Combining artificial intelligence services for the recognition of flora photographs: Uses in augmented reality and tourism

Guillermo Feierherd[1], Federico González[1], Leonel Viera[1], Rosina Soler[2], Lucas Romano[1], Lisandro Delía[3], Beatriz Depetris[1]

[1] Instituto de Desarrollo Económico e Innovación, Universidad Nacional de Tierra del Fuego Antártida e Islas del Atlántico Sur
Hipólito Irigoyen 880, 9410, Ushuaia, Tierra del Fuego, Argentina
{gfeierherd, fgonzalez, lviera, lromano, bdepetris}@untdf.edu.ar

[2] Centro Austral de Investigaciones Científicas, CONICET
Houssay 200, 9410, Ushuaia, Tierra del Fuego, Argentina
rosinas@cadic-conicet.gob.ar

[3] III-LIDI - Universidad Nacional de La Plata
50 y 120, 1900, La Plata, Buenos Aires, Argentina
ldelia@lidi.info.unlp.edu.ar

Abstract. Tourism information services are evolving rapidly. With Internet, tourists organize their trips by managing information before arriving at their destination. Nature is the main tourist attraction in Argentina. However, the information tools as field guides, have had few improvements in their digital version compared to printed ones. This work compares and combines machine learning services that includes deep learning, artificial intelligence and image recognition, to evaluate the app development for mobile phones that offer recognition of flora species in real time, in natural areas with low or no internet connectivity. Recognition of three *Nothofagus* tree species (with a dataset of 45 photos per species) were evaluated in the Tierra del Fuego National Park, using IBM Watson, Google Cloud and Microsoft Azure. Finally, we defined an algorithm combining those services to improve the results. Google Cloud was the service with the best performance recognizing all the tree species (83% effectiveness in average). The accuracy of Watson and Azure was lower than Google Cloud, and varied according to tree species. Combined algorithm improved the recognition with a 90% effectiveness in average. A next iteration of this work expects to increase the accuracy of recognition to get a total of 150 photos per specie into the dataset. We also expect to use assisted learning to improve the efficiency of the neural network obtained to know the adaptation capacities for each evaluated service.

Keywords: smart tourist destinations, computer vision, IBM Watson, Microsoft Azure, Google Cloud.

1 Introduction

1.1 Background

The present work is an extension of the one presented at the XXIV Argentine Congress of Computer Science (CACIC 2018) [1]. Both are part of the project "Virtual and Augmented Reality, Big Data and Mobile Devices: Applications in Tourism" that since 2017 is developed in the National University of Tierra del Fuego, Antarctica and South Atlantic Islands. The project seeks to reveal the uses that the tourism industry is making of these emerging technologies individually or in combination, to propose alternatives for application in the area of Tierra del Fuego.

In the previous work, machine learning and artificial intelligence services offered in the cloud by IBM and Microsoft were used in order to generate the bases for an application that would allow recognizing in real time the three most common Nothofagus species in the National Park of Tierra del Fuego. In addition, the benefits of the services compared with respect to the recognition of species were discussed.

In the current one, Google services have been added to the comparison and an algorithm has been proposed that, by combining the results of the three aforementioned services, substantially improves the recognition capacity.

1.2 Technical rationale

Considering that the main tourist attraction of Tierra del Fuego is wildlife (flora and fauna), tourists often enrich their trip with field guides, species guides or information brochures about the local flora and fauna. However, it would be much more interesting if they could recognize a species at the same moment of taking a photograph and in this context, the potential of augmented reality and the recognition of images associated with artificial intelligence, is promising.

Biological organisms are not always easy to photograph. Animals, whether they are walkers, fliers or swimmers, besides being mobile (and some of them very fast) are elusive to human presence or have cryptic behavior. Photographing plants, on the other hand, is much easier, accessible and attractive for anyone without technical knowledge of photography, even when they have only a smartphone. Moreover, throughout the world, the proportion of identified plant species is greater than any other biological group [2].

