

Automated photo-id for crested newts

Contents

Introduction	page 2
Getting started	page 4
Step 1 - immediate search for a match to a new image	page 4
Step 2 - batch processing	page 5
Step 3 - adding new photographs to the catalogue	page 7
Step 4 - testing the performance of the comparison algorithms	page 8
Extracting flank and hindleg patterns from an image	page 10
Step 1 - selecting and cropping the correct image from the photo	page 10
Step 2 - placing the "special points"	page 11
Step 3 - locating the margins	page 12
Step 4 - fitting the model	page 13
Step 5 - extracting the pattern	page 16
Step 6 - erasing obscured areas	page 18
Step 7 - stripe ends and joins	page 19
Comparing patterns	page 21
Batch compare	page 21
Visual confirm	page 21
Further visual checks?	page 22
Add the latest images to the library?	page 23
Visual compare	page 23
Testing	page 25
Standardising scores	page 27
How to enter images	page 28
Reviewing encounters, images and cells	page 32

Introduction

The following notes are provided as a guide to using the files downloaded from the www.conservationresearch.co.uk website to search a catalogue of crested newt images for a match to a new image.

Extracting the contents of the "crested_newt_folders.zip" file should have created a folder (named "\crested newts" by default), containing images and related files of a few newts in a number of subfolders and an Access database file called "crested newts.mdb". The name of the database file must match the name of the folder so if the default folder name has been changed the database file must also be renamed.

Running "setup" from "ExtractCompare_install.zip" should also have installed the ExtractCompare program. It can be run from the Programs list or from a button on the blue screen that appears when the database file is opened. But before the ExtractCompare program is run make sure that the two textboxes at the bottom of that screen contain the correct paths to the folder and to the program. Also, when the program is first run, make sure that the paths displayed at the top of its initial screen are correct.

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 "\crested newts" 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.

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 crested newts. 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 crested newts they show. If that data is available in spreadsheets or database tables it can be imported into the crested newts.mdb 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 an crested newt 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 crested newt 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 crested newts they were with is never known exactly and lies within a range that depends on the risk of failing to match images of different encounters with the same crested newt. The best we can do is to infer the number of crested newts 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 crested newt 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. In the case of big cats that are poached for their skins another potential function of the system is to match the pattern scanned from a skin to one of the catalogue images and reveal where and when that cat was last encountered.

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.

Getting started

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 crested newts.mdb database and click the "Extract/Compare" button. Elect to "Register Later", pick the "real time ident." option and click the "Start" button.

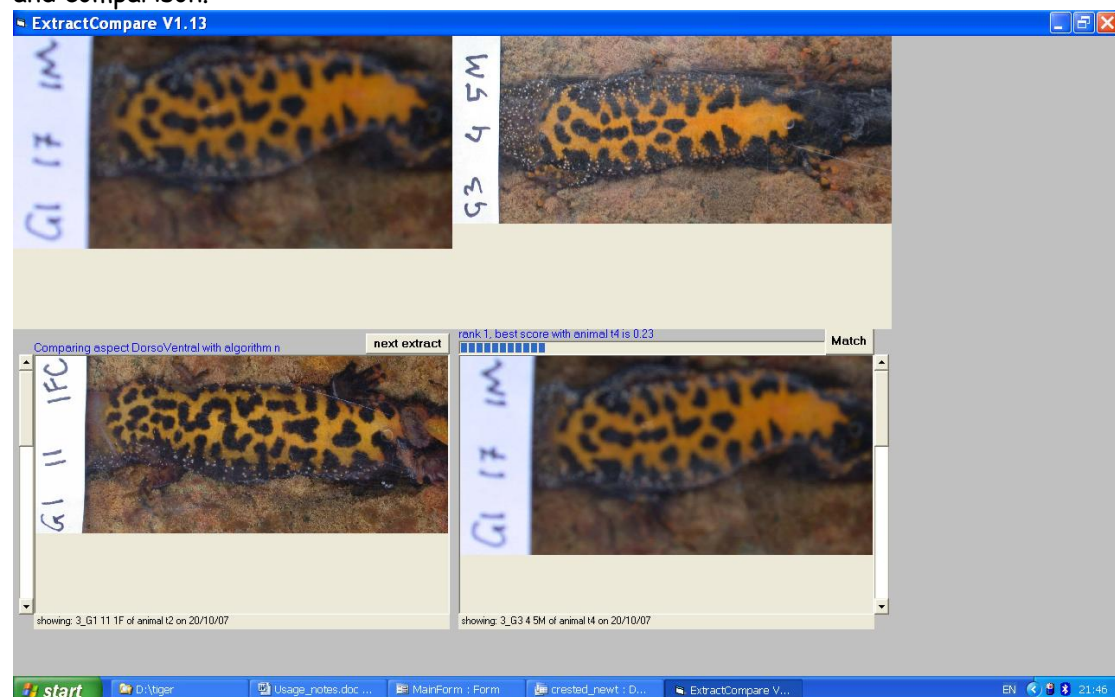
An crested newt image is displayed on the screen. A label near the top of the screen gives the image name as "3_G1 7 2Fp25" with a "_DV" extension indicating that the pattern is to be extracted from the ventral surface. The current ID assigned to the encounter of which this is one image is given as "t2a".

