

The author(s) shown below used Federal funds provided by the U.S. Department of Justice and prepared the following final report:

Document Title: The Information Content of Friction Ridge Impressions as Revealed by Human Experts

Author(s): Thomas Busey, Chen Yu

Document No.: 244567

Date Received: January 2014

Award Number: 2009-DN-BX-K226

This report has not been published by the U.S. Department of Justice. To provide better customer service, NCJRS has made this Federally-funded grant report available electronically.

Opinions or points of view expressed are those of the author(s) and do not necessarily reflect the official position or policies of the U.S. Department of Justice.

The information content of friction ridge impressions as revealed by human experts

Final Technical Report

NIJ grant #2009-DN-BX-K226

Thomas Busey and Chen Yu, Indiana University

This report describes the technical accomplishments of NIJ grant 2009-DN-BX-K226.

Executive Summary

The goal of this project is construct a quantitative representation of the information content in fingerprints. Such a representation has two benefits. First, it can reveal those areas that experts consider most diagnostic, which might help trainees or jurors decide the value of different regions. Second, it can help guide an expert as to how diagnostic a region may be if this diagnosticity differs from their expectations. For example, there are regions known as force pattern areas in which the collision of multiple ridges can create several minutiae that might appear similar in another print.

This quantitative approach takes on more importance as computer-based searches become more available and database sizes grow. If a print is developed based on evidence from a detective or informant, high similarity between a latent print and an exemplar from a suspect becomes very valuable evidence. However, if the suspect is developed through a database search using a system such as the Automated Fingerprint Information System (AFIS), high similarity between the latent and a candidate from AFIS should be viewed with skepticism. This is because AFIS, *by design*, must return similar looking prints (Busey, Silapiruti, & Vanderkolk, under review; Dror & Mnookin, 2010). Thus a quantitative representation of the information content in fingerprints would help address the issue of which regions are

most diagnostic, *given the statistics of the entire database*.

The challenge is to discover a feature representation that will allow statistical or quantitative analyses. The traditional approach has been to rely on the locations and orientations of minutiae, possibly including details such as the number of intervening ridges. This approach has proven to be quite promising, and produced statistical likelihood estimations that help identify the evidentiary value of a particular print (Egli, Champod, & Margot, 2007; Neumann et al., 2007).

Other approaches rely on computing the statistics of individual minutiae at locations relative to a standard landmark such as the core region. This allows for a generative model of fingerprint information that characterizes the distribution of minutiae across the fingerprint (Srihari & Su, 2008; Su & Srihari, 2008).

One limitation of the reliance on minutiae is that examiners report using features other than y-branchings and ridge endings. For example, they rely on overall pattern shape, ridge curvature, and even the shapes of individual pore elements that make up the ridges. However, defining what a feature is has proven extremely difficult, in part because much of perception is below the level of conscious awareness and is difficult to verbalize (Snodgrass, Bernat, & Shevrin, 2004; Vanselst & Merikle, 1993). This is self-evident by looking at a perceptual illusion. Despite the fact that the observer knows that their percept deviates from the actual image, they cannot use this knowledge to veridically

perceive the illusion. Thus we cannot simply ask examiners what information they rely on and expect to receive a complete answer, although in some instances talk-aloud procedures can and do provide useful information.

Because of the difficulty with verbalization of perceptual data, our approach collects eye gaze data from latent print examiners. We will use this gaze data to infer what regions or sources of information they consider most diagnostic or distinctive. These results will then be used to develop quantitative metrics that characterize the information available in latent prints. We will in fact describe two approaches that serve complimentary purposes, but each of which is validated against expert eye gaze data.

In this technical report we describe the nature of the eye tracking data collected from experts and novices, how we use this to train a set of intermediate-level descriptors known as *basis functions*, and how we can use the activations of these basis functions to provide a quantitative description of the information contained in friction ridge impressions.

This report is organized as follows. First, a major contribution of this project is the development of robust eye tracking methods that allow us to collect eye gaze data in the field as examiners conduct tasks that are similar to casework. Thus we will discuss the development of the tools that allow this data collection.

Second, we discuss how we construct an intermediate level feature representation that provides a means to represent image detail from both trained and novel fingerprint images. This intermediate representation is similar to an alphabet in language, in the sense that it contains elemental features that can be combined together to build more complex objects. We will use principled approaches to the construction of the basis set, which reflects the natural scene statistics of fingerprint

images as well as those regions that experts consider most important.

Third, we will use the temporal information contained in the eye gaze data to explore how experts knit together different regions into larger constructs, which they term ‘target groups’.

Finally, we describe extensions to the basis function modeling approach that learns the covariance among the activations. This is similar to second stage of visual processing in visual system (V2/V3) and therefore has a measure of both biological and computational plausibility.

Together the analyses provide estimates of the rarity of individual features, and do so by using the statistics of fingerprint impressions rather than a hand-picked set of features. This provides access to potentially a larger set of visual features than individual minutiae.

A. Eye tracking data collection and analysis

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.

Although commercial eye trackers are available, they are closed source and at the onset of the project were not very portable. We took advantage of the fact that small high-resolution cameras became available and thus we built our own eye tracker and software. Because this is open source and made cheaply from readily-available parts, these techniques can be duplicated by other researchers who are

interested in collecting eye tracking data in forensic disciplines.

The eyetracker is illustrated in Figure 1. One camera records the position of the eye relative to the head, and the other camera records the position of the head relative to the computer monitor that presents the fingerprints. As part of this project we developed the ExpertEyes software (<http://code.google.com/p/experteyes/>), which allows us to align the two video streams in time and establish a set of correspondences between the eye position and the scene camera. The output of this eye tracker is an estimate of the gaze location on each video frame captured during the experiment. Our latest eye tracker is very high definition and provides a resolution of 1280x720 pixels at 60hz, as shown in Figure 2.

The eye gaze data allows us to infer, at each point in time, the visual features that the expert considers to be diagnostic for the fingerprint examination task.

We have collected data from several different sets of images. Typically we record for about 20 minutes for each subject, and they do 30-39 trials during that period. Each trial consists of a latent print and an inked print shown side by side. We sometimes create simulated latent prints by adding noise to an inked print, as shown in Figure 3. This gives us access to the ground truth image detail while still asking the participant to conduct a procedure that is similar to latent print work.

Each experiment consists of at least 12 novices and 12 experts, and the eye gaze record consists of over 500,000 datapoints.

A.1. Hardware and recording devices

Our eye tracker consists of a set of cameras mounted on a set of safety glasses (see Figure 1). 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 4 demonstrates the software that fits the forward eye model.

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,

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 final 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 6 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.

Figure 7 shows the raw eye trace along with the results of the fixation finding algorithm.

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. 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.

B. Intermediate Feature Level Descriptions- Basis Functions

Eye tracking data tells you where the subject moved their eyes, but not necessarily what features or visual information they rely on. Some of this information is even below the level of visual awareness and is difficult to verbalize (Snodgrass et al., 2004; Vanselst & Merikle, 1993). As a result, we must use computational techniques to infer the nature of the visual information used by human experts. We do this by borrowing from the known computational properties of the visual system, which tends to break up the visual scene into individual components and then build back up to a larger, more complete, representation of an object or scene. This requires using a set of elemental features called *basis functions*, which are analogous to phonemes or morphemes in language as described next.

B.1 Analogies from spoken language

The basic premise behind dimensionality reduction and machine classification is to derive the basic building blocks or features that are used in the perception and matching of latent prints. Consider spoken language as an analogy. Words are made up of smaller units called phonemes, which from the Greek means "a sound uttered." The /k/ in *kit* is one example, although the [k] in *skill* is a different sound despite being the same letter. There are a little over 40 of these phonemes in the English language, and all words are produced by combinations of these phonemes.

What is surprising about phonemes is that even as young children, humans have a remarkable ability to differentiate different phonemes, and more importantly recognize that two identical phonemes spoken by different talkers are really the same speech sound and therefore convey the same meaning. A process called categorical

perception enables this form of speaker-independent. Once the categories for phonemes are formed, the phoneme can be recognized in noisy environments and from different speakers.

The downside to categorical perception is that the phoneme structure can be difficult to alter. This produces the common problem that native Japanese speakers have difficulty initially distinguishing between the /r/ and /l/ phonemes. This can be overcome with practice, which serves to reorganize the category boundaries between /r/ and /l/ so they no longer map to the same perceptual category.

B.2. Applications to visual perception

The same processes that allow the formation of categories in spoken language may also be at work in visual perception. Studies of the neural anatomy of the visual system suggest that the visual world is broken down into small pieces by the early stages of the visual system, and more complicated structures are then constructed from these simple building blocks. Importantly, the process of learning to see, perceive and interpret the visual world depends on learning the relation between these parts, as well as the likelihood of perceiving one feature given the presence of a related feature. For example, if you perceive the left half of a face, your experience tells you that you are extremely likely to see an eye on the right side of the face. In fact, not seeing an eye on the right side violates your expectation in such a way that your brain can often interpret this as damage, which forms the basis of many horror imagery. We internalize this missing eye as a result of imagery and seek to avoid this happening to us.

We infer the initial building blocks of visual perception by acquiring small pixel patches of fingerprint images that are centered on the eye fixations of experts and novice

subjects. Our working dataset contains about 22,000 fixations per subject group, which gives us about 44,000 pixel patches that are small, 38x38 pixel crops from the larger fingerprint images. These contain between 3 to 8 ridges depending on the scale of the images. This process is shown in Figure 8. These images represent the type of features that experts and novices fixate, and we will use dimensionality reduction procedures to find commonalities between the features.

Figure 9 illustrates example features extracted from experts, as well as features that come from random locations that still contain a majority of ridge detail. What is common among these features on the left that is different from those on the right? We start with the assumption that there is a common set of features that experts look at. These features may be decomposable into an even simpler feature set that provides regularities that can be discovered with the right algorithm. This is analogous to phonemes in spoken language described previously.

We will use a procedure called independent components analysis (ICA) to discover these elemental features, called the basis set. This will eventually provide a dimensionality for machine classification as a means to test whether our basis set is meaningful.

A basis set is simply a linear combination of pixel locations. It tells you how important each pixel is to that latent component. Typically there are fewer basis functions than pixels (we have 1444 pixels in each patch).

ICA has two general properties. It looks for a basis set that:

- 1) minimizes the mutual information between individual components, such that the information that one basis image tells you about the feature is independent from the information from other basis images.

- 2) maximizes the non-Gaussianity of the transformed (projected) data.

The central limit theory says that as you add non-Gaussian sources (like uniform or sparse distributions), the resulting signal looks more and more Gaussian.

ICA essentially reverses this to look for decompositions that provide signals that are as non-Gaussian as possible. These could be the original signals that were added together.

Applied to vision, you get two amazing facts for free:

- 1) The resulting components look very much like the receptive fields in the early visual system

- 2) The ICA components are sparse, just like neurons in the brain, in that a given stimulus tends to activate a relatively few neurons to a large degree, while the rest are not activated at all.

When we apply ICA to the image patches from experts and random patches, we obtain a basis set that looks like those shown in Figure 10. This basis set resembles fingerprint fragments, and while these individually do not contain features such as bifurcations or ridge endings, they can easily be combined to reproduce the original features (with some error) as shown in Figure 11. As such, these represent a reasonable starting point for a basis set that is designed to represent the common elements of expertise among experts. The relatively good reconstruction also illustrates that although we are reducing the dimensionality of the images, we still maintain enough information in the reduced representation to allow a fairly accurate reconstruction.

C. Region Clustering

The first step to evaluating our basis functions is to perform clustering on the activations.

C.1. Automatic Feature Segmentation of Fingerprints

Once a basis set has been selected and derived (similar to that shown in Figure 10), the basis set can then be applied to a new fingerprint and the activations of each basis function can be computed for each pixel location in the fingerprint. Essentially this is done by convolving the basis function with the image at each location. Each basis function in the basis set will then produce an *activation* at that location. Thus for an image of size 1024x768 and a basis set of size 16, the output is a 1024x768x16 matrix that contains all of the activations for each basis function at each location.

These basis activations are not random, but instead reflect the regularity of the image information. Each pixel is represented as a feature vector, which the value of the ICA activations filling the vector. So for a basis set of size 16, there will be 16 numbers for each pixel. This places the pixel in a 16-dimensional space. The locations of these pixels in this 16-dimensional space is not uniform or random, but reflects the image data, and we would like to cluster similar regions based on the fact that they produce similar patterns of activity via the ICA basis functions.

We are using an approach called Expectation Maximization (EM) which attempts to cluster the data based on multidimensional Gaussian functions that are designed to cover the locations of the pixel values in ICA activation space. We specify a number of Gaussians (typically between 25 and 100) and cluster the feature vectors according to their Euclidean distance. To allow for generalization, we combine the data across almost 40 images, sampling only 2% of the image data from each fingerprint to allow the clustering solution to converge.

The output of this function is a set of labels associated with each pixel that assigns the most probable cluster to that pixel. We

visualize these clusters by changing the color of the pixel, as shown in Figure 12.

What is remarkable about this clustering solution is that it tends to find very contiguous regions despite not knowing anything explicit about space. It is somewhat sensitive to ridge orientation, which tends to co-vary with the different regions. However, it also accurately captures the core and delta areas which do not have well-defined orientations.

This clustering algorithm generalizes to other fingerprints as well. Figure 13 shows the same clustering solution applied to 6 different fingerprints. Although the location of the core and delta vary from print to print, it still accurately finds these regions in a totally automated procedure. This demonstrates that the clustering approach based on ICA weights can detect equivalent regions in different fingerprint impressions.

The clustering approach is also surprisingly accurate at detecting corresponding regions in two fingerprints from the same finger. Figure 14 illustrates the clustering algorithm applied to two impressions from the same fingerprint. The contours of the clustering indicate that the solution finds very similar regions and therefore has good generality to novel instances of the same finger.

Together, these examples illustrate the robustness of the clustering approach based on ICA activations. The next step is to condition our measure of feature rarity based on the region that contains the feature. This will introduce a context dependency that will measure the feature rarity *given a particular location*.

D. Initial Classification Attempts Based on ICA Basis Activations

As a starting point we used the raw activations of the basis function to try to separate experts from novices. We should say up front that while this is a traditional

approach in classification research, it will not produce good generalization performance and we therefore use the failure of this approach to motivate two new approaches that use combinations of different feature activations to produce maps of diagnosticity that provide much stronger classification and generalization performance.

Each region that is fixated by an expert can be processed by a basis set to produce a set of activations. The set of activations visited by experts may be different than that of random activations or novices. We used a support vector machine (SVM) classifier to classify regions visited by experts from randomly-chosen regions based on the ICA activations at each patch. The friction ridge image detail centered at a fixation are cross-correlated against the basis functions to produce a set of activations that places each image patch in a high-dimensional space. The classifier is then asked to find a hyper plane (possibly using a non-linear kernel transformation) that separates the expert from random fixations.

We separated the image patches into a training set and a held-out testing set that came from novel images. The SVM was able to accurately classify 77% of the training image patches. Although this classification results in fairly accurate separation of the training sets, it does less well when generalizing to a new testing set. It only classifies approximately 57% of the novel image patches that make up the testing set. Given that chance performance is approximately 50%, this represents only a modest level of performance, and limits the utility of this approach to generalize to novel prints.

The high dimensional nature of the activations (up to 400 dimensions) may limit the ability of the classifier to find an appropriate transformation that doesn't overfit the training set. This may have produced the spuriously-high training results and poor

generalization performance. Thus we need a more principled way to reduce the dimensionality. In the next section we explore one transformation that not only combines across the different dimensions in a principled way, but also provides statistics about feature rarity and diagnosticity. In a later section (Section G), we explore an even more promising model that uses the image statistics of friction ridge impressions to combine across dimensions.