Despite this greater knowledge about plant species, in some cases the interspecific morphological differences are minimal and difficult to recognize with the naked eye. Even in herbaria that contain a large number of specimens, their identification is being carried out with artificial vision and machine learning approaches applied to scans of leaves or images of plants in the field [3]. The trees of the genus *Nothofagus* are the most representative of the forest of Tierra del Fuego, but the specific differences of their leaves are difficult to recognize when comparing them without expert human eye, which represents an ideal difficulty to test the augmented reality together with computer vision.

Considering the complexity and constant evolution of the technologies related to machine learning, deep learning, artificial intelligence and image recognition, we decided to use the services provided in the cloud to train the neural network from a database of photos to identify tree species typical of the forest of Tierra del Fuego. The ultimate goal is to use that knowledge and to develop an offline app in the future to recognize flora and fauna without connectivity, which is a peculiarity of the protected areas.

2 Information in smart tourist destinations

Tourist information systems have traditionally been organized in three chronological stages: promotion, planning and stay. In this way, the person who deals with the management of a tourist destination, first makes marketing, then provides information tools to plan the itinerary and, finally, guides the tourists with useful information to help them know the destination during their stay. In recent years, the concept of stay was reoriented to that of experience and a fourth component was added to the stages: sharing what was lived.

This work focuses on the stage of the experience, particularly when it takes place in Protected Areas. In those cases, the traditional methods to provide information were based on printed brochures (which could be obtained at a visitor center), on strategically located posters, signage of the trails, etc.

This scheme, with its advantages and disadvantages, but the only one possible until a few years ago, must be complemented and enriched with the resources that technology offers today. This also allows adapting them to the characteristics of younger visitors, digital natives, accustomed to using new technological resources to obtain the information they need.

Doug Lansky [4] questions the role of visitor centers. Their questions take into account the investment that many of them represent and the increasingly scarce utility of their services. In his opinion its strategic location makes many tourists visit them with the idea of collecting some free maps and brochures, or maybe get an air-conditioned space, shelter from rain or cold, or the need for toilets. But, if they were not there, that does not mean tourists could not get the same information services using their phones and other mobile devices.

Obviously there will be some people who do not carry a smartphone with them, but they are becoming a limited minority. And it is likely that the "analogue" visitors will already arrive prepared with a printed guide, newspaper clippings or the advice they have requested from the hotel receptionist.

It is known that smartphones have been replacing several useful old devices for tourists, such as photo cameras, video recorders, GPS among others. It is also true that many visitors consider the Protected Areas as a symbol of nature and feel that it is worthwhile to remain "disconnected" while they go through them.

People over 50 only use the telephone (Fig. 1) to perpetuate moments and understand the place -take pictures, look at a map- while younger generations incorporate communication tools to interact with social networks and for other recreational uses such as listening music while they visiting the protected area [5].

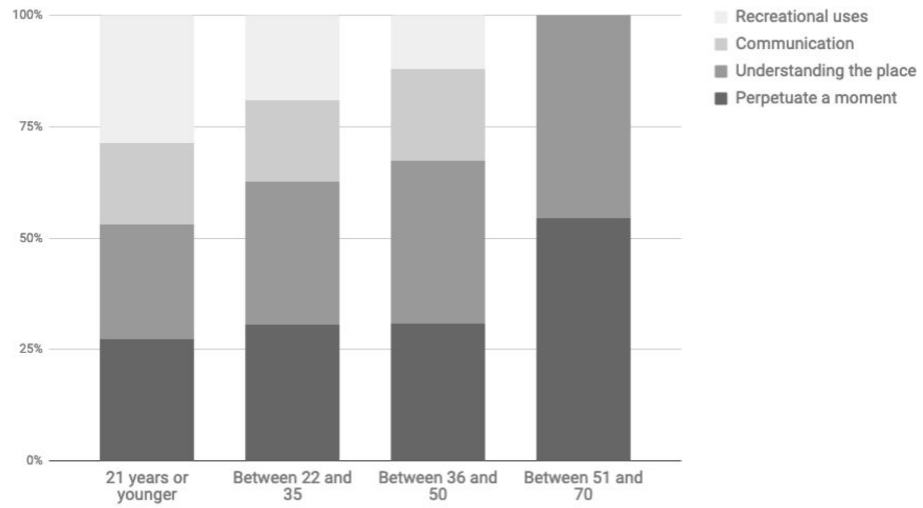

Fig. 1. Uses of the smartphones according to generations grouped by activity [4].

