https://doi.org/10.2352/ISSN.2470-1173.2020.4.MWSF-118

Checking the Integrity of Images with Signed Thumbnail Images

Martin Steinebach, Sebastian Jörg, Huajian Liu; Fraunhofer SIT, Darmstadt, Germany

Abstract

The integrity of images is an important and interesting field of research, since digital images are constantly encountered in everyday life today. The availability of image processing programs makes it possible for almost anyone to manipulate images without great effort. With the help of social media platforms, the hurdle for their distribution to a very large number of viewers has also been lowered. As a result, confidence in the integrity and authenticity of images, which was even stronger at the time of analogue photography, is dwindling.

The aim of this work is to develop and investigate a concept that counteracts the lost trust and creates an opportunity to check the integrity of processed images. The concept is based on a combination of signed thumbnails and the logging of possible processing steps. We show that this combination has advantages compared to the existing approaches.

Motivation

Integrity is a security goal in IT security and takes on different meanings in different contexts. In general, integrity is defined as a state where the subject is unchanged. The integrity of digital images can be considered at file or bit level. In this context, integrity requires an unchanged binary representation of the image. If one considers the content level of an image, i.e. what a human eye sees and processes, then integrity demands that the content or meaning of the image has not been changed. This is independent of the representation of the image in bits.

This work addresses the latter aspect, also called content, semantic, perceptual, visual or optical integrity. Cryptographic hashes can be used to detect the change in an image at the bit level but fail to distinguish between accepted operations like lossy compression and manipulations.

In research, there are several directions to address this problem. As early as 1993 Friedman [3] described an approach against the loss of trust in digital photos, in which a cryptographic hash is calculated directly from the photo in the digital camera and signed by the digital camera. Robust hashs should allow changes that do not alter the content of the image. The goal is to extract the information from an image that is critical for the content of the image. This is called a feature of images. In the literature different algorithms exist, which are based on completely different features. Fragile, semi-fragile or content-fragile watermarks are used to prove the integrity of an image. An approach defines a threshold above which a watermark that is no longer completely present categorizes an image as manipulated. Another possibility is to select content-relevant features, which are resistant to allowed operations, from the image and embed them as watermarks. The stored and actual marks are compared with each other. Image forensic detects manipulations in an image based on model of unaltered images or by recognizing typical traces of manipulations.

Research does not address logging and saving the permitted edits in order to allow them on a larger scale. Based on robust hash values, the thumbnails store the content information of the images. In contrast to the approaches of signing a photo directly in the camera and thus trusting the camera, the concept of this work requires trusting the first distributor of the image. The additional storage space required lies between that of a robust hash value and the extra storage of the complete image.

Objective

Our approach aims to provide a method to reliably verify the integrity of images without limiting usual image operations. We use a size reduced version of the image (called thumbnail) as an integrity reference point and a list of operations on the original image to be able to ensure that a current image can be the result of the operations executed on the original image. By this, we are significantly more robust than fragile or content-fragile watermarks, e.g. with respect to scaling.

We do not require lookup to a central database holding all original images. The security is based on well-established hash functions and public key infrastructures. Image file size increases due to the additional reference thumbnail limited to gray-scale images with a low resolution, the hash signature as well as a modification log.

There are many use cases for this approach. A photographer signs his photo and makes it available to an agency. The agency can now verify the integrity and authenticity of the photo. In the positive case the agency selects e.g. only a picture section and forwards this to an editorship. The editorial department can now check the integrity of the photo before publishing it. If the article is published on a website, then the reader would also be able to check the integrity of the photo.

In the field of medicine, the imaging devices can sign the image directly during the recording. A physician now examines the images and cuts the image only to the relevant parts of the findings. The attending physician is now able to check the integrity of this image section.

Perspective

We see this paper as an impulse to re-think current approaches for image integrity verification. So far known approaches are either too complex with respect to the required infrastructure or too fragile with respect to accepted operations. This leads to most images being unprotected against content manipulations. Only post-mortem image forensics provide some means to verify integrity, but results are often unreliable or depend on the processing steps executed after the actual manipulation, like scaling or lossy compression.