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" (tailbase and neck, in that order) and the red and blue dots tracing the top and bottom edges of the body (which must be oriented to face to the left of the screen). Click "Initial Fit". A 3D surface model of the crested newt body is superimposed on the crested newt image but because the limb base points have not been placed it is not correctly orientated. Check the "and roll" checkbox and click "Initial fit" again. Click "Fit" two or three times. The model changes its orientation and shape to fit to the edges of the crested newt image. Click "Extract". A red line demarcates a region of the image and the grey-shade stripe is scanned from that region and displayed in a small picturebox. Click "OK".

The pattern is displayed in three pictureboxes. The centre box shows a detrended version and the rightmost box a thresholded wavelets decomposition (the screen may go blank for a few seconds before the display is complete). The resulting shapes are thinned to a set of branching lines and small circles are placed at the line endings and joins. Red and blue circles are placed at ends and joins of lines running from right to left and left to right respectively. Green circles are those for which the algorithm could not determine the direction. Click "place seeds" to generate up to 9 sets of triples from the circle patterns, which will be used to initiate the search for spot pattern correspondence in one of the two algorithms that will next be used to 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 stripe pattern will appear at the bottom left of the screen and an image of crested newt "t2" will appear at the bottom right. One more image of "t2" will appear at the top of the screen - click on it to enlarge it at the bottom right. "t2" is the crested newt for which a pattern extract taken from one of its images gave the highest similarity score with the extract from the new image. The new image is clearly another image of crested newt "t2".

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 ("t4") will be displayed. Return to "t2" and click the "Match" button and then the "confirm match of this animal to this animal ->" button. Note the ID of the new image changes from "t2a" to "t2" and the scrollbar moves down to show the next highest-scoring library animal. It isn't a match so click "next extract". You are now shown the results of the second comparison algorithm (of type "c" - the first one was type "n") The right hand picture is blank because the type "c" algorithm also placed animal "t2" at the top of the list and it has already been matched to the new image. Use the scrollbar to check the next library animal on the list. It isn't a match so click "next extract". 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". The program stops because no more images were waiting for pattern extraction and comparison.



Step2 - batch processing.

Back at the blue start-up screen of the crested newt.mdb database click on the "view images" button. That displays a form listing the images that either are to have a pattern extracted or that have had a pattern extracted. The sixth pattern cell is the one that was just extracted. Columns seven and three show it was extracted from image "3_G1 7 2Fp25" that was taken at an encounter with crested newt "t2". The ID is "t2" because it was just updated from "t2a" when the match was confirmed. The AutoMatch status is "L" to identify this pattern cell as one that has been compared and retained in the library.

Edit the AutoMatch value from "L" to "E". Also edit the ID back from "t2" to "t2a".

Close the "view images" form, click "Extract/Compare" and elect to "Register Later" again. Choose the "batch extract" option and click "Retrieve" again. This time drag the stack of "special point" labels across the screen to see how it can be moved if it's obscuring the image and right click on any of the yellow dots that have been placed on

the crested newt image to highlight one of the labels on the stack and identify which of the special points is located by that dot. Click "Initial Fit" > "Fit" > "Extract" > "OK". Because the "batch extract" option was chosen the comparison algorithms do not run at this time and a "Next image" button appears.

When "Next image" is clicked the program stops because no more rows in the cells table have the AutoMatch value "E".

Open the "view images" form again. The AutoMatch value for the sixth record is now "P", which means that it's pending comparison with any other comparable "P" and "L" pattern cells. Close the form and run the ExtractCompare program using the "batch compare" option. The extract will be compared with the other extracts and the similarity scores will be stored in a temporary table. Close the ExtractCompare program and open the "view images" form again. The AutoMatch value has changed to "V", which means that the image is waiting for visual comparison to images of other crested newts in the catalogue. When the ExtractCompare program is run again using the "visual confirm" option the screen that appears is the same as the one that appeared when the "real time ident." option was selected. That's because that option just causes the extraction, comparison and visual confirmation steps to be run sequentially for one pattern cell instead of running each in turn on a batch of cells. Thus the match of "t2a" to "t2" can be confirmed as before.

