APPLYING PHOTOGRAPHS FILTERS TO IMPROVE PEOPLE AND OBJECTS RECOGNITION USING AN API

Oana CĂPLESCU [1*]

Costin-Anton BOIANGIU [2]

ABSTRACT

In a world where computer science started evolving so much that artificial intelligence is recognizing objects and people's reactions from a photo, relying on the technology makes surviving in a modern world so much easier. The most important step in automatic content recognition is the quality of the image - quality which can be improved in many ways. Having a high contrast level provided in a photograph is a characteristic that ensures that more objects would be recognized because Microsoft's Computer Vision API identifies content in the provided photographs using the shapes of the targets. It is clear that making an image's objects more "visible" by applying various effects to it will increase the algorithm's recognition.

KEYWORDS: *API, visual recognition, Computer Vision, voting system, knowledge library*

1. INTRODUCTION

Using the cloud-based Computer Vision API everyone can analyze data provided by the API. It also "provides developers with access to advanced algorithms for processing images and returning information" [1]. The API's "algorithms can analyze visual content in different ways based on inputs and user choices" [1]

    A. Faces

The computer vision algorithms can focus on many kinds of output information depending on the needed results. One of these algorithms [2] can detect human faces and analyze their characteristics like age (approximated using face gestures techniques), gender and displays the rectangle of the face - for pictures containing multiple persons.

All of this visual output is actually a subset of metadata generated for each face to describe its content.

The output for the following image is: ["age": 28, "gender": "Female", "faceRectangle": "left": 447, "top": 195, "width": 162, "height": 162 , "age": 10, "gender": "Male", "faceRectangle": "left": 355, "top": 87, "width": 143, "height": 143]

[1*]   corresponding author, Engineer, "Politehnica" University of Bucharest, Bucharest, Romania, oana.caplescu@gmail.com

[2] Professor PhD Eng., "Politehnica" University of Bucharest, Bucharest, Romania, costin.boiangiu@cs.pub.ro

Figure 1. Example of a picture's analysis resulted from Computer Vision API

B. Tagging images

The main point of this paper's scope is image tagging, which refers to the API's "understanding" (or better said matching) of various objects found in the provided images. The tagging consists of two main steps:

- Matching the object with its knowledge library: for an input image, Computer Vision API returns tags "based on more than 2000 recognizable objects" [3] for example: living beings, scenery, house belongings and actions.
- On each matching, the algorithm provides a confidence grade that describes how sure it is regarding the recognition of the elements inside the current tag.

Generally, the information provided by the algorithm has a confidence grade around 95-99%, but of course there are images which cannot be recognized at that level.

For example, on the following image not all the information is graded with a 99% confidence grade. The Computer Vision API outputs for the following picture: "tags": ["train", "platform", "station", "building", "indoor", "subway", "track", "walking", "waiting", "pulling", "board", "people", "man", "luggage", "standing", "holding", "large", "woman", "yellow", "suitcase"], "captions": ["text": "people waiting at a train station", "confidence": 0.8331026].

Figure 2. On each tag, a confidence information is also added which describes (as a percentage) how sure the algorithm is regarding its prediction.

The human operator can see and understand that in this photo the man with the little girl are not waiting for a train but are rather getting out of the metro platform. The reason why the algorithm suggested that they are waiting for the train (with a confidence grade of 83%) is that usually on a subway platform the people are waiting for the train or that in all of its previous pictures people were waiting for a train.

This is because the algorithm, based on artificial intelligence, is being continuously trained by the images it's processing, becoming more accurate after each category of photos.

According to Microsoft's documentation [3], "When tags are ambiguous or not common knowledge, the API response provides 'hints' to clarify the meaning of the tag in context of a known setting." Unfortunately, it appears that at this point English is the only supported language for image description.

C. Optical Character Recognition - OCR

Computer Vision also provides output (and thus support) for Optical Character Recognition technology which detects the text from input images and also extracts the identified text into a machine-readable character stream, which can be used in many ways.

According to Microsoft's documentation [4], "OCR supports 25 languages".

The supported languages are: "Arabic, Chinese Simplified, Chinese Traditional, Czech, Danish, Dutch, English, Finnish, French, German, Greek, Hungarian, Italian, Japanese, Korean, Norwegian, Polish, Portuguese, Romanian, Russian, Serbian (Cyrillic and Latin), Slovak, Spanish, Swedish, and Turkish."