Our actual implementation is currently somewhat limited, as the results presented later in this work will show. But there may be methods better suited for reducing thumbnail images and comparing both image versions to distinguish accepted operations from

manipulations. The general concept of providing a 'digital twin' of an image which is protected by cryptography and allows to recreate the changes in the image protocol is promising. From a security perspective, only the integrity of the thumbnail is relevant. If an attacker can modify the list of operations, he still can only execute accepted operations on the image. A masking or removal of operations is impossible or unproblematic, as at the end the integrity decision is based on the comparison of thumbnail and actual image.

Related Work

Our work aims to provide a method to verify the integrity of images. We therefore first discuss the concept of 'integrity'. It is a protection goal in IT security and has different meanings in different contexts. In general, integrity is considered to be a state where the subject is unaltered or changes are only allowed by authorized persons. Stored images can be viewed at the bit level. In this context, integrity requires unchanged Bits.

If, on the other hand, one looks at the perceived content level of an image, it is only relevant what the human eye sees. Here integrity requires that the content or meaning of the image has not been altered. This is independent of the representation of the image in bits. Integrity on this level can be seen as *optical, perceptual or visual integrity*. If integrity is used in the further course of this work in connection with an image, then optical integrity is always meant. Authenticity describes the genuineness and credibility of an image. In this work this means the proof of authorship of the creator of the image and thus to the trust placed in him.

Cryptographic Hashes

A cryptographic hash is a one-way function which calculates a short bit sequence of fixed length (hash value) from a file. A well known and recommended algorithm is SHA-3, for example, which calculates a completely different hash value if only a single bit of the file changes. The original file cannot be recalculated from the hash value. Due to the collision resistance, it is also virtually impossible to create two different files with the same hash value.

Cryptographic hashes can be used to detect bit-level changes in an image. However, if an image is re-compressed with a lossy compression such as Joint Photographic Experts Group (JPEG), then the bit representation of the image changes even though the optical integrity has not been changed. Thus the integrity check would fail even though the image represents the same content. This is addressed by robust or perceptual hashes.

Signatures

Signatures use cryptographic hashes and asymmetric cryptography in such a way that when applied to a digital file, they identify a person or a device, and any subsequent manipulation of the file must be recognisable. The focus of this work is not on signatures, certificates and the public key infrastructure. We use signatures to ensure that the thumbnails are not changed and they clearly originate from the stated person. To achieve this, we use standard libraries widely available.

Signing directly during recording

Friedman [3] describes an idea against the loss of trust in digital photos, in which a cryptographic hash of the photo is computed directly in the digital camera and this is signed by the digital camera. Also meta data, such as camera version, time, date and Global Positioning System (GPS) coordinates, can be written optically on the photo and are thus also signed. This idea was later patented [4] . Since a cryptographic hash is used, there is the already described restriction that the photo cannot be stored again in a lossy compressed format.

Hermann, Lampesberger, Heimberger and Altenhuber [14] combine the concept with a Trusted Platform Module (TPM) to create your TPM Image Signature System (TISS). The concept exists so far only as a prototype on a Linux notebook. Secure Boot and TPM provide the integrity and authenticity of the camera system. In addition, the private keys are protected against unauthorized Access protected. With the solutions from the camera manufacturers Canon and Nikon, weaknesses in the Key management for the disclosure of private keys [14]. Besides the actual image, metadata such as date and GPS coordinates signed before storing in memory

Recently approaches based on smart-phone apps have been proposed [24]. Two startups are transferring the idea of ensuring integrity directly in the camera [15]. Truepic [7] is an extra smartphone app for taking pictures. The photos and metadata are signed. Furthermore, the app tries to detect changes on the smartphone, in place and time. The complete photo and metadata are stored in a block chain at Truepic and a web interface allows the integrity of the image to be checked [16], [14]. The second app Serelay [17] is based on locally deriving 'data points' from the images and does not upload image on its servers.

