there is a wealth of other studies and analyses that could be done with the rich nature of eye gaze data, and additional scientists should be recruited to help address a variety of different questions related to the nature of expertise in latent print examiners.

D. Products, publications and presentations

D.1. Products

This project produced several different scientific advancements. The most fundamental achievement was the demonstration that eyetracking data could be reliably gathered in the field from active latent print examiners. This took a great deal of work to accomplish because no system existed that was reliable and portable for such an endeavor. Our ExpertEyes software is available to download for free, and we have several groups interested in adopting our system. The hardware schematics are also online.

The second major achievement is the demonstration that the eyetracking data acquired from experts and novices could be used to improve upon existing quantitative approaches to fingerprint detail. There are many different approaches, and we have endeavored to touch on each of these techniques. The best demonstration of the utility of our approach comes from the information theoretic techniques, in which the data from experts can be used to demonstrate which of several different computations of entropy are most likely to be used by experts.

In addition to these accomplishments, we have the following publications and presentations that reflect our dissemination efforts.

D.2. Publications

The following publications are made possible with the support of the current grant.

Busey, T.A. & Loftus, G. R. (2007). Cognitive science and the law. *Trends in Cognitive Sciences, 11,* 3, 111-117.

Busey, T.A. & Dror, I. (in press). Special Abilities and Vulnerabilities in Forensic Expertise. To appear in *The Latent Print Sourcebook*. Alan McRoberts (Ed). Published by the National Institutes of Justice.

Busey, T. A., Yu, C., Wyatte, D., Vanderkolk, J., Parada, F., & Akvipat, R. (Submitted Manuscript). Consistency and Variability Among Latent Print Examiners as Revealed by Eye Tracking Methodologies. *Journal of Forensic Identification.*

Busey, T. A. & Parada, F. (in press). The Nature of Expertise in Fingerprint Examiners. In press at *Psychonomic Bulletin & Review*.

D.3. Presentations

We believe that outreach activities in the form of presentations at scientific and educational conferences is crucial to dissemination activities. We have made every effort to present our work for critical appraisal and discussion, including the following 22 presentations, listed in chronological order:

Busey, T. A. & Vanderkolk, J. R. (2005). Expertise in Fingerprint Examiners. Paper presented at the New England Society for Identification annual meeting.

Busey, T. A. (2005). The role of configural processing in the development of visual expertise. Paper presented at the Annual Interdisciplinary Conference.

Busey, T. A., & Vanderkolk, J. R. (2005). Behavioral and electrophysiological evidence for configural processing in fingerprint experts [Abstract]. Journal of Vision, 5(8), 635a, http://journalofvision.org/5/8/635/

Vanderkolk, J. R., & Busey, T. A. (2005). Forensic individualization of images. Forensic Identification Seminar of Toronto, Canada.

Vanderkolk, J. R., & Busey, T. A. (2005). Expert and novice fingerprint studies. Fingerprint Society Lectures, Brighton, United Kingdom.

Busey, T. A. & Vanderkolk, J. R. (2005). Expertise in Fingerprint Examiners. Paper presented at the International Association of Firearm and Toolmark Examiners annual meeting.

Busey, T.A. (2006). How The Cognitive and Visual Sciences Might Help (and Hurt) in a Daubert Hearing. Paper presented at the 2006 meeting of Forensic Document Examiners Society.

Busey, T.A. (2006). Measuring Expertise in Latent Print Examiners to Improve the Quantitative Analyses of Latent Prints. Paper presented at the 2006 meeting of the International Association for Identification, Boston, MA.

Busey, T.A. (2006). Quantifying Human Expertise During Latent Print Examinations. Paper presented at the annual meeting of the National Institutes of Justice, Boston, MA.

Schneider, B., DeLong, J., Wyatte, D., James, K., & Busey, T. (2007). The neural correlates of face-like expertise in fingerprint examiners [Abstract]. Journal of Vision, 7(9):575, 575a, http://journalofvision.org/7/9/575/, doi:10.1167/7.9.575. Paper presented at the Vision Sciences Conference, Sarasota, FL.

Blaha, L., & Busey, T. (2007). Electrophysiological substrates of configural learning [Abstract]. Journal of Vision, 7(9):796, 796a, http://journalofvision.org/7/9/796/, doi:10.1167/7.9.796. Paper presented at the Vision Sciences Conference, Sarasota, FL.

Busey, T. A. (2007) Expertise in Latent Print Examiners. Invited paper presented at the Indiana Society for Identification annual meeting, Evansville, IN.

Busey, T.A. (2007). Measuring Expertise in Latent Print Examiners. Paper presented at the 2007 meeting of the International Association for Identification, San Diego, CA.

Busey, T.A. (2007). Extracting Expertise from Latent Print Examiners. Paper presented at the annual meeting of the National Institutes of Justice, San Diego, CA.

Busey, T. A. (2007) Cognitive psychology and fingerprint experts. Paper presented at the Illinois Society for Identification annual meeting, Peoria, IL.

Busey, T., Schneider, B. & Wyatte, D. (2008). Expertise and the width of the visual filter in fingerprint examiners. Volume 8, Number 6, Abstract 178, Page 178a http://journalofvision.org/8/6/178/

Schneider, B., Harman-James, K., Wyatte, D & Busey, T. (2008). A noise x inversion paradigm reveals the nature of fingerprint expertise for latent print examiners in EEG and fMRI. Volume 8, Number 6, Abstract 177, Page 177a http://journalofvision.org/8/6/177/

Wyatte, D. & Busey, T. (2008). Low and high level changes in eye gaze behavior as a result of expertise. Volume 8, Number 6, Abstract 112, Page 11 http://journalofvision.org/8/6/112/

Busey, T, & Vanderkolk, J. (2008). Measuring Expertise in Latent Print Examiners. Talk presented at the 2008 International Association for Identification Annual Meeting in Louisville, KY, August 2008.

Busey, T, Yu, C. & Vanderkolk, J. (2008). The Varieties of Expertise in Latent Print Examiners. Talk presented at the 2008 International Association for Identification Annual Meeting in Louisville, KY, August 2008.

Busey, T. (2008). Expertise in Latent Print Examiners As Revealed by Behavior, Electrophysiology, and Eyetracking. Invited Symposium presented at the 2008 Psychonomics Society Annual Meeting, Chicago, IL.

Busey, T, Yu, C., Akavipat, R. & Vanderkolk, J. (2009). Adding Human Expertise to the Quantitative Analysis of Fingerprints. Talk presented at the 2009 National Institutes of Justice Grantees meeting in Tampa, FL, August 2009

Appendix

Methods for Latent/Inked comparisons

The procedures below describe how we collected the data for our latent/inked comparisions.

Stimuli

The stimuli for Experiment 1 come from the National Institutes of Standards and Technology Special Database 27, which has previously determined identifications of latent and inked prints that are typical of what is traditionally found during casework. The images were grouped into *good*, *bad* and *ugly* prints by an expert with an obvious affinity for Clint Eastwood western movies. We used a mixture of all three types, which produced trials that could be quite challenging at times.

We created three lists of images, each of which has 30 pairs of images. We had relatively few non-matching prints since our analyses are less applicable to non-matching pairs. List 1 had 5 non-matches (exclusions), list 2 had 3 nonmatches, and list 3 had 8 nonmatches.

Participants

Our experts were recruited from forensic science laboratories in Indiana, Illinois and Nevada that were associated with state or large metropolitan agencies. They had an average of 15.3 years working as latent print examiners and were an average of 45 years old. There were 4 men and 2 women. Four of the six had trained other examiners. The novices were recruited from the Indiana University student body, had no prior experience with latent prints, and tended to be younger with a mean age of 23 years. There were three men and three women.

All participants were tested according to the procedures of the Human Subjects Protection committee of Indiana University.

Procedures

Participants were seated approximately 36 inches away from a 17" inch LCD monitor set at a resolution of 1024x768 pixels. The images were scaled so that the latent and inked prints together filled the horizontal dimension, with 138 pixel horizontal borders on the top and bottom of the prints. The monitor was part of a model 1750 Tobii eyetracking system (Tobii Technology) which uses infrared cameras positioned on the monitor to track the position of the eye gaze by monitor both eyes. After a calibration procedure that established the relation between the observer's eye position and positions on the screen, the participant was shown pairs of prints and asked to determine whether they came from the same source. They were given up to one minute to make this determination, and if they came to a conclusion sooner they stated this conclusion and proceeded to the next image. An experimenter recorded their response as either identification or exclusion. Due to a disk crash and a corrupted backup disk, the responses for three novices were lost.

