

LifeCLEF 2014: Multimedia Life Species Identification Challenges

Alexis Joly¹ Hervé Goëau² Hervé Glotin³ Concetto Spampinato⁴ Pierre Bonnet⁵ Willem-Pier Vellinga⁶ Robert Planque⁶ Andreas Rauber⁷ Bob Fisher⁸ Henning Müller⁹

¹ Inria, LIRMM, Montpellier, France

² Inria, Saclay, France

³ IUF & Univ. de Toulon, France

⁴ University of Catania, Italy

⁵ CIRAD, France

⁶ Xeno-canto foundation, The Netherlands

⁷ Vienna Univ. of Tech., Austria

⁸ Edinburgh Univ., UK

⁹ HES-SO, Switzerland

Abstract. Using multimedia identification tools is considered as one of the most promising solution to help bridging the taxonomic gap and build accurate knowledge of the identity, the geographic distribution and the evolution of living species. Large and structured communities of nature observers (e.g. eBird, Xeno-canto, Tela Botanica, etc.) as well as big monitoring equipments have actually started to produce outstanding collections of multimedia records. Unfortunately, the performance of the state-of-the-art analysis techniques on such data is still not well understood and is far from reaching the real world's requirements. 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. This paper presents more particularly the 2014 edition of LifeCLEF, i.e. the pilot one. For each of the three tasks, we report the methodology and the datasets as well as the official results and the main outcomes.

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 [58] 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 solution to help bridging the taxonomic gap [39, 19, 11, 55, 49, 1, 54, 32]. 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. eBird¹⁰, Xeno-canto¹¹, Tela Botanica¹², 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 [32]. 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 [7]).

1.2 Evaluated Tasks

The LifeCLEF lab proposes to evaluate these challenges in the continuity of the image-based plant identification task [33] 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 addition 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

¹⁰ <http://ebird.org/>

¹¹ <http://www.xeno-canto.org/>

¹² <http://www.tela-botanica.org/>

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 multi-lingual 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 Main contributions

The main 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



Fig. 1. Thematic map of the 127 registrants to LifeCLEF 2014

- foster technological ports from one domain to another and exchanges between the different communities (information retrieval, computer vision, bio-acoustic, machine learning, etc.)
- promote citizen science and nature observation as a way to describe, analyse and preserve biodiversity

In 2014, 127 research groups worldwide did registered to at least one task of the lab. Figure 1 displays the distribution of the registrants per task showing that some of them were interested specifically in one task whereas some others were interested in several or all of them. Of course, as in any evaluation campaign, only a small fraction of this raw audience did cross the finish line by submitting runs (actually 22 of them). But still, this shows the high attractiveness of the proposed datasets and challenges as well as the potential emergence of a wide community interested in life media analysis.

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 [22] or [37] for instance. We therefore see an increasing interest in this trans-disciplinary challenge in the multimedia community (e.g. in [26, 12, 36, 41, 28, 5]). 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 LeafSnap¹³. 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 [37] for more details). A step beyond was achieved recently by the Pl@ntNet project [32] which released a cross-platform application (iPhone [21], android¹⁴ and web¹⁵) 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 Botanica¹⁶ (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 precisely the goal of the ImageCLEF¹⁷ plant identification task organized since 2011 in the context of the worldwide evaluation forum CLEF¹⁸. 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 images-based plant identification systems (see [22], [23] and [33] for more details). Data mobilised during these 3 first years can be consulted at the following url¹⁹, geographic distribution of theses botanical records can be seen on Figure 2.