E. Self-Information Metric

The goal of the self-information metric described below is to develop a quantitative metric of the information content in fingerprints using extensions from Information Theory. We will use the feature set described previously to estimate the likelihood of observing individual features. This will provide a measure of the diagnosticity of the ridge detail for purposes of individualization (in the sense that it identifies which regions are most rare given a data set of images). This measure is *independent* of human examiners, with the exception that we will combine across different spatial scales and basis set sizes using feedback from experts. However, individual diagnosticity maps are a function solely of the statistics of the data set.

We will use the data from experts to assess whether the self-information metric identifies regions that experts also fixate. This would be the case if experts have determined, either individually or as a field, which regions tend to be the most diagnostic. While individual self-information calculations based on one set of basis functions and pixel patch size are entirely parameter free, we have several choices for the pixel patch size and basis set size. We will bootstrap our way out of this problem by using data from experts to select which basis set size and pixel patch size is most appropriate.

Here are the inference steps:

Experts are very accurate at this task.

The self-information metric is a statistical property of the fingerprints.

The self-information metric produces rarity maps that (as we will demonstrate) agree with experts eye gaze.

If experts can distinguish between matching and nonmatching prints, and they rely on similar features that the model considers to be diagnostic,

This analysis will produce a quantitative representation of the information content in friction ridge impressions that is based primarily on the statistical properties of ridge detail, guided somewhat by the use of expert data to determine the best parameterization of the self-information metric.

Our eye gaze data for this particular validation set was collected from 24 expert examiners with at least 2 years of unsupervised casework experience (median is 6.3 years of experience). Examiners viewed pairs of images, one of which was combined with visual noise to create the impression of a latent print. Of the 39 image pairs, 5 were non-matching prints. Experts were told to conduct a task similar to a traditional latent print comparison, with the option to say “match”, “non-match” or “too soon to tell” which we typically interpret as “inconclusive.” Recording was typically limited to about 20 minutes to avoid lengthy IR exposure from the eyetracker on the eye.

E.1. Constructing Feature Rarity Maps

The failure of the classification approach using raw ICA activations suggests that the activation levels of several features might be important to distinguish between important and less important regions.

In the approach described below, we detail a technique that uses Shannon Information Theory along with the statistics of fingerprint ridge detail to provide a measure of feature rarity. Because the rarest features are

the most diagnostic, we will argue that this approach is a measure of feature diagnosticity. It is important to point out at the onset that this measure is an *independent* of human experts. It does not depend on where experts move their eyes, with the exception that the data from experts was used to train the feature set (and in fact these feature sets do not depend critically on the training data in that feature sets look similar between experts, novices and random input image patch sets). Once the diagnosticity of each region is determined, it can then be compared against the eye fixations of experts. Because we determine the diagnosticity first and then compare against the fixations of experts, there is no circularity involved in our analyses.

Latent print examinations are an individualization task, and as such they benefit from the presence of rare features. Rare features have much more value when it comes to individualization (and therefore carry more information in an information theoretic sense). Rarity is defined relative to the global set of features, which has been acknowledged by the fingerprint field on web sites devoted to unusual prints.

An examiner’s eye has an intuitive sense for what is rare or common. This is based on experience with large numbers of prints, and would have to be repeated for any new skin surface such as palms, lips or elbows. The human expert has a general-purpose visual system that can adapt to these new domains in part because the neurons in the earliest stages of the visual system show response profiles that are very similar to the ICA basis functions in Figure 15 and Figure 16 (Olshausen & Field, 1997). Thus the ICA activations that are learned on the basis of the fingerprint patches at the fixation points of experts are mimicking the behavior of the earliest stages of the visual system.

The next step is to use the activations of these basis sets to determine whether a given patch of friction ridge skin is rare or common.

This requires three steps when applied to a new fingerprint, which are described below and summarized in Figure 17.

First, we determine the activation of each basis function at each location in the fingerprint. This is done through a process of convolution, and essentially determines how much a given patch of fingerprint resembles each basis function. Fingerprint patches that are similar to a given basis function produce high activations at that location. The output of this step is a value at each pixel location in the fingerprint that determines how active a given basis function is at that location. This is repeated for all basis functions in the basis set.

Second, we repeat this process for many fingerprints (we currently do this for over one hundred prints, but eventually plan several thousand). This allows estimation of the *activation distribution* for each basis function. Most of the time the activation will be close to zero, while occasionally it will be quite high or quite low if the basis function is a very close visual match to a particular patch of friction ridge skin.

Third, we use these activation distribution to determine the rarity of a particular region of friction ridge skin. Work by Bruce and Tsotsos (2009) demonstrated with visual search tasks that the self-information of a region could be used to estimate the rarity or diagnosticity of that region. The self-information is computed as follows. A given location in the friction ridge impression produces a set of activations across all of the basis images (essentially how well each basis function matches the patch of skin at that location). We can determine how likely it is to encounter an activation value for a given basis function by looking at the activation distribution estimated from the entire dataset.

Mathematically, this is the Shannon self-information measure (Shannon, 1997):

$$-\log(p(x))$$

where $p(x)$ is the probability of observing an activation value for that particular basis function. The nature of ICA basis function are such that they tend to be highly active only rarely. The smaller $p(x)$, the larger (in absolute terms) $-\log(p(x))$ will be.

These $-\log(p(x))$ values are summed up over all of the different basis function activations. Common features will produce very high values of $p(x)$ and therefore very low values of $-\log(p(x))$. However, rare features will produce activation values that fall in a range that are almost never encountered, and therefore will have a low $p(x)$ value. It is this relation that links self-information with feature rarity.

To visualize the self-information computed at each pixel, we have constructed masks that overlay the images. As shown in Figure 18 through Figure 22, visible regions are those that are considered to have high self-information and therefore be diagnostic with respect to feature rarity and individualization. Each of these figures are describe below.

E.2. Feature Rarity Maps

Figure 18 is perhaps our best example of how the self-information metric reveals what a human eye would consider as diagnostic features. The metric reveals the core and delta, but also minutia, breaks in the ridges and regions of high curvature. Straight regions without much minutia or other interesting activity have much less self-information by this metric (see the upper-right portion of the tip). Similar findings are shown for a second print in Figure 19, where straight regions without much variations are down weighted, which regions with curvature or breaks are considered more informative. This figure does reveal one limitation that will have to be addressed, which is that ridges with breaks between pore elements are considered more interesting, which may have be addressed using thinning and a threshold procedure.

Expanding the size of the basis set seems to lead to subtle but noticeable improvement in the diagnosticity maps. Figure 20 illustrates a self-information map that has the same patch size as previous figures, but a larger basis set of 360 (vs. 150 in Figure 18). The contrast between diagnostic and non-diagnostic regions appears even stronger, making this basis set perhaps the most useful of all we have tested.

The choice of basis set may prove critical. The next two figures, Figure 21 and Figure 22 illustrate two cases that do not seem to produce results that are consistent with what a human would label as interesting or rare. The large, 128x128 basis images used in Figure 21 are not revealing interesting features beyond the region above the core, while the small, 16x16 pixel basis images used in Figure 22 reveal little more than the edges of the ridges. These illustrate that our particular metric is sensitive to the underlying basis set, which gives us confidence that it is working appropriately when given the correct input information.

E.3. Validation of Information Measure against Expert Eye Gaze Data

The visualizations shown in Figure 18 through Figure 22 represent our initial foray into self-information analyses. The results for particular combinations of patch size and basis set size look quite promising. However, to validate these results (which are currently ongoing) can compare the regions considered to be most diagnostic against the fixations from experts.

Figure 24 illustrates how close the correspondence between the self-information and eye fixations can be. The dark regions are those the model considers most diagnostic, and the red dots are the fixations from experts. The experts tend to cluster their fixations in those regions the self-information metric deems most diagnostic.

The image pair in Figure 25 show a similar close correspondence between fixations and regions of high diagnosticity. Likewise, the self-information metric illustrated in Figure 26 shows close correspondence when a smaller ICA basis set is used, suggesting that self-information might have to be combined across different spatial scales and levels of redundancy reduction. We explore this below.

The self-information metric is based solely on the image data of prints and is therefore a measure of the natural image statistics inherent to fingerprints. To train the model, we seed it with image patches derived from human examiners, collected using eye tracking methods. By holding some images out of this training process, we can use the remaining eyetracking data to evaluate the metric on novel images.

To evaluate the metric, we use the eye gaze data from 30 novel images and measure the self-information computed at each fixation. We then compare that to the self-information from an equivalent number of image patches that were not fixated by experts. Rather than use a test set of fixations taken from the same images that were used for training, we held out a randomly-selected 30% of our images for testing. This allows us to argue that our classification results generalize to novel images rather than just to novel fixations from the training images.

We expect that the self-information of regions fixated by experts will be higher than those regions not fixated by experts. We are limited, however, by the fact that we must select an ICA basis set first before computing the self-information. As illustrated by Figure 23, ICA basis functions produced by image patches of different sizes tend to highlight different levels of ridge detail. Thus we do not know which level of detail the examiners are relying on.

The solution to this problem is to use the self-information metric from a variety of basis

functions and use a machine learning procedure (logistic regression) to discover which basis function contribute the most to the decision to fixate a particular location. The classification accuracy for individual ICA basis set are found in Figure 27. Individual ICA basis functions give classification accuracies in the range of 58-70%, while combining several sets together gives a classification accuracy of 75% with no loss of generalization to the testing set. Thus we are not over-fitting the training data and suggests that, as with the examiners, the optimal solution combines information across different spatial scales.

Based on this technique, we can correctly classify approximately 75% of the fixations made by experts, with very little loss of generality to novel prints. This is much higher than we saw with the raw ICA activations, and we have good correspondence between the fixations and rarity visualizations in Figure 24 through Figure 26.

This finding represents one of the central points of this research, and is worth summarizing. The activation values from the basis functions are used to identify those regions that are most diagnostic, using principles derived from Information Theory applied to the statistics of the fingerprint activations. These diagnosticity maps are then compared with the fixations of experts to see if they correspond. The only aspect of human expertise that enters the model is during the construction of the feature set, which is fairly generic, and when the different saliency maps are combined to allow information at different spatial scales to impact overall diagnosticity. The essential elements that determine diagnosticity of individual feature maps such as those shown in Figure 17 through Figure 26 is the statistical distribution of individual feature activations in fingerprints.

A limiting factor on accuracy is the inherent noise in eye gaze data (some fixations

are more meaningful than others, and the classifier is forced to classify all fixations with equal weight) and we address this in Section F below.

F. Temporal Dependencies From Automatic Translation

One way to characterize the important regions is to include temporal information. It is likely that human expertise includes a temporal element: it matters in which order you visit particular regions, not just that you visit them. The human visual system has a high capacity visual buffer, which lasts for less than a second before it fades. Matching one complex image patch to another requires moving the eyes back and forth to determine whether two regions are similar, and this must be done before the contents of visual memory fade.

To identify whether the temporal sequence is a factor in human expertise, we used procedures derived from machine translation, as illustrated in Figure 28. This requires four steps: 1) temporal fixation finding: reducing the continuous time series of raw gaze data into a sequence of eye fixations defined mostly by the speed of eye movements over time; 2) spatial clustering to calculate Regions of Interests (ROIs): clustering (x,y) gaze data points into several clusters/ROIs based on the spatial distribution of gaze data on the prints; 3) alignment: segmenting the ROI sequences into ink-latent fixation pairs based on temporal proximity; 4) using a machine translation method to compute the correspondences between ROIs in the inked and latent prints. As a result, we extract the patterns of which corresponding areas that experts examine back and forth between two prints and which areas that novices pay attention to when conducting the same matching task.

The machine translation algorithm works as follows. Our goal is to calculate correspondences between gazed regions in one

image with gazed regions in the other image as participants conducted the matching task. To do so, we view this task as similar to machine translation in natural language processing. The general idea of machine translation is this: assume that we have parallel texts from two languages, for example, “Harry Potter and the Order of the Phoenix” in both English and French, the goal of machine translation is to infer which two words in the two languages correspond. This inference can be done based on statistical information, such as how frequent “egg” in English and “oeuf” in French co-occur together and how frequent “egg” appears without “oeuf”. Intuitively, if a word in English always co-occurs with another word in French and that word in English appears only when the other word in French appears, then those two words are likely to correspond to each other. Most often an assumption in machine translation is a sentence-level assignment – which sentence in English maps to which one in French is known. Say it in other way, we have sentence pairs from two languages and use this data to infer word correspondences.

In the fingerprint-matching task, we conceptualize ROIs from one image as words in English, and ROIs on another print as words in French. Based on this conceptualization, the aim here is to find which gazed region in one print maps to which gazed region in the other print. To achieve this, we also need to segment continue gaze data generated by participants into “sentence” pairs. This is done based on the observation that participants may generate a few fixations on one image, switch to examine another image with more fixations to search for corresponding areas on the other image. In light of this, and as showed at the bottom of Figure 28, we first divided a whole sequence into several subsequences using the visual attention switches between two prints as breaking points, and then grouped those

subsequences into several pairs based on temporal proximity.

The outcome of this procedure is a set of fixation sequence pairs from which we further calculated which fixated areas in one image map to what fixated area in the other image in the next step. We call each pair of two fixation subsequences on two prints a searching instance as we assume that participants were comparing and matching regions between two prints through those eye fixations on both prints. Figure 29 and Figure 30 illustrate instances extracted from a continuous ROI sequence. To the degree to which experts will find matching features in both prints we will be able to discover these through machine translation, and in each case the machine translation algorithm found more correspondences for the expert than the novice.

The results of the machine translation analysis applied to all experts and novices are very clear. In implementation, our method produced all of the possible ROI-ROI mappings between fixations on the two images. We chose two criteria to select reliable ROI-ROI pairs. First, the co-occurring frequency is at least 2, meaning that a participant at least looked at one region in one image and subsequently look at another region in the other image, and repeated this again. Second, the overall mapping probability needs to be greater than 0.4. We use two criteria because with one visit between the two ROIs the link strength will be artificially high at 1.0, but trivially high.

Based on this selection, the experts have an average of 17.1 reliable mappings/links found, while the novices have an average of 7.1 links found ($t(34)=-6.73$; $p < 0.001$, $sd = 8.84$) from a dataset of latent/inked prints. For a dataset with clean prints, we found a similar result. The machine translation found an average of 11.1 links for experts and 8.3 links for novices ($t(29)=-3.18$; $p < 0.01$, $sd = 4.59$).

The results of this analysis demonstrate that one element of human expertise resides in the order in which different locations are visited. This suggests that temporal information is an important component of human abilities.

The correspondences revealed by machine translation are quite accurate despite the fact that the algorithm knows nothing about space directly. The average deviation between the corresponding location found by machine translation and the actual matching location as determined by a human experts is quite small: The deviation is about 1 degree of visual angle which for our images corresponds to about 2 ridge widths in distance. This is perhaps a surprisingly small number given that the machine translation algorithm does not know about space directly.

This modeling approach demonstrates that the temporal search processes of human experts reveals not only the process by which they identify regions in the two prints, but may also demonstrate important dependencies between regions in friction ridge impressions. Typical models of ridge information make independence assumptions, but the temporal aspect of human expert search may reveal important dependencies between regions. We discuss possible future extensions the modeling in Section H.1.2. In the section below, we explore dependencies between the ICA activations for patches that come from similar regions as another way to account for dependencies that might play an important role in representing image detail.

G. The CoVar Model- Modeling Covariance Among Activations

The ICA weights used with the self-information metric produced reasonably high classification accuracy for our data, especially if the different scales are combined together to reflect the fact that examiners likely use information at different spatial scales.

However, central to the self-information metric is the idea that individual basis functions are independent. This allows for the multiplication of probabilities that underlies the self-information computation and gives a statistical measure of feature rarity. This assumption is justified by the fact that the independent component analysis algorithm is *designed* to find components that have independent activations across the entire dataset.

