COMPUTER VISION PERFORMANCE AND IMAGE QUALITY METRICS: A RECIPROCAL RELATION

Christopher Haccius and Thorsten Herfet

Telecommunications Lab, Saarland University, Saarbrücken, Germany
{haccius/herfet}@nt.uni-saarland.de

ABSTRACT

Computer vision algorithms are essential components of many systems in operation today. Predicting the robustness of such algorithms for different visual distortions is a task which can be approached with known image quality measures. We evaluate the impact of several image distortions on object segmentation, tracking and detection, and analyze the predictability of this impact given by image statistics, error parameters and image quality metrics. We observe that existing image quality metrics have shortcomings when predicting the visual quality of virtual or augmented reality scenarios. These shortcomings can be overcome by integrating computer vision approaches into image quality metrics. We thus show that image quality metrics can be used to predict the success of computer vision approaches, and computer vision can be employed to enhance the prediction capability of image quality metrics – a reciprocal relation.

KEYWORDS

Computer Vision Performance, Image Quality Assessment, Subjective Quality

1. INTRODUCTION

In today's world computer vision systems have become a central part of modern life. Computer vision in cars reads street signs and markers, in assembly lines checks production and processes, and almost every camera uses computer vision for face detection or artistic effects. In most scenarios Computer Vision is employed to analyze visual information. However, Computer Vision is also increasingly used to generate visual information, for example in augmented reality applications.

For all the different scenarios of computer vision the robustness of the computer vision algorithms is important. As robustness we consider the impact that common types of image errors have on a given computer vision algorithm. Classical image errors stem from image acquisition, and are given by thermal noise or blur. Each computer vision system relying on cameras needs to be robust against such noise, at least to a certain degree. Compression artifacts, like JPEG blocking or JPEG2000 ringing artifacts, become a matter of concern as soon as data for computer vision algorithms is retrieved from space limited storage or after distribution over throughput-limited channels, which make data size and respectively compression critical.

Today we see an increasing amount of visual information that is synthetically generated. For such content novel types of errors occur, which are scene composition errors. Such scene composition

David C. Wyld et al. (Eds) : CSITA, ISPR, ARIN, DMAP, CCSIT, AISC, SIPP, PDCTA, SOEN - 2017
pp. 27– 37, 2017. © CS & IT-CSCP 2017 DOI : 10.5121/csit.2017.70104

errors occur when synthetic objects are algorithmically merged with captured content, and the synthetic addition is positioned incorrectly, scaled wrongly or not aligned with the captured environment.

This paper analyzes the impact of classical image errors and novel scene composition errors on standard computer vision approaches. For this analysis an image database is necessary, which contains both classical image distortions and scene composition errors, and comes with additional information like ground truth segmentation data, object information or subjective evaluations. Section 2 introduces this database.

We analyze the impact of distortions for three very basic computer vision algorithms, which are object segmentation, object tracking and object detection. In Section 3 we introduce these computer vision algorithms and the experiments we have set up to analyze the distortion impact.
Knowing the impact of a distortion on computer vision algorithms for a single image is interesting, yet far more interesting is the ability to predict the impact of distortions. A good prediction can enable system designers to define robustness levels for computer vision systems. In Section 4 we correlate the distortion impact to three image quality metrics, which are subjective opinion scores, scores based on image statistics and scores based on the human visual system.

Finally, in Section 5, we switch the perspective. After having evaluated how image quality metrics can be used to predict computer vision performance, we introduce an approach which employs computer vision algorithms for enhanced image quality assessment. We thus have image quality metrics to predict computer vision performance and computer vision to enhance image quality metrics - a reciprocal relation.

2. THE IMAGE DATABASE

In order to conduct experiments according to the above mentioned motivation an image database is necessary that fulfills several requirements. Most importantly, color images are required which are distorted by classical errors such as noise and compression artifacts. Second, the images need to contain synthetic objects which can be modified to model scene composition errors. Third, ground truth segmentation data needs to be available to conduct object segmentation and tracking experiments. Last, to allow comparisons to subjective assessments, mean opinion scores (MOS) need to be available for the distorted images.

Several image databases exist that fulfill one or few of these requirements. The LIVE and TID2013 database [1], [2] are databases based on real images which are distorted by classical image errors and subjectively evaluated. Both databases lack the ability to modify scene objects, and ground truth segmentation data is missing. The BSDS500 database [3] contains images and segmentation data, yet also lacks the ability to modify objects and has no subjective evaluations for the images. As - to our knowledge - no suitable database exists, we present the Synthetic Image Database SSID, which we have created from fully synthetic scenes with the goal to enable scene composition modifications, image distortions and additional data like depth maps and segmentation data. This database was evaluated by human assessors, and MOS have been calculated for all distorted images [4].

