https://doi.org/10.2352/ISSN.2470-1173.2021.4.MWSF-274

Automated Image Metadata Verification

Kyra Wittorf, Martin Steinebach, Huajian Liu; Fraunhofer SIT, Darmstadt, Germany

Abstract

Nowadays, almost everyone owns a device capable of photography. More and more photos are taken and distributed through the Internet. In the times of analogue photography, pictures were considered to be legally accepted evidence, but today, due to the multitude of possibilities to manipulate digital pictures, this is not necessarily the case. Metadata can provide information about the origin of the image. The prerequisite for this is that they have not been altered. This work shows possibilities how metadata can be extracted and verified. The additional meta information of an image and its standards are of central importance. We introduce a method of comparing metadata with the visual image content. For this purpose, we apply machine learning for automatically classifying information from the image. Finally, an exemplary verification of the metadata by means of the weather is carried out to provide a practical example of how the presented approach works. Based on this example and on the presented concept, verifiers for metadata that verify several aspects can be created in the future. These verifiers can help to detect forged metadata in a forensic investigation.

Motivation

Digital images play an important role in todays distribution channels of news. They are often used to support a written statement and act as an indicator that a given event actually happened. These news and images can come from official news agencies as well as social media networks. In the second case they are most often user generated content.

Independent from its provenience, any person planning to distribute the news and the images, especially a journalist, is well advised to verify the facts behind this statement. This verification can be executed in various ways and can include searching for more sources, talking to persons involved or having a closer look at the image.

Images can be used in a manipulative manner or can actually be manipulated. The former means that for example images from different events are used to support a news message. The latter is the case when image montages or content manipulations changing the content of the images have been executed. While image manipulation detection is a challenge widely addressed by media security experts and especially the risks coming from deep fakes are discussed today, a significant amount of misleading usage of images is still done by putting these images in the wrong context without any means of content manipulation [1]. Figure 1 shows an example of the types of manipulations. We see the original photo, one copy that has been manipulated by a copy-move tool, and another copy where only the metadata has been changed to claim another place of origin.

There are two possibilities of taking images out of their context: one is to re-used images already distributed in the public. In this case, inverse image searching can identify the images and their original context. The original images are detected with little effort, even in the case of montages of multiple images [2]. This is the case because mechanisms for finding copies of identical images or parts of these are well advanced and reliable.

But if an image which has not been published before is used out of context, finding the original reference must fail. For example, a photo of an aggressive crowd at an public protest is taken at a different place or a different time and stored away. Now it can be used together with a news text claiming a peaceful protest turned into a riot. The only ways to identify this manipulative usage of the photo are to either find sufficient alternative photos showing the same protest being peaceful or to identify aspects of the photo proofing it was not taken at the given time or place.

In this work we address the second strategy: We aim at automated cross-checking the content of the photo and its metadata. Of course there are many aspects of a photo that can be used for verification, and some of them have been used in this context already. Well-known is e.g. the relationship between metadata time and visible shadows on the photo as these are influenced by the position of the sun. We use 'weather' as an example, as it can be derived from the content of photos taken outsides and data about the weather at a given time and place is available freely.

Cross-checking the weather can be done manually: a journalist looks at the weather on the photo and uses the photos' metadata to look up the weather on a public database at the given time and place. If both differ, this is a good hint of a fraud. Identical weather from both photo content and metadata may not be sufficient to verify the photo due to the limited data space given by the types of weather. But this is only one example of possible fact-checking.

Our goal is to automate the process of comparing the weather shown at the photo with the weather derived from the metadata. We use machine learning to classify the weather on the photo and Open Source Intelligence (OSINT) to look up the weather at the given metadata. If this is possible for multiple aspects of a photo in the future, all these individual automated verification steps could be used to find out any inconsistency between photo and metadata to largely support the process of image verification.

