 doi:10.1145/3206004.3206007

Applicability of No-Reference Visual Quality Indices for Visual Security Assessment

Heinz Hofbauer[1] • Andreas Uhl[1]

[1]University of Salzburg, Department of Computer Sciences, {hhofbaue, uhl}@cosy.sbg.ac.at

Abstract

From literature it is known that full-reference visual quality indices are a poor fit for the estimation of visual security for selective encryption. The question remains whether no-reference visual quality indices can perform where full reference indices falter. Furthermore, no-reference visual quality indices frequently use machine learning to train a model of natural scene statistics. It would be of interest to be able to gauge the impact of learning statistics from selectively encrypted images on performance as quality estimators for encryption. In the following we will answer these two questions.

Contents

1 Introduction

Selective encryption (SE) is the encryption, utilizing state of the art ciphers like AES, of a *selected* part of a media file or stream. The goal is to secure the content, or parts thereof, while still maintaining the file format, that is the file is still usable as the media file or stream it actually is. An example of this is would be surveillance cameras, which are more and more common and pose some privacy concerns since a malicious actor could track a person. A solution for this is to encrypt identity revealing information like faces or license plates. On the other hand, an observer should still be able to see the rest of the video(-stream) in high quality to be able to identify suspicious behaviour. If such behaviour is identified and a law enforcement agent is notified the identity containing portion of the video can be decrypted. There are a number of papers dealing with problems like this, focusing on the technical side of performing encryption without destroying the rest of the video through drift errors and similar things, e.g., [1–4].

It is well known that SE is secure, Lookabaugh and Sicker [5] showed how it relates to Shannon's work [6]. Potential insecurity come from the information left in plaintext to achieve the other objectives, e.g., the rest of the video outside a region of identity should be error free and the media should be format-compliant. In essence some information will leak, this is unavoidable with the other goals in mind, the amount of leakage is another question though. The goal of SE can be stated as a destruction of quality in the encrypted part such that it can not be exploited in a malicious fashion. Which directly leads to the formulation of the security evaluation of the SE: "To what extent can the information left in plain text be used to reconstruct an image or video?"

It is expensive and time-consuming to use human observer experiments, to gauge recognizability and remaining quality, so image metrics or visual quality indices (VQI) are used instead as they represent a model of the human visual system. While such methods can be, and are, used to evaluate SE methods it has been shown that their performance suffers on low quality images [7]. Furthermore, there is a subset of VQIs specifically designed to evaluate SE methods, created in conjuction with a specific SE method. These SE-VQIs usually work well for the design target but are even worse than regular VQIs for all other applications [8]. Up to now the focus VQI regarding security evaluation are on full-reference VQIs (FR-VQI), i.e., metrics which have access to both the original and distorted (encrypted) image. Even evaluation of the fitness of VQI for security evaluation of SEs focus almost exclusively on FR-VQI, with one exception in [8] which is based on local entropy to evaluate the distance to a random signal.

An alternative are no-reference VQIs (NR-VQI) which only use the distorted image and statistics based on natural images. The benefit of NR-VQIs over FR-VQIs for security evaluation is threefold. NR-VQIs use a statistical approach which is better fit to deal with the artefacts resulting from encryption, which should look like a random signal, while FR-VQIs consist of models of image structure and the HVS. The second benefit of NR-VQIs is that the rely on a statistical model of images and distortions. This model is the result of a machine-learning process and can, as opposed to a fixed model (as used by most FR-VQIs), be adjusted by learning the statistics of encrypted images. Thirdly, the view of NR-VQIs is the same view an attacker has, i.e., only the encrypted image. As such it is, theoretically, a very accurate way of estimating residual content and structure in the encrypted image, which is exactly the goal of evaluating an SE method.

In this paper we will ask, and answer, two questions:

1. *How do NR-VQIs fare in the evaluation methodology for visual security metrics, as outlined in [8]?*
2. *Can the specific statistics of encrypted images be learned to improve the NR-VQIs for the evaluation of SE methods?*

Section 2 gives an overview over the VQIs used in this paper. Section 3 recaps the evaluation methodology from [8] and in Section 4.1 the VQIs are evaluated based on this methodology. In Section 4.2 we will use one VQI to learn, with different fitness criteria, the statistics of encrypted images and gauge the improvement for security evaluation. Finally, Section 5 will recap the results and conclude the paper.

