Semi-automated Identification of Leopard Frogs[*]

Dijana Petrovska-Delacrétaz[†,1], Aaron Edwards[2], John Chiasson[2], Gérard Chollet[2,3]
and David S. Pilliod[4,5]

[1]*Electronics and Physics (EPH), Department of the Mines Telecom SudParis, CNRS Samovar, Paris, France*
[2]*ECE Dept, Boise State University, Boise, 83725, ID, U.S.A.*
[3]*LTCI of CNRS, Institut Mines-Télécom, Paris, France*
[4]*U.S. Geological Survey, Forest and Rangeland Ecosystem, Science Center Boise, Boise, 83706, Idaho, U.S.A.*
[5]*Graduate Faculty of the Department of Biological Sciences at Boise State University, Boise, 83725, ID, U.S.A.*

Keywords: Animal Biometrics, Automatic Identification, Frogs, Principal Component Analysis.

Abstract: Principal component analysis is used to implement a semi-automatic recognition system to identify recaptured northern leopard frogs (*Lithobates pipiens*). Results of both open set and closed set experiments are given. The presented algorithm is shown to provide accurate identification of 209 individual leopard frogs from a total set of 1386 images.

1 INTRODUCTION

Identification of individual frogs in wild populations is important for biologists who are conducting demography studies used to evaluate the status and trends of endangered species. Wildlife biologists have used various methods to identify individuals in the wild, most of which involve some type of permanent or temporary mark or tag. These identification methods, while often reliable, may pose health risks to animals and thus there is a need for non-harmful alternatives. One of the most intriguing alternatives for animal identification is photography.

Photographically-based frog identification is conducted in the following manner. Biologists capture wild frogs from a study site (e.g., a pond), photograph them, and then release them back into the population. Later (e.g., days, weeks, months, or even annually), biologists return to the study site and capture another group of frogs, photograph them, and return them to the population. The biologists then try to match individual frogs from the second group (set) to individuals caught during the previous visit (or all pre-

vious visits). Individuals from the second group are then classified as "new" or "recaptured", depending on whether they were captured during previous surveys. This visual matching approach works well for small sets of frogs, but becomes burdensome or even impossible as the number of frogs captured increases.

The identification problem is to determine from the photograph if a captured frog is in the existing database of photographs or is a new frog. Humans can identify the frogs quite accurately based on the shape and location of spots or other features on their skin. For example, in (Lama et al., 2011) the tree frog *Scinax longilineus* was successfully identified by researchers simply looking at the collected photographs and they found that photo-identification was as accurate as tagging the animals. However, as databases of photographs become large, this visual matching approach is unrealistic. Instead, researchers are examining ways to automate this process through computer-aided pattern recognition.

One of the first steps in pattern recognition is to identify the area of an animal that will be used for pattern matching. To accomplish this we adopted an existing tool developed by a research team at Idaho State University (Velásquez, 2006), (Kelly, 2010). An example is shown in Figure 1 (Kelly, 2010) which shows the dorsal (i.e., back) side of the captured frog and indicates the area of its backside which is cutout for use in the identification. The cutout portion follows natu-

---

[*]Any use of trade, product, or firm names is for descriptive purposes only and does not imply endorsement by the U.S. Government.

[†]This work was done during a sabbatical stay of Dijana Petrovska-Delacrétaz in the ECE Dept at Boise State University, Boise ID 83725

ral contours of the frog's backside. This area (referred to as the *region of interest*) is then stretched to make a rectangular array of pixels as shown in Figure 2. The details of this stretching procedure are given in (Kelly, 2010).

Ideally, one wants an automatic procedure to identify the frog, i.e., determine whether or not it is in the current database. Here we use the terminology semi-automatic to mean identification of the frog based on user's manual selection of the region of interest as indicated in Figure 1. This manual intervention is quite easy in terms of the user's effort. In the recognition approach used in (Kelly, 2010), the cutouts were then manually segmented to identify the spots and then engineered features were developed for the identification procedures. However, the segmentation turned out to be a rather tedious task to perform on each photograph.
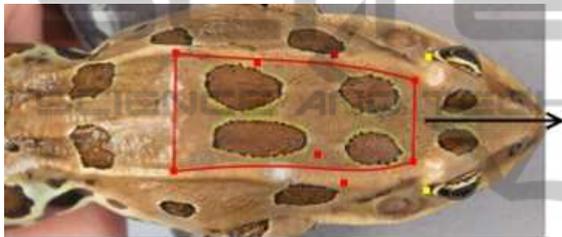

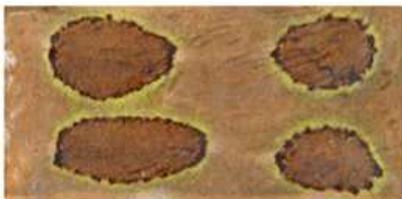Figure 1: Cutout along natural contours of the frog (Kelly, 2010).

