

# Research of Methods for Image Sharpness Evaluation in Photos of People

Victoria Vysotska<sup>1</sup>, Nataliia Sharonova<sup>2</sup>, Mariya Shirokopetleva<sup>3</sup>, Oleksandr Dolhanenko<sup>3</sup>, Anastasiya Chupryna<sup>3</sup> and Serhii Smelyakov<sup>3</sup>

<sup>1</sup> Lviv Polytechnic National University, Stepan Bandera, Lviv, 79013, Ukraine,

<sup>2</sup> National Technical University "KhPI", Kyrpychova str. 2, Kharkiv, 61002, Ukraine,

<sup>3</sup> Kharkiv National University of Radio Electronics, Nauky Ave. 14, Kharkiv, 61166, Ukraine

## Abstract

The subject matter of the article is image sharpness evaluation in photos of people.

The goal of the work is to analyse the existing methods of image sharpness evaluation, compare their performance and quality of results, suggest improvements for the use case of sharpness classification of photos of people, where large quantities of background blur is present due to the aperture effect.

In this article the methods for image sharpness evaluation were described and tested on a set of selected images. The images contained different subject sizes and types of blur. The following methods were used: Fast Fourier Transform (FFT), Variance of the Laplacian, Appearance-based face detection algorithms, metadata analysis, linear trendline analysis. As a result, the problem of naturally blurred background was demonstrated and conclusions were made. An alternative method of sharpness evaluation was de-scribed, which solves the mentioned problem. The suggested improved algorithm was tested to determine if it satisfies expectations and solves the identified problems. To implement the Fast Fourier Transform and Variance of the Laplacian methods, the OpenCV library was used. The following results were ob-tained – when using the default implementation, the FFT and Variance of the Laplacian methods are not reliable for evaluating sharpness for images containing large and unstable quantities of naturally blurred background (due to the open aperture) and when using different settings of the lens and camera.

The following conclusions were made: steps need to be taken to eliminate the factors of unstable quantities of naturally blurred background and camera preferences and in this way improve accuracy, reliability of sharpness evaluation. This means evaluating the sharpness of parts of the images and the images as a whole. These steps include but are not limited to face position detection, identifying faces that were supposed to be in focus when the photo was taken, sharpness evaluation of only areas that were in-tended to be in focus. Plans were set for further research and improvements of the suggested algorithm.

## Keywords

Sharpness, image, aperture, depth of field, focus, FFT, variance of the Laplacian

## 1. Introduction

Photography is very subjective. There are many ways to take a photo, since all of us see the world in very different ways. There are, however, some basic principles for achieving a generally well composed photograph. You may or may not follow these principles and still get a great photo that tells a story, but statistically speaking the best looking and most objectively recognised as professional and pleasing photos are the ones which have the subject in critical sharp focus.

Achieving focus on a photo is a very complicated process, which has multiple factors influencing the final result. A general smartphone user takes this process for granted, as the phone

---

COLINS-2024: 8th International Conference on Computational Linguistics and Intelligent Systems, April 12–13, 2024, Lviv, Ukraine

✉ victoria.a.vysotska@lpnu.ua (V. Vysotska); nvsharonova@ukr.net (N. Sharonova);

marija.shirokopetleva@nure.ua (M. Shirokopetleva); oleksandr.dolhanenko@nure.ua (O. Dolhanenko);

anastasiya.chupryna@nure.ua (A. Chupryna); serhii.smeliakov@nure.ua (S. Smelyakov)

ORCID 0000-0001-6417-3689 (V. Vysotska) 0000-0002-8161-552X (N. Sharonova) 0000-0002-7472-6045

(M. Shirokopetleva); 0000-0002-3996-6940 (O. Dolganenko); 0000-0003-0394-9900 (A. Chupryna); 0000-0002-5791-2479 (S. Smelyakov)



© 2024 Copyright for this paper by its authors.

Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

does all decision making and photo processing for us in real time, most of the time achieving pretty good results for general social media use. The algorithms used there are fine-tuned, subject oriented. Most of the time our phones use machine learning to detect the scene and tune the settings to achieve better results.

Our smartphones are getting better and better every year in terms of photography, there is even the ability to imitate expensive lenses by adding a fake depth-of-field to the photos [1]. However, any well-established professional photographer will say how important it is to be in control of the manual settings of the camera, being the decision maker and scene establisher, getting the exact result the creator wants, and not the tool on its own.

Moreover, they use professional hardware, which may contain some "smart" features, but generally is very exposed to manual overriding, giving the operator more flexibility. However, where humans are involved mistakes are present.