2 Overview Of Visual Quality Indices

In the following we will describe the visual quality indices/image metrics (VQI) which are used in this paper. The methods are primarily chosen based on available implementations, of those more recent were preferred. Further, VQIs based on different features or features from different domains, e.g., spatial versus DCT, were chosen over different VQIs using similar features.

BIQAA [9] is the blind image quality assessment based on anisotropy. The generalized Rènyi entropy is used to calculate local entropy histograms by associating a distribution for each pixel of the image. After normalization a windowed pseudo-Wigner distribution (PWD) can be approximated as a probability distribution function. This PWD is computed in a 1-D oriented window, allowing a measure of the entropy in a selected direction. Differences in the directional entropy are taken to measure image anisotropy and hence to estimate of the image quality.

BLIINDS-II [10] uses a Bayesian approach based on model parameters of generalized Gaussian distributions of various features. The features are derived from groups of AC coefficients of the local 2-D DCT transform of the image (on multiple scales). The AC coefficients are grouped into radial as well as circular clusters, orienting from the DC coefficient.

BRISQUE [11] is a quality evaluator which is based on learned regression, via scalable vector regression methods (SVR). The target values are the MOS values in the (LIVE) database and the features are based on localized luminance distribution and directional statistics. The features are calculated on two scales and are based on mean subtracted contrast normalized (MSNC) values. The MSNC values are taken directly as well as the difference values in horizontal, vertical and the two diagonal directions. An asymmetric generalized Gaussian distribution (AGGD) is fitted to each distribution and the values from the AGGD are used as features.

Global phase coherence (GPC),

Sharpness Index (sharp), and

Simplified sharpness index (SI) [12] are based on the observation that structural information in an image is encoded in the phase of its Fourier transformation. Impairment of the image affects the phase of the Fourier transform and changes phase coherence. Basically all three indices are based on the same model, they measure the (logarithmic) probability of the total variance, of the signal in phase space, diverging from the total variance of a 'random' image to denote structure (or quality). The notion of 'random' is where the three measures differ, the GPC uses a uniform random phase function as a model, sharp and SI use Gaussian white noise with the SI only using an approximation which changes it slightly but is much faster to calculate.

NIQE [13] uses directional MSNC values similar to BRISQUE. The main difference are a) that these values are taken from local patches which show a high local variance. And b) the model is a multivariate Gaussian based on the same statistics from regular images, which means unlike the BRISQUE no human observer information is taken into account.

SSEQ [14] uses the distribution, expressed by mean and skew, of spatial and spectral entropies as features. The entropies are calculated on a block by block partition of the image and on multiple scales. The features are trained using an SVR to conform to human judgement.

In addition to the NR-VQIs we also will present the two best full-reference VQIs (FR-VQI) from literature [8], the local edge gradient image metric (**LEG**) [15] and the visual information fidelity (**VIF**) [16]. The comparison will show how the NR-VQIs compare to the full-reference VQIs in a baseline configuration. Further, since the NR-VQIs are usually based on machine learning of statistical features, we will adapt one, the BRISQUE, and retrain for better performance on low-quality images. This will show the potential of using NR-VQIs when the target application of the VQI is known prior to training.

With respect to implementations, we used only image metrics for which the code is publicly available (for reproducible research). The three features GPC, SI, and sharp are available from the webpage of Lionel Moisan at the Université Paris Descaterse[1]. An implementation for BIQAA (version 1.0 by Gabarda was used) is available at MathWorks[2]. The remaining NR-VQIs and the VIF are available from the webpage of the Laboratory of Image and Video Encryption at the University of Texas (Austin)[3]. The LEG source code is available from the visual quality index implementation (VQI) of the University of Salzburg Wavelab group webpage[4].

[1] http://www.math-info.univ-paris5.fr/~moisan/sharpness/

[2] https://it.mathworks.com/matlabcentral/fileexchange/30800-blind-image-quality-assessment-through-anisotropy

[3] http://live.ece.utexas.edu/research/Quality/index.htm

[4] http://wavelab.at/sources/VQI

3 Evaluation Methodology

We will follow the proceedings described in [8], using the same methods and databases, however, we will briefly recap them here. Note that an illustrated guide to the evaluation method is available online [17].

The main goal of these evaluations is to find how well an visual quality index (VQI) reflects the human visual system (HVS) based on the mean opinion scores (MOS). The following assumes that a low MOS indicates low quality and that the VQIs are impairment indices, a low score implies good quality.