Microsoft based their advancement in development of these algorithms on output but they did not use voting as a form of analyzing the data.

2. VOTING PROCESS

To increase the optimal upshot, these voting systems have been used:

- majority voting: efficient, and it does not require a large set of data to be analyzed; its rule is: every vote has a fixed weight and a fixed probability of occurrence - considering the minimum confidence which is set at the beginning of the voting and a delta value which confirms if the confidence obtained after applying an effect is considered for the voting process as an input or not - considering that each effect produces its own result and improves its own recognized information within the photos.

- weighted voting: needs knowledge about the original pictures (considers previous inputs); in this case, the criteria is established as follows: if the output's confidence is higher than the original (thus the need from the previous input) it is considered for the voting process.

3. METHOD DESCRIPTION

Several processing based on voting have been developed over time [9][10][11][12]. The integrated ideas are, of course, specific to every application desired outcome, but there are also several key elements in common: the voting is performed on a processed version of the input and any of the used processing methods in itself cannot guarantee optimal results.

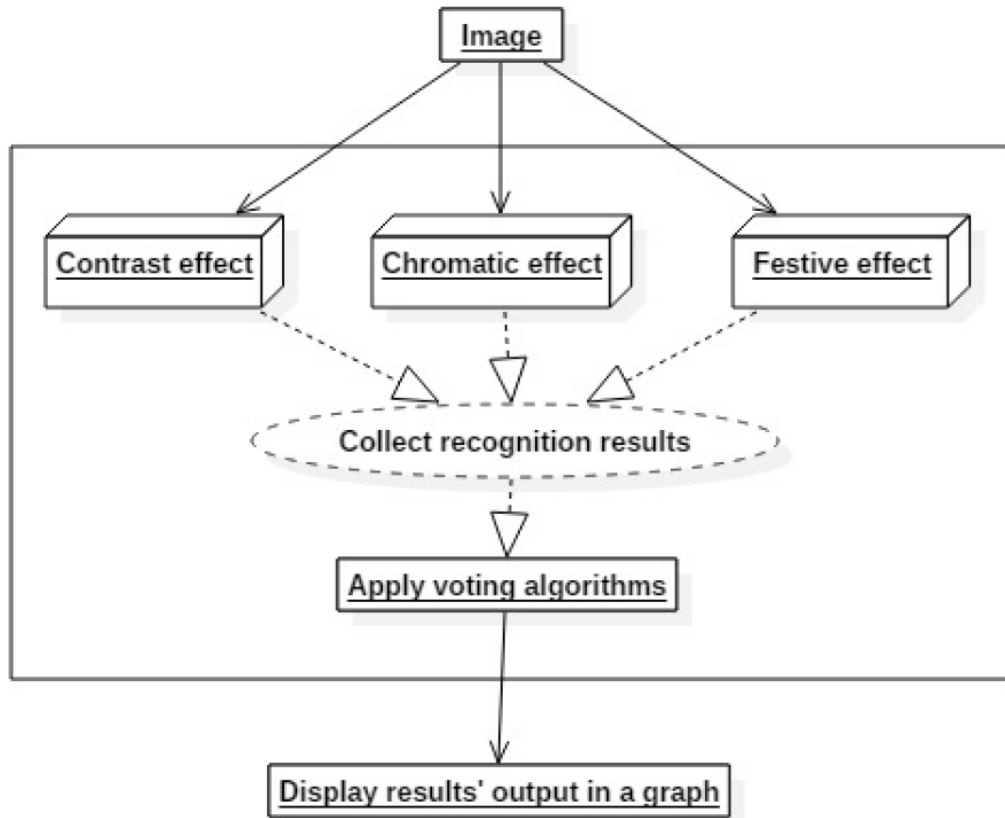The analyzing process is further explained in the following diagram:

Figure 3. Voting process from the initial image until the graphs which describe the improvement of the effects applied on the photos

A. Analyzing process

The API functions receives an image as an input (image that needs to have at most 4Mb), it starts analyzing it and displays in the form of JSON file, using tags, the following information:

- Description: an array of tags, generally describing the image (such as "outdoor", "water", "tree" etc.), captions, also describing the image (like "text": "a beach with palm trees and a body of water"), and the most important factor the "confidence" which is a real number, used instead of a percentage: 0.984310746)
- Tags: describe the objects found in the picture, in pairs of "name" and "confidence" values.
- Image format, image dimensions
- Black and white flag: Boolean value
- Dominant colors: for background, foreground, accent

Each information provided by the algorithms has a confidence value assigned to it which describes how confident is the algorithm of the information provided.