Objective

Metadata describes images beyond the pixel level. For example, they may include the location, time, and camera type of the image, thus providing information about the origin of the image. Thus, there are two sources of information in an image, namely the visual image content and the metadata. Verification of metadata means that the data is compared to the visual content and the overall image context. A verification example for weather: we look at the visual image content. Based on this, the weather on the photo is classified. Then location and time are extracted from the metadata and this data is used to determine the weather at the point where the photo was taken with the help of the Internet. The

Time	9:45 AM
Date	23.10.2020
Place	Frankfurt

Time	9:45 AM
Date	23.10.2020
Place	Frankfurt

Time	9:45 AM
Date	23.10.2020
Place	Berlin

Figure 1: Examples for manipulation types.

two pieces of information are now compared. If the weather data match, the metadata match the image in this respect.

The starting point of this work is the hypothesis that the metadata correspond to the truth, whereby the term hypothesis is understood in the present work as untested speculation or unproven assumption. Our aim is to develop a concept for the methods to be used to test the above-mentioned hypothesis. This is called "verification" if a confirming finding is found for this hypothesis, for example by means of a proof. In contrast, the term "falsification" is understood as the presence of a contradictory finding to a scientific statement.

Background

As stated already above, our work has the goal of designing a method to ensure the authenticity and integrity of the context in which images have been taken. Integrity here means the immutability of the data. Authenticity therefore ensures, among other things, that metadata of an image with respect to time and place is correct. The authors of [3] proof integrity and authenticity by a signature system integrated within a camera. This system links the identity of the camera and other metadata to the image in a way that cannot be falsified. Verification is not to be confused with image manipulation. There are many works (e.g. [4]) addressing the detection of manipulations within an image. Standard approaches are:

- Recognition of similar regions in an image created with cloning tool.
- A noise analyzer to detect manipulations such as deformations, warping or perspective corrected cloning.
- A luminance gradient, which can analyze the illumination of the image to detect whether an object has been inserted into the image by copying it.
- Error level analysis. This function compares the original image with a compressed version.

Software tools for manipulation detection are available, like 'Forensically' by Jonas Wagner , the 'Reveal Image Verification Assistant' from CERTH-ITI and Deutsche Welle and Amped Authenticate. What the tools have in common is that they detect whether an image has been manipulated on a visual level. What is missing is a check whether the metadata may have been altered and the image therefore shows something wrong. This is the research challenge of this work.

Design

As mentioned before, verification or falsification is based on a hypothesis. If a counterexample for this hypothesis is found, it is called a falsification. Otherwise, a positive argument is a sign of verification. If it is to be verified that a picture corresponds to the truth, two kinds of hypotheses are to be considered.

Hypotheses

The first hypothesis assumes that the visual image is true, i.e. that it has not been manipulated. For this hypothesis there are already possibilities of verification and therefore it will not be considered further.

The second hypothesis is the metadata correspond to the truth. To verify this, the credibility of the metadata can be inspected from different aspects.

- Verifying the validity of the metadata values.
- Verifying the consistency among different metadata.
- Verifying the consistency between the metadata and the image visual content.
- Verifying the consistency between the metadata and the testimony of a witness.

The hypothesis can be falsified by a contradictory finding or verified by confirmatory evidence. For full verification, it is appropriate to look at multiple factors. In these far-reaching verifications, an overall decision should be made by the verifier at the end from all partial results. In this paper, we will focus on the verification of the consitency between the metadata and the visual content.

Combination of Tools

We combine EXIF data extraction, open source intelligence (OSINT) and machine learning based attribution in a combined verification system, as shown in Figure 2. As an example, we show how to verify a photo with respect to weather conditions. The following steps are carried out: From the EXIF metadata we learn the time and the location, which indicate when and wehre the photo was taken. The time and location can then be used to look up weather information from Internet databases like Weatherbit, Accu Weather, Clima Cell, meteostat, OpenWeather or DarkSky. Finally, we compare the weather data retrieved from these databases with a machine learning based attribution of the photo.