One limitation of this approach is that although basis functions are independent across the entire dataset, they are likely not independent for smaller regions, because individual features may strongly activate several different basis functions. Thus there may be important correlations in the basis function activations within similar regions of an image.

This fact was recognized by Karklin and Lewicki (2009) during an investigation of natural scene statistics. They point out that the earliest stages of the visual system act much like the ICA basis decomposition, with individual neurons sensitive to different orientations and spatial frequency patches at particular locations. However, the visual system must be able to achieve a measure of positional invariance by pooling across similar detectors positioned at slightly different locations. This implies recognizing that different detectors at slightly different spatial positions are all processing similar features (i.e. same orientation and spatial frequency). This can be done by a second level of artificial neurons that learns a set of weights on the ICA basis functions to process the *correlations* between the individual basis function activations.

Karklin and Lewicki (2009) proposed a model, called the CoVar model, that could not only discover a set of weights that learned the correlations among the outputs of the basis functions, but also achieved positional and contrast invariance while remaining very

sensitive to orientation. This network could have a large number of basis functions (around 500) but have relatively few high-level latent neurons (around 25) that could represent more abstract features by selectively weighting the activations from the input layer. Essentially, the weights on the connections between the input layer (the basis functions) and the latent layer (the high-level neurons) learn a subspace decomposition of the high-level ICA activations that meaningfully represented the natural scene statistics of their images.

This particular model seems well-suited to the application to fingerprints, because the model naturally represents variations in texture appearance, orientation and spacing. This model is also a natural progression of the previous ICA approach, because the first layer of the CoVar model is very similar to the ICA decomposition. However, rather than assuming independence between the ICA activations as the Self-Information metric does, the model explicitly models these correlations in local subregions. This model is likely to reflect the behavior of human experts more closely, because it shares computational principles with the early stages of the visual system. However, it may not provide the feature rarity statistics of the self-information metric, and so in some ways the two approaches are complementary and could be used in different settings for different purposes.

G.1 CoVar Model Training

We trained the CoVar model using similar procedures as the ICA decomposition. We cropped out regions of clear fingerprints near where experts fixated, and used these image patches to train the model. The CoVar model is computationally very expensive, and to make the training tractable we limited the patches to 24x24 pixels (about 3 ridge widths). However, in recognition of the observation that spatial scale is an important element of the detail in friction ridges, we

used integer multiples of this patch size when extracting patches from our fingerprints, such that we used actual crops ranging from 48x48 pixels to 216x216 pixels. As with the ICA/Self-Information analysis, different spatial scales will likely represent different sources of information. This may range from level 3 detail such as idiosyncratic pore shapes, to level 1 configural information such as pattern type.

To evaluate the success of the model, we relied on similar procedures as described for the ICA/Self-Information metric. We computed the activation of the 25 latent neurons at each fixation from the experts. This places each fixation as a point in a 25 dimensional space. We then chose random points on the print that were not near expert fixations and computed the activations of the 25 latent neurons for each random fixation. We then submitted both sets of activations to a logistic regression classifier and a support vector machine, using a held-back test set of fixations from a separate set of images.

G.2 Classification Results

These analyses can be related to the Self-Information metric. The logistic regression classification achieved classification performance around 60-70% accuracy for individual Self-Information maps, and about 75% accuracy when all the maps are combined (see Figure 27).

Logistic modeling of the CoVar activations using the same expert/random comparison as with the self-information metric produces an improvement in the classification accuracy. Different spatial scales produce different classification accuracy, and these are summarized in Figure 31. Classification results for individual spatial scales are now in the low 80% range with strong generalization to novel prints (labeled Testing Data). This suggests that the latent layer of the model is capturing important covariance information that reflects how

humans perceive texture patterns. We are still exploring extensions to the classification procedures that combine across the individual scales, and we expect classification performance to improve even further as it did in Figure 27.

The classification results should be viewed in the context of the fact that eyetracking data is inherently noisy, not only from the perspective of the mechanisms of eye gaze collection (with accuracies in the 1-2° of visual angle, or about 3-5 ridge widths) but also in terms of where the experts choose to send their eyes. Some fixations are undoubtedly more informative than others, and the current approach treats all fixations equivalently. Given these sources of noise, it is unlikely that we would see classification accuracy above 90%.

These results suggest that the CoVar model represents an important tool to discover those features that experts rely on for comparing friction ridge impressions. In the next section we explore a set of visualizations that allow us to assess the adequacy of the model, as well as demonstrate how the CoVar model could be used to provide novice users or trainees an opportunity to view a novel latent print as if it was examined by an expert.

G.3 Saliency Visualization

The logistic regression analysis provides an opportunity to visualize those regions that experts consider to be diagnostic, even for prints that they have never seen before. To visualize the saliency of different regions for purposes of identification, we first compute the activations of the CoVar model at each point in the fingerprint image. This is somewhat computationally expensive, so we instead compute the activations in a grid of every 10th pixel and interpolate between the values. Explorations at finer scales revealed equivalent results at a cost of much more processing time.

Once we have these activations for a particular model and spatial scale, we can then multiply the activations by the weights from the logistic regression and sum up the weighted activations. This weighted sum computes the degree to which an examiner would consider that location particularly diagnostic (at least enough to warrant a fixation).

Example images are shown in Figure 32 through Figure 34 for training images (those that were used to train the logistic regression). Darker regions in the print are those deemed by the classifier to be more likely to be visited by human experts. The close correspondence between the distribution of fixations (red dots) and the darker regions illustrates how accurate the classifier can be when fitting human eye gaze data. The images shown in Figure 33 illustrate how reliable the classifier predictions can be across different exemplars of the same image.

As with the Self-Information metric, sometimes multiple spatial scales may be necessary to represent all of the information that examiners attend to. An example of this with the CoVar model is shown in Figure 35. The left panel is with a small scale, while the right side is with a large scale. Both capture fixations, but we may need a combined representation to fully account for human expert performance. We discuss this in a future section.

The logistic regression classifier's performance on the training data is impressive, but in order to be generally useful, it must generalize to new images. During the training portion of the logistic regression classifier we held back some of our images. This is the best test of generalization: can the classifier accurately predict where experts will send their eye gaze on images that were not used for training?

The images shown in Figure 36 through Figure 38 illustrate saliency predictions for the model on novel images. These demonstrate

that the model can readily generalize to novel images and accurately reflect the fixations for human experts.

Together the classification performance and the saliency maps in Figure 32 through Figure 38 illustrate that the CoVar model accurately reflects much of the perceptual mechanisms in human experts that drives their visual performance. Tools based on this algorithm can provide valuable information about human expertise to trainees and even suggest regions that one expert may not have seen but the model identifies as something that experts might look for.

In the next section we discuss applications of these approaches, to extend this basic science work into an applied domain.

H. Conclusions and Implications for policy and practice

Previous attempts to characterize the information in friction ridge impressions have relied primarily on Level 2 minutiae such as ridge endings, y-branchings and dots. The current approach is more agnostic about the nature of the features, and instead builds up a set of basis functions that represents a much larger set of visual details, including individual ridge shapes as well as global configural information.

We explored multiple spatial scales as well as the degree of redundancy reduction as expressed in the size of the basis set. This creates a set of tuned filters that are sensitive to the natural statistics of friction ridge impressions and also reflect those regions that experts consider most important.

The activations produced by the basis functions at each locations can be used to create a measure of the overall probability of observing that particular pattern in the entire database. This is relevant for feature diagnosticity computations, because the most informative features are the most rare.

These diagnosticity computations, computed via the self-information metric, were validated against the fixations from experts, which suggested that a combination of spatial scales was necessary to fully account for the behavior of experts. The validation procedure has a bootstrap nature, because there is no way to determine whether experts always look at the most diagnostic regions. However, the basis set that results from human eye fixations is fairly generic, and once it is specified the self-information metric follows directly. As with all measures of feature rarity, the computations depend on the initial representation, and the fact that we have strong correspondence with expert data suggests that we are tapping those features that experts also agree with. A particular strength of this analysis is that the self-information metric was able to generalize to new prints that were not part of the training set. This suggests that the representation consists of set of dimensions that meaningfully captures feature diagnosticity as understood by human experts. However, it is important to point out that this computation of feature diagnosticity is *independent* of the human data, and the computation would stand regardless of where humans sent their gaze. The fact that we see such close correspondence between the two suggests that we have an appropriate featural representation in the basis set, the self-information metric is the appropriate statistic to compute, and that humans have an intuitive sense of which features are most diagnostic.

It will be of particular interest to test regions that the self-information metric considers diagnostic but the experts have not yet looked at. We may find that experts realize that other regions they might not have considered are also worthy of inspection.

Two other extensions to the self-information metric were explored. First, we applied the CoVar model, which learns the covariance structure of the activations of the

individual basis functions. This allows the model to perform a subspace decomposition of the activations that best reflect the structure of individual patterns. This approach takes advantage that all patterns in a particular region are similar, and look dissimilar from patterns in other regions. For example, the core area has lots of circular structure, and each patch in this region will all contain this circular structure. This is different than the tip, where the ridges tend to have gentle concave ridges.

This model produces activations that were more diagnostic with respect to where the examiners send their gaze than the Self-Information metric. Part of this may come from the fact that the CoVar machine learning training uses the expert data along with the image features to develop the classification solution, while the Self-Information metric only uses the feature rarity statistics to determine which is most diagnostic. Part of the reason for the differences might be that experts may not look at the most diagnostic regions (as revealed by the Self-Information metric), or they may chain together several common features to compute a measure of feature rarity.

The CoVar approach is particularly useful for training applications, because it can be applied to a novel print and used to illustrate to a trainee where an expert would likely look if they were inspecting this print.

The second application we proposed uses temporal information in the eye gaze to determine which regions tend to correspond. We use extensions from Machine Translation to identify regions and correspondences between matching prints. This approach suggests that the temporal information, as well as which features are visited, is an important part of the human investigation. This also suggests that there are spatial dependencies between features in the latent print, which is something that present models may not account for. Below we explore extensions to

this approach that might account for spatial dependencies.

H.1. Extensions to Other Computations

The primary goal of the present proposal was to develop and validate a quantitative representation of the information content of friction ridge skin. The ICA basis functions and, the CoVar model extension, provide this quantitative representation because for any given image patch we can compute statistics such as its rarity and diagnosticity. We are also building this representation based on the statistics of friction ridge impressions and validating it against human expert data.

The quantitative representation afforded by the basis function approach is quite powerful, and in this section we will explore ways to extend this approach to other applications.

H.1.1. Discriminative Value of Particular Patches

The self-information metric tells us how likely it is to observe a particular pattern in a database given the image statistics of the images in that database. As such it constitutes a measure of feature rarity, and we saw close correspondence between feature rarity and the eye gaze of experts. However, there are additional statistics that might also be useful. For example, suppose that a particular pattern type was fairly rare, but that there were two prints in the database that were quite similar? This is known as the ‘close call’ situation among practitioners, and it would be useful to know the overall discriminative value of a feature relative to individual prints in the database. That is, there could be a very rare feature that has two strong matches in the database, and this might be distinguished from the case where a feature was rare and had only one close match in the database.

There are several approaches to computing the discriminative value of

individual patches, and there are lots of assumptions that would have to be worked out. For example, a spatial prior is likely an important assumption, because if the strong match is in a different location relative to some landmark such as the core, this strong match should probably be discounted.

One approach that seems promising is to compute the probability of a given feature being present in a particular exemplar print. Work by Ullman, Vidal-Naquet, and Sali (2002) suggests that fragments of intermediate size can be used to compute the probability of a fragment from a latent print coming from an exemplar inked print. It would also produce a measure of the discriminability of the feature given a particular database, because it would not only compute the likelihood of the latent and exemplar sharing a common source, but also the likelihood of having a similar print in the database. This would address the issue of having identical twins in the database who might have similar prints, or close non-matches that can occur in large database searches.

H.1.2. Temporal and Spatial Network Statistics

One current project that is in progress uses network statistics applied to the spatial and temporal dependencies in the expert eye gaze record. Human experts often describe creating ‘target groups’ by combining individual features into a group that presumably has higher discriminative value than individual features. This ignores possible dependencies between features, but likely increases the value of individual features for purposes of identification.

The creation of a target group likely represents an eye fixation to a central feature such as the core or a salient target, and then repeated saccades to nearby features, possibility with return saccades to the central feature. Once this set of features is placed into working memory, the examiner then looks to

the corresponding region in the other print to search for correspondences in image detail. The gaze behavior of experts supports this mechanism, because experts make many more within-image saccades than novices do, which is consistent with the idea of building up target groups with small saccades to nearby features.

The goal is to represent both the temporal information as well as spatial regions in a model that will allow us to compute network statistics to discover the ‘hub’ regions that likely underlie a target group. Figure 39 illustrates conceptually how we construct a Hidden Markov Model with a fixed number of hidden nodes (say 30, or 15 on each image of an image pair). The emission matrix of each hidden node is essentially a set of means and variances of Gaussian functions in the activation space, so with 20 ICA activations we have 20 sets of means and variances. These Gaussian functions determine the probability that the activation set of each fixation ‘belongs’ to a hidden node.

The model takes in the raw set of fixation locations, computes the ICA activations based on the image detail, and then fits both the transition probability matrix and the emission matrix to produce a representation of how experts search through the activation space. An example visualization of this process is shown in Figure 40. The different colors represent those fixations from experts that are identified with each hidden node (that is, they are most likely to have been produced by the hidden node). The blue and green traces show the strength of the probabilities in the transition probability matrix. The centroid of the clusters are determined by taking the median x-y locations of the fixations that are associated with each hidden node.

Several things can be observed from Figure 40. First, there is a strong tendency to associate similar regions across the two prints, which would be expected given the task.

Second, there are a large number of links within an image, suggesting that experts are looking at several regions to bind them together.

A strength of this approach is that it is computed across all experts *and* all images. That is, the solution is not specific to one individual or one image, and therefore has good generality to new images. Thus the solution is robust and generalizes such that it can predict the eye gaze record for future image pairs. It also simultaneously determines the spatial grouping (i.e. what is an important feature or region) as well as the temporal dependencies (i.e. how are these features temporally grouped together).

The next step is to use network statistics to discover the possible existence of ‘hub’ nodes that tend to be used to create target groups. For example, node 10 in the left image of Figure 40 has a large number of relatively connections within the left-hand print, and therefore may be identified as a hub.

This work is ongoing, as we need to determine the appropriate representation of the activation space (ICA or CoVar activations) as well as the appropriate spatial scale and number of hidden nodes. We will use likelihood statistics to determine the adequacy of each combinations of parameters. We anticipate finishing this work by the end of the grant period.

H.1.3. Self-Information and the CoVar model

A related effort extends the CoVar model to the self-information metric. In principle this is a straightforward application, because the CoVar activations can be gathered across many different image patches across hundreds of images and the distribution of activation values can be collected for each latent neuron. However, a key assumption underlying the self-information metric is that each basis function produces activations that are *independent* across the entire dataset. This was

verified empirically with the ICA basis functions, as well as expected because the ICA algorithm is designed to find basis functions that produce activations that are independent.

The CoVar model takes advantage of the fact that although the activations are independent across the entire dataset, they are highly correlated for local regions, because the nature of natural scene statistics tends to produce similar-looking image patches close together. The CoVar model learns the covariance structure to produce a principled sub-space decomposition of the ICA activations.

There is no guarantee, however, that the latent neuron activations will be independent. This must be empirically validated before the self-information metric could be applied in this case, or the metric will need to be generalized to take these dependencies into account. This is an active focus of our current efforts, because not only would this produce a metric with strong predictive ability for image patches, it would also reveal the feature rarity and diagnosticity of each image patch.