Having photographed an event, say, a wedding, the photographer usually spends hours and sometimes days looking at more than 2000 photos, filtering out the ones that are to be deleted and highlight the best ones for further editing. This is a very long and complicated process of comparing, what seem to be, identical photographs at first glance (but actually different in slight ways), which can be a big factor for the final result. The most important thing the photographer looks at is the sharpness of the subjects. It is a general rule, that, when photographing humans or animals, the eyes are the ones that need to be in critical sharp focus. Everything else in the photo can be changed - the lighting can be increased or decreased, some elements can be added or removed, but sharpening a blurry subject is a very destructive process which is highly discouraged.

The goal of this work is to analyse and compare different algorithms and program solutions for identifying the sharpness of a photo (containing humans as subjects) automatically and proposing improvements of the existing algorithms to solve the problems which will be described further on.

## 2. Related Works

When it comes down to solving a scientific problem, it is required to operate with objective terms and calculations. That is why it is important to define what a sharp photo is. The common way of measuring the sharpness is by the "rise distance" of edges within the image. That way, the sharpness is determined by the distance of a pixel level between 10% and 90%. Therefore, it can be stated that sharpness is measured by analysing the intensity of edge gradients [2]. However, the threshold level of intensity to classify a photo being sharp or not is very circumstantial and cannot be standardised. During the following experiments, a set of reference photos will be used to determine the minimum and maximum values of sharpness and thus, providing the threshold. Apart from that, the problem of image quality assessment [3] is well-known and was attempted to be generally solved by many.

In order to start detecting and classifying the sharpness of photos it is first needed to clarify which camera and lens settings influence it. The camera in combination with the lens has 3 main settings: aperture, shutter speed and ISO.

Aperture is the main setting of the lens. It is generally an opening that can be bigger or smaller and thus let in more or less light. It is usually preferred to let in as much light as possible, therefore "opening up", or "making the aperture wide open", but this has a side effect that can directly influence the sharpness of the subject. The more open the aperture - the more shallow the depth of field is. For cameras that can only focus on one object distance at a time, depth of field is the distance between the nearest and the furthest objects that are in acceptably sharp focus.

By knowing the DOF (Depth Of Field) we can understand what depth of the image had to be in focus. To give a more clear understanding of why this is important an example image is provided, shot with a very shallow DOF (see Figure 1).

Figure 1 (photo by Dolhanenko O.) demonstrates a shot with a very open aperture of 2.8 (the lower the number the more open the aperture is). The subject that lays within the shallow depth

of field is in critical sharp focus, however the secondary subjects which are before and after the field boundaries are not in focus at all (in this case only before the subject on the right).

By means of traditional photo sharpness detection algorithms, this is not a sharp photo, as the area in critical sharp focus is very small compared to the blurry part (this will be experimentally tested further on). However, if the subject is correctly identified among the two and is exclusively checked, then the photo is in fact in critical sharp focus and is acceptable for further editing.



**Figure 1:** Example of a photo with a shallow DOF

Another setting that influences the sharpness of a photo is the shutter speed. The rule is "the faster - the better" for general photography. The quicker the curtain collapses before the sensor - the less blur there will be on the photo, as each movement of subjects in the photo during the shot will cause their sharpness to decrease. Especially when shooting in low light conditions, when fast shutter speeds are not available (otherwise the photo will be too dark) this motion blur is quite noticeable due to minor movements.

The pursuit of accurate image sharpness assessment has led to the exploration of various approaches, encompassing spatial domain-based methods, spectral domain-based methods, learning-based methods [4], and a combination of these techniques, each presenting unique advantages and challenges, including utilization of methods such as Local Phase Coherence [5, 6], Edge Information analysis approach [7], Normal-Gradient-Based approach [8], Gradient Neighbourhood-Weighted approach [9] and others.

Zhu et al. (2023) [10] conducted a comprehensive review, offering insight into the current trends and performance comparisons of notable algorithms, revealing a landscape of ongoing innovation aimed at overcoming the shortcomings of existing methods.

Bielievtssov et al. (2018) [11] investigated network technology for the transmission of visual information, highlighting the importance of maintaining image quality in the context of digital communication and storage. This work is foundational, setting the stage for further research into image quality assessment methods that are critical for various applications, including but not limited to, facial recognition, social media, and professional photography.