Robust Hashs

Robust hashes are developed especially for their application area. In addition to robust hashes for images, there are also various other algorithms, for example for text [9] and audio [11]. Robust hashes should allow changes that do not change the content of the image. The aim is to extract as much information as possible from an image that is crucial for the content information of the image. This is called a feature of images. In the literature there are different algorithms that are based on very different features. In a further step, the information content of the selected feature has to be reduced so that they can be stored in a short hash value with a fixed length. To calculate how similar two images are, often the Hamming distance, which is a geometric model for error detection, is used. When comparing two hash values it gives the number of different bits back. Thus one receives a measure, how similar two hash values are and thus also how the two pictures are similar. A refinement of this model is the weighted Hamming distance. Here the individual bits that differ when comparing two hash values. It is examined how large is the distance to the point where the bits in the algorithm would tilt to the other value. This distance is then weighted in comparison. Small differences thus have less effect on the distance than larger ones.

Swaminathan, Mao and Wu Various compare various robust hash algorithms in their work [6]. Venkatesan, Koon, Jakubowski and Moulin [12] use a private key in addition to the image to calculate the hash value. As a feature, the subbands of a wave decomposition are randomly divided into rectangles and the hash value is calculated from these rectangles. This approach is e.g. robust against 10% scaling and against removal of 10% of the image area.

Weng and Preneel [5] choose a block approach to solve the problem of detecting the smallest changes. The private key is also used for calculation and therefore their approach can also be used as Message Authentication Code (MAC). The image is reduced to the size of 512x512 pixels and converted to grayscale. Then the image is split into blocks of 64x64 pixels. On each individual block a two-dimensional Discrete Fourier Transformation (DFT) is used for feature extraction. In our work [10] we propose a combination of a block hash and image segmentation. With the segmentation, individual larger objects of the image are identified and for all objects found an extra hash value is calculated. Thus one receives a set of hash values. If the cropped image contains one of these larger objects is present, it can be recognized.

Digital Watermarking

Digital watermarking is another well-known strategy to provide integrity protection [21]. Especially fragile watermarks [19] are designed to prove the integrity of an image. The decision if a manipulation has taken place is often made based on a threshold of the detected watermarking energy. Increasing the watermark robustness to allow certain accepted operations leads to semi-fragile watermarking [20]. Another approach is to embed features that are resistant to accepted operations but fragile to manipulations with a robust watermarking algorithm. An ideal watermark would be robust to all operations. Manipulations would only detected with the help of the embedded features. A third category are reversible watermarks [22] which allow to render the original unmarked state of an image. Their behavior is often similar to cryptographic hashes as they are fragile to even minimal changes. Typically in real-world designs, both features and watermarking can be fragile to accepted operations if the design is too fragile or they ignore content changes if it is too robust.

Multimedia Forensics

Unlike the previous areas, forensics is a passive technique. There is no knowledge of the original image and no calculations or information has been added to it before. The aim is to detect manipulations in an image. For different kinds of manipulations there are tool sets with different detection algorithms. In his overview, Farid [18] roughly groups the types of investigations into five areas. Statistical anomalies at pixel level, statistical correlations of lossy compression, conspicuous manufacturing under-tolerances of cameras, physical errors in the 3D and light model and geometric errors of the objects in relation to the camera perspective.

To summarize the state of the art, all known strategies for integrity protection have advantages and disadvantages. Table 1 shows an overview with respect to the most important aspects.

	sensitivity	**localization**	**robustness**
Cryptograhic hash	+	-	-
Robust hash	-	0	0
Fragile watermark	+	+	-
Semi-fragile watermark	0	+	0
Content-fragile watermark	0	+	0
Reversible watermark	+	0	-
Image forensics	0	+	0

Table 1: All known methods of image authentication have advantages and weaknesses

Concept

The concept can be divided into three phases. In the first phase, the creator of an image signs it. In the second phase, a third person is enabled to edit the image. Another person can now in a third phase verify the integrity of the image.