The similarity scores stored in the temporary table are used to rank the sets of images of existing IDs when the "visual confirmation" step is run, so that the set of most similar images is shown first. The final match decision is however up to the user, the program just saves time by showing the images that are most likely to match first. The similarity scores and sets of images are collated by ID. For example some images may have been matched before the automated system is used, so that the system may be required to check the identity of a crested newt photographed on a number of recent encounters against existing IDs. Thus the left hand picture box shows all those images, the scrollbar to its left being used to switch from one to another. The right hand picturebox shows all images of the existing ID for which one extract achieved the highest similarity score with one of the extracts from the new ID. Alternative images of that existing ID are displayed above as thumbnails that can be clicked to display in the right hand picture box.

To see where the special points were positioned on other images move down the rows in the "view images" form (for example by clicking at the left of a row and using the up/down keys) and edit the AutoMatch value for any row from "L" to "E". Run ExtractCompare using "batch extract" and clicking "Retrieve". Instead of having to fit the model the "scrap" button can be clicked to skip to the "Next image" stage. The AutoMatch value for that row will be set to "S" to indicate that this extract is no longer required and can then be edited back to "L".

Step3 - adding new photographs to the catalogue.

Two photographs have been downloaded to the c:\crested newt\download_photos folder to demonstrate the process of adding new photos to the catalogue. The objective is to make encounter records, grouping different encounters occurring at the same time and

place into a "sighting" and linking the crested newt images on photographs taken during that sighting to the encounters.

Click the "Enter Sightings" button. Sighting 1 is shown. Nine test images of crested newts were treated as separate encounters in this first sighting and those encounters are listed in the central "encounters in sighting" window. Highlighting a row in that window will show, to the right, any image that has been linked to that encounter.

Click the rightmost button of the "tape recorder" controls at the bottom of the screen 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 \crested newt\download_photos folder. 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 one of the images below as being of this encounter and click on, say, the head of the crested newt in that image. A new record will appear in the "images of encounter" window linking this crested newt 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 crested newt 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 crested newt in the photograph was linked to the encounter in case the photograph shows more than one crested newt image. A comment can also be added in the image row to identify which of the crested newt 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 crested newt image can be selected.

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 a "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 "aspect" for the image linked to the encounter is set by default at "DorsoVentral". If the aspect is changed to "none" the linked image will be displayed for visual comparison but will not be presented for pattern extraction.

If the download folder had contained another image of the first encounter (for example an image of the same crested newt from a different angle) it would be linked to the first encounter by clicking to the left of the first encounter row again and clicking on that second image. But in this case the other image is of a different crested newt so generate a second encounter, click at the left hand end of the new encounter row to highlight it and click on the head of the crested newt in the second image to link it to the second encounter.

Close the Enter Sightings form and open the "view images" button. The new cells should appear with AutoMatch status "E" if the DorsoVentral aspect was selected or "X" if the aspect was set at "none".

Repeat the pattern extraction and comparison process described in step2 for these new images. Because extracts have not taken previously from these images 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 images are 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 elect to "Register Late" again. 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 quite 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 crested newt, 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 crested newt 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 crested newt (or more accurately, of the same crested newt 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 crested newt 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

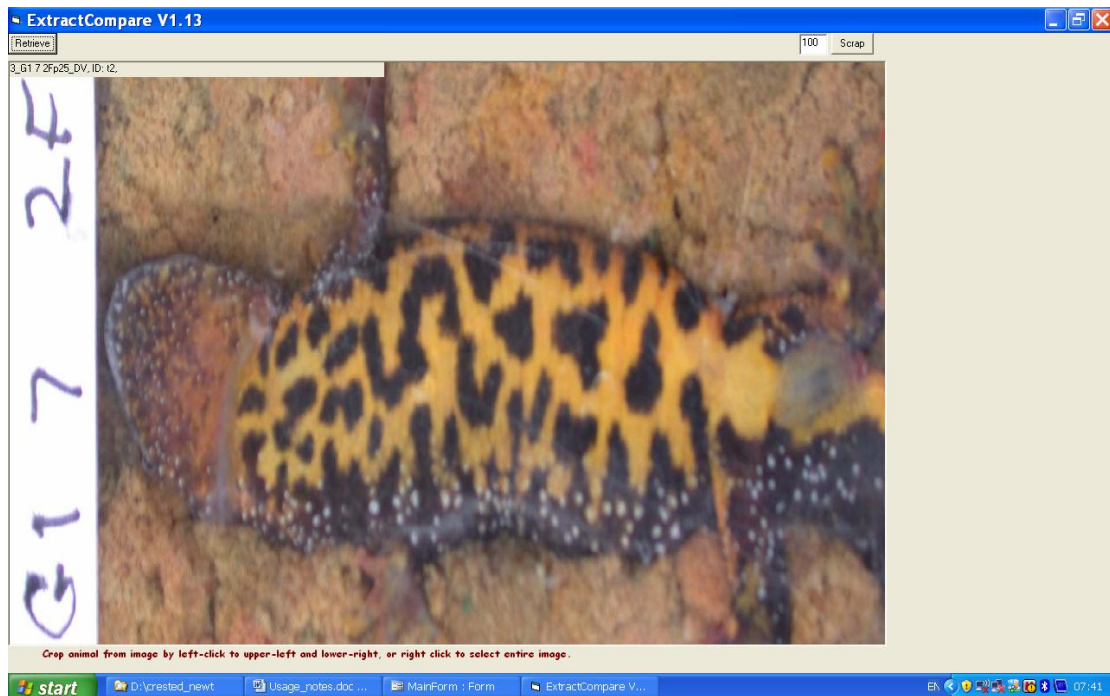
The following section describes in more detail the process of extracting the patterns.

