Eco. Env. & Cons. 27 (May Suppl. Issue) : 2021; pp. (S20-S34)

ISSN 0971–765X

Non - Invasive techniques for identification of individuals within a species: A computational review

Glen Bennet Hermon[a*] and Durgansh Sharma[b]

[a]School of Computer Science, [b]Department of Cybernetics, School of Computer Science, University of Petroleum and Energy Studies, Dehradun 248007, Uttarakhand, India

(Received 11 September, 2020; Accepted 5 November, 2020)

ABSTRACT

As technological advances are made to sustain the wildlife in our environment there is a requirement of robust non-invasive techniques equipped to distinctly identify individual animals within a species. Identification of individuals from species and tracking of the identified individuals have been conducted with mostly invasive techniques in the past. However, recently multiple methods for non-invasive identification of individuals in the species have emerged, using computer vision algorithms. In addition, the time complexity and space complexity of the technological approaches are far better than the previously used manual approach for individual identification. There exist a wide range of differences among these techniques, based on the patterns in consideration and the approaches used. This amounts to a large collection of techniques that have value and scope in the subject of pattern recognition and thus demand for a comparative study of these techniques, with a discussion on their accuracy, ease of use and their adaptability in various scenarios. This work reviews the different pattern matching techniques among a plethora of algorithms used to identify animal individuals, within various animal species along with their complexity, categorizing them based on the type of pattern used for recognition. The emphasis is towards the insights of computational techniques for various image-based animal biometrics with the intention to automate these processes. It is observed that mathematically modelled identification techniques with algorithmically specialized normalization techniques are the most efficient identification techniques in a broad range of scenarios. This work provides a foundation, so that better algorithms and techniques can be modelled with respect to the existing ones for identification of individuals in a specific species.

Key words : *Animal Biometrics, Animal Identification, Automated Photo Identification, Computer-Vision, Non-Invasive Techniques.*

Introduction

Unique Individual Identification within a species creates a foundation for the study of species demography, behavioural patterns, lifespan and other ecological parameters. These parameters help in the creation of various statistical models based on multiple individuals of the same species. The task of individual identification within a species has been handled utilizing various methodologies for various species with shifting degrees of accomplishment in the last few decades. Each species has certain highlights that end up being one of a kind among the specimens of that species. Such highlights like spot designs, bristle spots, coat stripes, and so forth, are used to comprehend the uniqueness of those examples with the goal that productive procedures and calculations can be performed for non-obtrusive identifications.

For unique identification of individuals within a population, various invasive techniques such as tattooing, attaching radio tags, etc. have been adopted. However, these practices involve physical contact with the animal, causing disturbance in the regularity of the animal's life-routine. In addition, the invasive methods have the potential to cause injury, disrupt the food chain and sometimes lead to separation of individuals from the community (Sharp *et al.,* 2007). On the other hand, non-invasively, capturing photographs makes it possible to record identification patterns unique to the individuals within the species. Researchers in the early times have also used manual drawing and techniques of highlighting the unique features, assigning identification cards to each individual within the focal species. With the advancement of technology, the older techniques proved to be slow and tedious even though they tend to have a higher amount of accuracy. Thus, it created a necessity for adoption of technology and computer systems to capture, store and retrieve features and information of individuals using various identification strategies.

The best and most used image-based animal identification models are those based on spot patterns, fluke or fin Patterns, coat patterns, whisker spots, face recognition and body part deformation patterns. Individuals of tigers are identified by the unique stripe patterns over the skin. The skin texture information has been used for individual identification (Hiby *et al.,* 2009). For Polar Bears, stating some of the key findings and the relevance of the work, a short analysis of areas in respect to the above discussions shows us that the identification system developed uses their whisker spot patterns. Three reference points are manually needed to be selected, to extract the region of the whiskers from the photographs (Anderson, 2007). Secondly, as for the non-invasive methods of identifying elephants, the nicks and cuts on the elephant's ears pose to be unique. A context-free approach of the Generalized Hough Transform, with the ability of handling non-connected curves is used to extract these patterns from the edges of an elephant's ears (Ardovini *et al.,* 2008). Further, in Humpback Whales and Sperm Whales, the fluke patterns such as notches, holes, spots, lines etc. are observed and recorded manually onto computer systems, with an observer based coding scheme (Mizroch *et al.,* 1990; Whitehead, H., 1990). Next, African Penguins have black spots on their chests that remain same during their adult life. The prominent black stripe on the chest that lies between the white areas of the neck and the chest provides as an initial factor to pick out adult penguins (Burghardt *et al.,* 2004). Progressively, Pinnipeds are carnivorous aquatic mammals like seals or walruses that can be identified through photographs of the flippers - fore or hind that display damage or abnormalities, large scars on the body and size of the lower canine (McConkey, 1999). Similarly, the individual identification of the yellow bellied toad species is done by capturing photographs of the belly patterns and generating an identification code. The code comprised of eight numbers that the first digit that denotes sex or the life stage and the remaining digits denoting the pattern and the presence of spots on different regions of the belly (Pla¢ia¸su*et al.,* 2005). Chimpanzees being primates, have many similar structural features to humans, the most common face recognition algorithms used for humans tested on chimpanzees to produced good results. Multiple modifications are applied to the basic algorithms used for humans, in which similarity scores computed over global features are combined with similarity scores computed over local features to generate a more accurate score (Loos and Ernst 2013). For Chimpanzees further techniques like Deep convolution Neural Networks (CNN), using Gabor magnitude pictures (GMPs), achieving dimensionality reduction by locality preserving projections (LPP) with classification techniques such as sparse representation classification (SRC), are used. The Stochastic Gradient Technique is also used along with Backpropagation to compute gradients of intermediate layers. The Main focus is to achieve fine-grained recognition, and for such tasks of animal identification the matrix logarithm method proved to increase accuracy, while in the domain of second order statistics (Freytag *et al.,* 2016). Lemur identification is also done by face recognition after normalizing the images to extract a local binary Multiscale pattern by a patch-wise method. Various techniques are used to normalize for facial hair and ambient lighting with Linear Discriminant Analysis (LDA) being the dimensionality reduction technique used (Crouse *et al.,* 2017). The Great White Sharks are identified by their fins with a multi-scale approach that gives the ability to have multiple labels for each image. The process collects contour information rather than colour or texture information, this is then used to generate an ultrametric contour map used for the identification process (Hughes and Burghardt, 2017). For Australian sea lions whisker spots are selected manually and stored in a semi-automatic system, with pairwise comparisons of the Chamfer distance transform under a minimal variation in angle (Osterrieder *et al.,* 2015). For Bewick Swans, the distinct identification feature is the yellow and black patterns found on the upper surface and either side of each bird's bill. These patterns are recorded on the identification card of each bird, along with other information such as the location, sex, and other features like gape etc (Scott, 1978). For leopards, the whisker spot patterns are taken for the identification of individuals. The algorithm used is the Oblique Principal component cluster analysis technique and to minimize correlation, unique character sets are generated (Miththapala *et al.,* 1989). When identifying cheetahs, their coat spot pattern is unique to every individual. A model of a 3D plane is made in such a way that it fits on the photograph after a few points are identified such as the position of the shoulder. An Identifier array which is a sample of grey-scale intensities stored as a matrix of numbers, serves as an identifier for individuals. Individuals are uniquely identified by calculating the correlation coefficient and similarity coefficients between the models of known individuals and the patterns in the identifier arrays of the image input in consideration (Kelly, 2001). Grey seals display spot patterns on their pelage and major patterns of the head and neck are taken into consideration to identify individuals. The matching is done manually by at least two independent observers (Vincent *et al.,* 2001). Multiple Marine Vertebrates like the Delphinid (dolphin like) species have nicks and notches on their dorsal fin's trailing edge, that provides for the matching process involved in identifying individuals. The methods that take advantage of these patterns are the curve matching and string-matching techniques. Curve matching plots the Euclidean distances of the depth of each notch in sequence and saves the pattern. The string-matching technique involves the assignment of different string characters for respective crests and troughs in the notch pattern and thus saves this string as the pattern. For a multispecies approach, a live wire algorithm to locate the fin boundary along with various noise reduction and smoothing functions are applied before storing the curves or strings for the respective methods (Anderson, 2007; Hillman *et al.,* 2003). Badgers are small carnivorous animals very similar to rats and squirrels that have peculiar tail patterns that encapsulate the uniqueness of individuals (Dixon, 2003). Whale Sharks are a species of fish whose individuals are identified by the spot pattern on the body of the fish. After extracting the region of interest and subsequently the blobs, an algorithm used in astronomy to find the position of stars, the Groth's algorithm, is used to generate a similarity score (Arzoumanian *et al.,* 2005). The identification procedure used for Zebras is like the fingerprint identification techniques used for human fingerprints (Foster *et al.,* 2007). For Marbled Salamanders, Long-Tailed Salamanders and Spotted Salamanders, recording the dorsal patterns enabled individual identification. Multi scale PCA, the ANOVA method and region based encoding techniques are used to extract and store features (Gamble *et al.,* 2008; Jonas *et al.,* 2011; Chase *el al.,* 2015).

