LifeCLEF: Multimedia Life Species Identification

Alexis Joly
INRIA, France
alexis.joly@inria.fr

Henning Müller
HES-SO, Switzerland
Henning.Mueller@hevs.ch

Hervé Goëau
INRIA, France
herve.goeau@inria.fr

Hervé Glotin
IUF & Univ. de Toulon, France
glotin@univ-tln.fr

Concetto Spampinato
University of Catania, Italy
cspampin@dieei.unict.it

Andreas Rauber
Vienna Univ. of Tech., Austria
rauber@ifs.tuwien.ac.at

Pierre Bonnet
CIRAD, France
pierre.bonnet@cirad.fr

Willem-Pier Vellinga
Xeno-canto foundation, The
Netherlands
wp@xeno-canto.org

Bob Fisher
Edinburgh Univ., UK
rbf@inf.ed.ac.uk

ABSTRACT

Building accurate knowledge of the identity, the geographic distribution and the evolution of living species is essential for a sustainable development of humanity as well as for biodiversity conservation. In this context, using multimedia identification tools is considered as one of the most promising solutions to help bridging the taxonomic gap. With the recent advances in digital devices/equipment, network bandwidth and information storage capacities, the production of multimedia big data has indeed become an easy task. In parallel, the emergence of citizen sciences and social networking tools has fostered the creation of large and structured communities of nature observers (e.g. eBird, Xeno-canto, Tela Botanica, etc.) that have started to produce outstanding collections of multimedia records. Unfortunately, the performance of the state-of-the-art multimedia analysis techniques on such data is still not well understood and is far from reaching the real world's requirements in terms of identification tools. The LifeCLEF lab proposes to evaluate these challenges around 3 tasks related to multimedia information retrieval and fine-grained classification problems in 3 living worlds. Each task is based on large and real-world data and the measured challenges are defined in collaboration with biologists and environmental stakeholders in order to reflect realistic usage scenarios.

1. LifeCLEF LAB OVERVIEW

1.1 Motivations

Building accurate knowledge of the identity, the geographic distribution and the evolution of living species is essential for a sustainable development of humanity as well as for biodiversity conservation. Unfortunately, such basic information is often only partially available for professional stakeholders, teachers, scientists and citizens, and more often incomplete for ecosystems that possess the highest diversity, such as tropical regions. A noticeable cause and consequence of this sparse knowledge is that identifying living plants or animals is usually impossible for the general public, and often a difficult task for professionals, such as farmers, fish farmers or foresters, and even also for the naturalists and specialists themselves. This taxonomic gap [31] was actually identified as one of the main ecological challenges to be solved during the Rio's United Nations Conference in 1992.

In this context, using multimedia identification tools is considered as one of the most promising solutions to help bridging the taxonomic gap [23, 12, 8, 30, 27, 1, 29, 19]. With the recent advances in digital devices, network bandwidth and information storage capacities, the production of multimedia data has indeed become an easy task. In parallel, the emergence of citizen sciences and social networking tools has fostered the creation of large and structured communities of nature observers (e.g. eBird[1], Xeno-canto[2], Tela Botanica[3], etc.) that have started to produce outstanding collections of multimedia records. Unfortunately, the performance of the state-of-the-art multimedia analysis techniques on such data is still not well understood and are far from reaching the real world's requirements in terms of identification tools [19]. Most existing studies or available tools typically identify a few tens or hundreds of species with moderate accuracy whereas they should be scaled-up to take one, two or three orders of magnitude more, in terms of number of species (the total number of living species on earth is estimated to be around 10K for birds, 30K for fishes, 300K for plants and more than 1.2M for invertebrates [6]).

