

Restoration of images of social networks that have undergone processing, on the example of the social network Instagram

M I Khotilin¹, N S Kravtsova¹, A V Kupriyanov^{1,2} and R A Paringer^{1,2}

¹Samara National Research University, Moskovskoe Shosse 34A, Samara, Russia, 443086

²Image Processing Systems Institute of RAS - Branch of the FSRC "Crystallography and Photonics" RAS, Molodogvardejskaya street 151, Samara, Russia, 443001

e-mail: turbomax.94@yandex.ru, natalyroom483@gmail.com, rusparinger@gmail.com

Abstract. This paper is devoted to finding the source images for the processed images of social networks. Existing methods and approaches that take place in the performance of this task are considered. The algorithms of image recovery is investigated and planned for finalization.

1. Introduction

In the modern world it is continuously generated a huge amount of data, whether the data received from the satellite, or sensors in the aircraft, bank transactions, patient diagnostic data, etc. A special place is occupied by social networks. The significance of social networks is due to the fact that, on the one hand they are the subject of socialization of people, and on the other - the most powerful and affordable political, ideological and economic instrument. A number of papers are dedicated to researches of social networks as systems, which contain large volumes of data.

Images have always occupied an important place in the life of people: whether it be cave paintings in primitive society, paintings by great artists, and especially photographs. With the advent of social networks, photos began to move from the category of personal to the category of public domain and became not only a new way of self-presentation, but also gave rise to other ways of thinking and vision [1]. People literally convey their individuality through photos: express their opinions, upload so-called "selfies", arrange discussions, share impressions and emotions in comments to photos or by color processing photos using various editors or filters. Social network "Instagram" began to play a huge role in the life of modern society. Data and images from it helps in various areas of life, for example, the police can solve crimes through photo data, psychologists can study various trends in the lives of modern people.

The color characteristic is one of the most significant characteristics describing the image. With it, you can extract various data that can be useful in many areas of human activity, such as psychology and sociology. In 2015, the Laboratory for the Study of Big Data, under the leadership of Lev Manovich, collected and analyzed 2.3 million photographs from the Instagram social network of large cities such as New York, Tokyo, Bangkok, San Francisco and eleven other global cities. With the help of the analysis conducted by the laboratory, it was possible to see the "biological rhythms" for which

the metropolis lives, as well as the life of a resident of one city differs from that of another, as the time frame and much more differ [2].

Also, the extracted data can be useful for example in medicine, in object selection and recognition of objects and in many other areas.

Image restoration is one of the most important tasks of computer vision and image processing. It finds its solution, for example, when detecting various kinds of tumors in medicine, when tracking objects in security systems, when restoring images, and in many others. For image restoration, various methods are used, both classical and modern, using work with neural networks and big data processing tools, or BigData.

In the future it is planned to transfer the implementation to such services as Apache Flume for data collection and Apache Spark for data processing as well as use Apache Cassandra for data storage.

This paper describes a method for finding the original image by determining the filter characteristic with which it was processed.

2. Getting the source data for image recovery

During the work, it was decided to use the social network Instagram, for analysis. Users of this social network often post photos and images that have been previously processed by various filters.

To obtain the original data, as a first step it was decided to write software that allows access to the social network through special API and obtain the necessary data from it, namely the image itself and additional service information containing the description and so-called tags indicating the filters used. To gain access to the images, it is necessary to authenticate, which requires special access keys. Authentication occurs through OAuth protocol version 2.0. This protocol is an open authorization protocol that allows limited access to protected resources to a third party (our software system), without the need to transfer a username and password to a third party (complex). After passing authorization and authentication, it becomes possible to access images and their descriptions and service information, and to download it to the computer as files.

3. Restoration of original images

Each image is a set of pixels characterized by its color brightness values and the whole image can be viewed as a three-layer table, each layer of which is a table consisting of pixels with specific brightness values of red, green, and blue. Images processed using special filters have a color-brightness distribution different from the original image.

In the course of the work, an educational sample of images was prepared, containing 100 pairs of images - the original and processed by a certain type of filter, as well as its name. An example of image data is shown in figure 1.

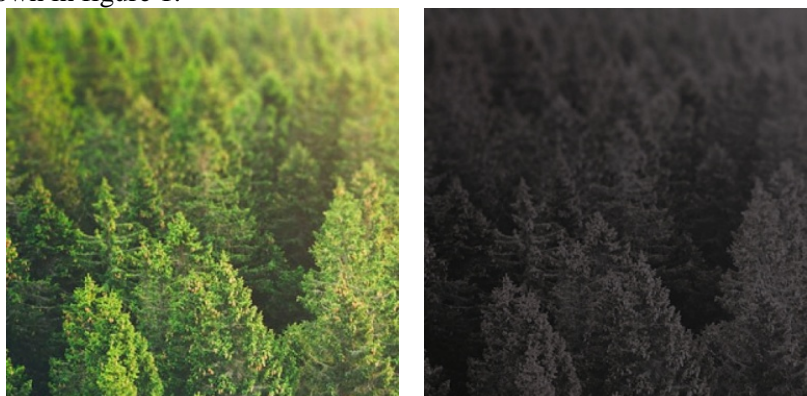


Figure 1. Used images sample.

Next, using the written analysis software, color analysis of the images was carried out: the distribution of colors by brightness was calculated and histograms were constructed for each of the used filters. An example of the initial histogram of color distribution by layers of a single image is presented in figure 2.