Figure 2: The cutout is stretched to form a rectangle (Kelly, 2010).

Work similar to that presented here was done by Gamble et al. (Gamble et al., 2008) who used principal component analysis (PCA) on normalized images of marbled salamanders (*Ambystoma opacum*). Specifically, they used a cutout of the back of the salamander as a vector in $\mathbb{R}^{640 \times 480}$ and then went through a series of preprocessing steps to handle nuisance variables to obtain $M = 625$ "new" images for each original image. Each of these images was then scaled 8 times (multi-scale in half-octaves from 1 to $8\sqrt{2}$) using a Gaussian filter and appended to the original image so that the feature vector was now in $\mathbb{R}^{9 \times 640 \times 480}$. Their database consisted of 366 different salamanders and a total of 1008 images. In their closed-set experiments, they reported a 95% correct

identification that the test image was in the top 10 matches.

In (Azhar et al., 2012) a texture based image feature descriptor called the Local Binary Patterns (LBP) was used for the (semi) automatic identification of the Great Crested Newt salamander (*Triturus cristatus*). They tested on a database of 40 newts and 153 images. Similar to the frog cutout procedure described above, they used normalized images and manually extracted a part of the belly images as the source of biometric information. They considered both open and closed set test procedures.

The goal of this paper is to provide a simple, fast, and efficient semi-automated pattern-recognition algorithm for a capture-recapture identification system for northern leopard frogs (*Lithobates pipiens*).

Section 2 describes the databases of frog pictures we used in the experiments. Section 3 discusses how a PCA algorithm is used to do the animal recognition (identification), Section 4 describes the evaluation protocols, Section 5 presents the experimental results and Section 6 gives the conclusions.

2 ANIMAL DATABASE

The database consisted of images of northern leopard frogs with 209 separate identities. The cutouts of the frogs described in the introductory section are all rectangular arrays of 256×128 pixels (see Figure 2) and converted to grayscale. This leopard frog database was provided by the research of Oksana Kelly (Kelly, 2010). Kelly obtained 209 frogs bred in captivity and photographed them. A photographic light diffusing dome (Cloud Dome, www.clouddome.com) was used to take an average of 3 to 4 images per frog for all 209 identities, although some frogs had up to 11 images. The light diffusing dome reduced glare from sunlight, which helped improve image quality. Compare Figure 3 with Figure 4. We had 966 images taken with the dome (hereafter, referred to as Shade Dome images). There were also 420 additional images taken of frog identities 109-209 that did not use the shaded dome (hereafter, referred to as No Dome images). These images were of significantly lower quality due to glare (See Figure 4). With the combination of the No Dome and Shade Dome images, the Captive Leopard frog database contained 1386 total images.

3 RECOGNITION PROCEDURE

The image capture follows the procedure discussed in the Introduction. We followed the "finger-

Figure 3: Photo taken using a shade dome (Kelly, 2010).

Figure 4: Photo taken without using a shade dome (Kelly, 2010).

print" extraction procedure as described in (Kelly, 2010). The open-source program IDENTIFROG (http://code.google.com/p/identifrog (Pilliod et al.,)) was used to obtain the rectangular cutouts made up of 256×128 pixels as shown in Figure 2. This was then converted to grayscale for use in the recognition procedure. We remark that these images of the dorsal pattern contained within the fingerprint boundaries can and do vary in size due to the original image scale variation, frog positioning, the user's selection of the lateral corners of the eyes, and boundary alignment to the dorsolateral folds.