Application Domain is the domain to apply the VQI, either on the decoding of an encrypted image (*encrypted domain*) or on the decoding of an 'attacked', i.e., reducing/removing the impact of the encryption, encrypted image (*extracted domain*).

This is done by generating a set of image pairs, consisting of two encrypted images with clearly different quality. Then the VQI is tasked to order each image pair based on quality. This is done in both extracted and encrypted domain, the percentage of correct orderings is used as score, 0.5 (50%) being the worst outcome, akin to a coin flip per pair. In case of referenced VQIs the groundtruth of the image is used as a reference for both cases.

Correspondence to HVS: Monotonicity measures the correlation of the VQI to the HVS. Since the HVS, and most VQIs, are non-linear rank order correlation is used instead of linear correlation, specifically, Spearman's rank order coefficient (SROC) [18].

Since high correlation over the full quality range does not imply a high correlation over the low quality range [7], we will give the SROC over the full, high and low quality range.

Correspondence to HVS: Confidence gives a more in depth analysis of the signal shape and divergence of VQI and the HVS which is only coarsely captured by the SROC. The confidence, for a given MOS value D, measures the VQI values $v(i)$ over all images i where zero false negatives, $V_{min}(D)$ such that $\forall i : MOS(i) > D \implies v(i) > V_{min}(D)$, and zero positives, $V_{max}(D)$ such that $\forall i : v(i) > V_{max}(D) \implies MOS(i) > D$, occur. The confidence score $\mathcal{C}$ for a given MOS value D is $\mathcal{C}_D := |V_{max}(D) - V_{min}(D)|$. This function is condensed into the following values for easier representation: the average and standard deviation over $\mathcal{C}_D$, $\mu_{D \in S}(\mathcal{C}_D)$ and $\sigma_{D \in S}(\mathcal{C}_D)$.

Similar to SROC the confidence is not constant over the full quality range. Therefore, we also give a signal shape which describes the shape of the $\mathcal{C}_D$ function. Outliers based on the z score of a data point D are calculated as $z_D = {}^{\mathcal{C}_D - \mu(\mathcal{C}_D)}/_{\sigma(\mathcal{C}_D)}$. If $z_D < -1$ it is a *high outlier*, and if $z_D > 1$ it is a *low outlier*. Based on the distribution of high and low outliers we can specify the shape of the signal as follows. A signal is **stable** if there are no outliers, and it is **biased** if the high and low outliers are separable by a single threshold (D_t). Specifically, the shape is denoted **biased towards high quality** if $z_D < -1 \implies D < D_t$ or **biased towards low quality** if $z_D > 1 \implies D < D_t$. A signal which is neither stable nor biased is considered **unstable**.

4 Evaluation

In this section we will evaluate the VQIs based on the same data as was used in [8]. Specifically, the LIVE and IVC-SelectEncrypt databases.

The LIVE database, [19], does not contain encrypted images. However, the images in the low quality range exhibit strong distortions which can be equated to encrypted images in the sense that strong distortions mask a lot of the visual information. The test sets contained in the LIVE database (and their abbreviation in plots and figures) are JPEG 2000 compression (jp2k), JPEG compression (jpeg), white noise (wn), Gaussian blur (gblur), and bit errors in JPEG2000 bit stream transmission over a simulated fast fading Rayleigh Channel (fastfading), for detailed information see [20]. The threshold to separate low and high quality is set to a mean opinion score (MOS) value of 40 for our experiments. The MOS is the mean quality judgment based on a number of human observers opinions, the range of MOS depends on the database.

The IVC-SelectEncrypt database [21] contains various instances of JPEG 2000 transparent encryption, using different encryption techniques. The test sets contained in the IVC-SelectEncrypt database (and their abbreviation) are traditional encryption (trad), truncation of the code stream (trunc), window encryption without error concealment (iwind_nec), window encryption with error concealment (iwind_ec), and wavelet packet encryption (res), for detailed information see [22]. The low/high quality threshold is set to a MOS of 3 for the following experiments.