When you run the 'ExtractCompare' program, either from the database screen or the Programs list, you are presented with a stack of six 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.

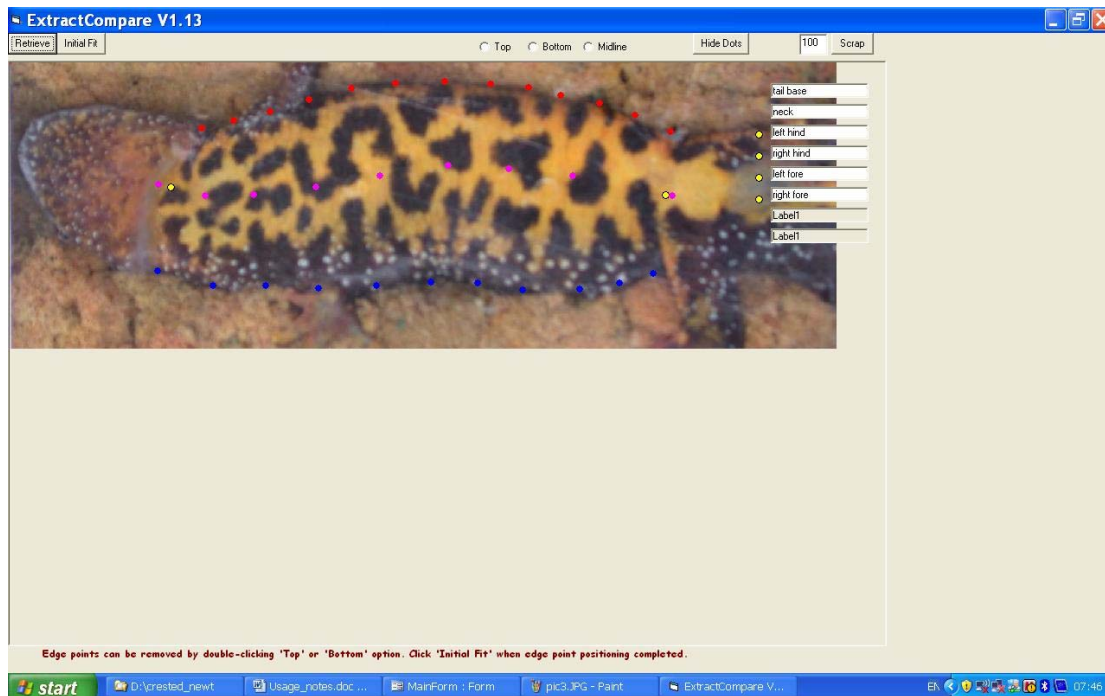
When you click on "Start" the first photo containing an image listed in the "crested newt.mdb" database as waiting for pattern extraction is displayed on the screen. The name of that image will be displayed in a label at the top of the screen. The "pxy" part of the image name will indicate the position on the photo of the image from which the extract is to be taken (in case the photo contains more than one crested newt 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 example below a pattern cell is to be extracted from the image "3_G1 7 2Fp25". In this case it's the only crested newt image on the photo but if there were more we would know to select the one with its head about 2 tenths of the way in from the left-hand edge of the photo and 3 tenths down from the top edge. As indicated by the instruction at the bottom of the screen you can expand the required crested newt 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.



Step 2 – placing the “special points”.

Drag the yellow “tail base” dot to the ventral midline level with the leading edge of the rear limbs and the “neck” point to the ventral midline on the neck. Further “special point” dots will then appear (plus two labels that are currently redundant). The extra dots can be dragged to the limb attachment points if necessary but usually it will be sufficient to use only the tail base and neck points and check the “and roll” checkbox at the next stage. In case of any confusion about which dot on the crested newt 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 crested newt).