3.1 Feature Extraction and Identification

We used Principal Component Analysis (PCA) which was developed over 100 years ago for statistical analysis (Pearson, 1901)(Hotelling, 1936). It is also a well-known method in Machine Learning (Barber, 2012), but for pattern recognition it relies on the images being normalized. This approach requires determining a set of images to make up the PCA space (covariance matrix), the choice of eigenvectors, and the choice of an appropriate distance measure. We use

one set of images (development images) to set these choices and then test on an independent set of images for the evaluation.

Each frog image $x^{(k)}$ is considered to be in $\mathbb{R}^d, d \triangleq 256 \times 128$ and with N_t the number of training images, the covariance of the training set is

$$C \triangleq \frac{1}{N_t - 1} \sum_{k=1}^{N_t} (x^{(k)} - x_m)(x^{(k)} - x_m)^T \in \mathbb{R}^{d \times d} \quad (1)$$

$$x_m \triangleq \frac{1}{N_t} \sum_{k=1}^{N_t} x^{(k)} \in \mathbb{R}^d. \quad (2)$$

The rank of C is less than or equal to $N_t - 1$ and in our case (typical) $N_t << d \triangleq 256 \times 128$. As C is a positive semi-definite symmetric matrix, there is an orthogonal matrix $Q \in \mathbb{R}^{d \times d}$ such that

$$C = Q diag(\lambda_1, ..., \lambda_{N_t-1}, \underbrace{0, ..., 0}_{d-(N_t-1)})Q^T. \quad (3)$$

That is, the ith column of Q is the ith eigenvector of C with eigenvalue λ_i. Further, $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_{N_t-1} \geq 0$. We can then represent any training image $x^{(k)} \in \mathbb{R}^d$ as

$$h^{(k)} \triangleq Q^T(x^{(k)} - x_m) \in \mathbb{R}^d \quad (4)$$

since we get the image x back by

$$x^{(k)} = x_m + Q h^{(k)} \in \mathbb{R}^d. \quad (5)$$

However, the point of this approach is to obtain a compressed representation of the image x by representing it by its first N eigenvectors where $N < N_t << d$. That is, the image is coded into $\mathbb{R}^N$ by

$$h_c \triangleq \begin{bmatrix} h_1 & h_2 & h_3 & \cdots & h_N \end{bmatrix}^T \in \mathbb{R}^N \quad (6)$$

where h_c is simply the first N components of $h \triangleq Q^T(x - x_m) \in \mathbb{R}^d$. With $Q_c \in \mathbb{R}^{d \times N}$ the first N columns of Q, the theory of PCA (Barber, 2012) tells us that reconstruction error is given by

$$\|x - Q_c h_c\|^2 = \lambda_{N+1}^2 + \cdots + \lambda_{N_t-1}^2. \quad (7)$$

N is chosen so that $\lambda_{N+1}^2 + \cdots + \lambda_{N_t-1}^2$ is small. Thus, as far as the Euclidean norm in concerned, the PCA representation $h_c \in \mathbb{R}^N$ is a much lower dimensional representation of the data than the original data vector $x \in \mathbb{R}^d$ yet provides an accurate reconstruction of the image.

Let the feature vector be

$$f = \begin{bmatrix} h_A & \cdots & h_N \end{bmatrix}^T \in \mathbb{R}^{N-A+1} \quad (8)$$

which indicates that we are representing the image by eigenvectors A through N. Our choice will turn out to be $A = 3, N = 120$, that is, we remove the first two

components h_1, h_2 from h_c to obtain better identification accuracy in contrast to reconstruction accuracy.

The basic test is as follows: Let $x^{(k)}$, $k = 1,...,N_t$ be the N_t images in the database and $f^{(k)}$ their corresponding feature vectors. Let x be any test image (recapture) with its feature f computed as above. For $k = 1,...,N_t$ compute the (cosine of the) minimum angle between the new image and the existing images, that is, compute

$$s^{(k)} \triangleq \frac{f^T f^{(k)}}{\|f\| \|f^{(k)}\|}. \tag{9}$$

This value $s^{(k)}$ is referred to as the *score* between the test image and the kth image in the database. Let k^* be defined by

$$k^* \triangleq \underset{k \in \{1,...,N_t\}}{\arg\max} \left\{ s^{(k)} \right\} \tag{10}$$

which we will refer to as the identified image.