Figure 1a shows a set of scenes contained in the database and presents a depth map (Figure 1b) and a segmentation map (Figure 1c) which can be easily rendered from the synthetic data. In the database Gaussian blur, white noise, JPEG and JPEG2000 coding artifacts as well as object scaling, rotation and translation are implemented as distortions. For the distorted images roughly 20.000 opinions have been obtained from 200 assessors in subjective evaluations, and mean

opinion scores have been calculated. The database is available online for image quality associated research [5].

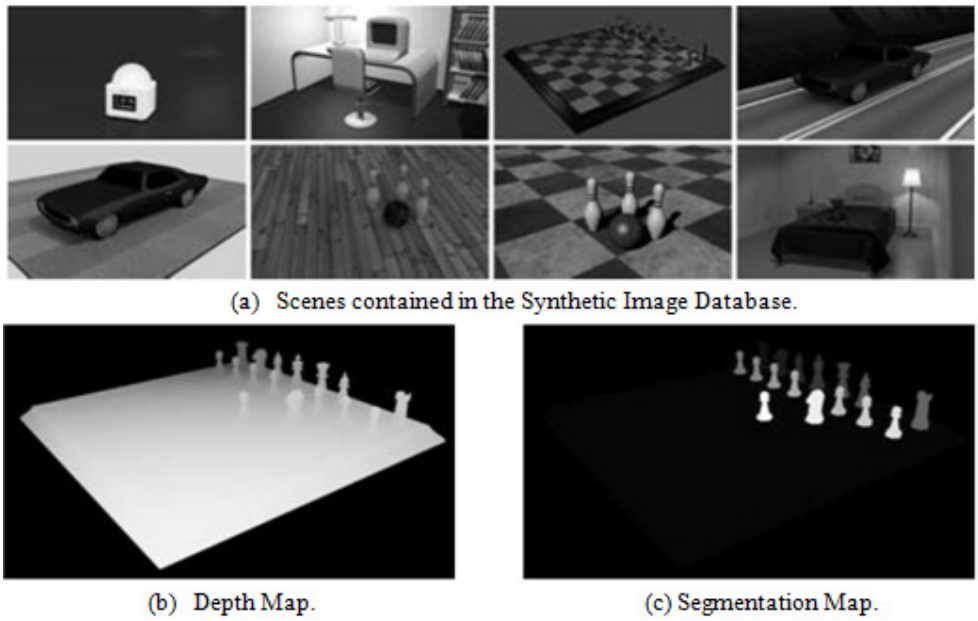

Figure 1.The Synthetic Image Database [4].

3. COMPUTER VISION – ALGORITHMS AND EXPERIMENTS

On the image database presented in Section 2 we test three computer vision algorithms, which are object segmentation, object tracking and object detection. By measuring the success of the computer vision algorithm we can derive the impact that a given image distortion has on a computer vision approach. For these tests we observe a certain object per scene (e.g. the computer, given in Figure 2 which is segmented, tracked and detected in the experiments described in the following sections.

3.1. Object Segmentation

A well established and widely used image segmentation approach is presented by Achanta et al. [6]. SLIC Superpixels are known for their high boundary recall at a low computational complexity, at the cost of oversegmentation. The low computational complexity makes SLIC Superpixel segmentation applicable even for real-time requirements. We evaluate experimentally, how much the boundary recall of SLIC Superpixel segmentation is affected by image distortions. Boundary recall br here is defined as the ratio between correctly recalled boundary pixels in the test image segmentation S and total number of boundary pixels in the ground truth segmentation T:

$$br(S,T) = \frac{bp(S,T)}{bp(T)} \quad (1)$$

where $bp(S,T)$ are the boundary pixels of ground truth T matching the boundary pixels of test image segmentation S, and $bp(T)$ are the boundary pixels of T only. We compare the boundary recall of each test image segmentation S to the boundary recall of its reference image segmentation R, and record the impact on segmentation $I_{segment}$ as the ratio

$$I_{segment} = \frac{br(S,T)}{br(R,T)} = \frac{bp(S,T)}{bp(R,T)}. \quad (2)$$

3.2. Object Tracking

Optical flow was introduced by Horn and Schunck in 1981 already [7] and presents a common basis for object tracking between frames. In our experiment we "track" scene objects from a reference image to the distorted test images. Using the ground truth segmentations we can evaluate how many object pixels of the distorted image are tracked correctly from the reference image. For a total number of pixels k and correctly tracked pixels t we calculate the impact on tracking I_{track} as the ratio

$$I_{track} = \frac{t}{k}. \quad (3)$$

Figure 2. Exemplary Training Object for Object Detection.