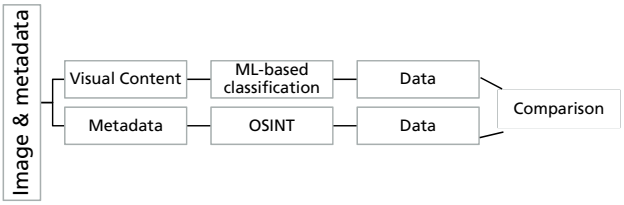


Figure 2: Abstract concept of our approach

Image Attribution

We use a pre-trained VGG16 network to learn the the photo attribution, which consists of 16 layers (convolution layers and fully connected layers). The three fully connected layers in the classifier block must be replaced for the task. The replacement is done with three other fully connected layers. For training, we used freely available image sets. One such set is the 65,000-element RSCM record [5]. All images are divided into six categories. The record contains about 10,000 images for each of these categories, taken from Flickr and Google. The categories are: sunny, cloudy, rainy, snowy, foggy and thundery.

GPS Data Correctness

One important issue is the accurarcy of the GPS data obtained and whether it can contribute to the determination of the location. In their work, the authors Dennis Zielstra and Hartwig H.Hochmair analyzed the positional accuracy of images on Flickr and Panoramio. Starting point for their investigation were 639 images from Flickr and 794 from Panoramio. In their study, the authors found that the average error distance of images on Flickr varies from continent to continent. The results indicate that in Asia an average error distance of 234 meters, in Europe of 58 meters, in Latin America of 1606 meters, and in North America of 46 meters [6]. This result shows that GPS tag inaccuracy should be expected and this should be taken into account when using the data.

Training Data

For training the image classification, we use the following predefined, labeled data sets.

MWI dataset The MWI dataset [7] contains 20,000 images, which were crawled from Flicker, Picasa, MojiWeather, Poco, Fengniao, among others. The images have been divided into five classes: snowy, sunny, rainy and foggy. The dataset was used by Zheng Zhan et al. and is not publicly available online.

Weather dataset This weather database has only two classes. The image set contains 5,000 sunny and 5,000 cloudy images, which are obtained from the Sun database, Labelme database and Flickr [8].

Image2Weather dataset The 'Image2Weather' database [9] contains a total of 183,798 images from Europe. These images are divided into five weather classes. There are 70,501 sunny images, 45,662 cloudy images, 1,252 snowy images, 1,369 rainy images and 357 foggy images. The creators explicitly made sure that the images contained at least 10 percent sky.

Table 1: Verification of the pristine images

No.	Timestamp	Actual Weather	Dark Sky	VGG16	Result
1	09.04.2020 10:49:44	sunny	sunny	sunny	verified
2	26.02.2020 14:55:11	cloudy	rainy	cloudy	falsified
3	09.04.2020 10:46:12	sunny	sunny	snowy	falsified
4	12.04.2020 17:55:27	rainy	cloudy	cloudy	verified
5	10.04.2020 11:32:52	sunny	sunny	sunny	verified
6	12.04.2020 17:46:21	rainy	cloudy	rainy	falsified
7	12.04.2020 17:42:47	cloudy	cloudy	cloudy	verified
8	12.04.2020 17:45:41	rainy	cloudy	foggy	falsified
9	27.02.2020 17:04:58	snowy	snowy	snowy	verified
10	12.04.2020 09:23:58	sunny	cloudy	sunny	falsified

Table 2: Verification of the intelligently manipulated images

No.	Timestamp	Location	Dark Sky	VGG16	Result
1	30.03.2020 16:00:00	Kiel	sunny	sunny	verified
2	29.08.2019 13:00:00	Stuttgart	cloudy	cloudy	verified
3	30.03.2020 16:00:00	Kiel	sunny	snowy	falsified
9	26.02.2020 15:00:00	Munich	snowy	snowy	verified
5	30.03.2020 16:00:00	Kiel	sunny	sunny	verified

RSCM dataset Another dataset is the 65,000-element RSCM dataset [10]. All images are divided into six categories. The dataset contains about 10,000 images for each of these categories, which come from Flickr and Google. The categories are: sunny, cloudy, rainy, snowy, foggy, and thundery. Accordingly, this dataset has one class, and that is "thundery", more than the 'Image2Weather' dataset. In the construction of the dataset, care was taken to ensure that they were photorealistic images of the outdoors and that they have a reasonable resolution.