1.2 Evaluated Tasks

The LifeCLEF lab proposes to evaluate these challenges in the continuity of the image-based plant identification task [20] that was run within ImageCLEF lab during the last three years with an increasing number of participants. It however radically enlarges the evaluated challenge towards multimodal data by (i) considering birds and fish in addition to plants (ii) considering audio and video contents in addi-

[1] http://ebird.org/
[2] http://www.xeno-canto.org/
[3] http://www.tela-botanica.org/

tion to images (iii) scaling-up the evaluation data to hundreds of thousands of life media records and thousands of living species. More concretely, the lab is organized around three tasks:

PlantCLEF: an image-based plant identification task

BirdCLEF: an audio-based bird identification task

FishCLEF: a video-based fish identification task

As described in more detail in the following sections, each task is based on big and real-world data and the measured challenges are defined in collaboration with biologists and environmental stakeholders so as to reflect realistic usage scenarios. For this pilot year, the three tasks are mainly concerned with species identification, i.e., helping users to retrieve the taxonomic name of an observed living plant or animal. Taxonomic names are actually the primary key to organize life species and to access all available information about them either on the web, or in herbariums, in scientific literature, books or magazines, etc. Identifying the taxon observed in a given multimedia record and aligning its name with a taxonomic reference is therefore a key step before any other indexing or information retrieval task. More focused or complex challenges (such as detecting species duplicates or ambiguous species) could be evaluated in coming years.

The three tasks are primarily focused on content-based approaches (i.e. on the automatic analyses of the audio and visual signals) rather than on interactive information retrieval approaches involving textual or graphical morphological attributes. The content-based approach to life species identification has several advantages. It is first intrinsically language-independent and solves many of the multilingual issues related to the use of classical text-based morphological keys that are strongly language dependent and understandable only by few experts in the world. Furthermore, an expert of one region or a specific taxonomic group does not necessarily know the vocabulary dedicated to another group of living organisms. A content-based approach can then be much more easily generalizable to new floras or faunas contrary to knowledge-based approaches that require building complex models manually (ontologies with rich descriptions, graphical illustrations of morphological attributes, etc.). On the other hand, LifeCLEF lab is inherently cross-modal through the presence of contextual and social data associated to the visual and audio contents. This includes geo-tags or location names, time information, author names, collaborative ratings or comments, vernacular names (common names of plants or animals), organ or picture type tags, etc. The rules regarding the use of these meta-data in the evaluated identification methods will be specified in the description of each task. Overall, these rules are always designed so as to reflect real possible usage scenarios while offering the largest diversity in the affordable approaches.

1.3 Expected Outcomes

The main expected outcomes of LifeCLEF evaluation campaign are the following:

- give a snapshot of the performances of state-of-the-art multimedia techniques towards building real-world life species identification systems

- provide large and original data sets of biological records, and then allow comparison of multimedia-based identification techniques

- boost research and innovation on this topic in the next few years and encourage multimedia researchers to work on trans-disciplinary challenges involving ecological and environmental data

- foster technological ports from one domain to another and exchanges between the different communities (information retrieval, computer vision, bio-accoustic, machine learning, etc.)

- promote citizen science and nature observation as a way to describe, analyse and preserve biodiversity

2. TASK1: PlantCLEF

2.1 Context

Content-based image retrieval approaches are nowadays considered to be one of the most promising solution to help bridge the botanical taxonomic gap, as discussed in [15] or [22] for instance. We therefore see an increasing interest in this trans-disciplinary challenge in the multimedia community (e.g. in [17, 9, 21, 24, 18, 4]). Beyond the raw identification performances achievable by state-of-the-art computer vision algorithms, the visual search approach offers much more efficient and interactive ways of browsing large floras than standard field guides or online web catalogs. Smartphone applications relying on such image-based identification services are particularly promising for setting-up massive ecological monitoring systems, involving hundreds of thousands of contributors at a very low cost. The first noticeable progress in this way was achieved by the US consortium at the origin of LeafSnap[4]. This popular iPhone application allows a fair identification of 185 common American plant species by simply shooting a cut leaf on a uniform background (see [22] for more details). A step beyond was achieved recently by the Pl@ntNet project [19] which released a cross-platform application (iPhone [14], android[5] and web [6]) allowing (i) to query the system with pictures of plants in their natural environment and (ii) to contribute to the dataset thanks to a collaborative data validation workflow involving Tela Botanica[7] (i.e. the largest botanical social network in Europe).