4.1 Evaluation of No-Reference VQI

For reasons of brevity all the results are summarized in Table 3. In tables the results are marked for good (bold) and bad (italics) performance. For the application domain results in $[0.4, 0.6]$ are considered bad and those in $[0, 0.1] \cup [0.9, 1]$ are considered good. For SROC values lower than 0.5 are considered bad and those higher then 0.9 are considered good results. For $\mu(\mathcal{C}_D)/\sigma(\mathcal{C}_D)$ 0.3/0.1 are good results and 0.5/0.2 are bad, with a difference of signal shapes over the two databases also being considered bad.

The final score is simply a sum of individual performances, starting at 0, bad (those in italics) performances distract 1, good (in bold) performances add 1. This allows for a simple ranking of the VQIs.

Application Domain The performance of the NR-VQIs in this test was extremely bad. It was expected that they perform badly in the encrypted domain, since errors introduced by the encryption have been shown to mislead VQIs. What is kind of interesting is the fact that the clearly different qualities in the extracted domain, see Figure 1, can not be sorted correctly. There are exceptions, BLIINDS-II does the sorting extremely well and SSEQ shows quite some improvement in the extraction domain. For the rest the improvements are slight (GPC, sharp, SI) to non-existing (NIQE, BIQAA, BRISQUE). That is not to say that the NR-VQIs can be applied in the encrypted domain, only that they (for the most part) also fail in the extraction domain. The most likely reason for this is that even the medium encryption still destroys much of the features the statistics are base on. However, the fact that BLIINDS-II does not make a single error in the extracted domain indicates that this is not true for all features.

Table 1: The Spearman rank order for the full- and low-quality range for the LIVE and IVC-SelectEncrypt databases.

LIVE		VQI	IVC-SelectEncrypt	
full	low		full	low
0.930	0.641	LEG	0.893	*0.492*
0.963	0.788	VIF	**0.914**	*0.285*
0.270	*0.098*	BIQAA	*0.139*	*0.100*
0.912	0.711	BLIINDS-II	0.501	0.690
0.950	0.765	BRISQUE	0.642	*0.304*
0.427	*0.080*	GPC	0.703	0.504
0.445	*0.086*	sharp	0.708	0.532
0.445	*0.086*	SI	0.708	0.532
0.907	0.585	NIQE	0.632	*0.479*
0.897	0.539	SSEQ	0.511	*0.285*

Figure 1: Residual information in the extracted domain of a (high, medium) and (medium, low) quality pair.

Correspondence to HVS: Monotonicity The SROC for the full and low-quality range is given in Table 1. The per test set SROC scores for the low-quality range is given in Table 3.

As pointed out in [7] the SROC over the full-quality range does not give any indication about the performance in the low-quality range. Compare the BLIINDS-II and the VIF on the IVC-SelectEncrypt database, the VIF low-quality performance is far worse than for the full-quality range, for the BLIINDS-II the performance over the low-quality range is higher than for the full-quality case. It should also be noted that, apart from the BLIINDS-II on the IVC-SelectEncrypt database, all VQIs perform worse on the low-quality range.

Another noteworthy tendency is the performance over regular impairments as given on the LIVE database versus the performance on encrypted images. All VQIs apart from those based on phase coherence perform worse for the encryption test sets. For most NR-VQIs this could be explained with their respective training with human judgement of regular image impairment (as given in the LIVE database). However, NIQE also does not utilize human judgement in the training phase and still performs worse on the IVC-SelectEncrypt database. The fact that phase coherence encodes structure and encryption is usually aimed at destroying structure would be a likely explanation. However, given that the phase coherence based

Table 2: Per test set comparison of SROC for the low-quality range as well as overall SROC for the full- and low-quality range on the IVC-SelectEncrypt database.

SROC on	BRISQUE	BRISQUE cross	BRISQUE low
iwind ec	0.598	0.485	0.408
iwind nec	0.143	0.709	0.676
resolution	0.107	0.321	0.393
trad	0.560	0.437	0.723
truncation	0.885	0.657	0.750
full-qualtiy	0.642	0.767	0.745
low-qualtiy	0.304	0.364	0.636

VQIs behave very differently on the iwind test set with and without error concealment, see Table 3, invalidates this argument. The error concealment should only reduce noise in the image and is not able to reconstruct structural information. As such it is as of yet unclear what the reason for this behaviour is, however, it should be noted that the phase coherence based VQIs show some desirable traits for visual security VQIs.

Correspondence to HVS: Confidence The confidence of the NR-VQIs overall is very bad, owing to a high average spread, the only exception is the NIQE, which is noteworthy by itself. Recall, that the confidence score gives the range of zero false negatives and zero false positives in relation to human judgement. The NIQE is the best of the NR-VQIs in this regard, even though it does not utilize information about human judgement during training.