The biggest difference between young and old generations is linked to recreational use and communication (with the outside world of the protected area). This mean that older generations are prone to disconnect themselves within protected areas, while digital natives prefer to remain connected and give intensive use to all the analyzed functions of the telephone when visiting protected areas.

From the above, it is clear that interactive identification of species using a smartphone can be a motivation against traditional brochures or guides with descriptions and figures.

3 Methodology

3.1 Morphological aspects of the compared species

For this work three tree species of the genus *Nothofagu*s were selected, which predominate in the Fuegian forest: lenga (*N. pumilio*), ñire (*N. antarctica*) and guindo (*N. betuloides*). The first two are deciduous and phylogenetically closest species [6], so hybridization between them is possible [7]. While the third species is an evergreen tree with morphological characteristics that differentiate it from its congeners. The main foliar characteristics of these trees (Table 1) allow to recognize with an ordinal eye, with a minimum of botanical knowledge, what kind of species they are [8] [9].

Table 1. Main foliar characteristics of the *Nothofagus* species considered in the study

Foliar characteristic	Lenga	Ñire	Guindo
Habit	deciduous	deciduous	evergreen
Shape	oblong-elliptical	oblong	oval
Length	2-3 cm	2-3 cm	1-2.5 cm
Width	1.5 cm	1.5 cm	1 cm
Apex	blunt	rounded	sharp
Edge	crenate	irregular lobed	regular serrated
Base	slightly asymmetric	asymmetric	symmetric
Petiole length	0.5 cm	0.5 cm	0.3 cm

3.2 Visual recognition services

Neural networks have proven effective at solving difficult problems but designing their architectures can be challenging, even for image classification problems alone. Different provider services aim to minimize human participation, employing evolutionary algorithms to discover such networks automatically [10]. Despite significant computational requirements, it is now possible to develop models with high accuracies.

Different complex artificial intelligence and machine learning services are being offered in the cloud by companies such as IBM, Microsoft, Google or Amazon. With different levels of complexity regarding its implementation, this paper compares the services of IBM Watson (Visual Recognition), Google Cloud (Vision), and Microsoft Azure (Custom Vision), since they are the only ones that incorporate integral products based on visual platforms that facilitate its implementation in great measure.

3.3 Collection of photographs

For each species of *Nothofagus* trees, 45 photographs of leaves or branches with different level of clearly differentiable leaves were provided, to be used as machine learning models. The photos were taken at different seasons of the year (except winter) by people without specific knowledge of photography, but with knowledge of plant species.

In addition to the datasets created and categorized, a stock of 30 photos (10 for each specie) that did not belong to the database and were used to verify the certainty of the results offered by Watson, Google and Azure was established.

The quality of this 30 photographs (Table 2) was variable to represent different situations into the database, all of them in jpg format.

Table 2. Main characteristics of photographs for each species used to verify the model for this study. Nb: guindo (*Nothofagus betuloides)*, Np: lenga (*N. pumilio)*, Na: ñire (*N. antarctica)*

Photo	Leaves/Branches	Noise	Human recognition	Photo quality	Colour
Nb1	35/10		easy	good	green
Nb2	70/10		easy	good	green
Nb3	50/5	fruits	easy	medium	green
Nb4	45/2		easy	good	green
Nb5	30/10	gall	easy	good	green
Nb6	60/5		easy	good	green
Nb7	80/10		easy	good	green
Nb8	50/15	fence	easy	good	green
Nb9	55/10		medium	good	green
Nb10	20/5	stems	easy	medium	green
Np1	85/2		easy	good	green
Np2	50/10		easy	good	yellow
Np3	20/5		easy	bad	green
Np4	60/15		easy	good	green
Np5	80/2		medium	medium	red
Np6	40/5		tricky	good	green
Np7	10/2		tricky	good	green
Np8	75/5		easy	good	green
Np9	60/2		easy	good	green
Np10	55/20		medium	good	green
Na1	60/10		tricky	bad	yellow
Na2	35/10		easy	good	green/yellow
Na3	80/0		easy	medium	green
Na4	75/2		easy	bad	green
Na5	55/5		easy	good	green
Na6	50/20	fruits	easy	good	green
Na7	75/2		easy	good	green
Na8	45/15		easy	good	green
Na9	20/15	flowers	easy	good	green
Na10	15/20	hand	easy	good	green

