Photo-identification in amphibian studies: a test of I³S Pattern

Marco Sannolo1,*, Francesca Gatti2, Marco Mangiacotti3,4, Stefano Scali3, Roberto Sacchi4

1 CIBIO, Research Centre in Biodiversity and Genetic Resources, InBIO, Universidade do Porto, Campus Agrário de Vairão, Vila do Conde, Portugal. *Corresponding author. E-mail: marco.sannolo@gmail.com
2 Via Garibaldi, 20090 Vimodrone, Milano, Italy
3 Museo Civico di Storia Naturale di Milano, Corso Venezia, 55, 20121 Milano, Italy
4 Dipartimento di Scienze della Terra e dell’Ambiente, Università di Pavia, Via Taramelli 24, 27100, Pavia, Italy

Submitted on 2015, 21st October; revised on 2016, 8th March; accepted on 2016, 9th March
Editor: Rocco Tiberti

Abstract. Photo-identification is used for individual recognition of several animal species. It gives the possibility to take photos of large species from a distance or to avoid invasive marking techniques in small animals. For amphibians, the use of non-invasive marking methods is even more relevant in the light of their global decline. Here we use the photo-identification data from a population of Triturus carnifex to validate the photo-identification software I³S Pattern. This recently developed utility has never been applied to amphibians. The software proved to be efficient and accurate for individual recognition for this species. Contrarily to the previous releases of the I³S family, I³S Pattern is particularly suitable for amphibians characterized by a complex individual pattern of large blotches or irregular spots, which are not readily identified by eye.

Keywords. Triturus carnifex, computer vision, capture-mark-recapture, photo-identification, SURF algorithm, non-parametric MANOVA.

INTRODUCTION

Individual recognition is important in ecological studies such as Capture-Mark-Recapture designs (CMR). This kind of research is of primary importance because it provides data on variables of interest like abundance, density, habitat use and dispersion, among others (Seber, 1973; Pradel, 1996; Beirincx et al., 2006; Mazerolle et al., 2007). CMR studies require the individual marking of captured animals, and rely on the assumption that the marking method unequivocally identifies each individual during the study period (Williams et al., 2002; Mazerolle et al., 2007).

Photo-identification (hereafter photo-ID) is a well-established technique for individual recognition in CMR studies: it has been successfully applied to invertebrates (Frisch and Hobbs, 2007; Caci et al., 2013), fish (Speed et al., 2007; Van Tienhoven et al., 2007), amphibians (Gamble et al., 2008; Ribeiro and Rebelo, 2011; Zaffaroni Caorsi et al., 2012; Bendik et al., 2013; Moya et al., 2015), reptiles (Perera and Perez-Mellado, 2004; Sacchi et al., 2010) and mammals (Hammond et al., 1990; Kelly, 2001). This approach is based on the identification of individually distinct and permanent features of a certain body region (e.g., patterns of colouration, scars), which are little variable over time or at least during the study period. Photo-ID has long been performed through by-eye comparison of photos. In the case of large photo-ID catalogues, this approach can be time-consuming and prone to recognition errors (Arntzen et al., 2004).

The steady development of digital cameras and computers allowed overcoming these issues, and in the past thirty years various software and algorithms had been developed and are now available (e.g., I³S, van Tienhoven...
et al., 2007; Wild-ID, Bolger et al., 2012; Mantha Matcher, Town et al., 2013; MYDAS, Carter et al., 2014; APHIS, Moya et al., 2015). Nevertheless, these utilities were often designed ad hoc for one or few species to maximize their efficiency, therefore their application to other species requires preliminary validation (Sacchi et al., 2016).

