Journal of ELECTRICAL ENGINEERING, VOL 73(2022), NO2, 73–87

sciendo

PAPERS

Study of subjective and objective quality assessment of infrared compressed images

Omar Zelmati[1], Boban Bondžulić[1], Boban Pavlović[1], Ivan Tot[1], Saad Merrouche[2]

Given the lack of accessible infrared compressed images' benchmarks annotated by human subjects, this work presents a new database with the aim of studying both subjective and objective image quality assessment (IQA) on compressed long wavelength infrared (LWIR) images. The database contains 20 reference (pristine) images and 200 distorted (degraded) images obtained by application of the most known compression algorithms used in multimedia and communication fields, namely: JPEG and JPEG-2000. Each compressed image is evaluated by 31 subjects having different levels of experience in LWIR images. Mean opinion scores (MOS) and natural scene statistics (NSS) of pristine and compressed images are elaborated to study the performance of the database. Five analyses are conducted on collected images and subjective scores, namely: analysis by compression type, analysis by file size, analysis by reference image, analysis by quality level and analysis by subject. Moreover, a wide set of objective IQA metrics is applied on the images and the obtained scores are compared with the collected subjective scores. Results show that objective IQA measures correlate with human subjective results with a degree of agreement up to 95 %, so this benchmark is promising to improve existing and develop new IQA measures for compressed LWIR images. Thanks to a real-world surveillance original images based on which we analyze how image compression and quality level affect the quality of compressed images, this database is primarily suitable for (military and civilian) surveillance applications. The database is accessible via the link: https://github.com/azedomar/compressed-LWIR-images-IQA-database. As a follow-up to this work, an extension of the database is underway to study other types of distortion in addition to compression.

K e y w o r d s: image compression, image quality assessment, infrared images, JPEG, JPEG-2000, LWIR

1 Introduction

Over the last decade, as an important field of research, the image quality assessment (IQA) has been attracting the attention of several researches [1]. The crucial role multimedia has been gaining in humans' life, has strongly prompted the need of assessing the quality of provided services, particularly for image and video data. As answer to this need, many works have been dedicated to provide researchers benchmarks and databases annotated by human subjects in order to help understanding human perception of the image quality.

The standard imaging devices used worldwide to capture images are mainly visual cameras. However, these cameras could not be used or they generate poor quality images in cases of complete darkness or low lighting scenes. It is also impossible for such cameras to capture objects throughout opaque materials. To overcome these limitations, non-illumination based imagers have been used. These devices utilize infrared (IR) radiation to generate images [2].

Thermal imaging cameras detect objects and background radiation from the IR spectral range in so-called atmospheric transmission windows, known as: near infrared (NIR, 0.7-1 μm), short wavelength infrared (SWIR, 1-2 μm), medium wavelength infrared (MWIR, 3-5 μm) and long wavelength infrared (LWIR, 7.5-14 μm). This work will focus on the LWIR range since most thermal imaging cameras operate in this atmospheric window [3].

LWIR images can be an asset in executing several assignments in many areas and applications such as firefighting [2], medicine [4], industry [5], security surveillance and the military [6, 7], which makes analyzing the images captured by this type of camera the focus of the present work.

Infrared imaging is used in many defense applications to enable high-resolution vision and identification in near and total darkness. Thermal cameras are nowadays combined with visible cameras (security cameras or embedded on aerial vehicles) to improve night vision for surveillance and to enhance automatic threats detection over wide areas [6]. In infrared imagery, night vision electronic sensors are also deployed by military troops; the objective is to provide visibility (day or night) in the presence of smoke, fog, dust or any other obscuring airborne particles. This can significantly facilitate decision making, monitoring and surveillance, and improve soldier performance [7].

The captured LWIR images are stored and/or transmitted through different media channels, making them

[1] University of Defence in Belgrade, Military Academy, Department of Telecommunications and Informatics, Serbia, zelmatiomar1991@gmail.com, bondzulici@yahoo.com, bobanpav@yahoo.com, totivan@gmail.com, [2] Ecole Militaire Polytechnique Chahid Abderrahmane Taleb, Algiers, Algeria, saaderic91@gmail.com

https://doi.org/10.2478/jee-2022-0011, Print (till 2015) ISSN 1335-3632, On-line ISSN 1339-309X
© This is an open access article licensed under the Creative Commons Attribution-NonCommercial-NoDerivs License (http: //creativecommons.org/licenses/by-nc-nd/3.0/).

prone to different types of signal distortions caused by the transmission media. Other artifacts may occur due to the imagers themselves or scene conditions. However, the most common degradation affecting image quality is caused by compression. The latter is utilized to remedy the problem of resource limitation such as weak transmission bandwidth and limited memory storage of devices which are usually used by end users. Transmitting or storing the data flow produced by IR imagers can be quite challenging. Truthfully, LWIR images are compressed using algorithms such as JPEG, JPEG-2000 and H.265 which are the most known image compression techniques. Therefore, assessing the quality of the compressed images, either by visual inspection (subjective) or by computer vision automation (objective), is an essential measure to be taken before further exploitation. In this regard, applying IQA techniques comes as a natural choice to fulfill the requirements related to the desired image quality.

In an effort towards predicting the quality of captured images as seen by humans, several IQA models have been developed to provide a numerical measurement of the distorted image quality. In cases where an original version of the degraded image is used to elaborate a comparison, the approach is called full-reference (FR-IQA). Conversely, no-reference (NR-IQA) techniques do not require access to the original non-distorted image, assuming that the reference image is unavailable. It is worthwhile to mention a third technique known as reduced-reference (RR-IQA), which only uses partial information of the reference image in order to measure the quality of its degraded version [8]. Considering the availability of the original version of compressed images, the approaches studied in this work belong to FR-IQA metrics family.

The accessible databases studying IQA of compressed LWIR images are very limited. Thereby, the research presented in this work tends to overcome this issue by providing a new publicly available database. The main objective of the work is to provide a benchmark of evaluated LWIR images. The obtained images are assessed by 31 subjects who are all military students and officers at the University of Defence in Belgrade. Subjects are chosen randomly so they have different levels of experience with infrared images. The evaluations have been made using a software developed by authors specifically to follow the International Telecommunication Union (ITU) recommendations released in 10/2019 [9]. The obtained scores have been studied and different analyses have been carried out in the aim of assessing the quality of the benchmark.

2 LWIR images and compression

LWIR imagers have many advantages over visual spectrum imaging devices, such as their ability to create images under poor lighting conditions or even in complete darkness. However, they are prone to several problems, particularly caused by IR detectors, as well as the influence of the captured objects themselves.

The visual grayscale or color camera uses lens to create images by taking all the light rays reflected by the objects in the scene and focusing them on the image sensor, resulting in a visual image, whereas, IR camera captures infrared radiation emitted by objects whose temperature is above absolute zero, and converts thermal radiation into visual images perceived by human eyes. Thus, artifacts like non-uniformity (NU) noise, halo effect and hotspot, which may occur on LWIR images, have been investigated in order to reduce their presence or to assess their effect on the image quality [10]. However, the main interest of this work is studying the effect of JPEG and JPEG-2000 compression without taking into consideration the above-mentioned artifacts.

Although thermal sensors have a lower resolution compared to visible camera, resorting to image compression may be still a necessary step in transmitting IR images and videos. This results in an efficient usage of the available bandwidth and helps enhance the communication system performance. In particular, LWIR images are transmitted from unmanned aerial vehicles to the control station via a narrowband wireless channel. In this case, image compression might be considered as a solution to the bandwidth availability problem [11].

Moreover, LWIR images use a different bit-depth than the one usually used for visible images. In fact, most thermal sensors nowadays offer a bit-depth of 16 bits, which is different from the classic 8 bits depth used for visible images. Exploiting the LWIR images in this format might be a complex task to achieve, and envisaging a compression algorithm would require a significant amount of computational resources. Thus, different techniques are utilized to solve this issue, namely tone mapping [12] and compression with a higher bit-depth [13].

As with visual spectrum images, compression of LWIR images highly depends on use case considerations. If images are transmitted in real time, as in video streaming, the compression algorithm should meet real-time performance, and buffering is required to match the output generated at the compression encoder with the transmission rate of the communication channel. Performing compression in real-time is not crucial for storage and post-processing purposes.

Image compression techniques can be divided into lossless and lossy categories [14]. In lossless techniques, compressed image is an exact replica of the original version. Lossless image compression is widely applied in several areas including: satellite imagery exploitation, forest fires monitoring, and medical imaging, where information loss may lead to an erroneous decision.

Unlike the lossless approaches, the compression ratio (CR) is significantly higher for lossy techniques [15]. Thereby, they are commonly used in image compression in the aim of increasing the CR to the detriment of the reconstructed image precision. Many compression applications tolerate loss of information taking advantage of human eye characteristics that allow compensation for some deformations of the resulted images. Lossy image

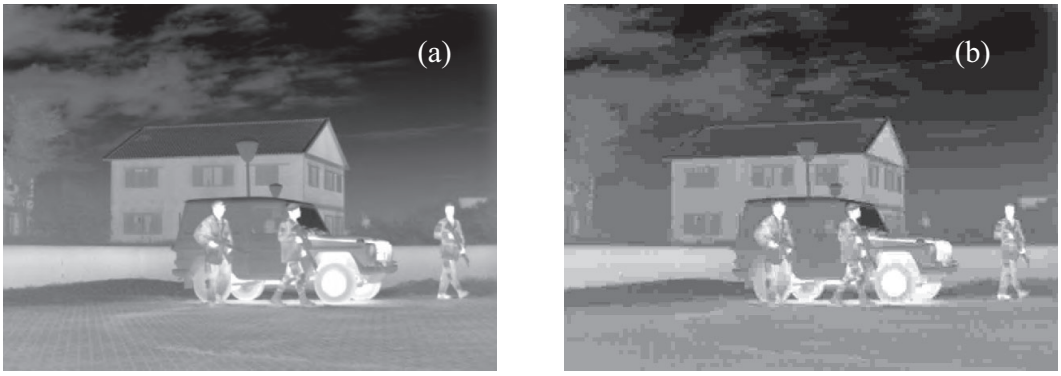

Fig. 1. Comparison of images: (a) – JPEG-2000 compressed, and (b) – JPEG compressed

compression is widely used in image transmission through the internet, in digital images storage and media imaging [15].