Sign In this first phase, a smaller gray scale variant is calculated from the original image. This thumbnail is then signed. Both The thumbnail and the signed hash value are attached to the original image. Figure 1 illustrates the process below:

1. The original image is resized to a given size or factor.
2. A gray scale image (the final thumbnail) is created from the resized image.
3. A cryptographic hash of the thumbnail is calculated.
4. The hash is signed by a private key creating a signature.
5. Original image, thumbnail and signature are stored together.

Edit In this phase, a special rudimentary image processing program is used to perform various permitted processing steps, which do not change the content of the picture. The following processing steps are supported:

- Vertical and horizontal mirroring
- Rotation in 90 degree steps
- Reduction of image size
- Cropping the image

All processing steps are logged in their exact order and stored in the metadata of the image. These do not have to be signed or protected, because if manipulated, the integrity check would fail in the next phase. The edited image can also be edited several times by the person or by other persons.

Verify The final phase is responsible for verifying the integrity of the image. To do this, it is necessary to perform the following steps:

1. From the stored thumbnail, a cryptographic hash value is calculated and compared with the stored and signed hash value.
2. The logged edits are read and applied to the saved thumbnail in order.
3. A new grayscale thumbnail is calculated from the current image.
4. The new thumbnail and the existing thumbnail are compared. If larger deviations are detected, a manipulation is assumed.

Results

During the evaluation, size reduction of images was the greatest challenge for integrity verification. In all other cases, a simple subtraction of the miniature images using a threshold value

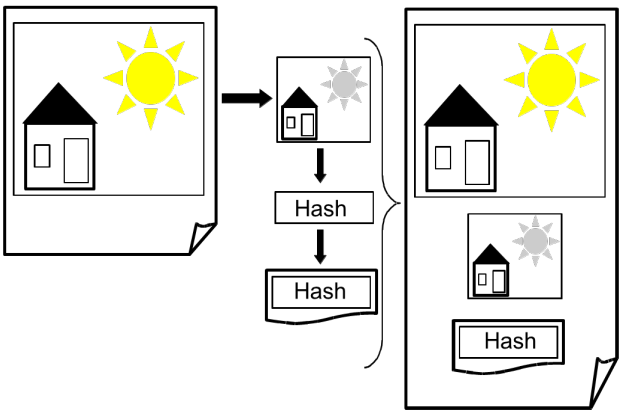


Figure 1. *Thumbnail Concept*

provided good results. The reduction caused too many pixel shifts and too much noise. On the one hand, a thumbnail image is created from the reduced image and on the other hand, the signed thumbnail image is reduced to the size of the new image.

If you calculate the robust hash value from the two thumbnails and then compare them, processed images are better recognized as equal than if the hash calculation is performed on the images (see table 2). However, this was not satisfactory for the detection of manipulations as processed and manipulated images often showed similar hamming distances.

In order to solve this problem, no singular hash value is used for the whole thumbnail, but a window of 32 x 32 pixels with an overlap of 16 pixels is moved over the thumbnail and for each resulting area a hash value is calculated and compared (see table 3). This procedure can also be used to determine the area in which the image has been manipulated. Figure 2 illustrates this by an example.

In addition, the noise is removed from the thumbnails before the hash values are calculated. Noise removal algorithms, such as Average Filter and Gaussian Averaging, result in a more blurred image, with edges very weakened, and thus crucial information is lost. On the other hand, the following two algorithms are better suited to preserve the edges in an image: A bilateral filter [1] preserves the edges by taking into account not only geometric proximity when replacing the color of a pixel, but also the color similarity of the surrounding pixels.

The Non-Local-Means filter is based on the simple principle of replacing the color of a pixel with the color of the average of similar pixels. However, similar pixels do not have to be nearby [2]. These described measures made it possible to detect manipulations and still allow extensive processing of the image. Existing approaches, on the other hand, consider permitted manipulations to be manipulations.

