Application of multi-species photo-ID database management systems: a user's perspective

Stephen C. Y. Chan [1,3], Scott Y. S. Chui [2,3], Leszek Karczmarski [1,2,3,4]

[1]Division of Cetacean Ecology, Cetacea Research Institute, Lantau, Hong Kong

[2]Division of Animal Behaviour, Cetacea Research Institute, Lantau, Hong Kong

[3]School of Biological Sciences, University of Hong Kong, Pokfulam, Hong Kong

[4]Department of Zoology and Entomology, University of Pretoria, Hatfield, South Africa

*Correspondence to Leszek Karczmarski. Email: leszek@cetacea-institute.org

Identification of individual animals in the field goes back to the 1930s and early studies of animal behaviour (Lorenz 1935, 1937). In the 1960s, it has been greatly popularised by the seminal work of J. Goodall on the Kasakela chimpanzee (*Pan troglodytes*) community at Gombe Stream Reserve (currently National Park), Tanzania (Goodall 1971, 1986). Nowadays, photographic identification of individual animals (individual photo-ID), based on naturally occurring, distinctive and individual-specific marking is a well-established and widely used research technique. It has been applied across a range of taxa, from insects (Coleoptera, e.g., Romiti et al. 2017; Díaz-Calafat et al. 2018), to amphibians (e.g., Bendik et al. 2013; Mettouris et al. 2016; Patel and Das 2020), reptiles (e.g., Treilibs et al. 2016; Gatto et al. 2018; Araujo et al. 2019; Desai et al. 2022), various species of elasmobranchs (e.g., Holmberg et al. 2009; Hughes and Burghardt 2017; Benjamins et al. 2018; Armstrong et al. 2019; Lewis et al. 2020), birds (e.g., Ferreira et al. 2020), and numerous species of mammals, both marine and terrestrial (e.g., Tavares et al. 2017; Muller 2018; Koivuniemi et al. 2019; Turner et al. 2021; Chan et al. 2022a, 2022b; Marneweck et al. 2022; Oberosler et al. 2022; Reed et al. 2022). While the type of individually distinctive marking may differ between taxa (e.g., pattern of spots and stripes, notches on flukes and fins, facial features or morphometric characteristics; e.g., Karczmarski et al. 2022a, 2022b), the technique has proven to be a powerful yet non-intrusive tool in field studies of animal behaviour, ecology, and various aspects of their biology.

First published accounts of computer-assisted photo-ID application were given in three case studies that appeared in a thematic collection of papers edited by Hammond et al. (1990). In one case, a custom-designed software was used to scan images of sperm whale (*Physeter macrocephalus*) flukes onto a digitiser tablet and generate a sum of similarities of user-defined descriptors (Whitehead 1990). In the next case, images of humpback whale (*Megaptera novaeangliae*) flukes were labelled against a 14-sector map of a fluke following a user-defined list of possible marks. Subsequently, a custom-designed software generated a similarity matrix to compare the individuals (Mizroch et al. 1990). The third case offered a first feature engineered individual-ID system (Hiby and Lovell 1990). Using photographs of animal head (in this case, grey seal *Halichoerus grypus*), a custom-designed software generated a grid-based 3-D model where individual-ID was represented by the summation of greyscale pixel values of each grid cell. A decade later, Kelly (2001) adopted the Hiby–Lovell approach to digitised images of cheetahs (*Acinonyx jubatus*), identifying points with known 3-D coordinates (e.g., shoulder and hip) to line up the 3-D model with a 2-dimensional photographic image, and used cheetahs' pelage pattern as the identifiable feature. Specific techniques for individual identification have since improved considerably, with ever-growing application of semi- and fully-automated systems (Schneider et al. 2019).

Recent years have witnessed notable developments in the application of computer vision, machine learning, and artificial intelligence (e.g., Parham et al. 2018; Miao et al. 2019; Tabak et al. 2020; Renò et al. 2020; Miele et al. 2021; Vidal et al. 2021; Bodesheim et al. 2022; Schneider et al. 2019, 2022). Using deep learning methods for object detection and similarity comparisons, current systems can extract animals from images and users can train deep learning classifiers to reliably (re-)identify species and individuals. There are currently numerous freely available systems for a wide range of taxa; some operate algorithms developed with a specific species in mind, while other offer a more generalised platform applicable (or adaptable) for a multi-species use (e.g., Berger-Wolf et al. 2017; Brust et al. 2017; Bogucki et al. 2018; Maglietta et al. 2020; Thompson et al. 2020; Miele et al. 2021; Blount et al. 2022; Clapham et al. 2020, 2022). Some have proven to be remarkably successful, facilitating automated comparisons of large photo-ID datasets (hundreds of thousands individual IDs) across large geographic scales (e.g., Cheeseman et al. 2022), while other remain at present impractical (e.g., de Silva et al. 2022) or insufficiently reliable (e.g., Morrison et al. 2016) for field research applications. Given the developmental changes in feature appearance and mark changes over time (e.g., Carlson and Mayo 1990; Waye 2013; Wattegedera et al. 2022), the concepts of active learning with human-in-the-loop and continuous lifelong learning (Käding et al. 2016; Bodesheim et al. 2022) come especially handy. With algorithms for deploying lifelong learning in photo-ID field studies available and ready for use (Bodesheim et al. 2022), this approach is particularly valuable in long-term studies of long-lived animals.

As digital technology is fast advancing, so is the sophistication and accuracy of computer-assisted platforms and fully automated systems. What seem to be lagging behind, however, are the means of data management of increasingly complex datasets built around the sighting histories of individuals. Be it environmental, behavioural, spatial, acoustic, genetic, or other, there are a wide range of data that are commonly gathered during photo-ID surveys. As many high-quality digital photographs can be obtained at low cost in a short space of time, vast amount of individual-ID associated data fast accumulates. Linking such different types of data with each individual-ID record provides a means to efficient data management, facilitating data processing and analyses that could be performed even in a remote field research location with restricted internet access. While important in any field research project, it is especially so in long-term individual-based studies.

With a few notable exceptions, most current individual identification systems offer specialised matching routines but are limited in the means of data integration and management and—to the best of our knowledge—none as yet offers an all-encompassing generalised system that could assist users at all stages of data collection, processing, ID-matching, and data management that would link sighting information with various user-defined parameters. A fully integrated and dynamic approach is therefore needed to provide a user-friendly platform that would assist users not only at the matching stage of individual photo-ID, but also at the multifaceted process of data management and analyses that take place after individual matching is completed.

Wildbook (https://wildme.org/#/wildbook), an autonomous web-based platform for data archiving and photo-identification of various animal species (Berger-Wolf et al. 2017) along with its taxa-specialised platforms such as Flukebook (https://www.flukebook.org; Blount et al. 2022), Sharkbook (https://www.sharkbook.ai), MantaMatcher (https://mantamatcher.org), GiraffeSpotter (https://giraffespotter.org), Whiskerbook (https://www.whiskerbook.org), and several other, offer appreciable though not extensive data-management tools, allowing input of a number of ecological, behavioural, biological, and other associated data. Users can work on unprocessed photographic data as well as import existing catalogues, and export their data in formats supporting third-party tools for further applications and analyses (e.g., mapping, mark–recapture, socio-behavioural analyses, etc.) and biodiversity data frameworks such as Ocean Biodiversity Information System (OBIS) and Global Biodiversity Information Facility (GBIF). In overall, the platform does a great job bridging the latest computer vision technology with the needs of field research, identifying individual animals with the latest algorithms for photographic ID of that taxonomic group. It facilitates integration of data from collaborative research and citizen science, and effective management of photo-ID catalogues along with basic data-management tools, though the latter remains open to further advancements.

A valuable contribution to efficient management of photo-ID datasets is the recently published R-based system dubbed 'catRlog' (https://github.com/ericmkeen/catRlog; Keen et al. 2022). As a self-contained computer directory with custom apps embedded throughout, catRlog serves as a workflow organiser that simplifies and streamlines photo-ID data processing. The apps are open-source, easily updated and customisable for the particulars of specific projects, providing interactive tools for data processing and management. The system generates datasets ready for further analyses (e.g., mark–recapture or socio-behavioural modelling) and third-party tools, and is sufficiently generalised to be applicable for almost any photo-ID project of any species in any habitat. The photo-ID matching component of catRlog, however, although neatly designed and computer-assisted throughout, is not automated and relies exclusively on the human analyst performing all matching procedures, which remains challenging when ID-matching across large datasets.

