

## ZOOMETRICS – Biometric Identification of Wildlife Using Natural Body Marks

S. Hoque, M.A.H.B. Azhar, and F. Deravi

*School of Engineering and Digital Arts, University of Kent, Canterbury, UK.  
{ s.hoque; m.h.b.azhar; f.deravi }@kent.ac.uk*

### **Abstract**

*Using physiological or behavioral characteristics to identify humans has been in use for quite some time now. Many wildlife animals also show distinctive natural body marks that can be used to identify them individually. Scientists in conservation research often use this approach but the process is manual and can be slow and error prone. This paper reports on an investigation to use biometric techniques for the identification of an important endangered species – The Great Crested Newt. The paper reports on novel techniques for extraction of the belly patterns of these animals as a source of biometric information. Features and classification techniques used for their automatic recognition are presented. The proposed approach is tested on a database of newts under investigation by conservationists. Preliminary studies are also reported on the ageing effects when belly images are compared over a number of years. The results suggest that such biometric techniques may be suitable for developing effective and flexible identification of wildlife in the field.*

**Keywords:** *zoometric, wildlife biometrics, pattern recognition, ageing, Great Crested Newt.*

### **1. Introduction**

Effective approaches for the conservation of wildlife require a sound knowledge of population demography. Such information is obtained through mark-recapture studies that enable following the lives of each individual animals. Marking techniques in practice include amputation, tagging, matching natural body marks, etc [1, 2]. All these methods have their pros and cons but the identification by body marks can be seen as least invasive. Although widely practiced, the method is mostly manual where newly captured images are manually compared with those in the library already. Ability to do so automatically can improve the speed and accuracy of their recognition as well as facilitate recognition in the field through the use of mobile devices.

Photographic identification method, albeit often manual, has been used successfully in many studies of animal behavior and population diversity. Successful applications have been seen on whales [3], polar bears [4], tigers [5], elephants [6], rattlesnakes [7], lizards, salamanders, frogs [8], and many more. Some researches towards automating the process by using computer aided techniques have also been reported. For example, Voros *et al* [9] reported a computer aided technique using the size and shape attributes of patches in the belly patterns in quantifying and identifying two European *Bombina* species and their hybrids. Van Tienhoven *et al* [10] used two-dimensional affine transform to compare the natural pigment marks on two individual ragged tooth sharks. Kelly [11] suggested a 3-D matching scheme to aid in matching Serengeti cheetahs. This latter study involved nearly 10000 photographs

taken over 25 years and showed that the approach can even be used for animals without visibly distinctive features.

This paper investigates a scheme to identify individual newts from their distinctive belly patterns. Great Crested Newt is very common in the UK mainland but their numbers are declining rapidly due to intensive farming practices, loss of habitat caused by urbanisation, etc. Conservation studies, therefore, aim to identify and monitor life-history characteristics of the newt populations, maintain existing habitats, increase populations through re-colonisation, and so on [12]. The process requires capture-recapture of the newts over a long period of time. The unique belly markings of great crested newts (see Figure 1) are generally used to identify individuals, especially when collecting long term population data. For this purpose newt belly images are acquired using limited restraint, and the photographs are hand-sorted for identification. Success of this process lies in the surveyor's ability to match recaptures which can become error prone and very time consuming especially when the population size is large. Variation due to factors such as animal positioning and quality of photograph further adds to the complexity.

In this preliminary study, we investigated the feasibility of automated biometric identification of newts by matching their unique belly patterns using computer vision techniques. In addition to the success of identification, we also investigated the effect of ageing of the newts on the matching performance.



**Figure 1. Distinctive belly pattern of Great Crested Newt**

## **2. Newt images database**

The newt images used for this pilot study were obtained from Durrel Institute of Conservation & Ecology (DICE) [13]. The database contains 30 images from 10 different newts. The images are of varying resolution, in 24-bit colour. The year of capture is also recorded. The ground truths of the newt identities were generated by human operators. Figure 2 illustrates some of the images from the database. Images in a column are from the same newt, but captured at different times. Despite the variations in imaging conditions and newt's posture, the intra-newt similarity is obvious.