Evaluation

To evaluate the proposed concept, three types of images are used in our test which are categorized into three groups. The first group consists of pristine images that are not manipulated. The second and the third group includes manipulated images. The shooting time or location, or both, in the metadata of the images in the second group are replaced with random values. In the third group, the manipulation is done in a more intelligent way. The adversary selects images with background corresponding to the weather condition at the shooting time and location.

Table 1 lists the verification results of sample images in the

Table 3: Classification results

Data Set	Standard	Fine-tuned
I2W	90.63%	94.20%
RSCM	83.41%	80.54%
WD	81.19%	83.81%

first group. Although all the images are not manipulated, some of them are recognized as falsified due to inaccurate weather data or wrong classification. In the second group, when only the shooting time is randomly altered, 50 percent images are erroneously verified. When the location is randomly replaced by one of ten selected cities in Germany, the results of 90 percent images remain unchanged because of coincidental weather data retrieved from DarkSky. When both the shooting time and location are changed, 72.5 percent manipulations are correctly detected, while 27.5 percent images are verified by mistake. The test results of the third group are listed in Table 2, in which only one image is detected as falsified.

Error Rate Estimation

The accuracy of the test results depends on the retrieved weather data, which is based on the metadata, and the classification of the visual content. Errors in either weather data retrieval or visual content classification may lead to incorrect results.

To estimate the error rate of the content classification, the correct rate of the used model on the 'Image2Weather' dataset (I2W) has first been determined, which is 90.63 percent on the model after feature extraction and improves to 94.2 percent on the fine-tuned one. Because the error rate might differ on other images, additional weather image databases have been tested with the model. At the end, the results are averaged to estimate the error rate of the model.

In the testing, the above-mentioned datasets have been used. For each of the five categories in 'Image2Weather', 2000 images are selected from the 'RSCM' dataset. The images tagged with the label "thundery" in the dataset are not taken into account because they cannot be optimally fit into the existing categories. The 'Weather' dataset (WD) consists of only two types of labeled images, so some classes in the model are not considered for these total 10,000 images. Thus, in addition to the 730 test images out of 'Image2Weather', a total of 20,000 images have been tested. The results are listed in Table 3. The two models have an average correct classification rate of 82.59 percent based on the data size of the three datasets, which results in an error rate of 17.41 percent.

In order to estimate the error rate of the weather data retrieval from DarkSky API, 92 queries have been examined, which are all for the location 'Darmstadt'. Four queries are made for the 1st and 15th of each month in the period from 1st May 2019 till 1st April 2020. These four queries are made at different times, namely 10, 12, 14 and 16 o'clock. The results are summarized in Table 4. The icon is used for weather determination and then converted to a weather class, which matches the 'Image2Weather' classes. The summary provides more human-readable information than the icon. To assess the accuracy, the weather class is compared with the summary. The average of all results gives an estimated overall accuracy of about 94 percent.

Furthermore, to examine which weather classes the model misclassifies, the confusion matrix of the fine-tuned model is

Table 4: Summary of DarkSky API Responses

Occur.	**Summary**	**Icon**	**Weather Class**	**Estimated Accuracy**
29	clear	clear-day	sunny	1
16	mostly cloudy	partly-cloudy-day	cloudy	1
9	light cloudy	partly-cloudy-day	cloudy	0.7
27	overcast	cloudy	cloudy	1
2	possible light rain	rainy	rainy	0.5
6	light rain	rainy	rainy	1
3	possible drizzle	rainy	rainy	0.5
92 times examined		Average:		0.9434

Table 5: Confusion matrix of the fine-tuned model

		Predicted Class				
		cloudy	foggy	rainy	snowy	sunny
Correct Class	cloudy	138	0	13	3	16
	foggy	0	41	5	4	0
	rainy	11	3	140	10	6
	snowy	1	1	4	162	2
	sunny	19	0	3	0	148

shown in Table 5, which gives the error distribution of the classification of the images from Image2Weather. Overall, it can be seen that most of the images are located on the diagonal and are thus correctly classified. Among misclassifications, cloudy images are mostly classified as rainy or sunny and vice versa sunny and rainy ones more often as cloudy. In addition, snowy images are sometimes recognized as rainy and incorrectly classified foggy images are also most likely to be assigned rainy.