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 the neck to level with the base base. Position the last red dot rearward of the yellow "tail base" point and the penultimate red dot ahead of the yellow "tail base" dot. Click the "bottom" option button and place blue dots at intervals along the lower margin. The order in which dots are placed does not matter. 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. Place midline dots along the ventral midline.

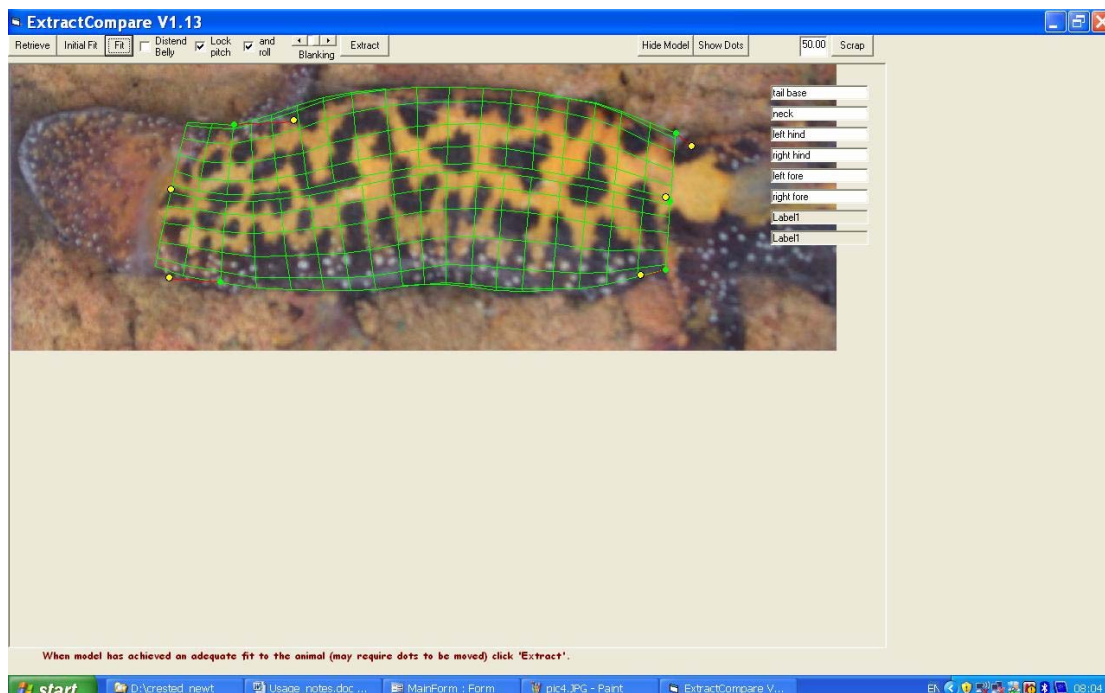
When an extract has previously been taken from a given image and that image is marked (via the "view images" form on the database screen) as needing pattern to be extracted again the "Retrieve" option will be enabled. Clicking the Retrieve button will display the image as it was originally cropped and the special and margin points in their original positions. Use this option with the sample images supplied to see where we placed the special and margin points in each case.

Step 4 – fitting the model.

Click "Initial Fit". A 3d surface model will be superimposed on the image of the crested newt but unless the limb points have been placed its orientation will not be correct. Check the "and roll" checkbox and click "Initial fit" again. The model will be roughly in the right place.