An older system, DISCOVERY (https://www.cetacea-institute.org/discovery and https://cetaecoresearch.com/research-software-discovery.html; Gailey and Karczmarski 2012), also offers dynamic functions that can be tailored to suit various project-specific requirements. Similarly as in catRlog, the photo-ID-matching component is not automated, and although the process is well-streamlined, it requires strong engagement of the photo-analysts and remains labour-intensive when ID-matching across large datasets. At its core, DISCOVERY was conceived as data-management system that can centralise and maintain a single database for research projects working on multiple species in multiple study areas, collecting data at large geographical scales and between multiple research teams working on different databases. A more recent component, Discovery Droid app developed for Android mobile phones (https://play.google.com/store/apps/details?id=com.cetaecoresearch.discoverydroid) helps to streamline computerised data collection in the field and stores the data (e.g., survey effort, geographic position, environmental data, sighting-associated information, and other user-defined parameters) in a database that can be imported into DISCOVERY software system. The system provides means to integrate sighting information with numerous user-defined parameters (e.g., environmental, geographic, and other), store photo-ID data along with a suite of associated information, process and manage all stored photographic and associated data, and link historic datasets with newly obtained data to form continuous complete datasets. The system offers basic analytical tools and graphic displays of data, and facilitates integration of all collected and stored data for further analyses (e.g., mark–recapture, movement, spatial, socio-behavioural, etc.) with third-party tools. See Appendix for a brief summary.

As potted concisely by Keen et al. (2022), data management of photo-ID projects is a complex, arduous, and time-intensive process that hinges upon rigorous organisation and customised database software to remain viable in the long term. While current advances in computer vision and deep learning techniques have advanced the proficiency of individual identification methods at times beyond the capabilities of a human observer, a greater integration of multi-layered versatile data-management tools into current individual identification systems is much needed and ripe for further developments. Existing database management systems can offer several useful suggestions. We encourage such developments as important next step that would significantly benefit individual-based field studies.

Appendix

The photo-ID database management system DISCOVERY was conceived in the mid-2000s, launched at the 19th Biennial Conference on the Biology of Marine Mammals in Tampa, Florida, USA in November 2011, and released online early the following year (Gailey and Karczmarski 2012). The principal programmer, G. Gailey, hosts the software on his website https://cetaecoresearch.com/research-software-discovery.html and any enquiry regarding program support should be directed through that web platform. The software is also available through a different interface at https://www.cetacea-institute.org/discovery. While much of the work on DISCOVERY was done without any financial support, the later stage of the development and the completion of the system were supported with funding from South African National Research Foundation (NRF) and L. Karczmarski's professorial start-up funding at University of Hong Kong,[1] both of which are gratefully acknowledged. Table A1 and the paragraphs below summarise the key functions of the system. Visual presentation of several of these functions can be viewed online at https://www.cetacea-institute.org/other-projects and at https://www.youtube.com/watch?v=n2GCwGhEBvs. Full documentation and manual of DISCOVERY can be found at the web addresses indicated above.

Table A1 Overview of main functions of DISCOVERY

File — Tools to create/open databases and import/export data
• *Create/Open Database*
• *Open*: Viewing an image file or GPS data file
• *Import*: Importing image files, sighting information, databases
• *Export*: Exporting data into various file formats, e.g., MARK and SOCPROG
• *Update*: Updating database and image metadata
Incoming — Tools for processing and structuring files from the field for subsequent comparison to the database
• *Filenaming*: Renaming images and storing field information into EXIF data
• *Filtering*: Image editing and processing tools
• *Matching*: Matching within a group and select the best quality ID-images for subsequent matching to catalogue
Database — Tools to add individuals and other associated sighting information to the database
• *Match ID*: Comparing to the catalogue of identified individuals
• *Sightings*: Recording group/survey sighting information
• *Survey Effort*: Recording survey effort of each survey
• *Others*: Recording user-specified information
Manage — Tools to manage the database and verify the data entered
• *Database*: Viewing and editing database
• *Verify*: Confirming proposed new/matched images (and associated information) of individuals
• *Review*: Reviewing the verified catalogue of individuals
View — Tools to visualise information in the database
• *Individuals*: Summary of catalogued individuals with their associated image and sighting information
• *Map*: Plotting the geographic location of individuals/groups/surveys, and importing/exporting/editing ESRI shapefiles
Data summary — Tools to process, summarise, and plot information from the database
• *Charts*: Summary of database in graphs and tables
• *Discover R*: Platform to interface the database with program R for a range of analytical and graphical functions from pre-defined or user-defined scripts
Setup — Options to define user preferences and parameters

Processing of incoming data

DISCOVERY offers built-in tools to assist in processing images collected during field surveys before they are matched to a database. Each of these tools works independently from the database and from each other.

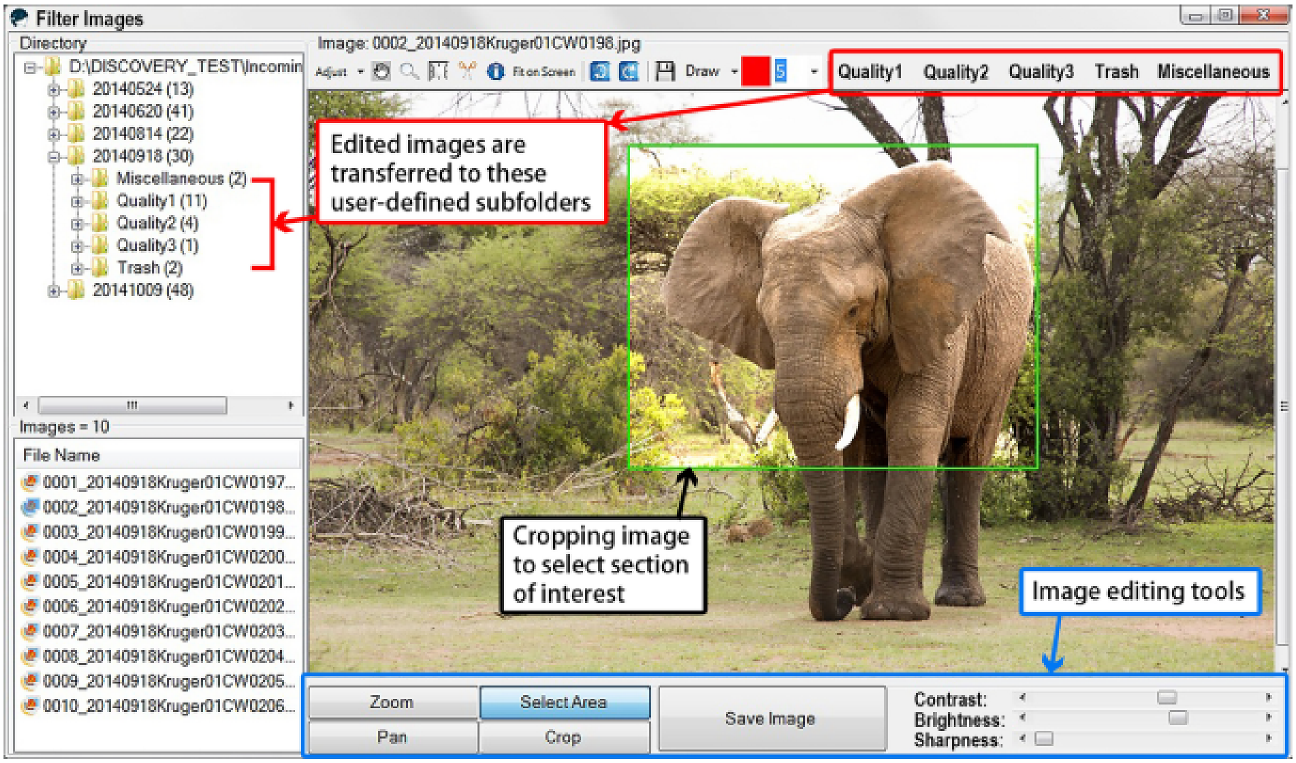


Fig. A1. A graphic summary of the filtering process of a field photograph of an African elephant (*Loxodonta africana*)

The filtering application allows to edit images by adjusting their contrast, brightness, and sharpness, and cropping images to select section(s) of interest (Fig. A1). Images can be transferred to user-specified subfolders to minimise the processing time at a later stage. After images have been filtered and transferred to a folder containing all identifiable individuals in a group/encounter, an incoming matching interface assists with finding the highest quality image of each individual in a group to later match this image to the existing catalogue.

Building database

Matching ID-images

Once the incoming images are processed, the individuals are ready to be matched to the existing catalogue of individuals in the database. Query images are compared to the "Type Specimen" images, which refer to the most representative images (the most recent and of the highest quality) of catalogued individual (Fig. A2). Only one type specimen may be assigned per aspect (e.g., left-side, right-side, fluke, etc.) of an individual, and has to be frequently updated throughout the study. Once an image is matched or determined to be a new individual, it is automatically moved to its ID-specific folder in the catalogue. At this time, various parameters can be added to the database, such as geographic location (automatically extracted from EXIF metadata of the image or imported from a GPX file), image quality, individual distinctiveness, descriptors, etc.

