Automated photo-id for seals

Contents

Introduction

On the CD you will find a file called "ExtractCompare_install.zip". Open the zip file and double-click on "Setup.exe" to install the ExtractCompare program. You will also find a folder called "seal_demo". Copy this folder to your C: drive. It contains a number of folders and files, one of which, "seal_demo.mdb", is an Access database. Provided you have a 2003 or later version of Access on your PC, double-clicking "seal_demo.mdb" will open the database (if you have an earlier version we can supply a converted copy of the database).

"seal_demo.mdb" and the other files and folders in the "seal_demo" folder constitute the demonstration catalogue. The ExtractCompare program is a tool that works with the database to reduce the number of visual image comparisons required to identify matches. It can be run from the Programs list or from a button on the blue screen that appears when the database file is opened. Before the ExtractCompare program is run make sure that the two textboxes at the bottom of that screen contain the correct paths i.e. that the path to local files is set to "c:\seal_demo\" and path to software is set to "c:\program files\ExtractCompare\". (Note: if setting the software up on a drive other than the C: drive, simply change the letter in each case as appropriate.)

The database screen is designed to provide access to the rest of the system: entering new images, reviewing those already in the catalogue, eliminating substandard images, retrieving "encounter histories" and so on. Editing, renaming or deleting any other files in the "\seal_demo" folder will corrupt the system and crash the program. The only exception is the "download_photos" subfolder which is provided as a convenient temporary location for new images that are to be added to the system. Because the system is in a state of continuous development the database itself is fragile, providing far more scope for the user to change values than is conventional. It should therefore be backed up very frequently so that, without losing much work, it is possible to return to the backup copy following any problem with the system. In the event of an error send a screen dump of the screen at the time of the error as a first step. If it is not possible to resolve the error in that way it may then be necessary to send a copy of the whole "seal_demo" folder.

After the initial period of becoming familiar with the system the user will want to add images and start using the system to search for matches. There may well be an existing catalogue of images of known seals. In that case there will be data relating to those images such as where and when they were obtained and the current names (IDs) assigned to the seals they show. If that data is available in spreadsheets or database tables it can be imported into the seal database to avoid the need to enter images and associated data individually. The images must then be available as JPEG files with names that match exactly the image names used in the data records and are unique throughout the catalogue. Thus if the images are currently available only as prints or negatives the first step is to digitise them, assigning the correct name to each. If data is currently available only in written form the images and associated data should be entered individually using the "Enter Sightings" form. That form should also be used in the future to enter new images and their associated data as they become available.

Although the physical catalogue consists of images, the database is designed to generate a record of encounters rather than images or animals. An "encounter" is the event that a seal has been at a certain place at a certain time. Although in the case of camera traps an encounter is known to have occurred only by the existence of images, in general an encounter could be recorded simply by someone having seen and perhaps recognised a seal without any images having been obtained. The images of an encounter are just one of its attributes, helping to establish its ID, which is another of its attributes. The encounter's ID establishes which individual the encounter was with in the same way as the recorded date and location establish when and where the encounter occurred. Although the number of encounters is known the number of different seals those encounters were with is never known exactly and lies within a range that depends on the number of seals known only from one side and the risk of failing to match images from the same side of different encounters with the same seal. The best we can do is to infer the number of seals from the frequency of each "encounter history" (analogous to but not the same as the "capture history" resulting from a mark/recapture experiment), allowing for the fact that a seal can generate more than one of those histories.

Each encounter has to be assigned an ID when it is entered into the database, either a default value generated by the Enter Sightings form or a more meaningful name provided by the user. The function of the automated system is to use the images of that encounter to update its ID to one assigned to a previous encounter and hence establish those encounter histories.

Automating the processing of photo-id data is inherently complicated and it is necessary to be quite pedantic in the following sections about the difference between sightings, encounters, photos, images and pattern cells. However the first "Getting started" section provides a brief run-through of the process of extracting a pattern from a new image and using it to find a match and hence update the ID of the new encounter to the previous ID. Please follow the steps in the "Getting started" section to check that the software installation has been successful and observe the system working before challenging the system with new images. It should be possible to use the system without having to read through the more detailed sections that are provided for reference, for example to check on what a control that is not mentioned in the "Getting started" section does or looking for a solution in case things go wrong.

Please note that the system is "automated" only in the use of algorithms to generate similarity scores between pattern extracts. The extraction of those patterns from images is via a user interface designed to capture the user's ability to interpret the image as showing a seal from a given angle, lying in a given posture, partly obscured by some vegetation or another seal, with some of the apparent pattern being the result of dirt or ruffled fur rather than persistent pelage colouration or scarring. When introducing the system to a new user it is important to allow some time for training in order to maintain consistency at the pattern extraction stage.

Getting started

Video tutorials.

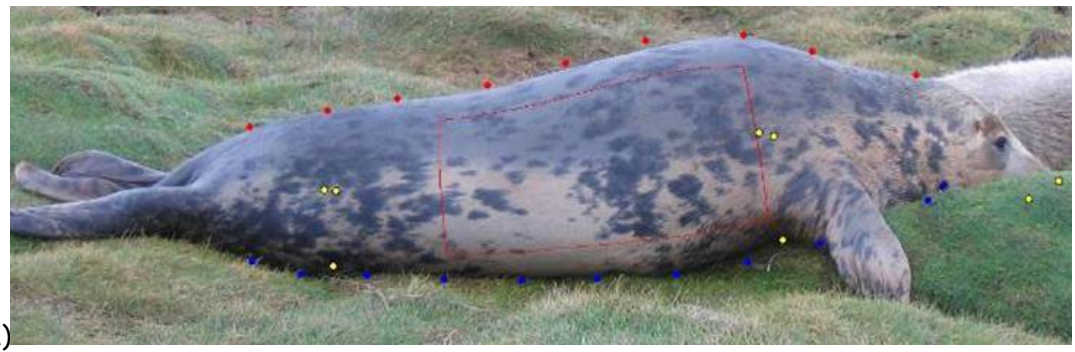
The example below describes in detail how to take a pattern extract from the flank region. To see video tutorials of extracts being taken from each aspect click on neck, flank, chest or abdomen. In the special case of head extracts click on land or water to see pattern extracts being taken from the head in those scenarios.