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 [32] 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

¹³ <http://leafsnap.com/>

¹⁴ <https://play.google.com/store/apps/details?id=org.plantnet>

¹⁵ <http://identify.plantnet-project.org/>

¹⁶ <http://www.tela-botanica.org/>

¹⁷ <http://www.imageclef.org/>

¹⁸ <http://www.clef-initiative.eu/>

¹⁹ <http://publish.plantnet-project.org/project/plantclef>

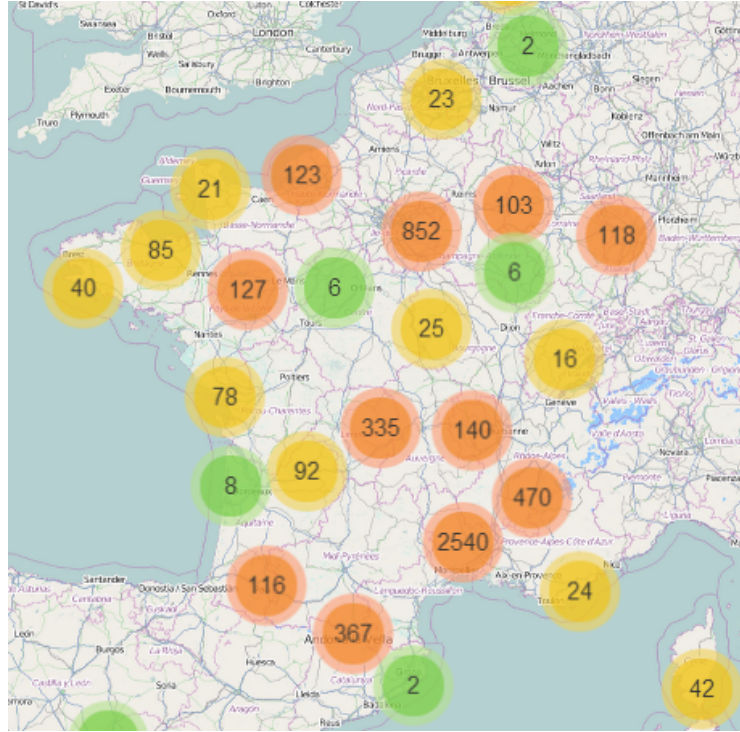


Fig. 2. Distribution map of botanical records of the Plant task 2013.

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 contributors. 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 [37, 6, 12] or in the more recent methods evaluated in [23]). Only few of them were focused on flower's images as in [42] or [4]. Leaves are far from being the only discriminant visual key between species but, due to their



Fig. 3. 6 plant species sharing the same common name for laurel in French, belonging to distinct species.

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 3 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 undisputably 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 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 4 gives some examples of pictures with decreasing averaged users ratings for the different types of views. Note that the users of the specialized social network creating these ratings (Tela Botanica) are explicitly asked to rate the



































Stars	★★★★★	★★★★☆	★★★☆☆	★★★☆☆	★★☆☆☆
Branch <i>Cercis siliquastrum</i> L.					
Entire <i>Quercus ilex</i> L.					
Leaf (photo) <i>Pittosporum tobira</i> L.					
Leaf (scan & scan-like) <i>Hedera helix</i> L.					
Flower <i>Papaver rhoeas</i> L.					
Fruit <i>Crataegus monogyna</i> L.					
Stem <i>Betula pendula</i> L.					

Fig. 4. Examples of PlantCLEF pictures with decreasing averaged users ratings for the different types of views

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 was 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). Semi-supervised and interactive approaches were allowed but as a variant of the task and therefore evaluated independently from the fully automatic methods. None of the participants, however, did use such approaches in the 2014 campaign.

In practice, the whole PlantCLEF dataset was split in two parts, one for training (and/or indexing) and one for testing. The training set was delivered to the participants in January 2014 and the test set two months later so that participants had some times to become familiar with the data and train their systems. After the delivery of the test set, participants had two additional months to run their system on the undetermined plant observations and finally send their results files. Participants were allowed to submit up to 4 distinct runs. More concretely, 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 were kept unaltered.

The metric used to evaluate the submitted runs was a score related to the rank of the correct species in the returned list. Each query observation was attributed with a score between 0 and 1 reflecting equal to the inverse of the rank of the correct species (equal to 1 if the correct species is the top-1 decreasing quickly while the rank of the correct species increases). An average score was then computed across all plant observation queries. A simple mean on all plant observation queries would however introduce some bias. Indeed, we remind that the PlantCLEF dataset was built in a collaborative manner. So that few contributors might have provided much more observations and pictures than many other contributors who provided few. Since we want to evaluate the ability of a system to provide the correct answers to all users, we rather measure the mean of the average classification rate per author. Finally, our primary metric was defined as the following average classification score S :

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

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 equals to the inverse of the rank of the correct species.