H.1.4. Tools for Examiners

The representations created by the ICA activations and the CoVar decomposition have a number of strengths. They accurately predict where experts will send their gaze, and they provide a measure of feature diagnosticity. There are several practical applications of these metrics. Although the focus of the current grant efforts were on basic science, below we describe how our results could be applied to casework.

As discussed previously, the self-information metric readily provides a measure of feature rarity and can be extended to produce estimates of the discriminative value of a particular patch. Our current modeling is conducted in Matlab, which allows for rapid prototyping and model construction, but is not a suitable platform for widespread production

software. Thus the self-information metric would have to be translated to code that would allow for distribution. However, this kind of application could easily be adopted by AFIS vendors as part of their application suite.

Another application of the basis functions might be noise reduction. The ICA basis weights (and those of the CoVar model) represent a set of tuned filters that characterize the image detail in a set of friction ridge patches. Matched filters are useful when extracting a signal from noise, because they are most sensitive to the image detail and exclude off-band noise.

An extension of this approach might be to include spatial dependencies in the ICA activations. This might allow a region of high noise to be cleaned up by using the image statistics of nearby cleaner regions to estimate the optimal filter to apply to the noisy region. The CoVar model might be especially appropriate here, although further testing would be required to fully develop this application.

These examples illustrate that the quantitative representation afforded by the ICA and CoVar representations have wide application, especially once the appropriate spatial scale parameters have been estimated from human experts.

H.2. Glossary

basis functions- a set of fundamental building blocks that when combined create more complex structures. In vision these are simple shapes that when combined produce complex patterns. These reflect the statistical regularities of the input images.

Clustering- A technique for grouping objects together based on their spatial proximity.

CoVar Model- Covariance model by Karlin & Lewicki (2009) that can be used to decompose a set of image patches into a set of basis functions, but also to model the covariances of the activation functions to

produce an even lower subspace representation.

dimensionality reduction- a procedure that eliminates redundancies within a set of input objects (in this case image patches) to find the most common statistical patterns.

Expectation Maximization- a procedure to fit a set of Gaussian curves to a set of data such that there is maximum overlap between the data and the Gaussian meant to reflect these points.

Feature space- The concept that a complex object such as an image patch can be represented by a set of basis functions. The activations of these basis functions can be considered as coordinates in a high dimensional space, the feature space. Similar points in this feature space likely correspond to two similar features in visual space.

Gaussians- Another term for the normal distribution (also called the bell curve).

Hidden Markov Model- A representation of both spatial and temporal information that models the sequence of steps through a set of hidden states that produce observable outcomes. An analogy might be a frog jumping from one lily pad to the next, but all you could observe was the way action on the shore and had to infer the sequence of jumps.

independent components analysis (ICA)- a method of dimensionality reduction that produces basis functions that have activations that are as independent as possible, in the sense that knowing the activation of one ICA component tells you very little about the activation of the other components.

logistic regression- A classification method that assigns linear weights to different basis function activations and then transforms the resulting weighted sum into the range 0-1, which corresponds to category membership.

Machine Translation- A technique to assign correspondences between tokens in different languages or images based on their co-occurrence across parallel texts.

non-Gaussianity- many things are normally distributed (i.e. Gaussian). However, brain signals and other systems tend to have non-Gaussian distributions, such that the output is usually small but occasionally becomes quite large.

Self-Information Metric- A technique to assign feature rarity to a region based on the self-information value calculated from the basis function activation probabilities.

Shannon self-information measure- The theorem that rare signals are the most informative.

support vector machine- A classification procedure that attempts to separate two groups by finding a hyperplane (decision criterion) that correctly assigns the category membership to as many objects as possible. This is typically done in a feature space.

H.3. Grant-Related Publications

Busey, T. A., & Parada, F. J. (2010).

The nature of expertise in fingerprint examiners. *Psychonomic Bulletin & Review*, 17(2), 155-160.

Spellman, B. A., & Busey, T. A.

(2010). Emerging Trends in Psychology and Law Research. *Psychonomic Bulletin & Review*, 17(2), 141-142.

Busey, T., & Dror, I. E. (2010).

Special Abilities and Vulnerabilities in Forensic Expertise. Scientific Working Group on Friction Ridge Analysis (Ed.), *The Fingerprint Sourcebook* (pp. 15.11-15.24). Washington DC: National Institute of Justice.

Yu, C., Busey, T., & Vanderkolk, J. R. (2010). *Discovering Correspondences between Fingerprints based on the Temporal Dynamics of Eye Movements from Experts*. Paper presented at the International Workshop on Computational Forensics, Tokyo, Japan.

Busey, T., Yu, C., Wyatte, D., Vanderkolk, J. R., Parada, F. J., & Akavipat, R. (2011). Consistency and variability among latent print examiners as revealed by eye tracking methodologies. *Journal of Forensic Identification*, 61(1), 60-91. Lead article in this volume.

Busey, T., Yu, C., Wyatte, D., & Vanderkolk, J. (2013).

Temporal sequences quantify the contributions of individual fixations in complex perceptual matching tasks. *Cogn Sci*, 37(4), 731-756. doi: 10.1111/cogs.12029

Parada, F. J., Busey, T.A., & Yu, C. (submitted). ExpertEyes: an open-source system for the collection and analysis of eye-tracking data. *Behavior Research Methods*.

Busey, T.A., Silapiruti, A. & Vanderkolk, J. (submitted). The Relation Between Sensitivity, Similar Non-Matches, and Database Size in AFIS Searches. *Submitted to Law, Probability and Risk*.

H.3. Grant-Related Presentations

- Busey, T. A., Yu, C. & Vanderkolk, J. (2010). Expertise in latent print examiners. Talk presented at the 2010 NIJ Grantee Conference, Washington DC.
- Busey, T. A., Yu, C. & Vanderkolk, J. (2010). Machine learning and fingerprints. Talk presented at the 2010 ASIC conference in Bend, OR.
- Busey, T. A., & Yu, C. The Nature Of Expertise In Fingerprint Examiners As Revealed By Eyetracking. Talk presented at the 2010 annual meeting of the Psychonomic Society, St. Louis, MO.
- Busey, T. A., Yu, C., & Parada, F.J. Features and Strategies used by Latent Print Examiners. Talk presented at the 2011 Annual Interdisciplinary Conference, Jackson, WY.
- Busey, T.A., Yu, C. & Vanderkolk, J. (2011). What Cognitive Psychology Can Tell Us about Latent Print Examinations? Talk presented at the International Association for Identification Educational Conference, Milwaukee, WI.
- Kitchell, L.M., Parada, F.J., Emerick, B. L. & Busey, T.A. (2011). Feature Selection Strategies and Perceptual Expertise in Configuration Search Tasks. Poster presented at the 2011 Psychonomics Society Annual Meeting, Seattle, WA.
- Busey, T.A., Yu, C. & Vanderkolk, J. (2012). Cognitive Science and Latent Print Examinations. Talk presented at the Maryland Division of the International Association for Identification.
- Busey, T.A., Yu, C. & Vanderkolk, J. (2012). What Cognitive Psychology Can Tell Us about Latent Print Examinations. Talk presented at the Illinois Division of the International Association for Identification.
- Parada, F.J. & Busey, T.A. (2012). ExpertEyes: Open source eyetracking software and hardware for research and teaching. Talk presented at the Society for Computers in Psychology conference, Minneapolis, MN.
- Busey, T. A., Yu, C., Parada, F. J., Emerick, B. R., & Vanderkolk, J. (2012). The Utility of an Intermediate Representation of Feature Space: Lessons From Fingerprint Examiners. Talk presented at the annual Psychonomics Society Meeting, Minneapolis, MN.
- Busey, T.A. & Vanderkolk, J. (2012). What Cognitive Psychology Can Tell Us about Latent Print Examinations. Talk presented at the Texas Division of the International Association for Identification.

Figures

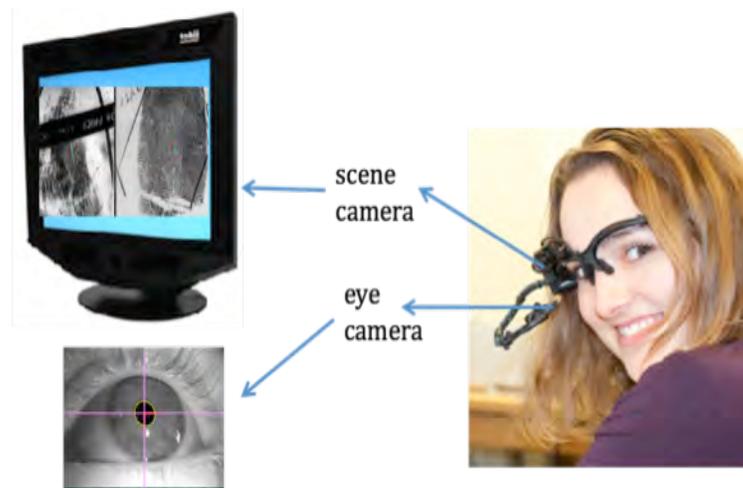


Figure 1. Our eye tracker worn by a participant in the eyetracking studies. It consists of one camera that monitors the position of the eye relative to the head, and a second camera that monitors the position of the head relative to the images being examined.



Figure 2. High Definition Eye camera that allows high-resolution movie capture during actual latent print examinations.



Figure 3. Simulated latent print with artificial texture noise added to the left print. These images look very similar to latent prints, but have the advantage of allowing access to the ground truth of the ridge detail in the clean version of the print.

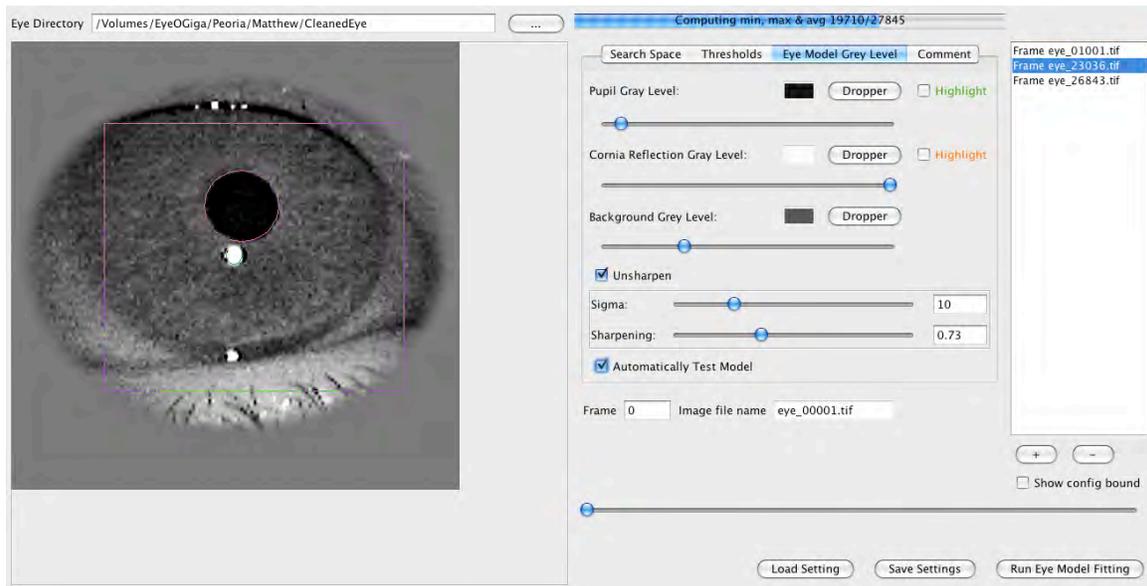


Figure 4. Forward eye model interface. The pupil (dark circle) and corneal reflection (white circle) are identified using a forward eye model that adjusts the parameter settings based on the location of the pupil. Once a set of parameters is found for each region, the program fits the eye model to the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

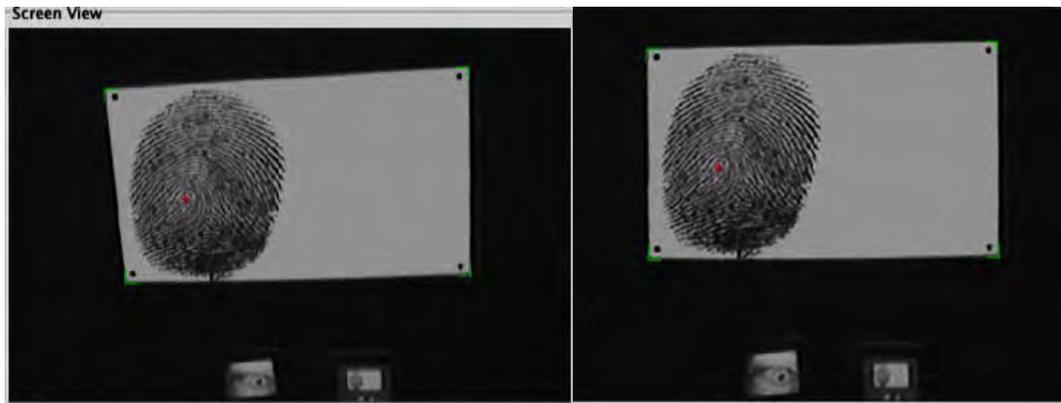


Figure 5. Eliminating image distortions. When the head-mounted video camera is positioned away from the center of the monitor, image distortions result. In this case the upper-right corner of the monitor in the image on the left is distorted. We created an undistortion algorithm that eliminates these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

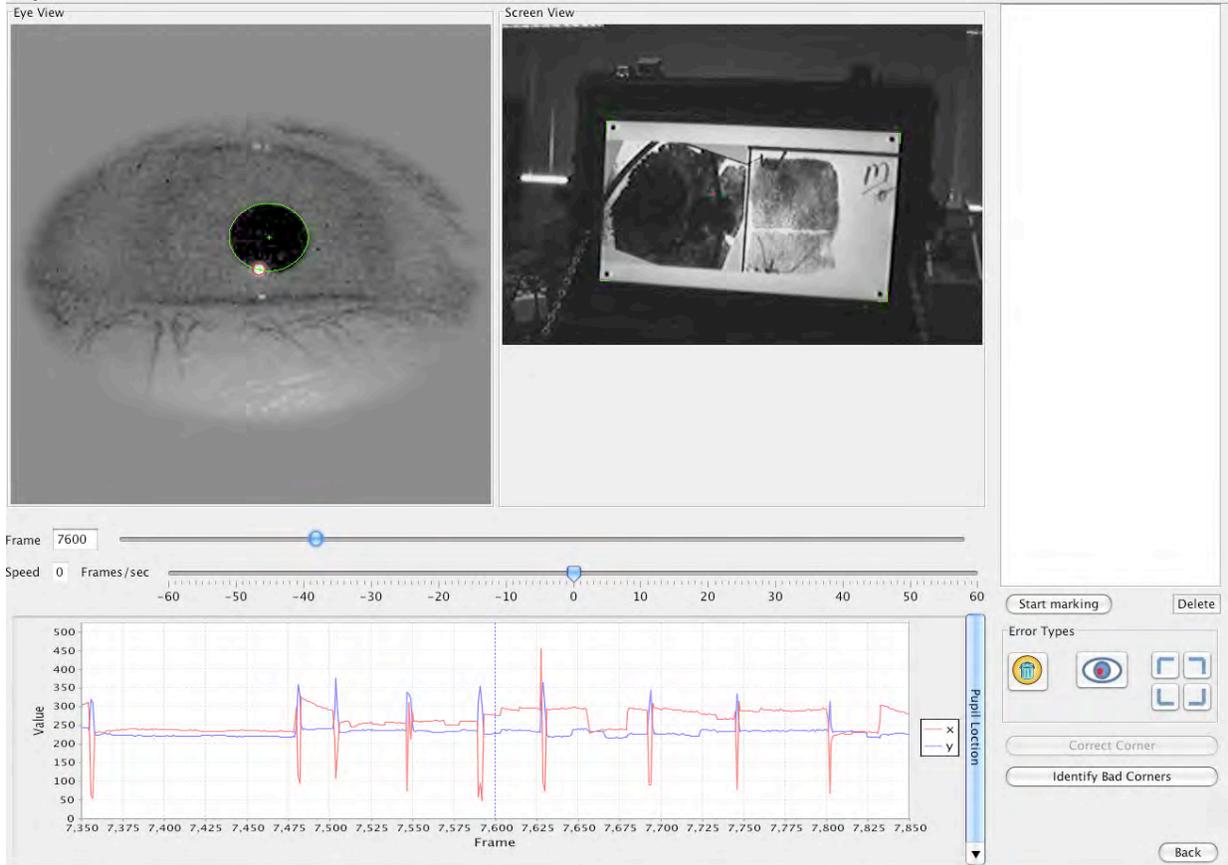


Figure 6. Screen that plots the eye data, along with the fitted pupil and corneal reflection parameters, along with the estimated eye gaze location in the scene camera (red plus). Graphs along the bottom show the pupil location for frames around the current frame. The large discontinuities in the blue and red curves are eye blinks.