1) Contrast increase: one of the most used effects that is added is the increase of contrast, which makes the output image a lot more clearer than the original input, therefore making it a lot more easy to understand (from both human eye's point of view but also from the Computer Vision's understanding), and the impact is that the number of recognized objects is increased

2) Chromatic effect: another important aspect for Microsoft's Computer Vision analysis is color. Because of the fact that objects are searched in images based on past recognition, the algorithms analyze color: a beach is either white to light brown, water is turquoise, green or blue shaded - all of these are patterns which train the algorithm on how to better recognize elements and objects.
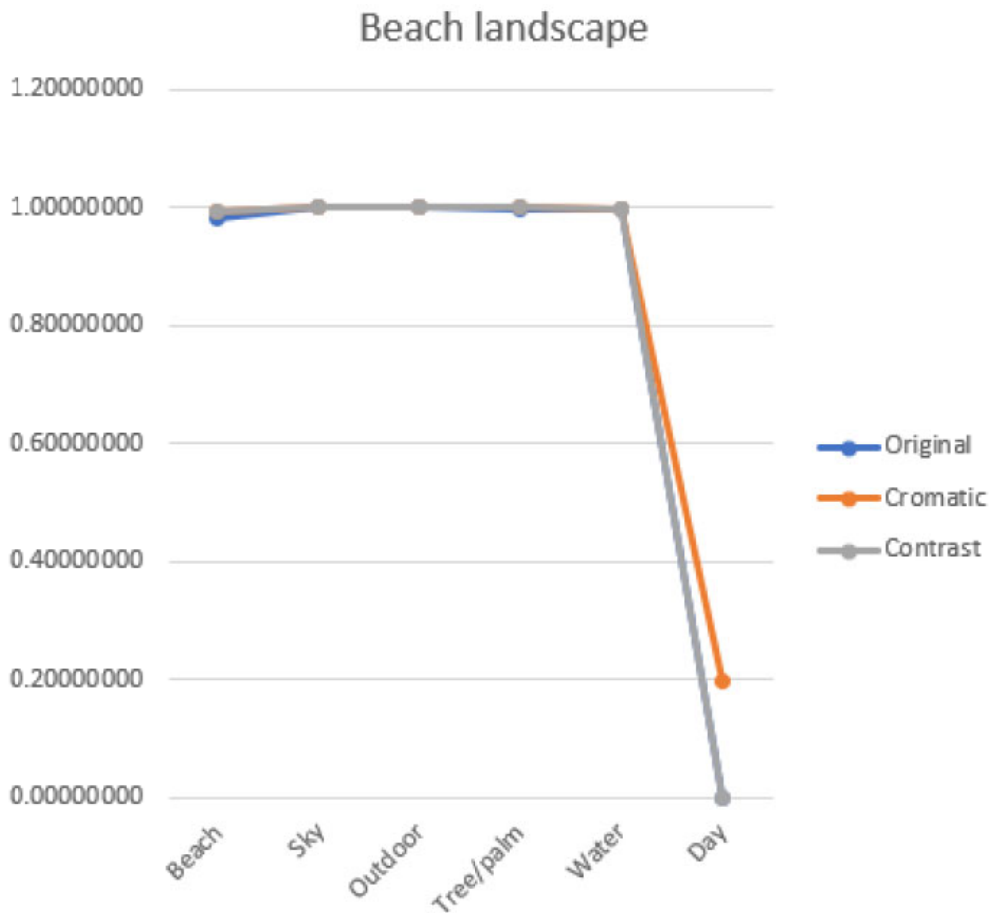
4. RESULTS

Figure 4. Voting process on a beachside photograph on which all above effects have been applied

In this case it is very important to observe that unlike the original photo and the chromatic effect, in the picture that had the contrast effect applied on the algorithm recognized another element: the fact that is a daytime picture. Of course this can only be voted as best result in the majority voting as it does not have a bigger confidence than 0.95 (value which was chosen for weighted voting).