2.4 Participants and Results

74 research groups worldwide registered to the plant task (31 of them being exclusively registered to the bird task). Among this large raw audience, 10 research groups did cross the finish line by submitting runs (from 1 to 4 depending on the teams). Details on the participants and the methods used in the runs are synthesised in the overview working note of the task [25] and further developed in the individual working notes of the participants who submitted one (BME

TMIT [53], FINKI [15], I3S [29], IBM AU [13], IV-Processing [18], MIRACL [34], PlantNet [27], QUT [52], Sabanci-Okan [59], SZTE [44]). We here only report the official scores of the 27 collected runs and discuss the main outcomes of the task.

Figure 7 therefore shows the main official score obtained by each run of the task.

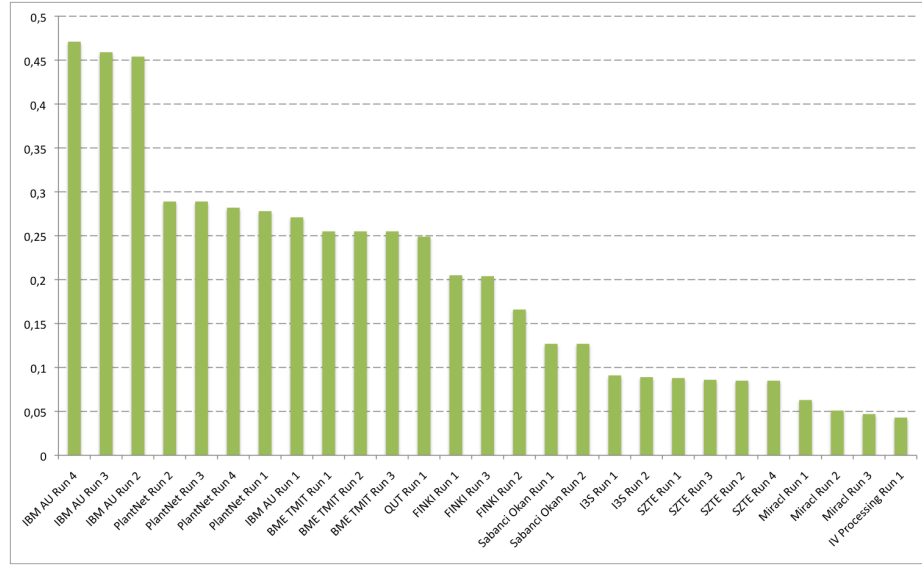


Fig. 5. Official results of the LifeCLEF 2014 Plant Identification Task.

The best results are indisputably obtained by the three last runs of the IBM AU team (*IBM AU run 2-4*). This confirms that using Fisher Vector encoding and linear support vector machines still provides the best state-of-the-art performances as in many other fine-grained image classification benchmarks. On the other side, the convolutional neural network used in the first run of the same team (*IBM AU run 1*) didn't succeed in outperforming the handcrafted visual features used in the 4 runs of the PlantNet team (whereas they are known to perform very well in generalist benchmarks such as ImageNET). The main reason, as discussed in the working note of IBM AU team [13], is that deep models usually require much training data to learn their millions of parameters and avoid overfitting (e.g. up to 1000 images per class within ImageNet). To solve this issue, deep neural networks are usually pre-trained on generalist classification tasks before being fine-tuned on the targeted task. But as using external training data was not authorized in PlantCLEF 2014, this approach could not be evaluated by the participants. Allowing such approaches in next campaigns

might be possible but is a tricky problem as we need to guaranty that none of the images of test set could be found somewhere on the web (queries of the 2014 campaign are for instance publically available on TelaBotanica website).

Despite the supremacy of IBM fisher vectors runs, it is surprising to see that the performances of BME TMIT runs, which are based on a very close training model, reached much lower performances. It demonstrates that different implementations and parameters tuning can bring very different performances (e.g. 512x60 fisher vectors dimensions for IBM AU vs. 258x80 for BME TMIT). Another outcome of the task was that the second best performing method from PlantNet was already among the best performing methods in previous plant identification challenges [2] although LifeCLEF dataset is much bigger and somehow more complex because of the social dimension of the data. This demonstrates the genericity and stability of the underlying matching method and features.

