Grand Challenges in Image Processing

Frédéric Dufaux

Grand Challenges in Image Processing

Frédéric Dufaux
Université Paris-Saclay, CNRS, CentraleSupélec, Laboratoire des signaux et systèmes
91190 Gif-sur-Yvette, France
frederic.dufaux@l2s.centralesupelec.fr

Introduction

The field of image processing has been the subject of intensive research and development activities for several decades. This broad area encompasses topics such as image/video processing, image/video analysis, image/video communications, image/video sensing, modeling and representation, computational imaging, electronic imaging, information forensics and security, 3D imaging, medical imaging, and machine learning applied to these respective topics. Hereafter, we will consider both image and video content (i.e. sequence of images), and more generally all forms of visual information.

Rapid technological advances, especially in terms of computing power and network transmission bandwidth, have resulted in many remarkable and successful applications. Nowadays, images are ubiquitous in our daily life. Entertainment is one class of applications that has greatly benefited, including digital TV (e.g. broadcast, cable, and satellite TV), Internet video streaming, digital cinema, and video games. Beyond entertainment, imaging technologies are central in many other applications, including digital photography, video conferencing, video monitoring and surveillance, satellite imaging, but also in more distant domains such as healthcare and medicine, distance learning, digital archiving, cultural heritage or the automotive industry.

In this paper, we highlight a few research grand challenges for future imaging and video systems, in order to achieve breakthroughs to meet the growing expectations of end users. Given the vastness of the field, this list is by no means exhaustive.

A brief historical perspective

We first briefly discuss a few key milestones in the field of image processing. Key inventions in the development of photography and motion pictures can be traced to the 19th century. The earliest surviving photograph of a real-world scene was made by Nicéphore Niépce in 1827 [Hirsch, 1999]. The Lumière brothers made the first cinematographic film in 1895, with a public screening the same year [Lumière, 1996]. After decades of remarkable developments, the second half of the 20th century saw the emergence of new technologies launching the digital revolution. While the first prototype digital camera using a Charge-Coupled Device (CCD) was demonstrated in 1975, the first commercial consumer digital cameras started appearing in the early 1990s. These digital cameras quickly surpassed cameras using films and the digital revolution in the field of imaging was underway. As a key consequence, the digital process enabled computational imaging, in other words the use of sophisticated processing algorithms in order to produce high quality images.

In 1992, the Joint Photographic Experts Group (JPEG) released the JPEG standard for still image coding [Wallace, 1992]. In parallel, in 1993, the Moving Picture Experts Group (MPEG) published its first standard for coding of moving pictures and associated audio, MPEG-1 [Le Gall, 1991], and a few years later MPEG-2 [Haskell et al., 1996]. By guaranteeing interoperability, these standards have been essential in many successful applications and services, for both the consumer and business markets. In

particular, it is remarkable that, almost 30 years later, JPEG remains the dominant format for still images and photographs.

In the late 2000s and early 2010s, we could observe a paradigm shift with the appearance of smartphones integrating a camera. Thanks to advances in computational photography, these new smartphones soon became capable of rivaling the quality of consumer digital cameras at the time. Moreover, these smartphones were also capable of acquiring video sequences. Almost concurrently, another key evolution was the development of high bandwidth networks. In particular, the launch of 4G wireless services circa 2010 enabled users to quickly and efficiently exchange multimedia content. From this point, most of us are carrying a camera, anywhere and anytime, allowing to capture images and videos at will and to seamlessly exchange them with our contacts.

As a direct consequence of the above developments, we are currently observing a boom in the usage of multimedia content. It is estimated that today 3.2 billion images are shared each day on social media platforms, and 300 hours of video are uploaded every minute on YouTube[1]. In a 2019 report, Cisco estimated that video content represented 75% of all Internet traffic in 2017, and this share is forecasted to grow to 82% in 2022 [Cisco, 2019]. While Internet video streaming and Over-The-Top (OTT) media services account for a significant bulk of this traffic, other applications are also expected to see significant increases, including video surveillance and Virtual Reality (VR) / Augmented Reality (AR).