The shift in identification techniques of individuals from the manual mode to the technological mode has improved both the time as well as the space complexity involved with the purpose but has also shown variation in the level of accuracy, adaptability and ease of use. This creates a demand for a comparative study between the various techniques.

In this paper, we present a comparative study of various techniques used for individual identification within the species. A computational comparison of the different techniques provides a structure for categorizing these models based on complexity, similarity and repeatability of algorithms used. One of the primary goals of the paper is to analyse different techniques and emphasise on the various issues faced by the stakeholders while implementing these techniques, which may be both physical and technical.

The techniques, when compared with each other, distinctly show that the ones that rely on mathematical modelling are chosen more often, hinting to the accuracy that it provides and proving that such techniques based on mathematical modelling are the better techniques. Such a mathematical model will be able to quantify the uniquely identifiable aspects of appearances by reducing noise and ambiguity and by implementing various normalization techniques to quantify observable properties.

Discussion

Unique individual identification amongst the species have been performed either by marking them manually or by recognizing the uniqueness of the texture of their body, which may include patterns, colour, etc. as demonstrated in Fig. 1 This unique identity and feature must be sufficient to prove individuality over the whole size of the population of that particular species in consideration.

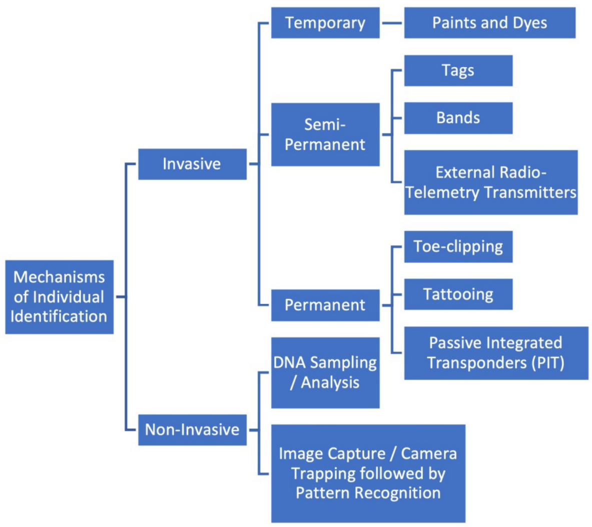

Fig. 1. Mechanisms of Individual Identification

In the past few decades, researchers have tried to identify several techniques to identify the unique patterns for individual tagging within the species with a considerable amount of accuracy. Initially, researchers used to manually draw and assign identification cards to each individual that stored values of various other features, which could not be drawn. With the advancement of technology, the traditional techniques appeared to be slow and complex even though they held a good amount of accuracy. This led to the adoption of upcoming technologies and computer systems, which captured stored and retrieved features and information of individuals from focal species using various identification strategies.

According to the analysis presented in (Pennycuick, 1978), larger the sample size, lower is the level of accuracy for identification of individuals. However, the accuracy happens to be directly proportional to information content, i.e. if the information content of a pattern is substantial to identify an individual out of a hypothetical sample size, then such a pattern is very reliable and tends for a higher level of accuracy in individual identification of the focal species. A proper literature review in the matters of advancements of Artificial Intelligence and computational aspects suggests that reliability assessments are difficult and provide better results by assuring the information content of a pattern.

The above discussion suggests that a good practice would be to assess the likelihood of a pattern replication. If the actual size of a population is large, then the complexity of the pattern used for identification should be considerably better in order to obtain the required information content. Hence, it suggests that the pattern complexity also holds a proportionate relationship with the population size of the focal species. Reliability hence can be expressed as the probability of a pattern being duplicated within a selected size of a population.