3.4 Combination of recognition services

After having generated the neural networks in each platform, the results obtained individually with each service were combined, with the aim of evaluating whether the results improved. A simple algorithm was developed which is described below:

1. If all 3 predictions matched, that result was considered valid.
2. If only 2 predictions matched (i.e., the majority), that result was considered valid.

3. If the predictions did not matchet, the one with the greatest accuracy reported was considered valid. This, as long as that accuracy is greater than the lower confidence reported by the models.
4. Otherwise the result was indeterminate and the species unknown.

4 Results

In general terms, the three services had good results when evaluating guindo and ñire. With lenga the case was different since Watson and Azure had more problems than Google Cloud.

Fig. 2. Confusion Matrix with recognition probability of tree species between Watson, Azure and Google Cloud services. Match cases are highlighted in gray over the diagonal. Nb: guindo, Np: lenga, Na: ñire

	Watson			Azure			Google		
	Nb	Np	Na	Nb	Np	Na	Nb	Np	Na
Guindo (Nb)	90%	0%	10%	80%	0%	20%	90%	0%	10%
Lenga (Np)	20%	40%	40%	20%	40%	40%	0%	70%	30%
Ñire (Na)	20%	0%	80%	30%	0%	70%	20%	0%	80%

Fig. 2 summarizes the results obtained with a confusion matrix. The diagonal marked in gray shows the percentage of hits for each platform, while the white cells indicate the degree of error and what was the confusion. For example: when evaluating lenga (Np), Azure confused 20% of the samples with guindo (Nb); Something similar happened to Google with ñire and guindo.

A quick analysis of the confusion matrix makes it easy to see that Google Cloud was the service with the best performance. It equaled Watson with a 90% effectiveness when analyzing guindo images, was highly superior to Watson and Azure in evaluating lenga (70% against 40% in the other two platforms), and finally equaled Watson by recognizing ñire correctly in 80% of cases.

Fig. 3. One matrix has been added to preliminary results, combining those models with the described algorithm.

	Watson			Azure			Google			Combined		
	Nb	Np	Na	Nb	Np	Na	Nb	Np	Na	Nb	Np	Na
Guindo (Nb)	90%	0%	10%	80%	0%	20%	90%	0%	10%	100%	0%	0%
Lenga (Np)	20%	40%	40%	20%	40%	40%	0%	70%	30%	10%	70%	20%
Ñire (Na)	20%	0%	80%	30%	0%	70%	20%	0%	80%	0%	0%	100%

The combination of the three models improved the individual results (Fig. 3). Analyzing the results obtained, it can be seen that the performance for guindo and ñire reached 100% effectiveness, while for lenga it was not possible to improve the results previously achieved with Google Cloud.

5 Conclusions

Watson and Azure services had similar performance, but Google was more successful in recognizing two of the three analysed tree species. Guindo (Nb) and ñire (Na) seem to have more defined morphological characteristics, while lenga (Np) had intermediate and similar morphological characteristics to other species and this generated high confusion for the systems. A previous work [10] compared just Watson and Azure, and the service of IBM was more accurate than Microsoft. This agree with our findings, but when we added Google Cloud service in the comparison we got better results.

Considering that the knowledge base was generated from 45 photographs for each species, we believe that the results have been highly satisfactory. A next iteration of this work will add 315 more *Nothofagus* photographs to be included into the dataset to compare the evolution capacity that each service has to improve their models. We expect to add 105 photos for each species, so the database will be a set of 150 photos per species. Regarding this last point, it is important to refine the recognition capabilities of the machine learning for the current database (but mainly for lenga), manually correcting the errors of the first iteration and testing a new group of photographs. At that time, the combination algorithm will be applied again to evaluate its potential in a more refined model than the current one.