Many lossy compression techniques have been introduced. However, the most known algorithms are JPEG and JPEG-2000. In fact, JPEG is not just one compression algorithm, but it is a standard including different algorithms and defined for several modes (baseline, lossless, progressive and hierarchical) [16]. The most popular JPEG mode is the baseline mode. It only supports lossy encoding and it is based on discrete cosine transform (DCT) and Huffman coding. The progressive and hierarchical modes of JPEG are both lossy and differ from baseline mode only in the way DCT coefficients are computed and coded. Meanwhile, the lossless mode is based on a completely different algorithm with a predictive scheme. The prediction error is entropy encoded with the Huffman algorithm [17].

According to Brislawn and Quirk [18], JPEG-2000 is an improvement of the JPEG standard. JPEG-2000 relies on providing enhanced compression performance and new functionalities by removing the subdivision step. As result, it eliminates the blocking artifact that characterizes DCT-based approximations at high compression ratio JPEG images [19, 20]. In Fig. 1, despite that JPEG image has a larger file size (9.2 kB), the JPEG-2000 image seems to have better image quality with much smaller file size (4.6 kB).

3 Image quality assessment

One of the major issues faced by lossy compression algorithms lies in assessing the quality of the resulted images [21]. Conventionally, the image quality is assessed using distortion. The later is defined by the difference between the reference and the compressed image. Therefore, smaller distortion corresponds to a better compression algorithm. In other words, the distortion quantifies the fidelity or quality of the compressed image compared to its original version [19]. The study of this fidelity leads to defining two types of IQA: objective IQA (OIQA) and subjective IQA (SIQA).

In practice, OIQA techniques rely on mathematical models that provide an error estimate of the compressed image compared to its original [1]. Meanwhile, SIQA approaches are introduced to collect human perception of the image quality, by creating databases that contain information about observers' judgment. Developing and exploiting such a database is the main contribution of the present work.

In order to determine the quality of the compressed (degraded) image **d** to its original (reference) **r**, some practical measures of distortion can be calculated. One of the most popular methods is the mean squared error (MSE) which measures the difference between **r** and **d** assuming that the images are two dimensional functions of $M \times N$ size. The MSE is

$$e = \frac{1}{M \times N} \sum_{j=1}^{N} \sum_{i=1}^{M} [r(i,j) - d(i,j)]^2 \,, \tag{1}$$

where i and j are respectively horizontal and vertical positions of a given pixel.

This measurement provides an average value of the energy loss in the compression of the original image **r**. Therefore, an observer will generally evaluate compressed images with lower MSE to be closer to the original image (for same type of compression). However, when comparing between two images compressed by different types of compression, the one with the smaller MSE is not necessarily the one that is perceptually closest or most similar to the original image.

Related peak signal-to-noise ratio (PSNR) is a simple squared error measure, measured in dB where interest is about the size of the error relative to the peak value of the image. For 8 bits per pixel images, PSNR is defined as

$$p = 10 \log_{10} \frac{255^2}{e}, \tag{2}$$

and it can be concluded that higher PSNR values indicate better image quality.

MSE and PSNR are appealing because of their simplicity and easy implementation, but, they do not match well with the human visual system (HVS) perception [22].

Thus, many alternative metrics have been introduced to overcome this limitation. A set of the most striking techniques will be discussed and used in this work.

Based on the known characteristics of HVS, particularly, its ability to extract structural information, an OIQA model based on a measure of structural similarity (SSIM) was proposed in [23]. This full-reference (FR) model takes advantage of the fact that the natural scenes exhibit peculiar structures with a pronounced inter-pixel dependency. Those structures hold essential information about the elements present in the natural scene. Therefore, measuring changes of those structures can offer a reliable assessment of the distorted image quality. The developed model compares between the original and the distorted version in terms of luminance, contrast and structure. Based on the three comparisons, the model is able to provide a quality map which accounts only for structural distortion. A simple computational model is used in pooling process. The later can then provide an overall score which is basically a similarity measurement varying from -1 to 1 from the worst to the best quality of the test image.

In practice, the subjective evaluation of an image might be significantly affected by the sampling density of the image signal and the distance from the image plane to the observer. To overcome this issue, a modified SSIM version known as multi-scale SSIM (MS-SSIM) was proposed in [24]. Practically, the proposed metric takes the reference and distorted images as input, then iteratively applies a low-pass filter and downsamples the filtered image by a factor of two. The contrast comparison and the structure comparison are calculated for each scale. Luminance comparison is applied just for the last scale. Thereby, the overall MS-SSIM score is obtained by combining evaluations from all the scales in the pooling phase.

In [25], authors treated the IQA pooling problem from an information theoretic point of view. Assuming that the components of the image that contain more information content would attract more visual attention, the information content weighted structural similarity (IW SSIM) measure is then developed based on a local information weighting model deployed in the pooling process. The proposed weighting model relies on multi-scale Laplacian pyramid transform (LPT) applied on both original and distorted images. Accordingly, for each scale, an information content weight (ICW) map is calculated by applying a 3×3 sliding window on the resulted LPT coefficients. Therefore, IW SSIM is defined as the combination of the ICWs and the MS-SSIM measure. Furthermore, in the same paper, authors defined the information content weighted MSE (IW MSE) and PSNR (IW PSNR).

The application of IQA in practice requires algorithms with low computational costs and high correlation with subjective evaluations. Thereby, authors in [26] proposed an IQA approach having a high prediction performance and a low computational cost. Given the fact that perceived quality is strongly correlated to the visual saliency, the novel algorithm, namely the spectral residual based similarity (SR SIM), is based on an effective and efficient visual saliency model (SRVS) [27]. Practically, SR SIM uses the SRVS map as a feature map that characterizes the image local quality, and as a weighting function that indicates the local region importance for the HVS.

A recent approach elaborating the visual saliency was proposed in [28]. The contrast and visual saliency similarity-induced index (CVSSI) is a FR-IQA based on the summation of a deviation-based pooling strategy for a local contrast similarity map and a global visual saliency similarity map.

Instead of treating the distortions indiscriminately, authors in [29] proposed to assess the quality of an image within two aspects, namely: detail losses and additive impairments. Detail losses are commonly a direct result of the most distortion types. Thereby, detail losses metric (DLM) is defined to assess the amount of losses. Meanwhile, additive impairments refer to the redundant visual information that appears in the distorted image but does not exist in the reference one. A wavelet domain decoupling algorithm was developed to provide the additive impairments metric (AIM). The overall quality measure ADM is yielded by adaptively combining DLM and AIM.

An IQA metric based on the edge/gradient similarity (GSM) was proposed in [30]. GSM proved that the gradient information provides more emphasis on distortions around the edge regions. Therefore, the luminance and contrast-structure comparisons were conducted by introducing the gradient information.

Inspired by gradient, authors in [31], proposed to use gradient-based local quality map for overall image quality measurement. The proposed metric, namely gradient magnitude similarity deviation (GMSD) is based on calculating the horizontal and vertical gradients of both original and distorted images by applying a 3×3 Sobel filter. The final quality value of an image is obtained via a pooling phase where the GMSD score represents the standard deviation of the local quality map.

The multi-scale contrast similarity deviation (MCSD) was proposed in [32], where authors utilized the contrast feature alone to design their metric. At each scale, the contrast similarity deviation (CSD) for both original and distorted images is calculated. Accordingly, the final quality score is obtained from pooling the CSDs from each scale.

The physiological and psychophysical evidence points towards the existence of visually discernable features that correspond to the points where the Fourier components have congruent phases. This fact motivated a low-level feature similarity metric, named FSIM [33]. Given the fact that phase congruency is a contrast invariant measure, the image gradient magnitude was introduced to account for the contrast variations.

Inspired by the FSIM model, the authors in [34] proposed the Haar wavelet-based perceptual similarity index (HaarPSI). The amplitude of high-frequency Haar transform coefficients are used to determine local similarities and the low-frequency coefficients to weigh the relevance

of similarities/dissimilarities at various locations in the image domain.

Based on the belief that visual information in an image is often redundant and the HVS tends to perceive an image on its low-level features at key locations, authors in [35] proposed the Riesz transform (RT) based feature similarity (RFSIM) index. RFSIM algorithm compares between Riesz transform features of both original and distorted image at key regions (regions of interest) in three stages. Unlike the RFSIM where RT is applied directly on both original and distorted images to determine the features and then estimate the quality, Riesz transform and visual contrast sensitivity-based feature similarity index (RVSIM) was proposed by combining RT with visual contrast sensitivity to provide the quality score [36].

The authors in [37] proposed an FR-IQA approach, center-emphasized quality index (CEQI), using visual saliency and contrast similarity. The idea behind this approach originates from the psychological vision research which states that the human vision is biased to the center area of an image, and any distortion in this area will be considerably perceived and compared to distortions which may affect other areas of the image.

The behavior of the HVS conducted the IQA proposed in [38]. The authors named their method the most apparent distortion (MAD). The HVS was supposed to behave differently based on the amount of distortion in the image. Thereby, two qualities were proposed, namely: detection-based quality and appearance-based quality. Firstly, a spatial-domain model is employed in order to take into consideration the contrast sensitivity function, luminance and contrast masking with distortion-type-specific adjustments. The MSE is then calculated between the original and distorted images in order to provide the detection-based quality. Finally, the overall quality of the degraded image is computed by taking a weighted geometric mean of the detection- and appearance-based scores.

Full-reference IQA models usually assume access to high quality pristine image. However, this assumption cannot be always guaranteed. Calculating a quality based on this kind of source image might produce an inaccurate score. Thereby, assessing the quality of the compressed image has been achieved using a two-step quality assessment approach (2stepQA) [39]. First, a no-reference (NR) model is applied to the source image by computing its perceptual distance from the space of high-quality natural images to assess the degree of possible distortion in this image. Then an FR model is applied to compare between the source and the compressed image. An overall score is obtained as the product of the two calculated scores. When the source image is of high quality, the obtained NR score will be sufficiently high, and thus the quality of the compressed image might be accurately assessed by the FR model. In the opposite case, if the source image is distorted, the NR score will serve as a basis to correct the obtained FR score.