Considering a photo with a higher quality (about 10 times higher than the previous one - 2Mb) it can be easily seen that the first three recognized objects have a confidence close to the ideal value - the 0.95 for majority - while still remaining bigger than it, whereas the only picture in which the furniture wasn't recognized is the original one (which indeed was a bit blurry - it becomes obvious that the applied effects improved this photograph).
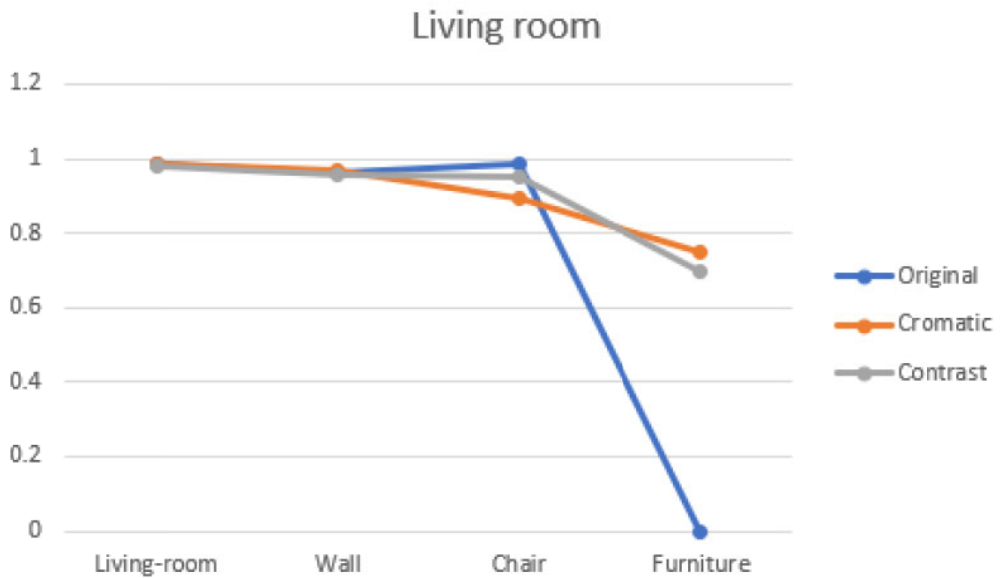

Figure 5. Voting process on a living room photograph on which all above effects have been applied

It can be seen that each face was easily recognized as either a man's or a female's face, yet the ages (which are not visible in the current photo because the space for these was too small for the algorithm) were more exact than in the original photo. Unfortunately not much else was recognized - as an example trees or other scenery objects. It can also be noticed that the boy's wasn't recognized, but this was also not recognized in the original photo, although it is safe to say that this iteration has a higher description confidence (0.9885462) than the original - 0.9802546. Another important aspect is that the age of the people in the photo was more accurate after applying both effects (comparing to applying only the contrast or only the chromatic one).
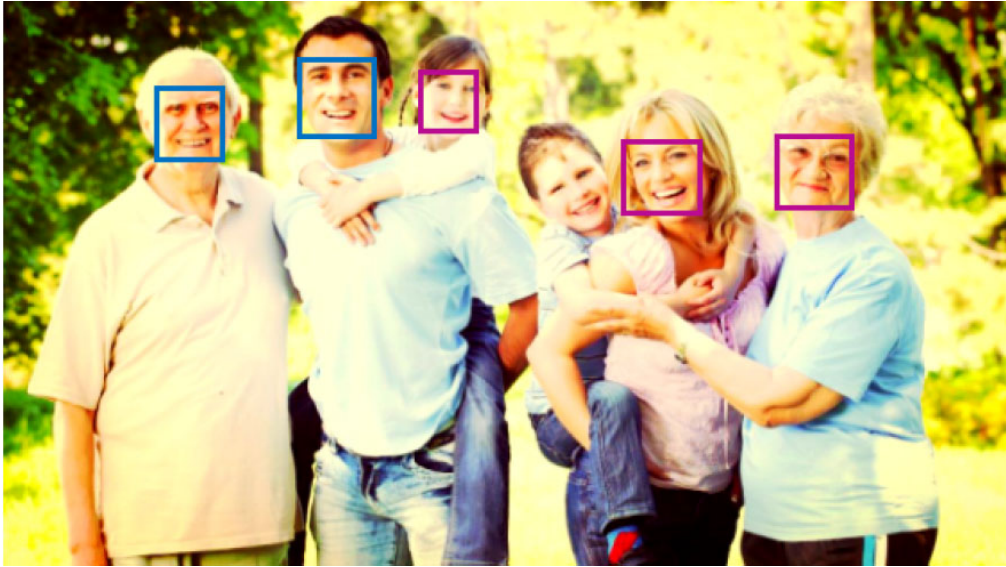
Figure 6. Output on both contrast added and chromatic effect on people's face recognition using Microsoft's Computer Vision