By following the steps described below the user can see how to extract a pattern from an image that has just been added to the system, how it is compared to extracts already incorporated into the "library", and how to confirm a match. Finally the user is invited to add a new image to the system and use it to search the library for a match.

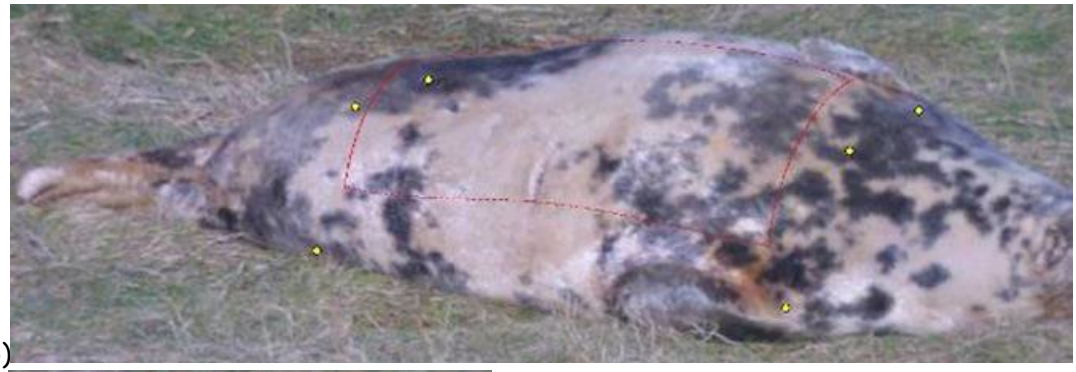
Step1 – immediate search for a match to a new image.

Open the seal_demo.mdb database and click the "View encounters/images" button. Click at the left hand end of a row to display an image. The data along the row show the ID that was originally assigned to the encounter with the seal in the image, its current ID, age and gender and the stage of any pup it was with. The "image" value is the name of the image and is assigned as the name of the photograph followed by "pxy" when the image is entered via the "sighting" form. x and y are rough coordinates for the head of the seal: x tenths in from the left of the photo and y tenths down from the top. In case there are more than one seal images on the photo that identifies which of the images the data refers to.

When the user entered the image he/she decided which "aspects" of the seal image displayed pattern suitable for individual identification: "flank" (Figure 1a, rearwards from the flippers to the pelvis); "abdomen" (Figure 1b, ventral surface rearwards from the flippers to the pelvis); "chest" (Figure 1c, forwards from the flippers to the narrowest part of the neck); "neck" (Figure 1d, forwards from the flippers to the ear).

(a) 

(b)



(c)

(d)

Figure 1. Examples of aspects used for pattern extraction showing the flank(a), abdomen(b), chest(c) and neck(d).

To see which aspects were selected for a given image and try to find the matching animal from the catalogue click the "reset" column on one of the rows. A relatively easy "flank" view is on row 9 of the form, which will be ordered by "orig_ID". The row is for image "05NR\215_1546p17" of a seal originally assigned an ID "34290_P" and since identified as an image of "Curly72". By clicking "reset" you will flag this image as needing an extract to be taken and change the ID to "X_Curly72" so that the comparison program will compare the extract with all the others, including other extracts of "Curly72". Close the database and start ExtractCompare.exe via shortcut or programs list. If the default location of the "seal_demo.mdb" database file at the top of the first screen is not correct then browse to the correct location. Then pick the "real time ident." option and click the "Start" button.

The selected seal image is displayed on the screen. A label near the top of the screen gives the image name with an extension indicating from which aspect the pattern is to be extracted. For example if it's from the flank the extension will be "_FL". The current ID assigned to the encounter of which this is one image follows the image name.

Click the "Retrieve" button (which is enabled because a pattern from this image was extracted previously). Note the yellow dots showing where we put the "special points" (starting with "pelvis" and "nose", in that order, for flank and abdomen patterns and "mid flipper" and "nose" for chest and neck patterns). The red and blue dots trace the top and bottom edges of the body. Click "Initial Fit". A 3D surface model of the seal body is superimposed on the seal image in roughly the right place. Click "Fit" two or three times. The model changes its orientation and shape to fit to the edges of the seal

image. Click "Extract". A red line demarcates a region of the image and the pattern is scanned from that region and displayed in a small picturebox. Click "OK".

The pattern is displayed in three pictureboxes. The centre box temporarily shows a detrended version and then it and the rightmost box show a wavelets decomposition (the screen may go blank for a few seconds before the display is complete). Green circles appear over the leftmost picturebox at locations corresponding to the darkest areas of the wavelet decomposition. Click "place seeds" to generate up to 9 sets of triples from the circles pattern, which will be used to initiate the search for pattern correspondence in the algorithm that will next compare the newly extracted pattern to previous extracts stored in the "library". Click "save extract" to initiate that process.

A progress bar indicates progress. Finally the image from which we just extracted a pattern will appear at the bottom left of the screen and an image of the most likely matching seal will appear at the bottom right. If you clicked "reset" on row 9 that should be seal "Curly72". More images of that seal will appear at the top of the screen if you click the "more lib pics" button – click on any image to enlarge it at the bottom right.