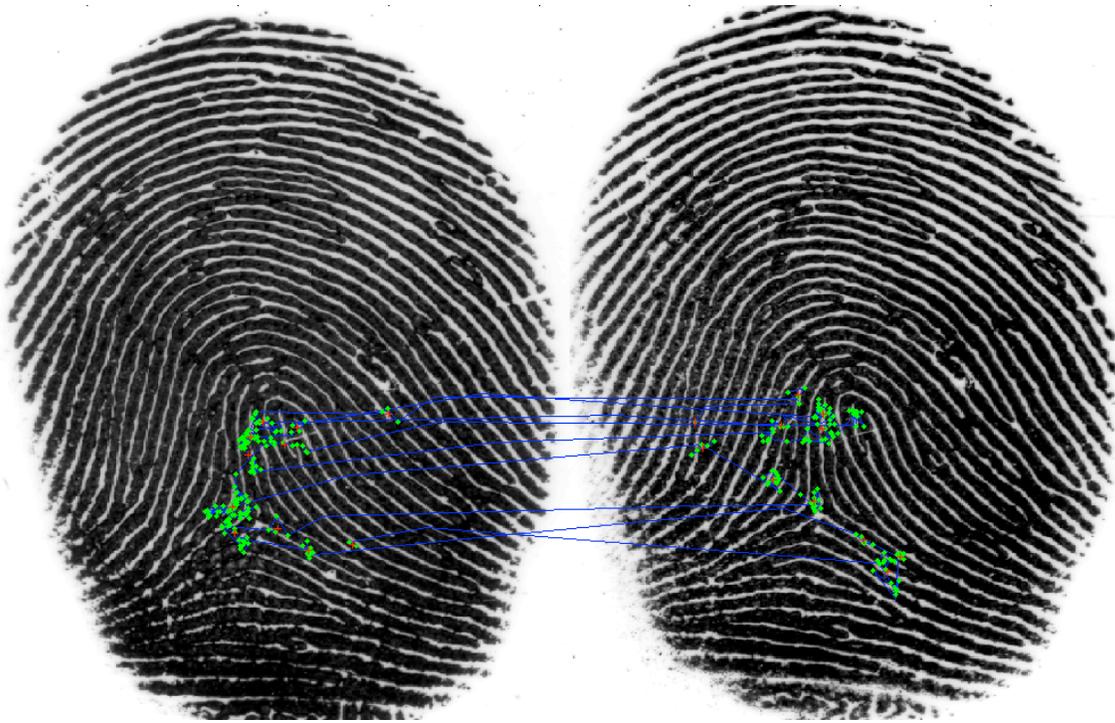


Figure 7. Gaze data from one subject. Green dots are raw gaze estimates, red dots are fixations that are determined using a velocity-based measure, and blue lines are saccades from one location to another.

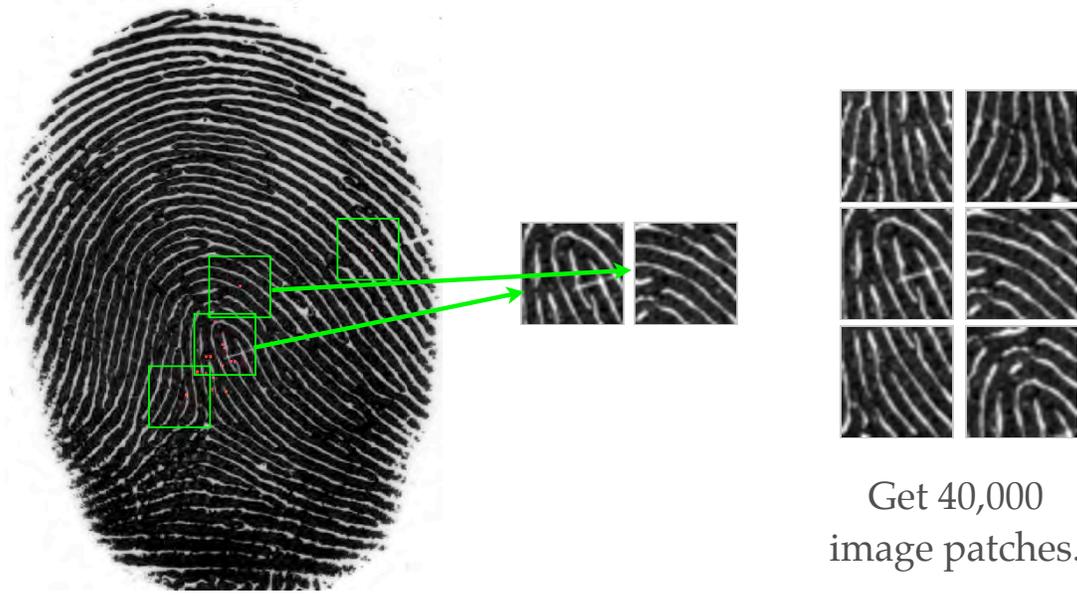
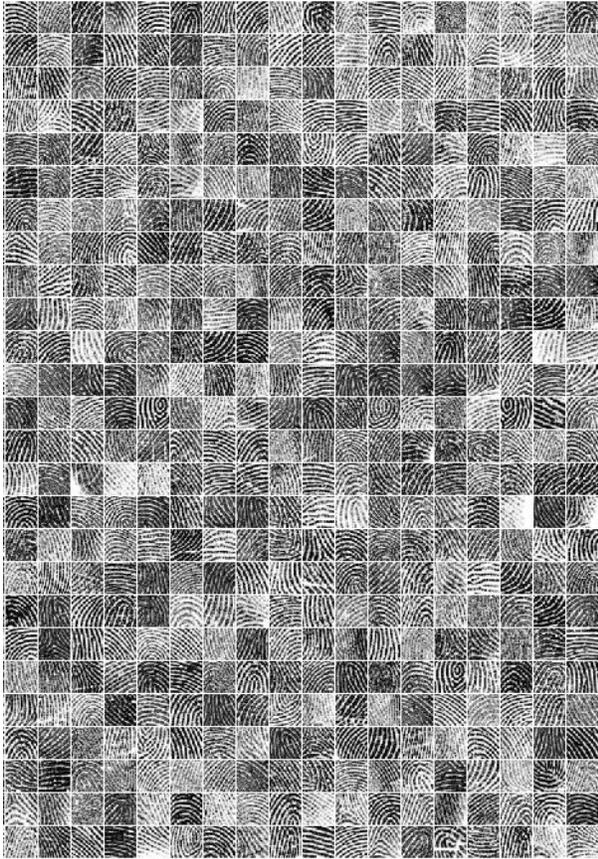


Figure 8. Illustration of image patch extraction centered on fixations from human observers. Each image patch is 38x38 pixels in size, and is extracted from a region of the latent or inked print. These image patches are then analyzed for commonalities using the Independent Components Analysis approach.

Patches from Experts



Patches from Random

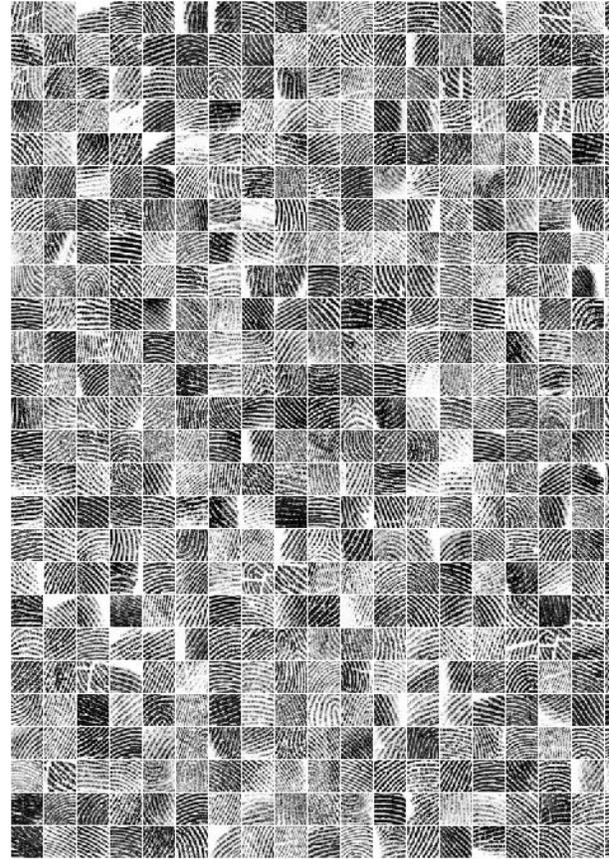


Figure 9. Example patches derived from experts and random locations.

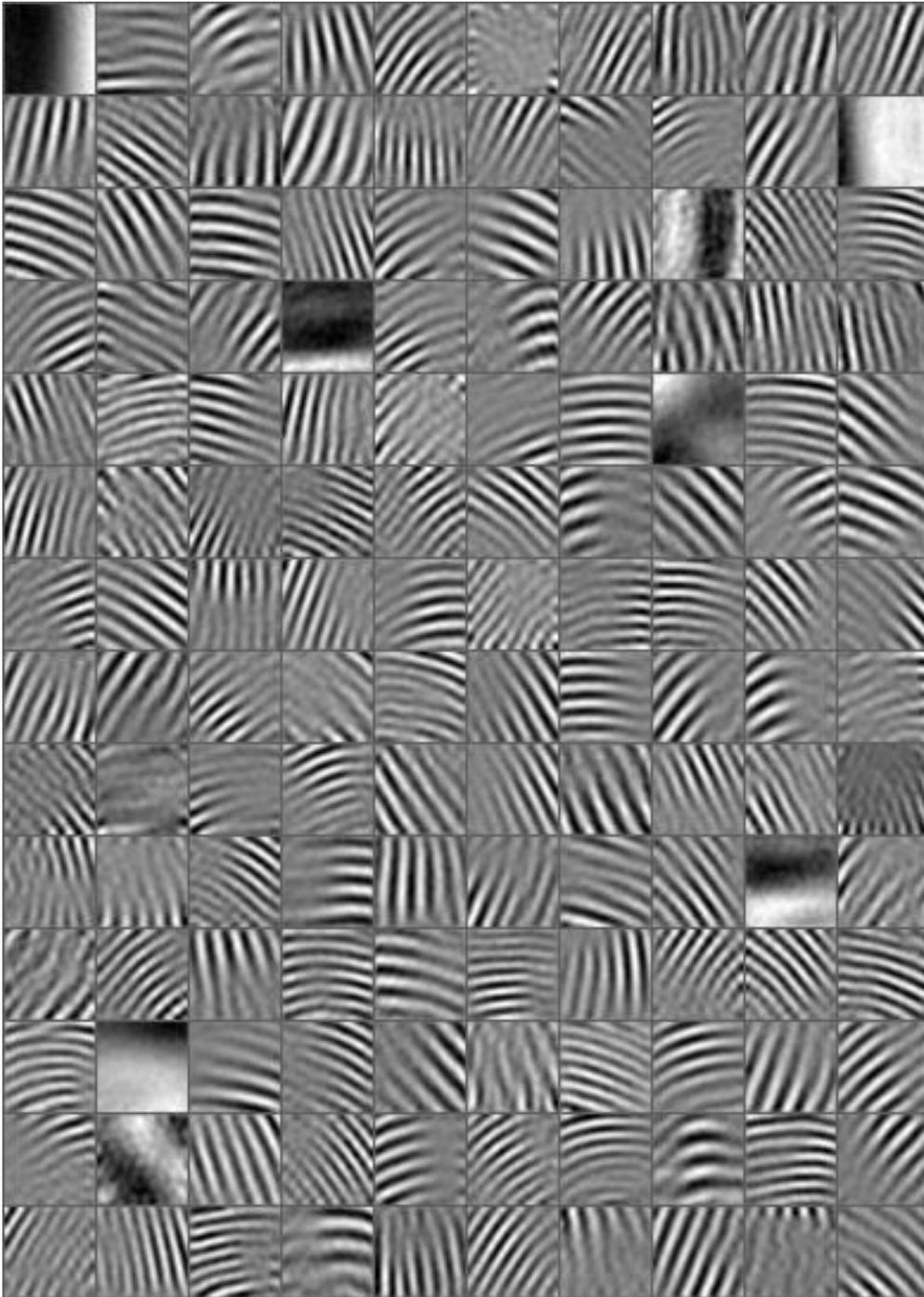


Figure 10. Basis set recovered from image patches taken from expert data and random locations. This set can be used to project the data into a high dimensional space for machine classification.

Relatively few basis images can reconstruct the original images (with some error)

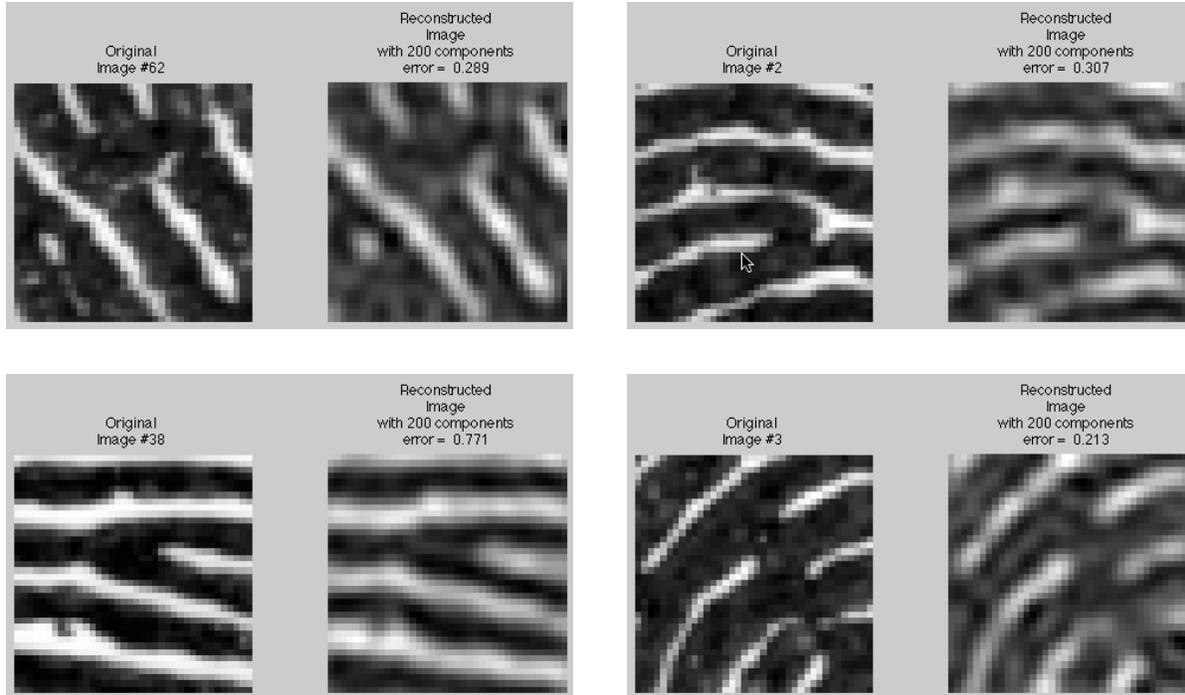


Figure 11. Relatively few basis functions can be used to reconstruct the original image patches, although with some error which mainly translates as blurring of level three details such as the shapes of the pores.