The visual information fidelity (VIF) index [40] views the IQA problem as an information fidelity problem, and the images are modeled using Gaussian scale mixtures to measure the amount of image information. The HVS is modeled as a channel with a single additive noise component that adds uncertainty to the image signal that flows through the HVS. Therefore, a distorted image is considered as an original image passed through the distortion channel. A local VIF score is calculated from both original and distorted images' signals processed by the HVS channel. At last, the final VIF score is a normalized weighing sum of those local VIF scores.

Authors in [41] were inspired by researches results on brain theory and neuroscience indicating that the human brain works with an internal generative mechanism (IGM) for visual information perception and understanding. The proposed approach adopts a Bayesian prediction model and the input image is decomposed into predicted and disorderly portions. An adaptive nonlinear procedure is applied to combine the results on the two portions and then estimate the final quality score.

In the sparse feature fidelity (SFF) metric [42], authors trained a feature detector on a set of natural images using an independent component analysis algorithm. The test image is divided into image patches and the mean value of each patch is calculated. Therefore, all the patch vectors with zero mean are used for the computation of feature similarity, while all the mean values serve to compute luminance correlation. The overall SFF score is obtained by combining the feature similarity and the luminance correlation.

In a recent research, a novel deep learning based IQA metric was proposed to unify structure and texture similarity [43]. The deep image structure and texture similarity (DISTS) approach is based on an innovative pre-trained object-recognition model named VGG. A parametric texture model was established by computing the global means of convolution responses at each stage.

According to [44], objective metrics are not necessarily correlated to human perception, but they are often used due to their independency of the observer and assessment conditions. Consequently, SIQA techniques are much better for image evaluation especially in cases where the goal is to reach high quality images as seen by end-users. Thus, in this work, a database of compressed LWIR JPEG and JPEG-2000 images was evaluated by human subjects in order to study the agreement between subjective and objective scores.

4 Database description

4.1 Sources of LWIR images

The new LWIR image database is formed using JPEG (JPG) and JPEG-2000 (J2K) compressed images issued from the database presented in [45]. This database was created by collecting images from four different databases: MORRIS [46], TRICLOBS [47], Military academy 1 and Military academy 2 [48]. A different thermal sensor is used in each database, and characteristics of those sensors are

Table 1. Characteristics of LWIR sensors

	MORRIS	TRICLOBS	Military academy 1	Military academy 2
Infrared sensor	Miricle 110KS	XenICs Gobi 384	FLIR SC620	ATIS
Type	uncooled micro-bolometer	uncooled micro-bolometer	uncooled micro-bolometer	cooled Stirling machine
Spectral range	7–14 μm	8–14 μm	7.5–13 μm	8–12 μm
Resolution	384 × 288	384 × 288	640 × 480	320 × 240
Pixel size	35 μm	25 μm	N/A	N/A
NETD	50 mK @ 30℃	50 mK @ 30℃	40 mK @30℃	N/A

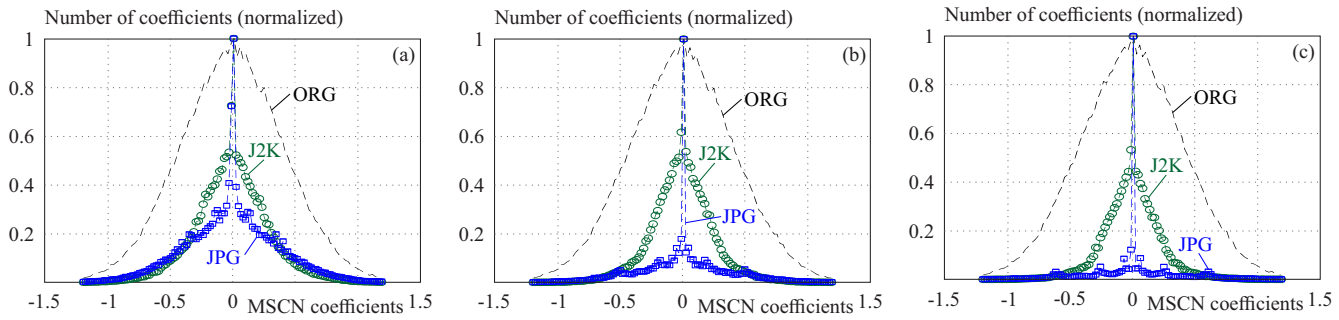

Fig. 2. MSCN histograms extracted from original LWIR images (ORG) and from JPEG and JPEG-2000 compressed images over three quality levels: (a) – level 5, (b) – level 3, (c) – level 1

presented in Tab. 1. The database contains images obtained from surveillance infrared cameras which justifies studying the effects of compression on the exploitability of LWIR images by human operators. In the aim of emphasizing different monitoring and surveillance scenarios under the outdoor conditions, the images have been chosen so that they include different targets. Therefore, the considered scenarios consist of: military and civilian people (stationary, walking or running, carrying various objects, group of soldiers, and a soldier in prone unsupported position), civilian vehicles, and military engines.

The database contains twenty reference images. Each pristine image is compressed applying JPEG and JPEG-2000 algorithms using five different quality parameters for each algorithm in order to obtain 200 degraded images with different subjective quality.

4.2 NSS of LWIR images

According to [10], highly successful IQA metrics have utilized band-pass statistical image models and particularly mean-subtracted contrast normalized (MSCN) coefficients and its variants. The MSCN is a result of a deeper study of natural scene statistics (NSS) on visible images conducted by Ruderman [49]. Results in [10] showed that LWIR images share a lot of features with visible images which makes using NSS models an appealing choice for LWIR images quality assessment. The NSS of LWIR images can successfully identify local distortions. Besides, they can offer an acceptable distortion estimate which may help in predicting the subjective assessment.

An assessment model based on NSS was suggested in [50]. The model was used to evaluate the quality of fused LWIR and visible images in the presence of additive white Gaussian noise (AWGN), NU distortions, blurring and JPEG compression. An opinion-distortion unaware and an opinion-aware fused image quality analyzer were developed to provide better correlation with the subjective scores.

Given an input image $\mathbf{r}$, the corresponding MSCN coefficients are denoted by $\hat{\mathbf{r}}$ and defined as [10]

$$\hat{r}(i,j) = \frac{r(i,j) - \mu(i,j)}{\sigma(i,j) + C}\,, \tag{3}$$

where μ and σ are weighted estimates of the local luminance mean and standard deviation

$$\mu(i,j) = \sum_{k=-K}^{K} \sum_{l=-L}^{L} \omega_{k,l} r_{k,l}(i,j)\,, \tag{4}$$

$$\sigma(i,j) = \sqrt{\sum_{k=-K}^{K} \sum_{l=-L}^{L} \omega_{k,l} \big(r_{k,l}(i,j) - \mu(i,j)\big)^2}\,, \tag{5}$$

with $\omega = \{\omega_{k,l} \mid k = -K, \ldots, K, l = -L, \ldots, L\}$ being a circularly-symmetric two-dimensional weighting function sampled out to three standard deviations and normalized to unit form. The constant C is used to prevent instabilities when σ tends towards zero.

For our database, the MSCN histograms averaged for both original and compressed LWIR images over multiple quality levels are depicted in Fig. 2. It can be noticed that the curves corresponding to both JPEG and JPEG-2000 are significantly tighter compared to original images.

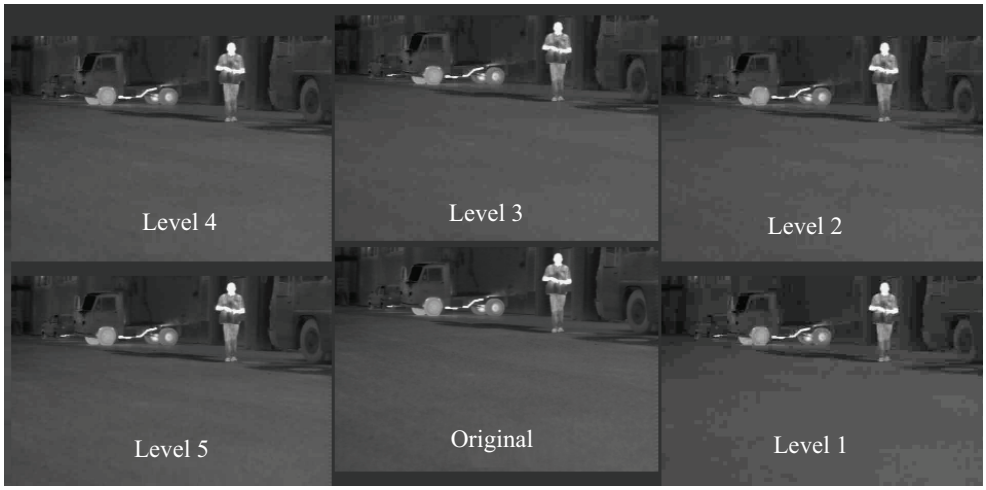

Fig. 3. Example of an original image and corresponding test images with different quality levels (training session)

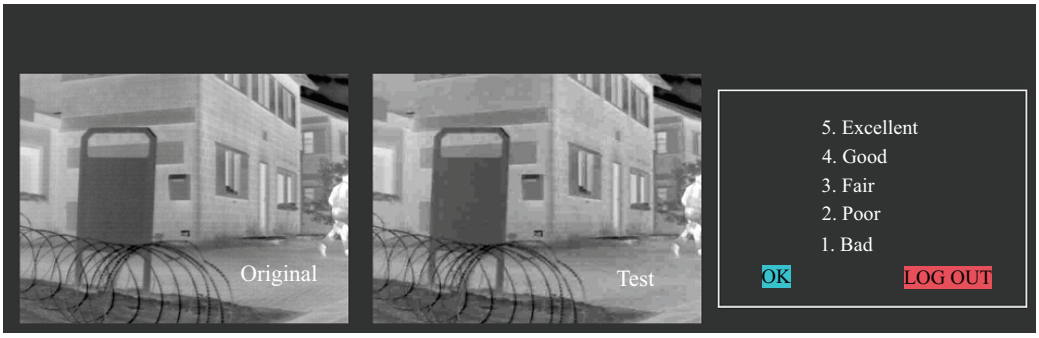

Fig. 4. Example of double stimulus evaluation

This result is more pronounced for JPEG algorithm with lower quality levels (3 and 1). Meanwhile, the MSCN histogram of JPEG-2000 images narrows slightly over different quality levels compared to JPEG. Furthermore, the pristine images exhibit a Gaussian-like appearance, while compressed images create a more Laplacian appearance.