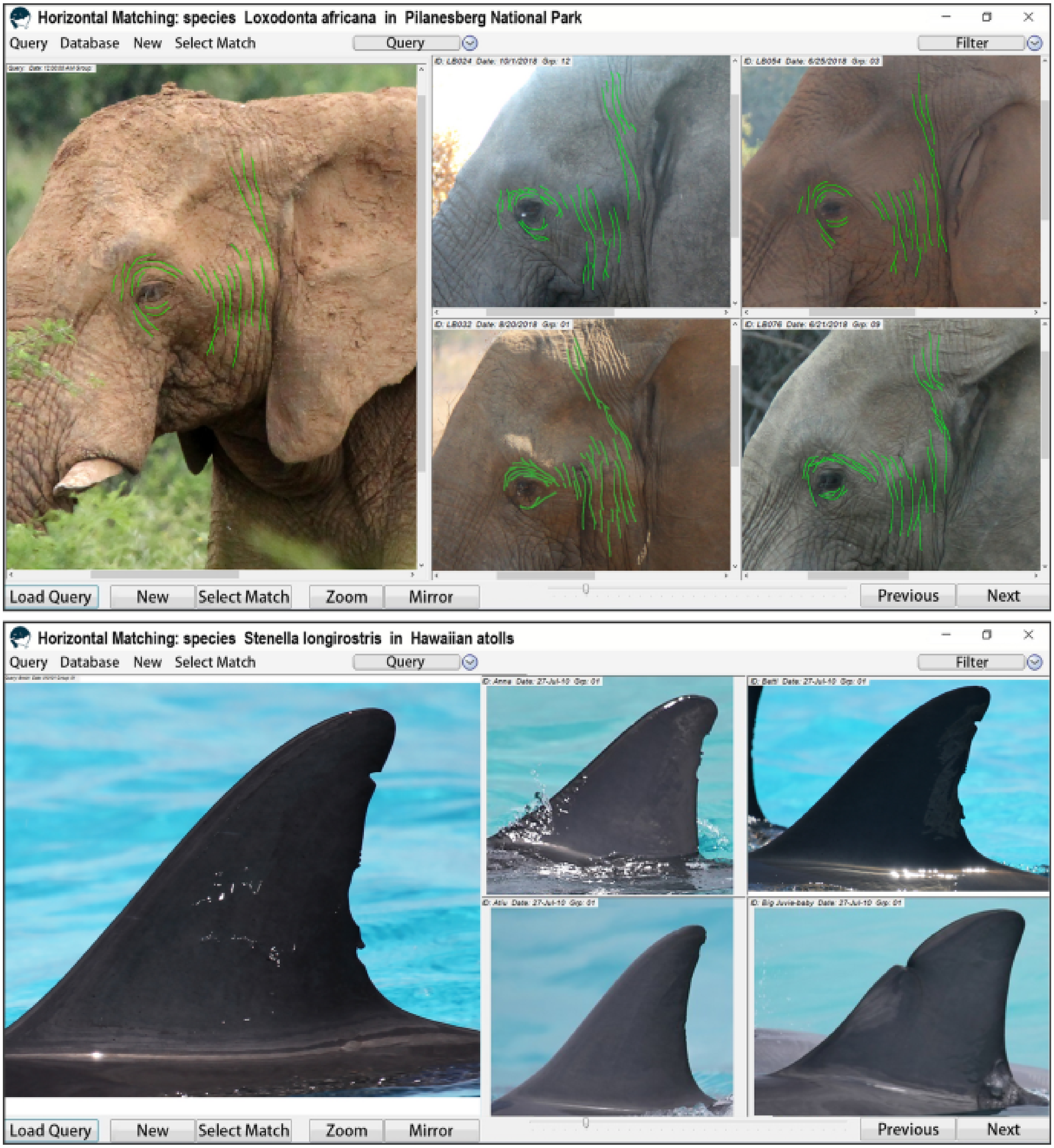


Fig. A2. Horizontal matching of (upper) facial wrinkles of African elephants (Chui and Karczmarski 2022) and (lower) Hawaiian spinner dolphins (Karczmarski et al. 2005) with a query image on the left and database images on the right. For species where a full flank of the animal's body has to be examined for individually distinctive features, as for example is the case with hyaenas, leopards, or gray whales (Gailey et al. 2020; Braczkowski et al.

2022; Portas et al. 2022; Schwarz et al. 2022; Spagnuolo et al. 2022) a vertically arranged matching window is also available; see Manual of DISCOVERY available at https://cetaecoresearch.com/research-software-discovery.html and https://www.cetacea-institute.org/discovery

As the number of individuals in a database increases, the time needed to find a potential match to a query image increases as well. DISCOVERY offers several catalogue filtering tools to facilitate the searching and matching of individuals, with customisable classification schemes that may for example prioritise 'aspects' (e.g., left-side/right-side), 'categories' (e.g., primary feature: number of notches), 'descriptors' (e.g., secondary features: scars, pigmentation, shape of tusks), individual distinctiveness, or other user-defined features. Morphometric topographies, such as positions of whisker spots, can also be applied if so user-defined; while for well-known individuals, an individual alpha-numeric ID can be used for a rapid search through the catalogue. However, as DISCOVERY does not offer automated image recognition, for large datasets, the matching process remains labour-intensive.

Sighting information

Other than photographic images of individual animals, a variety of information about each sighting (e.g., group size, age structure, predominant behaviour, geographic coordinates, and environmental data) are usually collected during photo-ID surveys. Some information is standard for photo-ID studies, while other parameters can be project-/species-/study- or site-specific. As such, DISCOVERY offers several standard features as well as dynamic user-defined features to create data forms for any specific needs (Fig. A3). The minimum standard information for a sighting includes study area, date, species, and group ID, which are used as a unique identifier for each sighting. Other standard sighting information is categorised as (i) *Group information*, with user-defined parameters and data types, which can include such details as number of adults/juveniles/calves, number of females/males, single-/multi-species group, etc., with a possibility of time-stamped multiple entries per sighting/group to facilitate recording changes in group membership over time; (ii) *Geographic positions*, which can be imported automatically from files downloaded from GPS devises (e.g., GPS Exchange Format, GPX files) or from Excel and Access files, or entered manually with optional multiple entries per sighting; (iii) *Environment data*, with user-defined parameter-specific data types, multiple entries per sighting, and full customisation of environmental parameters; and (iv) various other categories as described in the Manual of DISCOVERY available online at the web addresses indicated above. To minimise data entry errors, DISCOVERY offers an automatic error checking function that highlights any invalid entries (e.g., date or time).

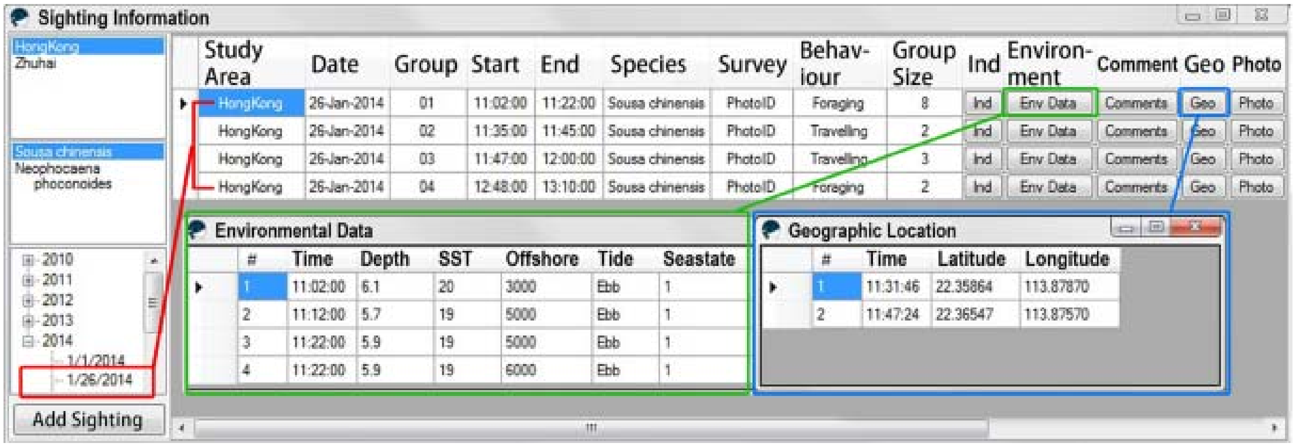


Fig. A3. An illustration of summarised sighting information in a centralised DISCOVERY database

Survey effort and other information

Survey effort recorded with GPS track data can be downloaded from a GPS device, generating a reference to the GPS/GPX file directory in the DISCOVERY database (Fig. A4). This serves as an overall record of the spatial and temporal effort for the number of surveys conducted for each study area and can be subsequently summarised and visualised. Various miscellaneous information related to the study or to a specific individual (e.g., newly recorded birth or death) but independent of the matching process or sighting/survey information can be entered as 'other info' and be associated with a specific sighting or survey for future data processing.