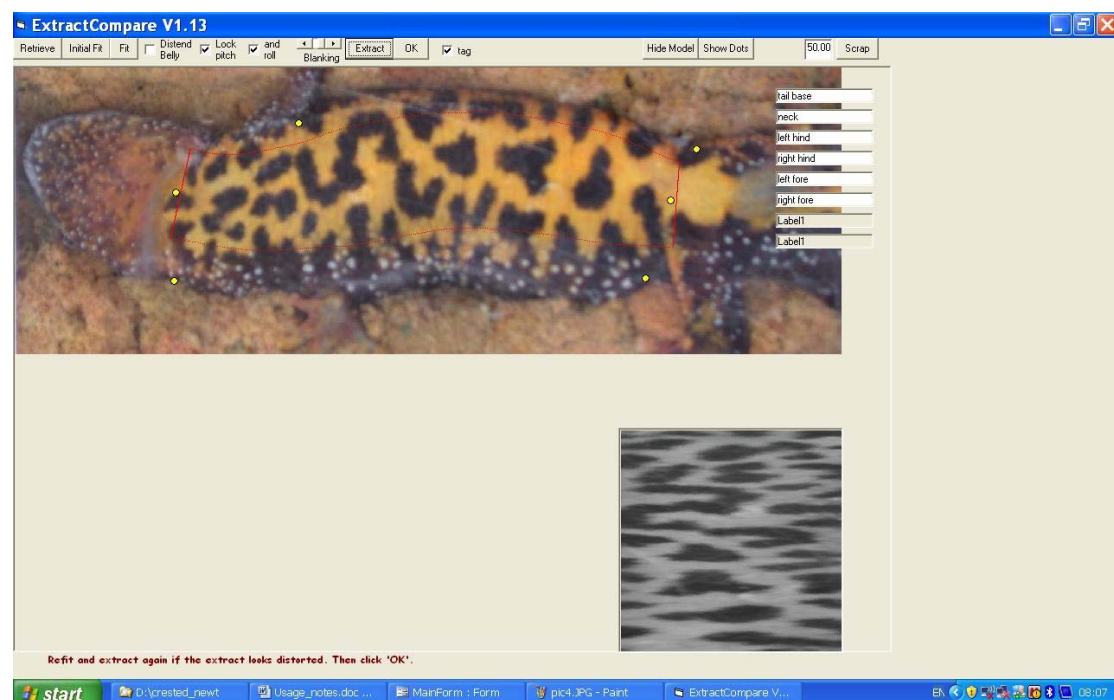
Click "Fit". The model should adjust its shape to fit closely to the edges of the body. Toggle the "Hide Model"/"Show Model" button to check on the fit of the model to the body edges, if necessary click "Fit" again. It may be necessary to adjust some of the margin dots to get a good fit - decrease the distance between successive dots to attract the model to that section of the edge. Two or more red or blue dots rearwards of the yellow "tail base" dot will pull the model rearwards. 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.



In the example above the limb points have been used to encourage the model to bend more at the neck and tail base.

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 head. The pattern cell is displayed in a moveable window and the margins of the array are indicated by the pattern of red lines drawn on the crested newt.

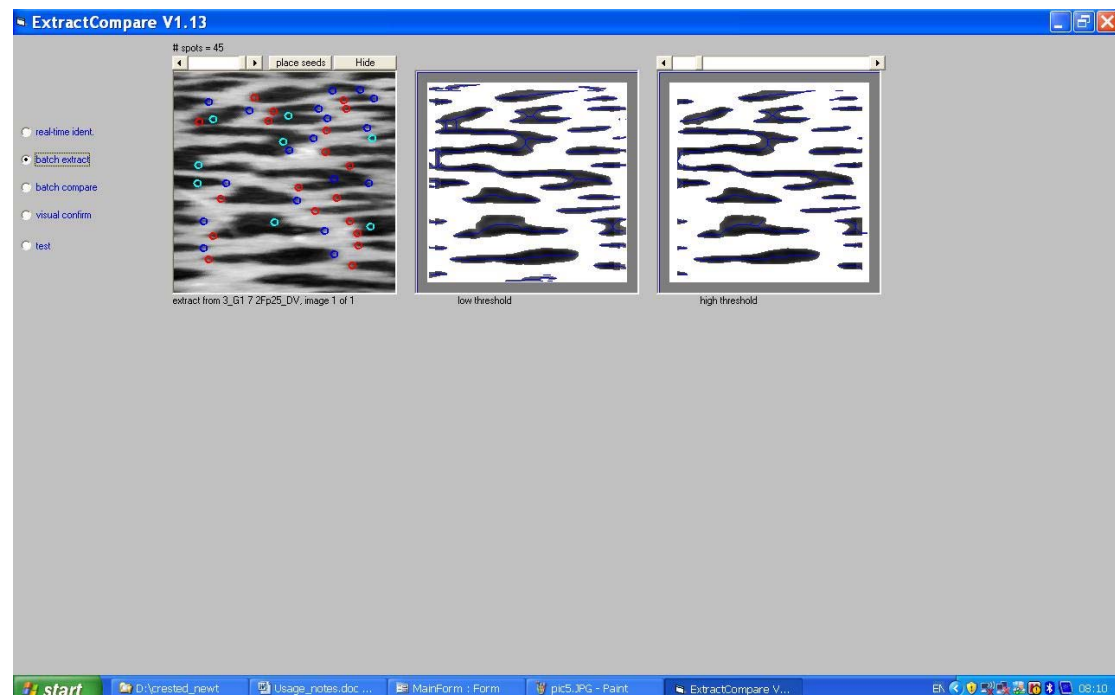


Step 6 - erasing obscured areas.

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.

Step 7.