In a *closed-set* protocol the test image of the frog is assumed to be in the database, such as when a frog is recaptured during a second sampling event. One then identifies the test image x as the image $x^{(k^*)}$. In practice one typically finds the (say) 10 images in the database that score the highest with x and then checks which of the 10 match the test image.

In an *open-set* protocol the test image may or may not correspond to any frog from the reference database, which is a more realistic test when a frog is captured and its identity unknown. We again compute k^* as just explained and, with γ some pre-determined threshold, we check if

$$s^{(k)} \geq \gamma. \tag{11}$$

If this is true then x is identified as the image $x^{(k^*)}$ else we say x is a new identity. Again, in practice, one typically finds the (say) 10 images in the database that score closest to the test image and then visually checks if it matches these already known identities.

Figure 5 shows the mean frog image x_m and the first five eigenvectors (eigenfrog images) of the covariance matrix constructed using all of the 1386 frog images (shade dome and no shade dome). Note that the spots are quite blurred in the first two eigenfrog images hinting that the first two eigenvectors may not contribute much to differentiation among frog identities.

4 EVALUATION PROTOCOLS

We consider both open and closed evaluation protocols. To explain the evaluation protocols, we describe them for the database consisting of the 209 captive
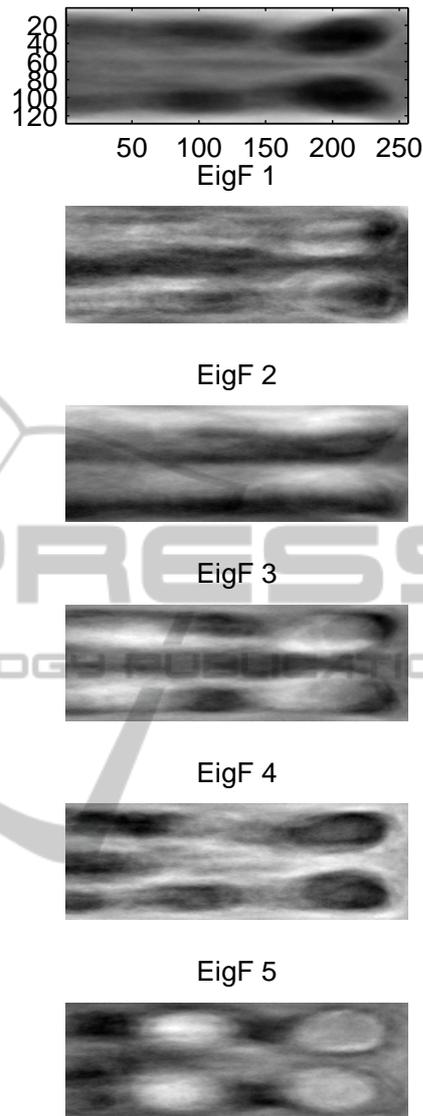

EigF 1

EigF 2

EigF 3

EigF 4

EigF 5

Figure 5: Eigenfrogs.

frog identities with a total of 966 frog images taken with the shade dome. The number of images for each frog identity ranged from 2 to 11 with the majority of the frog identities having 3 to 4 images.

4.1 Closed Set Evaluation Protocol

The closed set protocol assumes that a test frog is in the database. These frog images are in a file listed with the number of the frog and the number of its image. For example, R_001_01, R_001_02, R_001_03, R_001_04 are the 4 images we have of frog 1, R_002_01, R_002_02, R_002_03, R_002_04, ..., R_002_11 are the 11 images we have of frog 2, etc. We then distributed the images

into 5 bins as follows: We put R_001_1 into bin 1, R_001_2 into bin 2, R_001_3 into bin3, R_001_4 into bin 4, R_002_1 into bin 5, R_002_2 into bin 1, R_002_3 into bin 2, R_002_4, into bin 3, etc. This was done in order to mix the images of each frog identity well among the bins. This mixing results in the bin distribution given in Table 1.

Table 1: Bin distribution for the closed set 5-fold protocol.

Bin Number	Number of Images
bin 1	194
bin 2	193
bin 3	193
bin 4	193
bin 5	193

After putting all the frog images into the 5 bins as just described, we used the first four bins to compute the covariance matrix C (PCA subspace). C was therefore constructed from $N_t = 3 \times 193 + 194 = 773$ images. The images in the 5th bin were used for testing. We take the identified image to be $x^{(k^*)}$ where k^* is as given in equation (10). The 5th bin had 193 images[3] and three of the images were incorrectly identified for an accuracy of $190/193 = 98.5\%$. We then repeated the procedure four more times using a different bin as the test bin, that is, 5-fold test. The results are in in Table 2.