3.3. Object Detection

A common way of detecting objects in images is to compare image features. The Scale-Invariant Feature Transform (SIFT) was introduced by Lowe in 1999 [8], and in 2006 the faster Speeded-Up Robust Features (SURF) were made public by Hay et al. [9]. SURFs can be learned on a reference object, and then be used to detect the same object in a scene. For our experiment we create renderings of objects outside of their scene and train SURFs on this image. We than compare the matched SURFs between training object and reference image m to the matched SURFs between training object and test image n. The distortion impact on object detection I_{detect} is than

$$I_{detect} = \frac{m}{n}. \quad (4)$$

For all impact measures I it is I = 1 if segmentation, tracking and detection remains as good in the test image as in the reference, and I < 1 if the computer vision results are deteriorated in the test cases compared to the reference.

4. IMAGE QUALITY – METRICS AND RESULTS

Image quality is usually assessed using image statistics or methods modeling the human visual system. To predict the impact of image distortions on computer vision algorithms we calculate two algorithmic image quality metrics and compare to subjective quality scores as well. Mean Opinion Scores (MOS) are already available through subjective tests for the database described in Section 2.

For a reference image R and a test image T with dimensions $x \times y$ an average value expressing the overall statistical image error is the Mean Squared Error (MSE) calculated as

$$MSE(T,R) = \frac{1}{x \cdot y} \sum_{i=0}^{x} \sum_{j=0}^{y} \big(R(i,j) - T(i,j)\big)^2. \quad (5)$$

In the Peak Signal to Noise Ratio ($PSNR$) the MSE is related to the amplitude of the original signal:

$$PSNR(T,R) = \frac{max_{i\in[0,x]}\big(max_{j\in[0,y]}(R(i,j)^2)\big)}{MSE}. \quad (6)$$

Table 1. Table of Spearman Rank Correlations between Distortion Impact and Quality Measures.

	Noise	Compression	Transformation	All
$I_{segment}$ to MOS	0.429	0.557	0.365	0.415
$I_{segment}$ to $PSNR$	0.697	**0.574**	0.655	**0.650**
$I_{segment}$ to $SSIM$	0.649	0.371	**0.680**	0.545
$I_{segment}$ to $PARA$	**0.708**	0.015	0.023	0.171
I_{track} to MOS	0.372	0.636	0.349	0.592
I_{track} to $PSNR$	0.849	**0.693**	0.880	0.746
I_{track} to $SSIM$	**0.850**	0.626	**0.957**	**0.875**
I_{track} to $PARA$	0.649	0.095	0.039	0.056
I_{detect} to MOS	0.450	0.639	0.218	0.408
I_{detect} to $PSNR$	0.141	**0.643**	**0.571**	**0.490**
I_{detect} to $SSIM$	0.032	0.628	0.532	0.403
I_{detect} to $PARA$	**0.494**	0.112	0.085	0.165

The $PSNR$ is still a very common metric for image quality analysis. It can be easily implemented and has a very low computational complexity, which is an important criterion for real-time applications. For image quality assessment $PSNR$ has been shown to relate poorly to subjective image quality findings [10]. Therefore metrics based on the human visual system have been developed, of which the Structural Similarity ($SSIM$) is a widely established one [11]. The Structural Similarity index ($SSIM$) compares three different image components: luminance, contrast and structure. Structural similarity $SSIM$ between a test image T and a reference image R is calculated as the weighted product of luminance l, contrast c and structure s:

$$SSIM(T,R) = l(T,R)^{\alpha} \cdot c(T,R)^{\beta} \cdot s(R,T)^{\gamma} \quad (7)$$

with $0 < \alpha, \beta, \gamma$.

A fourth measure for image quality which we analyze in the context of this work is the parameter which was used to distort the image. As the scene composition errors are assigned with error parameters for 3 dimensions, we map the three parameters to one error parameter $PARA$ by calculating the absolute rotation angle, absolute size deviation and vector sum of transitions.