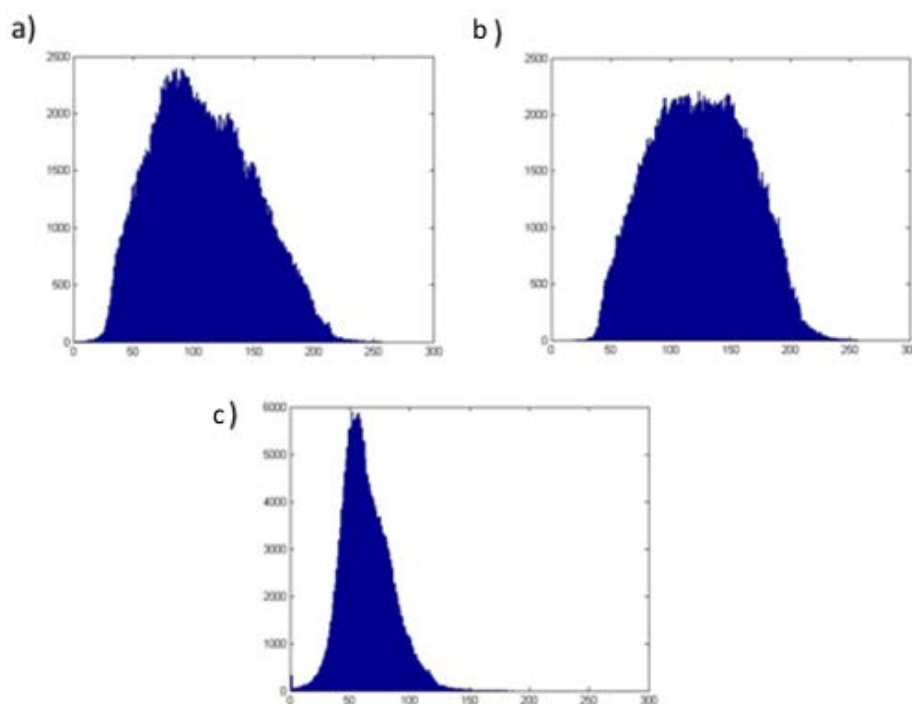


Figure 2. Example of color distribution of one image (a) - red, b) - green, c) - blue).

The distribution for the whole class of images processed by a certain filter is similarly constructed.

Further, since the filter uses the same brightness transformations, then by subtracting the image processed by the filter from the matrix of the original image from the matrix, the filter matrix can be obtained. By adding all the results of calculating the matrices of filters for pairs of images processed by the same filters, and also by normalizing, we obtain a common matrix for a filter of a certain type.

Hence, since the service information obtained from the social network in the previous step contains data on the applied filter, it is not difficult to restore the original image using the filter matrix.

To test this hypothesis, a test sample was organized, containing about 1000 images received from the Instagram network, and analysis and restoration of the original images was carried out. An example of the result of the work can be seen in figure 3.



Figure 3. Example of processed and original reconstructed image (a) - processed, b) - original reconstructed).

4. Conclusion

Definition of source images is one of the most significant tasks of computer vision and image processing. It finds its solution in various fields of human activity. Often, existing image processing

methods work fine with relatively small amounts of raw data. Processing of large arrays of source images takes considerable time, which is absolutely unacceptable in a number of tasks.

Currently, work is underway to adapt the above methods to work with large volumes of data or BigData, and software is being developed that allows for the proper processing of image arrays and the recognition of objects on them.

5. References

- [1] Manovich L Z 2015 Instagram is a window in the mind and imagination of a person Instagram is a window in the mind and imagination of a person Instagram is a window in the mind and imagination of a person *MediaProfi*
- [2] Khaykin S 2008 *Neural networks: full course* (Williams Publishing House) p 1131
- [3] Sergeev V V 1998 Application of the methodology of pattern recognition in digital image processing tasks *Autometry* **2** 63-76
- [4] Verhagen K, Doyne R and Groon F 1985 *Pattern Recognition: Status and Prospects* (M.: R&S) p 104
- [5] Samal D I 2002 Algorithms for human identification from a photo portrait based on geometric transformations *Thesis for the degree of Candidate of Science. Institute of Technical Sciences. Cybern* (NAS of Belarus, Minsk) p 170
- [6] Graphic filters based on the torsion matrix URL: <https://habrahabr.ru/post/43895> (01.05.2019)
- [7] Kravtsova N S, Paringer R A and Kupriyanov A V 2017 Parallel implementation of the informative areas generation method in the spatial spectrum domain *Computer Optics* **41(4)** 585-587 DOI: 10.18287/2412-6179-2017-41-4-585-587
- [8] Khotilin M, Blagov A 2016 Visualisation and cluster analysis of social networks *CEUR Workshop Proceedings* **1638** 843-850
- [9] Reece A, Danforth C 2019 Instagram photos reveal predictive markers of depression *EPJ Data Science* URL: <https://epjdatascience.springeropen.com/articles/10.1140/epjds/s13688-017-0110-z> (07.05.2019)
- [10] *Image Pre-processing – Towards Data Science* URL: <https://towardsdatascience.com/image-pre-processing-c1aec0be3edf> (17.05.2019)
- [11] Krishnaveni C, Udhayakumar U 2018 Big Data Analytics: Challenges and Applications for Social Media Data *International Journal of Engineering and Techniques - National Conference On Emerging Trends in Computing Technologies* **107** 1-6

Acknowledgments

This work was financially supported by the Russian Foundation for Basic Research under grant # 19-29-01135, # 18-37-00418, # 17-01-00972 and by the Ministry of Science and Higher Education within the State assignment to the FSRC “Crystallography and Photonics” RAS No. 007-GZ/Ch3363/26 (theoretical results).