Methods for Inked/Inked comparisons

The methods below describe how we collected data on inked/inked comparisons.

Stimuli

We created 30 pairs of images by obtaining clean impressions from volunteers at Indiana University. These were made using ink stamp pads and placed on glossy ink jet photo paper. These were scanned using an Epson 4870 scanner at 1200 pixels per inch, and then cropped so that they fit inside a rectangle of 840x1050 pixels. Two image pairs were selected to be non-matching. The non-matching prints were selected by left-right reflecting or image reversing a print from the same finger of the other hand. These often have the same general ridge flow but will differ in the exact ridge details.

Participants

We tested 6 expert and 7 novice participants, including one expert who had participated in Experiment 1. The experts were recruited at forensic identification conferences in Michigan, Illinois and Indiana, while the novices were members of the Indiana University community. The mean age of the experts was 43 and the mean number of years of experience was 9.3. There were 5 men and 1 woman. All had self-reported 20/20 vision (corrected or uncorrected).

The novices had a mean age of 32 years and there were 4 men and 3 women.

Procedures

Participants were seated approximately 60 cm (~24 inches) away from a 21" LCD monitor. To ensure the quality of the data we used a chinrest for some participants to reduce head movements. Participants wore a head-mounted eyetracker which uses two small cameras to monitor the eye and the view of the scene respectively according to the hardware proposed by (Babcock & Pelz, 2004). Both cameras are mounted and specially located on a pair of lightweight safety glasses. One infrared light is located next to the eye-camera in order to illuminate the eye properly. This light provides us a constant spot of white light known as the first corneal reflection, which will be used for further offline analysis using the 'ExpertEyes Software', an open source approach for analyzing eye-tracker data (http://code.google.com/p/experteyes/wiki/ExpertEyes) developed by our research group.

Using this setup we record the streaming from both cameras, which is split later into image sequences. These images are used by the two modules of our software for further temporal alignment, calibration and gaze estimation. The first module uses the images from the eye stream in order to calculate the relationship in time between the pupil and the corneal reflection and fit the eye model. The second module uses the images from both streams and the eye model data to synchronize and calibrate both streams.

The ExpertEyes eyetracking system allows the computation of the average error of the eye tracker. The Tobii system reports the average error of 0.5 degrees of visual angle under typical use. We found that our eye tracker produced values in a similar range. The mean error for experts was .71 degrees, while the mean error for novices was 0.60 degrees. These values were not significantly different for the two groups ($t(11) = 2.02$; $p > 0.05$). Thus we are confident that our eye tracking results from the inked/inked comparisions are comparable in accuracy to those of the commercial system used for the latent/inked comparisions.

Figures

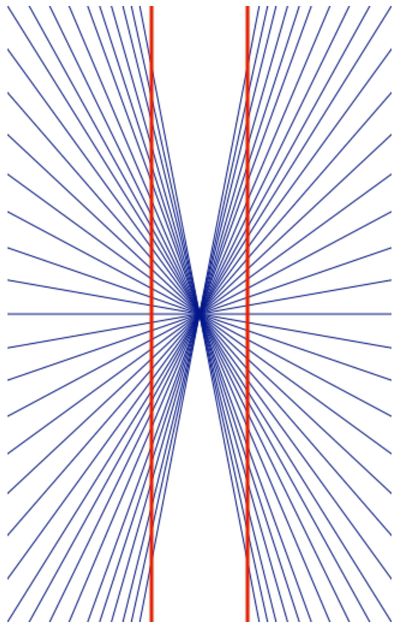

Figure 1. Example of a perceptual illusion that cannot be overcome through knowledge. The red lines are actually parallel, but knowledge of this does not make them appear straight. However, a sheet of paper next to the line illustrates that it is straight.

Figure 2. Our eye tracker worn by a participant in the eyetracking studies. It consists of one camera that monitors the position of the eye relative to the head, and a second camera that monitors the position of the head relative to the images being examined.

Figure 3. Example eye model fit which estimates the locations of the pupil (dark circle) and corneal reflection (white circle). The left panel shows the masked eye camera frame, while the right panel shows the fit of the eye model, which adjusts the locations of the black and white ovals to find the best overlap with the pupil and corneal reflection. The centers of each oval define the position of the eye.

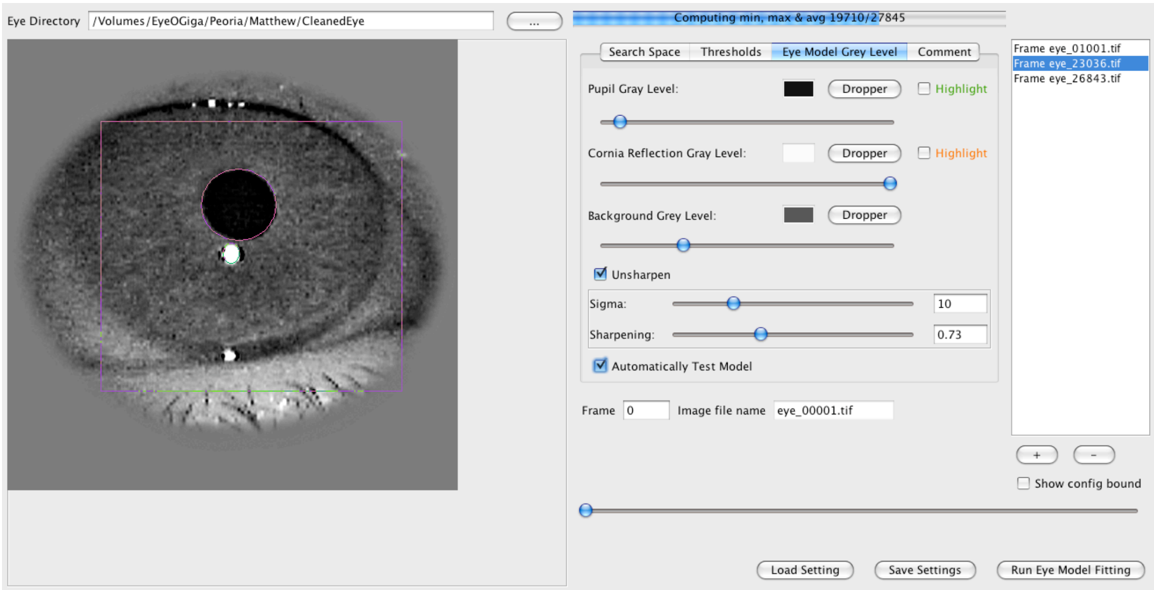

Figure 4. Forward eye model interface. The pupil (dark circle) and corneal reflection (white circle) are identified using a forward eye model that adjusts the parameter settings based on the location of the pupil. Once a set of parameters is found for each region, the program fits the eye model to the entire dataset, which can take 2-16 hours of CPU time on a 3 GHz computer.

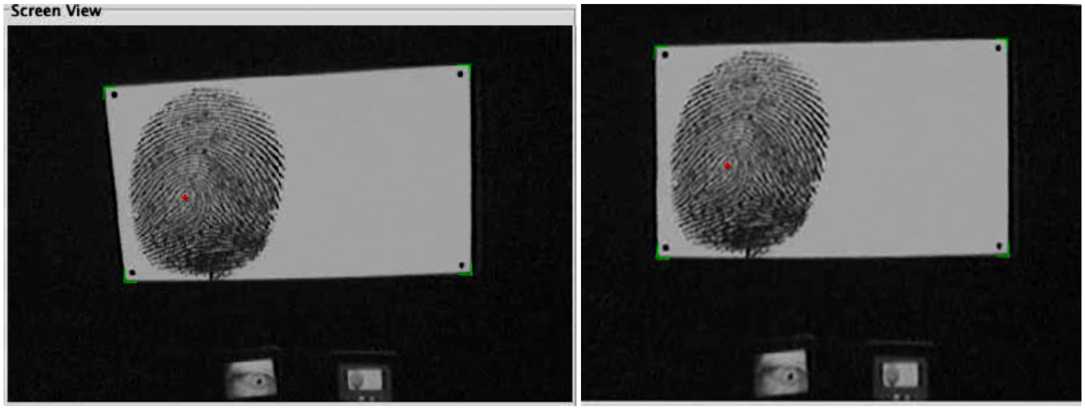

Figure 5. Eliminating image distortions. When the head-mounted video camera is positioned away from the center of the monitor, image distortions result. In this case the upper-right corner of the monitor in the image on the left is distorted. We created an undistortion algorithm that eliminates these distortions and allows us to accurately project the eye location onto the original image (right panel).