The concept also works well for integrity checks of smaller image sections. For this purpose, of course, sufficient pixels must represent the section on the thumbnail. So the size must be a multiple of the image thumbnail factor. Alternatively, it is possible to use a larger thumbnail. The grayscale thumbnails store a lot of information about the content. Of course, this also comes with greater memory consumption compared to a robust hash value. The thumbnail images also allow for the use of other approaches

Figure 2. *Example of detection*

	Edited		Manipulated	
	Miniature	Full	Miniature	Full
001.jpg	7	35	8	36
002.jpg	6	20	6	23
003.JPG	8	27	9	27
004.jpg	1	23	4	27
005.jpg	4	16	8	19
006.jpg	1	23	2	23
007.jpg	0	34	1	33
008.jpg	5	11	15	17
009.JPG	0	21	0	22
010.jpg	0	30	1	26
011.jpg	0	22	3	20
012.JPG	0	29	0	29
013.JPG	1	34	7	31
015.jpg	4	35	4	35
016.jpg	0	30	0	31
017.jpg	0	36	3	36

Table 2: Comparison of the complete image by a robust hash leads to hamming distances similar for edited and manipulated images.

to compare the miniature images be able to rate the image.

Figure 3 shows a comparison of the hamming distances from table 2 and 3. Figure 4 better illustrates the improved performance of the windowing approach. The difference between allowed operations and manipulation is much higher with the windowing approach.

Conclusion

In this work we suggest an alternative to existing approaches for image integrity protection. A signed thumbnail of the original image is added to the image file. For verification, the thumbnail of the current image and the original thumbnail are compared. By logging the permitted edits, these can be easily reproduced on the thumbnail. This ensures that when the two thumbnails are compared, the information is approximately the same locally. This is a big advantage for the comparison. The block-wise comparison includes much more information than just a single hash value is available. In addition, the position of possible manipulations can be displayed. Of course the concept reaches its limits if the manipulations are very small. This could be counteracted to a certain extent by larger thumbnails and is a general challenge of all concepts beyond cryptographic hashes.

Since the thumbnail, the signature and the protocol can be easily removed from the image file, a statement about the integrity