This year, few teams attempted to explore the use of metadata. The date was exploited in the Sabanki-Okan runs, only on flowers or fruits, but we don't have a point of comparison in order to see if the use of this information was useful or not. Miracl team attempted to combine the whole textual and structural informations contained in the xml files, but it has been showed to degrade the performances of their pure visual approach. Note that for the first year, after three years of unsuccessful attempts during the previous ImageCLEF Plant Identification Tasks, none of the teams tried to use the locality and GPS information.

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 [11, 55, 54, 10]. 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 group²⁰, the university of Toulon and the National Natural History Museum of Paris [20]. 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

²⁰ <http://sabiod.org>

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 [3], including for some of them reference to some useful scripts.

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 by audio recordings collected by Xeno-canto (XC)²¹. 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 6 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 [16]) 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 including 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.

²¹ <http://www.xeno-canto.org/>

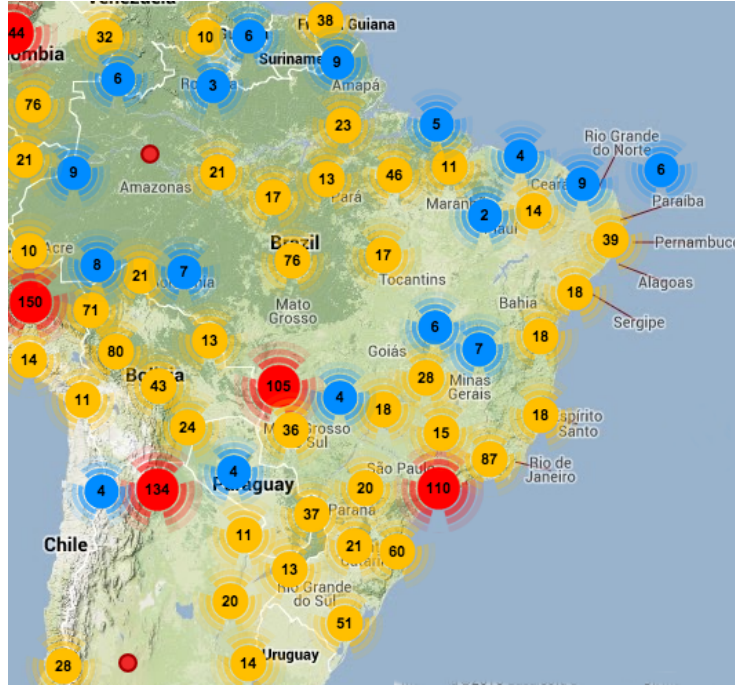


Fig. 6. Xeno-canto audio recordings distribution centered around Brazil area

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.

3.4 Participants and Results

87 research groups worldwide registered to the bird task (42 of them being exclusively registered to the bird task). Among this large raw audience, 10 research groups, coming from 9 distinct countries, did cross the finish line by submitting runs (from 1 to 4 depending on the teams). Details on the participants and the methods used in the runs are synthesised in the overview working note of the task [24] and further developed in the individual working notes of the participants who submitted one (MNB TSA [38], QMUL [51], Inria Zenith [31], HTL [46], Utrecht Univ. [57], Golem [40], SCS [43]). We here only report the official scores of the 29 collected runs and discuss the main outcomes of the task.