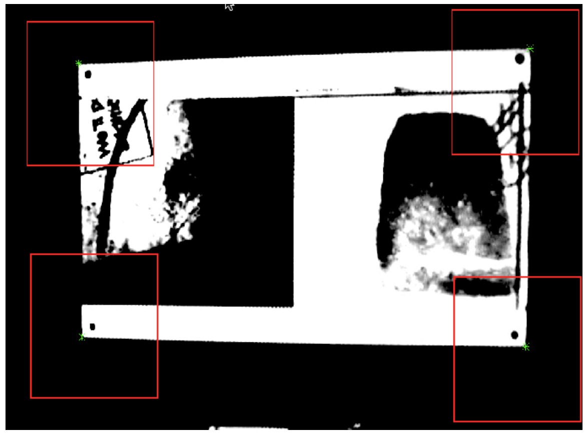

Figure 6. Search algorithm to identify the corners of the monitor in the scene camera after barrel distortions have been removed from the image. Red squares indicate our search space and green marks indicate the identified corner regions.

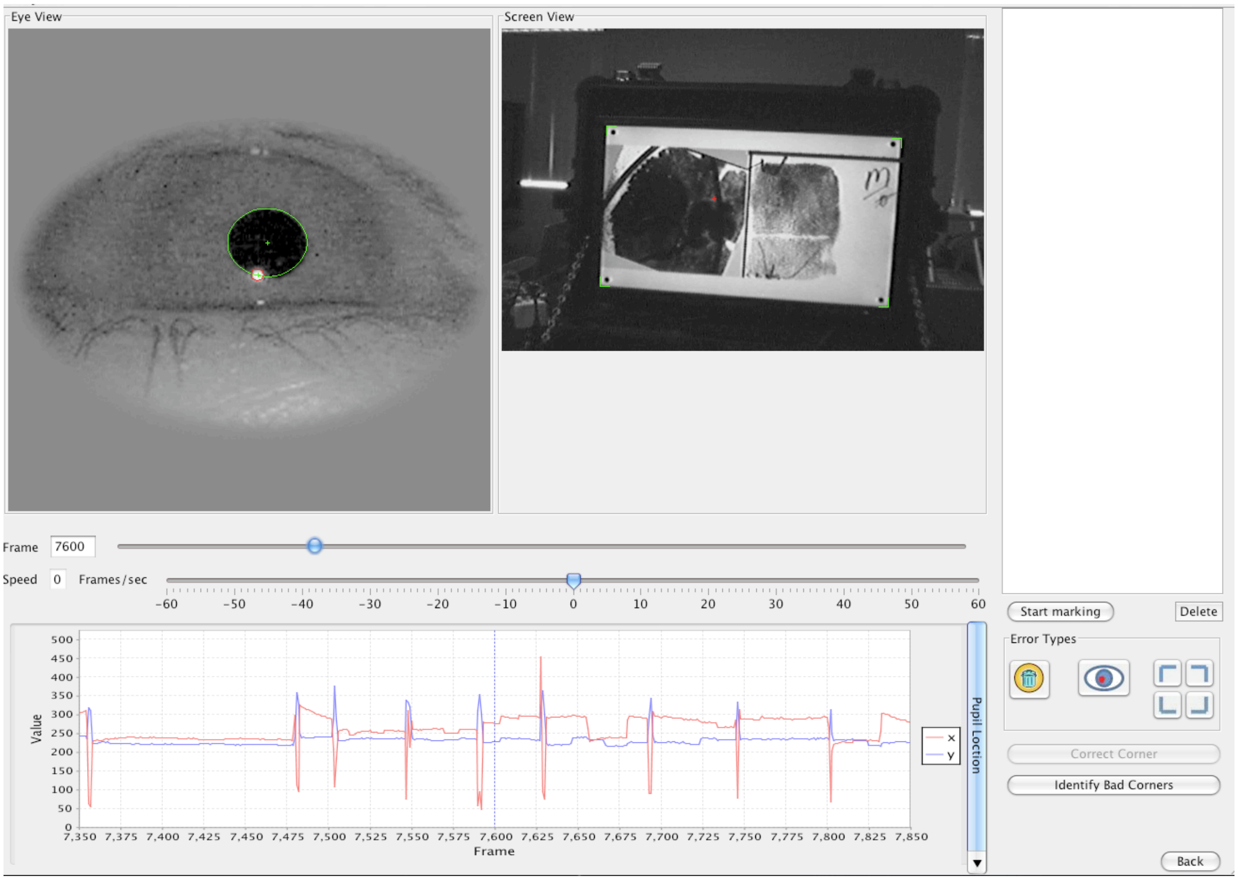

Figure 7. Screen that plots the eye data, along with the fitted pupil and corneal reflection parameters, along with the estimated eye gaze location in the scene camera (red plus). Graphs along the bottom show the pupil location for frames around the current frame. The large discontinuities in the blue and red curves are eye blinks. This particular screen allows the user to identify and correct portions of the data that have missing or incorrect data.

Figure 8. Gaze data from one subject. Green dots are raw gaze estimates, red dots are fixations that are determined using a velocity-based measure, and blue lines are saccades from one location to another.

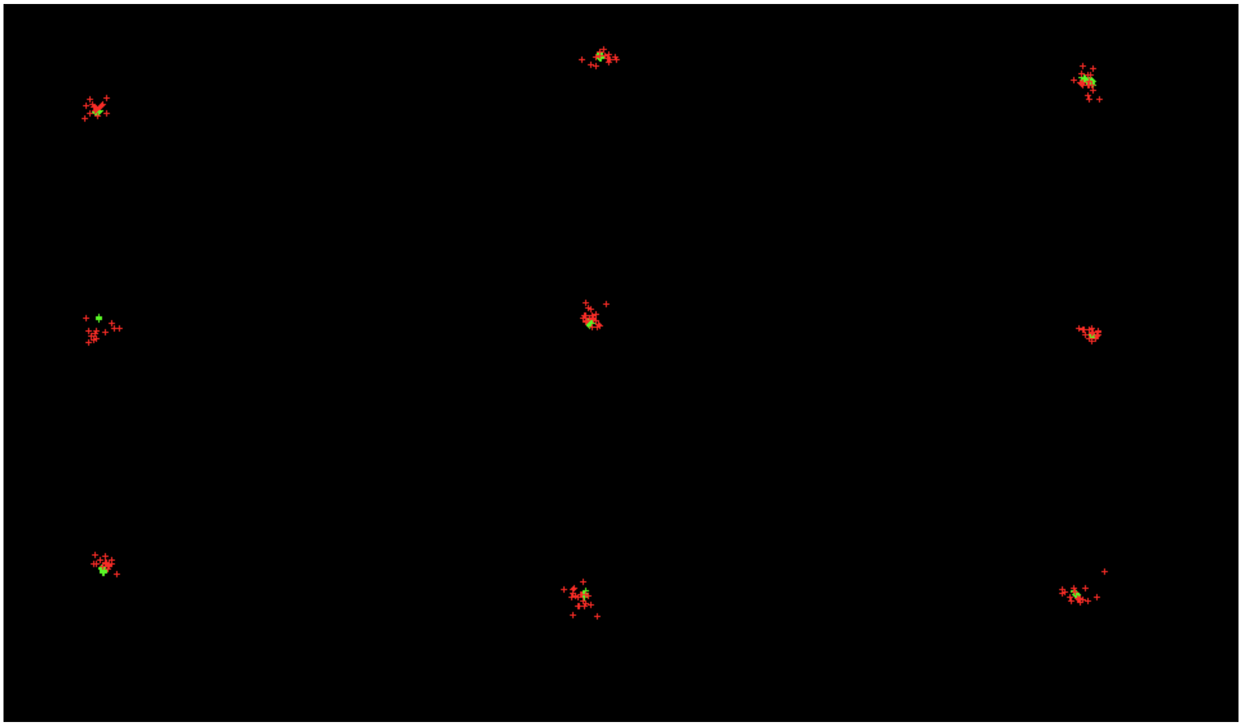

Figure 9. Results of the calibration procedures. Green dots are locations on the monitor that were fixated, and red dots are the estimated eye positions. The close clustering of each set of red dots illustrates the accuracy of our system, which is comparable to commercial systems. We typically see error on the projected monitor of less than 1 degree of visual angle and often less than .5 degrees, which is similar to commercial systems such as the Tobii system.