The authors (Pennycuick, 1978) suggest that a starting iteration would be, odds of 100:1 chosen and the addition of a few bits (1.7 bits) to make the odds 1000:1 to be preferred, and adding a similar step-up of more information in bits (1.7 bits) in order to increase reliability to the odds of 10000:1 chosen as a preferred reliability for most identification procedures. As most identification procedures also record locations of detection, even in cases of migratory avian species and Cetaceans, the reliability scores and the level of accuracy can be made higher. A similar argument had also been in (Scott, 1978) which states that, while considering all animals and birds, the questions that can be asked depends on the number of individuals that can be represented and recognized (Scott, 1978).

Considering all the various species that the research community has worked with, a broad categorization of the methods used for identification based on the similarity of the type of patterns that are used for unique identification is possible, as shown in Table 1.

Described below are detailed accounts on these approaches mentioned in Table 1, that are used to uniquely identify individuals of different birds and animals. The time complexity and space complexity of the documented techniques have been generated based on the computation model described for each process, giving a reasonable insight into each technique.

The animals that are identified by quantifiable Spot Patterns that cover only specific regions of the body are African Penguins, Whale sharks, Spotted Salamanders, Long tailed salamanders and Grey Seals.

Grey seals display spot patterns on their pelage.

To identify individuals, major patterns of the head and neck are taken into consideration. The patterns include black patterns over white. The whole coat in general becomes darker as the seals get older and due to sexual dimorphism, it becomes difficult to identify male individuals as the coat usually becomes totally black deeming it impossible to identify many males individually (Vincent *et al.,* 2001). The patterns on females get more prominent with time, making it easier. The matching is done manually by observers and the quality of the pattern is classified into good, medium and bad.

Whale Sharks are a species of fish that are vulnerable to extinction due to excess targeted fishing due to its international trade value (Arzoumanian *et al.,* 2005). A unique technique using numerical pattern matching is used to identify individual whale sharks. Previous to this usage of the numerical algorithm it is majorly used in the field of astronomy to compare patterns of stars. This algorithm is majorly based on the Groth's algorithm that helps to identify the positions of stars. To extract the region of interest the image is rotated so that the vertebral column is horizontal. After a generic method of Blob extraction, the algorithm basically finds out all possible triangles between any two coordinate points, these triangles should also be of a specific shape, having specified the location of the shortest, intermediate and longest edges between respective vertices. Many unwanted triangles are also filtered out, such as a threshold of triangles too small to be useful. The matching procedure after the generation of these triangles are based on the length ratios and the internal angle cosine for each triangle. A magnification factor is computed for each triangle compared, which leads to the computation of the similarity score.

For African Penguins, to extract the black spots on their chest right below the prominent black stripe of the neck (Burghardt *et al.,* 2004). During the acquisition of images, constraints such as weather conditions of no rain and neither images of dusk nor dawn are chosen. The chest is outlined by first starting with an initial seed outline and then expanded to the actual borders of the chest. The spot patterns are extracted by a series of image processing techniques that includes dilation, morphological differencing, spot kernel convolution, thresholding to generate a unique identifier. While querying with an identifier for a new image with the identifiers stored in the system, the weighted sum of matching every spot in the identifiers under comparison generates the similarity score for that pair. The database entries are also sorted based on similarity so that comparisons are fast. The natural markings on African penguins are the spots over their chest plumage that prove to be unique, and in a later work (Sherley *et al.,* 2010), the Animal ID software (Burghardt, 2008) is used for the enrolment of the penguins into the system, by selecting penguin images that the camera could capture with near-frontal positions, being orthogonal to the camera within 20^o to 30^o. The artificial intelligence algorithms in the Animal ID software handle the identification and matching operations. Three different matching techniques are executed to observe how they perform. A basic benchmark of performing a binary task to classify authenticatable images (images of individuals enrolled in the database) apart from those that aren't, creates an error statistic that is called a receiver operating characteristic (ROC). The first step involves a technique that normalizes scale, shift and rotation before calculating the mean square error between landmarks that are closest. This is a Procrustes analysis method that results in a rigid alignment technique. The second technique is by creating

Table 1. Features used for Non-Invasive Identification of Animals

Feature	Animals Identifiable by said feature
Unique Spot Patterns	African Penguins, Whale sharks, Spotted Salamanders, Long tailed salamanders, Grey Seals
Fluke/ Fin Deformation Patterns	Great White sharks, Dusky dolphins, Spinner dolphin,Bottlenose dolphin, Long finned pilot whale,Sperm whale, Humpback whale, other cetaceans
Coat Patterns	Tigers, Cheetahs, Amphibians (yellow bellied toad),Zebras, Marbled salamanders
Whisker Spots	Leopards, Polar Bears, Lions, Sea Lions
Deformation of body part	Elephant, New Zealand Sea Lions,
Face Recognition	Chimpanzees, Lemurs.
Unique bodyfeatures	Swans, Badgers

shape contexts of the spot neighborhoods by their polar histograms, sets of such histograms are used for comparison between individuals as a matching technique. The third method is an extension to shape contexts by creating a distribution context that is distortion-specific, which is produced by generating a model for the landmark location uncertainty. To further quantify the system's performance statistical values like Genuine Acceptance Rate (GAR) and Failure to Enroll (FTE) rate are also formulated. Methods like attentional cascades and fast rectangular pixel sum is used so that the process can be sped up to be used in video as well.

Long-Tailed Salamanders have unique spot patterns on the dorsal side of their body. A spot recognition method can be performed to recognize individuals with the help of the overall pattern along with the spot count (Jonas *et al.,* 2011). It is based on the use of just the number of spots present on the head region, which are the dorsal spots that are on the region that starts from the nose area to the area just at the beginning of the forelimbs, not considering the spots that occur laterally on the sides. The ANOVA method is used to determine if there exists a relationship of the number of spots with the length of the salamander. The Least Significant Differences mean separation test revealed that the number of spots increased with the increase in length. Even though it could be inferred that spot patterns increased as the salamander grew, the general patterns of the spots remained unchanged and unique. It is also observed that there is a higher tendency for spots to appear than disappear. The pattern recognition of spots hence proved to be a more useful technique.