## **3. Segmentation of the region of interest (RoI)**

The belly of the newt displays a distinctive pattern of small black patches. For a successful match, this area needs to be isolated from the background. However, in order to minimise distress while capturing the live newt images, only a limited restraint was applied. Thus the newt could stretch/twist in any direction, thus deforming the belly pattern between subsequent

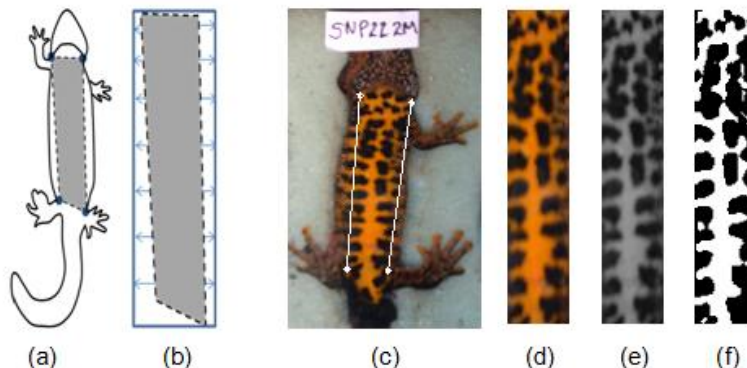
captures. The semi-automated segmentation of the RoI proposed here involves minimal human intervention.



**Figure 2. Some sample images from the database.**

### 3.1. The 4-point segmentation scheme

This scheme relies on human user locating 4 key points in the newt image – the points where the upper and lower limbs joins the body, as shown in Figure 3(a). The resulting quadrilateral is then stretched to fit a pre-defined rectangular area (200x40 pixels for this study), as shown in Figure 3(b). This produces a common sized belly pattern from all the newts, thus making comparison simple and efficient. Figures 3(c) and 3(d) shows the RoI and reshaped belly pattern of a real newt image. Figure 3(e) and 3(f) shows the gray-scale and binary versions of 3(d) which were used in the two matching schemes implemented in this study. The selection of the 4 key points marked by the human operator may vary to some extent without adversely affecting the matching scheme.



**Figure 3. The 4-point segmentation scheme and resulting outputs; (a) limb joints used as the anchor points, (b) RoI is stretched to form a rectangle of fixed size, (c) & (d) when used on a newt image, (e) & (f) Gray-scale and binary equivalent of (d)**

It is obvious from Figure 3(c) that a significant portion of the belly pattern may remain outside the selected RoI and this may have a detrimental impact on the matching. To overcome this, a 6-point segmentation scheme has been developed.

### 3.2. The 6-point segmentation scheme

In this scheme, the human operator needs to mark two points (midway between the front and rear limbs, as shown in Figure 4(a)) in addition to four limb joints. A cubic spline curve is fitted to these points to identify the RoI. This area is then linearly stretched to fit a pre-defined rectangle. By comparing Figures 3 and 4, it is obvious that the latter approach extracts most of the belly region visible in the newt image. Like the previous scheme, the extracted image converted to grayscale and binary for matching purposes. No other noise removal or image enhancement techniques are applied in this implementation.