When compared to full reference VQIs the NR-VQIs exhibit a very bad performance, except for the NIQE. Interestingly, the NR-VQIs, with the exception of those based on phase coherence, maintain their signal shape; this is something the FR-VQIs frequently fail at.

4.2 Learning SE statistics

Since a lot of the NR-VQIs are based on machine learning, to better conform to human judgement, an obvious question is how much the performance improves when the learning is done on a database with encrypted distortions. For this we will use the BRISQUE, which uses the LIBSVM, and learn it on the IVC-SelectEncrypt database. In order to fairly evaluate it we will use cross validation, where we iteratively evaluate a test set which is disjoint from the training set. The folds are constructed by taking an original image and all its distorted versions for evaluation and all other images for training.

The parameters for the scalable vector regression (SVR) were searched on a grid in logarithmic space with different fitness functions. The basic version *BRISQUE cross* uses the full-quality SROC and the *BRISQUE low* uses the SROC over the low-quality range as fitness function.

The results, given in Table 2, expose an interesting behaviour. The improvement is a tradeoff in which test sets with horrible performances (iwind nec and resolution) are greatly improved at the cost of a performance reduction on relatively well performing test sets (truncation). This basically shifts all test sets towards a rather mediocre score. This is true for both target fitness functions. As such the training clearly improves the performance by a lot, compare e.g. the full- and low-quality results of BRISQUE low with the pre-trained BRISQUE, but at the cost of individual strength. As such it is recommended to train as specifically or as generally, i.e., on a large database with different distortions, as possible.

In conclusion the overall performance of LEG and VIF is still better. However, if the application scenario is known the learning approach can yield a metric which is better than LEG and VIF which are, by nature, fixed in their performance.

5 Conclusion

We initially asked two questions, and the answer can be summarized as follows.

How do NR-VQIs fare in the evaluation methodology for visual security metrics, as outlined in [8]? The no-reference VQIs behave overall very similar to most FR-VQIs. Some are more fit as VQIs for visual security (NIQE), some less (GPC, sharp, SI, BIQAA). But more importantly VIF and LEG also outperform the NR-VQIs, so the recommendation is to use VIF or LEG, if time is a constraint, still stands.

Can the specific statistics of encrypted images be learned to improve the NR-VQIs for the evaluation of SE methods? We used the BRISQUE, which uses LIBSVM with a SVR kernel, to try and learn the specific distortions as introduced the IVC-SelectEncrypt database. It is clear that the learning improves the overall performance quite a lot. Specifically, there are test sets where the learned version of the BRISQUE outperforms LEG and VIF. So if the specific application is known and training data is available the trained NR-VQI can be a better choice. As a general purpose VQI for security metrics the VIF and LEG still are a better choice.

Future Work

Certain metrics show desirable traits: good confidence (NIQE), good overall and low quality performance on encrypted images (phase coherence based) or good performance for specific (hard) test sets like SSEQ on the iwind ec test set. It would be interesting to see if the specific traits of single metrics can be transfered and combined by using a bag of features approach.

Acknowledgements

This work was partially supported by the Austrian Science Fund, project no. P27776.

Table 3: Summary of Evaluation

Good performances are marked in **bold**, bad ones in *italic*. Application Domain: results in $[0.4, 0.6]$ are bad / results in $[0, 0.1] \cup [0.9, 1]$ are good. SROC: values lower than 0.5 are bad / those higher then 0.9 are good. Confidence: $\mu(\mathcal{C}_D) < 0.3$ and $\sigma(\mathcal{C}_D) < 0.1$ are good / and $\mu(\mathcal{C}_D) > 0.5$ $\sigma(\mathcal{C}_D) > 0.2$ are bad. A difference of signal shapes over the two databases is also bad.