4.3 Database annotation

Dedicated software with a graphical user interface (GUI) was developed in order to facilitate the subjective quality assessment process. The software was designed as a Web application whose code is written in PHP 7.1.0 under the framework Laravel 2.4.0, and MySQL 5.6.17 database management system is used to store the data collected from the subjects.

The application was designed in accordance with recommendations specified by ITU-R BT.500-14 [9]. Therefore, the developed evaluation system meets the following requirements:

- Before beginning the evaluation session, the observer has to be familiarized with the annotation system, the compression types and levels which will be used. The images shown to the observer during this phase shouldn't be part of the assessment process. Figure 3 depicts an example used in the training session where the original image is presented with test images compressed using different compression levels.
- Both pristine and compressed (test) image are displayed as a couple one beside the other on the same screen to fulfill the double stimulus requirement. This choice is motivated by the assumption that the subjects are mainly unfamiliar with LWIR images and should be guided through the exposure to the original and degraded versions of the image.
- The user is aware of the image type (original or degraded) and the original image position is kept the same during the whole testing session (left side).
- User's age and diopter correction (if any) are stored.
- The two images display lasts for a maximum of 20 seconds. If the user does not annotate the image quality within this duration, another pair of images will be randomly chosen and displayed.
- A black screen which lasts for 2 seconds is displayed between two consecutive couples of images.
- A session maximum length of 30 minutes is set. The user's assessment results are stored after the elapsed session and he is allowed to continue the evaluation process after a break.

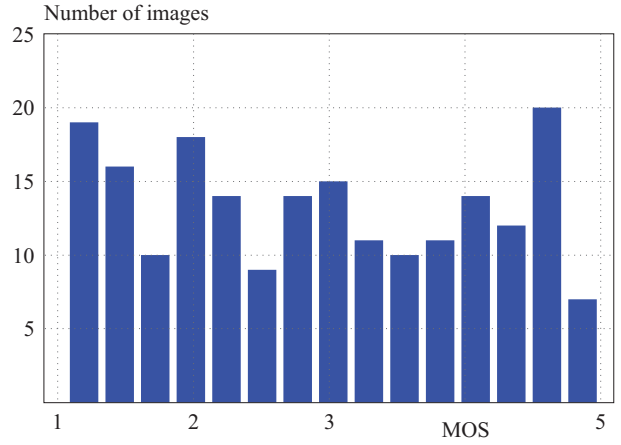

Fig. 5. Histogram of MOS scores for the complete database

- Two consecutive couples of images are subject to some constraints which are: the original image has to be different from the one previously displayed and the compression level must differ from the previous as well as the compression type.

Figure 4 shows the developed interface which displays a pair of images for evaluation. The evaluation scale contains five quality levels from "1" to "5", where "5" corresponds to the highest quality and "1" corresponds to the lowest. Depending on the perceived quality and taking into account the original image, the user can decide whether the quality of the test image is "bad", "poor", "fair", "good" or "excellent".

The evaluation process was carried out by 31 users (soldiers, officers and military students from the Military Academy, University of Defence in Belgrade) having different levels of familiarization with LWIR images and using the same setup. The age range of the users was from 20 to 45. In the testing phase, each user was exposed to a set of 200 images. All the users have successfully finished the scoring process in a maximum of two sessions (maxi-

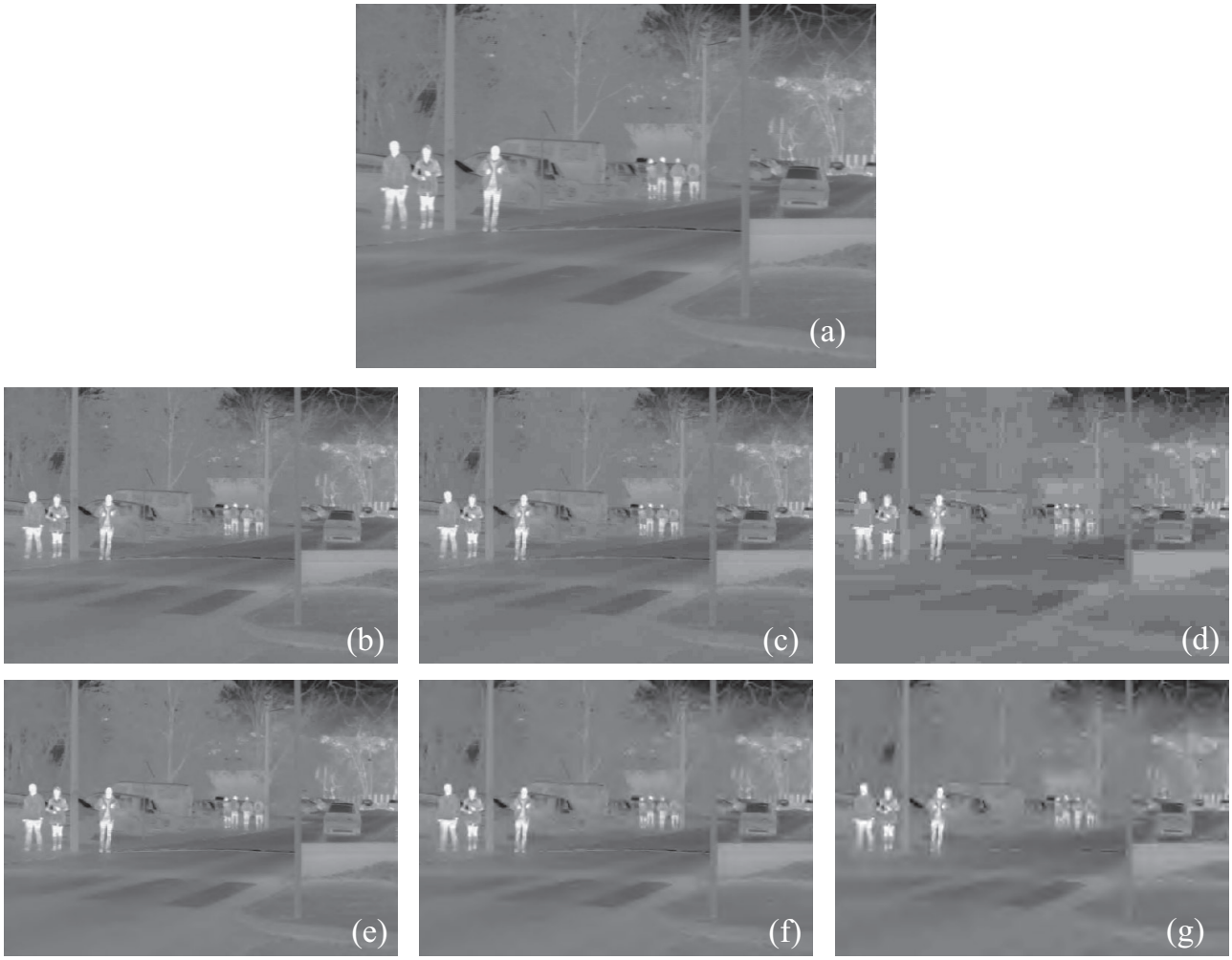

Fig. 6. Examples of JPEG and JPEG-2000 compressed images (quality levels 5, 3 and 1) from the database (FS refers to the file size in kB): (a) – original image (FS=266 kB), (b) – JPEG, quality level 5 (MOS=4.35, PSNR=39.86, FS=11 kB), (c) – JPEG, quality level 3 (MOS=3.52, PSNR=36.79, FS=8 kB), (d) – JPEG, quality level 1 (MOS=1.19, PSNR=30.9, FS=5 kB), (e) – JPEG-2000, quality level 5 (MOS=4.23, PSNR=39.94, FS=7 kB), (f) – JPEG-2000, quality level 3 (MOS=2.23, PSNR=35.78, FS=3 kB), and (g) – JPEG-2000, quality level 1 (MOS=1.19, PSNR=33.01, FS=2 kB)