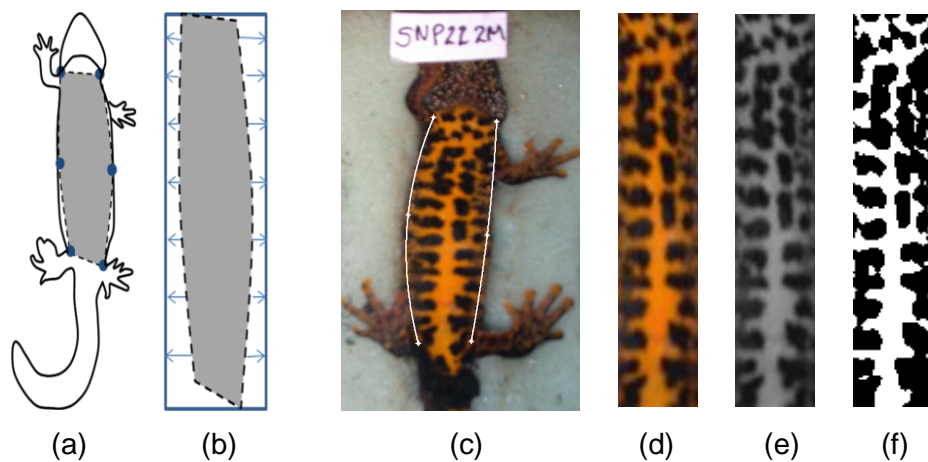


Figure 4. The 6-pt segmentation scheme and resulting outputs

## 4. Comparing the Regions of Interest

In order to find the similarity between two newts, the corresponding RoIs are compared. To compensate the misalignment due to flexible segmentation mechanism, a subsection of the RoI is isolated from one image that scans over the second image. A score is generated for each overlapping region and the overall comparison score is the one that shows maximum similarity. Two different methods were used for this comparison; first is the correlation coefficient (CC) of the gray scale pixel intensities and the second is the Hamming distance (HD) between the binary image segments.

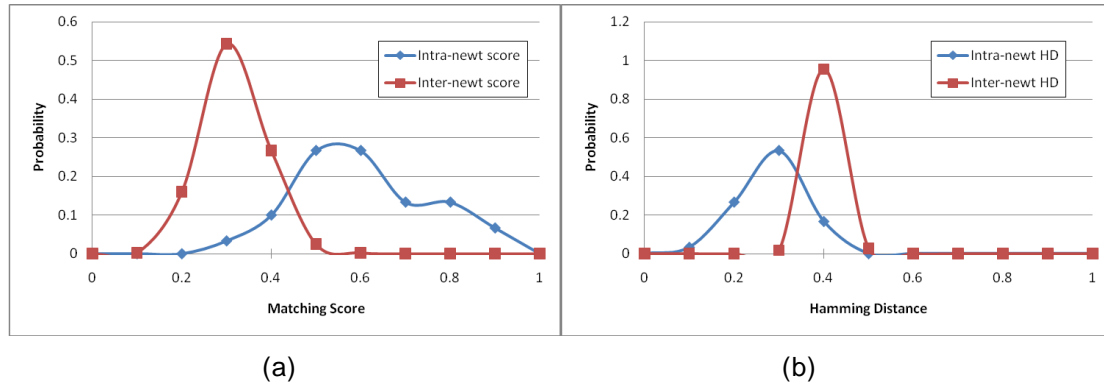
Size of the chosen RoI subsection (or sub-window) has an impact on the comparison scores generated. Different window sizes were, therefore, investigated and the results presented here refer to a window size which is about 50% of the originally segmented RoI.

## 5. Experimental Results

A number of experiments/analyses were carried out to ascertain the feasibility of an automated system for photographic identification of newts.

### 5.1. Distribution of comparison scores

Figure 5 illustrates the distribution of the comparison scores (CC or HD) between images of the same newts (intra-newt score) as well as between those of different newts (inter-newt score). It is obvious that the intra-newt and inter-newt score clusters are distinguishable and hence establishes the viability of image based identification of newts. But in both the plots, there is a significant overlap. For the current selection of features and comparison schemes, some errors are inevitable. The optimum threshold will be the point where the two curves intersect each other and the overall error will then be minimum.



**Figure 5. Score distributions for (a) Matching score and (b) Hamming distance**

### 5.2. Comparison in a static population

The experiment below looks into the accuracy of identification in a newt population where every newt had been captured before and thus has an image in the database. All the test images (i.e., recaptured newts) have been compared with all those already in the database, and the label of the newt generating the best similarity score is assigned to the test pattern. For this study, of the newt images available, 1 (or 2) images per newt were picked at random to be included in the database. The remaining images were tested against this database. Any image given a wrong label is an error and the error rate is the proportion of test images mislabeled by the system (see eqn. 1).

$$\text{Errorrate} = \frac{\text{Total no.of incorrect labels}}{\text{Total no.of testimages}} \quad (1)$$

Table 1 shows the error rates experienced under different configurations. The figures shown are the averages of 100 test runs. It is clear that the Hamming distance approach produced less error and having several images in the database per newt significantly reduced (nearly halved) the error rates.