Before clicking the "match" button click the arrow at the bottom of the vertical scrollbar to the right of the right hand picturebox. The next highest-scoring library animal will be displayed. Return to the highest-scoring library animal and if it is a match click the "Match" button and then the "confirm match of this animal to this animal ->" button. The scrollbar to the right of the right hand picturebox moves down to show the next highest-scoring library animal in case it's also a match. Click "next extract". You are shown the results of a second (type c) comparison algorithm that runs in case the first (type n) fails. If the highest-scoring animal is a match click the "Match" button. The second algorithm may well show the same highest-scoring library animal as the first and if it was already recorded as a match there's no need to click the "Match" button again but doing so will not cause any problems. Click "next extract" again. You are invited to add this image to the library. Add it (although in practise you might not want to retain too many images of the same animal in the library). Click "Next image". If another aspect was selected for this image as showing suitable pattern then the same image will be displayed again, this time for pattern extraction and comparison using that alternative aspect. Otherwise the program will stop and return to the database screen.

It may seem laborious to have to review the results generated by each algorithm comparing patterns extracted from each aspect. However, once a reasonably large sample of patterns has been extracted from each aspect the "standardise" and "combine" options can be checked to combine the individual results and present a single ranked list of potential matching animals to each new image.

Step2 – batch processing.

Back at the "View encounters/images" form uncheck the previous "reset" and check "reset" on one or more different rows to repeat the process described above and see where the "special point" and "edge" dots need to be placed to extract patterns from the different aspects. Also, instead of selecting the "real-time ident." option, try the

batch processing method by completing the selected extracts using the "batch extract" option. Close the program and run it again using the "batch compare" option to compare the just-extracted patterns to each other and the library. Finally run it again using the "visual confirm" option" to review the results and confirm the matches. Batch processing is convenient when a large batch of new images is to be compared to the library because the "batch compare" process can be left to run overnight if necessary.

Step3 - adding new photographs to the catalogue.

The c:\seal_demo\download_photos folder contains a "try_to_enter.jpg" photograph so you can try out the process of adding new photos to the catalogue. To enter a photograph you need to make a new "sighting" table entry to record where and when you encountered the seal, a new "encounter" table entry to record details of the seal and new entries in the "image" and "cells" tables to record the image name and the aspects from which you want patterns to be extracted. All that is accomplished using the "sightings" form.

Click the "Enter Sightings" button. Sighting 1 of 32 is shown. Move on to the third sighting using the "tape recorder" controls at the bottom of the screen. Two seals, currently called "Curly72" and "Robyn", were encountered in this third sighting at location "NR" on 4th October 2004 and those encounters are listed in the central "encounters in sighting" window. Highlighting a row in that window will show, in the lower "images of encounter" window, the names of any images taken of that encounter and the aspects selected for pattern extraction from each image. Aspects "chest" and "abdomen" were selected for the image of the Curly72 and "chest", "abdomen" and flank" for the image of Robyn. Highlighting a row in the lower window displays the image itself to the right of the windows.

Click the rightmost button of the "tape recorder" controls to start a new sighting. Select a "location" at the top of the screen. The date will default to today's date and a new sighting number will be generated (sighting numbers need not be sequential). Click on the camera icon at the bottom of the screen to display the photos stored in the \seal_demo\download_photos directory. The download_photos directory currently contains only the "try_to_enter" photo, in a subdirectory called "00XX". Select a gender in the first row of the "encounters in sighting" window. Highlight this encounter row by clicking at its left hand end. The row will go black and the ID will default to the next "encounter number". Select the image below as being of this encounter and click on the head of the seal in that image. A new record will appear in the "images of encounter" window linking this seal image to the highlighted encounter. The image name is the name of the photo followed by "pxy" where x and y are rough coordinates of the seal head in the photo – x tenths of the way across from the left edge and y tenths of the way down from the top. That identifies which seal in the photograph was linked to the encounter in case the photograph shows more than one seal image. A comment can also be added in the image row to identify which of the seal images in the photo is linked to the encounter – that comment will be displayed when the photo is presented for pattern extraction so that the correct seal image can be selected. By checking the "include last subdirectory in image name" checkbox you can have a subdirectory name

like "00XX" included in the image name, in case the existing photo names are unique only within the subdirectories.

A copy of the selected photograph will be saved in the "newpic" subfolder so that once all images in that photo have been linked to encounters the original photo can be removed from the download_photos folder. Just above the "images of encounter" window is an "add catalogue identifier and sighting number to saved photo file name?" checkbox. Checking that box before linking the image will prefix the name of the copy in "newpic" with the catalogue letter specified on the database screen and the current sighting number. That means that provided they are unique within the sighting any arbitrary names (for example those assigned automatically by a digital camera) can be used for the original photos (the "download_photos" folder could also be the digital camera currently attached via a USB link).

If the download folder had contained another image of the same encounter (for example an image taken from a different angle) it could be linked to that same encounter by clicking to the left of the encounter row again and clicking on that second image. But if the other image was of a different seal it would be linked to a different encounter.

Close the Enter Sightings form and open the "view images/cells" button. The new cell should appear with AutoMatch status "E".

Repeat the pattern extraction and comparison process described in step2 for this new image. Because extracts have not been taken previously from this image the "Retrieve" option will be disabled so you will need to position the "special point" yellow dots and the top and bottom margin red and blue dots before clicking "Initial Fit". At the "visual confirm" stage images of the existing IDs will still be presented in order of decreasing similarity with the images of the new IDs but high scores are unlikely because the new image is not a match for any of the existing images.

To delete the new sighting open the "Enter Sightings" form again, move to the new sighting, highlight the vertical bar at the far left and click the Delete key. You will be warned that records in related tables will be deleted. Those are the encounter, image and pattern cell records that you made and that are linked to this sighting. Click "OK".

Step4 - testing the performance of the comparison algorithms.

Whenever the "visual confirm" stage has been completed and existing matches confirmed the temporary table of similarity scores will be empty and the catalogue should be in a state in which its ability to find matches can be tested.