Using photo-ID as a non-invasive marking technique is particularly recommended for amphibians since they are facing a worldwide decline (Blaustein and Wake, 1990; Pounds et al., 2006) and many other marking techniques could be harmful (Bloch and Irschick, 2004; Heemeyer et al., 2007; Waddle et al., 2008) and ethically debated (May, 2004). However, many amphibian species possess complex skin patterns, often characterized by large blotches and irregular spots. On the one hand such skin properties allow the unique identification of each individual. On the other hand, if the study species is characterized by extremely complex pattern, error rate might increase sensibly. Therefore, there is the need of software specifically developed to perform such task. In particular, we focused our attention on I3S Pattern (http://www.reijns.com/i3s/), the latest release of the I3S family (van Tienhoven et al., 2007) because of five main reasons: i) it has been specially designed for those species characterized by big and irregular spots or blotches, such as many amphibians species; ii) it is apparently robust to non-standard photographic settings (as in the case of many field works); iii) it requires little image pre-processing procedure (saving operator time); iv) being freeware, it could be addressed to a broad audience among herpetologists; v) it has not yet been validated on amphibians.

I3S Pattern employs the algorithm SURF (Speeded-Up Robust Features) developed in the field of computer vision (Bay et al., 2007). This free utility allows a fast comparison of the image of the individual to be identified with all the previously stored images and ranks the images from the best to the worst match. To do this, it requires the user to process each image only once by a two steps protocol where: i) setting three homologous “reference points” that allow translating all images in the same two-dimensional reference system (see also Sacchi et al., 2010 for further details); ii) defining a rectangular area from which the software will automatically search and extract the “interest points” on which a unique profile of the image is built (Bay et al., 2007). The chosen area should obviously refer to the same body region, but it has not necessarily be perfectly replicated in each image. The two steps require only few seconds, and a “.fgp” file containing the pattern (i.e., position of reference points and interests points) is created and stored. When a new image is processed and its pattern extracted, the software compares it with the database and computes the dissimilarity score between the new and the stored patterns (see software manual for computational details; http://www.reijns.com/i3s/download/I3S Pattern. pdf). At the end of the comparison process, the software shows side by side the new image and all possible matches ranked by increased dissimilarity score. All proposed matches can be visually inspected to assess if the match is correct.

We used the photo-ID data from an ongoing long-term study on the Italian crested newt Triturus carnifex to test the performance of I3S Pattern. T. carnifex is a good model species to test the software because adult newts show a ventral pattern of irregular spots and large blotches that is unique for each individual and stable over time (Arntzen and Wallis, 1999). Therefore, the aims of this research are to assess i) if I3S Pattern successfully recognizes Triturus carnifex individuals, and ii) its potential application to CMR herpetological studies.

MATERIAL AND METHODS

Study species and sampling sessions

We sampled 324 T. carnifex in the Groane Regional Park (Lombardy, Northern Italy, 45°38’N 9°6’E), in a natural pond located in a typical moor area (Gatti and Sannolo, 2014). Newts were caught using funnel traps during nine sampling sessions every week between March and June 2014. We measured the snout-to-vent-length (SVL) and the tail length (TL) of each newt and we noted their sex (Table 1). We took photos of the ventral pattern of each newt in standard conditions: each newt was put on graph paper at 20 cm from the lens (using a Nikon D-90 with an 18-55 mm lens) while keeping its body as straight as possible; each newt was kept wet, and we conducted the whole operation with wet hands and as fast as possible, in order to avoid the injuries and minimize any stress. To simulate a recapture, after one hour we took a second photo of each individual using the same procedure. Between the two photo sessions, newts were kept in individual 1L plastic box filled with water. Following measurements and photo identification, newts were released at the exact point of capture. Two authors (MS and FG) identified by eye every individual captured during all field sessions, checking for recaptures and providing a double-checked reference against which the software performances were evaluated (Arntzen et al., 2004; Bendik et al., 2013). All images were then processed using I3S Pattern.