**Table 1: Error rates in identifying newts from a known population**

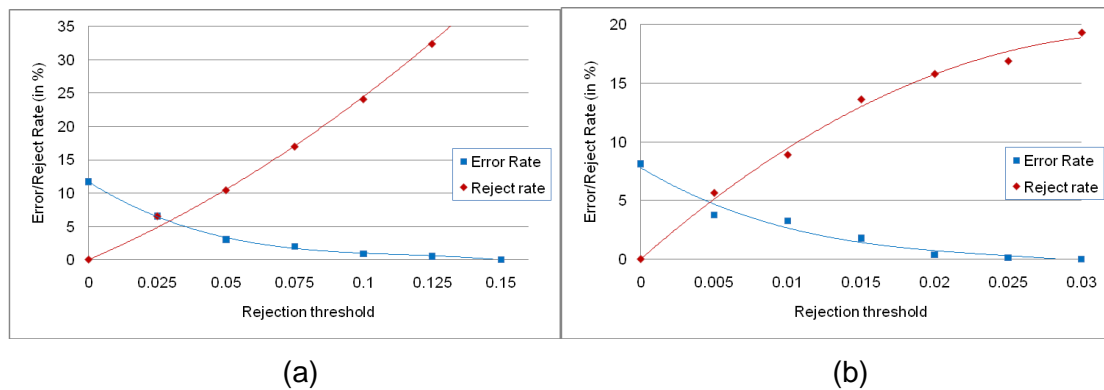
Comparison Method	Images per newt in the database	
	1	2
Correlation Coefficient	11.7%	6.6%
Hamming Distance	8.1%	4.4%

When two newts in the database appear similar to a test newt, they produce very similar comparison scores. When the two (or more) best scores are similar, the confidence in the labeling becomes low and such cases should be dealt by a more sophisticated algorithm or by a human operator to prevent error from propagating to the subsequent stages. A ‘rejection threshold’ sets the minimum margin needed for a label to be deemed reliable. Otherwise, the system marks these images as ‘Rejections’. In the estimation of error rates, the images ‘rejected’ are not included (see eqn. (2)). In addition, the proportion of the images rejected can also be estimated (see eqn (3)).

$$\text{Error rate} = \frac{\text{Total no. of incorrect labels}}{\text{Total no. of test images} - \text{Total no. of rejections}} \quad (2)$$

$$\text{Reject rate} = \frac{\text{Total no. of rejections}}{\text{Total no. of test images}} \quad (3)$$

Figure 6 illustrates the effect of such “rejection” on performance. It is clear that by selecting an appropriate rejection threshold, error rate can be reduced significantly at the expense of a higher reject rate.



**Figure 6. Error rates and Reject rates for various thresholds for (a) the CC scheme and (b) for the HD scheme. There is one image per newt in the album**

### 5.3. Comparison in a dynamic population

In almost all real scenarios, the population is dynamic. Previously unknown newts may be captured (denoted ‘new captures’) as new generations are bred, or migrate from nearby habitats, or simply may not have been captured before for whatever reason. Therefore, the database will have a finite population already registered. New captures will be noticed along with recaptures of previously seen newts. The following test simulates this scenario.

It is, therefore, necessary to first establish whether the newt under investigation is a new capture. If not, it is a recaptured newt which will have to be assigned to one of the labels already in the database. For simplicity of implementation, new captures were not added to the database. Any failure to identify a new capture gives a ‘false positive’ output whereas identifying a re-captured newt as new capture generates a ‘false negative’ outcome. All recaptures are given a label based on their similarity to those in the database. Mislabeling may occur at this stage too. All these contribute towards the overall error of the system. The following error rates are calculated for this scenario.