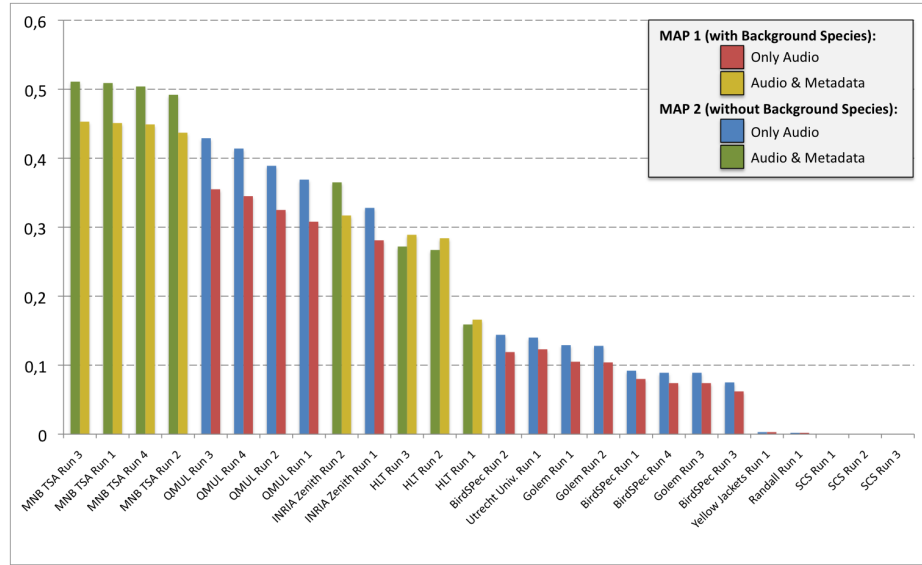


Fig. 7. Official scores of the LifeCLEF Bird Identification Task. mAP 1 is the Mean Average Precision averaged across all queries taking into account the Background species (while mAP2 is considering only the foreground species).

Figure 7 therefore displays the two distinct measured mean Average Precision (mAP) for each run, the first one (mAP1) considering only the foreground

specie of each test recording and the other (mAP2) considering additionally the species listed in the *Background species* field of the metadata. Note that different colors have been used to easily differentiate the methods making use of the metadata from the purely audio-based methods.

The first main outcome is that the two best performing methods were already among the best performing methods in previous bird identification challenges [3, 20] although LifeCLEF dataset is much bigger and somehow more complex because of the social dimension of the data. This clearly demonstrates the genericity and stability of the underlying methods. The best performing runs of MNB TSA group notably confirmed that using matching probabilities of segments as features was once again a good choice. In their working note [38], Lassek et al. actually show that the use of such Segment-Probabilities clearly outperforms the other feature sets they used (0.49 mAP compared to 0.30 for the OpenSmile features [17] and 0.12 for the metadata features). The approach however remains very time consuming as several days on 4 computers were required to process the whole LifeCLEF dataset.

Then, the best performing (purely) audio-based runs of QMUL confirmed that unsupervised feature learning is a simple and effective method to boost classification performance by learning spectro-temporal regularities in the data. They actually show in their working note that their pooling method based on spherical k-means actually produces much more effective features than the raw initial low level features (MFCC based features). The principal practical issue with such unsupervised feature learning is that it requires large data volumes to be effective. However, this exhibits a synergy with the large data volumes used within LifeCLEF. This might also explain the rather good performances obtained by the runs of Inria ZENITH group who used hash-based indexing techniques of MFCC features and approximate nearest neighbours classifiers. The underlying hash-based partition and embedding method actually works as an unsupervised feature learning method.

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. Automatic fish identification in videos is therefore of crucial importance, in order to estimate fish existence and quantity [50, 49, 47]. 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

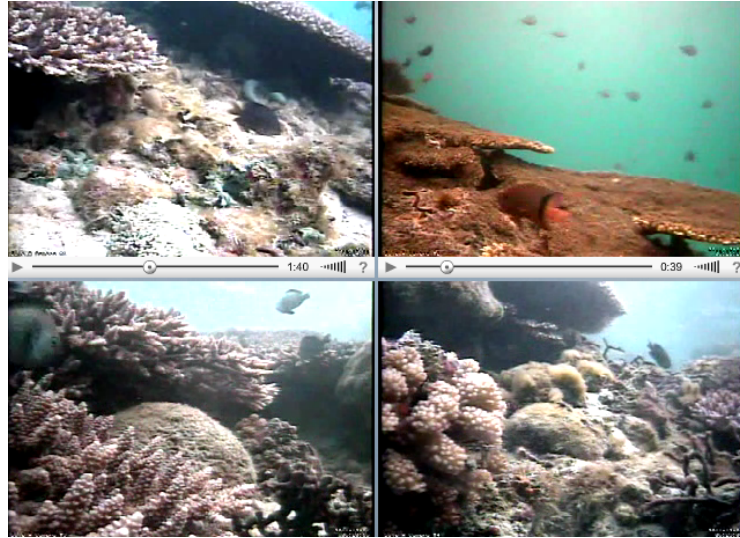


Fig. 8. 4 snapshots of 4 cameras monitoring the Taiwan’s Kenting site

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 entertainment industry (e.g. in aquarium).