Click "Extract/Compare" and choose the "test" option. The program will advise you to back up the database, which should be done frequently in case of a crash during an extended run. Given the small size of the sample catalogue this test run should complete very quickly. The test code temporarily changes all the AutoMatch "L" codes to "P", runs the "batch compare" code and then ranks the results. A frequency histogram is displayed, showing the probability that a seal, currently represented by a single extract in the library, will be identified as a match to a new image if the similarity score

results are inspected down to rank r during the "visual confirm" procedure (for r from 1 to 10). With the current extracts the histogram shows a probability of 100% of confirming a match at rank 1 for the sample images. That probability will reduce as the catalogue size increases but is greatly increased if a seal is represented by more than one extract in the library.

Instead of inspecting to a predetermined rank r the user can continue to check the potential matches down to the bottom of the list or until a match is found. In that case using the automated system only reduces the time taken to find matches if they exist (because they will normally be at or near the top of the list), i.e. when the system has been used for long enough for a significant proportion of the population to be represented in the catalogue.

To see any pairs that failed to get rank 1 click the "View Test Results" button. The form shows every pair of images known to be of the same seal (or more accurately, of the same seal that has an ID that does not start with a number, because animals which have been assigned a name rather than a default number ID are usually known animals or animals selected for testing the system). It displays the images, their names, the seal ID and the rank. Click in the "rank" textbox and click the "Z>A" Access icon to order the pairs in decreasing rank order. Use the "tape recorder" controls to work through the pairs that failed to get rank 1. Pairs that fail usually contain at least one poor image.

Extracting the pattern from an image

The following section describes in more detail the process of extracting the patterns.

When you run the 'ExtractCompare' program, either from the database screen or the Programs list, you are presented with a stack of five option buttons at the left of the screen. Having selected one of those you can click a "Start" command button to run the program but check first that the path to your data files, displayed at the top of the screen, is correct. This example illustrates what happens when you choose the "batch extract" option to extract a pattern from a new image. The first "real time ident." option would immediately compare the new extract to the existing library whereas with the "batch extract" option any images waiting for pattern extraction are presented before the extracts are compared (to each other and the existing library). The advantage, when a number of images are to be processed, is that the comparison stage can be left to run unattended, for example overnight.

Step1 – selecting and cropping the correct image from the photo.

When you click on "Start" the first photo containing an image listed in the "seal_demo.mdb" database as waiting for pattern extraction is displayed on the screen. The name of that image will be displayed in a label at the top of the screen. The "pxy" part of the image name will indicate the position on the photo of the image from which the extract is to be taken (in case the photo contains more than one seal image). The head of the correct image will be x tenths of the way in from the left-hand edge of the photo and y tenths of the way down from the top edge. If a comment describing the location of the image on the photo was included when the image was entered it will also be displayed. In case the label obscures part of the image click on it to remove it.

In the examples below a pattern cell is to be extracted from an image selected by checking the "reset" box on the sixth row of the "View encounters/images" form. The ID currently assigned to the encounter of which this is an image is "Curly72". This will become "X_Curly72" when the "reset" box is ticked and you have exited the form. After exiting the "View encounters/images" form start the ExtractCompare program in real time ident. Mode. You will see that you are working with an image that has been named "04NR\IMG_8767p84_FL,ID: X_Curly72," (Figure 2). In this case it's the only seal image on the photo but if there were more we would know to select the one with its head about 8 tenths of the way in from the left-hand edge of the photo and 4 tenths down from the top edge. The "_FL" extension indicates that the pattern is to be extracted from the flank region.

In the ExtractCompare program ignore the fact that there is the option to retrieve a previous fit for "04NR\IMG_8767p84_FL,ID: X_Curly72," and work with this image as if it was newly entered. As indicated by the instruction at the bottom of the screen you can expand the required seal image by left-clicking to its upper-left and lower-right. If it's already big enough to work with comfortably in step 2 then right-click anywhere on the photo.

Figure 2. ExtractCompare program showing "04NR\IMG_8767p84_FL,ID: X_Curly72,".

Drag the yellow "pelvis" dot to the ventral midline level with the pelvis and the "nose" dot to tip of the nose (Figure 3). Further "special point" dots will then appear.

Figure 3. Example of where dots should be placed for a flank extract.

Step 2 – placing the "special points".

Envisage a line encircling the circumference of the animal's body level with the midpoint of where the fore flipper meets the trunk of the body. The following "flipper" points dots should be dragged to level with this line. The "mid flipper" point should be placed on the ventral midline while the "left flipper" point and "right flipper" point should be placed half way up the body. (Note: in the case of "chest" and "abdomen aspects these points are placed at the front of the fore flipper where the fore flipper meets the trunk of the body). The "left pelvis" and "right pelvis" points are half way up the body, level with the pelvis and the "chin" point at the tip of the lower jaw If you have any doubts as to where these points should be placed look at the examples given in these usage notes or alternatively go back to Step 1 in the Getting started section to view examples of where points have been placed previously.

In case of any confusion about which dot on the seal is supposed to locate which feature, right-clicking on any dot will identify it by highlighting one of the stack of special point labels (the stack can be dragged aside if it's obscuring the seal).

Step 3 – locating the margins.

You now need to mark the body margins. Click the "top" option button and use the cursor and left mouse button to place red dots at intervals along the upper margin of the body, from level with the pelvis to the neck (Figure 3) (Note: for "chest" or "neck" aspects top and bottom dots should start level with the fore flippers and end at the tip of the nose). Click the "bottom" option button and place blue dots at intervals along the lower margin. The position of any dot can be adjusted by dragging or, to place all the dots again, they can be removed by double clicking the relevant option button.

The order in which dots are placed does not matter in general because they are re-ordered across the screen by the program, however that process may fail if the image is aligned vertically rather then across the screen. In that case the model may "lock up" instead of fitting well to the marked margins. The solution is to place the dots in clockwise order around the image (thus from left to right for the upper dots and right to left for the lower dots) and uncheck the "Auto order" box.

Step 4 – fitting the model.

Click "Initial Fit". A 3d surface model will be superimposed on the image of the seal in roughly the right place (Figure 5).