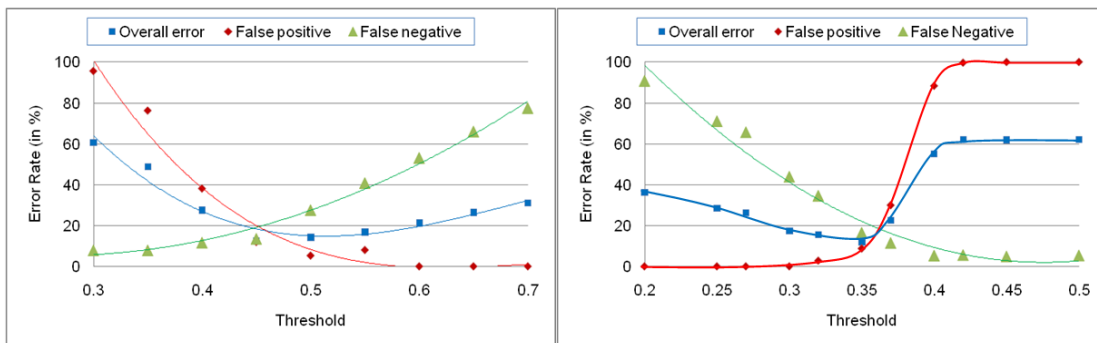
$$\text{Overall Errorrate} = \frac{\text{Total no.of incorrect labels}}{\text{Total no.of test images}} \quad (4)$$

$$\text{False Positiverate} = \frac{\text{Total no.of new captures labelled as recapture}}{\text{Total no.of new captures}} \quad (5)$$

$$\text{False Negativerate} = \frac{\text{Total no.of recaptures labelled as new capture}}{\text{Total no.of recaptures}} \quad (6)$$

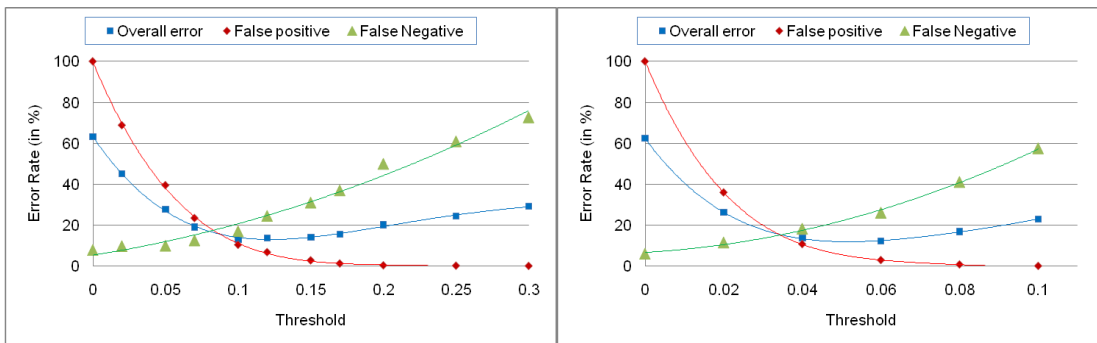
Three different schemes were implemented to identify a new capture. For experiments in this section, we randomly picked 5 different newt images and added them to the database album. The remaining 25 were used as test data. For each of these test images, we first tested whether they are re-captures or not, and then all recaptures were matched with those in the album for identification.

**5.3.1. Scheme A:** The matching score is compared against a threshold and if it is beyond the range, the newt in question is treated as a recapture. Figure 7 illustrates the error rates observed for different thresholds. The plots identify the thresholds that produce the lowest overall error.



**Figure 7. Variation in various error rates at different thresholds (Scheme A)**

**5.3.2. Scheme B:** The confidence in decision is used as the criterion for distinction between new and re-captures. The difference between the top two match scores is the confidence measure and if this is below a pre-defined threshold, the test image is from a new capture. Figure 8 shows the error rates observed for different thresholds.