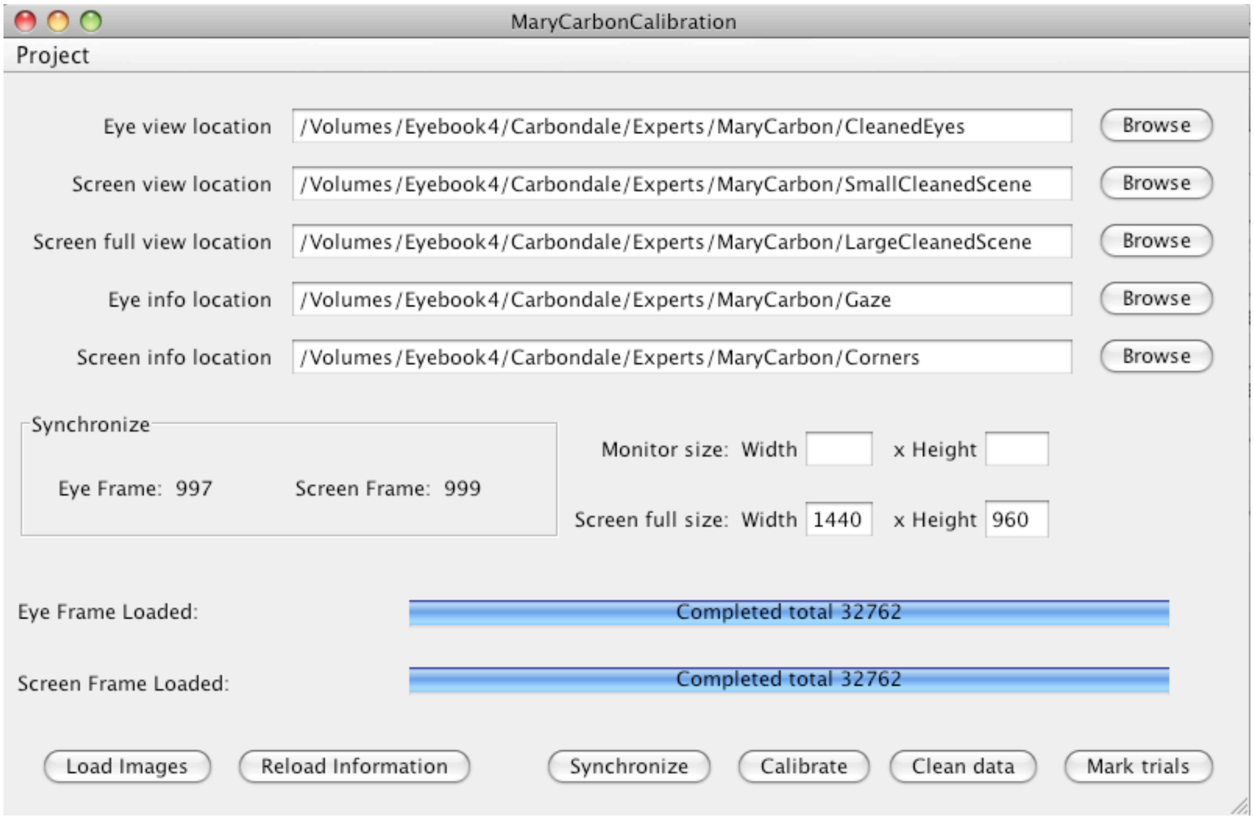

Figure 10. System integration screen which binds together folders containing scene information, eye information, gaze and corner information, and allows the user to synchronize the video streams, calibrate the gaze information, clean bad frames, and mark events from the experiment.

Expert Data **Novice Data**

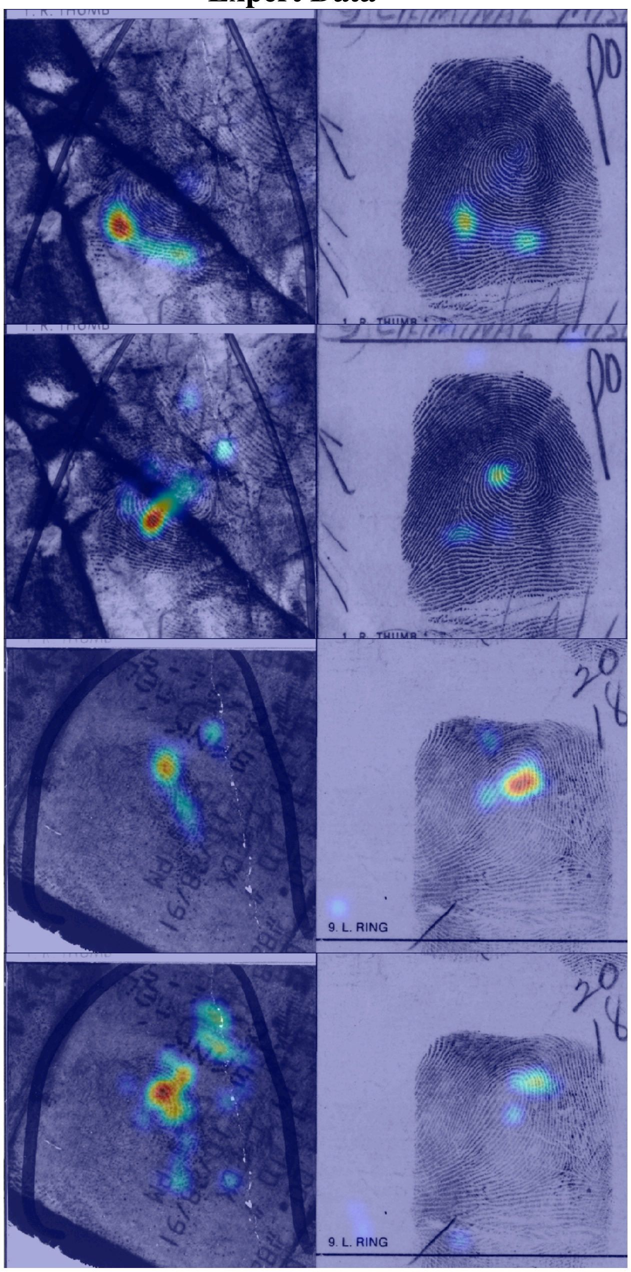

Figure 11. Examples of heat maps. Color indicates the amount of time spent at each location. The left column contains data from a latent print expert on two pairs of images, while the right column contains data from a novice. In general, experts show much tighter concentrations of fixations around relatively few locations, while novices direct their attention to many more locations.

Figure 12. Graphical illustration of counting nearby minutiae. The Universal Latent Workstation was used to computer-code the locations of the minutiae, shown here as green crosses. We then determined the fixations for each user (shown as red crosses) and counted the number of minutiae contained in a circle centered on each fixation (green circles).

Figure 13. Graphical illustration of fixations of all novices (in blue) and all experts (in red) for one representative image pair. The expert fixations tend to focus at and below the core, while novice fixation tend to distribute along the top of the fingerprint. The core tends to have many more minutiae, leading to a difference between the groups in terms of the number of nearby minutiae.

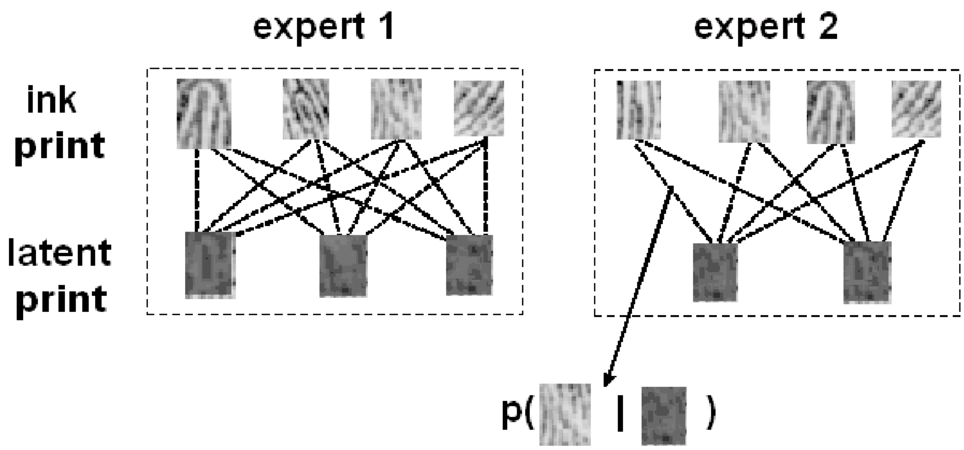

Figure 14. Parallel translation allows for the inference of correspondences between latent and inked prints.