Statistical validation of the software

We first evaluated the effect of the two primary sources of error in which the operator could incur, following Sacchi et al. (2010): the error in selecting reference points and research area (Drep1), and the prospective error due to non-perfect alignment and positioning of the newt under the camera (Drep2).
newts, 15 males, and 15 females. The NP-MANOVA was performed using the function “adonis” of the “vegan” R package (Oksanen et al., 2013) in R ver. 3.1.0 (R Core Team, 2014). The P-values were assessed by setting the number of permutations to 999, and we tested the assumption of homogeneity of dispersions among distance (Anderson, 2001; Warton et al., 2012) and Dpop clearly decreased reaching its minimum at 45 interest points on distance (P < 0.001), but not of the interaction term (P = 0.64). Sex has a significant effect on Drep2 (Kruskal-Wallis test, KW = 13.5, P < 0.001, df = 1) and on Dpop (KW = 4.4, P < 0.001, df = 1); female dissimilarity scores were always greater than male ones (Table 1). The effect of interest points was significant only at Dpop level (KW = 122, P < 0.001, df = 4). Increasing the number of interest points, the difference between Drep2 and Dpop clearly decreased reaching its minimum at 45 interest points, but it was always highly significant (KS test at 45 interest points = 1, P < 0.001, one-tailed, Fig. 1).

### RESULTS

Descriptive statistics for the calculated distances are reported in Table 1. The NP-MANOVA detected a significant main effect of both sex (P < 0.001) and interest points on distance (P < 0.001), but not of the interaction term (P = 0.64). Sex has a significant effect on Drep2 (Kruskal-Wallis test, KW = 13.5, P < 0.001, df = 1) and on Dpop (KW = 4.4, P < 0.001, df = 1); female dissimilarity scores were always greater than male ones (Table 1). The effect of interest points was significant only at Dpop level (KW = 122, P < 0.001, df = 4). Increasing the number of interest points, the difference between Drep2 and Dpop clearly decreased reaching its minimum at 45 interest points, but it was always highly significant (KS test at 45 interest points = 1, P < 0.001, one-tailed, Fig. 1).

The three scores were clearly separated (Fig. 1). In particular, Dpop was always greater than Drep2, which in turn is greater than Drep1. This result means that the

### Table 1. Mean dissimilarity scores ± standard deviation in relation to sex and number of interest points. Note that females’ scores are always larger than males’ ones.

<table>
<thead>
<tr>
<th>no. of points</th>
<th>25</th>
<th>30</th>
<th>35</th>
<th>40</th>
<th>45</th>
</tr>
</thead>
<tbody>
<tr>
<td>Drep1 males</td>
<td>0.85 ± 0.44</td>
<td>0.78 ± 0.46</td>
<td>0.72 ± 0.52</td>
<td>0.72 ± 0.46</td>
<td>0.79 ± 0.41</td>
</tr>
<tr>
<td>females</td>
<td>0.95 ± 0.76</td>
<td>0.94 ± 0.70</td>
<td>0.77 ± 0.44</td>
<td>0.85 ± 0.52</td>
<td>0.83 ± 0.43</td>
</tr>
<tr>
<td>Drep2 males</td>
<td>9.73 ± 5.51</td>
<td>8.24 ± 3.50</td>
<td>7.83 ± 3.85</td>
<td>7.99 ± 4.19</td>
<td>7.39 ± 3.54</td>
</tr>
<tr>
<td>females</td>
<td>13.4 ± 5.47</td>
<td>11.8 ± 4.97</td>
<td>10.7 ± 4.85</td>
<td>9.88 ± 3.86</td>
<td>8.91 ± 3.17</td>
</tr>
<tr>
<td>Dpop males</td>
<td>48.8 ± 5.97</td>
<td>36.4 ± 3.74</td>
<td>30.3 ± 3.27</td>
<td>25.3 ± 3.06</td>
<td>22.2 ± 2.62</td>
</tr>
<tr>
<td>females</td>
<td>57.6 ± 9.65</td>
<td>42.1 ± 6.34</td>
<td>35.2 ± 5.61</td>
<td>27.9 ± 3.73</td>
<td>23.9 ± 2.56</td>
</tr>
</tbody>
</table>