As promising as these applications are, their performances are however still far from the requirements of a real-world social-based ecological surveillance scenario. Allowing the mass of citizens to produce accurate plant observations requires to equip them with much more accurate identification tools. Measuring and boosting the performances of content-based identification tools is therefore crucial. This was pre-
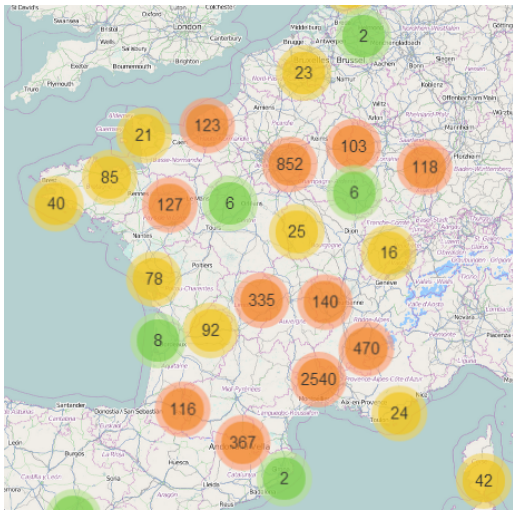
---

[4] http://leafsnap.com/
[5] https://play.google.com/store/apps/details?id=org.plantnet&hl=fr
[6] http://identify.plantnet-project.org/
[7] http://www.tela-botanica.org/

Figure 1: Distribution map of botanical records of the Plant task 2013.

Figure 2: 6 plant species sharing the same common name for laurel in French, belonging to distinct species.

cisely the goal of the ImageCLEF[8] plant identification task organized since 2011 in the context of the worldwide evaluation forum CLEF[9]. In 2011, 2012 and 2013 respectively 8, 10 and 12 international research groups did cross the finish line of this large collaborative evaluation by benchmarking their image-based plant identification systems (see [15], [16] and [20] for more details). Data mobilised during these 3 first years can be consulted at the following url[10], geographic distribution of these botanical records can be seen on Figure 1.

Contrary to previous evaluations reported in the literature, the key objective was to build a realistic task closer to real-world conditions (different users, cameras, areas, periods of the year, individual plants, etc.). This was initially achieved through a citizen science initiative initiated 4 years ago in the context of the Pl@ntNet project in order to boost the image production of Tela Botanica social network. The evaluation data was enriched each year with the new contributions and progressively diversified with other input feeds (Annotation and cleaning of older data, contributions made through Pl@ntNet mobile applications). The plant task of LifeCLEF 2014 is directly in the continuity of this effort. Main novelties compared to the last years are the following: (i) an explicit multi-image query scenario (ii) the supply of user ratings on image quality in the meta-data (iii) a new type of view called "Branch" additionally to the 6 previous ones (iv) basically more species (about 500 which is an important step towards covering the entire flora of a given region).

2.2 Dataset

More precisely, PlantCLEF 2014 dataset is composed of 60,962 pictures belonging to 19,504 observations of 500 species of trees, herbs and ferns living in a European region centered around France. This data was collected by 1608 distinct con-

tributors. Each picture belongs to one and only one of the 7 types of view reported in the meta-data (entire plant, fruit, leaf, flower, stem, branch, leaf scan) and is associated with a single plant observation identifier allowing to link it with the other pictures of the same individual plant (observed the same day by the same person). It is noticeable that most image-based identification methods and evaluation data proposed in the past were so far based on leaf images (e.g. in [22, 5, 9] or in the more recent methods evaluated in [16]). Only few of them were focused on flower's images as in [25] or [3]. Leaves are far from being the only discriminant visual key between species but, due to their shape and size, they have the advantage to be easily observed, captured and described. More diverse parts of the plants however have to be considered for accurate identification. As an example, the 6 species depicted in Figure 2 share the same French common name of *"laurier"* even though they belong to different taxonomic groups (4 families, 6 genera).

The main reason is that these shrubs, often used in hedges, share leaves with more or less the same-sized elliptic shape. Identifying a *laurel* can be very difficult for a novice by just observing leaves, while it is indisputably easier with flowers. Beyond identification performances, the use of leaves alone has also some practical and botanical limitations. Leaves are not visible all over the year for a large fraction of plant species. Deciduous species, distributed from temperate to tropical regions, can't be identified by the use of their leaves over different periods of the year. Leaves can be absent (ie. leafless species), too young or too much degraded (by pathogen or insect attacks), to be exploited efficiently. Moreover, leaves of many species are intrinsically not enough informative or very difficult to capture (needles of pines, thin leaves of grasses, huge leaves of palms, ...).