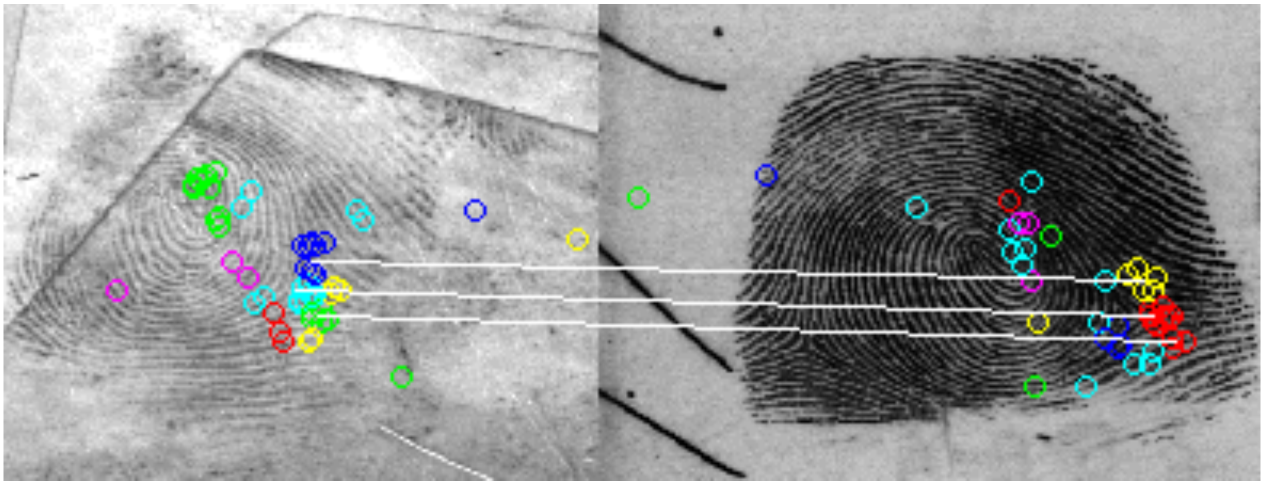

Figure 15. Fixations (circles) are clustered based on spatial location and a distance-based criterion. Fixations that fall in the same cluster are colored using the same color. Machine translation algorithms can be used to assign correspondences between clusters in the two prints by assuming that experts are looking at corresponding locations one after another. The machine translation algorithm uses the temporal information in the gaze sequence to identify corresponding clusters.

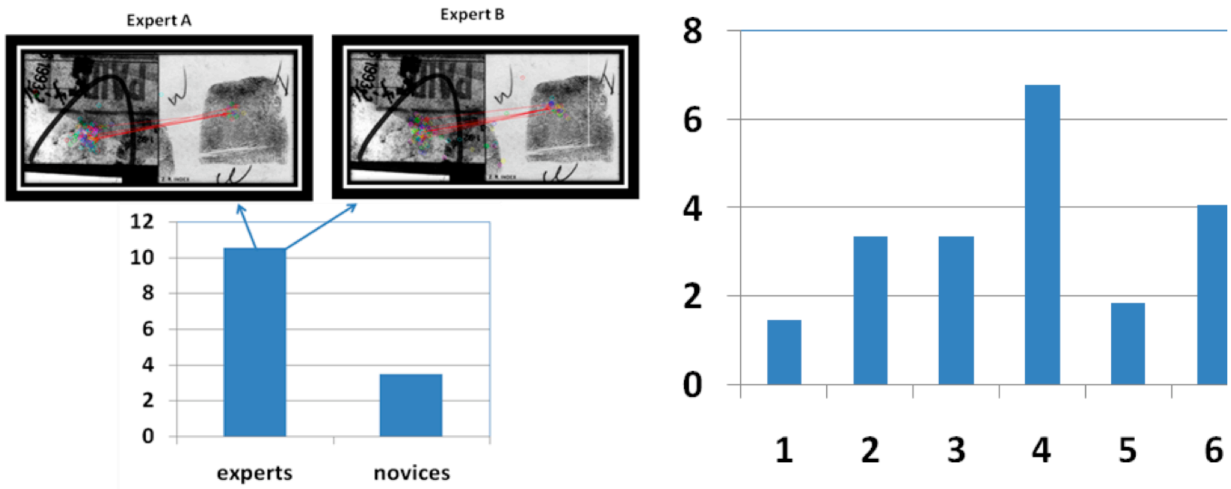

Figure 16. Left panel: Results of the machine translation analysis demonstrating that experts had many more correspondences than novices. Right panel: Histogram of the number of correspondences for each of our 6 novices. There is an outliner (number 4) who turns out to be a graduate student working on the fingerprint project and therefore not a true novice.

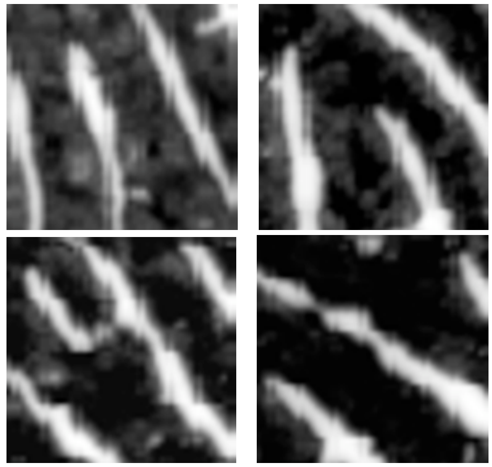

Figure 17. Example image window patches extracted at eye fixation locations that can be used to induce the nature of the features inspected by examiners.