We than calculate the Spearman rank correlation [12] between the distortion impact on segmentation $I_{segment}$, tracking I_{track} and detection I_{detect} and the quality measures MOS, $PSNR$, $SSIM$ and $PARA$. Table 1 presents the correlation values for different image distortion classes. These distortion classes are noise (including white noise and Gaussian blur), compression artifacts (including JPEG and JPEG2000 compression artifacts), transformation errors (including object rotation, scaling and translation errors) and all (superclass of all previous classes). A bold font indicates the best correlation in each error class for one computer vision approach.

Table 1 indicates that $PSNR$ and $SSIM$ are good measures to predict the success of computer vision algorithms. Especially object tracking is very well correlated to $SSIM$ and $PSNR$. Only for the noise error class the error parameter (if obtainable) is suggested as a success indicator, while $SSIM$ has almost no correlation to the impact of noise on object detection. At the same time, when correlating the presented quality metrics $PSNR$, $SSIM$ and $PARA$ to the subjective MOS, it becomes clear that neither those presented metrics (nor any other metrics known to us) present suitable predictors for image distortions resulting from scene composition errors. This observation is confirmed by Caviedes at al. who note that subjective quality is more aesthetically-oriented whereas computer vision may have different quality requirements [13].

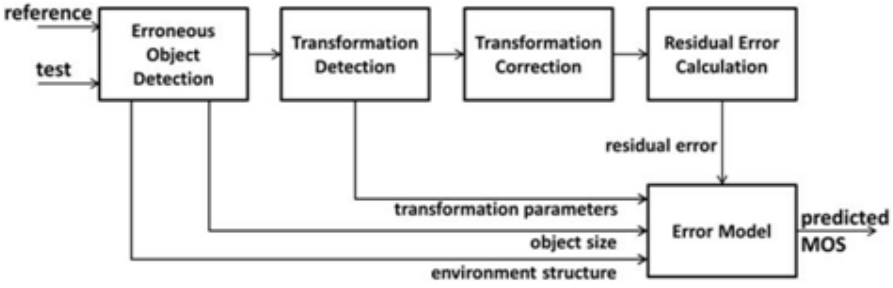

Figure 3. Structure of our proposed SC-VQM.

5. A RECIPROCAL RELATION

In [14] we approach this problem of predicting subjective image quality and develop a Visual Quality Metric for Synthetic Content (SC-VQM) that employs computer vision algorithms to better predict subjective image quality in virtual worlds or augmented reality scenarios. The approach to achieve this goal is straight forward: We detect object changes and correct those before calculating a residual error. This idea is outlined in the block diagram in Figure 3 and described by the following six steps:

1. Erroneous object detection: Distorted objects in a scene composition are detected
2. Erroneous object matching: Objects in test image are matched with objects in reference image
3. Object size calculation: The portion of the image affected by the distorted object is calculated
4. Environment structure analysis: The environment of the distorted objects is analyzed for the amount of structures contained
5. Object correction: The object in the test image is corrected according to the reference object, transformation parameters are recorded
6. Residual error calculation: The residual error between corrected object and reference image is calculated
7. Approximate MOS by detected parameters: All parameters from the previous analysis steps are combined in an error model to predict a MOS

5.1. Implementation

From a computer vision perspective the following three steps are most interesting: object detection, object matching and object correction. They closely relate to the above mentioned computer vision approaches of object segmentation, object detection and object tracking. To illustrate these three steps we employ a sample image, which is introduced in Figure 4.

a) Erroneous Object Detection: A characteristic of erroneous objects is that image errors accumulate in the areas of these objects. We use this characteristic and in a first step compute the

average image error as the Mean Square Error. If objects are misplaced the image error in these areas is above the average image error, while the error is below in other areas. By filtering the error areas with a disk-shaped stencil, object areas can be distinguished. Two things are important to note: First, the object outline is only rough, but covers the whole area in which an object is misplaced with respect to the original. Second, the averaging disk size depends on image size and viewing conditions, to differentiate between noise and relevant objects.

(a) Rerefence Image. (b) Test Image.