Hyper-realistic and immersive imaging

A major direction and key driver to research and development activities over the years has been the objective to deliver an ever-improving image quality and user experience.

For instance, in the realm of video, we have observed constantly increasing spatial and temporal resolutions, with the emergence nowadays of Ultra High Definition (UHD). Another aim has been to provide a sense of the depth in the scene. For this purpose, various 3D video representations have been explored, including stereoscopic 3D and multi-view [Dufaux et al., 2013].

In this context, the ultimate goal is to be able to faithfully represent the physical world and to deliver an immersive and perceptually hyperrealist experience. For this purpose, we discuss hereafter some emerging innovations. These developments are also very relevant in VR and AR applications [Slater, 2014]. Finally, while this paper is only focusing on the visual information processing aspects, it is obvious that emerging display technologies [Masia et al., 2013] and audio also plays key roles in many application scenarios.

Light fields, point clouds, volumetric imaging

In order to wholly represent a scene, the light information coming from all the directions has to be represented. For this purpose, the 7D plenoptic function is a key concept [Adelson and Bergen, 1991], although it is unmanageable in practice.

By introducing additional constraints, the light field representation collects radiance from rays in all directions. Therefore, it contains a much richer information, when compared to traditional 2D imaging that captures a 2D projection of the light in the scene integrating the angular domain. For instance, this allows post-capture processing such as refocusing and changing the viewpoint. However, it also entails several technical challenges, in terms of acquisition and calibration, as well as computational image processing steps including depth estimation, super-resolution, compression and image synthesis

[1] https://www.brandwatch.com/blog/amazing-social-media-statistics-and-facts/ (accessed on Feb. 23, 2021)

[Ihrke et al., 2016; Wu et al., 2017]. The resolution trade-off between spatial and angular resolutions is a fundamental issue. With a significant fraction of the earlier work focusing on static light fields, it is also expected that dynamic light field videos will stimulate more interest in the future. In particular, dense multi-camera arrays are becoming more tractable. Finally, the development of efficient light field compression and streaming techniques is a key enabler in many applications [Conti et al., 2020].

Another promising direction is to consider a point cloud representation. A point cloud is a set of points in the 3D space represented by their spatial coordinates and additional attributes, including color pixel values, normals, or reflectance. They are often very large, easily ranging in the millions of points, and are typically sparse. One major distinguishing feature of point clouds is that, unlike images, they do not have a regular structure, calling for new algorithms. To remove the noise often present in acquired data, while preserving the intrinsic characteristics, effective 3D point cloud filtering approaches are needed [Han et al., 2017]. It is also important to develop efficient techniques for Point Cloud Compression (PCC). For this purpose, MPEG is developing two standards: Geometry-based PCC (G-PCC) and Video-based PCC (V-PCC) [Graziosi et al., 2020]. G-PCC considers the point cloud in its native form and compress it using 3D data structures such as octrees. Conversely, V-PCC projects the point cloud onto 2D planes and then applies existing video coding schemes. More recently, deep learning-based approaches for PCC have been shown to be effective [Guarda et al., 2020]. Another challenge is to develop generic and robust solutions able to handle potentially widely varying characteristics of point clouds, e.g. in terms of size and non-uniform density. Efficient solutions for dynamic point clouds are also needed. Finally, while many techniques focus on the geometric information or the attributes independently, it is paramount to process them jointly.

High Dynamic Range and Wide Color Gamut

The human visual system is able to perceive, using various adaptation mechanisms, a broad range of luminous intensities, from very bright to very dark, as experienced every day in the real world. Nonetheless, current imaging technologies are still limited in terms of capturing or rendering such a wide range of conditions. High Dynamic Range (HDR) imaging aims at addressing this issue. Wide Color Gamut (WCG) is also often associated with HDR in order to provide a wider colorimetry.

HDR has reached some levels of maturity in the context of photography. However, extending HDR to video sequences raises scientific challenges in order to provide high quality and cost-effective solutions, impacting the whole imaging processing pipeline, including content acquisition, tone reproduction, color management, coding, and display [Dufaux et al., 2016; Chalmers and Debattista, 2017]. Backward compatibility with legacy content and traditional systems is another issue. Despite recent progress, the potential of HDR has not been fully exploited yet.