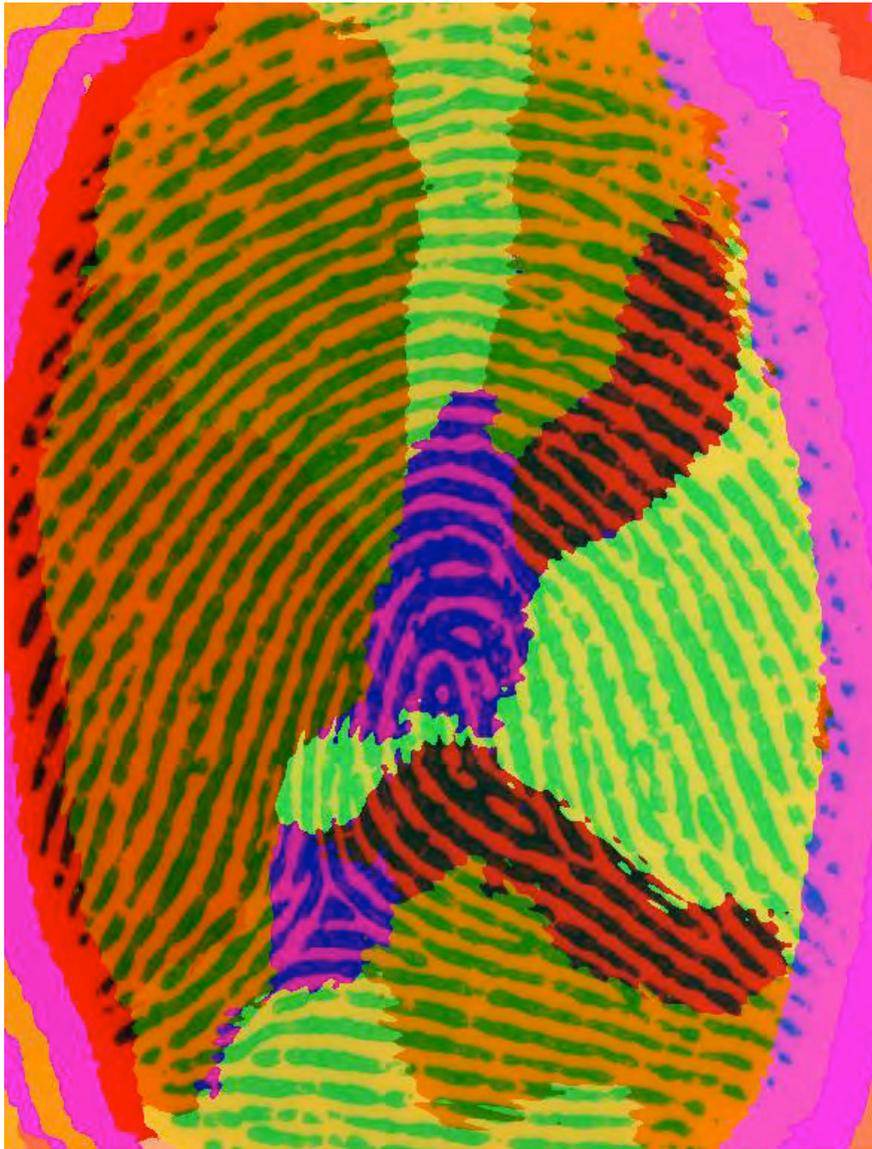


Figure 12. Example clustering solution based on a pixel patch size of 128x128, a basis set size of 50, using 30 different clusters. Regions of color correspond to collections of pixels that are associated with the same cluster of ICA basis activations. Despite the fact that this algorithm knows nothing of space, it tends to find that contiguous regions of pixels are associated with the same cluster. In addition, it tends to find the core and delta regions of each fingerprint. Although this approach awaits validation from human experts, it does seem to capture the different regions of the fingerprint and we will use it to define sub-regions for the self-information metric of feature diagnosticity.

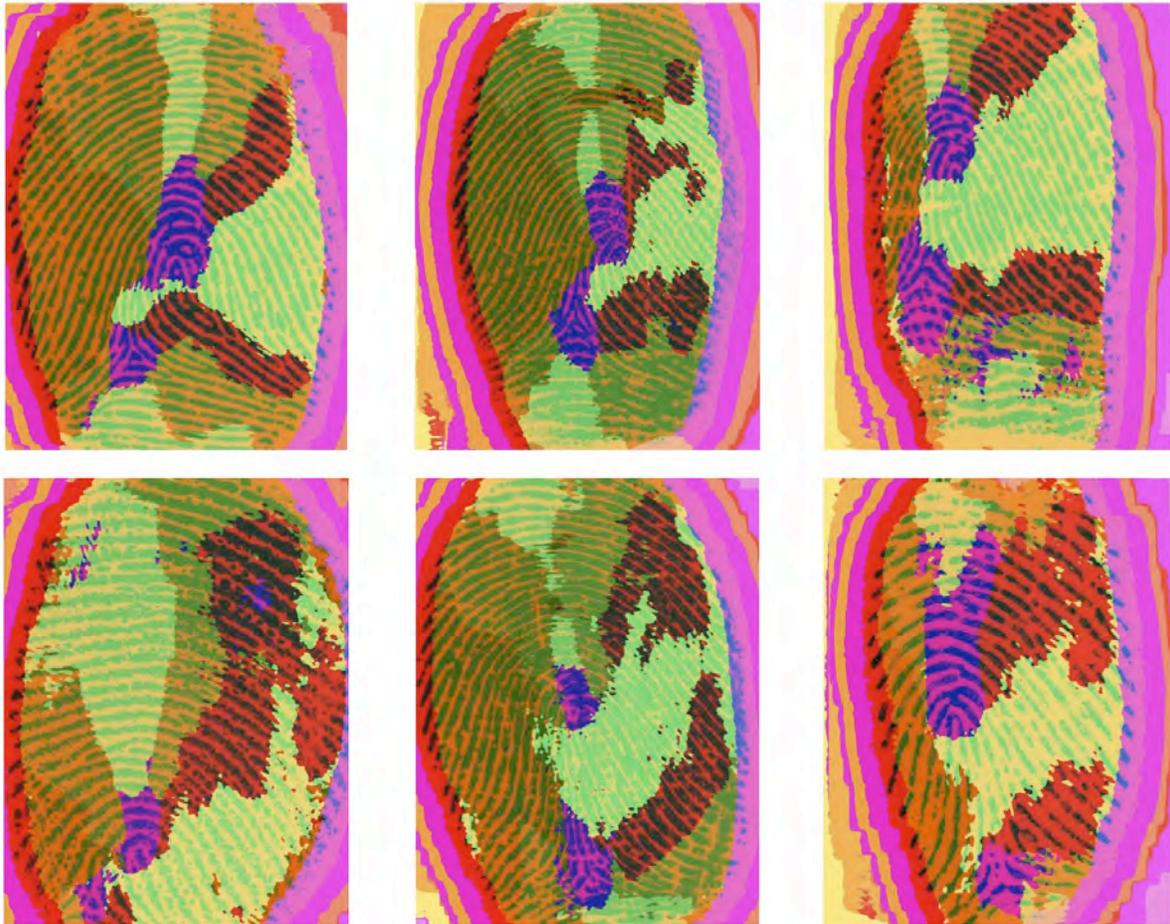


Figure 13. Clustering solution applied to 6 different fingerprints, illustrating how the algorithm can identify similar regions in different fingerprints regardless of the exact location of the core or delta (see purple areas).

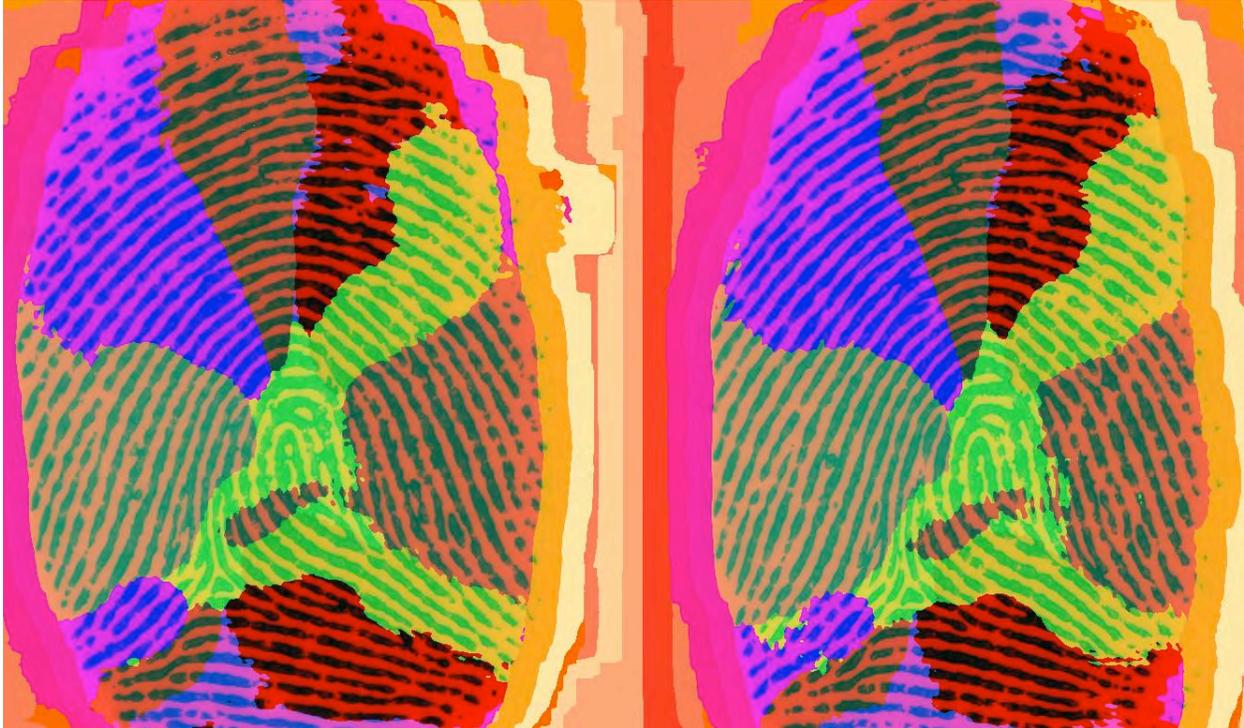


Figure 14. In addition to finding the same general feature on different fingerprints, the clustering algorithm will also automatically identify corresponding regions on two impressions from the same fingerprint. This illustrates the power of the ICA basis function approach, especially when clustering is used to identify regions.

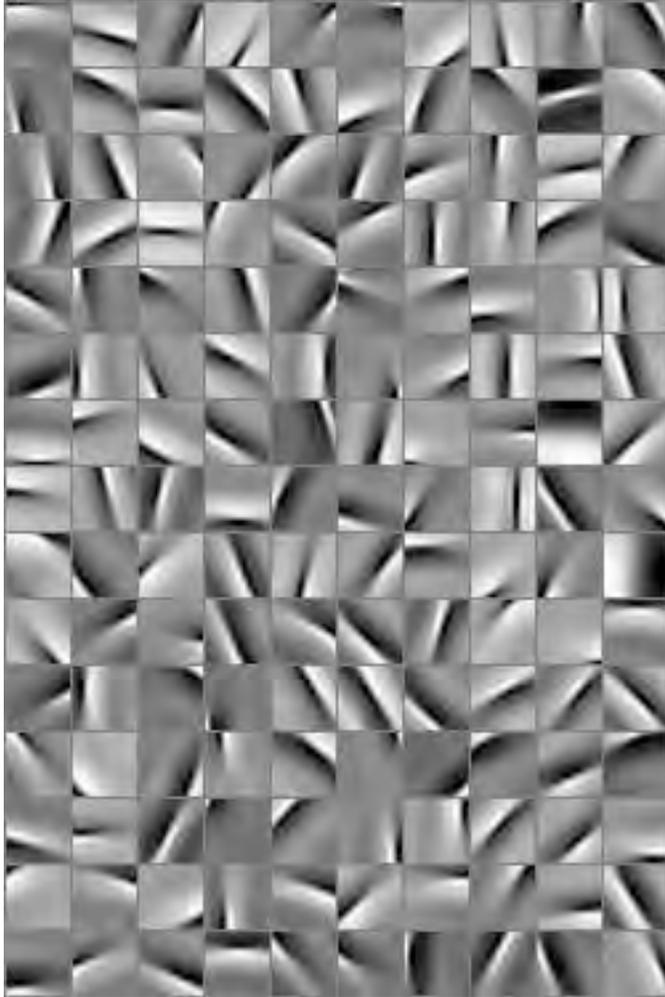


Figure 15. Examples of other ICA basis functions. This was constructed from a 24x24 pixel basis images, 150 basis functions.

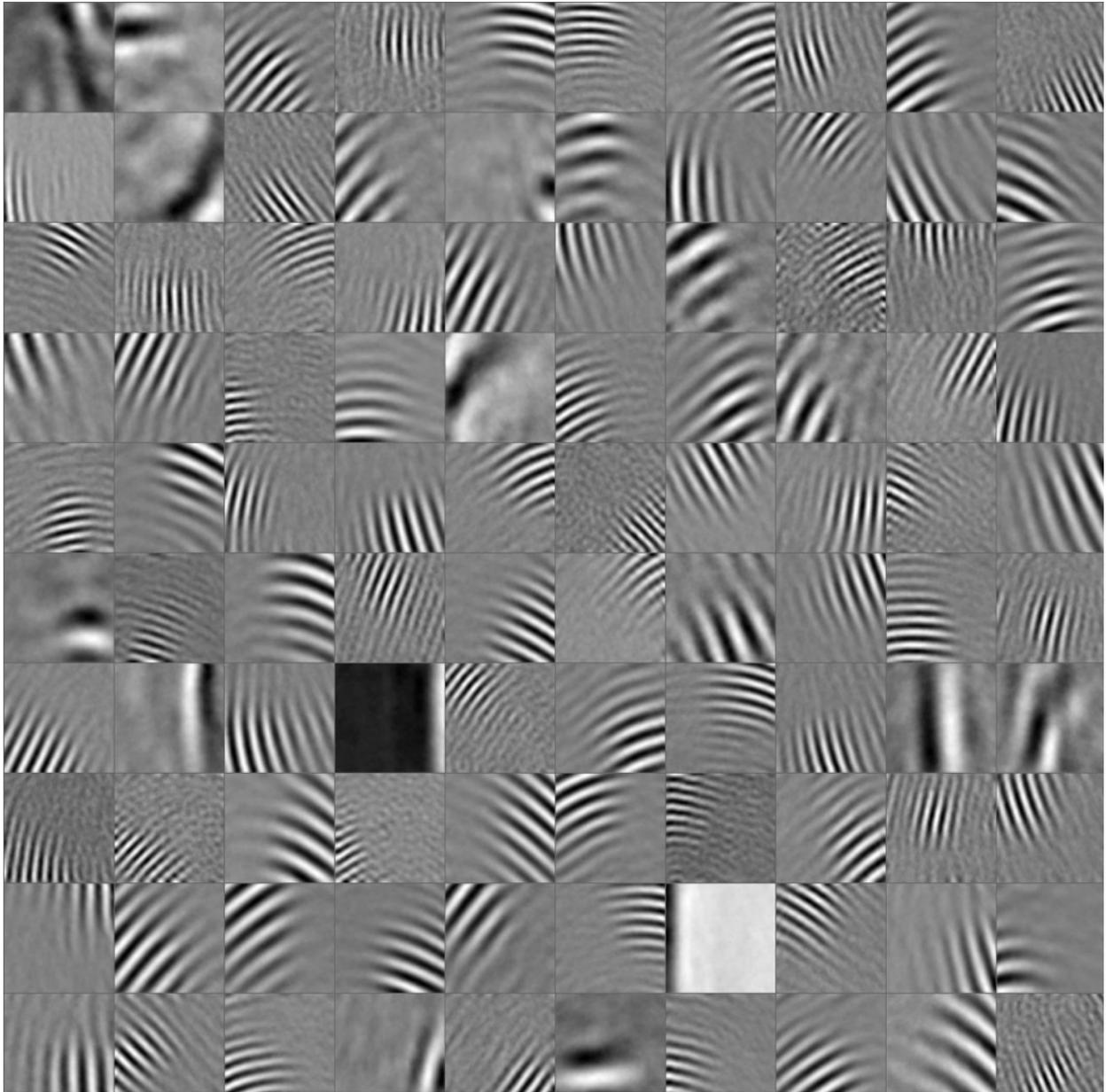


Figure 16. Examples of other ICA basis functions. This was constructed from a 128x128 pixel basis images, 450 basis functions (only some shown).