Drep1 is, therefore, the dissimilarity score between the same image processed two times, while Drep2 is the dissimilarity between two images of the same individual taken in one session of sampling (i.e., a simulated recapture). We also calculated a third measure, Dpop, that is the mean dissimilarity score of each newt with all other individuals. If the software is able to discriminate consistently among individuals, Drep2 should be lower than Dpop. We also expected Drep1 being always greater than zero and Drep2 always greater than Drep1 since it combines the errors of digitalization with the variability due to both newt positioning under the camera and photo quality (e.g., light conditions). The software allows adjusting nine additional parameters for a fine tuning of the algorithm. Changing the number of interest points sampled in each photo might be the single parameter that affect more heavily the calculated distances among individuals. Thus, we focused on the effect of the number of interest points used to calculate the dissimilarity score. The default number is 35 and has been optimized on sea turtles. Therefore, we resampled the database also at 25, 30, 40, 45 interest points.

To assess if the number of interest points might affect the dissimilarity score, we applied a non-parametric distance-based MANOVA (Anderson, 2001; McArdle and Anderson, 2001) using the three scores (Drep1, Drep2 and Dpop) as dependent variables and the number of interest points, the sex, and their interaction as the independents. For this test, we used 30 adult newts, 15 males, and 15 females. The NP-MANOVA was performed using the function “adonis” of the “vegan” R package (Oksanen et al., 2013) in R ver. 3.1.0 (R Core Team, 2014). The P-values were assessed by setting the number of permutations to 999, and we tested the assumption of homogeneity of dispersions among distance (Anderson, 2001; Warton et al., 2012) using the functions “betadisper” and “permutest” (Anderson, 2006; Borcard et al., 2011) of the “vegan” R package.

### Table 2. Mean ± standard deviation snout-to-vent-length (SVL) and tail length of Triturus carnifex divided by sex and age.

<table>
<thead>
<tr>
<th>Sex</th>
<th>n</th>
<th>SVL ± 0.38</th>
<th>TL ± 0.28</th>
</tr>
</thead>
<tbody>
<tr>
<td>Males</td>
<td>195</td>
<td>6.56 ± 0.38</td>
<td>4.69 ± 0.28</td>
</tr>
<tr>
<td>Females</td>
<td>129</td>
<td>6.89 ± 0.39</td>
<td>5.51 ± 0.37</td>
</tr>
<tr>
<td>Juveniles</td>
<td>12</td>
<td>4.64 ± 0.36</td>
<td>3.73 ± 0.35</td>
</tr>
</tbody>
</table>

Field validation of the software

The ability of the software to identify correctly each individual in the context of a CMR study was then assessed using the newts’ images obtained during the nine session of field capture-mark-recapture. We processed 852 photos relative to 324 adults of *T. carnifex* sampled in the field between March and May 2014 (Table 2). Among them, 55 individuals were captured at least twice (34 males and 21 females). For these individuals, we randomly selected one photo and searched in the whole database for the matching. We noted if the software correctly matched the individual, the assigned rank, and the dissimilarity score between the sample image and the first correct match in the database. This score is equivalent to Drep2 calculated in the previous test. While the false positive rate (FAR) is usually very low and easy to control using a countercheck method (by-eye comparison), the false rejection rate (FRR) is a major issue with photo-identification (Bolger et al., 2012). FRR has therefore been calculated as the ratio between the number of correct matches on the number of total attempts. We accepted a match as correct if it was at least one of the first 10 listed images.
software can correctly distinguish among different newts. Both Drep1 and Drep2 decreased across groups (Table 1), but the difference was not significant. Dpop showed an exponential decrease for the mean values.

The sex had a significant effect in the model, while the interaction term sex×no. of points was not significant, which means either that the sex effect was constant across groups or that the test lacked enough statistical power. In Table 1 it is shown that female means and variances were always larger than male ones.