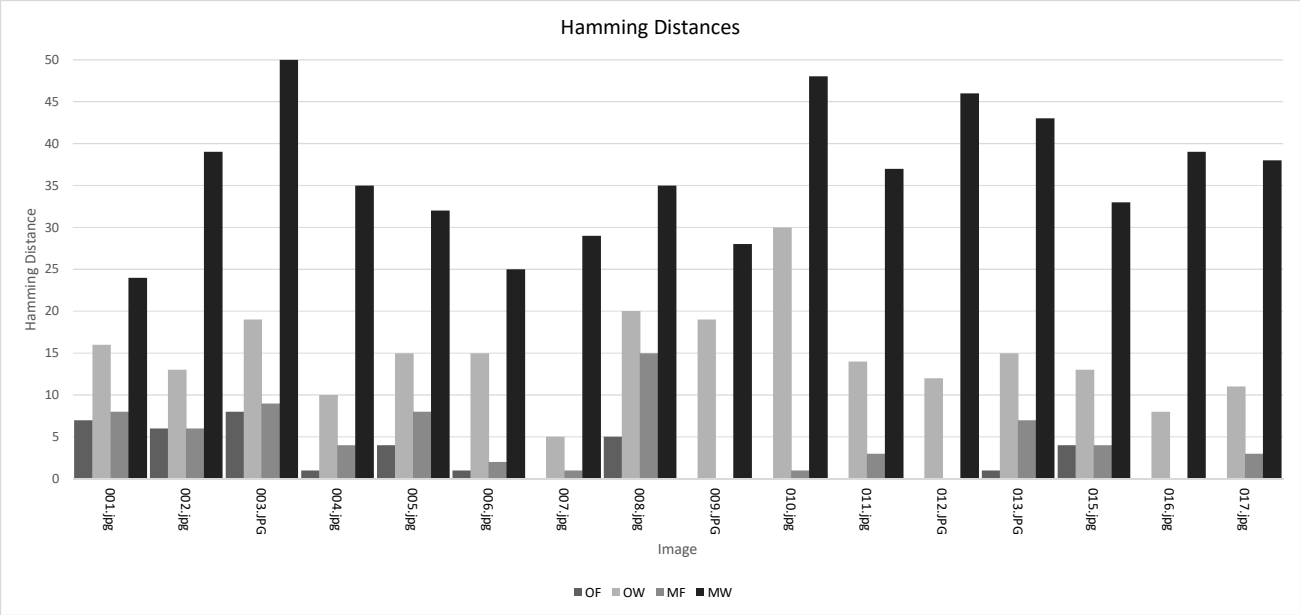


Figure 3. *Comparison of hamming distances. OF stands for only allowed operations and full image compared, OW for allowed operations and windowing, MF for for manipulations and full image compared and MW for for only allowed operations and windowing.*

	Edited		Manipulated	
	Miniature	Full	Miniature	Full
001.jpg	16	35	24	35
002.jpg	13	20	39	23
003.JPG	19	27	50	25
004.jpg	10	23	35	23
005.jpg	15	16	32	18
006.jpg	15	23	25	24
007.jpg	5	34	29	33
008.jpg	20	11	35	11
009.JPG	19	21	28	18
010.jpg	30	30	48	28
011.jpg	14	22	37	22
012.JPG	12	29	46	28
013.JPG	15	34	43	34
015.jpg	13	35	33	35
016.jpg	8	30	39	29
017.jpg	11	36	38	35

Table 3: Windowing the robust hash comparison allows to better distinguish edited and manipulated images

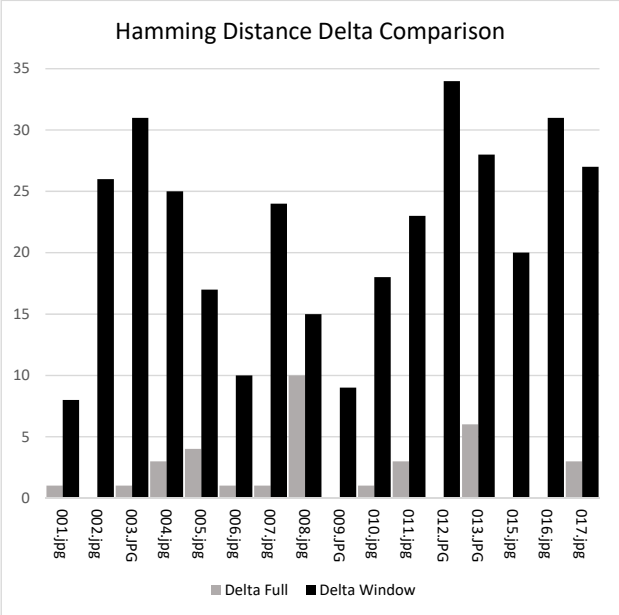


Figure 4. *Comparison of hamming distance deltas.*

of the image can only be made if this information is available. This is a shortcoming shared with all other active approaches for integrity protection. It can only be solved by either forensics or by mandatory usage of the integrity verification protocol.

The current state of our work must be seen as a starting point. Further research needs to address if the thumbnail is the most efficient and effective way to store as much information about the original image. Feature-based alternatives may provide better results here. Also the comparison strategy could be supported by an estimation of the likelihood of local changes due to image characteristics. Image registration methods to align both thumbnails before comparison my reduce the negative effect of image scaling. One advantage of our strategy is that improved verification algorithms does not require new creating of hashes and signatures.