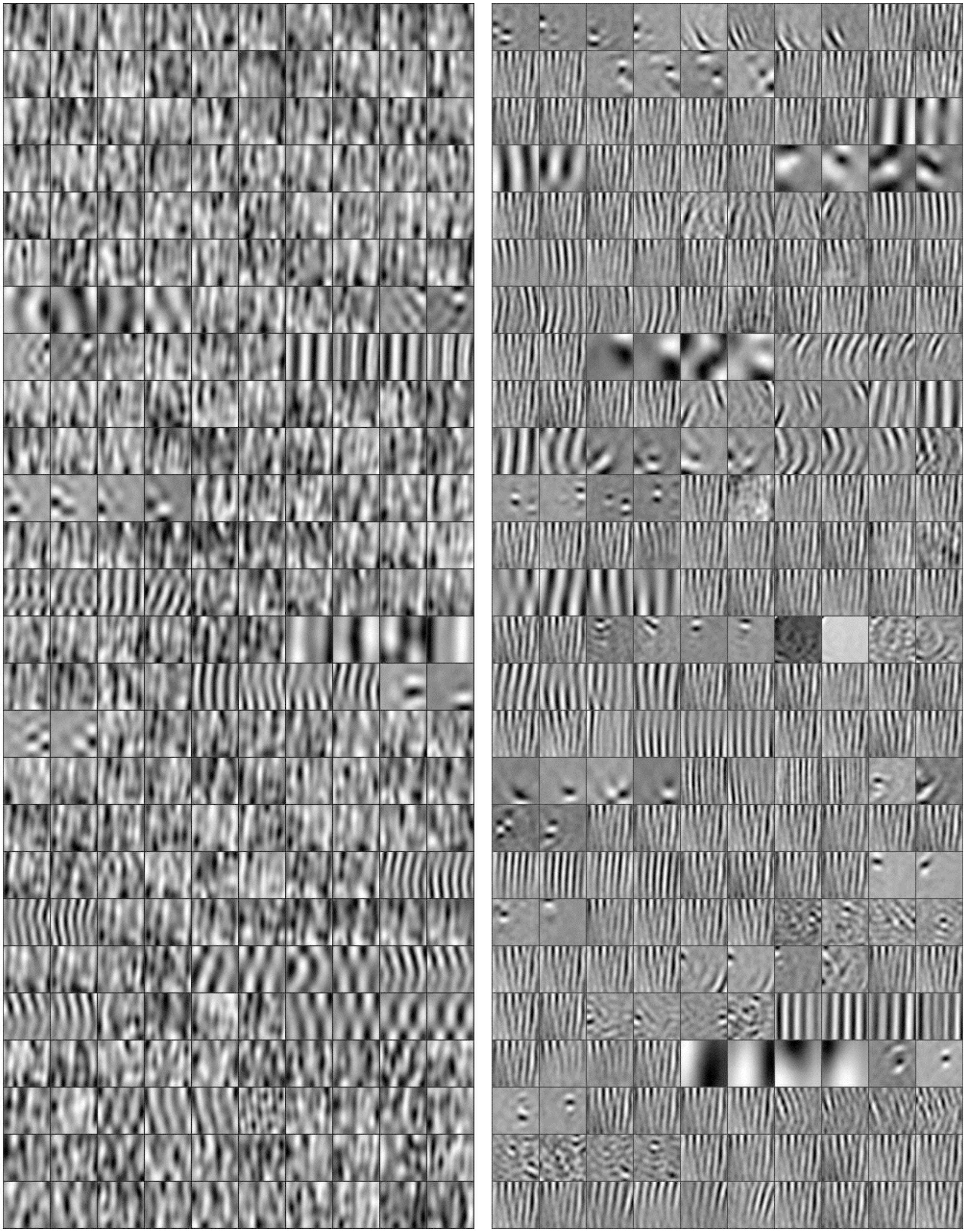

Figure 18. Feature basis set derived from expert data (left) and novice data (right).

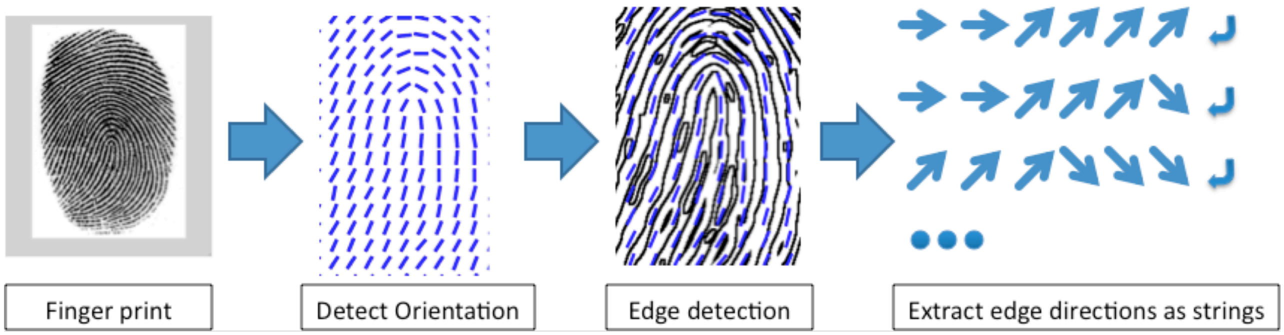

Figure 19. The process of converting fingerprint ridges to strings of direction orientation information and converting edges to strings of orientations.

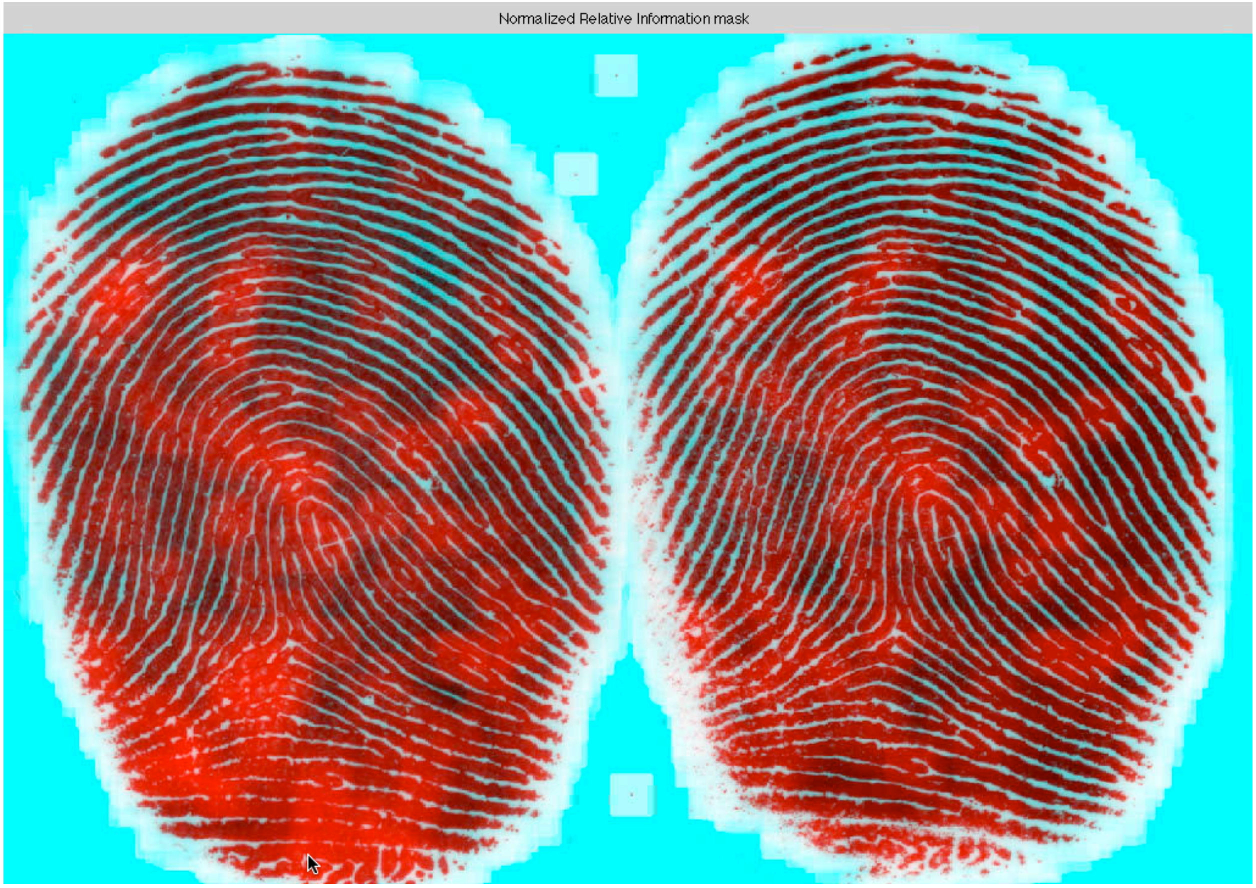

Figure 20. Amount of information contained at each location, as determined by the information contained in relative angle measurements. The amount of red at each location is a visualization of how much information that location contains by this measure. This particular measurement looks at the orientation changes that occur along a ridge.

Figure 21. The information contained in the absolute orientation of ridge elements. Red regions are those that are more informative by this measure.