5. FUTURE WORK

After experiencing very good results on applying a number of effects on a single photo (such as contrast, saturation and highlights combined with another complete effect – like chromatic [5] or Ludwig [6]) it can be concluded that in order to achieve a better recognition the image's resolution needs to be as high as possible after the effects have been applied - considering that the effects applied using Canva Editor [7] and Be Funky photo effects [8] are deteriorating the image's quality.

As future work a good approach would be to try Photoshop in applying these effects as the quality won't get deteriorated and visible impact on Computer Vision's recognition is expected.

6. CONCLUSIONS

As far as the chromatic effect's results were confident at first point, we can concur that the output isn't of that much impact - being an important part of the weighted voting process, majority voting couldn't be concluded as the confidence exposed by this effect didn't offer any results whatsoever.

From the contrast effect's "point of view", results were clear and defined in both the weighted voting and majority voting, which concludes an important improvement in Computer Vision's algorithms results.

REFERENCES

[1]   Computer Vision API - Microsoft documentation, Available at: https://docs.microsoft.com/en-us/computer-vision, Accessed on 1 March 2018

[2]   Detecting and analyzing faces - Microsoft documentation, Available at: https://docs.microsoft.com/en-us/computer-vision/faces, Accessed on 1 March 2018

[3]   Tagging images - Microsoft documentation, Available at: https://docs.microsoft.com/en-us/computer-vision/tagging, Accessed on 1 March 2018

[4]   Detecting and analyzing printed text found in images – Microsoft documentation, Available at: https://docs.microsoft.com/en-us/computer-vision/ocr, Accessed on 1 March 2018

[5]   How do I replicate the Chromatic effect of the Instagram filter, Available at: https://www.youtube.com/chromatic-effect, Accessed on 1 March 2018

[6]   How do I replicate the Ludwig Instag effect of the Instagram filter, Available at: https://www.quora.com/Instagram-filter-Ludwig, Accessed on 1 March 2018

[7]   Canva - Photo Editor, Available at: https://www.canva.com/photo-editor/, Accessed on 1 March 2018

[8]   Canva - Photo Editor, Available at: https://www.befunky.com/photo-effects/, Accessed on 1 March 2018

[9]   Costin-Anton Boiangiu, Radu Ioanitescu, Razvan-Costin Dragomir, "Voting-Based OCR System", The Proceedings of Journal ISOM, Vol. 10 No. 2 / December 2016 (Journal of Information Systems, Operations Management), pp 470-486, ISSN 1843-4711.

[10]   Costin-Anton Boiangiu, Mihai Simion, Vlad Lionte, Zaharescu Mihai – "Voting Based Image Binarization" - , The Proceedings of Journal ISOM Vol. 8 No. 2 / December 2014 (Journal of Information Systems, Operations Management), pp. 343-351, ISSN 1843-4711.

[11]   Costin-Anton Boiangiu, Paul Boglis, Georgiana Simion, Radu Ioanitescu, "Voting-Based Layout Analysis", The Proceedings of Journal ISOM Vol. 8 No. 1 / June 2014 (Journal of Information Systems, Operations Management), pp. 39-47, ISSN 1843-4711.

[12]   Costin-Anton Boiangiu, Radu Ioanitescu, "Voting-Based Image Segmentation", The Proceedings of Journal ISOM Vol. 7 No. 2 / December 2013 (Journal of Information Systems, Operations Management), pp. 211-220, ISSN 1843-4711.