For spotted salamanders a method to generate a code based on the spot patterns on the individuals is used (Chase *et al.,* 2015). These codes are generated by recording the number of spots present on six regions on the dorsal part of these salamanders. Since there exists an issue of pattern codes being repeated among few individuals, there is an extra bit of infor-

Table 2. Algorithms and techniques used for Animals with Spot Patterns

Algorithms and techniques	Time Complexity	Space Complexity	Remarks
Haar Transform	$O(\log_2 N)$	$O(N)$	Haar-like feature extraction methods are used to train classifiers to detect the area of interest along with the patterns in them by learning simple local luminance features, in African Penguins
AdaBoost	$O(N^2)$	$O(N)$	Used to take advantage of the combined performance of all the strong classifiers by taking classification and regression trees remodelled from the feature extraction methods as the input to the algorithm, for African Penguins
Sobel feature extraction	$O(N^2 \log N^2)$	$O(N^2)$	To make sure that there exists a penguin chest in the image the Sobel-like kernal detection techniques are used so that a detailed chest outline could be generated.
Procrustes Algorithm	$O(N^3 \log N)$	$O(N^3)$	To compare for a match, a form of the Procrustes algorithm is used, in African Penguins.
Polar Transform	$O(N \log N)$	$O(N^3)$	A polar transform of the chest pattern, is used to compare for a match for African Penguins
Groth's Algorithm	$O(N \log N)$	$O(N)$	The triangles and their internal angles, generated from this algorithm is used for the comparison for a match, in whale sharks.
Global Thresholding	$O(N^2)$	$O(N)$	To reduce noise, used for both Whale sharks and African penguins.
ANOVA	$O(N^2)$	$O(N)$	Used to determine the relationship between the spot pattern and the length of the tail in long-tailed sala manders.
Logistic Regression	$O(N^4)$	$O(N^2)$	Used to determine the accuracy of identification in spotted salamanders.
Manual observation	$O(N)$	$O(1)$	Done for the spot patterns in Grey Seals.

mation recorded for such individuals. To determine the accuracy of the identification technique, logistic regression is performed in the software R. The Binary Response variable in this case is defined over how likely the system is able to identify an individual correctly, a yes or no response and the likelihood of this outcome as well.

The animals that are identified by fluke or fin patterns are Great White sharks, Dusky dolphins, Spinner dolphins, Bottlenose dolphins, Long finned pilot whales, Sperm whales, Humpback whales, and other cetaceans.

The identification for dolphins is done by creating a curvature description of its dorsal fins, rather than measuring and comparing the curvature itself (Araabi *et al.,* 2000), by assigning a string of characters that represent significant and insignificant primitives. This provides as a kind of feature reduction approach to dimensionality reduction. Every fin is thus assigned character strings, which are matched to new fins to provide similarity scores. Further calculations provide us with a family of semantic / syntactic distances between two strings, which yields a composite distance to measure dissimilarity. For a multispecies approach, a live wire algorithm to locate the fin boundary along with the help of noise reduction techniques and various smoothing functions are applied. Even though the data entry required experts to handle the inputs the search mechanism proved that computers are more effective in terms of time saved (Hillman *et al.* 2003). The number of incorrect matches served as a criterion to measure the performance of the system. The issues faced by this system are mostly due to noise in the images posed by environmental artefacts like glare, splashes etc. Hence quantification of image quality is a required parameter that has a direct effect on the performance of the system.

For Humpback Whales the fluke patterns such as notches, holes, spots, lines etc. are observed and recorded manually onto computer systems as the storage search and retrieval media (Mizroch *et al.,* 1990). The coding scheme is also totally observer based like the techniques used for identifying swans (Scott, 1978). Independent SCAN and MATCH functions are created for an efficient retrieval of individual information from the search space with the MATCH function searching for the exact match given the encoded identification scheme and the SCAN function that returns groups of results that stand true to the partial encoded information provided (e.g., all whales in the database with striped flukes). The main issue is the quality of the image captured and displayed for the observer who makes the categorization codes for each of the whales. This introduces many inconsistencies like observation and observer ambiguities. The system developed (Mizroch *et al.,* 1990) records the recognition quality for each observation into four categories of excellent, moderate, poor and 'not coded' – acting like a new observable measure. This method gives us a quantifiable approach to confidence in the system. The other valuable insight is that a single photograph of a unique individual is not enough to make a new entry in the database, there should be at least two images as a minimum to describe an individual. The measurement of quality of the data collected is an important aspect when it comes to identifying individuals. An interesting observation duly noted (Friday *et al.,* 2000) is that, highly distinctive individuals are more easily identifiable from photographs of lesser quality, but this isn't really a strong point in support of photographs of lesser quality due to the fact that less distinctive individuals are totally left out. Thus, the quality of photographs also depends on how distinctive each individual is from the rest. To quantify both the aspects of quality of photographs and the distinctiveness of an individual, an evaluation criterion involving people as judges to rate each aspect is proposed. The issue of how each judge is supposed to differentiate between both aspects is resolved by breaking each aspect into various variables that are rated individually that contributing to the overall score towards each aspect. The process evaluated the agreement between the judges by evaluating the average kappa statistic score for each judge derived from the agreement statistics of every possible pair of judges – an inter-rater agreement.

For Sperm whales, an operator enters values to the computer system in a unified and standardized structure and the computer is programed to process the identification based on the formulas and techniques specified. A technique involving the calculation of value of match and accuracy of match to derive the match coefficient proves to be an efficient method (Whitehead, 1990). These values are possible only after a process that involves an operator to look at the images, in this case the fluke images of sperm whales, and then enter a coded representation, as identifiers of the various marks and shapes on the edge of the fluke by their proportional distance, generating a unique code of fixed length for each individual. The measures of orientation, tilt and resolution are used as parameters that indicate the quality of the image.

For Great White Sharks the paradigm of non-invasive identification of individuals is approached by a non-linear model that incorporates multi-scale segmentation with the uniqueness selected over the space created over scale variations (Hughes and Burghardt, 2017). This approach gives the ability to have multiple labels for each image. Contour information is collected rather than colour or texture information, which is then used to generate an ultra metric contour map. The contour information helps to generate multiple boundary descriptors that make up a Bag of Boundaries (BoB), which is used to detect the presence of the shark fin by classification techniques. After this a Bag of Normals (BoN) is generated, which encompass the shape information along with the regional information. While encoding the contour information the Difference of Gaussian (DoG) is used as a filter to normalize the curves. The training is done with the ground truth contour labelled by a human. The quality score is generated by thresholding of the L2 normalized values of opponent-SIFT (scale invariant Feature Transform) for local appearance which is independent of direction and the threshold values of contour shape by the histogram of boundary normals which are dependent on direction. The final matching process is handled by a random forest structure, that ranks the various contours. This ensures that the search time is lesser than a linear search time. The standard DoG norm, which is rotation invariant, is combined with LNBNN (Local naive Bayes nearest neighbor), so that a recognition baseline is generated, after which a scoring pattern can be generated. To achieve scale invariance, classification is done separately for each scale and then combined, also affine-covariant sift descriptors are used. The overall model automates the process of obtaining the identity of the animal from the natural image provided to the system. The results of the system are tested on the basis of Average Precision (AP) over a Precision-recall Curve (PR). They are also able to create a population wide fin-space, which is like a single model for the whole species, with unique patterns for individual distinctiveness. This is done so that the system doesn't learn patterns of uniqueness separately for each individual and that the learning is not overfitted over individuals. To achieve this, a standard for partitioning each fin contour is devised. Thus, a separate score is generated for each fin-space as well as for each class.

