

3-17-2014

A Reference System for Animal Biometrics: Application to the Northern Leopard Frog

D. Petrovska-Delacrétaz
Institut Mines-Télécom

A. Edwards
Boise State University

J. Chiasson
Boise State University

G. Chollet
Boise State University

D. Pilliod
Boise State University

A REFERENCE SYSTEM FOR ANIMAL BIOMETRICS: APPLICATION TO THE NORTHERN LEOPARD FROG

D. Petrovska-Delacrétaz¹, A. Edwards and J. Chiasson², G. Chollet^{2,3}, D. S. Pilliod^{4,5}

¹ Institut Mines-Télécom; Télécom SudParis; CNRS UMR5157 SAMOVAR

²ECE Department, Boise State University, Boise ID USA 83725

³Institut Mines-Télécom; Télécom ParisTech; CNRS LTCI

⁴Research Ecologist with the U.S. Geological Survey, Forest and
Rangeland Ecosystem Science Center, Boise Idaho 83706

⁵Department of Biological Sciences, Boise State University, Boise ID USA, 83725.

dijana.petrovska@telecom-sudparis.eu, aaronedwards2@u.boisestate.edu,
johnchiasson@boisestate.edu, chollet@telecom-paristech.fr, dpilliod@usgs.gov

ABSTRACT

Reference systems and public databases are available for human biometrics, but to our knowledge nothing is available for animal biometrics. This is surprising because animals are not required to give their agreement to be in a database. This paper proposes a reference system and database for the northern leopard frog (*Lithobates pipiens*). Both are available for reproducible experiments. Results of both open set and closed set experiments are given.

Index Terms— Automatic Identification, Frogs, Principal Component Analysis, Public Database, Reference Systems

1. INTRODUCTION

To evaluate the status and trends of endangered species biologists conduct demographic studies of animal populations in the wild. To do so requires capturing the animals, identifying them and coming back later to recapture and reidentify them in order to determine how the species is doing in a particular habitat by estimating demographic parameters such as survival, recruitment, and population growth rates. This identification of the individual animals has been done in the past by using a permanent or temporary mark or tag. This identification method, while often reliable, may pose health risks

The authors gratefully acknowledge the Department of Electrical and Computer Engineering at Boise State University for providing funding for this project. Dr. Petrovska-Delacrétaz worked on this project during her sabbatical stay at Boise State University and is grateful to the Department of Electrical and Computer Engineering at Boise State University for their hospitality.

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

to animals and thus there is a need for non-harmful alternatives. One of the most intriguing alternatives for animal identification is photography. Here we consider the photographic identification of frogs.

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 [1] 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 [2], [3]. An example is shown in Figure 1 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 natural 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 [3].

Ideally, one wants an automatic procedure to identify the frog, i.e., determine whether or not it is in the current database. By fully automatic procedure we mean the identification is done with the full photograph of the frog (Figure 1). In this paper 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 [3], the cutouts were then 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.

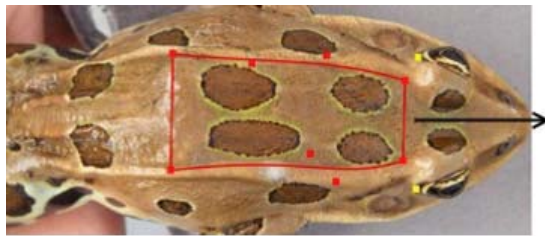


Fig. 1. Cutout along natural contours of the frog [3].

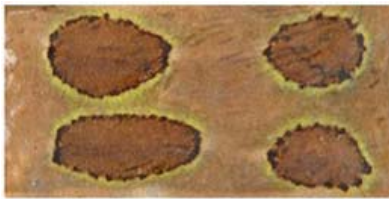


Fig. 2. The cutout is stretched to form a rectangle [3].

Work similar to that presented here was done by Gamble et al. [4] 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 [5] 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.