Figure 5. Example of the 3d model being applied after clicking "Initial Fit".

Then click "Fit". The model should adjust its shape to fit closely to the edges of the body (Figure 6). Toggle the "Hide Model"/"Show Model" button to check on the fit of the model to the body edges, if necessary click "Fit" again. It may be necessary to adjust some of the margin dots to get a good fit – decrease the distance between successive dots to attract the model to that section of the edge. The model fit is a compromise between trying to get the model edges to run along the margins of the body and trying to get the model positions for the special points (the green dots) to lie close to where the user has put them on the screen – a red line shows the displacement between the two positions.

Figure 6. Example of 3d surface model after it has been properly fitted to the margins of the animal's body.

Step 5 – extracting the pattern.

Click "Extract". The program generates the pattern cell by reading the grey-shade values from the photo at an array of points defined by the overlying model to compensate for the orientation and shape of the body. The pattern cell is displayed in a moveable window and the margins of the array are indicated by the rectangle of red lines drawn on the seal (Figure 7).

An option is provided to adjust the position of the pattern extract if it seems too far forward/backward or high/low. An extra blue dot should be visible below the model. Moving the mouse cursor left and right will move that blue dot backwards and forwards over the surface of the model, moving the cursor up and down will move the dot around the model. The dot is moved by moving the mouse cursor over the screen, not by dragging the dot as was done to place the special points and margins. If the cursor is moved too far the dot may slide around to the rear of the model, which should sound an audible warning and cause the dot to move up and down in the opposite direction to the cursor.

Holding don the Shift key, move the blue dot to where you want the centre of the leading edge of the pattern cell to be. Click "Extract" again to reposition the leading edge. Release the Shift key and move the blue dot to where you want the centre of the trailing edge of the pattern cell to be. Click "Extract" again to reposition the trailing edge. Moving the blue dot rapidly may cause its colour to change from blue to grey but its functioning will not be effected.

Figure 7. The red rectangle on the flank of the animal indicates where the pattern extraction is taken from. The square window is a greyscale image of that pattern extraction.

Step 6 – erasing obscured areas.

To erase regions of the flank pattern that are occluded or obscured by dirt drag the cursor with the right mouse button held down. The slider at the centre top of the frame adjusts the size of the eraser. The erased areas on the image will be blue. Click "Extract" again. The erased areas within the pattern extract will have an average grey colour and be ignored by the pattern comparison algorithms. To remove a blue area that has been placed in error click on the eraser slider and press the "E" key while holding down the Shift key. An alternative method for erasing obscured areas is to erase parts of the pattern extract itself. Pressing "Extract" again will make any erasing done in this way disappear so that you may start again. Try both methods to see which one suits you best.

Step 7 - save.

When a satisfactory extract has been obtained click "OK" to save it and switch to the next screen (clicking "Scrap" at any stage will skip a photo from which a satisfactory extract can not be obtained and bring up the next photo from the "newpic" folder). Before clicking "OK", the "tag" checkbox can be checked to record the pattern as being reliable for identification, so that the seal is recorded as "tagged" via the pattern in this image. At the top of the next screen the extract will be displayed in the leftmost

of three windows (Figure 8). The central window displays a detrended version of the extract and the rightmost window displays a wavelets decomposition of the extract.

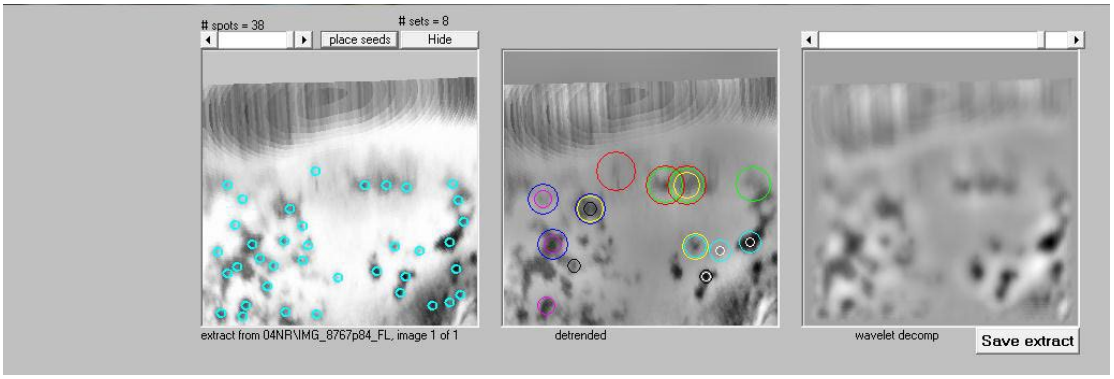


Figure 8. This is the screen displayed when a satisfactory extract has been taken. From left to right are the pattern extract, a detrended version of that extract and a wavelets decomposition version of that extract.

Click "place seeds" and then "save extract" to save the extract for subsequent comparison using the comparison algorithms.

Comparing patterns

The following section describes in more detail how the pattern extracts are compared and the resulting similarity scores used to guide visual comparison of images.

Batch compare

When a flank pattern is extracted the AutoMatch field in the database record is updated to "P" and files are created in the newlib and oldlib folders. When the ExtractCompare program is run using the "Batch compare" option those records with AutoMatch status "P" are selected and the corresponding library files compared to each other using type "n" and type "c" algorithms. Each file is also compared to the files corresponding to records with AutoMatch status "L", which are those that have been compared previously and are now in the library. Comparisons can also restricted by gender assignment but by default all extracts are compared.

All the resulting scores are stored in a "temp" table in the database. For example, if there are N extracts in the library and n extracts have been taken from new images using "Batch extract" then "Batch compare" will generate $n(n-1)/2 + nN$ scores. The program can be left unattended to calculate the comparison scores, a progressbar shows what stage the comparison process has reached for each comparison algorithm. A maximum of 200 new extracts can be compared in one run of "Batch compare" and they will be compared with successive batches of 200 library extracts, as indicated by the progressbar.