Acknowledgment

This work was funded by the Federal Ministry of Education and Research (BMBF) within project DORIAN. This research work has also 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] C. Tomasi, R. Manduchi. (1998). Bilateral filtering for gray and color images. In Sixth International Conference on Computer Vision (IEEE Cat. No.98CH36271) (S. 839–846). doi:10.1109/ICCV.1998.710815

[2] Buades, A., Coll, B., Morel, J.-M. (2011). Non-Local Means Denoising. Image Processing On Line, 1, 208–212

[3] Friedman, G. L. (1993). The trustworthy digital camera: restoring credibility to the photographic image. IEEE Transactions on Consumer Electronics, 39(4), 905–910. doi:10.1109/30.267415

[4] Friedman, G. L. (1996). Digital camera with apparatus for authentication of images produced from an image file. US005499294. https://ntrs.nasa.gov/archive/nasa/casi.ntrs.nasa.gov/19960034301.pdf

[5] Weng, L., Preneel, B. (2011). A Secure Perceptual Hash Algorithm for Image Content Authentication. In B. de Decker, J. Lapon, V. Naessens, A. Uhl (Hrsg.), Communications and Multimedia Security (Bd. 7025, S. 108–121). Lecture Notes in Computer Science. Berlin, Heidelberg: Springer Berlin Heidelberg. doi:10.1007/978-3-642-24712-5

[6] Swaminathan, A., Mao, Y., Wu, M. (2006). Robust and Secure Image Hashing. IEEE Transactions on Information Forensics and Security, 1(2), 215–230. doi:10.1109/TIFS.2006.873601

[7] Truepic. (2019). Truepic. https://truepic.com/technology/

[8] Birajdar, G. K., Mankar, V. H. (2013). Digital image forgery detection using passive techniques: A survey. Digital Investigation, 10(3), 226–245. doi:10.1016/j.diin.2013.04.007

[9] Steinebach, M., Klöckner, P., Reimers, N., Wienand, D., Wolf, P. (2013). Robust Hash Algorithms for Text. In B. de Decker, J. Dittmann, C. Kraetzer, C. Vielhauer (Hrsg.), Communications and Multimedia Security (S. 135–144). Berlin, Heidelberg: Springer Berlin Heidelberg.

[10] Steinebach, M., Liu, H., Yannikos, Y. (2014). Efficient Cropping-Resistant Robust Image Hashing. In Proceedings of the 2014 Ninth International Conference on Availability, Reliability and Security (S. 579–585). ARES '14. Washington, DC, USA: IEEE Computer Society. doi:10.1109/ARES.2014.85

[11] Haitsma, J., Kalker, T., Oostveen, J. (2001). Robust audio hashing for content identification. In International Workshop on Content-Based Multimedia Indexing (Bd. 4, S. 117–124).

[12] Venkatesan, R., Koon, Jakubowski, M. H.,Moulin, P. (2000). Robust image hashing. In Proceedings 2000 InternationalConference on Image Processing (Cat. No.00CH37101) (Bd. 3, 664–666 vol.3). IEEE. doi:10.1109/ICIP.2000.899541

[13] ISO 16684-1:2019, Graphic technology — Extensible metadata platform (XMP) — Part 1: Data model, serialization and core properties

[14] Hermann, E., Lampesberger, H., Heimberger, L., Altenhuber, M. (2019). Authentizität und Integrität des Aufnahmekontextes von Bildern. Datenschutz und Datensicherheit - DuD, 43(5), 281–286. doi:10.1007/s11623-019-1108-4

[15] Karen Hao. (2018). Deepfake-busting apps can spot even a single pixel out of place, https://www.technologyreview.com/s/612357/deepfake-busting-apps-can-spot-even-a-single-pixel-out-of-place/

[16] Melcher, P. (2018). 10 Questions for a Founder : TruePic. Zugriff unter https://kaptur.co/10-questions-for-a-foundertruepic/

[17] Serelay (2019), https://www.serelay.com/

[18] Farid, H. (2009). Image forgery detection - A survey. IEEE Signal Processing Magazine, 26(2), 16–25. doi:10.1109/MSP.2008.931079

[19] Sreenivas, K., and V. Kamkshi Prasad. "Fragile watermarking schemes for image authentication: a survey." International Journal of Machine Learning and Cybernetics 9.7 (2018): 1193-1218.