To the best of our knowledge, this is the first worldwide initiative on automatic image and video based fish species identification.

4.2 Dataset

The underwater video dataset used for FishCLEF, is derived from the Fish4Knowledge²² 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 ²³. 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 8 shows 4 snapshots of 4 cameras monitoring the coral reef by Taiwan’s Kenting site and it illustrates the complexity of automatic fish detection and recognition in real-life settings.

More specifically, the FishCLEF dataset consists of about 3000 videos with several thousands of detected fish. The fish detections were obtained by pro-

²² www.fish4knowledge.eu

²³ <http://fishdb.sinica.edu.tw/>

cessing such underwater videos with video analysis tools [48] and then manually labeled using the system in [35].

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 all subtasks. Although tasks 2 and 3 are based on still images, participants are invited to exploit motion information extracted from videos to support their strategies.

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 PASCAL score between it and the corresponding object in the ground truth is over 0.7;
- Average recall and recall per fish species for the fish recognition subtask.

The participants to the above tasks will be asked to produce several runs containing a list of detected fish together with their species (only for subtask 3). When dealing fish species identification, a ranked list of the most probable species (and the related likelihood values) for each detected fish must be provided.

4.4 Participants and Results

About 50 teams registered to the fish task, but only two of them finally submitted runs: one, the I3S team, for subtask 3 and one, the LSIS/DYNI team, for subtask 4.

The strategy employed by the I3S team [9] for fish identification and recognition (subtask 3) consisted of, first, applying a background modeling approach based on Mixture of Gaussian for moving object segmentation. SVM learning using keyframes of species as positive entries and background of current video as negative entries was used for fish species classification. The results achieved by the I3S team were compared to the baseline provided by the organizers (ViBe [8] background modeling approach for fish detection combined to VLFeat BoW [56] for fish species recognition). While the average recall obtained by the I3S team was lower than the baseline’s recall, the precision was improved, thus implying that their fish species classification approach was reliable more than the

fish detection approach. On average More detailed results can be found in the working note of the task [14].

The LSIS/DYNI team submitted three runs for subtask 4 [30]. Each run followed the strategy proposed in [45] which, basically, consisted of extracting low level features, patch encoding, pooling with spatial pyramid for local analysis and a linear large-scale supervised classification by averaging posterior probabilities estimated through linear regression of linear SVM's outputs. No image specific pre-processing regarding illumination correction or background subtraction was performed. Results show that the method of LSIS/DYNI team clearly outperforms the baseline (VLFeat BoW [56]) and achieves near-perfect classification on several species. It is however important to note that the image-based recognition task (subtask 4) was easier than subtask 3 since (i) it didn't need any fish detection module (which is the most complex part in video-based fish identification) and (ii) only ten fish species were included in the ground truth.

5 Conclusions and Perspectives

With more than 120 hundreds research groups who downloaded LifeCLEF datasets and 22 of them who submitted runs, the pilot edition of LifeCLEF was a success showing a high interest of the proposed challenges in several communities (computer vision, multimedia, bio-acoustic, machine learning). The results of the plant and the bird tasks did show that very promising identification performances can be reached even with such an unprecedented number of species in the repsective training sets (i.e. 500 species for each task). This is clearly a good news with regard to the ecological urgency in building effective identification tools. However, we believe that some consistent progresses are still needed if we would like to use such tools for automatically monitoring real-world ecosystems. One of the key challenge is notably to deal with the long tail of species that are represented with much less images than the top-500 most common species that we targeted within BirdCLEF and PlantCLEF 2014. For the next camapigns, we will notably discuss the possibility of using the whole Pl@ntNet dataset that covers more than 5000 species but in which many species are represented with very few samples. Concerning the fish task, we believe that the main reason of the lower participation is its highest complexity. Video contents are actually much harder to manage and implies several difficult subtasks before being able to apply classical image classification techniques. Also, the cost of annotating the raw video contents makes it difficult to produce large-scale ground-truth and training data. But on the other side, this shows the importance of building automatic methods for processing such huge data.