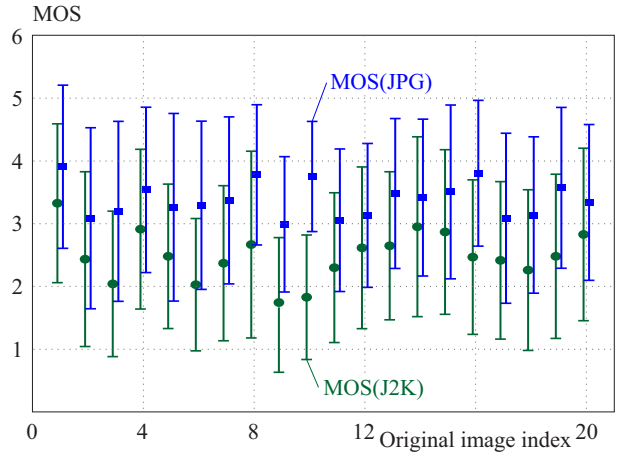

Fig. 7. MOS scores (±STD) per original image

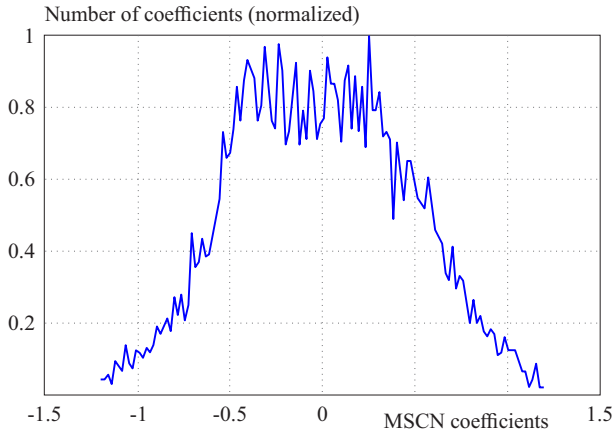

Fig. 8. Histogram of MSCN coefficients for the original image 10

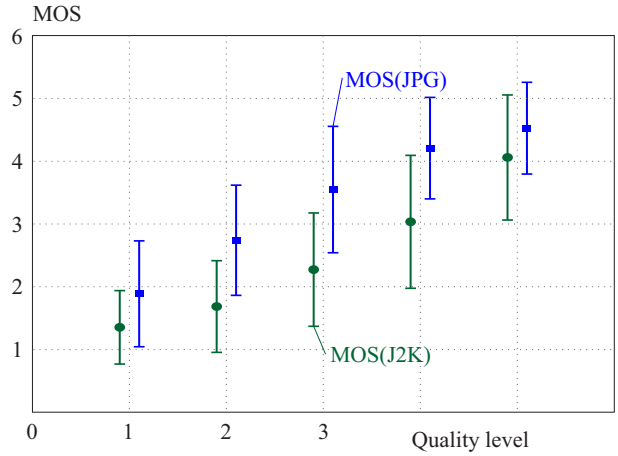

Fig. 9. MOS scores (± STD) as a function of quality level

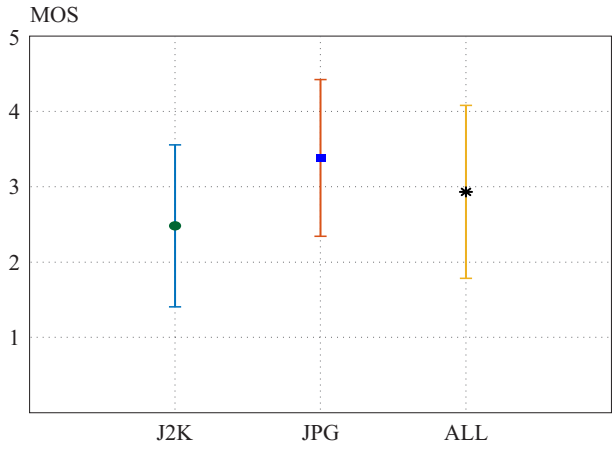

Fig. 10. MOS scores (± STD) as a function of compression type

mum of 30 minutes per session). Finally, a total number of 6200 scores were gathered for later exploitation.

The results interpretation relies on the mean opinion score (MOS). A plot of the histogram of the MOS scores for the complete database is represented in Fig. 5, showing an almost uniform distribution of the MOS scores for the full range of image quality.

Figure 6 depicts a comparison between JPEG and JPEG-2000 compressed images from our database. Three quality levels are taken into consideration (5, 3 and 1). It can be noticed that for the smallest degree of degradation of the original image (quality level 5), the MOS scores and PSNR values of the compressed images have approximately the same values. JPEG-2000 image with quality level 3 has a significant lower MOS compared to the JPEG image at the same quality level, and the advantage in JPEG compressed image quality was properly evaluated with PSNR objective measure. The MOS scores of the compressed images are equal for the maximum degradation of the original image (quality level 1). At the same time, the PSNR objective measure favors the image with JPEG-2000 compression (the difference is about 2 dB). Moreover, the file size of JPEG-2000 images is significantly lower than for JPEG images upon all three levels. Thereby, for this example, JPEG-2000 algorithm may take place if the user requires a higher compression ratio with a quality compromise.

5 Results and discussion

Five analyses have been carried out in order to better utilize and interpret the collected scores and images. The conducted analyses are the following: analysis by original image, analysis by quality level, analysis by compression type, analysis by file size and analysis by subject.

5.1 Analysis by original image

We represented the MOS scores and their standard deviations (MOS±STD) as a function of the original image index for the two types of compression (Fig. 7). For each original image, MOS scores were averaged over all five quality levels. It can be noticed that users prefer JPEG over JPEG-2000 images for all originals (MOS(JPG)>MOS(J2K)). Furthermore, it could be observed that the values of MOS(JPG)±STD overlap with

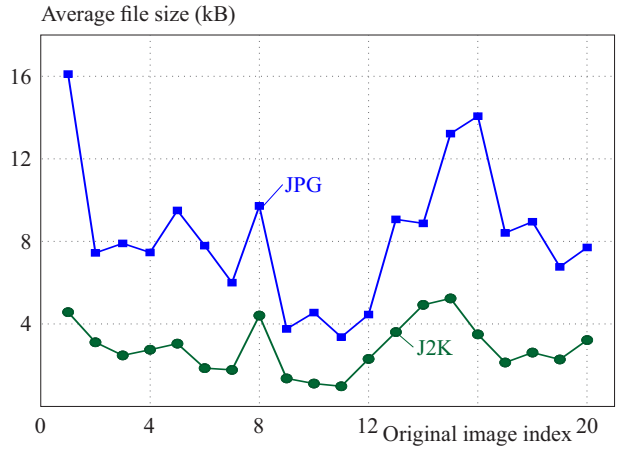

Fig. 11. Average file size per original image

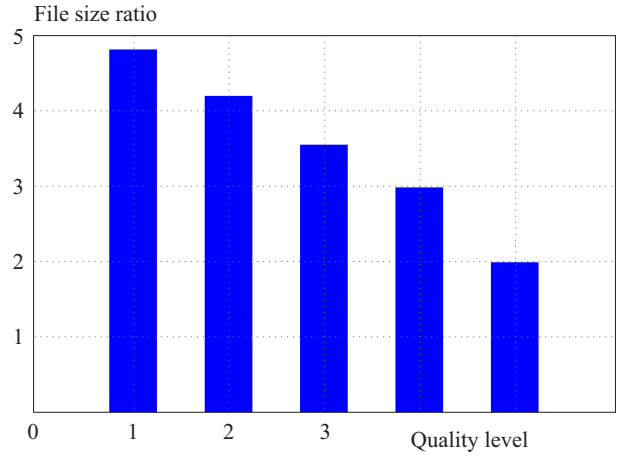

Fig. 12. Average file size ratio (JPEG to JPEG-2000) as function of quality level

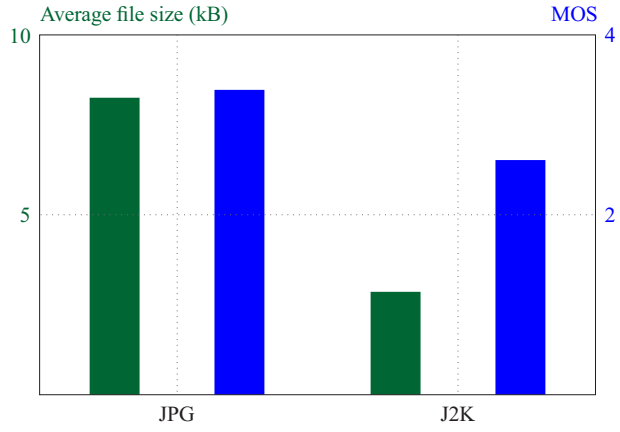

Fig. 13. Average file size and MOS scores for JPEG and JPEG-2000 images

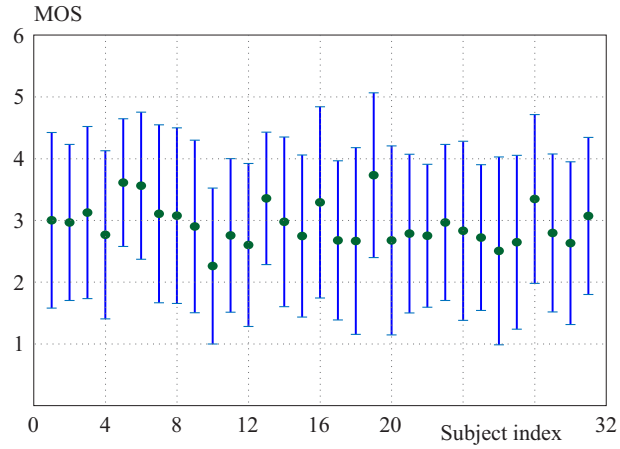

Fig. 14. MOS scores (± STD) per subject for the entire database

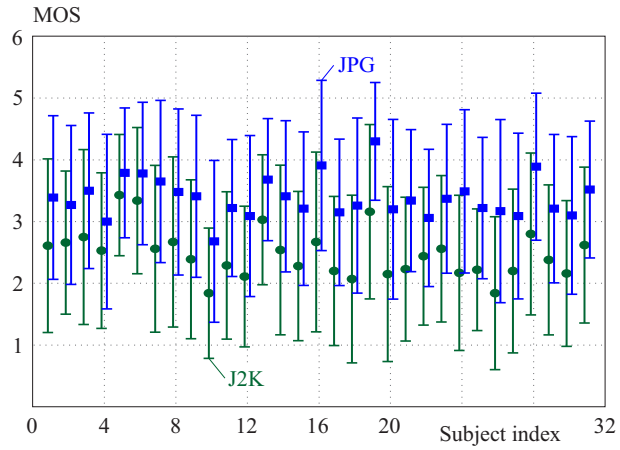

Fig. 15. MOS scores (± STD) for JPEG and JPEG-2000 images per subject

those of MOS(J2K) ± STD with the exception of original image 10. This could be explained by the histogram given in Fig. 8, where it is shown that the MSCN coefficients distribution behaves broader compared to the rest of images (MSCN histogram for the original images is given in Fig. 2). Despite this observation, scores for image 10 are kept to assure the fidelity of the database regarding subjects evaluation.

5.2 Analysis by quality level

Figure 9 depicts MOS scores and their standard deviations as a function of quality (degradation/compression) level for the entire compressed database. In this case, averaging of MOS scores was performed on all original images (20 values averaging). It is clear that the MOS scores of JPEG compressed images are significantly higher than those of JPEG-2000. In other words, for a given degradation level MOS(JPG) > MOS(J2K). Moreover, it is noticed that the quality difference between JPEG-2000 and JPEG images is lower for both bad and excellent quality levels (levels 1 and 5) since the observers can easily decide about image quality for those levels.