[20] Egorova, Anna, and Victor Fedoseev. "Semi-Fragile Watermarking for JPEG Image Authentication: A Comparative Study." 2019 7th International Symposium on Digital Forensics and Security (ISDFS). IEEE, 2019.

[21] Steinebach, Liu; Fragile and authentication watermarks, in: Katzenbeisser, S.: Information Hiding Norwood/Mass.: Artech House, 2016 ISBN: 978-1-60807-928-5 ISBN: 1-60807-928-7

[22] Khan, Asifullah, et al. "A recent survey of reversible watermarking techniques." Information sciences 279 (2014): 251-272.

[23] Zheng, Lilei, Ying Zhang, and Vrizlynn LL Thing. "A survey on image tampering and its detection in real-world photos." Journal of Visual Communication and Image Representation 58 (2019): 380-399.

[24] Azoulay, Roy. "Verification of data captured by a consumer electronic device." U.S. Patent Application No. 16/476,005.

Author Biography

Prof. Dr. 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.

Sebastian Jörg received his B.S. degree in computer science from the Technical University of Darmstadt in 2019 and he is currently studying for his M.S. degree in IT security at the Technical University of 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 research scientist at Fraunhofer Institute for Secure Information Technology (SIT). His major research interests include information security, digital watermarking, robust hashing and digital forensics.

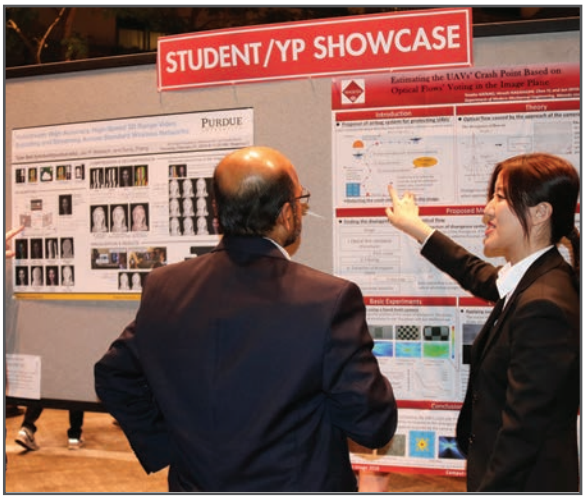
JOIN US AT THE NEXT EI!

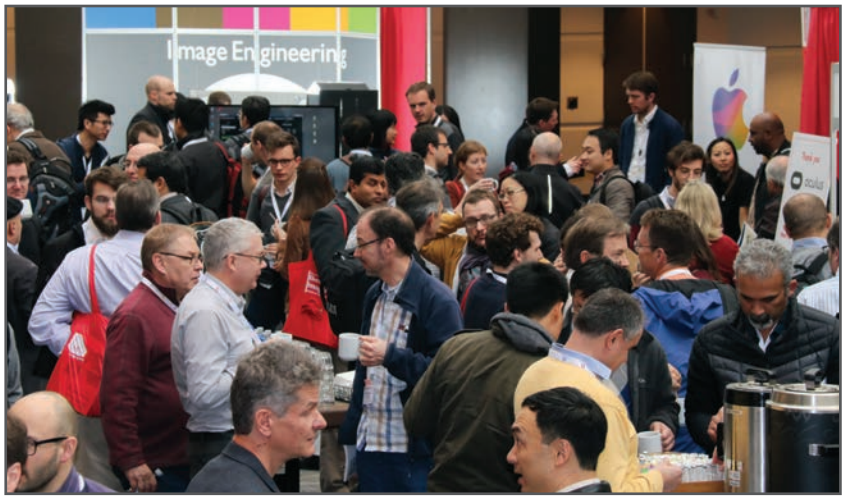
IS&T International Symposium on

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