Coding and transmission

Three decades of standardization activities have continuously improved the hybrid video coding scheme based on the principles of transform coding and predictive coding. The Versatile Video Coding (VVC) standard has been finalized in 2020 [Bross et al., 2021], achieving approximately 50% bit rate reduction for the same subjective quality when compared to its predecessor, High Efficiency Video Coding (HEVC). While substantially outperforming VVC in the short term may be difficult, one encouraging direction is to rely on improved perceptual models to further optimize compression in terms of visual quality. Another direction, which has already shown promising results, is to apply deep learning-based approaches [Ding et al., 2021]. Here, one key issue is the ability to generalize these deep models to a wide diversity of video content. The second key issue is the implementation complexity, both in terms of computation and memory requirements, which is a significant obstacle to a widespread deployment. Besides, the emergence of new video formats targeting immersive communications is also calling for new coding schemes [Wien et al., 2019].

Considering that in many application scenarios, videos are processed by intelligent analytic algorithms rather than viewed by users, another interesting track is the development of video coding for machines [Duan et al., 2020]. In this context, the compression is optimized taking into account the performance of video analysis tasks.

The push towards hyper-realistic and immersive visual communications entails most often an increasing raw data rate. Despite improved compression schemes, more transmission bandwidth is needed. Moreover, some emerging applications, such as VR/AR, autonomous driving, and Industry 4.0, bring a strong requirement for low latency transmission, with implications on both the imaging processing pipeline and the transmission channel. In this context, the emergence of 5G wireless networks will positively contribute to the deployment of new multimedia applications, and the development of future wireless communication technologies points towards promising advances [Da Costa and Yang, 2020].

Human perception and visual quality assessment

It is important to develop effective models of human perception. On the one hand, it can contribute to the development of perceptually inspired algorithms. On the other hand, perceptual quality assessment methods are needed in order to optimize and validate new imaging solutions.

The notion of Quality of Experience (QoE) relates to the degree of delight or annoyance of the user of an application or service [Le Callet et al., 2012]. QoE is strongly linked to subjective and objective quality assessment methods. Many years of research have resulted in the successful development of perceptual visual quality metrics based on models of human perception [Lin and Kuo, 2011; Bovik, 2013]. More recently, deep learning-based approaches have also been successfully applied to this problem [Bosse et al., 2017]. While these perceptual quality metrics have achieved good performances, several significant challenges remain. First, when applied to video sequences, most current perceptual metrics are applied on individual images, neglecting temporal modeling. Second, whereas color is a key attribute, there are currently no widely accepted perceptual quality metrics explicitly considering color. Finally, new modalities, such as 360° videos, light fields, point clouds, and HDR, require new approaches.

Another closely related topic is image aesthetic assessment [Deng et al., 2017]. The aesthetic quality of an image is affected by numerous factors, such as lighting, color, contrast, and composition. It is useful in different application scenarios such as image retrieval and ranking, recommendation, and photos enhancement. While earlier attempts have used handcrafted features, most recent techniques to predict aesthetic quality are data driven and based on deep learning approaches, leveraging the availability of large annotated datasets for training [Murray et al., 2012]. One key challenge is the inherently subjective nature of aesthetics assessment, resulting in ambiguity in the ground-truth labels. Another important issue is to explain the behavior of deep aesthetic prediction models.

Analysis, interpretation and understanding

Another major research direction has been the objective to efficiently analyze, interpret and understand visual data. This goal is challenging, due to the high diversity and complexity of visual data. This has led to many research activities, involving both low-level and high-level analysis, addressing topics such as image classification and segmentation, optical flow, image indexing and retrieval, object detection and tracking, and scene interpretation and understanding. Hereafter, we discuss some trends and challenges.