The animals that are identified by their coat patterns are Tigers, Cheetahs, Amphibians (yellow bellied toad), Zebras and Marbled salamanders.

Tigers can be identified by differences in their stripe patterns (Karanth and Nichols, 1998). Since tigers and leopards occur syntopically, the estimation of their densities, can be performed by capture-recapture theory. The tiger expresses surface patterns on its body coat which proves to be unique (Hiby *et al.,* 2009). A 3-D surface model is proposed to solve issues related to the angle of the photograph and the posture of the tiger. The patterns are

Table 3. Algorithms and techniques used for Animals with Fluke or Fin Patterns

Algorithms and techniques	Time Complexity	Space Complexity	Remarks
String Matching or Encoding	O(N)	O(N)	Arbitrary Encoding used for sperm whales, hump back whales and Dolphins
Kappa Statistic Score	O(1)	O(1)	Derived from the statistic of multiple judges for the quality of the photographs.
Live Wire Algorithm	O(N)	O(N)	For a multispecies fin identification approach
Affine Transform	$O(N^4)$	$O(N^3)$	An additional method for scale invariance used in the identification of great white sharks.
Difference of Gaussian (DoG)	$O(N^3 \log N)$	$O(N^2)$	To achieve rotational invariance for the fins of the great white sharks.
SIFT (Scale Invariant Feature Transform)	$O(N^2)$	O(N)	To normalize over scale of the fins in great white sharks.
Randomforest algorithm	$O(N^2 \log N)$	$O(N^2)$	For the final matching in great white sharks.
Local Naive Bayes nearest neighbor (LBNN)	$O(N^3)$	$O(N^2)$	To create a baseline for the fin recognition in great white sharks.

matched to produce a similarity score with the help of two algorithms that complement each other. It is basically a semi-automated system to extract the pattern, and the software goes by the name 'Extract Compare'. Due to the method relying on user inputs there is a risk of subjectivity in the case of model fitting, which is being looked into for further automation.

For Cheetahs the technique of having a 3-D mathematical model (Kelly, 2001) of the coat or body part in consideration is a direct solution to resolve multiple issues that exist in the available 2D images such as, position of the animal, angle of the camera, tilt, scaling, orientation, and other issues like partial capture. The model is made in such a way that it fits on the photograph after a few points are identified such as the position of the shoulder.

For Zebras, six location points are used to define the region of interest (Foster *et al.*, 2007). After the selection of region of interest, the stripes within them are skeletonized to single pixel lines. Comparing the complexity of the patterns occurring on human fingerprints, the patterns on the zebra coat contain lesser complexity, granting the fact that human fingerprints are way more feasible for individual identification procedures.

Moving towards marbled salamanders, with the widespread reduction of pond basins in the ecological habitat of marbled salamanders, there is a risk of this species to be endangered (Gamble *et al.*, 2008). The traditional technique of capture-mark-recapture proved to be less accurate and very time consuming with significant intrusion on the regular life of these salamanders. Therefore, the method of photographically recording the dorsal patterns while individuals are captured is found to be more feasible. These dorsal patterns act as unique fingerprints. The process is semi-automated as one has to digitally mark the dorsal midline after which the system pre-processes the image to straighten the midline in effect straightening the whole image warping it from its original form. But the pattern is preserved, and this process makes all the comparisons to have a standard form. Multi scale PCA is used to extract and store features, also the whole database of images is ranked on the similarity with each other. The outcome of this research provides new information on migratory patterns and total time spent within basins, due to the information recorded based on individual identification.

The animals that are identified by their whisker spots are Leopards, Polar Bears, Lions and Sea Lions.

The study on leopards is done on captive leopards making the setting quite invasive (Miththapala *et al.*, 1989), the readings being taken while the animals are under anesthesia. The algorithm used here is the Oblique Principal component cluster analysis. To minimize correlation, unique character sets are generated. The Binomial Theorem is used to check these character set spot clusters for the sum of probabilities of zero occurrence and that of a single occurrence to be at a maximum (more than 0.95), to be feasible to be used as a characteristic of identification used in the process.

For Australian sea lions the whisker spots are selected manually and stored in the system with only the right side of the sea-lion's face chosen for

Table 4. Algorithms and techniques used for Animals with Coat Patterns

Algorithms and techniques	Time Complexity	Space Complexity	Remarks
PCA /Multi Scale PCA	$O(N^3)$	$O(N^2)$	Extracting and storing features in marbled salamanders.
String Matching or Encoding	$O(N)$	$O(N)$	Encoding the presence or absence of patterns by the region of the body (yellow bellied toad).
Edge Detection	$O(N^2 \log N^2)$	$O(N^2)$	Used in the tool that extracts the features in the stripes of tigers.
3-D mathematical model superimposition	$O(N^6)$	$O(N^5)$	Solves issues related to the angle of the photo graph and the posture of tigers. A 3D plane is used in the case of cheetahs.
Correlation coefficient	$O(N)$	$O(1)$	Created from the models of known individuals of cheetahs.
Fingerprint Identification Technique	$O(N^4)$	$O(N^2)$	Patterns on zebras are similar to human fingerprints.

the identification process. Scars are not used as the sealions undergo molting. To standardize the images 3 manual reference points are taken. To calculate similarity scores, the Euclidian distance between points in the 1st image in comparison and the nearest points in the 2nd image and then the reverse distance from the 2nd image to the 1st is calculated, also to improve the similarity score the image is shifted manually in steps. An adapted Chamfer distance transform is also used. To improve the system, the use of Groth algorithm is suggested so that a geometric relationship is made between the spots. The encoding and comparison is achieved by superimposing a grid over the manually marked spots with the selection of grid-cell size by accounting for the angle of the photographs. The grid is filled based on the presence or absence of spots. The accuracy of the system decreased when the photograph is not at the 90° of the right face-profile of the sealion.