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. Inria's participation at ImageCLEF 2013 Plant Identification Task. In *CLEF (Online Working Notes/Labs/Workshop) 2013*, Valencia, Espagne, 2013.
3. *Proc. of the first workshop on Machine Learning for Bioacoustics*, 2013.
4. A. Angelova, S. Zhu, Y. Lin, J. Wong, and C. Shpecht. Development and deployment of a large-scale flower recognition mobile app, December 2012.
5. E. Aptoula and B. Yanikoglu. Morphological features for leaf based plant recognition. In *Proc. IEEE Int. Conf. Image Process., Melbourne, Australia*, page 7, 2013.
6. 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.
7. 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.
8. O. Barnich and M. Van Droogenbroeck. Vibe: A universal background subtraction algorithm for video sequences. *Image Processing, IEEE Transactions on*, 20(6):1709–1724, 2011.
9. K. Blanc, D. Lingrand, and F. Precioso. Fish species recognition from video using svm classifier. In *Working notes of CLEF 2014 conference*, 2014.
10. 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.
11. 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.
12. 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.
13. Q. Chen, M. Abedini, R. Garnavi, and X. Liang. Ibm research australia at life-clef2014: Plant identification task. In *Working notes of CLEF 2014 conference*, 2014.
14. S. Conchetto, B. Fisher, and B. Boom. Lifeclef fish identification task 2014. In *CLEF working notes 2014*, 2014.
15. I. Dimitrovski, G. Madjarov, P. Lameski, and D. Kocev. Maestra at lifeclef 2014 plant task: Plant identification using visual data. In *Working notes of CLEF 2014 conference*, 2014.
16. O. Dufour, T. Artieres, H. GLOTIN, and P. Giraudet. Clusterized mel filter cepstral coefficients and support vector machines for bird song identification. 2013.
17. F. Eyben, M. Wöllmer, and B. Schuller. Opensmile: the munich versatile and fast open-source audio feature extractor. In *Proceedings of the international conference on Multimedia*, pages 1459–1462. ACM, 2010.
18. S. Fakhfakh, B. Akrouf, M. Tmar, and W. Mahdi. A visual search of multimedia documents in lifeclef 2014. In *Working notes of CLEF 2014 conference*, 2014.
19. K. J. Gaston and M. A. O'Neill. Automated species identification: why not? 359(1444):655–667, 2004.
20. H. Glotin and J. Sueur. Overview of the first international challenge on bird classification. 2013.
21. 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.

22. 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.
23. 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.
24. H. Goëau, H. Glotin, W.-P. Vellinga, and A. Rauber. Lifeclef bird identification task 2014.
25. H. Goëau, A. Joly, P. Bonnet, J.-F. Molino, D. Barthélémy, and N. Boujemaa. Lifeclef plant identification task 2014.
26. H. Goëau, A. Joly, S. Selmi, P. Bonnet, E. Mouysset, L. Joyeux, J.-F. Molino, P. Birnbaum, D. Barthelemy, and N. Boujemaa. Visual-based plant species identification from crowdsourced data. In *ACM conference on Multimedia*, pages 813–814, 2011.
27. H. Goëau, A. Joly, I. Yahiaoui, V. Bakić, and V.-B. Anne. Pl@ntnet’s participation at lifeclef 2014 plant identification task. In *Working notes of CLEF 2014 conference*, 2014.
28. 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.
29. M. Issolah, D. Lingrand, and F. Precioso. Plant species recognition using bag-of-word with svm classifier in the context of the lifeclef challenge. In *Working notes of CLEF 2014 conference*, 2014.
30. P.-H. Joalland, S. Paris, and H. Glotin. Efficient instance-based fish species visual identification by global representation. In *Working notes of CLEF 2014 conference*, 2014.
31. A. Joly, J. Champ, and O. Buisson. Instance-based bird species identification with indiscriminant features pruning - lifeclef2014. In *Working notes of CLEF 2014 conference*, 2014.
32. 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.
33. 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.
34. H. Karamti, S. Fakhfakh, M. Tmar, and F. Gargouri. Miracl at lifeclef 2014: Multi-organ observation for plant identification. In *Working notes of CLEF 2014 conference*, 2014.
35. I. Kavasidis, S. Palazzo, R. Salvo, D. Giordano, and C. Spampinato. An innovative web-based collaborative platform for video annotation. *Multimedia Tools and Applications*, pages 1–20, 2013.
36. H. Kebapci, B. Yanikoglu, and G. Unal. Plant image retrieval using color, shape and texture features. *The Computer Journal*, 54(9):1475–1490, 2011.
37. 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.
38. M. Lasseck. Large-scale identification of birds in audio recordings. In *Working notes of CLEF 2014 conference*, 2014.
39. 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.