Keypoints detection and local descriptors

Local imaging matching has been the cornerstone of many analysis tasks. It involves the detection of keypoints, i.e. salient visual points that can be robustly and repeatedly detected, and descriptors, i.e. a

compact signature locally describing the visual features at each keypoint. It allows to subsequently compute pairwise matching between the features to reveal local correspondences. In this context, several frameworks have been proposed, including Scale Invariant Feature Transform (SIFT) [Lowe, 2004] and Speeded Up Robust Features (SURF) [Bay et al., 2008], and later binary variants including Binary Robust Independent Elementary Feature (BRIEF) [Calonder et al., 2010], Oriented FAST and Rotated BRIEF (ORB) [Rublee et al., 2011] and Binary Robust Invariant Scalable Keypoints (BRISK) [Leutenegger et al., 2011]. Although these approaches exhibit scale and rotation invariance, they are less suited to deal with large 3D distortions such as perspective deformations, out-of-plane rotations, and significant viewpoint changes. Besides, they tend to fail under significantly varying and challenging illumination conditions.

These traditional approaches based on handcrafted features have been successfully applied to problems such as image and video retrieval, object detection, visual Simultaneous Localization And Mapping (SLAM), and visual odometry. Besides, the emergence of new imaging modalities as introduced above can also be beneficial for image analysis tasks, including light fields [Galdi et al., 2019], point clouds [Guo et al., 2020], and HDR [Rana et al., 2018]. However, when applied to high-dimensional visual data for semantic analysis and understanding, these approaches based on handcrafted features have been supplanted in recent years by approaches based on deep learning.

Deep learning-based methods

Data-driven deep learning-based approaches [LeCun et al., 2015], and in particular the Convolutional Neural Network (CNN) architecture, represent nowadays the state-of-the-art in terms of performances for complex pattern recognition tasks in scene analysis and understanding. By combining multiple processing layers, deep models are able to learn data representations with different levels of abstraction.

Supervised learning is the most common form of deep learning. It requires a large and fully labeled training dataset, a typically time-consuming and expensive process needed whenever tackling a new application scenario. Moreover, in some specialized domains, e.g. medical data, it can be very difficult to obtain annotations. To alleviate this major burden, methods such as transfer learning and weakly supervised learning have been proposed.

In another direction, deep models have been shown to be vulnerable to adversarial attacks [Akhtar and Mian, 2018]. Those attacks consist in introducing subtle perturbations to the input, such that the model predicts an incorrect output. For instance, in the case of images, imperceptible pixel differences are able to fool deep learning models. Such adversarial attacks are definitively an important obstacle to the successful deployment of deep learning, especially in applications where safety and security are critical. While some early solutions have been proposed, a significant challenge is to develop effective defense mechanisms against those attacks.

Finally, another challenge is to enable low complexity and efficient implementations. This is especially important for mobile or embedded applications. For this purpose, further interactions between signal processing and machine learning can potentially bring additional benefits. For instance, one direction is to compress deep neural networks in order to enable their more efficient handling. Moreover, by combining traditional processing techniques with deep learning models, it is possible to develop low complexity solutions while preserving high performance.

Explainability in deep learning

While data-driven deep learning models often achieve impressive performances on many visual analysis tasks, their black-box nature often makes it inherently very difficult to understand how they reach a predicted output and how it relates to particular characteristics of the input data. However, this is a major impediment in many decision-critical application scenarios. Moreover, it is important not

only to have confidence in the proposed solution, but also to gain further insights from it. Based on these considerations, some deep learning systems aim at promoting explainability [Adadi and Berrada, 2018; Xie et al., 2020]. This can be achieved by exhibiting traits related to confidence, trust, safety, and ethics.

However, explainable deep learning is still in its early phase. More developments are needed, in particular to develop a systematic theory of model explanation. Important aspects include the need to understand and quantify risk, to comprehend how the model makes predictions for transparency and trustworthiness, and to quantify the uncertainty in the model prediction. This challenge is key in order to deploy and use deep learning-based solutions in an accountable way, for instance in application domains such as healthcare or autonomous driving.