**Figure 8. Variation in various error rates at different thresholds (Scheme B)**

**5.3.3. Scheme C:** This is a combination of Schemes A & B. If either of the criteria is fulfilled, a newt is marked as a new capture. The results presented in Table 2 correspond to the particular threshold pairs shown in the table and the error rates are the averages from 10 runs. When compared with the corresponding figures in Figures 6 and 7, it is obvious that the False Positive Rate and the overall Error Rate has gone down significantly, although no significant improvements were noticed in the False Negative Rates.

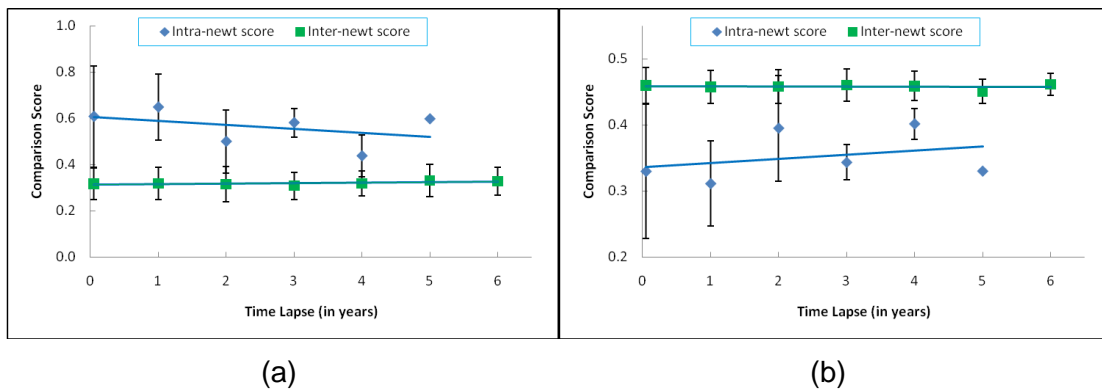
**Table 2. Error rates in identifying newts in a dynamic population (Scheme C)**

Comparison Method	Incorrect Decisions		
	False Positive Rate	False Negative Rate	Overall Error Rate
Correlation Coefficient (Thresholds=0.4, 0.1)	5.3%	19%	10.8%
Hamming Distance (Thresholds=0.35, 0.03)	4.7%	18%	10%

**5.4. Ageing profile**

With ageing, the belly pattern of the newt changes to some extent. For example, spots grow in size, often nearby marks join. This, therefore, makes the belly patterns look dissimilar even when these are from the same newt. Figure 9 illustrates how the matching scores gradually change when the time lapse between the two captures increases. It should be noted here that only the ‘year of capture’ data was available. As such, newt captured in the same calendar year has been shown as having 0 (zero) time-lapse. The variation of inter-newt scores over these years was not significant.

This study suggests that a variable threshold may be used in the classification process to accommodate this phenomenon.



**Figure 9. Score changes with lapse of time between captures (a) for CC and (b) for HD**



## 6. Conclusion

Despite the simplicity of the algorithms, the study generated very interesting results which definitely justify the viability of the automated technique. It is obvious that the error rates, at this stage, are relatively high but this is partly due to the very small image database being available. Some of the newts were photographed while the belly was not properly aligned and their inclusion in this small dataset made the task difficult. A repeat investigation with a much larger dataset can be expected to produce significantly lower error rates. From the algorithmic point of view, more distinctive feature vectors and sophisticated classification scheme should improve the error rates.

A much larger image database of newt belly patterns (as well as several other wildlife creatures) is currently being accumulated. These databases will be publicly available to facilitate research community interested in zoometrics.

## Acknowledgment

The authors gratefully acknowledge the support of Prof Richard Griffiths and his team at DICE for providing the newt images for this study.