Visual confirm

Initially the "standardise" checkbox must be unchecked during the "Batch compare" run. The "combine" checkbox must then be unchecked during the "Visual confirm" run so that results will be presented separately for each combination of aspect and comparison algorithm (type "n" and type "c"). Once sufficient extracts have been compared summary statistics will have accumulated to allow the checkboxes to be checked and results presented for all aspects and algorithms combined.

At the "Visual confirm" stage the scores are collated by animal ID so that, for each existing ID, the maximum of the scores between extracts from images linked to encounters with that ID and the comparable extracts linked to the new ID is selected. An image of the new ID is shown at the bottom left of the screen and its maximum scores with existing IDs ranked so that an image of the most similar ID is shown for comparison at the bottom right (the highest-scoring images of the two IDs will be shown). Any alternative images of that most similar ID can be shown as a raster scan of up to eight thumbnails at the top of the screen and clicking on a thumbnail will show it enlarged at the bottom right, along with its name and date. If there are more than eight alternative images available the scan can be moved along to show the additional images. Using the checkbox on the initial ExtractCompare screen either cropped or

uncropped images can be displayed. Alternative images of the new ID can be displayed in the picturebox at the bottom left by using the vertical scrollbar to its left.

Scrolling down the vertical scrollbar to the right of the picturebox at the bottom right displays the next most similar ID, then the next and so on through all the existing IDs that got a score with the new one. The question of how far down the list it's worth searching for a match is covered in the section on testing, below.

If visual comparison of the images linked to the new ID with those linked to an existing ID shows that the two sets of images are of the same animal click the "Match" command button above the right hand picture box. Only click the "Confirm this match…" command button that then appears if you are absolutely sure that the two sets of images show the same animal, otherwise click "don't match". On confirming a match the new ID will be updated in the database to the matching ID and the next most similar ID displayed. That next ID may also be a match to the new one if a previous match was missed. This process can therefore lead to updates of existing IDs in the database to reflect the latest information about identities. For example, suppose two current IDs, 1000 and 2000, are actually the same seal. Suppose ID 2000 is the most similar to the new set of images that show that same seal, then as a result of the first match confirmation the ID of the new set will be updated to 2000. The set of images linked to ID 1000 is then also seen to match the new set and on confirming that second match ID 2000 will be updated to 1000. Because that ID will be updated wherever it occurs in the database the ID 2000, which may have existed for some time, will be no longer exist and all the images will have ID 1000 (i.e. they will all be linked to encounters with "1000" in the ID field).

By convention the temporary ID assigned to the new set of images is replaced by the ID of the matching set from the library except that an ID starting with a letter always takes precedence over one starting with a number. This is because the default IDs assigned when encounters are entered via the sighting form are based on the encounter number whereas if the encounter is with a known animal its existing name can be used as the encounter ID. Although these conventions attempt to preserve the most meaningful names the loss of certain IDs via the updating process may be undesirable. Therefore a "prev_ID" field is assigned a copy of the ID entered via the sighting form and it will not be updated, thus preserving a record of the name originally assigned to the new encounter.

Once the image (or set of images) linked to the new ID has been visually compared to enough existing IDs click the "next extract" command button above the left hand picture. Scores generated by the type "c" algorithm will present a second ranked list of existing IDs (unless, as explained above, the "standardise" and "combine" options are used, in which case the original ranked list of library IDs will already be based on both algorithms and all aspects). If the current new image (or set of new images) was not matched via either algorithm the new encounter's ID will remain at its current value and hence represent a seal that is new to the catalogue. The risk that those images are actually of a seal already in the catalogue because the match was missed depends on the number of existing and new images of that seal and on their quality. Software is available which can interface with the database to calculate that risk given the results of the test application described below and hence allow for that risk in estimating population size and survival.

The "test" option described below will provide guidance how many existing IDs in the ranked list of similarity scores it's worth comparing to visually before clicking the "next extract" button.

Add the latest images to the library?

When the "next extract" command button is clicked following inspection of the second algorithm's results the program provides the option to add the new image/s to the library, in which case the AutoMatch status of each extract is set to "L" and those extracts will be included in future automated comparisons, or not to add them, in which case the status is set to "R" and those extracts will not be included in future automated comparisons (the encounter record is retained in either case). Usually the images would be added, either because they have not been matched and may therefore represent an encounter with a new ID or because they have been matched and additional extracts linked to an existing ID reduce the risk that a match to a future new ID will be missed.

Testing

The following section describes in more detail the procedure for testing the performance of the automated system.