For Polar Bears to validate the reliability of the patterns, the complexity and information content is measured (Anderson, 2007). To compute the similarity scores between images, the Chamfer distance transform is used. Images of only one side of the bear's muzzle is used for analysis, due to the notion of a possible correlation of the patterns of both sides that might result in a statistical bias. The whole procedure of analysis including, selection of photographs, locating the spots, information content and reliability, mutual exclusiveness are repeated across 3 judges over the same number of photographs to account for consistency. The variations in accuracy of identification is analysed based on the angle and quality of the photographs. Affine transformation, global thresholding, adaptive thresholding, histogram specification, logarithmic transformation, are the algorithms used for pre-processing. Based on the similarity scores, and the errors of matching, a tolerance graph of false positives is generated from which the similarity threshold is derived. Even though three reference points are manually selected, the time taken for the processing on one image is under a minute.

The animals that are identified by the unique deformations in various body parts are Elephants and New Zealand Sea Lions.

For the New Zealand Sea Lion, while trying to capture photographs of the abnormalities in the flippers and other features like large scars and the size of the lower canine, it is found that the animal may not comply, may be aggressive or the animal may depart before the photograph is taken (McConkey, 1999). Also the recorded features of body and facial scars may change by the next moult (shedding feathers).

For Elephants, to extract the patterns of the nicks and cuts on the elephant's ears that pose to be unique, a common edge detector wouldn't be feasible to detect them due to the uniform color and texture over the elephant's body, over which the images of the ears overlap (Ardovini *et al.*, 2008). And most of the curve detector algorithms handle these patterns by assuming connected curves.

Table 5. Algorithms and techniques used for Animals with Whisker Spots

Algorithms and techniques	Time Complexity	Space Complexity	Remarks
PCA /Multi Scale PCA	$O(N^3)$	$O(N^2)$	Cluster analysis for leopards.
Chamfer distance transform	$O(N^2)$	$O(N)$	Pairwise comparison of these distances are possible only under minimal variation in angle in the photographs of Australian sea lions.For Polar Bears Chamfer distance transform is used to generate similarity scores between images.
Groth's Algorithm	O(NlogN)	$O(N)$	To improve the system for Australian sea lions by creating a geometric relationship between the spots.
Affine Transform	$O(N^4)$	$O(N^3)$	The various processes used to extract the whisker spots from the images of polar bears.
Global Thresholding	$O(N^2)$	$O(N)$	
Adaptive Thresholding	$O(N^3)$	$O(N^2)$	
Histogram Specification	$O(N)$	$O(1)$	
Logarithmic Transform	$O(N)$	$O(N)$	

Hence the choice of Generalized Hough Transform, with the ability of handling non-connected curves – a context-free approach. With core efforts to automate the process their system is semi-automated with users required to input positional information like the position of the head and various reference points like the location of the eyes, etc. After this step a canny edge detection algorithm charts out the edges that it sees, the user then has to input the start and end point of every curve and nick. These patterns are stored as a sequence for that specific image. The matching algorithm first compares to check for portions of matching sequences of just the presence of these nicks or cuts in the same sequence. By increasing the number of cuts for every successful sequence portion matched (eg. Set of 2 cuts matched then going on for sets of 3 cuts in the matching portion of the sequence), the whole sequence is checked. After this each nick or cut is compared by the shape difference algorithm. Every curve is compared to give a final average dissimilarity value, as the output of the matching process. This method proved to be much more accurate than other curve-matching algorithms.

The animals that are identified by Face recognition are Chimpanzees and Lemurs.

For Chimpanzees that have similar structural features to humans, common face recognition algorithms are observed to produce good results. The popular real-time object detection algorithm by Viola and Jones, combined with techniques that take advantage of Gabor descriptors along with SURF descriptors (taking into account texture and shape features), and Sparse Representation Classification (SRC) over global features along with support vector machines (SVM) over local features for the required classification routines are also used for a fine-tuned recognition algorithm (Loos and Ernst, 2013). The Locality Preserving Projection (LPP) method is the dimensionality reduction procedure adopted. With all these finer adjustments to the basic face recognition algorithms, there is an improvement in the accuracy for non-frontal images, which is an issue in a previous version of the system. Techniques like Deep convolution Neural Networks (CNN), using Gabor magnitude pictures (GMPs) are also used to improve efficiency. The Stochastic Gradient Technique is also used along with Backpropagation to compute gradients of intermediate layers (Freytag *et al.,* 2016).

For Lemurs images are normalized to extract a local binary Multiscale pattern by a patch-wise method (Crouse *et al.,* 2017). To normalize for facial hair and ambient lighting, various techniques are used. Inter-class similarity and intra-class variability are tested by considering each class as an individual. Linear Discriminant Analysis (LDA) is the dimensionality reduction technique used. The percentage of correct matches is termed as True Accept Rate (TAR). It is also observed that, in previous related studies of individual identification of chimpanzees, the Gabor features are feasibly used due to the lack of hair on the face compared to lemurs. Observations point that, clustering with classification can improve the identification system over unknown individuals in the wild.

The animals with other unique body features covered below are that Bewick Swans and Badgers.

The unique features of the Bewick Swans that described their individuality are drawn manually and various other features are recorded on identity cards designed specifically for this purpose (Scott,

Table 6. Algorithms and techniques used for Animals with unique Deformations

Algorithms and techniques	Time Complexity	Space Complexity	Remarks
String Matching or Encoding	O(N)	O(N)	Strings that describe the curvature of the dorsal fins of dolphins. Patterns in the elephant's ears are also stored as a sequence of strings
Curve Matching	$O(N^2)$	O(N)	Used to match the curve sequences on the edges of the elephant's ears.
Edge Detection	$O(N^2 \log N^2)$	$O(N^2)$	To extract the curves of the edges in the elephant's ears.
Generalized Hough Transform	$O(N^3 \log N)$	$O(N^3)$	For the ability to handle non-connected curves in the elephant's ear edges.
Context-Free Algorithms	$O(N^3)$	$O(N^2)$	Useful when it comes to edges of elephants ears, because of the undifferentiable colour.