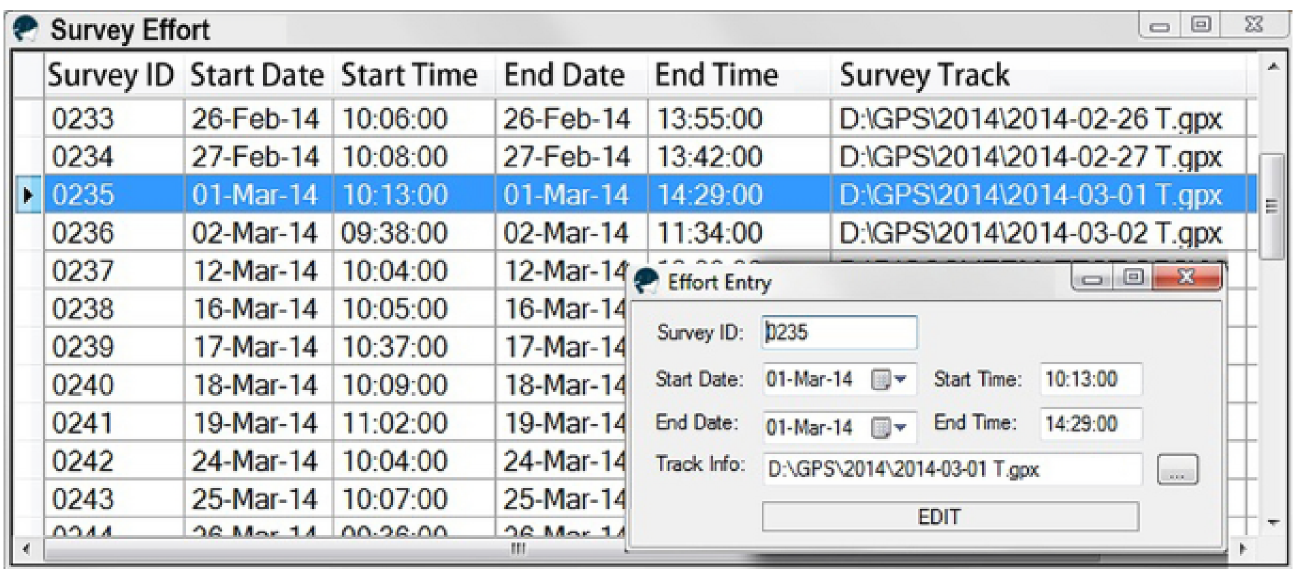


Fig. A4. An illustration of survey effort which can be recorded and viewed in a single table

Managing database

Viewing and editing database

DISCOVERY stores all data into a Microsoft Access database file and offers built-in options to visualise, edit, and manage the database. This function allows users who do not have Microsoft Access to partition, search, edit, and update the data directly without the need to open the database through the Access software (Fig. A5).

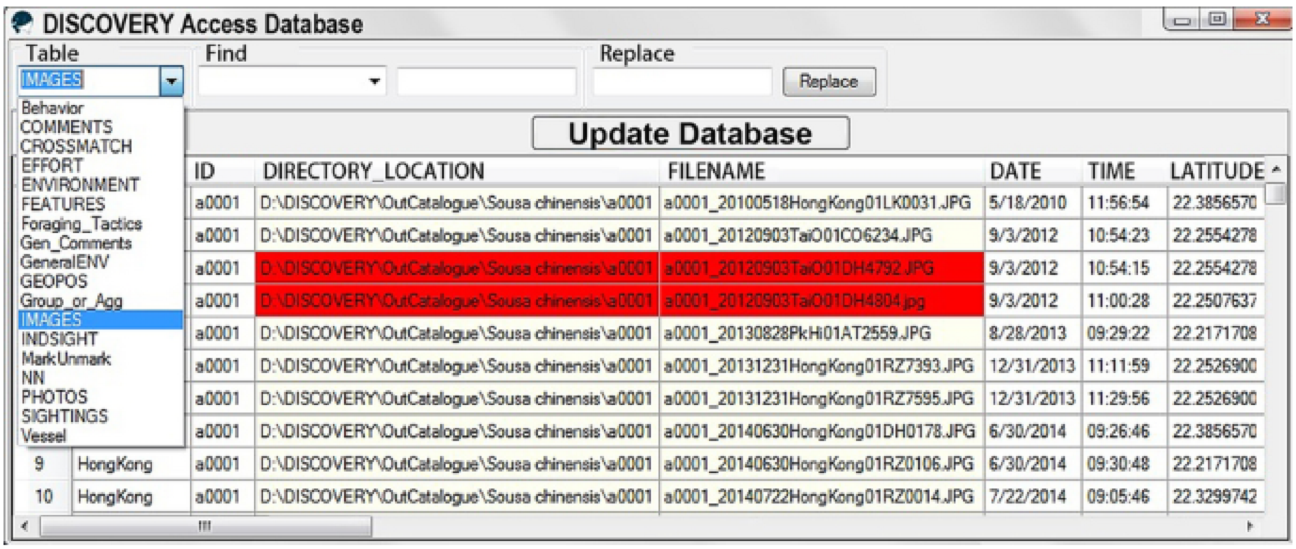


Fig. A5. An illustration of viewing and editing the Access database in DISCOVERY

Updating database and image metadata

Built-in functions provide options to quickly synchronise image-related information by updating the EXIF data of images and updating the GPS locations (latitude and longitude) of all catalogued image records in the DISCOVERY database.

Verification

The verification function is particularly handy for research teams with varying degree of experience among team members. To ensure that each individual-ID and relevant information (e.g., distinctiveness) are correct and have been correctly entered to the database, all newly catalogued images have to be double-checked and verified (usually by a project leader or database manager). An ID is entered into the database as "new individual" only if no images for that ID have previously been verified, or as a "matched individual" if at least one image for that ID has been previously verified.

While for some animals their primary identifiable features do not change in appearance whether viewed from the left or right sides, such as the pattern of notches on the dorsal fin of many dolphin species (e.g., Würsig and Würsig 1977; Read et al. 2003; Verborgh et al. 2022); for many other, such as elephants where identification of individuals is based on the pattern of natural markings on their membranous ears (e.g., Ardovini et al. 2008; de Silva et al. 2022) or the pattern of facial wrinkles (Whitehouse and Hall-Martin 2000; Chui and Karczmarski 2022), or the coat pattern of giraffes (Bolger et al. 2012; Muller 2018), or that of several species of felids (Oberosler et al. 2022; Pereira et al. 2022) and canids (Dorning and Harris 2019; Marneweck et al. 2022), or pigmentation and scars of some poorly marked cetaceans (Karczmarski and Cockcroft 1998; Elliser et al. 2022), these primary identifiable features display different patterns on the left and right sides of the animal. For that reason, DISCOVERY offers options to verify new IDs by either a particular aspect (Fig. A6) or several aspects combined (see the Manual for details).

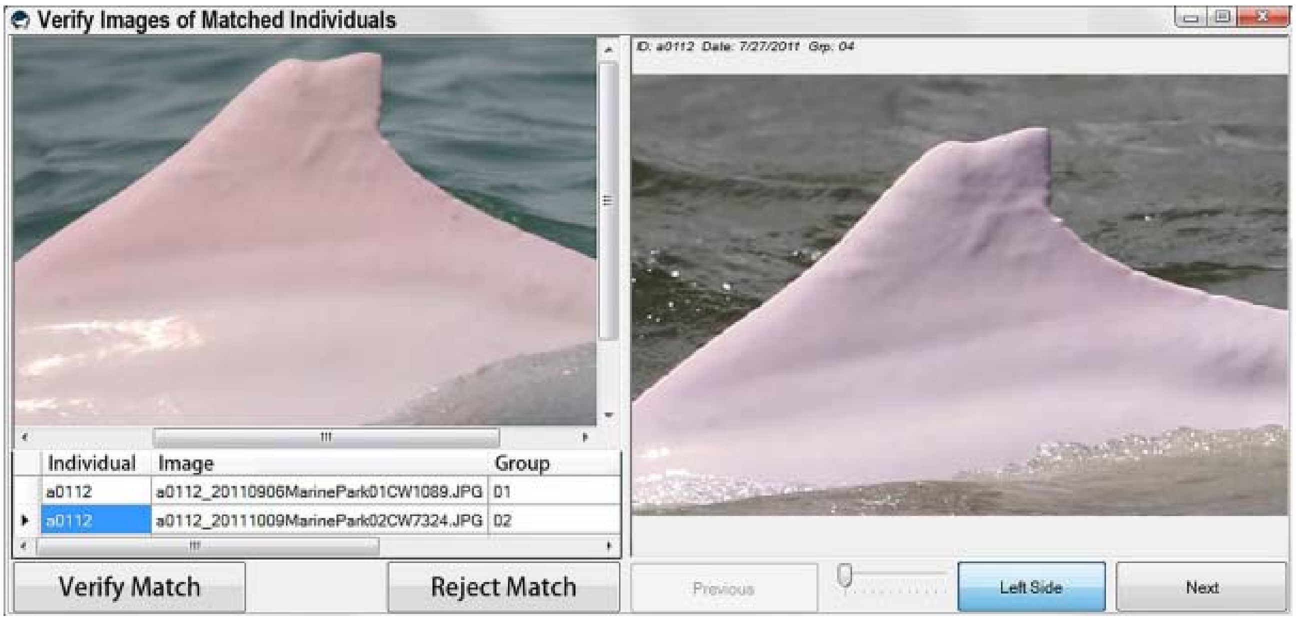


Fig. A6. An example of verifying images of matched individuals, in this case Indo-Pacific humpback dolphin (*Sousa chinensis*) in Hong Kong waters (Chan and Karczmarski 2017, 2019)

As every individual-ID catalogue has to be regularly reviewed and updated, DISCOVERY offers a review tool to examine the verified catalogue and update type specimen images to the most recent and highest quality representative image of that individual for each aspect.