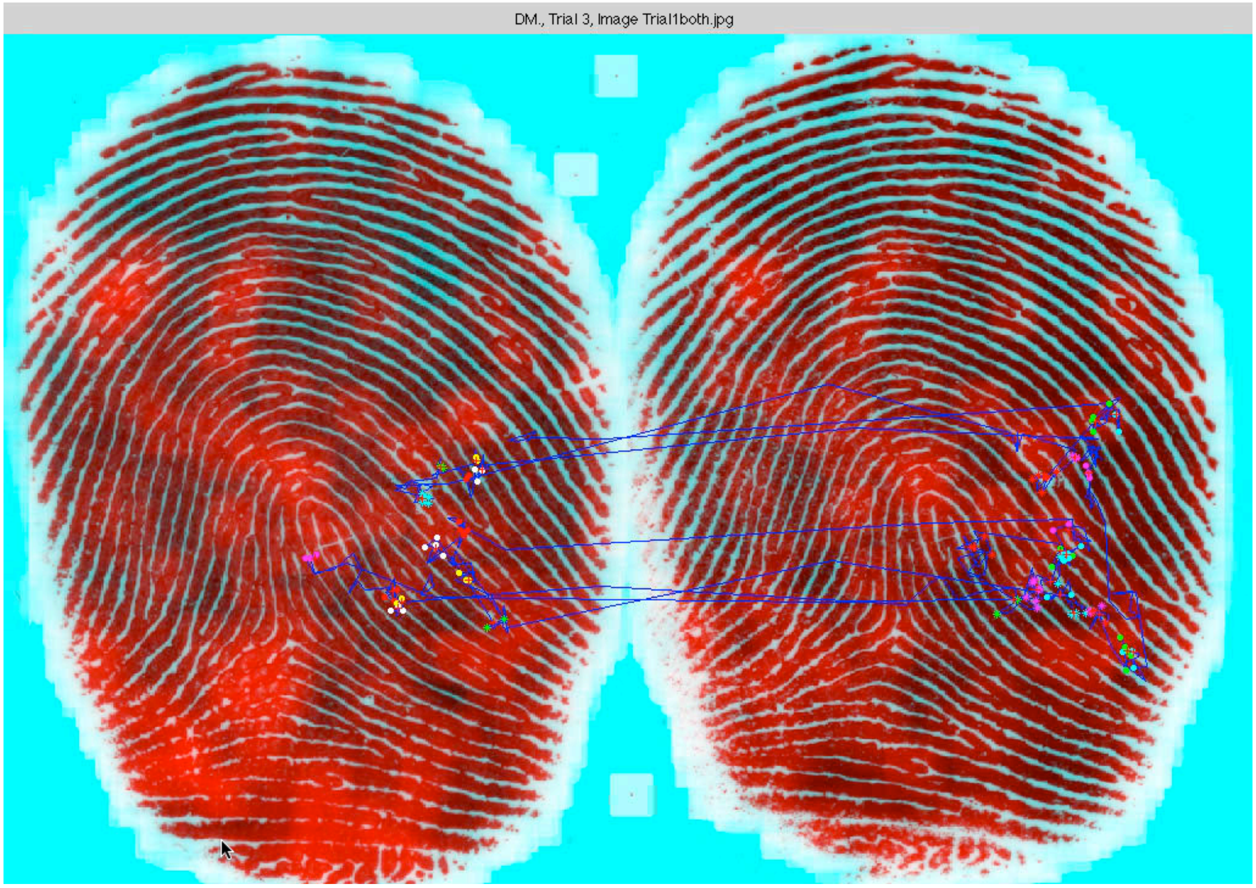

Figure 22. Fixation data from one expert overlaid on one information metric (relative information). Fixations tend to occur in higher-information regions.

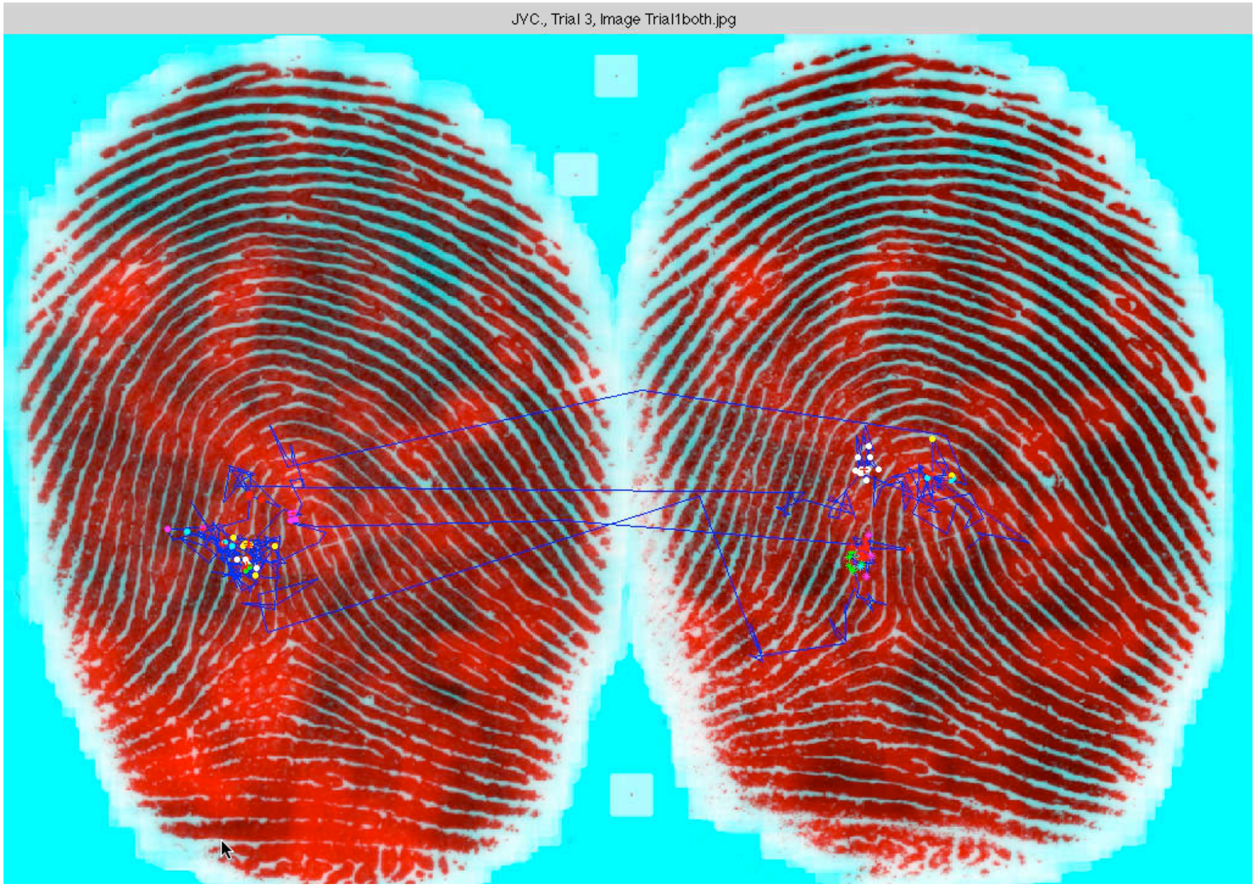

Figure 23. A second expert on the same set of prints and information metric as Figure 22. Again we see the fixation from experts in higher-information regions.

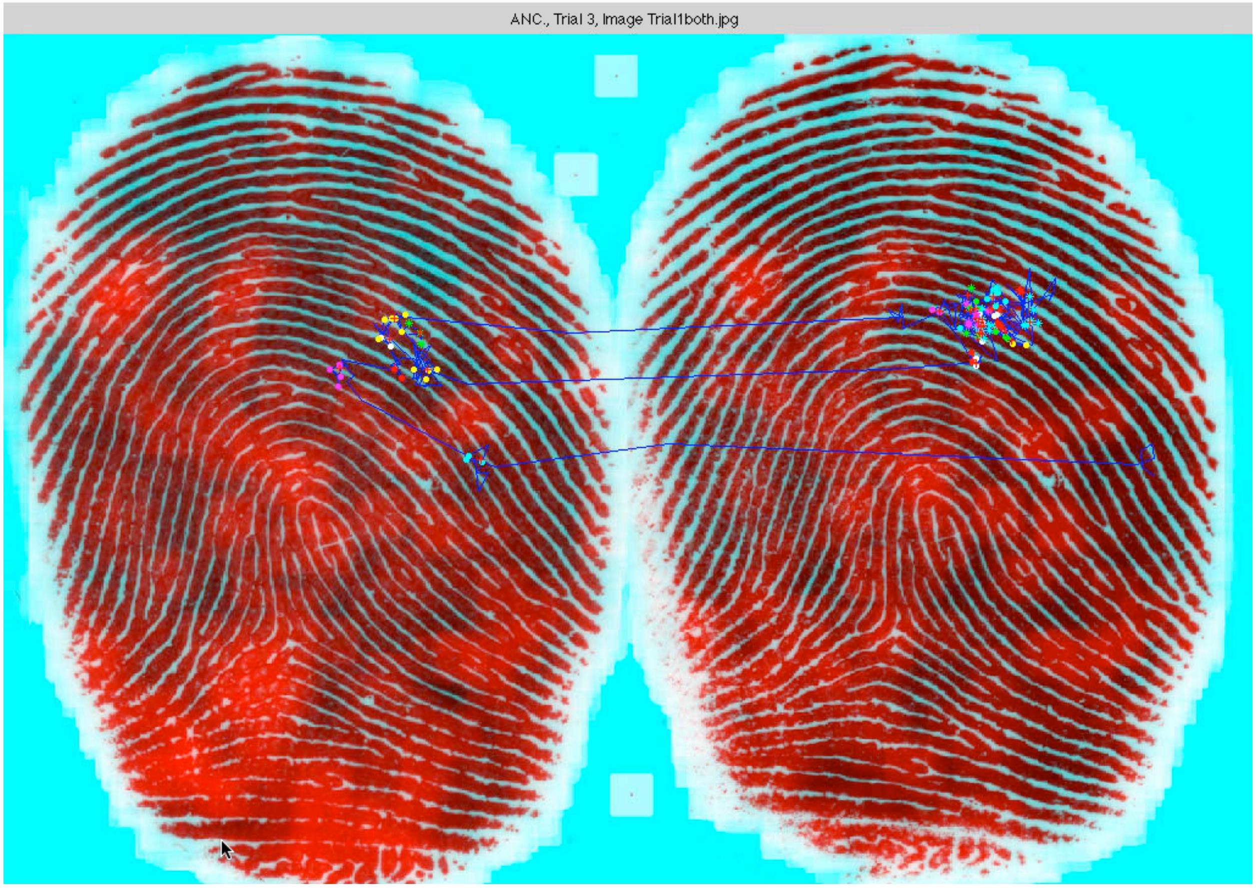

Figure 24. Data from a novice on the same set of prints and information metric as Figure 22 and Figure 23. The fixations appear to land in lower-information regions.