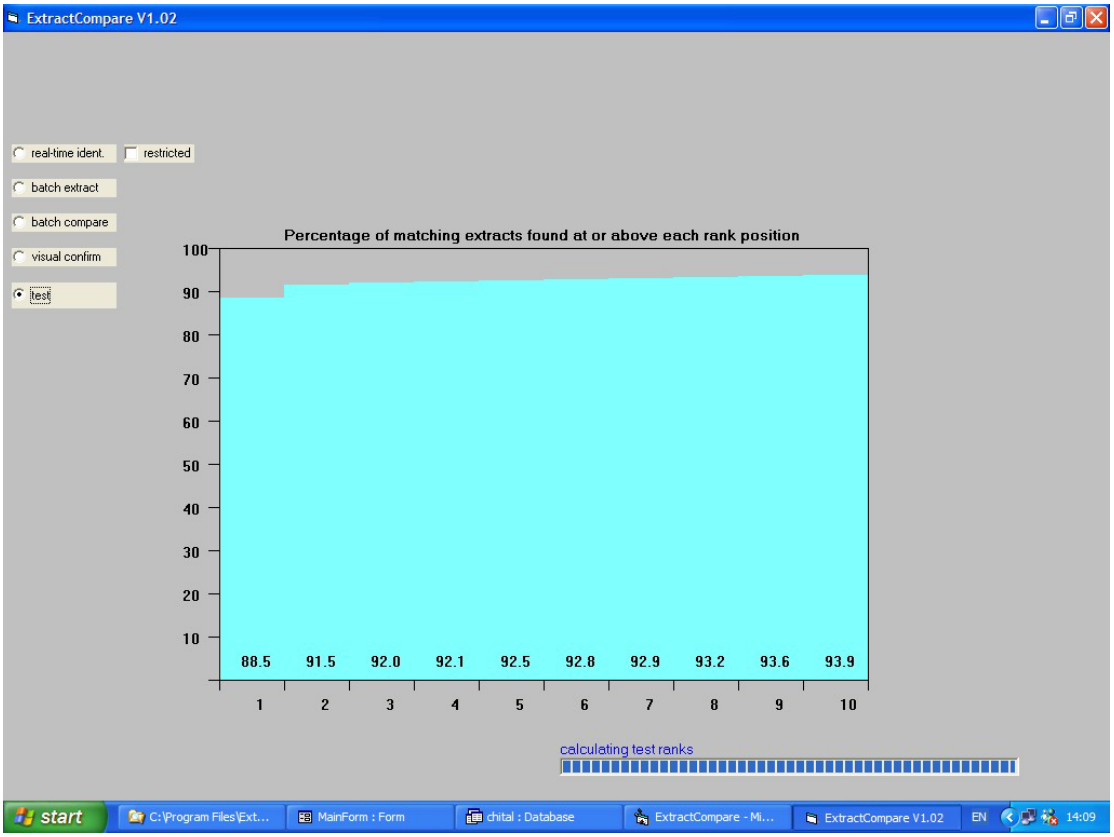
When images of existing IDs are displayed for visual comparison to images of a new ID the existing IDs are ordered by decreasing maximum similarity score. The user searches for a match to an existing ID down to a certain position in that ranked list. So if the maximum similarity score with images of an existing ID that is in fact a match to the new one (i.e. the two sets of images show encounters with the same seal) is at a lower position in the list the match will be missed and the new ID will not be updated.

The risk of missing a match depends on the position in the ranked list at which the visual comparisons are abandoned, the number of images of the existing matching ID and the new ID and the quality of those images. To start to evaluate the risk the "test" option in the ExtractCompare program provides an estimate of what that risk would be if a single image was available for the new ID and a single image was available for the existing matching ID. It estimates the risk for searches down to ranks 1 to 10. By filtering the test images included in the test it estimates how the risk is affected by image quality.

The raw data for the test is images that were known, prior to using the automated system, to show the same seal. n such images of the same seal yield n(n-1)/2 pairs, each of which provides a test of where the similarity score between that pair would fall in the ranked list of scores that each gets with the rest of the library. Suppose the library contains images of seals A to Z. Consider images a_1 to a_n of seal A. The similarity score between extracts from images a_1 and a_2 is compared with the maximum scores the extract from a_1 gets with extracts from images of seals B to Z. It's also

compared with the maximum scores the extract from a_2 gets with extracts from images of seals B to Z. That one pair of images thus yields to two rank positions. Combining the rank positions from all available image pairs provides a frequency histogram for rank position 1, rank position 1 or 2, rank position 1 or 2 or 3 and so on. Inspecting the histogram shows what percentage of matches would be missed by searching to rank position r if the new and existing ID of the seal were each represented by a single image. It suggests to what position, as a percentage of the size of the library, the search for a match should be continued in order to limit that risk to a given value. The shape of the histogram also shows from what rank position further search provides little further reduction in the risk. The histogram tends rapidly to an asymptote because the distribution of similarity scores between extracts from images of the same seal is bimodal: either a high score and thus a high rank position, or a score that is random selection from the distribution of non-matching scores that could therefore rank at any position.

The following illustration shows test results for chital (spotted deer) flank patterns. 88.5% of the test pairs achieved rank position 1, 91.5% rank position 1 or 2 and 92% rank position 1, 2 or 3. From then on the percentage increases only gradually suggesting search beyond rank position 3 would not be worthwhile for the type of images included in the test.



A test can be run whenever a "visual confirm" run is completed so that the temporary table holding similarity scores is empty. All extracts from images linked to encounters with IDs that do not start with a number and that currently have AutoMatch value "L" have the AutoMatch value changed temporarily to "P". A "batch compare" run of the ExtractCompare program is completed and the resulting similarity scores processed to

generate the pairwise rank scores and histogram. The method thus assumes that encounters with known seals have been assigned IDs that do not start with a letter, the encounters with other seals having the default numerical IDs. If that assumption does not hold run the test using a copy of the database with the IDs updated so that all test pairs will be derived from images of known seals. The test can be restricted to extracts of a certain type or of a certain quality by updating the AutoMatch value of other extracts to some other value such as "J" to exclude them from the test.

How to enter new images

This section describes in more detail the way new seal images are entered into the system along with basic data such as where and when they were taken, gender, age and so on. It is necessary to enter images and data via a form to ensure that they go to the right folders and database tables. To understand the "Enter Sightings" form the database design is outlined below. It was designed around a few basic concepts concerning the population.

The population consists of a number of animals, some or all of which have permanent markings that provide the potential for individual recognition. The researchers make "sightings" of groups of animals (the group may consist of a single animal) that occur at a time and place. In the case of camera traps the sighting is the set of photos and data collected when the cameras are tripped by one or more animals at a certain time. The set of photos and data that are known at that time to relate to one animal is called an "encounter", thus a sighting consists of one or more encounters. It is possible for there to be more encounters than animals if an animal was photographed from different sides and it was not known at that time that those photos were of the same animal.

Each new encounter is assigned an ID. The animal may be recognized immediately, in which case the ID assigned is the name of that animal, otherwise the ID is any unique alphanumeric that will be assigned by the database if not entered by the user. If the encounter is then "matched" (identified as being with an animal encountered previously) its ID is updated to the ID currently assigned to the previous encounter/s.