This work focused on the interpretation of images in the foreground (leaves, branches) since for tourist purposes it is of great value for the visitors to be able to know what they have at their fingertips. Beyond this objective, it would also be interesting and possible in future studies to broaden the point of view towards the interpretation of specimens at medium distances to answer questions like "what kind of trees are those?" or to count "how many individuals of each species are there?" in a determined group.

It is necessary to incorporate new species in the training process in the near future to generate a database with knowledge about local flora of high interest for tourists. The three species studied in this work are the most representative trees of the Tierra del Fuego National Park, and also the best known by people. The recognition of other plants as forbs, grasses, or ferns is a major challenge from a technical point of view, since a greater variety of species very similar among them could potentially generate more confusion in the recognition or less accuracy of results.

Finally, and as final outcome of our project “Virtual and Augmented Reality, Big Data and Mobile Devices: Applications in Tourism” we expect to develop a public offline app that incorporates image recognition and offers results from the compared platforms, while generating at the same time new databases to refine knowledge with assisted machine learning. Offline mode is a key feature for the usefulness of apps in countries such as Argentina, where the main destination of international tourism are the National Parks (Perito Moreno Glacier, Iguazú Falls). Although Argentina is one of the countries with higher access to the internet in Latin America, such connectivity occurs in cities or urban areas where population is concentrated [11]. The protected areas do not have any type of internet connection and in most cases, no access to mobile phone network [12]. To achieve this, the application must be able to identify the species offline without external services. With the previous training of a neural network based on machine learning services, we hope to have enough potential to achieve it.

References

1. Feierherd, G., González, F., Viera, L., Soler, R., Romano, L., Delía, L., Depetris, B.: Comparison of services for the recognition of flora images. Uses in augmented reality and tourism. In: Proceedings of the XXIV National Congress of Computer Science, pp. 1152-1159. 8-12th October, Tandil, Buenos Aires, Argentina (2018)
2. Mora, C., Tittensor, D.P., Adl, S., Simpson, A.G.B., Worm, B.: How many species are there on earth and in the ocean? PLoS Biol. 9(8): e1001127 (2011)
3. Carranza-Rojas, J., Goeau, H., Bonnet, P., Mata-Montero, E., Joly, A.: Going deeper in the automated identification of herbarium specimens. BMC Evolutionary Biology 17:181 (2017)
4. Lansky, D.: Has The Smartphone Killed The Tourist Office? Digital Tourism Think Tank, (2016) http://thinkdigital.travel/opinion/has-the-smartphone-killed-the-tourist-office/
5. González, F.: Digital tourist information tools in Protected Areas for Millennials and Gen Z. Master Thesis, University of Girona, Spain (2017)
6. Jordan, G., Hill, R.S.: The phylogenetic affinities of *Nothofagus* (Nothogaceae) leaf fossils based on combined molecular and morphological data. Int. J Plant Sci. 160: 1177-1188 (1999)
7. Burns, S.L., Cellini, J.M., Lencinas, M.V., Martínez Pastur, G.J., Rivera, S.M.: Descripción de posibles híbridos naturales entre *Nothofagus pumilio* y *N. antarctica* en Patagonia Sur (Argentina). Bosque: 31: 9-16 (2010)
8. Moore, D.M.: Flora of Tierra del Fuego. Anthony Nelson, England (1983)
9. Donoso, C.: Las especies arbóreas de los bosques templados de Chile y Argentina. Autoecología. Marisa Cuneo Ediciones, Chile (2006)

10. Real, E., Moore, S., Selle, A., Saxena, S., Suematsu, Y., Tan J., Le, Q., Kurakin, A.: Large-Scale Evolution of Image Classifiers. Proceedings of the 34th International Conference on Machine Learning, Sydney, Australia, PMLR 70, (2017)
11. Open Signal: Global cell coverage maps, Argentina (2018) https://opensignal.com/networks?z=4&minLat=-56.9&maxLat=-16.6&minLng=-109.7&maxLng=-18.3&s=&t=4
12. APN: Mapa oficial con polígonos de las Áreas Protegidas Nacionales de Argentina. Administración de Parques Nacionales (2018) http://mapas.parquesnacionales.gob.ar/layers/geonode%3Aapn_areasprotegidas_01