Table 2: Closed set 5-fold protocol.

test bin	bin 5	bin 4	bin 3
accuracy	$\frac{190}{193} = 98.5\%$	$\frac{188}{193} = 97.4\%$	$\frac{187}{193} = 96.8\%$

test bin	bin 2	bin 1
accuracy	$\frac{187}{193} = 96.8\%$	$\frac{192}{194} = 99\%$

This shows a total of 22 errors over the five folding (changing the bins five times) tests on the 966 images.

4.2 Open Set Evaluation Protocol

The open set protocol refers to the situation where the test image (a frog) may or may not be in the database. In this case we take all 966 of the frog images (209 identities) and separate them into two groups: 804 known frog images with 151 identities and 162 unknown frog images with 58 identities. The identities in these two sets are *disjoint*. As in the closed-set protocol, the 804 known frog images are distributed into 5 bins of approximately the same size (804/5 giving 161 or 160 images per bin). See Table 3. The covariance matrix (PCA space) is computed using 4 of the 5 bins in the known group. Then the 5th bin and the unknown or 6th bin were used to test.

---

[3]966/5 = 193 remainder 1. So 4 of the bins had 193 images and the other bin has 194 images.

Table 3: Bin distribution for the open set protocol.

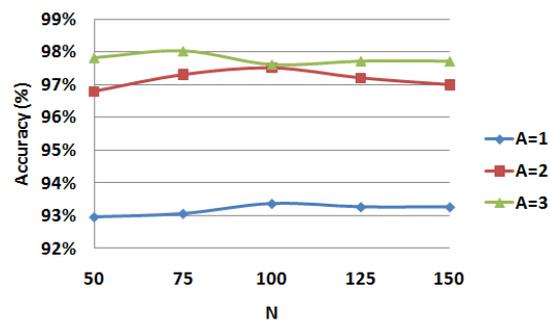| Bin Number	Number of Images
bin 1	161
bin 2	161
bin 3	161
bin 4	161
bin 5	160
bin 6 (unknown frogs)	162

This was repeated a total of five times by permuting bins 1 through 5 made up of images from the known group. The threshold in (11) was chosen to be $\gamma = 0.5$. An error occurs in one of two ways: (*k*-known) the test image is in the database, but the identified image $x^{(k^*)}$ is not the correct one or (*u*-unknown) the test image and its identified image $x^{(k^*)}$ satisfy the threshold, but the test image in *not* in the database. The results are given in Table 4 where u says the error was made on an unknown frog while k says that a known frog was misidentified.

Table 4: Open set 5-fold protocol.

test bins	bins 5&6	bins 4&6
errors	16 u, 6 k	16 u, 5 k
accuracy	$\frac{322-22}{322} = 0.93\%$	$\frac{323-21}{323} = 93.5\%$

test bin	bins 3&6	bins 2&6	bins 1&6
errors	11 u, 4 k	14 u, 5 k	15 u, 3 k
accuracy	$\frac{323-15}{323} = 93.4\%$	$\frac{323-19}{323} = 94.1\%$	$\frac{323-18}{323} = 94.4\%$

4.3 Determination of the PCA Space

In using the PCA test we used the feature vector given in (8) with $A = 3, N = 125$. That is, we represented the image as a linear combination of eigenvectors 3 through 125 with the coefficients of this representation the feature vector. To make this determination we

Figure 6: PCA accuracy as A, N vary.

repeated the closed set evaluation of Subsection 4.1 using eigenvectors A to N where A was varied from 1 to 3 and, for each value of A, N was varied from 50 to 150. The results are shown in the graph of Figure 6. The graph shows $A = 3, N = 125$ give good results.

5 EXPERIMENTAL RESULTS

Using the open and closed set protocols explained in the previous section, we performed tests on our database of images.