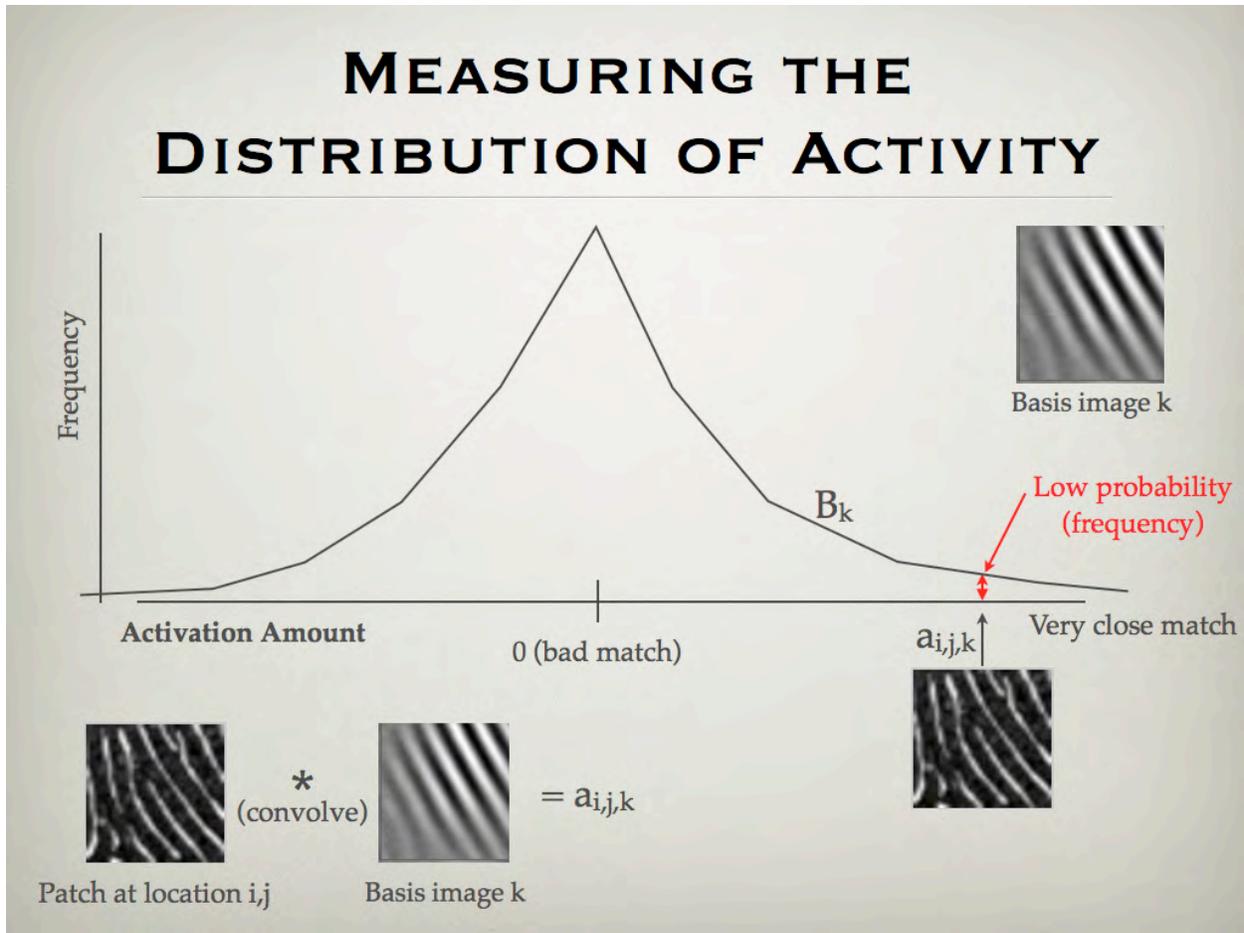


Figure 17. Computing self-information. By cross-correlating the basis images (the k 's) across hundreds of thousands of image patches in hundreds of images, we construct an estimate of how likely that basis image will take on a particular value of activation (B_k). For a particular activation $a_{i,j,k}$, we can then estimate the probability of observing an activation of that magnitude. This produces a probability estimate for this activation, which can then be combined with all other basis activations to produce the likelihood of observing that particular patch given the statistics of the database. This combination is through multiplication, following the independence assumption of ICA.



Figure 18. Diagnostic regions as revealed by the self-information metric applied using a 38x38 pixel patch size and a basis set size of 150 pixels. Dark regions are regions that the metric determines are most diagnostic, while light regions are less diagnostic. Edge effects are common with these techniques and can be ignored. The dark regions seem to correspond to features that humans would consider to be interesting, such as minutiae. However, the method also highlights regions of severe curvature even in the absence of minutiae, which also may be diagnostic.



Figure 19. A second image showing diagnostic regions with the same basis set used in Figure 18. Again the method seems to reveal regions that human observers would consider diagnostic.



Figure 20. Self-information map similar to that in Figure 18, except using a basis set of 360 basis functions rather than just 150.



Figure 21. A rarity map that is relatively unsuccessful at revealing diagnostic features. This particular map was generated using a basis set of 128x128, which is probably too large to reveal interesting features. This particular basis set does tend to reveal the region above the core in multiple images (not shown) which is often where examiners start their search. This example illustrates that the choice of basis set is relatively important for the self-information metric to work.



Figure 22. Another example of a basis set that fails to produce interesting results. In this case the patch size was only 16x16 pixels, which is too small to adequately capture detail at larger scales. This analysis is successful only at highlighting edges. This demonstrates that the choice of basis set size and patch size is critical.



Figure 23. Self-information maps for one impression, processed using basis function of different sizes. All maps were constructed using 150 ICA basis functions, and illustrate that different patch sizes highlight different kinds of information, ranging from fine detail to relatively large regions. Top left: Map with basis set size of 24x24 pixels highlighting local features. Top right: Map with basis set size of 64x64 pixels. Bottom left: Map with basis set size of 128x128 pixels. Bottom right: Map with basis set size of 160x160 pixels highlighting broad regions.



Figure 24. Strong predictability for a 160x160 pixel patch size basis set with 150 basis functions. The red dots corresponds to expert eye fixations, and there is close correspondence between the locations of the dots and the regions identified by the metric as diagnostic.



Figure 25. As in Figure 24 but with a different image to show generality. Again there is close correspondence between the eye fixations of experts and the regions identified as diagnostic.



Figure 26. Smaller numbers of basis functions may pick up different types of features. This is a 128x128 pixel patch basis set with only 16 basis images. It appears to be more specific in the regions it identifies.

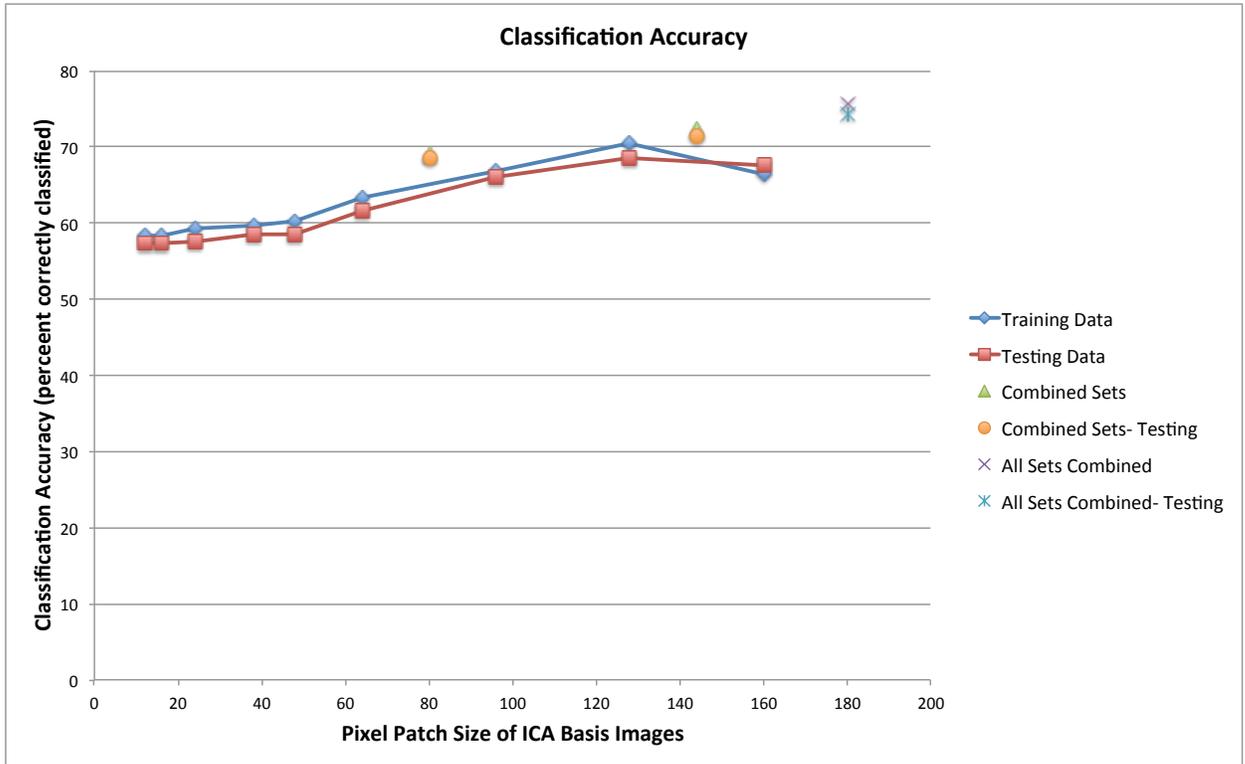


Figure 27. Classification accuracy as a function of pixel patch size for both training and testing sets. In this graph, the x-axis is the pixel patch size, and the y-axis is the classification accuracy for both training and testing sets. The non-connected points at 75, 150 and 180 correspond to combinations of pixel patches. The 75 combines 60 and 100, while the 150 combines the 128 and 160. The value shown at 180 represents all pixel patch sets combined.

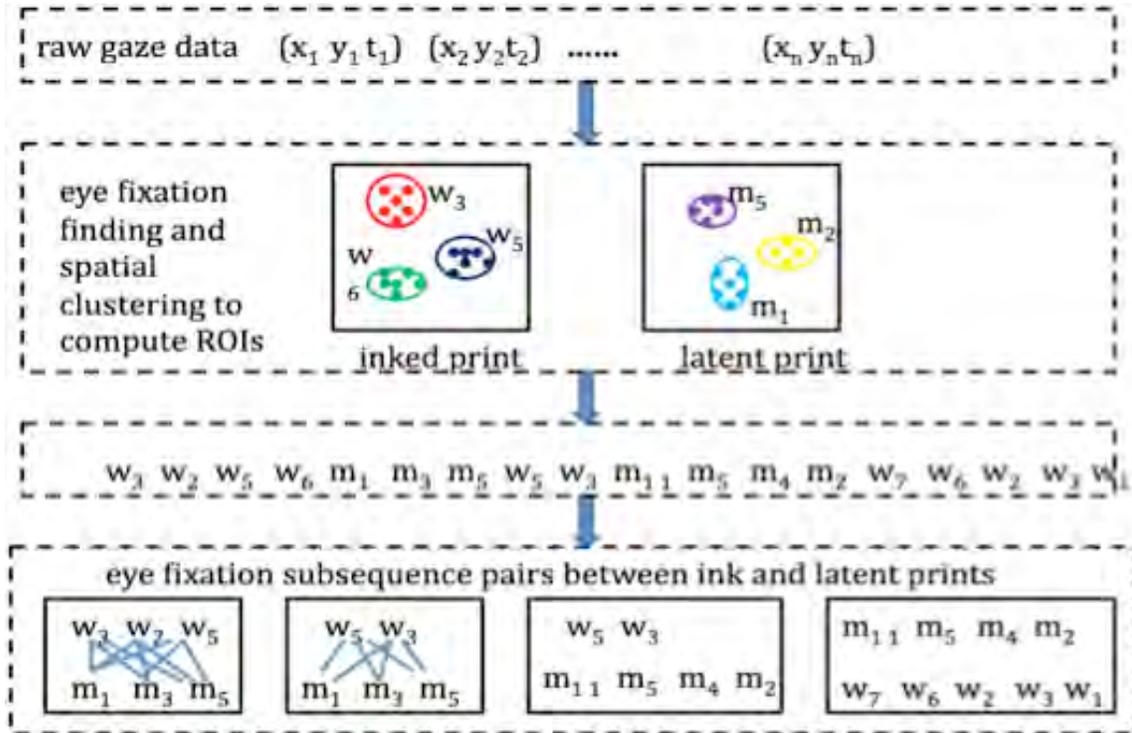


Figure 28. Overview of data processing for the extraction of temporal dependencies that can be used by machine translation procedures to identify regions of correspondence between two fingerprints.

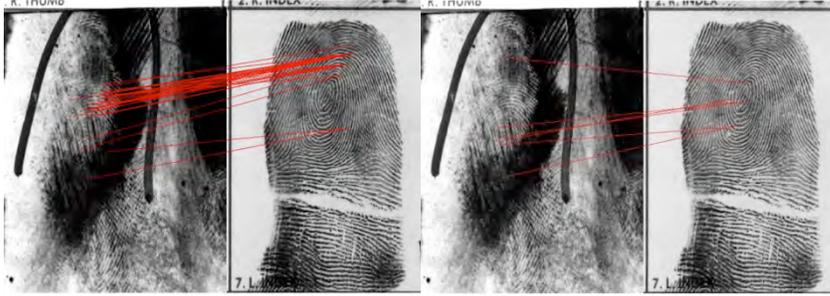


Figure 29. The corresponding regions from Data Set 1 with inked and latent prints. Left: an example result from experts. Right: an example result from novices.