Figure 4. Example Image Set illustrating the implementation

The result of this detection step is a mask with outlined areas. If multiple objects in an image are moved, all of these areas are marked and noted. For the sample images shown in Figure 4 the object detection mask is given in Figure 5.

b) Erroneous Object Matching: To match objects between test and reference images there are two possible cases: a transformed object may be overlapping in reference and test image (only one erroneous region detected) or they may be spatially distinct (two erroneous regions). With the additional possibility to have several wrong objects in an image, we need to match each region with itself and with all other error regions. For region matching we employ Scale Invariant Features (SIFT) as proposed by Lowe [8]. For each area detected in the previous step we record the closest match between reference and test image. Figure 6 shows detected features between reference (top) and test image (bottom). The translation of the car between test and reference image can already clearly be seen by the feature lines (white) running slightly tilted between both images.

c) Object Correction: Reallocating the distorted object from the test image to its original position in the reference image is an important task to calculate the visual disturbance of the picture irrespective of any transformations. Initially, we remove the misplaced object from the test image and fill the created hole with an inpainting algorithm. Second, we use the SIFT feature correspondences to get a rough registration of the object in the test image [8]. As SIFT feature matching leaves inaccuracies in the order of single pixels we employ a Levenberg-Marquardt least-square optimization with a Fourier-Mellin transform module to achieve an image registration with sub-pixel precision for exact object placement [15]. The order of applying the SIFT registration before the Fourier-Mellin transform based registration is advantageous, as the SIFT registration works robustly, but with a certain inaccuracy, while the Fourier-Mellin transform becomes unstable for images that are too different from each other but works with a high precision when images are closely aligned already. Our implemented concatenation is both robust and precise. Finally, the registered object is fitted onto the filled background image. Filled background image and test image after object registration are shown in Figure 7. Next to the registered image this step retrieves the scaling, translation and rotation values between reference and test object.

(a) Mask outlining erroneous object. (b) Environment of erroneous object

Figure 5. Mask and Environment of Erroneous Object

Figure 6. SIFT Matching between test and reference image

5.2. Metric Results and Relation to Computer Vision

The SC-VQM analyzes scene objects for transformations, and employs detected transformation parameters as well as the object size and its environment structure for visual quality prediction. We have tested this metric on the $SSID$ database, presented in Section 2. A comparison of correlations between the different metrics shows that the SC-VQM increases the correlation between MOS and predicted MOS for transformation errors by 28% compared to currently existing and established metrics. This result is visualized exemplary in Figure 8. While SSIM assigns a MOS score of "Fair" to the image ($MOS_p = 3$), our metric evaluated the test image close to "Excellent" ($MOS_p = 4.6$). The statistical error map (Figure 8.c) indicates why traditional metrics fail: a shift in image textures causes large parts of the image to be fully wrong, yet the human brain judges this error to be fairly unimportant.

We therefore observe a reciprocal relation between Computer Vision Performance and Image Quality Metrics. Image Quality Metrics can be used to predict the performance of Computer Vision algorithms; Image Quality Metrics can therefore play an essential role in the design and development of Computer Vision algorithms. At the same time, ideas from Computer Vision are employed in Image Quality Metrics to increase the correlation between predicted quality and subjective evaluations.

(a) Filled Background. (b) Registered Object.

Figure 7. Filled Background and Registered Object for Example Image Set.

(a) Original. (b) Wrong Shirt. (c) Error Map.

Figure 8. Perceived quality does not always correspond to statistics.

6. CONCLUSION

In the previous sections we have shown that there exists a reciprocal relationship between image quality and computer vision. As a basis for research connecting both domains we have developed an image database, SSID, which contains synthetic images with classical image distortions and scene composition errors. This database is subjectively evaluated and contains additional data such as depth and segmentation maps, as well as the raw data to produce further information.

We have introduced three basic computer vision algorithms and four quality measures for visual information and have analyzed the image quality measures concerning their suitability for computer vision success prediction. Especially $PSNR$ and $SSIM$ were found to predict the impact of image distortions on computer vision algorithms well.

On the other hand we have observed that image quality metrics fail for visual content produced with computer vision approaches. Therefore a novel visual quality metric, SC-VQM, was developed, which is especially designed to analyze synthetic contents in virtual worlds or augmented reality scenarios. This metric can increase the quality prediction by 28% compared to current standard quality metrics.

Thus image quality metrics can be used to predict the success of computer vision approaches and computer vision can be employed to enhance the prediction capability of image quality metrics - a reciprocal relation.

REFERENCES

[1] Hamid R Sheikh, Zhou Wang, Lawrence Cormack, and Alan C Bovik, “Live image quality assessment database release 2,” 2005.

[2] Nikolay Ponomarenko, Oleg Ieremeiev, Vladimir Lukin, Karen Egiazarian, Lukui Jin, Jaakko Astola, Benoit Vozel, Kacem Chehdi, Marco Carli, Federica Battisti, et al., "Color image database tid2013: Peculiarities and preliminary results," in Visual Information Processing (EUVIP), 2013 4th European Workshop on. IEEE, 2013, pp. 106–111.