Self-supervised learning

Self-supervised learning refers to methods that learn general visual features from large-scale unlabeled data, without the need for manual annotations. Self-supervised learning is therefore very appealing, as it allows exploiting the vast amount of unlabeled images and videos available. Moreover, it is widely believed that it is closer to how humans actually learn. One common approach is to use the data to provide the supervision, leveraging its structure. More generally, a pretext task can be defined, e.g. image inpainting, colorizing grayscale images, predicting future frames in videos, by withholding some parts of the data and by training the neural network to predict it [Jing and Tian, 2020]. By learning an objective function corresponding to the pretext task, the network is forced to learn relevant visual features in order to solve the problem. Self-supervised learning has also been successfully applied to autonomous vehicles perception. More specifically, the complementarity between analytical and learning methods can be exploited to address various autonomous driving perception tasks, without the prerequisite of an annotated data set [Chiaroni et al., 2021].

While good performances have already been obtained using self-supervised learning, further work is still needed. A few promising directions are outlined hereafter. Combining self-supervised learning with other learning methods is a first interesting path. For instance, semi-supervised learning [Van Engelen and Hoos, 2020] and few-short learning [Fei-Fei et al., 2006] methods have been proposed for scenarios where limited labeled data is available. The performance of these methods can potentially be boosted by incorporating a self-supervised pre-training. The pretext task can also serve to add regularization. Another interesting trend in self-supervised learning is to train neural networks with synthetic data. The challenge here is to bridge the domain gap between the synthetic and real data. Finally, another compelling direction is to exploit data from different modalities. A simple example is to consider both the video and audio signals in a video sequence. In another example in the context of autonomous driving, vehicles are typically equipped with multiple sensors, including cameras, LIght Detection And Ranging (LIDAR), Global Positioning System (GPS), and Inertial Measurement Units (IMU). In such cases, it is easy to acquire large unlabeled multimodal datasets, where the different modalities can be effectively exploited in self-supervised learning methods.

Reproducible research and large public datasets

The reproducible research initiative is another way to further ensure high-quality research for the benefit of our community [Vandewalle et al., 2009]. Reproducibility, referring to the ability by someone else working independently to accurately reproduce the results of an experiment, is a key principle of the scientific method. In the context of image and video processing, it is usually not sufficient to provide a detailed description of the proposed algorithm. Most often, it is essential to also provide access to the code and data. This is even more imperative in the case of deep learning-based models.

In parallel, the availability of large public datasets is also highly desirable in order to support research activities. This is especially critical for new emerging modalities or specific application scenarios, where it is difficult to get access to relevant data. Moreover, with the emergence of deep learning, large datasets, along with labels, are often needed for training, which can be another burden.

Conclusion and perspectives

The field of image processing is very broad and rich, with many successful applications in both the consumer and business markets. However, many technical challenges remain in order to further push the limits in imaging technologies. Two main trends are on the one hand to always improve the quality and realism of image and video content, and on the other hand to be able to effectively interpret and understand this vast and complex amount of visual data. However, the list is certainly not exhaustive and there are many other interesting problems, e.g. related to computational imaging, information security and forensics, or medical imaging. Key innovations will be found at the crossroad of image processing, optics, psychophysics, communication, computer vision, artificial intelligence, and computer graphics. Multi-disciplinary collaborations are therefore critical moving forward, involving actors from both academia and the industry, in order to drive these breakthroughs.

The “Image Processing” section of Frontier in Signal Processing aims at giving to the research community a forum to exchange, discuss and improve new ideas, with the goal to contribute to the further advancement of the field of image processing and to bring exciting innovations in the foreseeable future.

References

Adadi, A., & Berrada, M. (2018). Peeking inside the black-box: a survey on explainable artificial intelligence (XAI). IEEE access, 6, 52138-52160.

Adelson, E. H. and Bergen, J. R. (1991). The plenoptic function and the elements of early vision, In *Computational Models of Visual Processing*, MIT Press, Cambridge, MA.

Akhtar, N., & Mian, A. (2018). Threat of adversarial attacks on deep learning in computer vision: A survey. IEEE Access, 6, 14410-14430.

Bay, H., Ess, A., Tuytelaars, T., & Van Gool, L. (2008). Speeded-up robust features (SURF). Computer vision and image understanding, 110(3), 346-359.

Bosse, S., Maniry, D., Müller, K. R., Wiegand, T., & Samek, W. (2017). Deep neural networks for no-reference and full-reference image quality assessment. IEEE Transactions on image processing, 27(1), 206-219.