5.3 Analysis by compression type

The database contains 100 images with JPEG and JPEG-2000 compression. The mean values and standard deviations of the subjective scores for both types of compression and the complete database are given in Fig. 10.

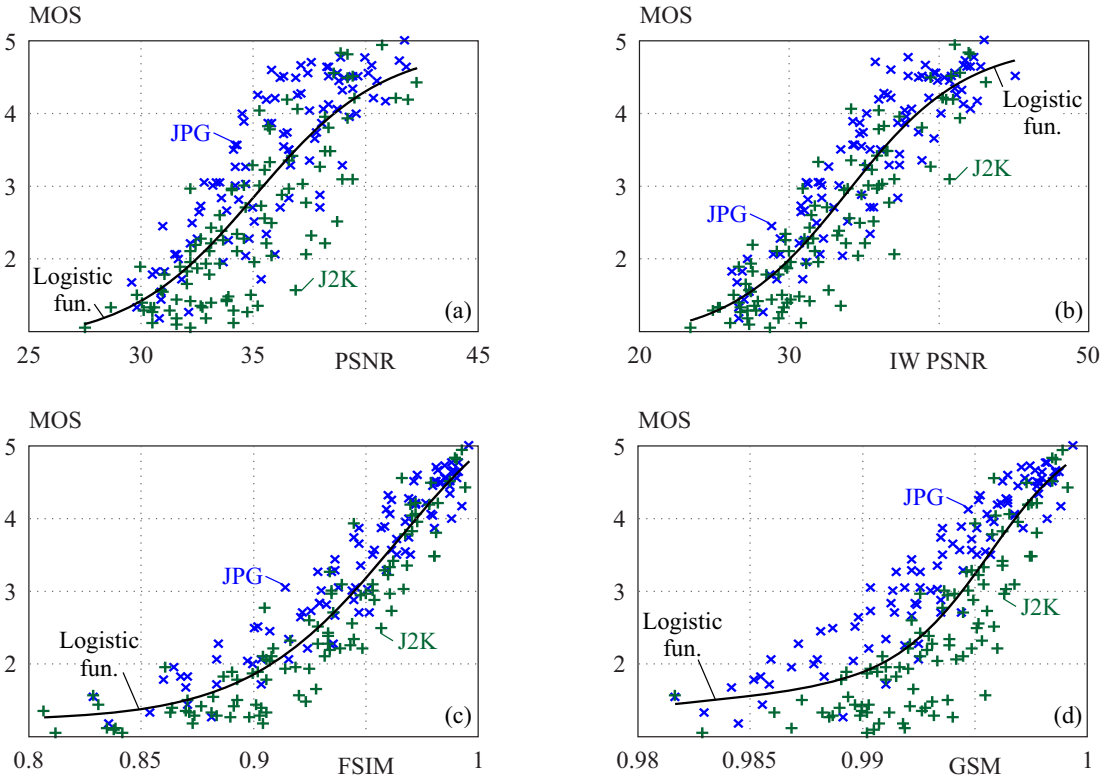

Fig. 16. Scatter plots of subjective MOS against chosen objective metrics (PSNR, IW PSNR, FSIM and GSM)

Averaging was performed on all 20 originals and five levels of image degradation. From Fig. 10, it can be seen that the average MOS for the whole database is 2.98, whereas the mean MOS score of images with JPEG compression is higher than the mean of MOS scores of images with JPEG-2000 compression (3.47 *vs* 2.48).

5.4 Analysis by file size

Similar to the described analyzes where the averaging of subjective scores was performed according to the original image, level and type of compression, the averaging of the sizes of the compressed images was also performed.

Figure 11 presents the average file size for JPEG and JPEG-2000 compression per original image, where it can be noticed that the curve of the average sizes of JPEG compressed images is always above the corresponding curve of the average values of images with JPEG-2000 compression.

File size ratio of compressed JPEG to JPEG-2000 images is inversely proportional to the level of quality as shown in Fig. 12. The ratio is between 5 (bad quality images) and 2 (excellent quality images).

Figure 13 depicts the average file size of all compressed JPEG and JPEG-2000 images and their corresponding average MOS scores. It can be noticed that JPEG-2000 algorithm is more beneficial in terms of compression ratio with moderate average MOS compared to JPEG algorithm.

5.5 Analysis by subject

The mean values of subjective quality evaluations for each subject (observer), on the complete database and by compression types, are shown in Fig. 14 and Fig. 15.

From Fig. 14, one can observe that the obtained MOS scores are above 2.5 for all the subjects except for subject 10. It can also be highlighted that subject 19 is not strict in evaluation. To account for individual preferences and image content differences, we have resorted to the outlier elimination procedure as explained in [9]. No outliers were found which means no need for rejecting any subject. In addition, all the subjects expressed a preference towards the JPEG over JPEG-2000 images (Fig. 15).

6 Objective quality assessment of the database

In the objective analysis, the performance of 23 conventional and state-of-the-art widely cited quality metrics is tested and results are summarized in Tab. 2. For the analysis of the degree of agreement between subjective and objective quality scores, five standard measures are employed, namely: linear correlation coefficient (LCC), Spearman's rank-order correlation coefficient (SROCC), mean absolute error (MAE), root mean square error (RMSE) and outlier ratio (OR). More details about the definitions and explanations of these quantitative indicators can be found in [8]. A logistic function with four parameters is used.

Table 2. Comparison of the objective IQA metrics' performance on the entire LWIR image database

Metric	LCC	SROCC	MAE	RMSE	OR (%)
PSNR	0.8144	0.8150	0.5229	0.6643	64.00
SSIM	0.8028	0.8031	0.5453	0.6826	70.50
MS-SSIM	0.8327	0.8276	0.5165	0.6339	68.50
VIF	0.9071	0.9019	0.3964	0.4817	60.00
MAD	0.8931	0.8847	0.4124	0.5149	63.00
GMSD	0.8860	0.8813	0.4291	0.5309	62.50
ADM	0.8815	0.8845	0.4275	0.5406	61.00
FSIM	0.9392	0.9370	0.3162	0.3931	48.50
GSM	0.8623	0.8526	0.4575	0.5797	64.00
IW SSIM	0.9043	0.9023	0.3935	0.4886	60.00
IW MSE	0.9008	0.9007	0.3948	0.4971	61.00
IW PSNR	0.9020	0.9007	0.3904	0.4942	60.00
CVSSI	0.8789	0.8729	0.4385	0.5460	62.50
MCSD	0.8678	0.8613	0.4562	0.5688	64.50
SR SIM	0.9112	0.9101	0.3759	0.4717	56.50
RFSIM	0.8698	0.8702	0.4679	0.5648	68.00
IGM	0.8641	0.8589	0.4615	0.5761	65.00
RVSIM	0.8973	0.8923	0.3861	0.5053	54.50
SFF	0.9032	0.9024	0.3962	0.4912	63.00
2stepQA	0.7730	0.7707	0.5487	0.7262	61.50
CEQI	0.9045	0.8991	0.3863	0.4882	57.00
HaarPSI	0.8591	0.8520	0.4704	0.5860	65.50
DISTS	0.8996	0.8833	0.3700	0.4671	56.00

Table 3. Comparison of the objective IQA metrics' performance on JPEG compressed LWIR images

Metric	LCC	SROCC	MAE	RMSE	OR (%)
PSNR	0.8377	0.8146	0.4362	0.5650	59.00
SSIM	0.8472	0.8403	0.4408	0.5496	66.00
MS-SSIM	0.8649	0.8423	0.4322	0.5193	64.00
VIF	0.8899	0.8708	0.3880	0.4719	60.00
MAD	0.9390	0.9202	0.2693	0.3558	40.00
GMSD	0.8940	0.8724	0.3738	0.4636	61.00
ADM	0.9051	0.9070	0.3472	0.4398	53.00
FSIM	0.9532	0.9445	0.2496	0.3129	40.00
GSM	0.9461	0.9406	0.2568	0.3350	42.00
IW SSIM	0.8963	0.8669	0.3712	0.4587	57.00
IW MSE	0.9050	0.8860	0.3569	0.4402	53.00
IW PSNR	0.9054	0.8860	0.3556	0.4392	57.00
CVSSI	0.8858	0.8606	0.3937	0.4800	61.00
MCSD	0.9113	0.8910	0.3380	0.4259	51.00
SR SIM	0.9115	0.9083	0.3299	0.4256	48.00
RFSIM	0.8762	0.8715	0.4044	0.4987	61.00
IGM	0.8756	0.8517	0.4106	0.4997	68.00
RVSIM	0.8843	0.8520	0.3728	0.4831	52.00
SFF	0.8917	0.8696	0.3789	0.4684	63.00
2stepQA	0.7844	0.7532	0.4897	0.6417	63.00
CEQI	0.9246	0.9151	0.3197	0.3941	53.00
HaarPSI	0.9107	0.8990	0.3463	0.4273	60.00
DISTS	0.9271	0.9123	0.2924	0.3879	42.00

Table 4. Comparison of the objective IQA metrics' performance on JPEG-2000 compressed LWIR images

Metric	LCC	SROCC	MAE	RMSE	OR (%)
PSNR	0.8145	0.7998	0.4879	0.6205	62.00
SSIM	0.7943	0.7810	0.5272	0.6499	70.00
MS-SSIM	0.8308	0.8031	0.4978	0.5953	72.00
VIF	0.8893	0.8666	0.4023	0.4892	64.00
MAD	0.9231	0.9073	0.3179	0.4114	49.00
GMSD	0.8856	0.8630	0.3968	0.4968	55.00
ADM	0.9249	0.9153	0.3199	0.4068	50.00
FSIM	0.9369	0.9283	0.2890	0.3741	51.00
GSM	0.8902	0.8597	0.3917	0.4874	63.00
IW SSIM	0.8948	0.8789	0.3854	0.4777	66.00
IW MSE	0.8937	0.8800	0.3775	0.4800	58.00
IW PSNR	0.8979	0.8800	0.3637	0.4709	57.00
CVSSI	0.8848	0.8602	0.4116	0.4985	69.00
MCSD	0.8975	0.8770	0.3772	0.4719	62.00
SR SIM	0.9351	0.9251	0.2921	0.3790	48.00
RFSIM	0.8840	0.8634	0.4089	0.5001	65.00
IGM	0.8616	0.8339	0.4423	0.5431	66.00
RVSIM	0.8893	0.8749	0.3615	0.4892	51.00
SFF	0.9013	0.8900	0.3732	0.4635	63.00
2stepQA	0.7941	0.7658	0.5022	0.6501	64.00
CEQI	0.9226	0.9028	0.3477	0.4127	61.00
HaarPSI	0.8846	0.8620	0.4135	0.4989	70.00
DISTS	0.8996	0.8833	0.3700	0.4671	56.00

Besides studying on the entire database, the chosen metrics are applied to test their performance over compression algorithms (JPEG and JPEG-2000) separately. Results for JPEG images are given in Tab. 3, while Tab. 4 summarizes results for JPEG-2000 compressed images.