Another originality of PlantCLEF dataset is that its social nature makes it closer to the conditions of a real-world identification scenario: (i) images of the same species are coming from distinct plants living in distinct areas (ii) pictures are taken by different users that might not have used the same protocol to acquire the images (iii) pictures are taken at different periods in the year. Each image of the dataset is associated with contextual meta-data (author, date, locality name, plant id) and social data (user ratings on image quality, collaboratively validated taxon names, vernacular names) provided in a structured xml file. The gps geo-localization and the device settings are available only for some of the images.

Table 3 gives some examples of pictures with decreasing averaged user ratings for the different types of views. Note that the users of the specialized social network creating these rat-

---

Figure 3: Examples of PlantCLEF pictures with decreasing averaged users ratings for the different types of views

ings (Tela Botanica) are explicitely asked to rate the images according to their plant identification ability and their accordance to the pre-defined acquisition protocol for each view type. This is not an aesthetic or general interest judgement as in most social image sharing sites.

2.3 Task Description

The task will be evaluated as a plant species retrieval task based on multi-image plant observations queries. The goal is to retrieve the correct plant species among the top results of a ranked list of species returned by the evaluated system. Contrary to previous plant identification benchmarks, queries are not defined as single images but as *plant observations*, meaning a set of one to several images depicting the same individual plant, observed by the same person, the same day. Each image of a query observation is associated with a single view type (entire plant, branch, leaf, fruit, flower, stem or leaf scan) and with contextual meta-data (data, location, author). Each participating group is allowed to submit up to 4 runs built from different methods. Semi-supervised and interactive approaches (particularly for segmenting parts of the plant from the background), are allowed but will be compared independently from fully automatic methods. Any human assistance in the processing of the test queries has therefore to be signaled in the submitted runs meta-data.

In practice, the whole PlantCLEF dataset is split in two parts, one for training (and/or indexing) and one for testing. The test set was built by randomly choosing 1/3 of the observations of each species whereas the remaining observations were kept in the reference training set. The xml files containing the meta-data of the *query* images were purged so as to erase the taxon name (the ground truth), the vernacular name (common name of the plant) and the image quality ratings (that would not be available at query stage in a real-world mobile application). Meta-data of the observations in the training set are kept unaltered.

The metric used to evaluate the submitted runs will be a score related to the rank of the correct species in the returned list. Each query observation will be attributed with a score between 0 and 1 equal to the inverse of the rank of the correct species (i.e. decreasing quickly when the rank of the correct species increases). The relevance score of a complete run is the computed as the mean of the per-user average score, i.e:

$$S = \frac{1}{U} \sum_{u=1}^{U} \frac{1}{P_u} \sum_{p=1}^{P_u} \frac{1}{N_{u,p}} s_{u,p}$$

where U is the number of users, P_u the number of individual plants observed by the u-th user, $N_{u,p}$ the number of pictures of the p-th plant observation of the u-th user, $s_{u,p}$ is the score between 1 and 0 equal to the inverse of the rank of the correct species. Using S rather than a flat mean across all query observations is justified by the long tail distribution of the user's data. As in most real-world collaborative datasets, few contributors actually recorded much more pictures than the majority of users who recorded few. Computing the per-user average score before the overall mean allows considering all users equally within the evaluation of the identification performances.

3. TASK2: BirdCLEF

3.1 Context

The bird and the plant identification tasks share similar usage scenarios. The general public as well as professionals like park rangers, ecology consultants, and of course, the ornithologists themselves might actually be users of an automated bird identifying system, typically in the context of wider initiatives related to ecological surveillance or biodiversity conservation. Using audio records rather than bird pictures is justified by current practices [8, 30, 29, 7]. Birds are actually not easy to photograph as they are most of the time hidden, perched high in a tree or frightened by human presence, and they can fly very quickly, whereas audio calls and songs have proved to be easier to collect and very discriminant.