Discussion

There are already many (forensic) methods that can detect manipulations based on the pixel level of the photo. In addition, there is a concept of how a special camera can be used to ensure the integrity of the recording context/metadata through signatures.

What is missing here is a process for arbitrary images that recognizes manipulations of the recording context. To fill this gap, we developed a concept how metadata can be verified and thus the authenticity of the recording context can be examined. In this work, we compare metadata with information derived from the visual information of an image in order to uncover inconsistencies. In doing so, the metadata can be checked for the correct format and among themselves for plausibility. If both sources information are in conflict with each other, this indicates a falsification, otherwise it is a verification.

The results and the perspectives coming from this work need to be seen as as proof of concept: we did show that the idea of automated verification is possible due to the potential of OSINT data gathering and machine learning based classification. The weather-based classification is only a small example of other verification aspects. This may include environmental information (does the style of a building match with the GPS info or other photos from that place?), traffic information (was there a traffic jam shown on the image at that time and place?), city plans (does the pattern of buildings shown on the image match with the map data for this

place?) and many others.

The limiting factors will be (a) the data that can be extracted from an image to be matched and (b) the data available for OSINT retrieval. Only if (a) and (b) both provide sufficient information and reliability, a working solution is possible. In addition, more data sources could be considered: verification often takes place while evaluating news; data elements extracted from the news text can also be compared to both OSINT and classification data. A trained net could for example recognize uniforms of soldiers and attribute these to countries. An image showing military activities together with a text claiming these activities to be executed by the army of country A could be falsified if the uniforms are classified as belonging to country B.

From a technical perspective, the process of verification may require are large amount of parallel processing: each aspect to be verified requires classification as well as OSINT data retrieval. It seems most feasible to first analyze the image with respect to elements that can be classified and then start the individual verification of these elements. After completion, an aggregated result can be provided to the use pointing him to potential inconsistencies within the photo.

Therefore, future work needs to address additional verification pairs of classification and OSINT as well as designing a framework to assign and handle all pairs relevant for the photo under analysis.

Acknowledgment

This research work has been funded by BMBF and the Hessen State Ministry for Higher Education, Research and the Arts within their joint support of the National Research Center for Applied Cybersecurity ATHENE.

References

[1] M. Steinebach, Bader Katarina, L. Rinsdorf, N. Krämer, and Alexander Roßnagel, editors. *Desinformation aufdecken und bekämpfen – Interdisziplinäre Ansätze gegen Desinformationskampagnen und für Meinungspluralität*. Nomos Verlag, Baden-Baden, 2020.

[2] Martin Steinebach, Karol Gotkowski, and Hujian Liu. Fake news detection by image montage recognition. In *Proceedings of the 14th International Conference on Availability, Reliability and Security*, pages 1–9, 2019.

[3] Eckehard Hermann, Harald Lampesberger, Lena Heimberger, and Michael Altenhuber. Authentizität und integrität des aufnahmekontextes von bildern. *Datenschutz und Datensicherheit-DuD*, 43(5):281–286, 2019.

[4] Hany Farid. *Photo forensics*. MIT press, 2016.

[5] Di Lin, Cewu Lu, Hui Huang, and Jiaya Jia. Rscm: Region selection and concurrency model for multi-class weather recognition. *IEEE Transactions on Image Processing*, 26(9):4154–4167, 2017.

[6] Dennis Zielstra and Hartwig H Hochmair. Positional accuracy analysis of flickr and panoramio images for selected world regions. *Journal of Spatial Science*, 58(2):251–273, 2013.

[7] Zheng Zhang and Huadong Ma. Multi-class weather classification on single images. In *2015 IEEE International Conference on Image Processing (ICIP)*, pages 4396–4400. IEEE, 2015.

[8] Cewu Lu, Di Lin, Jiaya Jia, and Chi-Keung Tang. Two-class weather classification. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 3718–3725, 2014.