Figure 16 depicts the scatter plots of subjective MOS against objective metrics scores obtained for the interesting metrics (PSNR, IW PSNR, FSIM and GSM) applied on the entire database. The choice of these metrics is basically motivated by their reputation and the results given in Tab. 2, Tab. 3 and Tab. 4. As already mentioned, PSNR belongs to the group of metrics which are widely used because of their simplicity and ease of implementation. However, Fig. 16(a) shows that PSNR does not fit well the subjective scores. The corresponding SROCC is 81.5% (Tab. 2). Therefore, a modified PSNR version is discussed, namely: IW PSNR. Inversely, this PSNR-based metric provides an acceptable interpretation of the MOS scores, Fig. 16(b), with SROCC of about 90 % (Tab. 2). This can be explained by the fact that IW PSNR measure takes into consideration information theoretic analysis of visual information content and fidelity in pooling stage.

From the obtained results, it is observed that the FSIM metric exhibits the best fit with the subjective data in comparison with other studied metrics which can be proved by the dense scatter plot shown in Fig. 16(c). Table 2 shows that FSIM outscored the other objective IQA metrics for the entire database where LCC and SROCC are about 94 % with the lowest MAE and RMSE. The same result is observed for both types of compression in Tab. 3 and Tab. 4, with summarized measures performance calculated for the objective IQA metrics for JPEG and JPEG-2000 images respectively.

GSM metric is selected due to the significant discrepancy between results obtained for JPEG and JPEG-2000 image subsets. This measure, in terms of performance on the JPEG subset of images, is ranked second (LCC=94.61%), while its poorer performance on the subset of images with JPEG-2000 compression (LCC=89.02%) has also led to poor performance on complete database (LCC=86.23%). It is also worthwhile to mention that this metric has a narrow dynamic range, which can be noticed in Fig. 16(d).

The less significant results are obtained for the 2stepQA metric [39] with an SROCC of 77.07 %. A plausible explanation for this observation stems from the fact that 2stepQA is based on a database trained on visible images, which may not be a suitable option for LWIR images quality assessment.

For the JPEG images, MAD metric is also a suitable candidate for the quality assessment. In this case, the LCC is 93.9 %. Furthermore, the computational efficient SR SIM metric provided an acceptable agreement score for the JPEG-2000 images. The corresponding LCC is 93.51 %.

7 Conclusion

In this work, we studied the quality of infrared (LWIR) images of a novel database which we provided to the research community. The database contains 20 pristine images and 200 JPEG and JPEG-2000 compressed images, with five degradation levels considered. The quality of the images was assessed by observers and by applying objective measures. The subjective evaluation was conducted by 31 subjects who are all military soldiers, students and officers from the Military Academy, University of Defence in Belgrade. Different analyses were elaborated in order to evaluate the collected scores and images. The obtained results were statistically satisfying.

Several objective full-reference image quality assessment measures were examined and among them the FSIM model performed the best for both types of compression (JPEG and JPEG-2000) and on complete database. When the entire database is considered, the FSIM obtained degree of agreement with the results of subjective tests, measured through the linear correlation coefficient, of 94%. Therefore, FSIM better interprets the subjective evaluations on compressed infrared images.

The database used in this work includes images originated from different surveillance scenarios while putting an emphasis on various types of human activity. Besides, this work provides experimental, publicly available results related to the image quality assessment of two widely used compression algorithms. The subjective results can be considered as an important benchmark for developing objective image quality algorithms which highly correlates with the human perception. Developing such a tool can be beneficial to surveillance systems given the fact that it is capable of assessing the quality of compressed LWIR images.

As a follow-up to this work, an extension of the database is underway to study other types of degradation in addition to compression (blurring, additive white noise and non-uniformity).

Acknowledgements

This research has been a part of the project No. VA-TT/3/20-22 supported by the Ministry of Defence, Republic of Serbia.

References

[1] D. R. Bull and F. Zhang, *Intelligent image and video compression: communicating pictures*, Second Edition, Academic Press, 2021.

[2] R. Gade and T. B. Moeslund, "Thermal cameras and applications: a survey", *Machine Vision and Applications*, vol. 25, no. 1, pp. 245–262, doi: 10.1007/s00138–013–0570–5, 2014.

[3] D. Peric, B. Livada, M. Peric, and S. Vujic, "Thermal imager range: predictions, expectations, and reality", *Sensors*, vol. 19, no. 15, pp. 3313-1-23, doi: 10.3390/s19153313, 2019.

[4] M. Diakides, J. D. Bronzino, and D. R. Peterson, *Medical infrared imaging: principles and practices*, First Edition, CRC Press, 2012.

[5] R. Usamentiaga, P. Venegas, J. Guerediaga, L. Vega, J. Molleda, and F. G. Bulnes, "Infrared thermography for temperature measurement and non–destructive testing", *Sensors*, vol. 14, no. 7, pp. 12305–12348, doi: 10.3390/s140712305, 2014.

[6] M. J. Haque and M. Muntjir, "Night vision technology: an overview", *International Journal of Computer Applications*, vol. 167, no. 13, pp. 8887–37–42, doi: 10.5120/ijca914562, 2017.

[7] G. Chen and W. Wang, "Target recognition in infrared circumferential scanning system via deep convolutional neural networks", *Sensors*, vol. 20, no. 7, pp. 1922–1–18, doi: 10.3390/s20071922, 2020.

[8] B. P. Bondžulić, B. Z. Pavlović, and V. S. Petrović, "Performance analysis of full–reference objective image and video quality assessment metrics", *Vojnotehnicki glasnik / Military Technical Courier*, vol. 66, no. 2, pp. 322–350, doi: 10.5937/vojtehg66-12708, 2018.

[9] *Methodologies for the subjective assessment of the quality of television images*, Recommendation ITU–R, BT, 500–14 10/2019, International Telecommunication Union, Geneva, Switzerland, 2020.

[10] T. R. Goodall, A. C. Bovik, and N. G. Paulter, "Tasking on natural statistics of infrared images", *IEEE Transactions on Image Processing*, vol. 25, no. 1 pp. 65–79, doi: 10.1109/TIP.2496289, 2015.

[11] K. Hossain, C. Mantel, and S. O. Forchhammer, "No-reference prediction of quality metrics for H. 264-compressed infrared sequences for unmanned aerial vehicle applications", *Journal of Electronic Imaging*, vol. 28, no. 4, pp. 043012–1–14, doi: 10.1117/1.JEI.28.4.043012, 2019.

[12] M. Teutsch, S. Sedelmaier, S. Moosbauer, G. Eilertsen, and T. Walter, "An evaluation of objective image quality assessment for thermal infrared video tone mapping", *IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, Seattle, WA, USA, June 14–19, pp. 108–109, doi: 10.1109/CVPRW50498.00062, 2020.

[13] E. Belyaev, C. Mantel, and S. Forchhammer, "High bit depth infrared image compression via low bit depth codecs", *SPIE 10403, 15th Infrared Remote Sensing and Instrumentation*, San Diego, California, USA, August 7–8, pp. 104030A–1–8, doi: 10.1117/12.2275542, 2017.

[14] D. Salomon and G. Motta, *Handbook of data compression*, Fifth Edition, Springer, 2010.

[15] A. J. Hussain, A. Al-Fayadh, and N. Radi "Image compression techniques: a survey in lossless and lossy algorithms", *Neurocomputing*, vol. 300, pp. 44–69, doi: 10.1016/j.neucom.02.094, 2018.

[16] M. S. Al-Ani and F. H. Awad "The JPEG image compression algorithm", *International Journal of Advances in Engineering & Technology*, vol. 6, no. 3, pp. 1055–1062, 2013.

[17] R. C. Gonzalez and R. E. Woods, *Digital image processing*, Third Edition, Pearson, Ch 10, pp. 743–747, 2008.

[18] C. M. Brislawn and M. D. Quirk, *Image compression: JPEG-2000 standard*, Encyclopedia of Optical and Photonic Engineering, Second Edition, CRC Press, Ch 83, 2015.

[19] F. Ebrahimi, M. Chamik, and S. Winkler, "JPEG vs. JPEG 2000: an objective comparison of image encoding quality", *Optical Science and Technology, the SPIE 49th Annual Meeting*, Denver, Colorado, USA, August 2–6, pp. 300–308, doi: 10.1117/12.564835, 2004.

[20] R. Maini and S. Mehra "A review on JPEG2000 image compression", *International Journal of Computer Applications*, vol. 11, no. 9, pp. 43–47, doi: 10.5120/1607-2159, 2000.

[21] Y. Q. Shi and H. Sun, *Image and video compression for multimedia engineering: fundamentals, algorithms, and standards*, Third Edition, CRC Press, 2019.

[22] Z. Wang and A. C. Bovik, "Mean squared error: love it or leave it? A new look at signal fidelity measures", *IEEE Signal Processing Magazine*, vol. 26, no. 1, pp. 98–117, doi: 10.1109/MSP.2008.930649, 2009.