1978). Hence the number of details that needed to be recorded, depended on the total population in consideration. This is initially done by recording the differences of one individual bird from the others and then resulted in maintaining an exhaustive record of individual features recorded systematically for all birds by having identity cards assigned to each of them. The unique features are the yellow and black patterns found on the upper surface and on either side of each bird's bill. To account for referencing and placing an individual into a fixed location in the population, a code is devised that has types of features and ranges of variations within those features with ranges having arbitrary order of precedence. This method is not a fixed standard, but it allows for a better method to search for an individual and also allowed for a method that can be used for computerization in later stages. The outcomes made possible are, social structure, seasonal and daily movements, life history, similarity of bill patterns in parents and offspring. Limitations: The observations are heavily dependent on the observer making these observations, hence introducing ambiguity amongst multiple observers and their recordings. The coding technique that is developed to quantify the shapes and patterns encodes observer biases into the code itself. And the system being physical involves a physical search through a huge list of encoded identifiers to deem whether the entry is new or not. Also, it is established that it is very difficult to assign a method for the identification of Cygnets (young swans).

For Badgers and their tail patterns, an accuracy of 95% is achieved for the tail patterns, by direct identification looking at the photographs (Dixon, 2003). Also, the variations in the appearance of the tail at any specific moment, apart from the patterns present on them, signifies various social interactions among individuals. Identification can also be done by the skull shape with less accuracy.

Conclusion

In order to prepare better models for the identifica-

Table 7. Algorithms and techniques used for identifying Animals by Face Recognition

Algorithms and techniques	Time Complexity	Space Complexity	Remarks
Locality Preserving Projection (LPP)	O(N)	O(N)	The dimentionality reduction technique used for chimpanzees.
Linear Discriminant Analysis (LDA)	$O(N^3)$	$O(N^2)$	The dimentionality reduction technique used for lemurs.
Gabordescriptors	$O(N^4)$	$O(N^2)$	Various techniques used for both chimpanzees and Lemurs to conduct the facial recognition task.
SURFdescriptors	$O(N^2)$	O(N)	
Viola and Jones object detection	O(N)	O(N)	
Sparce Representation Classification (SRC)	$O(N^3)$	O(N)	
Support Vector Machines (SVM)	$O(N^3)$	$O(N^2)$	
Stochastic Gradient Technique	$O(N^3)$	$O(N^2)$	
Matrix Logarithm Technique	$O(N^2)$	O(N)	
Deep Convolution Neural Networks	(CNN)	$O(N^5)$	$O(N^4)$

Table 8. Algorithms and techniques used for Animals with other Unique features

Algorithms and techniques	Time Complexity	Space Complexity	Remarks
String Matching or Encoding	O(N)	O(N)	An arbitrary code for all the variations recorded in Bewick Swans is manually generated.
Manual observation	O(N)	O(1)	Recording the tail patterns of badgers manually

tion of individuals in a focal species, a better understanding of traditional algorithms must be taken into consideration. Hence, this review article presents a study of various traditional and available algorithms in order to obtain a foundation for preparing better algorithms both in time and space complexities, with respect to the existing ones.

What has been observed from the previously available sections in the article, suggests that there are a few major takeaways that keep showing up in most of the individual identification approaches.

Semi-automated methods have issues of observer ambiguity or operator subjectivity just as the manual method, suggesting observer biases to be carried over to the observations being made, and in turn creating ambiguities in the information recorded. The workarounds observed are that there must be least two independent observers/operators that are required to record the information for identification (Vincent *et al.,* 2001). The issue extends even further, with questionable consistency of collected data over time due to different data collection methods. Also, identification based on researcher knowledge requires a lot of training, which is expensive and still has the problem of ambiguity due to biases.

The quality of the image depends on a variety of factors and plagues almost all the non-invasive techniques that use images or image based approached as the basis for identification. This makes image quality the main factor that determines identifiability, creating all kinds of observer and observation ambiguities. Another factor related to the quality of the photographs is the distinctiveness of an individual, by which highly distinctive individuals don't need as much photographic quality as much as those with lesser distinctiveness. The method of workaround used for this issue is by having multiple judges to rate both aspects on multiple variables, and then deriving an agreement statistic of every possible pair of judges (Friday*et al.,* 2000).

Weather conditions and factors like the time of the day also need to be taken into consideration for the best quality of the image captured, with preferences for no rain, and neither images of dusk nor dawn (Sherley *et al.,* 2010).

An issue that still prevails is that there is no system in place to quantify the degree of change in appearance over time in an individual animal (Crouse *et al., 2017*).

The preliminary studies and the takeaways from the discussion suggest that, improvement in the various individual identification systems can be achieved with a few workarounds for the various processes involved. Such as, the indexing process of the identity system can speed up the process of searching (Scott, 1978). Also, while handling multiple observers without proper data recording standards, a fixed length identification code can be used to reduce the ambiguity (Whitehead, 1990).

There are few workarounds observed to combat the issue of the quality of the image which are mentioned here. One approach is the creation of a new measure of image quality and using it as a metric for evaluation, so that there may be an improvement in the confidence of the system. The idea behind this workaround is to quantify the various problems in image quality by generating a score based on the various factors like, the position of the animal, the angle of the camera, the tilt, scaling, orientation, and partial capture involving glare or environmental artefacts, resolution and other image quality problems (Kelly, 2001).

An additional specification that at least two images of an individual is required for creating a new entry in the database, reducing the scope of ambiguity due to image quality (Mizroch *et al.,* 1990). This specification of minimum number of inputs for a new entry is a necessary factor while looking into approaches that implement machine learning to create mathematical models for each individual as templates.

During the process of training the system, most of the times there is the issue of overfitting unique patterns over the model created for each individual. A novel workaround observed to address this problem is the generation of a species-space model, like a single model for the whole species (Hughes and Burghardt, 2017).

For animals with an anatomy of having a muzzle, there is an issue of the possibility of correlation of patterns on both sides of the animal's muzzle, that has been addressed in previous research and the only workaround used was to consider the patterns of only one side of the muzzle for identification processes (Anderson, 2007; Osterrieder *et al.,* 2015). A suggested proposition would be that of having the combined pattern of both the sides taken to be used for identification. A requirement of two images each for both sides of the muzzle as well as two front facing images could be imposed as a minimum to create an entry for a new individual.

A major improvement in such a system would be the capability that could be added by a mathematical model that quantifies the degree of change in appearance over time for an individual animal (Crouse *et al.,* 2017). This increases the accuracy for longitudinal studies that conduct repeated observation of parameters over time. These studies shed light on life history variables and the heritability of traits.

Therefore, a standardized system model that encompasses all such requirements that wards off most ambiguities across different time sessions and across multiple researchers is the way forward. Such a system also reduces the required training of the researchers within this field, with an effect of potential acceleration in training. It is required to be able to quantify the uniquely identifiable aspects of appearances with higher accuracy by reducing noise and possible ambiguity, hence, various normalization techniques must be implemented.