Bovik, A. C. (2013). Automatic prediction of perceptual image and video quality. Proceedings of the IEEE, 101(9), 2008-2024.

Bross, B., Chen, J., Ohm, J. R., Sullivan, G. J., & Wang, Y. K. (2021). Developments in international video coding standardization after AVC, with an overview of Versatile Video Coding (VVC). Proceedings of the IEEE.

Calonder, M., Lepetit, V., Strecha, C., & Fua, P. (2010). Brief: Binary robust independent elementary features. In European conference on computer vision (pp. 778-792). Springer, Berlin, Heidelberg.

Chiaroni, F., Rahal, M. C., Hueber, N., & Dufaux, F. (2021). Self-supervised learning for autonomous vehicles perception: A conciliation between analytical and learning methods. IEEE Signal Processing Magazine, 38(1), 31-41.

Cisco (2019). Cisco Visual Networking Index: Forecast and Trends, 2017-2022 (white paper), Feb. 2019.

Chalmers, A. and Debattista, K. (2017). HDR video past, present and future: A perspective. Signal Processing: Image Communication 54, pp 49-55, 2017.

Conti, C., Soares, L. D. and Nunes, P. (2020). Dense Light Field Coding: A Survey, in IEEE Access, vol. 8, pp. 49244-49284, 2020, doi: 10.1109/ACCESS.2020.2977767.

Da Costa, D.B. and Yang, H.-C. (2020). Grand Challenges in Wireless Communications. Frontiers in Communications and Networks, vol. 1, p. 1, 2020.

Deng, Y., Loy, C. C., & Tang, X. (2017). Image aesthetic assessment: An experimental survey. IEEE Signal Processing Magazine, 34(4), 80-106.

Ding, D., Ma, Z., Chen, D., Chen, Q., Liu, Z. and Zhu, F. (2021). Advances In Video Compression System Using Deep Neural Network: A Review And Case Studies, arXiv, 2021.

Duan, L., Liu, J., Yang, W., Huang, T., and Gao, W. (2020). Video coding for machines: A paradigm of collaborative compression and intelligent analytics. IEEE Transactions on Image Processing, 29, 8680-8695.

Dufaux, F., Pesquet-Popescu, B. and Cagnazzo, M. (2013) Emerging Technologies for 3D Video: Creation, Coding, Transmission and Rendering, Wiley.

Dufaux, F., Le Callet, P., Mantiuk, R. and Mrak, M. (2016). High Dynamic Range Video - From Acquisition, to Display and Applications, Academic Press.

Fei-Fei, L., Fergus, R., & Perona, P. (2006). One-shot learning of object categories. IEEE transactions on pattern analysis and machine intelligence, 28(4), 594-611.

Galdi, C., Chiesa, V., Busch, C., Lobato Correia, P., Dugelay, J. L., & Guillemot, C. (2019). Light fields for face analysis. Sensors, 19(12), 2687.

Graziosi, D., Nakagami, O., Kuma, S., Zaghetto, A., Suzuki, T., & Tabatabai, A. (2020). An overview of ongoing point cloud compression standardization activities: Video-based (V-PCC) and geometry-based (G-PCC), APSIPA Transactions on Signal and Information Processing, 9, 2020, doi:10.1017/ATSIP.2020.12

Guarda, A., Rodrigues, N., & Pereira, F. (2020). Adaptive Deep Learning-based Point Cloud Geometry Coding. IEEE Journal of Selected Topics in Signal Processing.

Guo, Y., Wang, H., Hu, Q., Liu, H., Liu, L., & Bennamoun, M. (2020). Deep learning for 3D point clouds: A survey. IEEE transactions on pattern analysis and machine intelligence

Han, X. F., Jin, J. S., Wang, M. J., Jiang, W., Gao, L., and Xiao, L. (2017). A review of algorithms for filtering the 3D point cloud. Signal Processing: Image Communication, 57, 103-112.

Haskell, B. G., Puri, A., & Netravali, A. N. (1996). Digital video: an introduction to MPEG-2. Springer Science & Business Media.