In this paper we consider an identification procedure based on principal component analysis (PCA) of the cutout frog photographs. We test the proposed identification procedure by dividing the database into mutually exclusive development and evaluation sets as discussed in [6]. Determining the (hyper) parameters of the identification algorithm using the development set and then using these parameters to test the identification procedure on the evaluation set, the method was found to provide a simple, fast, and accurate semi-automated pattern-recognition algorithm. That is, it gave an accurate 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 presents the experimental results using the development database while Section 5 presents the experimental results using the evaluation database. Finally Section 6 gives our conclusions.

2. ANIMAL DATABASE

The database consists 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 [3]. Kelly obtained 209 frogs bred in captivity and photographed them. A photographic light diffusing dome (Cloud Dome, www.cloudome.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.

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. With the combination of the No Dome and Shade Dome images, the Captive Leopard frog database contained 1386 total images.

2.1. Development and Evaluation Databases

To have both a development and an evaluation database, the frog identities were split into two (essentially) equal sets. We put all the images of frog identities 1 – 50 and 156 – 209 into the development database resulting in 105 identities and a total of 697 images. The frog identities 51 – 155 were put into the evaluation database resulting in 104 identities and a total of 689 images. As only the frog identities 109 – 209 had no dome images as well as shade dome images, this distribu-

tion of identities between the development and the evaluation databases ensures that an equal number of frog identities with (the additional) no dome images are present in both databases.

3. RECOGNITION PROCEDURE

The image capture follows the procedure discussed in the Introduction. We followed the “fingerprint” extraction procedure as described in [3]. The open-source program IDENTIFROG (<http://code.google.com/p/identifrog> [7]) 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.

3.1. Feature Extraction and Identification

We used Principal Component Analysis (PCA) which was developed over 100 years ago for statistical analysis. It is also a well-known method in Machine Learning [8], 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) $N_t \ll d \triangleq 256 \times 128$. As C is a positive semi-definite symmetric matrix, there is an orthogonal matrix $Q \in \mathbb{R}^{d \times d}$ such that

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

That is, the i^{th} column of Q is the i^{th} eigenvector of C with eigenvalue λ_i . Further, $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_{N_t-1} \geq 0$. We can then represent any the 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

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

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 $N < N_t \ll d$. That is, the image is coded into \mathbb{R}^N by

$$h_c \triangleq [h_1 \ h_2 \ h_3 \ \dots \ h_N]^T \in \mathbb{R}^N \quad (6)$$

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 [8] tells us that reconstruction error is given by

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

N is chosen so that $\lambda_{N+1}^2 + \dots + \lambda_{N_t-1}^2$ is small. Thus, as far as the Euclidean norm is 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

$$f = [h_A \ \dots \ h_N]^T \in \mathbb{R}^{N-A+1} \quad (8)$$

which indicates that we are representing the image by eigenvectors A through N . Our choice will turn out to be $A = 3$, $N = 125$, 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 $x^{(k)}$, $k = 1, \dots, N_t$ 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 $k = 1, \dots, N_t$ compute the (cosine of the) minimum angle between the new image and the existing images, that is, compute

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

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

$$k^* \triangleq \arg \max_{k \in \{1, \dots, N_t\}} \{s^{(k)}\} \quad (10)$$

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 them matches the test image.

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

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

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.

The (hyper) parameters of the identification algorithm is the selection of the feature vector as given in (8), that is, the determination of A and N in (8) and the selection of threshold value for γ . This will be done using the development database. Once these parameters are chosen, they will be fixed when testing is done using the evaluation database.

4. EXPERIMENTAL RESULTS - DEV DATABASE

An open set test protocol does not assume that a test frog is in the database. We first performed an open set experiment on the development database which consists of 105 frog identities and a total of 697 images. We further subdivided these 105 frog identities into “known” and “unknown” frog identities. The “known” database was made up of 76 of the frog identities with 605 total images while the “unknown” database was made up of the remaining 29 frog identities with 92 total images.

The 697 “known” 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 these 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 procedure resulted in bins 1 – 5 having 121 images each. The 92 images of the “unknown” frogs were then put into a 6th bin.