Only three noticeable previous initiatives on bird species identification based on their songs or calls in the context of worldwide evaluation took place, in 2013. The first one was the ICML4B bird challenge joint to the international Conference on Machine Learning in Atlanta, June 2013. It was initiated by the SABIOD MASTODONS CNRS group[11], the university of Toulon and the National Natural History Museum of Paris [13]. It included 35 species, and 76 participants submitted their 400 runs on the Kaggle interface. The second challenge was conducted by F. Brigs at MLSP 2013 workshop, with 15 species, and 79 participants in August 2013. The third challenge, and biggest in 2013, was organised by University of Toulon, SABIOD and Biotope, with 80 species from the Provence, France. More than thirty teams participated, reaching 92% of average AUC. The description of the ICML4B best systems are given into the on-line book [2] [12], including for some of them reference to some useful scripts.
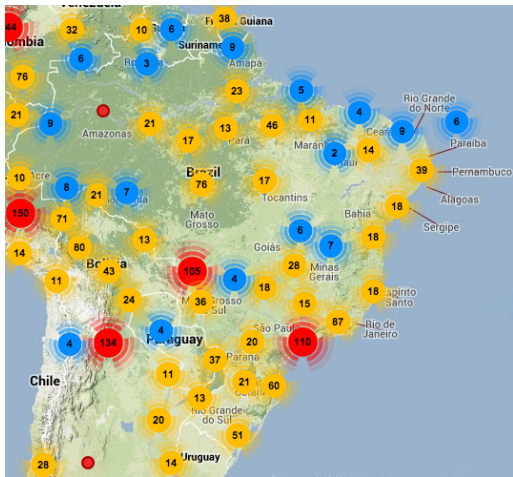

Figure 4: Xeno-canto audio recordings distribution centered around Brazil area

In collaboration with the organizers of these previous challenges, BirdCLEF 2014 goes one step further by (i) significantly increasing the species number by almost an order of magnitude (ii) working on real-world social data built from hundreds of recordists (iii) moving to a more usage-driven and system-oriented benchmark by allowing the use of meta-data and defining information retrieval oriented metrics. Overall, the task is expected to be much more difficult than previous benchmarks because of the higher confusion risk between the classes, the higher background noise and the higher diversity in the acquisition conditions (devices, recordists uses, contexts diversity, etc.). It will therefore probably produce substantially lower scores and offer a better progression margin towards building real-world generalist identification tools.

3.2 Dataset

The training and test data of the bird task is composed of audio recordings collected by Xeno-canto (XC)[13]. Xeno-canto is a web-based community of bird sound recordists worldwide with about 1500 active contributors that have already collected more than 150,000 recordings of about 9000 species. Nearly 500 species from Brazilian forests are used in the BirdCLEF dataset, representing the 500 species of that region with the highest number of recordings, totalling about 14,000 recordings produced by hundreds of users. Figure 4 illustrates the geographical distribution of the dataset samples.

To avoid any bias in the evaluation related to the used audio devices, each audio file has been normalized to a constant bandwidth of 44.1 kHz and coded over 16 bits in wav mono format (the right channel is selected by default). The conversion from the original Xeno-canto data set was done using ffmpeg, sox and matlab scripts. The optimized 16 Mel Filter Cepstrum Coefficients for bird identification (according to an extended benchmark [10]) have been computed with their first and second temporal derivatives on the whole set. They were used in the best systems run in ICML4B and NIPS4B challenges.

Audio records are associated with various meta-data in-

---

cluding the species of the most active singing bird, the species of the other birds audible in the background, the type of sound (call, song, alarm, flight, etc.), the date and location of the observations (from which rich statistics on species distribution can be derived), some textual comments of the authors, multilingual common names and collaborative quality ratings. All of them were produced collaboratively by Xeno-canto community.

3.3 Task Description

Participants are asked to determine the species of the most active singing birds in each query file. The background noise can be used as any other meta-data, but it is forbidden to correlate the test set of the challenge with the original annotated Xeno-canto data base (or with any external content as many of them are circulating on the web). More precisely and similarly to the plant task, the whole BirdCLEF dataset has been split in two parts, one for training (and/or indexing) and one for testing. The test set was built by randomly choosing 1/3 of the observations of each species whereas the remaining observations were kept in the reference training set. Recordings of the same species done by the same person the same day are considered as being part of the same observation and cannot be split across the test and training set. The xml files containing the meta-data of the *query* recordings were purged so as to erase the taxon name (the ground truth), the vernacular name (common name of the bird) and the collaborative quality ratings (that would not be available at query stage in a real-world mobile application). Meta-data of the recordings in the training set are kept unaltered.