## References

- [1] M.A.Donnelly *et al.* Techniques for marking amphibians. In W.R. Heyer *et al.* (eds.), *Measuring and monitoring biological diversity: standard methods for amphibians*. pp. 277-284. Smithsonian Institute Press, Washington DC. 1994.
- [2] J.W. Ferner, A Review of Marking Techniques for Amphibians and Reptiles. Herpetological Circular 9. Society for the Study of Amphibians and Reptiles Publications, USA. 1979.
- [3] R. Payne *et al.* External features in southern right whales (*Eubalaena australis*) and their use in identifying individuals. In R. Payne (ed.). *Communication and behavior of Whales*. AAAS Selected Symposium 76. Westview Press, Boulder, Colorado. 1983.
- [4] C.J.R. Anderson. Individual identification of polar bears by whisker spot patterns. MSc Thesis. University of Central Florida, Orlando, Florida. 2007.
- [5] L. Hiby *et al.* A tiger cannot change its stripes: using a three-dimensional model to match images of living tigers and tiger skins. *Biology Letters*. Vol.5, pp.383-386. 2009.
- [6] A. Ardovini *et al.* Identifying elephant photos by multi-curve matching. *Pattern Recognition*. Vol.41, pp.1867 – 1877. 2008.
- [7] B.R. Moon *et al.* Identifying Individual Rattlesnakes Using Tail Pattern Variation. *Herpetological Review*. Vol.35(2), pp.154–156. 2004.
- [8] K.S. Bradfield. Photographic identification of individual Archey's frogs, *Leiopelma archeyi*, from natural markings. DOC Science Internal Series 191. Department of Conservation, Wellington. New Zealand. 2004.
- [9] J. Voros *et al.* A new method for quantitative pattern analysis applied to two European *Bombina* species. *Herpetological Journal*. Vol.17, pp.97-103. 2007.
- [10] A.M. Van Tienhoven *et al.* A computer-aided program for pattern-matching of natural marks on the spotted raggedtooth shark *Carcharias Taurus*. *Journal of Applied Ecology*. Vol 44, pp.273-280. 2007.
- [11] M.J. Kelly. Computer-aided photograph matching in studies using individual identification: an example from Serengeti cheetahs. *Journal of Mammalogy*. Vol.82(2), pp.440-449. 2001.
- [12] B. Lewis *et al.*, Field Assessment of Great Crested Newt *Triturus Cristatus* Mitigation Projects in England. Natural England Research Report NERR001. Natural England, UK. 2007.
- [13] <http://www.kent.ac.uk/dice/>

## Authors



**Dr Sanaul Hoque** obtained his first degree from Bangladesh University of Engineering & Technology (BUET) in Electrical & Electronic Engineering in 1991. He also obtained an MSc in Computer Engineering from the same institute. He was awarded a PhD in Electronic Engineering at University of Kent, UK where he is currently working as a faculty. Dr Hoque's research interests include image processing & computer vision, OCR, biometrics, biometric encryption, multi-expert fusion, document modeling, and so on.



Currently a Research Associate in Computer Vision and Image Processing at the University of Kent, UK **Dr M. A. Hannan Bin Azhar** obtained his first degree in Electrical and Electronic Engineering from Bangladesh University of Engineering and Technology in 1999, where he joined as academic staff between 2000 and 2001 in the Department of Electrical and Electronic Engineering. Dr. Azhar then received his MSc Research degree in the area of Robotics in 2002 from the University of Kent, UK where he also obtained his PhD in 2008 in Electronic Engineering. His research interests cover a broad area of intelligent and biologically/nature inspired learning and adaptive systems and include a wide range of machine learning and hybrid intelligent techniques encompassing learning and adaptation methods in pattern recognition and robotic modelling.



**Dr Farzin Deravi** obtained his first degree in Engineering and Economics from the University of Oxford in 1981 and his M.Sc. in Electronic Engineering from Imperial College, University of London in 1982. From 1983 to 1987 he worked as a research assistant at the University of Wales, Swansea where he obtained my Ph.D. In 1987 he joined the academic staff at Swansea where at the Department of Electrical and Electronic Engineering. In 1998 he joined the then Department of Electronics, now School of Engineering and Digital Arts - at the University of Kent where he is Reader in Information Engineering. He is a Member of the IEEE, the Institution of Engineering and Technology, and the British Machine Vision Association. He was the founding chair of the IET Professional Network on Visual Information Engineering and currently Editor-in-Chief of the IET Image Processing journal. He is currently the Vice President of the BioSecure Association and also serves on BSI and ISO committees on biometric standardisation.