40. R. Martinez, L. Silvan, E. V. Villarreal, G. Fuentes, and I. Meza. Svm candidates and sparse representation for bird identification. In *Working notes of CLEF 2014 conference*, 2014.
41. 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.
42. 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.
43. J. Northcott. Overview of the lifeclef 2014 bird task. In *Working notes of CLEF 2014 conference*, 2014.
44. D. Paczolat, A. Bánhalmi, L. Nyúl, V. Bilicki, and Á. Sárosi. Wlab of university of szeged at lifeclef 2014 plant identification task. In *Working notes of CLEF 2014 conference*, 2014.
45. S. Paris, X. Halkias, and H. Glotin. Sparse coding for histograms of local binary patterns applied for image categorization: toward a bag-of-scenes analysis. In *Pattern Recognition (ICPR), 2012 21st International Conference on*, pages 2817–2820. IEEE, 2012.
46. L. Y. Ren, J. William Dennis, and T. Huy Dat. Bird classification using ensemble classifiers. In *Working notes of CLEF 2014 conference*, 2014.
47. M. R. Shortis, M. Ravanbaksh, 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.
48. C. Spampinato, E. Beauxis-Aussalet, S. Palazzo, C. Beyan, J. Ossenbruggen, J. He, B. Boom, and X. Huang. A rule-based event detection system for real-life underwater domain. *Machine Vision and Applications*, 25(1):99–117, 2014.
49. 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.
50. 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 ACM ARTEMIS 2010*, pages 45–50. ACM, 2010.
51. D. Stowell and M. D. Plumbley. Audio-only bird classification using unsupervised feature learning. In *Working notes of CLEF 2014 conference*, 2014.
52. N. Sunderhauf, C. McCool, B. Uppcroft, and P. Tristan. Fine-grained plant classification using convolutional neural networks for feature extraction. In *Working notes of CLEF 2014 conference*, 2014.
53. G. Szűcs, P. Dávid, and D. Lovas. Viewpoints combined classification method in image-based plant identification task. In *Working notes of CLEF 2014 conference*, 2014.
54. M. Towsey, B. Planitz, A. Nantes, J. Wimmer, and P. Roe. A toolbox for animal call recognition. *Bioacoustics*, 21(2):107–125, 2012.
55. 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.
56. A. Vedaldi and B. Fulkerson. Vlfeat: An open and portable library of computer vision algorithms. In *Proceedings of the international conference on Multimedia*, pages 1469–1472. ACM, 2010.

57. H. Vincent Koops, J. van Balen, and F. Wiering. A deep neural network approach to the lifeclef 2014 bird task. In *Working notes of CLEF 2014 conference*, 2014.
58. Q. D. Wheeler, P. H. Raven, and E. O. Wilson. Taxonomy: Impediment or expedient? *Science*, 303(5656):285, 2004.
59. B. Yanikoglu, Y. S. Tolga, C. Tirkaz, and E. FuenCagliartes. Sabanci-okan system at lifeclef 2014 plant identification competition. In *Working notes of CLEF 2014 conference*, 2014.