When the I3S was applied in the context of a CMR study, it was able to correctly match all 55 recaptured newts out of 324 individuals during the nine sampling sessions. In 42 cases the software proposed the correct match as the first one in the list, while for the other 13 cases, the correct match was between the second and the fifth position in the list. The FRR of the I3S was 0.24, but it rapidly dropped to 0 after the first five matches were visually inspected. The mean distance between the tested image and the first correct match (equivalent to Drep2), was 9.55 ± 3.17 SD.

**DISCUSSION**

According to our results, I3S Pattern performs reliable images classification and can be applied to photo-ID of *T. carnifex*. As expected from a good classifier, Dpop was greater than Drep2, and this latter greater than Drep1. This outcome means that a greater dissimilarity was calculated when comparing the images of two different individuals with respect to images representing the same newt (Table 1, Fig. 1). The ability of the software to assign a smaller score to the pair of images representing the same newt should translate into a correctly proposed match as output. Indeed, we found that I3S Pattern always suggested the correct match as one of the first five candidates, and most often as the first one. Moreover, this software proved to be robust against false negatives. In fact, one of the central issues using a new software or algorithm is that the rate of misidentification is unknown, especially the false negative rate (Sacchi et al., 2010, 2016), which could systematically bias the estimate of demographic parameters (Davis and Ovaska, 2001; McCarthy et al., 2009). Our estimation of FRR, instead, varied from 0 to 0.24 depending on how restrictive was the selection of the false negative. These values are similar to those estimated by other studies (Bolger et al., 2012).

A second outcome is that the number of interest points had a significant effect on the dissimilarity score: increasing the number of interest points from the default setting of 35 to 45 did not lead to an increase in performance. Therefore, at least for this species, 35 interest points covered the vast majority of inter-individual variation, managing to tell apart different newts. Apparently, increasing the number of interest points led to sample newt belly in uninformative areas, making more similar two images that represented different individuals. For example, Fig. 2 (on the right newt) summarizes as reflected light may be recognized as a coloration pattern by the software and sampled as interest points. On the contrary, reducing the interest points led to increasing the distance between Drep2 and Dpop. So, for this species, and maybe for other similar ones (sharing the same pattern), it is possible that reducing the number of interest points could improve the performance of I3S Pattern, leading to the sampling of highly informative areas only.

Sex had a significant effect in our analysis, and both the mean and the variance were greater for females than for males (Tab. 1). This result might be due to phenotypic differences in male and female bellies. Indeed, in our population the female pattern is less contrasted, and the blotches are often less sharp in comparison with male ones. If this difference is consistent among the sexes, as we believe, it is possible that I3S Pattern repeatedly calculated a greater dissimilarity scores for female.

In the light of the results of the present study, we are confident that I3S Pattern and its underlined algorithm SURF could be successfully applied not only to species characterized by a regular and clearly distinctive pattern, but also to species keeping highly complex and irregular patterns. This is the case for many species of Bufonidae,
Photo-identification in amphibian studies

Photo-identification in amphibian studies

Ranidae, Salamandridae and Ambystomatidae, which exhibit various cryptic or aposematic patterns that are not easily matched by eye (Vitt and Caldwell, 2013).

ACKNOWLEDGEMENTS

The present study received no financial support from any funding agency in the public, commercial or not-for-profit sectors. The research was conducted with respect to the current Italian laws in relation to amphibians capture, marking and detention (Aut.Min.Conc. Prot. 0008880/PNM). The authors would thank Gariboldi L., Wayne, B., and the “Elba Emperor”, Zuffi M.A.L., for their help and advice during the research period. Two anonymous reviewers provided valuable comments that improve an earlier version of the manuscript.

REFERENCES


Gamble, L., Ravela, S., McGarigal, K. (2008): Multi-scale features for identifying individuals in large biological...