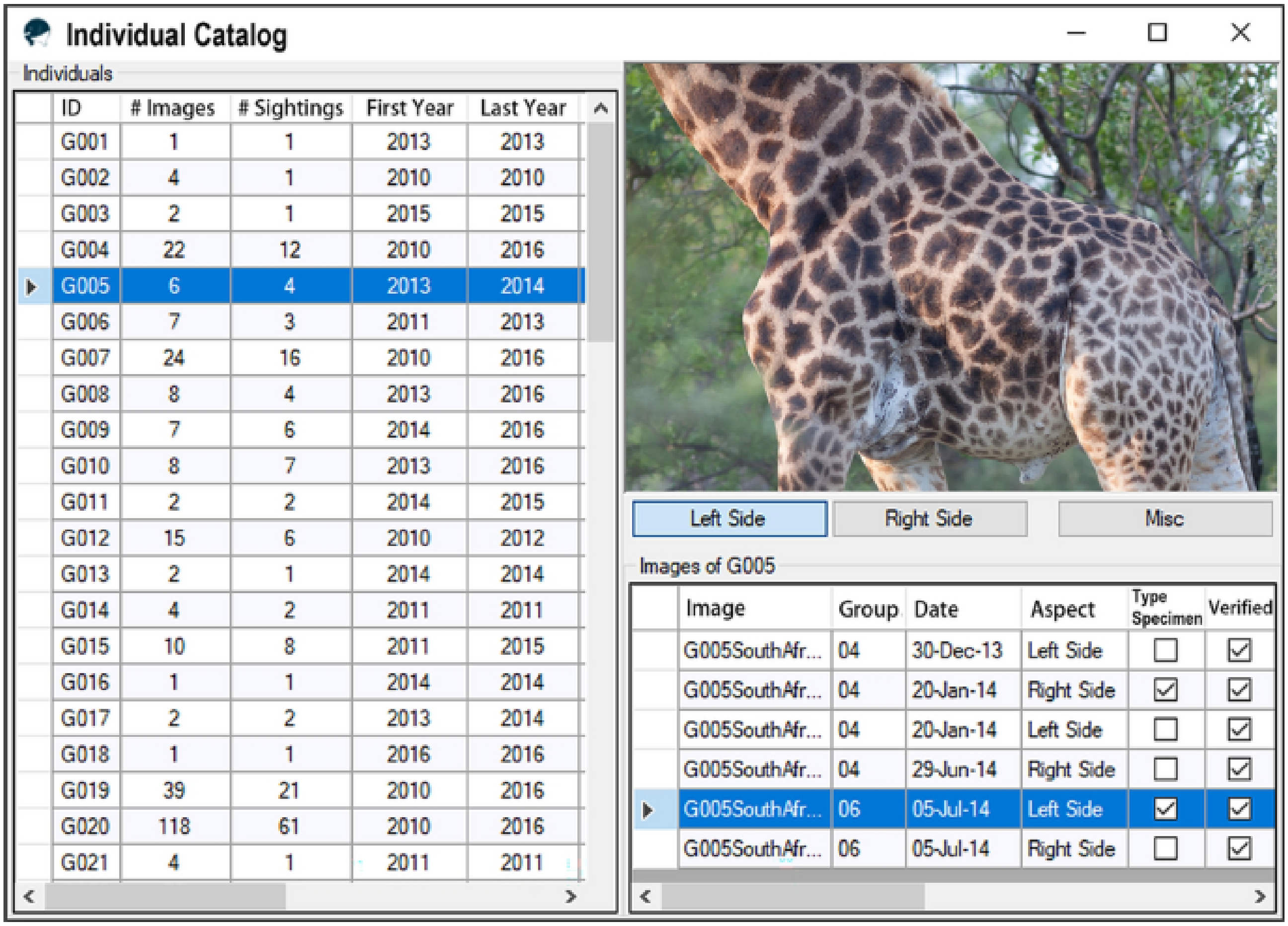
Crossmatching

DISCOVERY offers a crossmatching function to facilitate collaboration between research teams by centralising individual catalogues from different study areas. These catalogues are maintained as separate by assigning different *Study Area* in the initial Setup function (see the Manual for specific details). Each centralised catalogue can retain its own identification numbering system and be kept separately, while images from various catalogues can be compared with one another; a useful feature when working with species that exhibit long-range movements (e.g., Acevedo et al. 2022; Genov et al. 2022). However, as DISCOVERY does not offer automated image recognition, the crossmatching process is not as time-efficient as in recently developed AI-based matching systems (e.g., Cheeseman et al. 2022; Khan et al. 2022), especially when working with large datasets.

Database visualisation and other functions

Catalogue of individuals

DISCOVERY offers a quick overview of the individuals in the existing catalogue, as indicated in Fig. A7.



ID	# Images	# Sightings	First Year	Last Year
G001	1	1	2013	2013
G002	4	1	2010	2010
G003	2	1	2015	2015
G004	22	12	2010	2016
G005	6	4	2013	2014
G006	7	3	2011	2013
G007	24	16	2010	2016
G008	8	4	2013	2016
G009	7	6	2014	2016
G010	8	7	2013	2016
G011	2	2	2014	2015
G012	15	6	2010	2012
G013	2	1	2014	2014
G014	4	2	2011	2011
G015	10	8	2011	2015
G016	1	1	2014	2014
G017	2	2	2013	2014
G018	1	1	2016	2016
G019	39	21	2010	2016
G020	118	61	2010	2016
G021	4	1	2011	2011

Image	Group	Date	Aspect	Type Specimen	Verified
G005SouthAfr...	04	30-Dec-13	Left Side	☐	☑
G005SouthAfr...	04	20-Jan-14	Right Side	☑	☑
G005SouthAfr...	04	20-Jan-14	Left Side	☐	☑
G005SouthAfr...	04	29-Jun-14	Right Side	☐	☑
G005SouthAfr...	06	05-Jul-14	Left Side	☑	☑
G005SouthAfr...	06	05-Jul-14	Right Side	☐	☑

Fig. A7. Form illustrating *View Individuals* function in an existing catalogue of giraffes (*Giraffa camelopardalis*; L. Karczmarski, unpublished data) with the individuals listed in the left-hand table along with summary information (e.g., number of images, number of sightings, and first/last year of sighting), while all images of an individual along with the image details are displayed in the table at the bottom right

GIS viewer and shapefile export

This mapping function in DISCOVERY plots the geographically-referenced data stored in the database (e.g., locations of individual sighting records) along with sighting information and survey efforts as map layers, while the relevant fields from the data database are included in the attribute table of the shapefile. As in most GIS applications, specific features (e.g., point, line, or polygon data) can be selected by attributes to modify display settings of each layer, and export the layers as shapefiles for further spatial analyses.

Database summary

A built-in tool generates analytical summary of the database for each year of the study and the overall total, including for example the number of images in the database, the overall number of individuals, number of new individuals per year/season, re-sightings (min, max, and mean), percentage of re-sightings per aspects (e.g., left, right, fluke, etc.), group size (mean and range); as well as graphical display of summarised data such as cumulative discovery curve, number of sightings/re-sightings, individual sighting histories, etc. Frequency distribution of group sizes, predominant behaviour, species seen, etc. can be displayed as pie or frequency charts; and charts summarising the information are displayed in excel-like grids that can be copied/pasted to other applications to further processing and data display.

Interface with R

Discover R is a platform that interfaces the information in the DISCOVERY database with the freely available statistical software R, where a wide range of analytical and graphical functions can be used to suit the specific needs of different studies. For users who are not familiar with writing or modifying R scripts, there are a number of basic functions already built-in upon installation of DISCOVERY through the interactive interface in *Discover R*. For advanced users with previous experience in R, *Discover R* provides a platform to import any additional functions or to develop tailor-made functions.

Importing data

There are several options to import pre-existing database of ID-images and sighting information into DISCOVERY database. The system is sufficiently flexible to accept data by importing folders of images (and extracting information from EXIF metadata), Excel spreadsheets, Access databases, or other DISCOVERY databases.

Exporting data

The system offers tools to export data into Excel and other file formats that are ready for further use as input files into analytical tools, such as MARK (White and Burnham 1999), SOCPROG (Whitehead 2009), and various GIS applications (e.g., ArcGIS, QGIS; ESRI 2020; QGIS Development Team 2022).

The MARK export interface (Fig. A8) allows to pool data in user-defined capture occasions and specify the output file that meets the requirements of various population models (e.g., single-state closed/open, multi-state, or robust design models). INP files (.inp) are generated based on the images entered into the DISCOVERY database, with appropriate filtering of images performed when prompted at the time of generating the output file.