Hirsch, R. (1999). Seizing the light: a history of photography. McGraw-Hill. ISBN 978-0697143617.

Ihrke, I., Restrepo, J. and Mignard-Debise, L. (2016). Principles of Light Field Imaging: Briefly revisiting 25 years of research, in IEEE Signal Processing Magazine, vol. 33, no. 5, pp. 59-69, Sept. 2016, doi: 10.1109/MSP.2016.2582220.

Jing, L., & Tian, Y. (2020). Self-supervised visual feature learning with deep neural networks: A survey. IEEE Transactions on Pattern Analysis and Machine Intelligence.

Le Callet, P., Möller, S., and Perkis, A. (2012). Qualinet white paper on definitions of quality of experience. European network on quality of experience in multimedia systems and services (COST Action IC 1003), 3(2012)

Le Gall, D. (1991). MPEG: a video compression standard for multimedia applications. Commun. ACM 34, 4 (April 1991), 46–58. DOI: 10.1145/103085.103090

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. nature, 521(7553), 436-444.

Leutenegger, S., Chli, M., & Siegwart, R. Y. (2011, November). BRISK: Binary robust invariant scalable keypoints. In 2011 IEEE International conference on computer vision (pp. 2548-2555).

Lin, W., & Kuo, C. C. J. (2011). Perceptual visual quality metrics: A survey. Journal of visual communication and image representation, 22(4), 297-312.

Lowe, D. G. (2004). Distinctive image features from scale-invariant keypoints. International journal of computer vision, 60(2), 91-110.

Lumiere, L. (1996). 1936 the Lumière Cinematograph. SMPTE Journal, 105(10), 608-611.

Masia, B., Wetzstein, G., Didyk, P. and Gutierrez, D. (2013). A survey on computational displays: Pushing the boundaries of optics, computation, and perception, Computers & Graphics, Volume 37, Issue 8, 2013, Pages 1012-1038, doi: 10.1016/j.cag.2013.10.003.

Murray, N., Marchesotti, L., & Perronnin, F. (2012, June). AVA: A large-scale database for aesthetic visual analysis. In 2012 IEEE Conference on Computer Vision and Pattern Recognition (pp. 2408-2415). IEEE.

Rana, A., Valenzise, G., & Dufaux, F. (2018). Learning-based tone mapping operator for efficient image matching. IEEE Transactions on Multimedia, 21(1), 256-268.

Rublee, E., Rabaud, V., Konolige, K., & Bradski, G. (2011, November). ORB: An efficient alternative to SIFT or SURF. In 2011 IEEE International conference on computer vision (pp. 2564-2571).

Slater, M. (2014). Grand challenges in virtual environments. Frontiers in Robotics and AI, vol. 1, p. 3.

Van Engelen, J. E., & Hoos, H. H. (2020). A survey on semi-supervised learning. Machine Learning, 109(2), 373-440.

Vandewalle, P., Kovacevic, J., & Vetterli, M. (2009). Reproducible research in signal processing. IEEE Signal Processing Magazine, 26(3), 37-47.

Wallace, G. K. (1992). The JPEG still picture compression standard, in IEEE Transactions on Consumer Electronics, vol. 38, no. 1, pp. xviii-xxxiv, Feb. 1992, doi: 10.1109/30.125072.

Wien, M., Boyce, J. M., Stockhammer, T. and Peng, W. (2019). Standardization Status of Immersive Video Coding, in IEEE Journal on Emerging and Selected Topics in Circuits and Systems, vol. 9, no. 1, pp. 5-17, March 2019, doi: 10.1109/JETCAS.2019.2898948.

Wu, G., Masia, B., Jarabo, A., Zhang, Y., Wang, L., Dai, Q., Chai, T. and Liu, Y. (2017). Light Field Image Processing: An Overview, in IEEE Journal of Selected Topics in Signal Processing, vol. 11, no. 7, pp. 926-954, Oct. 2017, doi: 10.1109/JSTSP.2017.2747126.

Xie, N., Ras, G., van Gerven, M., & Doran, D. (2020). Explainable deep learning: A field guide for the uninitiated. arXiv preprint arXiv:2004.14545.