After putting all the “known” frog images of the development database 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 $N_t = 4 \times 121 = 484$ images.) The images in the 5th bin were used for testing “known” frogs against the “known” frogs in bins 1 – 4 while bin 6 was used for testing “unknown” frogs against the “known” frogs in bins 1 – 4. We take the identified image to be $x^{(k^*)}$ where k^* is as given in equation (10). We then permuted bins 1 – 5 so that each time a different set of 4 bins was used to compute the covariance matrix and the remaining bin was used as the “known” frog test bin.

4.1. Determination of the Parameter A and N

In using the PCA test we used the feature vector given in (8) with $A = 3, N = 125$. To make this determination we repeated the “known” frog test on all 105 frog identities of the development database using eigenvectors A to N where A

was varied from 1 to 6 and, for each value of A , N was varied from 50 to 200. The results are shown in the graph of Figure 3. The graph shows $A = 3$ gives a 98% accuracy over the range $N = 100 - 200$ as well as being as least as good as the other values of A over the range $N = 125 - 175$.

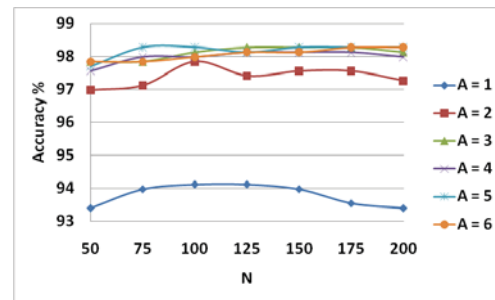


Fig. 3. PCA accuracy as A, N vary.

4.2. Testing the “known” frogs

As shown in Table 1 the 5th bin had 121 images and 1 of the images was incorrectly identified for an accuracy of $120/121 = 99.17\%$. We then repeated the procedure four more times using a different bin as the test bin, that is, 5-fold test [9]. The results are in in Table 1. This shows a total of 7

Table 1. 5-fold test protocol of “known” frogs.

test bin	bin 5	bin 4	bin 3
accuracy	$\frac{120}{121} = 99.17\%$	$\frac{119}{121} = 98.35\%$	$\frac{117}{121} = 96.7\%$
test bin	bin 2	bin 1	
accuracy	$\frac{121}{121} = 100\%$	$\frac{121}{121} = 100\%$	

errors over the five folding (permuting bins 1 through 5) tests on the 605 images of the “known” frogs. A more relevant (to the biologist) list of errors is given in Table 2. The first row of Table 2 ($n = 1$) shows that over the 5-folds there were 7 frogs in the test bin that were misidentified with a frog identity from the other 4 bins. The second row of Table 2 ($n = 5$) shows the results of finding the 5 frog identities among the 4 bins whose scored the highest with the test frog image. Specifically it shows that over the 5-folds there were 4 frogs whose identities were not among the top 5 retrieved identities (top 5 identity scores). Finally, the third row ($n = 10$) shows that over the 5-folds there were only 3 frogs which did not have their identities found among the top 10 retrieved identities (top 10 identity scores).

The biologist requires that the detection probability be very high while the false positive probability is of much less concern. That is, given a test image, the biologist wants to know with high probability if the identity is already in the database (and thus not a new frog identity).

Table 2. Incorrect identification errors in top n scores.

Top n	Errors
1	7/605 = 1.16%
5	4/605 = 0.66%
10	3/605 = 0.5%

4.3. Determination of the Threshold γ

We chose the threshold $\gamma = 0.36$ based on the data in the development database as is now explained. In Figure 4 the top 10 “unknown” pdf was calculated as follows: An “unknown” frog was tested against the “known” frogs and the top 10 identity scores $s_1 > s_2 > \dots > s_{10}$ were found. The value of s_{10} was saved as we are interested if this value is greater than or less than a given threshold γ . That is, the biologist wants to know if it is only necessary to look at the images of the top 10 identity scores with the “known” frogs to determine if the test frog is a new identity or not. A histogram of these scores was then normalized to become the “unknown” frog pdf.