[3] D. Martin, C. Fowlkes, D. Tal, and J. Malik, "A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics," in Proc. 8th Int'l Conf. Computer Vision, July 2001, vol. 2, pp. 416–423.

[4] Christopher Haccius and Thorsten Herfet, "An image database for design and evaluation of visual quality metrics in synthetic scenarios," in International Conference on Image Analysis and Recognition, ICIAR, Póvoa de Varzim, Portugal. IEEE, July 2016.

[5] Christopher Haccius, "SSID - saarbru¨cken synthetic image database," http://ssid.nt.uni-saarland.de/, Apr. 2016, Accessed: 2016-06-14.

[6] Radhakrishna Achanta, Appu Shaji, Kevin Smith, Aurelien Lucchi, Pascal Fua, and Sabine Susstrunk, "Slic superpixels compared to state-of-the-art superpixel methods," Pattern Analysis and Machine Intelligence, IEEE Transactions on, vol. 34, no. 11, pp. 2274–2282, 2012.

[7] Berthold K Horn and Brian G Schunck, "Determining optical flow," in 1981 Technical symposium east. International Society for Optics and Photonics, 1981, pp. 319–331.

[8] David G Lowe, "Object recognition from local scale-invariant features," in Computer vision, 1999. The proceedings of the seventh IEEE international conference on. Ieee, 1999, vol. 2, pp. 1150–1157.

[9] Herbert Bay, Tinne Tuytelaars, and Luc Van Gool, "Surf: Speeded up robust features," in Computer vision–ECCV 2006, pp. 404–417. Springer, 2006.

[10] Zhou Wang, Hamid R Sheikh, and Alan C Bovik, "No-reference perceptual quality assessment of JPEG compressed images," in Image Processing. 2002. Proceedings. 2002 International Conference on. IEEE, 2002, vol. 1, pp. I–477.

[11] Zhou Wang, Alan C Bovik, Hamid R Sheikh, and Eero P Simoncelli, "Image quality assessment: from error visibility to structural similarity," Image Processing, IEEE Transactions on, vol. 13, pp. 600–612, 2004.

[12] Charles Spearman, "The proof and measurement of association between two things," The American journal of psychology, vol. 15, no. 1, pp. 72–101, 1904.

[13] Kalpana Seshadrinathan, Jorge E Caviedes, Audrey C Younkin, and Philip J Corriveau, "Visual quality management in consumer video r&d," in International Workshop on Video Processing and Quality Metrics (VPQM), 2010.

[14] Christopher Haccius and Thorsten Herfet, "SC-VQM - a visual quality metric for synthetic contents," in Picture Coding Symposium, PCS, Nuremberg, Germany (submitted to). IEEE, Dec. 2016.

[15] George Wolberg and Siavash Zokai, "Robust image registration using log-polar transform," in Image Processing, 2000. Proceedings. 2000 International Conference on. IEEE, 2000, vol. 1, pp. 493–496.

AUTHORS

Prof. Thorsten Herfet was born in Bochum, Germany, on April 26th, 1963. Thorsten received a Diploma in Electrical Engineering in 1988 and a Ph.D. in telecommunication in 1992. Having been a PostDoc for 4 years he joined industry in 1996, finally being appointed Director of Research & Innovation of GRUNDIG. In 2004 he rejoined academia and became Full University Professor at Saarland University. His fields of research are cyber-physical networking, low latency streaming, computational videography and high mobility.
Prof. Herfet 2006-2008 served as the Dean for Informatics and Mathematics, in 2009 has been appointed Director of Research and Operations of the Intel Visual Computing Institute at Saarland University and since 2014 is Saarland University's Vice President for Research and Technology-Transfer. Thorsten published more than 100 papers, holds 15+ patents and has initiated and led several multi-million € collaborative research projects. He is a Senior Member of the IEEE, member of ACM SIGGRAPH, member of the German VDI/FKTG and serves as Steering Board member and Curator for various consortia and institutes.

Christopher Haccius was born in Lünen, Germany, on December 7, 1986. After his university-entrance diploma he received a BSc in Computer Science from the International University in Germany, Bruchsal, in 2009 and a MSc in Computer Science with focus on computer vision and telecommunications from Saarland University, Saarbrücken, Germany, in 2013. As a researcher his major fields of study are computer vision and telecommunications.
Mr. Haccius has work experience as a software developer with Fluid Operations and as a researcher at the telecommunications lab of Saarland University. Currently he is working for Continental Automotive in R&D of vehicle electronics.