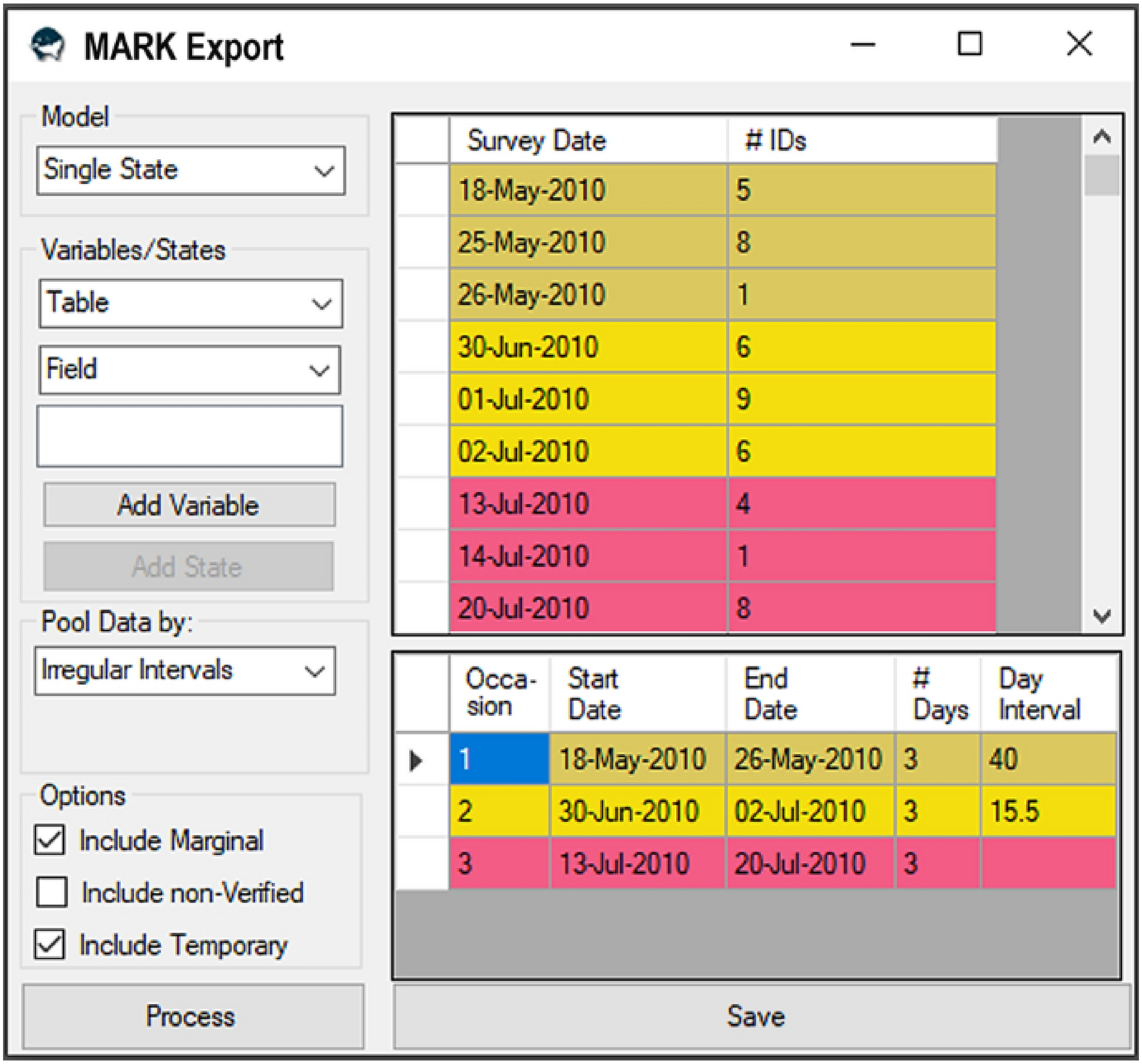


Fig. A8. An example illustrating pooling individual sighting histories (in survey days) by user-defined irregular intervals for MARK output files

The export interface for SOCPROG offers customisable output files that meet specific format of primary (sighting histories by individuals or by groups) and supplemental (individual features) input files required by SOCPROG. Microsoft Excel files are generated based on the images entered into the DISCOVERY database, with appropriate filtering of images performed when prompted at the time of generating the output file.

Notes

1. http://www.biosch.hku.hk/ecology/staffhp/lk/Discovery/index.html.

References

Acevedo J, Aguayo-Lobo A, Beeman P, Cheeseman T, Olavarría C (2022) From the Antarctic Peninsula to eastern Australia: the longest migration of a humpback whale through the South Pacific Ocean. Mamm Biol (Special Issue) 102(4). https://doi.org/10.1007/s42991-021-00195-2

Araujo G, Legaspi CGM, Ferber S, Murray R, Burdett K, Grundy S, Labaja J, Snow S, Yaptinchay A, Ponzo A (2019) In-water methods reveal population dynamics of a green turtle *Chelonia mydas* foraging aggregation in the Philippines. Endanger Species Res 40:207–218. https://doi.org/10.3354/esr00989

Ardovini A, Cinque L, Sangineto E (2008) Identifying elephant photos by multi-curve matching. Pattern Recogn 41:1867–1877. https://doi.org/10.1016/j.patcog.2007.11.010

Armstrong AO, Armstrong AJ, Bennett MB, Richardson AJ, Townsend KA, Dudgeon CL (2019) Photographic identification and citizen science combine to reveal long distance movements of individual reef manta rays *Mobula alfredi* along Australia's east coast. Mar Biodivers Rec 12:14. https://doi.org/10.1186/s41200-019-0173-6

Bendik NF, Morrison TA, Gluesenkamp AG, Sanders MS, O'Donnell LJ (2013) Computer-assisted photo identification outperforms visible implant elastomers in an endangered salamander, *Eurycea tonkawae*. PLoS ONE 8(3):e59424. https://doi.org/10.1371/journal.pone.0059424

Benjamins S, Dodd J, Thorburn J, Milway VA, Campbell R, Bailey DM (2018) Evaluating the potential of photo-identification as a monitoring tool for flapper skate (*Dipturus intermedius*). Aquatic Conserv 28(6):1360–1373. https://doi.org/10.1002/aqc.2937

Berger-Wolf TY, Rubenstein DI, Stewart CV, Holmberg JA, Parham J, Menon S (2017) Wildbook: Crowdsourcing, computer vision, and data science for conservation. arXiv:1710.08880v1 [cs.CY]. https://doi.org/10.48550/arXiv.1710.08880

Blount D, Gero S, Van Oast J, Parham J, Kingen C, Scheiner B, Stere T, Fisher M, Minton G, Khan C, Dulau V, Thompson J, Moskvyak O, Berger-Wolf T, Stewart CV, Holmberg J, Levenson JJ (2022) Flukebook: an open-source AI platform for cetacean photo identification. Mamm Biol (Special Issue) 102(3). https://doi.org/10.1007/s42991-021-00221-3

Bodesheim P, Blunk J, Körschens M, Brust C-A, Käding C, Denzler J (2022) Pre-trained models are not enough: active and lifelong learning is important for long-term visual monitoring of mammals in biodiversity research—Individual identification and attribute prediction with image features from deep neural networks and decoupled decision models applied to elephants and great apes. Mamm Biol (Special Issue) 102(3). https://doi.org/10.1007/s42991-022-00224-8

Bogucki R, Cygan M, Khan CB, Klimek M, Milczek JK, Mucha M (2018) Applying deep learning to right whale photo identification. Conserv Biol 33(3):676–684. https://doi.org/10.1111/cobi.13226

Bolger DT, Morrison TA, Vance B, Lee D, Farid H (2012) A computer-assisted system for photographic mark–recapture analysis. Methods Ecol Evol 3:813–822. https://doi.org/10.1111/j.2041-210X.2012.00212.x

Braczkowski A, Gopalaswamy AM, Fattebert J, Isoke S, Bezzina A, Maron M (2022) Spatially explicit population estimates of African leopards and spotted hyenas in the Queen Elizabeth Conservation Area of southwestern Uganda. Mamm Biol (Special Issue) 102(4). https://doi.org/10.1007/s42991-022-00324-5

Brust CA, Burghardt T, Groenenberg M, Käding C, Kühl H, Manguette ML, Denzler J (2017) Towards automated visual monitoring of individual gorillas in the wild. In: IEEE international conference on computer vision workshops (ICCVW), ICCV workshop on visual wildlife monitoring, pp 2820–2830. https://doi.org/10.1109/ICCVW.2017.333

Carlson CA, Mayo CA (1990) Changes in the ventral fluke pattern of the humpback whale (*Megaptera novaeangliae*), and its effect on matching; evaluation of its significance to photo-identification research. Rep Int Whal Commn (Special Issue 12):105–111

Chan SCY, Karczmarski L (2017) Indo-Pacific humpback dolphins (*Sousa chinensis*) in Hong Kong: Modelling demographic parameters with mark-recapture techniques. PLoS ONE 12(3):e0174029. https://doi.org/10.1371/journal.pone.0174029

Chan SCY, Karczmarski L (2019) Epidermal lesions and injuries of coastal dolphins as indicators of ecological health. EcoHealth 16(3):576–582. https://doi.org/10.1007/s10393-019-01428-0

Chan SCY, Chui SYS, Pretorius Y, Karczmarski L (2022a) Estimating population parameters of African elephants: a photographic mark-recapture application in a South African protected area. Mamm Biol (Special Issue) 102(4). https://doi.org/10.1007/s42991-022-00334-3