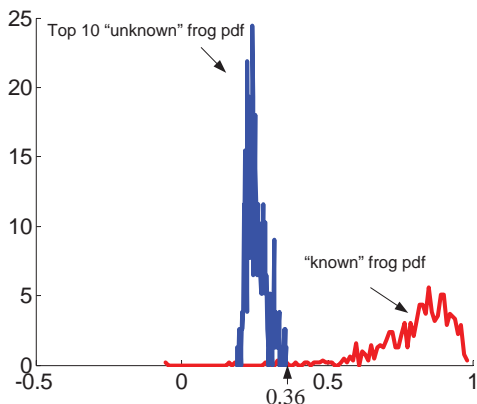


Fig. 4. pdf for the “known” and “unknown” frogs.

In contrast, the “known” frog pdf of Figure 4 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 above, the 605 “known” frog images were put into 5 equal sized bins. The PCA space (covariance matrix) was built from 4 of the bins. Then each identity in the 5th 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 (5-fold) each time using a different bin as the test bin. A histogram made of these scores was normalized to become the known frog pdf. Figure 4 shows the most of the “probability” of the known frog pdf is above the threshold. That is, with high probability, the biologist will only have to visually inspect 10 images to check for a match.

4.4. Testing the “unknown” frogs

We also tested the 92 images of the “unknown” frogs in the development database. We chose $\gamma = 0.36$ as the threshold value. We are interested in the n identities in the “known” database whose scores with an “unknown” test frog are highest, i.e., the top n identities scores. We did this for $n = 1, 2, \dots, 10$ and averaged over the 5-folds of bins 1 – 5. (For each fold the PCA space is changed and the “unknown” frog is tested using this PCA space against the “known” frogs in the 4 bins used to construct the PCA space). The average number (over the 5-folds) of frogs whose scores are greater than γ is shown in Table 3.

Table 3. Avg. no. of “unknown” frogs above threshold.

Top n	Avg. No. over 5-folds
1	91.8/92 = 99.8%
2	39.8/92 = 43.3%
3	19.6/92 = 21.3%
4	8.6/92 = 9.4%
5	4.2/92 = 5.6%
6	1.8/92 = 2%
7	0.8/92 = 0.9%
8	0.4/92 = 0.4%
9	0/92 = 0.0%
10	0/92 = 0.0%

To explain Table 3, the first row ($n = 1$) means that on average 91.8 of the 92 images in bin 6 (“unknown” frogs) had a score of 0.36 with at least one frog identity in the “known” database. Essentially, every “unknown” frog was taken to be in the database. The second row ($n = 2$) tell us that on average 39.8 of the 92 “unknown” images had a score of 0.36 with at least two *different* frog identities in the “known” database. Continuing, the ninth row tells that on average *none* of the 92 “unknown” images has a score of 0.36 with 9 different frog identities in the “known” database. From a practical point of view this means that the biologist would have the recognition system bring up only the 10 top identity scores and visually inspect whether or not they match the test frog image. If there is no visual match, then the biologist is reasonably assured that the test image is a new identity.

5. EXPERIMENTAL RESULTS - EVAL DATABASE

We follow the procedure given in Section 4, but now using the evaluation database. In the evaluation database we have 104 frog identities with a total of 689 frog images. We further subdivided these 104 frog identities into “known” and “unknown” frog identities. The “known” database was made up of 75 of the frog identities with 595 total images while the “unknown” database was made up of the remaining 29 frog identities with 94 total images.

The 595 “known” were distributed into 5 bins in the same manner as explained in Section 4 resulting in 119 images in

each bin. The remaining 94 images of the “unknown” frogs were then put into a 6th bin. The 5-fold testing etc. follows the same procedure as for the development database.

5.1. Testing the “known” frogs

Analogous to Subsection 4.2 (see Table 2), Table 4 shows the number of misidentifications within the top n identity scores over the 5-folds. The first row of Table 4 shows that over the 5-folds there were 8 frogs in the “known” frog test bin that were misidentified with a frog identity from the other 4 bins. The second row of Table 4 ($n = 5$) shows the results of finding the 5 frog identities among the 4 bins whose scored the highest with the test frog image. Specifically it shows that over the 5-folds there were 4 frogs whose identities were not among the top 5 retrieved identities (top 5 identity scores). Finally, the third row ($n = 10$) shows that over the 5-folds there were only 2 frogs which did not have their identities found among the top 10 retrieved identities.