5.1 Closed Set of Experiments

5-fold Testing on Shade Dome Images

Consider again the closed set evaluation with the captive frogs consisting of 209 identities, 966 total images all taken with the shade dome. We have already given the results of the first test. However, up to now we have considered the identified image to be that one in the database with the highest score [See equations (9) and (10)]. We now will compare the test image with the database images with the n highest scores where we typically take $n = 1, 5$, or 10. In the present case we have

Table 5: Shade dome - closed set - 5 fold.

Top n	1	5	10
Avg. errors/fold	4.4	1.6	1.4
Accuracy	97.7%	99.2%	99.3%

For example, in subsection 4.1 we reported the errors for each fold. There were a total of 22 errors in which the test image did not match the highest score. This results in average number of errors over the 5 folds given by $22/5 = 4.4$. In the case where $n = 10$, we see that there were $5 \times 1.4 = 7$ times over the five folds that test image was not one of the top 10 scores.

PCA Constructed from Shade Dome Images Tested on No Dome Images

In the next experiment we used the 966 *shade dome* images to build the PCA space and then tested on the 420 *no dome* frog images. As previously mentioned, the no dome images only contained the identities of frogs 109-209. In Table 6 we report the total number of misidentification errors.

Table 6: PCA constructed from shade dome images. Tested on no dome images.

Top n	1	5	10
No. errors	51	31	24
Accuracy	87.9%	92.6%	94.3%

5.2 Open Set Experiments

We next performed an open set experiment using the 966 captive shade dome frogs with 209 identities. As previously explained in Subsection 4.1, we chose 58 identities with 162 images to be the "unknown" frogs.

The remaining 804 images with 151 identities made up the "known" frogs. The 804 "known" frogs were then split into a 5 fold (bins). This was used to make 5 folds (iterations) where for each fold one of the "known" bins (˜161 frogs) along with the 162 "unknown" frogs were used for testing while the remaining ˜643 frogs were used for development of the PCA space (covariance matrix). The threshold was set as $\gamma = 0.5$. In Table 7 are the results of only testing the known frogs and thus is simply the closed set protocol (γ not used) for the known frogs.

Table 7: Shade dome - open set - 5 fold.

Known frogs - Top n	1	5	10
Avg. errors/fold	2.6	0.4	0.2
Accuracy	98.38%	99.75%	99.88%

However, with the threshold $\gamma = 0.5$ it was found during the 5-fold testing that on average 1.8 of the known frogs scored below this threshold and thus would be categorized as a unknown (new) frog.

For the known frogs who met the threshold, Table 8 gives the error results.

Table 8: Known frogs with scores $s^{(k)} \geq \gamma = 0.5$.

Top n	1	5	10
Avg. Errors/fold	1.4	0.0	0.0
Accuracy	99.13%	100%	100%

With the unknown frogs an error occurs when its score $s^{(k)}$ meets the threshold, i.e., $s^{(k)} \geq \gamma = 0.5$ because it is then taken to be in the database when of course it is not there. Our results are below.

Table 9: Unknown frogs with $s^{(k)} \geq \gamma = 0.5$.

Top n	1	2	3	4	5
Avg. errors/fold	14.4	5.2	2.4	0.6	0

To explain Table 9, the second column ($n = 1$) means that on average there were 14.4 unknown frog images whose top score with some known frog image was greater than 0.5. The third column ($n = 2$) of this table means that on average there were 5.2 unknown frog images whose top 2 scores with some known frog images were greater than 0.5. Similarly for the remaining columns. The point here is that if an unknown frog has a score $s^{(k)}$ against the known database with $s^{(k)} \geq \gamma = 0.05$ then there can be at most four images in the database that satisfy this threshold and the biologist need only look at these four images to visually make the determination that the frog is not in the database.

5.3 Threshold

We chose the threshold $\gamma = 0.5$. This is based on the data given in the open set evaluation test discussed in the previous subsection (Subsection 5.2). Figure 7 shows two probability density functions (pdf). The pdf in Figure 7 labeled "unknown" frog pdf was obtained by taking the image of each known frog and computing its score with each unknown frog image. A histogram of these scores was then normalized to become the unknown frog pdf.