The groups participating to the task will be asked to produce up to 4 runs containing a ranked list of the most probable species for each query records of the test set. Each species will have to be associated with a normalized score in the range $[0; 1]$ reflecting the likelihood that this species is singing in the sample. The primary metric used to compare the runs will be the Mean Average Precision averaged across all queries. Additionally, to allow easy comparisons with the previous Kaggle ICML4B and NIPS4B benchmarks, the AUC under the ROC curve will be computed for each species, and averaged over all species.

4. TASK3: FishCLEF

4.1 Context

Underwater video monitoring has been widely used in recent years for marine video surveillance, as opposed to human manned photography or net-casting methods, since it does not influence fish behavior and provides a large amount of material at the same time. However, it is impractical for humans to manually analyze the massive quantity of video data daily generated, because it requires much time and concentration and it is also error prone. An automatic fish identification system in videos is therefore of crucial importance, in order to estimate fish existence and quantity [28, 27, 26]. Moreover, it would help supporting marine biologists to understand the natural underwater environment, promote its preservation, and study behaviors and interactions between marine animals that are part of it. Beyond this, video-based fish species identification finds applications in many other contexts: from education (e.g. primary/high schools) to the
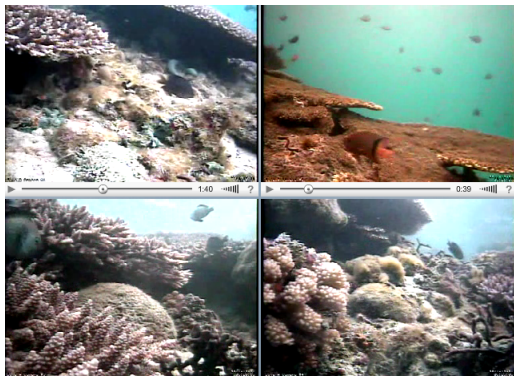
Figure 5: 4 snapshots of 4 cameras monitoring the Taiwan's Kenting site

entertainment industry (e.g. in aquarium).

4.2 Dataset

The underwater video dataset used for FishCLEF, is derived from the Fish4Knowledge[14] video repository, which contains about 700k 10-minute video clips that were taken in the past five years to monitor Taiwan coral reefs. The Taiwan area is particularly interesting for studying the marine ecosystem, as it holds one of the largest fish biodiversities of the world with more than 3000 different fish species whose taxonomy is available at [15]. The dataset contains videos recorded from sunrise to sunset showing several phenomena, e.g. murky water, algae on camera lens, etc., which makes the fish identification task more complex. Each video has a resolution of 320x240 with 8 fps and comes with some additional metadata including date and localization of the recordings. Figure 5 displays 4 snapshots of 4 cameras monitoring the Taiwan's Kenting site. It illustrates the difficulty in detecting and recognizing the fishes.

More specifically, the FishCLEF dataset consists of about 4000 videos with several thousands of detected fish. Only for a small portion (about 20K) of the detected fish, the species are provided. In addition, only 10 fish species will be considered as they represent almost the totality of the fish observed by the Taiwanese coral reef. For each annotation, several metadata will be available in an XML file: video_id, frame_id, bounding_box, and fish_species.

4.3 Task Description

The dataset for the video-based fish identification task will be released in two times: the participants will first have access to the training set and a few months later, they will be provided with the testing set. The goal is to automatically detect fish and its species. The task comprises three sub-tasks: 1) identifying moving objects in videos by either background modeling or object detection methods, 2) detecting fish instances in video frames and then 3) identifying species (taken from a subset of the most seen fish species) of fish detected in video frames.

Participants could decide to compete for only one subtask or to all subtasks. Although tasks 2 and 3 are based on still images, participants are invited to exploit motion infor-

---

[14]www.fish4knowledge.eu

[15]http://fishdb.sinica.edu.tw/

mation extracted from videos to support the identification tasks.

As scoring functions, the authors are asked to produce:

- ROC curves for sub-task one. In particular, precision, recall and F-measures measured when comparing, on a pixel basis, the ground truth binary masks and the output masks of the object detection methods are required;

- Recall for fish detection in still images as a function of bounding box overlap percentage: a detection is considered true positive if the overlap ratio between it and the corresponding object in the ground truth is over 0.7 (similarly to what is done in the object detection task of PASCAL VOC challenge [11]).

- Average recall and precision for each fish species for the fish recognition task. For this task, the recall for fish identification in video frames is also asked.