Towards the creation of fully automated individual identification systems, it can thus be concluded that, for non-invasive individual animal identification, modelling the system mathematically with the problem specific computational complexity in mind is necessary because there does not exist a 'one size fits all' approach (Crouse *et al.,* 2017).

References

Anderson, C. 2007. Individual identification of polar bears by whisker spot patterns.

Araabi, B.N., Kehtarnavaz, N., McKinney, T., Hillman, G. and Wu¨rsig, B. 2000. A string matching computer-assisted system for dolphin photoidentification. *Annals of Biomedical Engineering.* 28 (10) : 1269-1279.

Ardovini, A., Cinque, L. and Sangineto, E. 2008. Identifying elephant photos by multi-curve matching. *Pattern Recognition.* 41(6) : 1867-1877.

Arzoumanian, Z., Holmberg, J. and Norman, B. 2005. An astronomical pattern-matching algorithm for computer-aided identification of whale sharks Rhincodon typus. *Journal of Applied Ecology.* 42(6) : 999-1011.

Burghardt, T. 2008. Visual animal biometrics: automatic detection and individual identification by coat pattern (Doctoral dissertation, University of Bristol).

Burghardt, T., Thomas, B., Barham, P. J. and Calic, J. 2004. Automated visual recognition of individual african penguins.

Chase, F.W., Hardie, B.E., Kern, M.M., Pittman, S. E., Harden, L.A. and Dorcas, M.E. 2015. Evaluation of two individual identification techniques for spotted salamanders (*Ambystoma maculatum*). *Herpetological Review*. 46 : 192-196.

Crouse, D., Jacobs, R. L., Richardson, Z., Klum, S., Jain, A., Baden, A. L. and Tecot, S. R. 2017. LemurFaceID: a face recognition system to facilitate individual identification of lemurs. *Bmc Zoology.* 2 (1): 2.

Dixon, D.R. 2003. A non-invasive technique for identifying individual badgers Melesmeles. *Mammal Review.* 33(1) : 92-94.

Foster, G., Krijger, H. and Bangay, S. 2007. Zebra fingerprints: towards a computer-aided identification system for individual zebra. *African Journal of Ecology*. 45(2) : 225-227.

Freytag, A., Rodner, E., Simon, M., Loos, A., Ku¨hl, H. S. and Denzler, J. 2016. September). Chimpanzee faces in the wild: Log-euclideancnns for predicting identities and attributes of primates. In *:German Conference on Pattern Recognition* (pp. 51-63). Springer, Cham.

Friday, N., Smith, T. D., Stevick, P. T. and Allen, J. 2000. Measurement of photographic quality and individual distinctiveness for the photographic identification of humpback whales, Megaptera novaeangliae. *Marine Mammal Science.* 16(2) : 355-374.

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 (1) : 170-180.

Hiby, L., Lovell, P., Patil, N., Kumar, N.S., Gopalaswamy, A.M. and Karanth, K.U. 2009. A tiger cannot change its stripes: using a three-dimensional model to match images of living tigers and tiger skins. *Biology Letters.* 5(3) : 383-386.

Hillman, G.R., Wursig, B., Gailey, G. A., Kehtarnavaz, N., Drobyshevsky, A., Araabi, B. N. and Weller, D.W. 2003. Computer-assisted photo-identification of individual marine vertebrates: a multi-species system. *Aquatic Mammals.* 29 (1) : 117-123.

Hughes, B. and Burghardt, T. 2017. Automated visual fin identification of individual great white sharks. *International Journal of Computer Vision.* 122(3) : 542-557.

Jonas, M., Bowman, J.L. and Nazdrowicz, N.H. 2011. Using spot pattern to identify individual Long-tailed salamanders. *Herpetol. Rev.* 42 (4) : 520-522.

Karanth, K.U. and Nichols, J.D. 1998. Estimation of tiger densities in India using photographic captures and recaptures. *Ecology*. 79 (8) : 2852-2862.

Kelly, M.J. 2001. Computer-aided photograph matching in studies using individual identification: an example from Serengeti cheetahs. *Journal of Mammalogy.* 82(2), 440-449.

Loos, A. and Ernst, A. 2013. An automated chimpanzee identification system using face detection and recognition. *EURASIP Journal on Image and Video Processing.* 2013 (1) : 49.

McConkey, S.D. 1999. Photographic identification of the New Zealand sea lion: a new technique.

Miththapala, S., Seidensticker, J., Phillips, L.G., Fernando, S.B.U. and Smallwood, J. A. 1989. Identification of individual leopards (*Panthera parduskotiya*) using spot pattern variation. *Journal of Zoology*. 218 (4) : 527-536.

Mizroch, S. A., Beard, J. A. and Lynde, M. 1990. Computer assisted photo-identification of humpback whales. *Report of the International Whaling Commission.* 12: 63-70.

Osterrieder, S. K., Salgado Kent, C., Anderson, C. J., Parnum, I. M. and Robinson, R. W. (2015). Whisker spot patterns: a noninvasive method of individual identification of Australian sea lions (Neophocacinerea). *Journal of Mammalogy*. 96(5) : 988-997.

Pennycuick, C.J. 1978. Identification using natural markings. In :*Animal Marking* (pp. 147-159). Palgrave, London.

Plaia¸su, R., Hartel, T., Bancil'a, R. I. and Coga¢lniceanu, D. 2005. The use of digital images for the individual identification of amphibians. *Studii ¸siCercet¢ari.* 10: 137-140.

Scott, D. K. 1978. Identification of individual Bewick's swans by bill patterns. In : Animal marking (pp. 160-168). *Palgrave*, London.

Sharp, T., Saunders, G. and Mitchell, B. 2007. Marking of pest animals used in research. *Institute for Laboratory Animal Research Journal*, 44 : 1-15.

Sherley, R. B., Burghardt, T., Barham, P. J., Campbell, N. and Cuthill, I.C. 2010. Spotting the difference: towards fully-automated population monitoring of African penguins Spheniscus demersus. *Endangered Species Research.* 11 (2) : 101-111.

Vincent, C., Meynier, L. and Ridoux, V. 2001. Photo-identification in grey seals: legibility and stability of natural markings. *Mammalia.* 65 (3) : 363-372.

Whitehead, H. 1990. Computer assisted individual identification of sperm whale flukes. *Report of the International Whaling Commission.* 12 : 71-77.