[9] Wei-Ta Chu, Xiang-You Zheng, and Ding-Shiuan Ding. Image2weather: A large-scale image dataset for weather property estimation. In *2016 IEEE Second International Conference on Multimedia Big Data (BigMM)*, pages 137–144. IEEE, 2016.

[10] Di Lin, Cewu Lu, Hui Huang, and Jiaya Jia. Rscm: Region selection and concurrency model for multi-class weather recognition. *IEEE Transactions on Image Processing*, 26(9):4154–4167, 2017.

Author Biography

Kyra Wittorf received her B.S. degree in computer science from Technical University of Darmstadt in 2020. Today she studies M.S. IT Security at the Technical University of Darmstadt.

Martin Steinebach is the manager of the Media Security and IT Forensics division at Fraunhofer SIT. In 2003 he received his PhD at the Technical University of Darmstadt for this work on digital audio watermarking. In 2016 he became honorary professor at the TU Darmstadt.

Huajian Liu received his B.S. and M.S. degrees in electronic engineering from Dalian University of Technology, China, in 1999 and 2002, respectively, and his Ph.D. degree in computer science from Technical University Darmstadt, Germany, in 2008. He is currently a senior scientist at Fraunhofer SIT. His major research interests include multimedia security, digital watermarking, robust hashing and digital forensics.

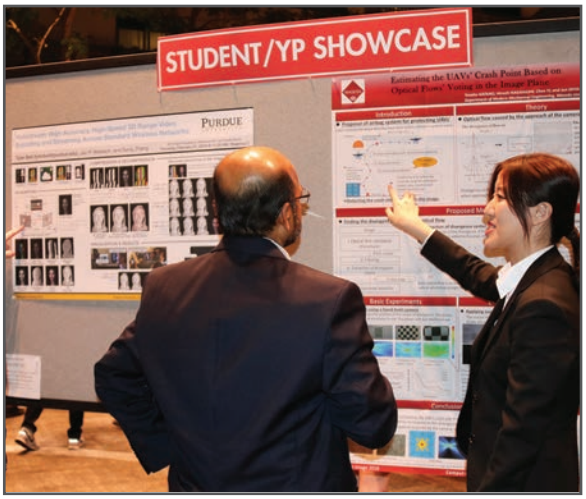
JOIN US AT THE NEXT EI!

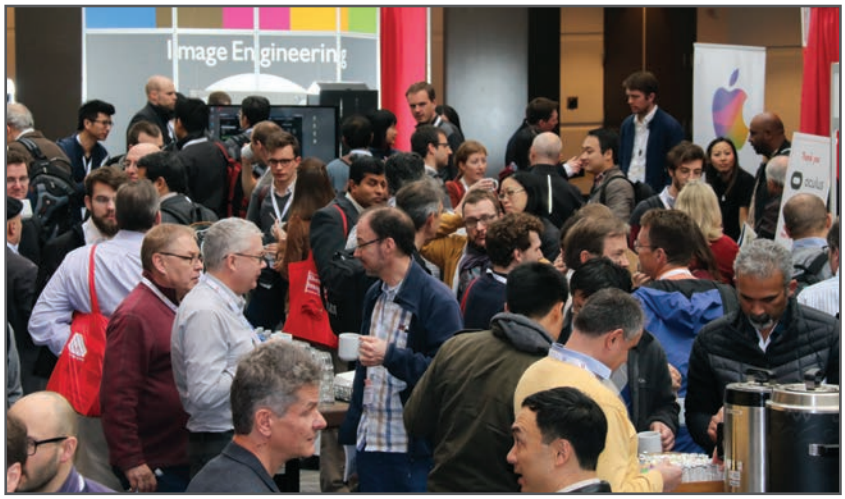
IS&T International Symposium on

Electronic Imaging

SCIENCE AND TECHNOLOGY

Imaging across applications . . . Where industry and academia meet!





• SHORT COURSES • EXHIBITS • DEMONSTRATION SESSION • PLENARY TALKS •
• INTERACTIVE PAPER SESSION • SPECIAL EVENTS • TECHNICAL SESSIONS •

www.electronicimaging.org

IS&T
imaging.org