5. SCHEDULE AND PERSPECTIVES

LifeCLEF 2014 registrations opened in December 2013 and will close at the end of April 2014. At the time of writing, already 61 research groups registered to at least one of the three task and this number will continue growing. As in any evaluation campaign, many of the registered groups won't cross the finish line be submitting official runs but this reflects at least their interest in LifeCLEF data and the related challenges. The schedule of the ongoing and remaining steps of LifeCLEF 2014 campaign is the following:

31.01.2014: training data ready and shared
15.03.2014: test data ready and shared
01.05.2014: deadline for submission of runs
15.05.2014: release of raw results
15.06.2014: submission of working notes describing each participant systems and runs
15.07.2014: overall tasks reports including results analysis and main findings
16.09.2014: 1 day LifeCLEF workshop at CLEF 2014 Conference (Sheffield, UK)

The organisation of a new campaign in 2015, as well as its precise content, will depend on the outcomes of the 2014 edition and on the feedback received from the registered groups.

6. ADDITIONAL AUTHORS

Robert Planqué (Xeno-Canto, The Netherlands, email: r.planque@vu.nl)

7. REFERENCES

[1] *MAED '12: Proceedings of the 1st ACM International Workshop on Multimedia Analysis for Ecological Data*, New York, NY, USA, 2012. ACM. 433127.

[2] *Proc. of the first workshop on Machine Learning for Bioacoustics*, 2013.

[3] A. Angelova, S. Zhu, Y. Lin, J. Wong, and C. Shpecht. Development and deployment of a large-scale flower recognition mobile app, December 2012.

[4] E. Aptoula and B. Yanikoglu. Morphological features for leaf based plant recognition. In *Proc. IEEE Int. Conf. Image Process., Melbourne, Australia*, page 7, 2013.

[5] A. R. Backes, D. Casanova, and O. M. Bruno. Plant leaf identification based on volumetric fractal dimension. *International Journal of Pattern Recognition and Artificial Intelligence*, 23(6):1145–1160, 2009.

[6] H.-T. C. Baillie, J.E.M. and S. Stuart. 2004 iucn red list of threatened species. a global species assessment. IUCN, Gland, Switzerland and Cambridge, UK, 2004.

[7] F. Briggs, B. Lakshminarayanan, L. Neal, X. Z. Fern, R. Raich, S. J. Hadley, A. S. Hadley, and M. G. Betts. Acoustic classification of multiple simultaneous bird species: A multi-instance multi-label approach. *The Journal of the Acoustical Society of America*, 131:4640, 2012.

[8] J. Cai, D. Ee, B. Pham, P. Roe, and J. Zhang. Sensor network for the monitoring of ecosystem: Bird species recognition. In *Intelligent Sensors, Sensor Networks and Information, 2007. ISSNIP 2007. 3rd International Conference on*, pages 293–298, Dec 2007.

[9] G. Cerutti, L. Tougne, A. Vacavant, and D. Coquin. A parametric active polygon for leaf segmentation and shape estimation. In *International Symposium on Visual Computing*, pages 202–213, 2011.

[10] O. Dufour, T. Artieres, H. GLOTIN, and P. Giraudet. Clusterized mel filter cepstral coefficients and support vector machines for bird song idenfication. 2013.

[11] M. Everingham, L. Van Gool, C. K. I. Williams, J. Winn, and A. Zisserman. The pascal visual object classes (voc) challenge. *International Journal of Computer Vision*, 88(2):303–338, June 2010.

[12] K. J. Gaston and M. A. O'Neill. Automated species identification: why not? *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, 359(1444):655–667, 2004.

[13] H. Glotin and J. Sueur. Overview of the first international challenge on bird classification. 2013.

[14] H. Goëau, P. Bonnet, A. Joly, V. Bakić, J. Barbe, I. Yahiaoui, S. Selmi, J. Carré, D. Barthélémy, N. Boujemaa, et al. Pl@ ntnet mobile app. In *Proceedings of the 21st ACM international conference on Multimedia*, pages 423–424. ACM, 2013.

[15] H. Goëau, P. Bonnet, A. Joly, N. Boujemaa, D. Barthélémy, J.-F. Molino, P. Birnbaum, E. Mouysset, and M. Picard. The ImageCLEF 2011 plant images classification task. In *CLEF working notes*, 2011.