	LEG	VIF	BIQAA	BLIINDS-II	BRISQUE	GPC	sharp	SI	NIQE	SSEQ
Application Domain										
Encryption	0.334	*0.454*	*0.555*	*0.512*	*0.504*	*0.496*	*0.452*	*0.453*	*0.473*	*0.457*
Extraction	**0.994**	**0.988**	*0.535*	**0.000**	*0.503*	*0.435*	*0.432*	*0.428*	*0.537*	0.754
Confidence on the LIVE database										
$\mu(\mathcal{C}_D)$	**0.291**	**0.285**	*0.934*	*0.504*	0.387	*0.724*	*0.855*	*0.857*	**0.174**	*0.564*
$\sigma(\mathcal{C}_D)$	**0.070**	0.110	**0.089**	0.122	0.137	*0.202*	0.101	0.100	0.168	0.125
Signal Shape	*Bias Low*	*Bias Low*	Stable	Bias High	Bias High	*Bias Low*	*Bias Low*	*Bias Low*	Bias Low	Bias High
Confidence on the IVC-SelectEncrypt database										
$\mu(\mathcal{C}_D)$	**0.268**	**0.277**	0.322	*0.569*	*0.597*	*0.668*	*0.699*	*0.699*	0.415	*0.554*
$\sigma(\mathcal{C}_D)$	**0.077**	**0.098**	*0.239*	*0.201*	0.190	*0.306*	*0.333*	*0.334*	0.166	0.130
Signal Shape	*Bias High*	*Bias High*	Stable	Bias High	Bias High	*Stable*	*Stable*	*Stable*	Bias Low	Bias High
Low Quality SROC on the LIVE database										
fastfading	0.893	**0.937**	0.527	*0.496*	*0.465*	0.511	0.546	0.550	0.699	0.528
gblur	0.872	**0.920**	0.720	0.681	0.873	0.845	0.819	0.821	0.872	0.845
jp2k	0.617	0.646	*0.122*	0.599	*0.367*	0.534	*0.436*	*0.439*	*0.469*	0.616
jpeg	0.699	0.829	*0.287*	0.715	0.819	*0.380*	*0.368*	*0.368*	0.768	*0.196*
wn	0.804	**0.911**	0.554	0.826	**0.957**	0.725	0.874	0.873	**0.908**	0.817
Low Quality SROC on the IVC-SelectEncrypt database										
iwind ec	*0.141*	0.518	*0.194*	0.584	0.598	*0.098*	*0.001*	*0.001*	*0.012*	0.699
iwind nec	0.823	0.732	0.687	0.717	*0.143*	0.615	0.516	0.516	0.676	0.648
resolution	*0.490*	0.823	*0.393*	*0.286*	*0.107*	0.571	*0.429*	0.607	*0.036*	*0.107*
trad	0.652	**0.913**	0.805	0.805	0.560	0.676	0.876	**0.907**	0.764	*0.437*
truncation	*0.181*	0.832	*0.407*	0.685	0.885	0.868	0.868	0.868	0.797	0.558
Comparison Score, -1 or +1 for insufficient or good performance, -1 for conflict in signal shape										
	LEG	VIF	BIQAA	BLIINDS-II	BRISQUE	GPC	sharp	SI	NIQE	SSEQ
Score	1	**6**	-8	-5	-6	-9	-10	-8	-3	-6

References

[1] T. E. Boult, "Pico: Privacy through invertible cryptographic obscuration," in *IEEE/NFS Workshop on Computer Vision for Interactive and Intelligent Environments*, Lexington, KY, USA, Nov. 2005, pp. 27–38 (cit. on p. 2).

[2] C. Bergeron and C. Lamy-Bergor, "Compliant selective encryption for H.264/AVC video streams," in *Proceedings of the IEEE Workshop on Multimedia Signal Processing, MMSP'05*, Oct. 2005, pp. 1–4. DOI: 10.1109/MMSP.2005.248641 (cit. on p. 2).

[3] F. Dufaux and T. Ebrahimi, "Scrambling for privacy protection in video surveillance systems," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 18, no. 8, pp. 1168–1174, 2008. DOI: 10.1109/TCSVT.2008.928225 (cit. on p. 2).

[4] A. Unterweger, K. Van Ryckegem, D. Engel, and A. Uhl, "Building a post-compression region-of-interest encryption framework for existing video surveillance systems – challenges, obstacles and practical concerns," *Multimedia Systems*, vol. 22, no. 5, pp. 617–639, 2015. DOI: 10.1007/s00530-015-0473-6 (cit. on p. 2).

[5] T. D. Lookabaugh and D. C. Sicker, "Selective encryption for consumer applications," *IEEE Communications Magazine*, vol. 42, no. 5, pp. 124–129, 2004 (cit. on p. 2).