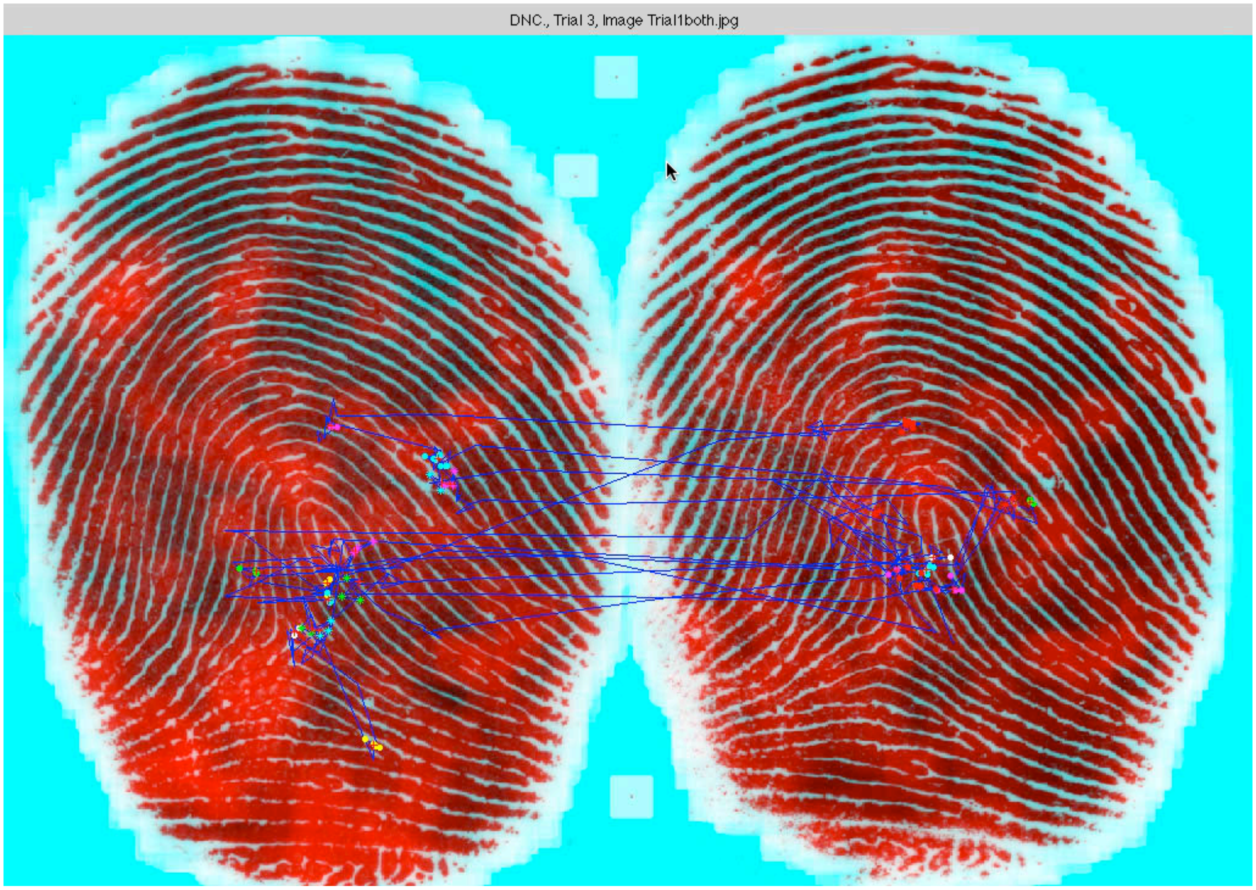

Figure 25. Similar to Figure 24, this is data from a second novice with fixations that tend to land in lower-information regions.

References

Ahissar, M., & Hochstein, S. (1997). Task difficulty and the specificity of perceptual learning. *Nature, 387*(6631), 401-406.

Babcock, J. S., & Pelz, J. (2004). *Building a lightweight eyetracking headgear*. Paper presented at the ETRA 2004: Eye Tracking Research and Applications Symposium.

Busey, T. A., & Vanderkolk, J. R. (2005). Behavioral and electrophysiological evidence for configural processing in fingerprint experts. [Article]. *Vision Research, 45*(4), 431-448.

Busey, T. A., Yu, C., Wyatte, D., Vanderkolk, J., Parada, F., & Akvipat, R. (Submitted Manuscript). Consistency and Variability Among Latent Print Examiners as Revealed by Eye Tracking Methodologies. *Journal of Forensic Identification*.

Dror, I. E., & Charlton, D. (2006). Why experts make errors. *Journal of Forensic Identification, 56*(4), 600-616.

Dror, I. E., Charlton, D., & Peron, A. E. (2006). Contextual information renders experts vulnerable to making erroneous identifications. *Forensic Science International, 156*(1), 74-78.

Dror, I. E., Peron, A. E., Hind, S. L., & Charlton, D. (2005). When emotions get the better of us: The effect of contextual top-down processing on matching fingerprints. *Applied Cognitive Psychology, 19*(6), 799-809.

Dyer, A. G., Found, B., & Rogers, D. (2008). An insight into forensic document examiner expertise for discriminating between forged and disguised signatures. *Journal of Forensic Sciences, 53*(5), 1154-1159.

Egli, N. M., Champod, C., & Margot, P. (2007). Evidence evaluation in fingerprint comparison and automated fingerprint identification systems- Modelling within finger variability. *Forensic Science International, 167*(2-3), 189-195.

Hong, L., Wan, Y. F., & Jain, A. (1998). Fingerprint image enhancement: Algorithm and performance evaluation. *Ieee Transactions on Pattern Analysis and Machine Intelligence, 20*(8), 777-789.

Krupinski, E. A. (1996). Visual scanning patterns of radiologists searching mammograms. *Academic Radiology, 3*(2), 137-144.

Neumann, C., Champod, C., Puch-Solis, R., Egli, N., Anthonioz, A., & Bromage-Griffiths, A. (2007). Computation of likelihood ratios in fingerprint identification for configurations of any number of minutiae. *Journal of Forensic Sciences, 52*(1), 54-64.

Portilla, J., Strela, V., Wainwright, M. J., & Simoncelli, E. P. (2003). Image denoising using scale mixtures of Gaussians in the wavelet domain. *Ieee Transactions on Image Processing, 12*(11), 1338-1351.

Reneau, R. D. (2003). Unusual Latent Print Examinations. *Journal of Forensic Identification, 53*(5), 531-537.

Shannon, C. E., & Weaver, W. (1949). *The mathematical theory of communication*. Urbana,: University of Illinois Press.

Snodgrass, M., Bernat, E., & Shevrin, H. (2004). Unconscious perception at the objective detection threshold exists. *Perception & Psychophysics, 66*(5), 888-895.

Srihari, S. N., & Su, C. (2008). *Computational Methods for Determining Individuality*. Paper presented at the Proceedings of the 2nd international workshop on Computational Forensics.

Su, C., & Srihari, S. N. (2008, December 8-11, 2008). *Generative Models for Fingerprint Individuality Using Ridge Models*. Paper presented at the International Conference on Pattern Recognition, Tampa, Fl.

Vanselst, M., & Merikle, P. M. (1993). Perception Below the Objective Threshold. *Consciousness and Cognition, 2*(3), 194-203.