[16] H. Goëau, P. Bonnet, A. Joly, I. Yahiaoui, D. Barthelemy, N. Boujemaa, and J.-F. Molino. The imageclef 2012 plant identification task. In *CLEF working notes*, 2012.

[17] H. Goëau, A. Joly, S. Selmi, P. Bonnet, E. Mouysset, L. Joyeux, J.-F. Molino, P. Birnbaum, D. Bathelemy, and N. Boujemaa. Visual-based plant species identification from crowdsourced data. In *ACM conference on Multimedia*, pages 813–814, 2011.

[18] A. Hazra, K. Deb, S. Kundu, P. Hazra, et al. Shape oriented feature selection for tomato plant identification. *International Journal of Computer Applications Technology and Research*, 2(4):449–meta, 2013.

[19] A. Joly, H. Goeau, P. Bonnet, V. Bakić, J. Barbe, S. Selmi, I. Yahiaoui, J. Carré, E. Mouysset, J.-F. Molino, N. Boujemaa, and D. Barthélémy. Interactive plant identification based on social image data. *Ecological Informatics*, 2013.

[20] A. Joly, H. Goëau, P. Bonnet, V. Bakic, J.-F. Molino, D. Barthélémy, and N. Boujemaa. The Imageclef Plant Identification Task 2013. In *International workshop on Multimedia analysis for ecological data*, Barcelone, Espagne, Oct. 2013.

[21] H. Kebapci, B. Yanikoglu, and G. Unal. Plant image retrieval using color, shape and texture features. *The Computer Journal*, 54(9):1475–1490, 2011.

[22] N. Kumar, P. N. Belhumeur, A. Biswas, D. W. Jacobs, W. J. Kress, I. C. Lopez, and J. V. B. Soares. Leafsnap: A computer vision system for automatic plant species identification. In *European Conference on Computer Vision*, pages 502–516, 2012.

[23] D.-J. Lee, R. B. Schoenberger, D. Shiozawa, X. Xu, and P. Zhan. Contour matching for a fish recognition and migration-monitoring system. In *Optics East*, pages 37–48. International Society for Optics and Photonics, 2004.

[24] S. Mouine, I. Yahiaoui, and A. Verroust-Blondet. Advanced shape context for plant species identification using leaf image retrieval. In *ACM International Conference on Multimedia Retrieval*, pages 49:1–49:8, 2012.

[25] M.-E. Nilsback and A. Zisserman. Automated flower classification over a large number of classes. In *Indian Conference on Computer Vision, Graphics and Image Processing*, pages 722–729, 2008.

[26] M. R. Shortis, M. Ravanbakksh, F. Shaifat, E. S. Harvey, A. Mian, J. W. Seager, P. F. Culverhouse, D. E. Cline, and D. R. Edgington. A review of techniques for the identification and measurement of fish in underwater stereo-video image sequences. In *SPIE Optical Metrology 2013*, pages 87910G–87910G. International Society for Optics and Photonics, 2013.

[27] C. Spampinato, Y.-H. Chen-Burger, G. Nadarajan, and R. B. Fisher. Detecting, tracking and counting fish in low quality unconstrained underwater videos. In *VISAPP (2)*, pages 514–519. Citeseer, 2008.

[28] C. Spampinato, D. Giordano, R. Di Salvo, Y.-H. J. Chen-Burger, R. B. Fisher, and G. Nadarajan. Automatic fish classification for underwater species behavior understanding. In *Proceedings of the first ACM international workshop on Analysis and retrieval of tracked events and motion in imagery streams*, pages 45–50. ACM, 2010.

[29] M. Towsey, B. Planitz, A. Nantes, J. Wimmer, and P. Roe. A toolbox for animal call recognition. *Bioacoustics*, 21(2):107–125, 2012.

[30] V. M. Trifa, A. N. Kirschel, C. E. Taylor, and E. E. Vallejo. Automated species recognition of antbirds in a mexican rainforest using hidden markov models. *The Journal of the Acoustical Society of America*, 123:2424, 2008.

[31] Q. D. Wheeler, P. H. Raven, and E. O. Wilson. Taxonomy: Impediment or expedient? *Science*, 303(5656):285, 2004.