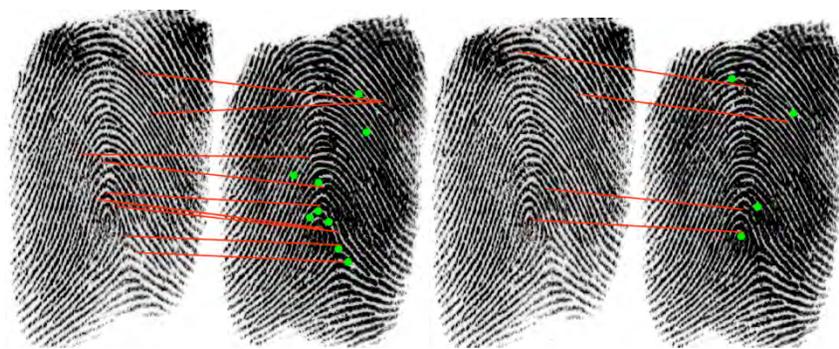


Figure 30. The corresponding regions from Dataset 2. Left: an example result from experts. Right: an example result from novices.

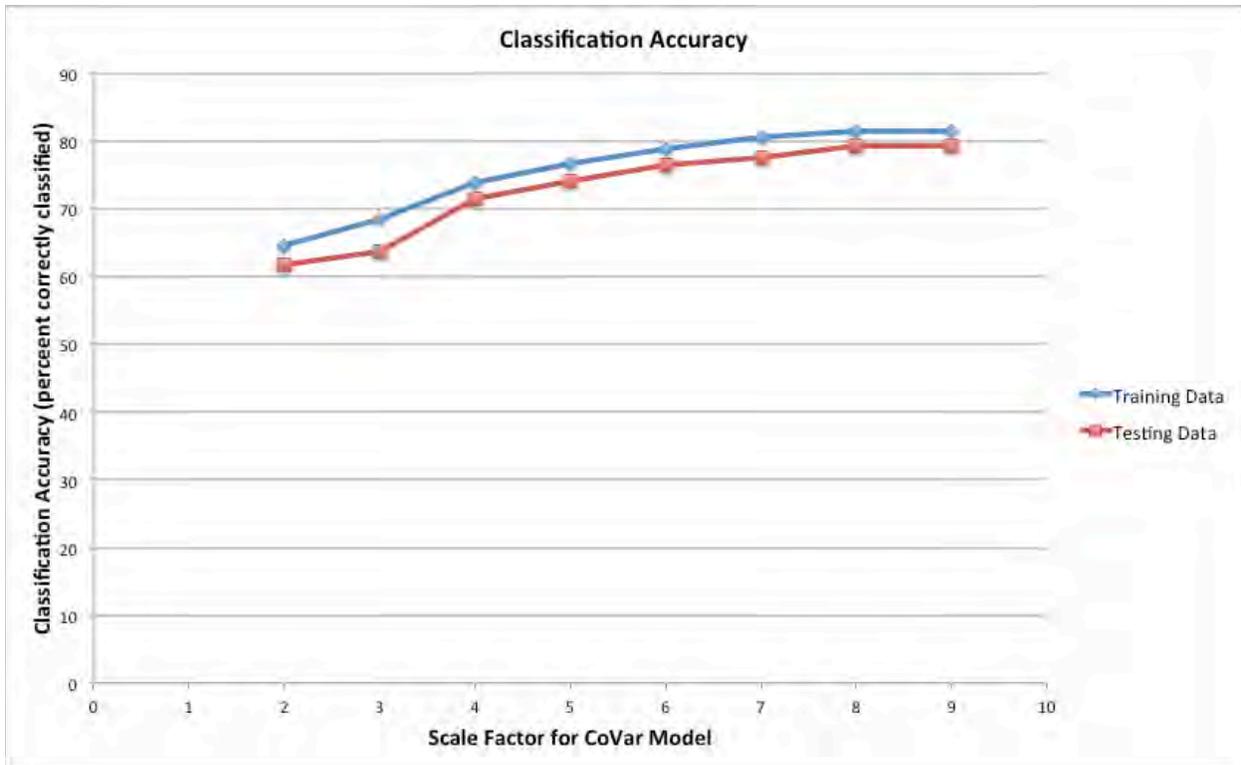


Figure 31. Classification accuracy as a function of spatial scale factor from the CoVar model both training and testing sets. Classification accuracy improves as the spatial scale increases, suggesting that examiners are relying on configural information in addition to simple features. The training and testing classification performance is similar, suggesting that the model readily generalizes to novel prints that were not used to train the system.



Figure 32. Example Saliency map for Spatial Scale 6 for a training image. Darker regions are those regions that the logistic model predicts experts will find most diagnostic, and the red dots are fixations from experts. The close correspondence between the two suggests that the model accurately captures those features that attract the eye gaze of experts.



Figure 33. Saliency map showing good correspondence at spatial scale 7 between the fixations and the regions deemed most likely to be associated with expert eye gaze by the logistic model (darker regions). These are training images.



Figure 34. Example Saliency map for spatial scale 7 for a pair of training (and in this case nonmatching) images. The model captures the tendency for the experts to focus on the core and region above the core, plus down the tail, while ignoring detail in the tip.



Figure 35. Left panel- spatial scale 2. Right panel- same image at spatial scale 6. Different spatial scales represent different information, and both seem to be necessary to capture the distribution of fixations for this print. These are training images.



Figure 36. Example Saliency map a pair of testing images (i.e. one that was not use to train the logistic regression classifier). Darker regions are those regions that the logistic model predicts experts will find most diagnostic, and the red dots are fixations from experts. The close correspondence between the two suggests that the model is able to readily generalize to new fingerprints and predict which regions will be visited by examiners.



Figure 37. Example Saliency map a pair of testing images (i.e. one that was not use to train the logistic regression classifier).



Figure 38. Example Saliency map a pair of testing images (i.e. one that was not use to train the logistic regression classifier). These are also non-matching images.

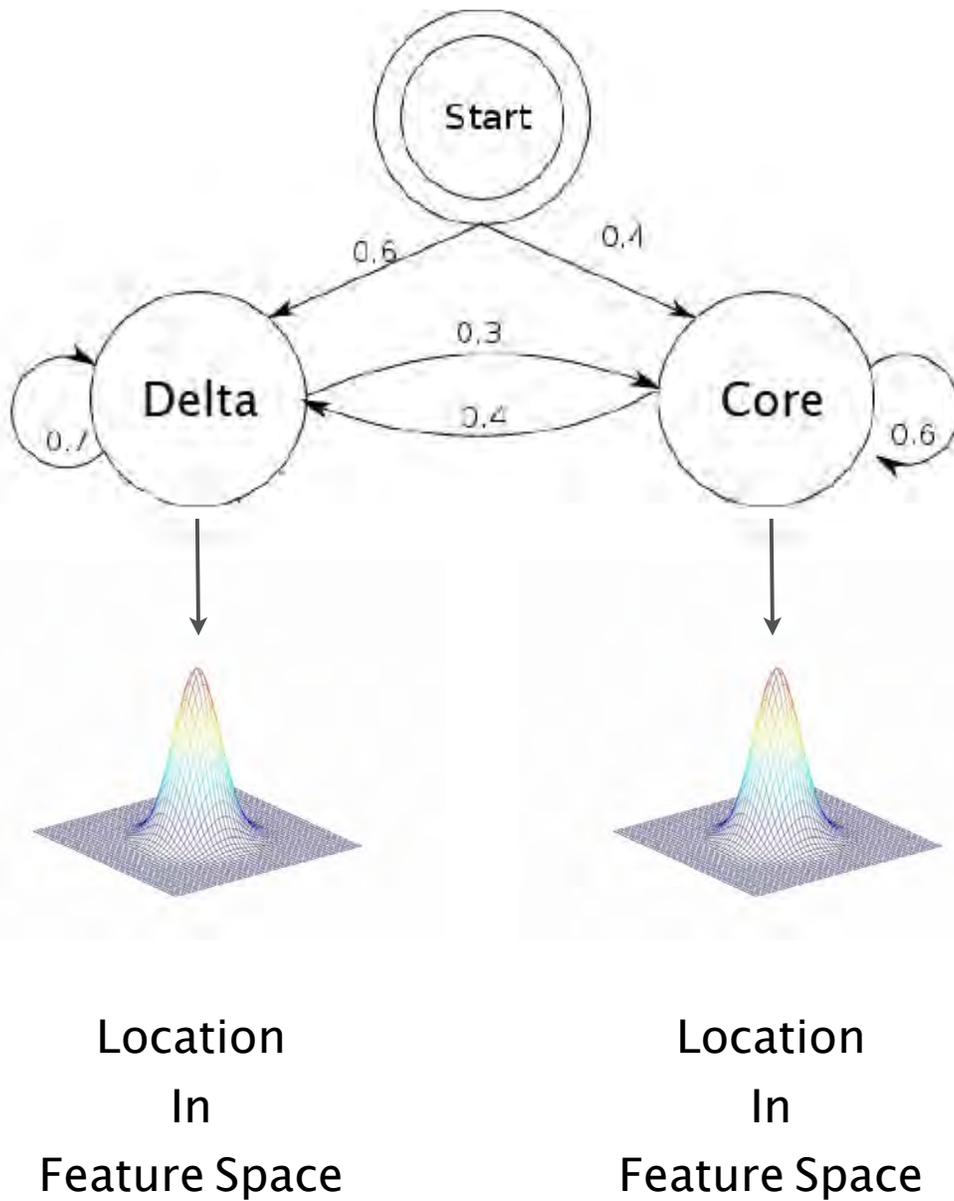


Figure 39. Illustration of the Hidden Markov Model that has two hidden nodes (here labeled Delta and Core but could represent any abstract area). Each hidden node has an emission matrix that is a multi-dimensional Gaussian projecting into the feature space. The fixations have a point in the feature space as defined by the basis function activations, and the emission matrix determines the probability that each fixation belongs to a hidden node. The HMM training fits both the transition probability matrix (as determined by the probabilities on the arrows) as well as the location of the Gaussians in the high-dimensional feature space. Thus the model fits both the temporal relations and determines what constitutes a region or feature for purposes of the HMM.

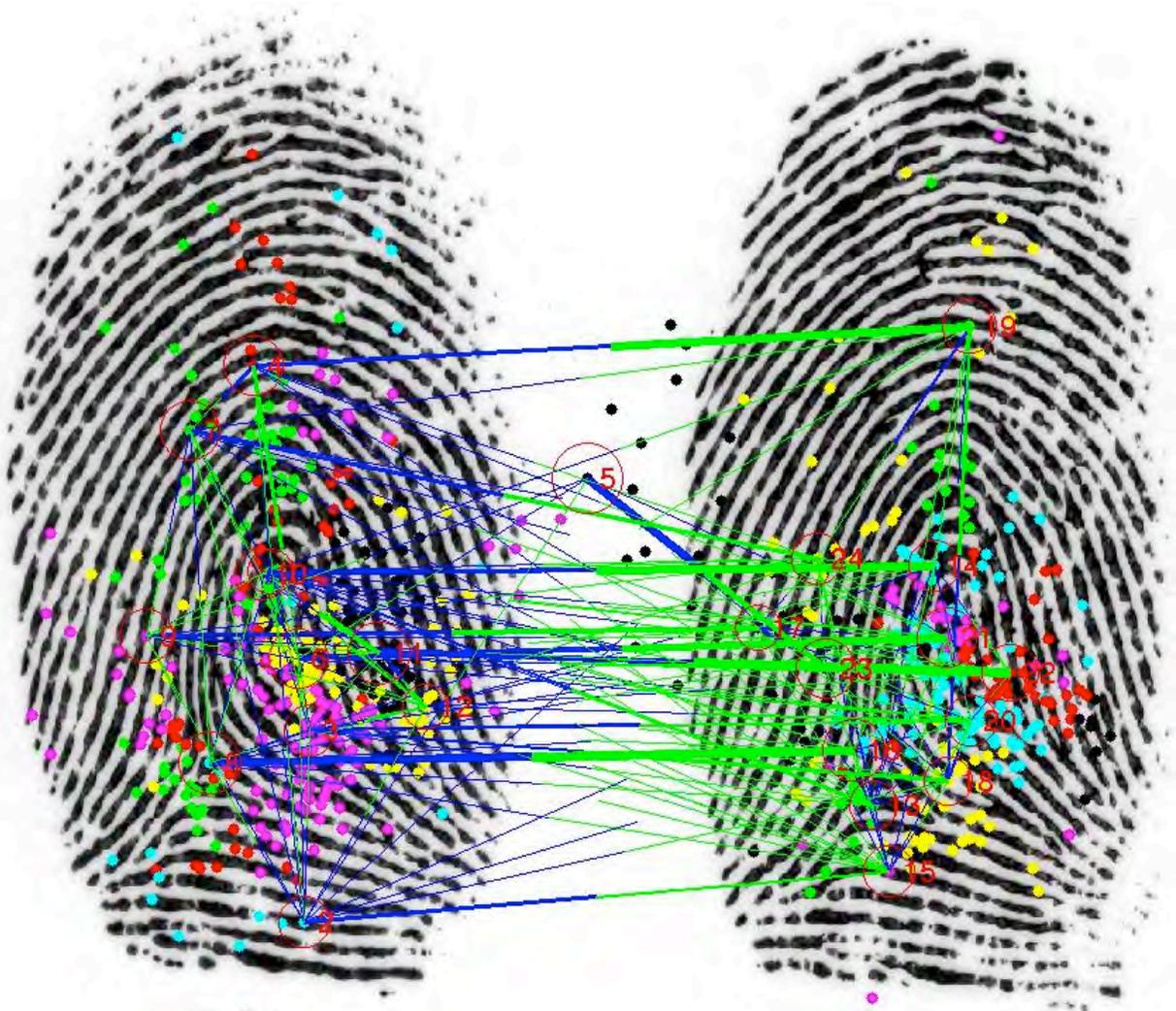


Figure 40. Transition probability matrix of the regions visited by fingerprint examiners as discovered by a Hidden Markov Model (HMM). The HMM uses the activations of each fixation to do spatial clustering while simultaneously fitting a set of transition probabilities between different hidden states using a Hidden Markov Model. Node 5 in the middle comes from the starting point at each trial, where the gaze is still in the middle of the screen after initiating the trial.

References

- Bruce, N. D. B., & Tsotsos, J. K. (2009). Saliency, attention, and visual search: An information theoretic approach. *Journal of Vision, 9*(3), -. doi: Artn 5
Doi 10.1167/9.3.5
- Busey, T., Silapiruti, A., & Vanderkolk, J. (under review). The Relation Between Sensitivity, Similar Non-Matches, and Database Size in AFIS Searches. *Law, Probability and Risk*.
- Dror, I.E., & Mnookin, J.L. (2010). The use of technology in human expert domains: challenges and risks arising from the use of automated fingerprint identification systems in forensic science. *Law, Probability and Risk, 9*(1), 47.
- 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. doi: 10.1016/j.forsciint.2006.06.054
- Karklin, Y., & Lewicki, M. S. (2009). Emergence of complex cell properties by learning to generalize in natural scenes. *Nature, 457*(7225), 83-U85. doi: Doi 10.1038/Nature07481
- 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. doi: Doi 10.1111/J.1556-4029.2006.00327.X
- Olshausen, B. A., & Field, D. J. (1997). Sparse coding with an overcomplete basis set: a strategy employed by V1? *Vision research, 37*(23), 3311-3325.
- Shannon, C. E. (1997). The mathematical theory of communication (Reprinted). *M D Computing, 14*(4), 306-317.
- Snodgrass, M., Bernat, E., & Shevrin, H. (2004). Unconscious perception at the objective detection threshold exists. *Perception & Psychophysics, 66*(5), 888-895.
- Srihari, S., & Su, C. (2008). Computational Methods for Determining Individuality. *Computational Forensics, 11-21*.
- Su, Chang, & 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.
- Ullman, S., Vidal-Naquet, M., & Sali, E. (2002). Visual features of intermediate complexity and their use in classification. *Nat Neurosci, 5*(7), 682-687. doi: 10.1038/nn870
- Vanselst, M., & Merikle, P. M. (1993). Perception Below the Objective Threshold. *Consciousness and Cognition, 2*(3), 194-203.