Table 4. Incorrect identification errors in top n scores.

Top n	Errors
1	8/595 = 1.18%
5	4/595 = 0.67%
10	2/595 = 0.34%

5.2. Testing the “unknown” frogs

We tested the 94 images of the “unknown” frogs in the evaluation database. As explained in Subsection 4.4 we are actually interested in the n identities in the “known” database whose scores with the “unknown” test frog are highest, i.e., the top n identities scores (recall Table 3). These results are shown in Table 5 for $n = 1, \dots, 10$.

Table 5. Avg. no. of “unknown” frogs above threshold.

Top n	Avg. No. of Frogs
1	93.6/94 = 99.6%
2	39.6/94 = 43%
3	20.6/94 = 22%
4	11.6/94 = 12.3%
5	6.6/94 = 7%
6	5.8/94 = 6%
7	3.6/94 = 3.8%
8	1.8/94 = 1.9%
9	1.4/94 = 1.5%
10	0.4/94 = 0.4%

6. CONCLUSIONS AND PERSPECTIVES

This work was originally motivated by the previous work of Velásquez [2] and Kelly [3]. Tables 2 and 4 show that our

recognition procedure identifies a “known” frog with the correct identity in the top 10 retrievals with a greater than 99% accuracy. Further Tables 4 and Table 5 show that to test if an “unknown” frog is in the existing database, 99% of the time one need only look at the top 10 identity scores for the visual test. We are now exploring the recent work reported in [10] [11] (and references therein) of new techniques that allows one to do fully automatic identification.

Please contact the authors to obtain the reference system, as well as the development and evaluation databases.

7. REFERENCES

- [1] Fernanda Del Lama, Michelle Drummond Roca, Miguel Angelo Andrade, and Luciana Barreto Nascimento, “The use of photography to identify individual tree frogs by their natural marks”, *South American Journal of Herpetology*, vol. 6, no. 3, pp. 198–204, 2011.
- [2] Maria Elena Velásquez, *Wavelets: Theory and Applications*, PhD thesis, Idaho State University, Pocatello, Idaho, USA, 2006.
- [3] Oksana Vladimirovna Kelly, *Automated Digital Individual Identification System with an Application to the Northern Leopard Frog Lithobates pipiens*, PhD thesis, Idaho State University, 2010.
- [4] Lloyd Gamble, Sai Ravela, and Kevin McGarigal, “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*, vol. 45, pp. 170–1180, 2008.
- [5] M.A.H.B. Azhar, S. Hoque, and F. Deravi, “Automatic identification of wildlife using local binary patterns”, in *IET Conference on Image Processing (IPR 2012)*, July 2012, London UK.
- [6] Dijana Petrovska-Delacrétaz, Gérard Chollet, and Bernadette Dorizzi, *Guide to Biometric Reference Systems and Performance Evaluation*, Springer-Verlag, 2009.
- [7] David Pilliod, Elena Velasquez, Ken Bosworth, Hidayatullah Ahsan, and Oksana Kelly, “Identifrog: An automated pattern recognition program for leopard frogs”, <http://code.google.com/p/identifrog/>.
- [8] David Barber, *Bayesian Reasoning and Machine Learning*, Cambridge, 2012.
- [9] Yoshua Bengio and Yves Grandvalet, “No unbiased estimator of the variance of k-fold cross-validation”, *Journal of Machine Learning Research*, vol. 5, pp. 1089–1105, September 2004.
- [10] Douglas T. Bolger, Bennet Vance, Thomas A. Morrison, and Hany Farid, “Wild-ID user guide: Pattern extraction and matching software for computer-assisted photographic mark-recapture analysis”, January 2011.
- [11] Hjalmar S. Kühl and Tilo Burghardt, “Animal biometrics: Quantifying and detecting phenotypic appearance”, *Trends in Ecology and Evolution*, vol. 28, no. 7, pp. 432–441, July 2013.