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 crested newt 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. The central window displays a detrended version of the extract, then the central and the rightmost windows display a wavelets decomposition of the extract at two levels of threshold and thinned to a set of blue lines. Coloured circles on the leftmost window indicate line ends and joins, the colours distinguishing lines ending and joining from left and right (the colour is green where the algorithm has failed to detect the line direction).



Click "place seeds" and then "save extract" to save the extract for subsequent comparison using two 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 oldlib and newlib folders. When the ExtractCompare program is run using the "Batch compare" option those records with AutoMatch status "P" are selected, the corresponding oldlib files compared to each other using a type "c" algorithm and those in newlib compared using a type "n" algorithm. 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 be 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 "oldlib" library and n extracts have been taken from new images using "Batch extract" then "Batch compare" will generate $n(n-1)/2 + nN$ scores using the type "c" algorithm. The two algorithms are to some extent complementary in the way they work so that by using both the risk of missing an existing match is reduced. 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

The scores stored in the "temp" table can be standardised so that scores from different algorithms are comparable, however initially there will be too few scores to calculate the required summary statistics so that the "standardise" and "combine" checkboxes must be unchecked during the "Batch compare" and "Visual confirm" runs. When the ExtractCompare program is run using the "Visual confirm" option the scores are collated by animal ID so that, for each existing ID, the maximum score between all extracts from images linked to encounters with that ID and all extracts linked to a new ID can be 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 can be shown for comparison at the bottom right (its most recent image will be shown). Any alternative images of that most similar ID are 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 crested newt. Suppose ID 2000 is the most similar to the new set of images that show that same crested newt, 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 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).

Even if, in this example, the ID 1000 had been the most similar to the new set of images the final ID assigned to all encounters would still have been 1000 rather than 2000 because the convention on updating IDs is that the later ID is updated to the earlier one. By convention an ID starting with a letter also takes precedence over one starting with a number, on the grounds that 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. The "view images" form can be used at any stage to enter a value in to the "prev_ID" field for an image, which will not be updated and hence be retained as a record of a name originally used for that image.

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. If the current new image (or set of new images) was not matched the new encounter's ID will remain at its current value and hence represent a crested newt that is new to the catalogue. The risk that those images are actually of a crested newt already in the catalogue because the match was missed depends on the number of existing and new images of that crested newt 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. 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.

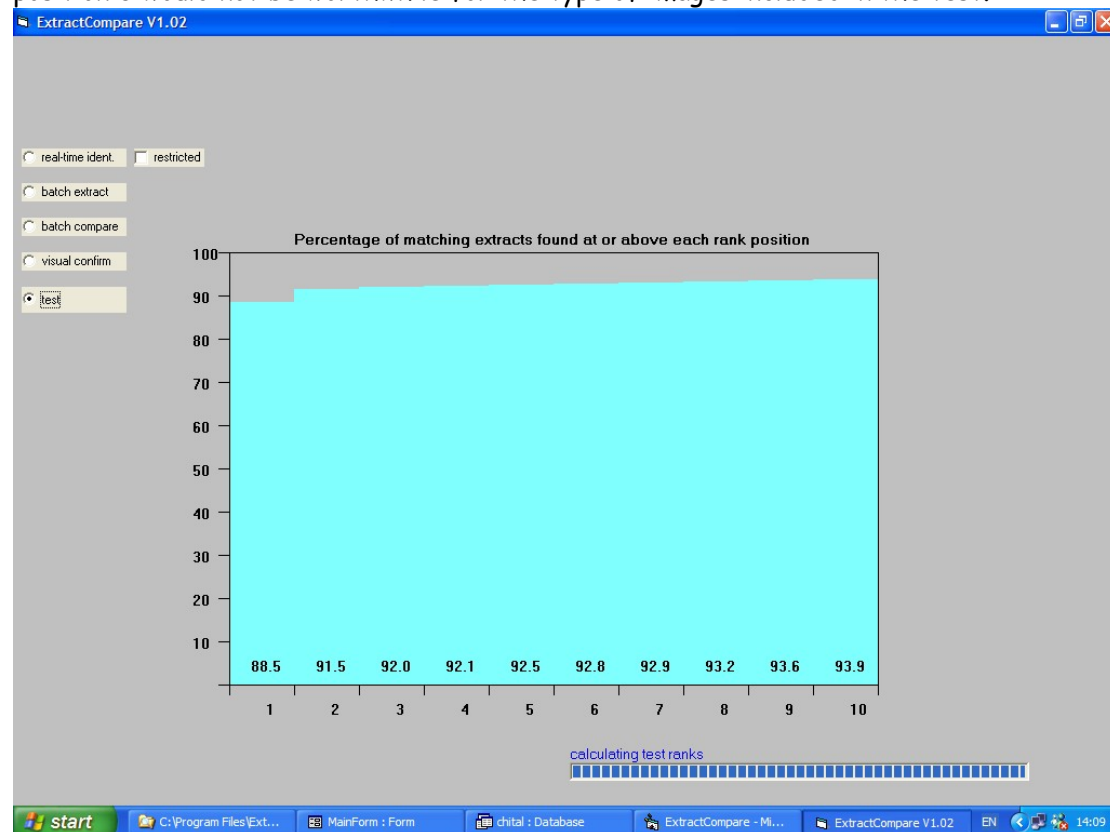
When images of existing IDs are displayed for visual comparison to images of a new ID the existing IDs are ordered by decreasing maximum similarity score. The user searches for a match to an existing ID down to a certain position in that ranked list. So if the maximum similarity score with images of 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 crested newt) 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 crested newt. n such images of the same crested newt 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 crested newts A to Z. Consider images a_1 to a_n of crested newt 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 crested newts B to Z. It's also compared with the maximum scores the extract from