Chan SCY, Karczmarski L, Lin W, Zheng R, Ho Y-W, Guo L, Mo Y, Lee ATL, Or CKM, Wu Y (2022b) An unknown component of a well-known population: socio-demographic parameters of Indo-Pacific humpback dolphins (*Sousa chinensis*) at the western reaches of the Pearl River Delta region. Mamm Biol (Special Issue) 102(4). https://doi.org/10.1007/s42991-022-00335-2

Cheeseman T, Southerland K, Park J, Olio M, Flynn K, Calambokidis J, Jones L, Garrigue C, Jordán AF, Howard A, Reade W, Neilson J, Gabriele C, Clapham P (2022) Advanced image recognition: a fully automated, high-accuracy photo-identification matching system for humpback whales. Mamm Biol (Special Issue) 102(3). https://doi.org/10.1007/s42991-021-00180-9

Chui SYS, Karczmarski L (2022) Everyone matters: identification with facial wrinkles allows more accurate inference of elephant social dynamics. Mamm Biol (Special Issue) 102(3). https://doi.org/10.1007/s42991-022-00257-z

Clapham M, Miller E, Nguyen M, Darimont CT (2020) Automated facial recognition for wildlife that lack unique markings: a deep learning approach for brown bears. Ecol Evol 10:12883–12892. https://doi.org/10.1002/ece3.6840

Clapham M, Miller E, Nguyen M, Van Horn RC (2022) Multispecies facial detection for individual identification of wildlife: a case study across ursids. Mamm Biol (Special Issue) 102(3). https://doi.org/10.1007/s42991-021-00168-5

de Silva EM, Kumarasinghe P, Indrajith KKDAK, Pushpakumara TV, Vimukhi RDY, de Zoysa K, Gunawardana K, de Silva S (2022) Feasibility of using convolutional neural networks for individual-identification of wild Asian elephants. Mamm Biol (Special Issue) 102(3). https://doi.org/10.1007/s42991-021-00206-2

Desai B, Patel A, Patel V, Shah S, Raval MS, Ghosal R (2022) Identification of free-ranging mugger crocodiles by applying deep learning methods on UAV imagery. Ecol Informat 72:101874

Díaz-Calafat J, Ribas-Marqués E, Jaume-Ramis S, Martínez-Nuñez S, Sharapova A, Pinya S (2018) Individual unique colour patterns of the pronotum of *Rhynchophorus ferrugineus* (Coleoptera: Curculionidae) allow for photographic identification methods (PIM). J Asia-Pacific Entomol 21(2):519–526. https://doi.org/10.1016/j.aspen.2018.03.002

Dorning J, Harris S (2019) The challenges of recognising individuals with few distinguishing features: Identifying red foxes *Vulpes vulpes* from camera-trap photos. PLoS ONE 14(5):e0216531. https://doi.org/10.1371/journal.pone.0216531

Elliser CR, van der Linde K, MacIver K (2022) Adapting photo-identification methods to study poorly marked cetaceans: a case study for common dolphins and harbor porpoises. Mamm Biol (Special Issue) 102(3). https://doi.org/10.1007/s42991-021-00194-3

ESRI (2020) ArcGIS Desktop: Release 10.8. Redlands, CA: Environmental Systems Research Institute

Ferreira AC, Silva LR, Renna F, Brandl HB, Renoult JP, Farine DR, Covas R, Doutrelant C (2020) Deep learning-based methods for individual recognition in small birds. Methods Ecol Evol 11:1072–1085. https://doi.org/10.1111/2041-210X.13436

Gailey G, Sychenko O, Tyurneva O, Yakovlev Y, Vertyankin V, van der Wolf P, Drozdov K, Zhmaev I (2020) Effects of sea ice on growth rates of an endangered population of gray whales. Sci Rep 10:1553. https://doi.org/10.1038/s41598-020-58435-3

Gailey G, Karczmarski L (2012) DISCOVERY: Photo-identification data-management system for individually recognizable animals. Software and manual available at https://www.cetacea-institute.org/discovery and https://cetaecoresearch.com/research-software-discovery.html

Gatto CR, Rotger A, Robinson NJ, Tomillo PS (2018) A novel method for photo-identification of sea turtles using scale patterns on the front flippers. J Exper Mar Biol Ecol 506:18–24. https://doi.org/10.1016/j.jembe.2018.05.007

Genov T, Železnik J, Bruno C, Ascheri D, Fontanesi E, Blasi MF (2022) The longest recorded movement of an inshore common bottlenose dolphin (*Tursiops truncatus*). Mamm Biol (Special Issue) 102(4). https://doi.org/10.1007/s42991-022-00316-5

Goodall J (1971) In the shadow of man. Houghton Mifflin, Boston

Goodall J (1986) The chimpanzees of Gombe: Patterns of behaviour. Harvard University Press, Boston

Hammond PS, Mizroch SA, Donovan GP (1990) Individual recognition of cetaceans: use of photo-identification and other techniques to estimate population parameters. Rep Int Whal Commn (Special Issue 12)

Hiby L, Lovell P (1990) Computer aided matching of natural markings: a prototype system for grey seals. Rep Int Whal Commn (Special Issue 12):57–61

Holmberg J, Norman B, Arzoumanian Z (2009) Estimating population size, structure, and residency time for whale sharks *Rhincodon typus* through collaborative photo-identification. Endanger Species Res 7:39–53. https://doi.org/10.3354/esr00186

Hughes B, Burghardt T (2017) Automated visual fin identification of individual great white sharks. Int J Comput Vis 122(3):542–557. https://doi.org/10.1007/s11263-016-0961-y

Käding C, Rodner E, Freytag A, Denzler J (2016) Watch, ask, learn, and improve: a lifelong learning cycle for visual recognition. In: European Symposium on Artificial Neural Networks (ESANN), pp 381–386. https://www.esann.org/sites/default/files/proceedings/legacy/es2016-91.pdf

Karczmarski L, Cockcroft VG (1998) Matrix photo-identification technique applied in studies of free-ranging bottlenose and humpback dolphins. Aquat Mamm 24:143–147

Karczmarski L, Würsig B, Gailey G, Larson KW, Vanderlip C (2005) Spinner dolphins in a remote Hawaiian atoll: Social grouping and population structure. Behav Ecol 16(4):675–685. https://doi.org/10.1093/beheco/ari028

Karczmarski L, Chan SCY, Rubenstein DI, Chui SYS, Cameron EZ (2022a) Individual identification and photographic techniques in mammalian ecological and behavioural research – Part 1: Methods and concepts. Mamm Biol (Special Issue) 102(3). https://link.springer.com/journal/42991/volumes-and-issues/102-3

Karczmarski L, Chan SCY, Chui SYS, Cameron EZ (2022b) Individual identification and photographic techniques in mammalian ecological and behavioural research – Part 2: Field studies and applications. Mamm Biol (Special Issue) 102(4). https://link.springer.com/journal/42991/volumes-and-issues/102-4

Keen EM, Wren J, O'Mahony É, Wray J (2022) catRlog: a photo-identification project management system based in R. Mamm Biol (Special Issue) 102(3). https://doi.org/10.1007/s42991-021-00158-7

Kelly MJ (2001) Computer-aided photograph matching in studies using individual identification: An example from Serengeti cheetahs. J Mammalogy 82(2):440–449

Khan C, Blount D, Parham J, Holmberg J, Hamilton P, Charlton C, Christiansen F, Johnston D, Rayment W, Dawson S, Vermeulen E, Rowntree V, Groch K, Levenson JJ, Bogucki R (2022) Artificial intelligence for right whale photo identification: from data science competition to worldwide collaboration. Mamm Biol (Special Issue) 102(3). https://doi.org/10.1007/s42991-022-00253-3

Koivuniemi M, Kurkilahti M, Niemi M, Auttila M, Kunnasranta M (2019) A mark–recapture approach for estimating population size of the endangered ringed seal (*Phoca hispida saimensis*). PLoS ONE 14:e0214269. https://doi.org/10.1371/journal.pone.0214269

Lewis R, Dawson S, Rayment W (2020) Estimating population parameters of broadnose sevengill sharks (*Notorynchus cepedianus*) using photo identification capture-recapture. J Fish Biol 97(4):987–995. https://doi.org/10.1111/jfb.14453

Lorenz K (1935) Der Kumpan in der Umwelt des Vogels. Der Artgenosse als auslösendes Moment sozialer Verhaltungsweisen [The companion in the bird's world. The fellow-member of the species as releasing factor of social behavior]. Journ für Ornith 83:137–213. https://doi.org/10.1007/BF01905355

Lorenz K (1937) The companion in the bird's world. Auk 54(1):245–273. http://klha.at/papers/1937-BirdCompanion.pdf

Maglietta R, Renò V, Caccioppoli R, Seller E, Bellomo S, Santacesaria FC, Colella R, Cipriano G, Stella E, Hartman K, Fanizza C, Dimauro G, Carlucci R (2020) Convolutional neural networks for Risso's dolphins identification. IEEE Access 8:80195–80206. https://doi.org/10.1109/ACCESS.2020.2990427

Marneweck DG, Druce DJ, Cromsigt JPGM, le Roux E, Somers MJ (2022) The relative role of intrinsic and extrinsic drivers in regulating population change and survival of African wild dogs (*Lycaon pictus*). Mamm Biol (Special Issue) 102(4). https://doi.org/10.1007/s42991-022-00281-z