We assume that the frame number *range* of photos taken during a sighting was recorded. Frame numbers for individual animals may not have been so the database must allow encounter records to be linked by inspection to images in the photos taken during that sighting – the "Enter Sightings" form was designed to facilitate that.

Entering Images

Step 1.
Download new digital images to the PC – the default location for those images is the "c:\seal_demo\download_photos" folder.

Step 2.
Open the "seal_demo.mdb" database and click on the "Enter Sightings" button

Step 3.
Move to a new sighting by clicking on the rightmost of the "tape recorder" buttons at the bottom of the screen. At the top of the screen the date (defaulted to today's date) and location code can be entered. As either of those values is entered a new sighting number will be assigned.

Step 4.

Display the set of downloaded images that relate to this sighting. The current location of those images is in the leftmost textbox at the bottom of the screen – edit it if it differs from the default. Click the camera icon to the right of the rightmost box and the images in the specified range will be displayed, three at a time. Use the arrow icons to move through larger sets of images relating to this sighting (avoid moving along the set faster than the PC can display new images).

Step 5.
Enter the first encounter record, i.e. the values for sex, age and comments relating to the first of the seals photographed as part of this sighting, into the first row below the sighting time and location. (A non-photographed encounter can also be recorded here) Either type in a unique ID or leave the ID field blank. Highlight the row by clicking in the space to its left. If the ID field was left blank a number will appear in that field (the current "encounter number", which is automatically incremented for each new encounter, will be used). Optionally the default ID will be followed by the catalogue identification letter/s as specified on the start-up screen so that if catalogues are compared later there will be no overlapping IDs.

Step 6.
Determine which image is of the seal in the highlighted row and left-click on the head of that seal image (there may be more than one seal image in a photo). The photo will be saved to the "\newpic" folder (optionally, the catalogue identifier and "sighting number" can be added automatically at the front of its file name to ensure that photo file names are unique within and between catalogues) and a new image record will be created linking the encounter record with the image that was clicked. That image record will be displayed in the list below the encounter record list.

The name assigned in the database for the image is the same as the name of the photo copied to the newpic folder except that a "pxy" string is appended – this indicates that the flank of the seal image is x tenths of the way in from the left-hand edge of the photo and y tenths down from the top edge of the photo. If the photo contains more than one seal image the image name will indicate from which image the pattern is to be extracted when the pattern extract/compare program is run. If the images are too close together on the photo for this mechanism to indicate which image is to be used and to guarantee unique image names then separate the image clicks and include a descriptive note in the "comments" field of the image record.

Clicking twice on a photo at the same position will, in attempting to generate a non-unique image name, cause an error "3022" – click "end" on the error dialog box to resume working (and refresh the list of images from the download folder). The sighting form can be closed (use the black-on-white cross at the upper right) before all the data and images for that sighting have been entered. To continue entering data re-open the sighting form and use the tape recorder buttons at the bottom of the screen to move to the last record (rather than a new record), then continue to add encounters and images.

Once a new image record has been created use the drop down arrow in the "aspect" field to select from which aspect a pattern "cell" is to be extracted. A database record for each specified cell will be created when the sighting form is closed. The name assigned in the database for the pattern cell is the name of the image with "_FL", "_AB", "_CH"

or "_NK" to indicate the aspect. That name is displayed when the photo copied to newpic is presented for pattern extraction.

For example, given an original photo "**040726_98.jpg**" stored in the download folder the names of the photo file copied to newpic, the image on that photo and the lateral cell to be extracted from that image are:

040726_98	A1_040726_98	A1_040726_98p64	A1_040726_98p64_FL
original photo file name	file name of photo copied to newpic folder	name of image selected from photo	name of cell extracted from image

Step 7.
If there is an image of another seal in this sighting (whether in the same photo as contained the previous image or in a different photo) make a new encounter record, highlight it and click on the new image. Repeat steps 6 and 7 until all encounters have been entered and linked to images. Different images of the same animal must be linked to the same encounter record and it is worthwhile linking it to additional images if an area obscured in one image is visible in another. Obviously the same image should not be linked to different records because animals in different encounter records are definitely or potentially (if the encounters are from different sides) different.

Step 8.
When the sighting form is closed a cell record for each image is created. If the sighting form is re-opened and moved to a recently entered encounter the lowest window will then list that cell record with AutoMatch status "E". The "E" status causes the photo showing the image from which that cell is to be extracted to be displayed when the pattern extract/compare program is run. Once all images from a photo have been linked to encounter records the photo can be removed from the download folder.

Reviewing encounters, images and cells

The "view images" form lists image records and the current ID of the encounters to which they were linked and displays those images. Double-clicking an image will duplicate it to a picture frame at the right of the screen. This is useful for manual pre-sorting of a set of photos, for example of captive seals used for testing the system, as it allows photos to be displayed side by side. The camera icons below the picture frames will copy the corresponding images to the "c:\seal_demo\scratch" folder from where they can be expanded/rotated/enhanced etc by image processing software to aid further inspection.

The form can be used, for example, to edit the AutoMatch status of a pattern cell extracted from an image. Once a cell has been extracted and compared the AutoMatch status of its database record changes from E to L (from "extract" to "library"). If the user wants to re-extract the pattern (for example to demonstrate the process) the AutoMatch status would be edited from L back to E. To prevent an existing cell being used in future comparisons (for example a cell judged to be superfluous or too distorted to be useful) the AutoMatch can be edited to J (for "junk"). Images used only for visual comparison are given an "X" status. If automated pattern comparison is in batches (as opposed to using the "real time ident." option in the extract/compare program) the status changes from E to P (for Pending automated comparison) to V (similarity scores ranked to display potential matches for Visual comparison) to L. Any status letter can be edited and edits in the order L > V > P > E are "safe" in the sense that a previous step will simply be repeated. Edits in the opposite direction are unsafe in that a step may be skipped resulting in a crash.