[23] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: from error visibility to structural similarity", *IEEE Transactions on Image Processing*, vol. 13, no. 4, pp. 600–612, doi: 10.1109/TIP.2003.819861, 2004.

[24] Z. Wang, E. P. Simoncelli, and A. C. Bovik, "Multiscale structural similarity for image quality assessment", *The* 37^{th} *Asilomar Conference on Signals, Systems & Computers*, Pacific Grove, CA, USA, November 9-12, pp. 1398–1402, 2003.

[25] Z. Wang and Q. Li, "Information content weighting for perceptual image quality assessment", *IEEE Transactions on Image Processing*, vol. 20, no. 5, pp. 1185–1198, doi: 10.1109/TIP.2092435, 2010.

[26] L. Zhang and H. Li, "SR SIM: a fast and high performance IQA index based on spectral residual", *19th IEEE International Conference on Image Processing*, Orlando, FL, USA, September 30-October 3, pp. 1473–1476, doi: 10.1109/ICIP.6467149, 2012.

[27] X. Hou and L. Zhang, "Saliency detection: a spectral residual approach", *IEEE Conference on Computer Vision and Pattern Recognition*, Minneapolis, MN, USA, June 17-22, pp. 1–8, doi: 10.1109/CVPR.383267, 2007.

[28] H. Jia, L. Zhang, and T. Wang, "Contrast and visual saliency similarity–induced index for assessing image quality", *IEEE Access*, vol. 6, pp. 65885–65893, doi: 10.1109/ACCESS.2878739, 2018.

[29] S. Li, F. Zhang, L. Ma, and K. N. Ngan, "Image quality assessment by separately evaluating detail losses and additive impairments", *IEEE Transactions on Multimedia*, vol. 13, no. 5, pp. 935–949, doi: 10.1109/TMM.2152382, 2011.

[30] A. Liu, W. Lin, and M. Narwaria, "Image quality assessment based on gradient similarity", *IEEE Transactions on Image Processing*, vol. 21, no. 4, pp. 1500–1512, doi: 10.1109/TIP.2175935, 2011.

[31] W. Xue, L. Zhang, X. Mou, and A. C. Bovik, "Gradient magnitude similarity deviation: a highly efficient perceptual image quality index", *IEEE Transactions on Image Processing*, vol. 23, no. 2, pp. 684–695, doi: 10.1109/TIP.2293423, 2013.

[32] T. Wang, L. Zhang, H. Jia, B. Li, and H. Shu, "Multiscale contrast similarity deviation: an effective and efficient index for perceptual image quality assessment", *Signal Processing: Image Communication*, vol. 45, pp. 1–9, doi: 10.1016/j.image.04.005, 2016.

[33] L. Zhang, L. Zhang, X. Mou, and D. Zhang, "FSIM: a feature similarity index for image quality assessment", *IEEE Transactions on Image Processing*, vol. 20, no. 8, pp. 2378–2386, doi: 10.1109/TIP.2109730, 2011.

[34] R. Reisenhofer, S. Bosse, G. Kutyniok, and T. Wiegand, "A Haar wavelet–based perceptual similarity index for image quality assessment", *Signal Processing: Image Communication*, vol. 61, pp. 33–43, doi: 10.1016/j.image.2017.11.001, 2018.

[35] L. Zhang, L. Zhang, and X. Mou, "RFSIM: a feature based image quality assessment metric using Riesz transforms", *17th IEEE International Conference on Image Processing*, Hong Kong, Hong Kong, September 12-15, pp. 321–324, doi:10.1109/ICIP.5649275, 2010.

[36] G. Yang, D. Li, F. Lu, Y. Liao, and W. Yang, "RVSIM: a feature similarity method for full-reference image quality assessment", *EURASIP Journal on Image and Video Processing*, vol. 2018, no. 1, pp. 1–15, doi: 10.1186/s13640–018–0246–1, 2018.

[37] M. Layek, A. F. M. Uddin, T. P. Le, T. Chung, and E. N. Huh, "Center-emphasized visual saliency and a contrast-based full reference image quality index", *Symmetry*, vol. 11, no. 3, pp. 296–1–14, doi: 10.3390/sym11030296, 2019.

[38] E. C. Larson and D. M. Chandler, "Most apparent distortion: full-reference image quality assessment and the role of strategy", *Journal of Electronic Imaging*, vol. 19, no. 1, pp. 011006–1–21, doi: 10.1117/1.3267105, 2010.

[39] X. Yu, C. G. Bampis, P. Gupta, and A. C. Bovik, "Predicting the quality of images compressed after distortion in two steps", *IEEE Transactions on Image Processing*, vol. 28, no. 12, pp. 5757–5770, doi: 10.1109/TIP.2922850, 2019.

[40] H. R. Sheikh and A. C. Bovik, "Image information and visual quality", *IEEE Transactions on Image Processing*, vol. 15, no. 2, pp. 430–444, doi: 10.1109/TIP.2005.859378, 2006.

[41] J. Wu, W. Lin, G. Shi, and A. Liu, "Perceptual quality metric with internal generative mechanism", *IEEE Transactions on Image Processing*, vol. 22, no. 1, pp. 43–54, doi: 10.1109/TIP.2214048, 2012.

[42] H. W. Chang, H. Yang, Y. Gan, and M. H. Wang, "Sparse feature fidelity for perceptual image quality assessment", *IEEE Transactions on Image Processing*, vol. 22, no. 10, pp. 4007–4018, doi: 10.1109/TIP.2266579, 2013.

[43] K. Ding, K. Ma, S. Wang, and E. P. Simoncelli, "Image quality assessment: unifying structure and texture similarity", *IEEE Transactions on Pattern Analysis & Machine Intelligence*, vol. 44, no. 5, pp. 2567-2581, doi: 10.1109/TPAMI.2020.3045810, 2022.

[44] A. Mittal, R. Soundararajan, and A. C. Bovik, "Making a "completely blind" image quality analyzer", *IEEE Signal Processing Letters*, vol. 20, no. 3, pp. 209–212, doi: 10.1109/LSP.2227726, 2012.

[45] S. Merrouche, D. Bujakovic, M. Andric, and B. Bondulic, "Classification of infrared image distortions using contrast–based features", *International Symposium on Signals, Circuits and Systems (ISSCS)*, Iasi, Romania, July 11-12, pp. 1–4, doi: 10.1109/ISSCS.8801761, 2019.

[46] N. J. Morris, S. Avidan, W. Matusik, and H. Pfister, "Statistics of infrared images", *IEEE Conference on Computer Vision and Pattern Recognition*, Minneapolis, MN, USA, June 17-22, pp. 1–7, doi: 10.1109/CVPR.383003, 2007.

[47] A. Toet, M. A. Hogervorst, and A. R. Pinkus, "The TRICLOBS dynamic multi-band image data set for the development and evaluation of image fusion methods", *PloS one*, vol. 11, no. 12, e0165016, doi: 10.1371/journal.pone.0165016, 2016.

[48] B. Bondžulić and V. Petrović, "Multisensor background extraction and updating for moving target detection", *11th International Conference on Information Fusion*, Cologne, Germany, 30 June-3 July, pp. 1–8, doi: 10.1109/ICIF.4632435, 2008.

[49] D. L. Ruderman, "The statistics of natural images", *Network: Computation in Neural Systems*, vol. 5, no. 4, pp. 517–548, 1994.

[50] D. E. Moreno-Villamarín, H. D. Bentez–Restrepo, and A. C. Bovik, "Predicting the quality of fused long wave infrared and visible light images", *IEEE Transactions on Image Processing*, vol. 26, no. 7, pp. 3479–3491, doi: 10.1109/TIP.2695898, 2017.

Received 14 February 2022

Omar Zelmati received the undergraduate scientific degree at National Preparatory School for Engineering Studies, Algiers, Algeria in 2013 and MSc degree in computer science engineering from Ecole Militaire Polytechnique Chahid Abderrahmane Taleb, Algiers, Algeria in 2016. He is a doctoral student at Military Academy, University of Defence in Belgrade. His research interests are image quality assessment and QoS improvement. He has published 8 papers in national and international conferences and journals.

Boban Bondžulić received the BSc degree in electrical engineering from Military Technical Academy, Serbia in 2000, MSc degree at the School of Electrical Engineering, University of Belgrade, Serbia in 2005 and PhD degree from Faculty of Technical Sciences, University of Novi Sad, Serbia in 2016. He is an associate professor at the Department of Telecommunications and Informatics, Military Academy, University of Defence in Belgrade. His research interests include information fusion, image and video quality evaluation, detection and tracking of moving objects, pattern recognition. He has published more than 70 papers in national and international conferences and journals.

Boban Pavlović received the BSc degree in electrical engineering from Military Technical Academy, Serbia, in 1999, MSc degree from School of Electrical Engineering, University of Belgrade, Serbia in 2006 and Ph.D. from Military Academy, University of Defence in Belgrade, Serbia in 2010. Currently, he is working as an associate professor at the Department of Telecommunications and Informatics, Military Academy, University of Defence in Belgrade, Serbia. His research interest includes: telecommunication networks, quality of service, traffic engineering, image and video quality evaluation, and QoS improvement. He has published more than 50 research papers.

Ivan Tot received the BSc degree in electrical engineering from Military Technical Academy, Serbia in 1999, MSc degree at the Faculty of Technical Sciences Cacak, University of Kragujevac, Serbia in 2008 and PhD degree from the Military Academy, University of Defence, Serbia in 2010. He is an associate professor at the Department of Telecommunications and Informatics, Military Academy, University of Defence in Belgrade. His research interests include database and application development, security of computer systems. He has published more than 70 papers in national and international conferences and journals.

Saad Merrouche received the undergraduate scientific degree at National Preparatory School for Engineering Studies, Algiers, Algeria in 2011, MSc degree in electrical engineering from Ecole Militaire Polytechnique, Algiers, Algeria in 2015 and PhD degree from the Military Academy, University of Defence in Belgrade. His research interests include stereo vision, image quality assessment, distance measurement of objects and classification. He has published 10 papers in national and international conferences and journals.