a_2 gets with extracts from images of crested newts 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 crested newt 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 crested newt 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 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 crested newts have been assigned IDs that do not start with a

letter, the encounters with other crested newts 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 crested newts. 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.

Standardising scores

This section describes the method currently used to standardise the scores generated by the type "c" and type "n" algorithms in order to list those scores in a single ranked list. The scores are standardised by using them to calculate the probability that the two extracts that gave that score are from images of the same crested newt.

Let $p\langle m|s \rangle$ be the probability that the extracts are a match given the score.

$p\langle m|s \rangle$ is a conditional probability that equals $p(m \wedge s)/p(s)$

$$= p\langle s|m \rangle p\langle m \rangle / (p\langle s|m \rangle p\langle m \rangle + p\langle s|n \rangle p\langle n \rangle)$$

$$= 1 / (1 + (p\langle s|n \rangle / p\langle s|m \rangle) (p\langle n \rangle / p\langle m \rangle))$$

where $p\langle s|n \rangle$ and $p\langle s|m \rangle$ are, respectively, the probabilities of getting a score of s if the extracts are a non-match and match, and $p\langle n \rangle$ and $p\langle m \rangle$ are, respectively, the prior probabilities of the extracts being a non-match or a match.

Currently the prior probability that a pair of extracts are from images of the same crested newt is set at the reciprocal of the assumed population size. The default value of 0.005 can be edited via the "prior P(m)" textbox that appears when the "visual confirm" option is selected. $p\langle n \rangle$ equals $1 - p\langle m \rangle$. As the prior probability is the same for all extract pairs the value of the $p\langle n \rangle / p\langle m \rangle$ ratio does not currently effect the rank order. However, the prior probability that a pair of extracts are a match could also incorporate information on the probability that, for example, extracts from images recorded as being of crested newts having different genders could still be a match because of recording errors. Estimated age and location are other factors that could affect the prior probability. This standardisation method could therefore provide a way to incorporate additional information with the algorithm scores to give the optimal ranking of potential matches. Currently, for example, an extract from the opposite end of the country will be ranked above one from an image obtained locally even if the score with the local extract is only slightly lower than the score from the distant extract.

$p\langle s|n \rangle$ and $p\langle s|m \rangle$ are the observed score distributions for all non-matches and matches for the relevant algorithm type. $p\langle s|n \rangle$ can also be based on the distribution of non-match scores for each extract individually to improve standardisation when certain extracts tend to get higher non-match scores than others.

How to enter new images

This section describes in more detail the way new crested newt 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:\crested newt\download_photos" folder, which is created when the "crested newt_folders.zip" file is unzipped.

Step 2.

Open the "crested newt.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), time, site and location codes and lat/long of the "sighting" can be entered. As any one 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. Type the prefix assigned by your camera to those image files into the next textbox and their frame range into the next two textboxes. Leave the textboxes blank to display all the images in the folder. 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 crested newts 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 crested newt in the highlighted row and left-click on the head of that crested newt image (there may be more than one crested newt 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 head of the crested newt 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 crested newt 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 whether a pattern "cell" is to be extracted or the image is only to be included

for visual comparison. 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 "_DV" to indicate a ventral extract. That name is displayed when the photo copied to newpic is presented for pattern extraction. If the image is entered only for visual comparison set the aspect field at "none".

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 ventral cell to be extracted from that image are:

040726_98	A1_040726_98	A1_040726_98p64	A1_040726_98p64_DV
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 crested newt 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 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 crested newts 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:\crested newt\scratch" folder from where they can be expanded/rotated/enhanced etc by image processing software to aid further inspection.

The "images/cells" 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.