Mettouris O, Megremis G, Giokas S (2016) A newt does not change its spots: using pattern mapping for the identification of individuals in large populations of newt species. Ecol Res 31(3):483–489. https://doi.org/10.1007/s11284-016-1346-y

Miao Z, Gaynor KM, Wang J, Liu Z, Muellerklein O, Norouzzadeh MS, McInturff A, Bowie RCK, Nathan R, Yu SX, Getz WM (2019) Insights and approaches using deep learning to classify wildlife. Sci Rep 9:8137. https://doi.org/10.1038/s41598-019-44565-w

Miele V, Dussert G, Spataro B, Chamaille-Jammes S, Allaine D, Bonenfant C (2021) Revisiting animal photo-identification using deep metric learning and network analysis. Methods Ecol Evol 12:863–873. https://doi.org/10.1111/2041-210X.13577

Mizroch SA, Beard JA, Lynde M (1990) Computer assisted photo-identification of humpback whales. Rep Int Whal Commn (Special Issue 12):63–70

Morrison TA, Keinath D, Estes-Zumpf W, Crall JP, Stewart CV (2016) Individual identification of the endangered Wyoming toad *Anaxyrus baxteri* and implications for monitoring species recovery. J Herpetol 50(1):44–49. https://doi.org/10.1670/14-155

Muller Z (2018) Population structure of giraffes is affected by management in the Great Rift Valley. Kenya. PLoS ONE. 13(1):e0189678. https://doi.org/10.1371/journal.pone.0189678

Oberosler V, Tenan S, Groff C, Krofel M, Augugliaro C, Munkhtsog B, Rovero F (2022) First spatially-explicit density estimate for a snow leopard population in the Altai Mountains. Biodivers Conserv 31:261–275. https://doi.org/10.1007/s10531-021-02333-1

Parham J, Stewart C, Crall J, Rubenstein D, Holmberg J, Berger-Wolf T (2018) An animal detection pipeline for identification. In: IEEE Winter Conf on Appl Comput Vis (WACV), pp 1075–1083. https://doi.org/10.1109/WACV.2018.00123

Patel NG, Das A (2020) Shot the spots: A reliable field method for individual identification of *Amolops formosus* (Anura, Ranidae). Herpetozoa 33:7–15. https://doi.org/10.3897/herpetozoa.33.e47279

Pereira KS, Gibson L, Biggs D, Samarasinghe D, Braczkowski AR (2022) Individual identification of large felids in field studies: common methods, challenges, and implications for conservation science. Front Ecol Evol 10:866403. https://doi.org/10.3389/fevo.2022.866403

Portas R, Wachter B, Beytell P, Uiseb KH, Melzheimer J, Edwards S (2022) Leopard *Panthera pardus* camera trap surveys in the arid environments of northern Namibia. Mamm Biol (Special Issue) 102(4). https://doi.org/10.1007/s42991-022-00237-3

QGIS Development Team (2022) QGIS Geographic Information System. QGIS Association. http://www.qgis.org

Read AJ, Urian KW, Wilson B, Waples DM (2003) Abundance of bottlenose dolphins in the bays, sounds, and estuaries of North Carolina. Mar Mamm Sci 19(1):59–73. https://doi.org/10.1111/j.1748-7692.2003.tb01092.x

Reed J, New L, Corkeron P, Harcourt P (2022) Multi-event modeling of true reproductive states of individual female right whales provides new insights into their decline. Front Mar Sci 9:994481. https://doi.org/10.3389/fmars.2022.994481

Renò V, Losapio G, Forenza F, Politi T, Stella E, Fanizza C, Hartman K, Carlucci R, Dimauro G, Maglietta R (2020) Combined color semantics and deep learning for the automatic detection of dolphin dorsal fins. Electronics 9:758. https://doi.org/10.3390/electronics9050758

Romiti F, Bissattini AM, Buono V, Cifarelli C, Della Rocca F, Eniang EA, Akani GC, Luiselli L, Superti V, Carpaneto GM, Vignoli L (2017) Photographic identification method (PIM) using natural body marks: A simple tool to make a long story short. Zool Anz 266:136–147. https://doi.org/10.1016/j.jcz.2016.11.006

Schneider S, Taylor GW, Linquist S, Kremer SC (2019) Past, present and future approaches using computer vision for animal re-identification from camera trap data. Methods Ecol Evol 10(4):461–470. https://doi.org/10.1111/2041-210X.13133

Schneider S, Taylor GW, Kremer SC (2022) Similarity learning networks for animal individual re-identification: an ecological perspective. Mamm Biol (Special Issue) 102(3). https://doi.org/10.1007/s42991-021-00215-1

Schwarz LK, Gailey G, Tyurneva O, Yakovlev Y, Sychenko O, van der Wolf P, Vertyankin VV (2022) Western gray whales on their summer feeding ground off Sakhalin Island in 2015: who is foraging where? Environ Monit Assess 194(Suppl 1):738. https://doi.org/10.1007/s10661-022-10022-x

Spagnuolo OSB, Lemerle MA, Holekamp KE, Wiesel I (2022) The value of individual identification in studies of free-living hyenas and aardwolves. Mamm Biol (Special Issue) 102(4). https://doi.org/10.1007/s42991-022-00309-4

Tabak MA, Norouzzadeh MS, Wolfson DW, Newton EJ, Boughton RK, Ivan JS, Odell EA, Newkirk ES, Corney RY, Stenglein J, Iannarilli F, Erb J, Brook RK, Davis AJ, Lewis J, Walsh DP, Beasley JC, VerCauteren KC, Clune J, Miller RS (2020) Improving the accessibility and transferability of machine learning algorithms for identification of animals in camera trap images: MLWIC2. Ecol Evol 10:10374–10383. https://doi.org/10.1002/ece3.6692

Tavares SB, Samarra FIP, Miller PJO (2017) A multilevel society of herring-eating killer whales indicates adaptation to prey characteristics. Bahav Biol 28(2):500–514. https://doi.org/10.1093/beheco/arw179

Thompson JW, Zero VH, Schwacke LH, Speakman TR, Quigley BM, Morey JS, McDonald TL (2022) finFindR: Automated recognition and identification of marine mammal dorsal fins using residual convolutional neural networks. Mar Mamm Sci 38:139–150. https://doi.org/10.1111/mms.12849

Treilibs CE, Pavey CR, Hutchinson MN, Bull CM (2016) Photographic identification of individuals of a free-ranging, small terrestrial vertebrate. Ecol Evol 6:800–809. https://doi.org/10.1002/ece3.1883

Turner JW, Robitaille AL, Bills PS, Holekamp KE (2021) Early-life relationships matter: social position during early life predicts fitness among female spotted hyenas. J Anim Ecol 90:183–196. https://doi.org/10.1111/1365-2656.13282

Verborgh P, Janssen EH, Esteban R, Gauffier P, Freitas L (2022) Proposing a framework for monitoring demographic parameters in local cetacean populations: the case of short-finned pilot whales in Madeira. Mamm Biol (Special Issue) 102(4). https://doi.org/10.1007/s42991-022-00266-y

Vidal M, Wolf N, Rosenberg B, Harris BP, Mathis A (2021) Perspectives on individual animal identification from biology and computer vision. Integr Comp Biol 61(3):900–916. https://doi.org/10.1093/icb/icab107

Wattegedera M, Silva D, Sooriyabandara C, Wimaladasa P, Siriwardena R, Piyasena M, Marasinghe RMSLRP, Hathurusinghe BM, Nilanthi RMR, Gunawardena S, Gunawardena S, Heshan Peiris H, Seneviratne P, Sendanayake PC, Dushmantha C, Chandrasena S, Gooneratne SS, Premaratne P, Wickremaratne S, Mahela M (2022) A multi-point identification approach for the recognition of individual leopards (*Panthera pardus kotiya*). Animals 12:660. https://doi.org/10.3390/ani12050660

Waye HL (2013) Can a tiger change its spots? A test of the stability of spot patterns for identification of individual tiger salamanders (*Ambystoma tigrinum*). Herpet Conserv Biol 8(2):419−425. https://digitalcommons.morris.umn.edu/cgi/viewcontent.cgi?article=1002&context=biology

White GC, Burnham KP (1999) Program MARK: survival estimation from populations of marked animals. Bird Study 46(S1):S120–S139. https://doi.org/10.1080/00063659909477239

Whitehead H (1990) Computer assisted individual identification of sperm whale flukes. Rep Int Whal Commn (Special Issue 12):71–77

Whitehead H (2009) SOCPROG programs: analysing animal social structures. Behav Ecol Sociobiol 63(5):765–778. https://doi.org/10.1007/s00265-008-0697-y

Whitehouse AM, Hall-Martin AJ (2000) Elephants in Addo Elephant National Park, South Africa: reconstruction of the population's history. Oryx 34:46–55. https://doi.org/10.1046/j.1365-3008.2000.00093.x

Würsig B, Würsig M (1977) The photographic determination of group size, composition, and stability of coastal porpoises (*Tursiops truncatus*). Science 198(4318):755–756. https://doi.org/10.1126/science.198.4318.755