In contrast, the "known" frog pdf of Figure 7 was obtained by taking each known frog and computing its score with itself (i.e., with all possible images of its identity) and keeping the highest score. More precisely, as explained in (the previous) subsection 5.2, the 804 known frogs were put into (essentially) 5 equal sized bins. The PCA space (covariance matrix) was built from four of the bins. Then each identity in the remaining (test) bin had its score computed with images of the *same* identity in the other 4 bins. The highest score was kept. This was done five times each time using a different bin as the test bin. A histogram made of these scores was normalized to become the known frog pdf. From the pdfs of Figure 7 the prob-
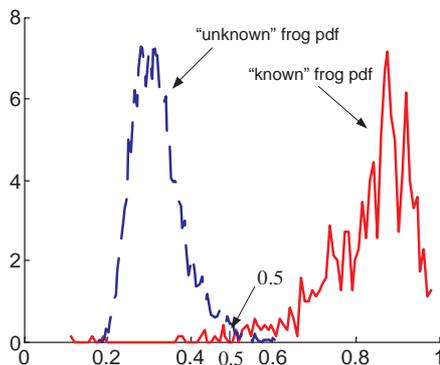

Figure 7: PDF of the scores of known frogs with unknown frogs and PDF of the scores of known frogs with themselves.

ability of detection of a known/unknown frog and the probability of a false alarm of a known/unknown frog were computed and are given in Table 10. For example, the area under the "unknown" frogs pdf from $-\infty$ to $\gamma = 0.5$ is 0.9836 which is the probability that the score of an unknown frog with a known frog is less than 0.5. The area under the "known" frogs pdf from 0.5 to ∞ is 0.9838 which is the probability that the score of a known frog with a known frog is greater than 0.5.

Of course one should have a separate database for an evaluation to determine a threshold. However, our

Table 10: Probability of detection.

Test\Truth	Unknown	Known
Unknown	98.36%	1.62%
Known	1.64%	98.38%

limited amount of images precluded such an opportunity. In practice, the wildlife biologist will not use a threshold. Instead, the biologist would typically take a new image and bring up the (say) 10 images in the database with the top 10 scores. Then a visual check would be made to determine whether or not the frog is a new identity. Another way to say this is that the closed set results matter most to the biologist.

6 CONCLUSIONS

This work was originally motivated by the previous work of Velásquez (Velásquez, 2006) and Kelly (Kelly, 2010). Here we have reported in Table 5 quite good closed set results which we surmise is due to the quality of the images. However, table 6 shows that when "training" (i.e., constructing the PCA space) on high quality (shade dome) images and then testing on lower quality (no dome) images the identification accuracy deteriorates. Though the open set results given in Table 4 are not nearly as good as the closed set results given in Table 2, Table 9 shows that the biologist needs to visually check less than five images in the database to determine if the captured frog is a new identity or not.

We are in the process of collecting more data to have enough for both development and evaluation databases. Another goal is to provide a reference system (based on PCA) with a publicly available reference database, so that the other researchers can compare their results to ours.

ACKNOWLEDGEMENTS

REFERENCES

Azhar, M., Hoque, S., and Deravi, F. (2012). Automatic identification of wildlife using local binary patterns. In *IET Conference on Image Processing (IPR 2012)*. London UK.

Barber, D. (2012). *Bayesian Reasoning and Machine Learning*. Cambridge.

Gamble, L., Ravela, S., and McGarigal, K. (2008). Multi-scale features for identifying individuals in large biological databases: An application of pattern recognition technology to the marbled salamander *Ambystoma Opacum*. *Journal of Applied Ecology*, 45:170–1180.

Hotelling, H. (1936). Relations between two sets of variates. *Biometrika*, pages 321–377.

Kelly, O. V. (2010). *Automated Digital Individual Identification System with an Application to the Northern Leopard Frog* Lithobates pipiens. PhD thesis, Idaho State University.

Lama, F. D., Roca, M. D., Andrade, M. A., and Nascimento, L. B. (2011). The use of photography to identify individual tree frogs by their natural marks. *South American Journal of Herpetology*, 6(3):198–204.

Pearson, K. (1901). On lines and planes of closest fit to systems of points in space. *Philosophical Magazine*, 2(11):559–572.

Pilliod, D., Velasquez, E., Bosworth, K., Ahsan, H., and Kelly, O. Identifrog: An automated pattern recognition program for leopard frogs. http://code.google.com/p/identifrog/.

Velásquez, M. E. (2006). *Wavelets: Theory and Applications*. PhD thesis, Idaho State University, Pocatello, Idaho, USA.