[6] C. E. E. Shannon, "Communication theory of secrecy systems," *Bell System Technical Journal*, vol. 28, pp. 656–715, Oct. 1949 (cit. on p. 2).

[7] H. Hofbauer and A. Uhl, "Visual quality indices and low quality images," in *IEEE 2nd European Workshop on Visual Information Processing*, Paris, France, Jul. 2010, pp. 171–176 (cit. on pp. 2–4).

[8] ——, "Identifying deficits of visual security metrics for images," *Signal Processing: Image Communication*, vol. 46, pp. 60–75, 2016, ISSN: 0923-5965. DOI: 10.1016/j.image.2016.05.001 (cit. on pp. 2–5).

[9] S. Gabarda and G. Cristóbal, "Blind image quality assessment through anisotropy," *Journal of the Optical Society of America A*, vol. 24, no. 12, B42–B51, 2007. DOI: 10.1364/JOSAA.24.000B42 (cit. on p. 2).

[10] M. A. Saad, A. C. Bovik, and C. Charrier, "Blind image quality assessment: A natural scene statistics approach in the DCT domain," *Image Processing, IEEE Transactions on*, vol. 21, no. 8, pp. 3339–3352, 2012 (cit. on p. 2).

[11] A. Mittal, A. K. Moorthy, and A. C. Bovik, "Blind/referenceless image spatial quality evaluator," in *In Asilomar Conference on Signals, Systems and Computers*, IEEE, 2011, pp. 723–727 (cit. on p. 2).

[12] A. Leclaire and L. Moisan, "No-reference image quality assessment and blind deblurring with sharpness metrics exploiting fourier phase information," *Journal of Mathematical Imaging and Vision*, vol. 52, no. 1, pp. 145–172, 2015, ISSN: 1573-7683. DOI: 10.1007/s10851-015-0560-5 (cit. on p. 3).

[13] A. Mittal, R. Soundararajan, and A. C. Bovik, "Making a "completely blind" image quality analyzer," *Signal Processing Letters, IEEE*, vol. 20, no. 3, pp. 209–212, 2013 (cit. on p. 3).

[14] L. Liu, B. Liu, H. Huang, and A. C. Bovik, "No-reference image quality assessment based on spatial and spectral entropies," *Signal Processing: Image Communications*, vol. 29, no. 8, pp. 856–863, 2014. DOI: https://doi.org/10.1016/j.image.2014.06.006 (cit. on p. 3).

[15] H. Hofbauer and A. Uhl, "An effective and efficient visual quality index based on local edge gradients," in *IEEE 3rd European Workshop on Visual Information Processing*, Paris, France, Jul. 2011, 6pp. (Cit. on p. 3).

[16] H. R. Sheikh and A. C. Bovik, "Image information and visual quality," *IEEE Transactions on Image Processing*, vol. 15, no. 2, pp. 430–444, May 2006 (cit. on p. 3).

[17] H. Hofbauer and A. Uhl, *How to calculate the spearman rank order correlation, confidence and signal shape*, Online at http://wavelab.at/papers_supplement/Hofbauer16c_supplement.pdf, 2017 (cit. on p. 3).

[18] C. Spearman, "The proof and measurement of association between two things," *The American Journal of Psychology*, vol. 100, no. 3/4, pp. 441–471, 1904 (cit. on p. 3).

[19] H. R. Sheikh, Z. Wang, L. Cormack, and A. C. Bovik, *LIVE image quality assessment database release 2*, http://live.ece.utexas.edu/research/quality (cit. on p. 4).

[20] H. R. Sheikh, M. F. Sabir, and A. C. Bovik, "A statistical evaluation of recent full reference image quality assessment algorithms," *IEEE Transactions on Image Processing*, vol. 15, no. 11, pp. 3440–3451, Nov. 2006 (cit. on p. 4).

[21] F. Autrusseau, T. Stütz, and V. Pankajakshan, *Subjective quality assessment of selective encryption techniques*, http://www.irccyn.ec-nantes.fr/~autrusse/Databases/, 2010 (cit. on p. 4).

[22] T. Stütz, V. Pankajakshan, F. Autrusseau, A. Uhl, and H. Hofbauer, "Subjective and objective quality assessment of transparently encrypted JPEG2000 images," in *Proceedings of the ACM Multimedia and Security Workshop (MMSEC '10)*, Rome, Italy: ACM, Sep. 2010, pp. 247–252 (cit. on p. 4).