In the domain of image search and retrieval, Smelyakov et al. (2020) [12] introduced an innovative approach to image engine search for big data warehouses, emphasizing the necessity of high-quality image processing for efficient and accurate image retrieval. This development is particularly relevant to our research as it underscores the significance of image sharpness in enhancing the performance of search engines, which often rely on visual content analysis to function effectively.

Furthermore, the effectiveness of preprocessing algorithms for natural language processing applications was explored by Smelyakov et al. (2020) [13], illustrating the broad applications of image and signal processing techniques across various fields of computer science. Although focused on natural language processing, the principles of preprocessing and quality enhancement are applicable to the domain of image processing, providing insights into methods that could potentially improve image sharpness evaluation algorithms.

The development of no-reference (NR) [14] sharpness metrics has been particularly noteworthy, with Duan et al. (2021) [15] introducing an efficient NR objective sharpness assessment metric designed for images with shallow depth of field, a common characteristic in portraits and photos emphasizing human subjects. This metric, which calculates sharpness based on bidirectional pixel intensity differences, addresses the limitations of traditional sharpness assessment tools when applied to such images.

Research by Her and Yang (2019) [16] on image sharpness assessment algorithms for autofocus systems further exemplifies the field's evolution. Their work evaluates the performance of several spatial domain functions, highlighting the scene adaptability and anti-jamming capabilities of the Benner algorithm and the sensitivity of the Laplace algorithm, among others. This research underscores the critical role of sharpness evaluation functions in enhancing the quality of images captured by various imaging systems.

The development of advanced artificial intelligence systems, as explored by Kyrychenko, Tereshchenko, Proniuk, and Geseleva (2023) [17], through the use of predicate clustering methods, presents potential avenues for refining image sharpness evaluation techniques

Moreover, advancements in image quality assessment for zoom photos, as investigated by Han et al. (2023) [18], reveal the challenges posed by small sensor sizes and fixed focal lengths in smartphones. Their novel no-reference zoom quality metric incorporates traditional sharpness estimation with image naturalness concepts, demonstrating significant improvements in assessing image quality over traditional metrics.

### 3. Methods

#### 3.1. Existing methods

The Fast Fourier Transform is a convenient mathematical algorithm for computing the Discrete Fourier Transform (DFT). It is used for converting a signal from one domain into another [19].

The FFT is used in different areas, such as mathematics, music, engineering, etc. This method is widely used, as sometimes calculations are much easier performed when the time-series signals are converted into the frequency domain. This method can also be used to convert the frequency domain back to the original format. When talking about FFT in image processing and computer vision, it is important to note that the image is represented in the Fourier and Spatial domains. So, the image is represented in both imaginary and real components.

The obtained values can be analysed to perform blurring or blur detection, edge detection, analysis of textures, etc.

There is a sampled Fourier Transform, which is called DFT. It contains only the set of image samples which are enough to fully represent the spatial domain image [20] (which is often used for further quality metrics extraction).

Given an image with size  $N*N$ , the resulting DFT matrix can be defined as follows:

$$EF(k, l) = \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} f(i, j) e^{-i2\pi\left(\frac{ki}{N} + \frac{lj}{N}\right)} \quad (1)$$

where  $f(a,b)$  is the image in the spatial domain and the exponential term is the basis function corresponding to each; point  $F(k,l)$  in the Fourier space.

The equation can be interpreted as: the value of each point  $F(k,l)$  is obtained by multiplying the spatial image with the corresponding base function and summing the result. At the base level functions, operations are represented as sine and cosine waves which have increasing frequencies. For example,  $F(0,0)$  represents the DC-component of the image (average image brightness) and  $F(N-1,N-1)$  is the highest frequency of the image.

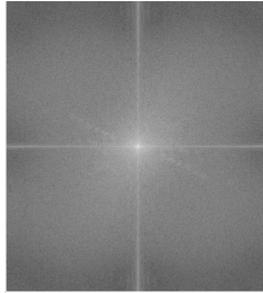
The ordinary one-dimensional DFT has  $N^2$  complexity. If the Fast Fourier Transform (FFT) is used, the complexity can be reduced to  $N\log_2 N$ . For computing large images this improvement is crucial. However, some forms of the FFT may restrict the maximum size of the input to  $N=2^n$ .

The result is an output image represented with complex numbers. This image can be displayed in two states: either with the real and imaginary part or with magnitude and phase (see Fig. 2).

When solving problems in the area of image processing, commonly only the magnitude of the Fourier Transform is displayed. The example image of the result of such transformation is illustrated above.

For clear and reliable results, a contrast detection threshold must be calculated beforehand [21] in order to understand if the image has enough contrast for further analysis. In case of contrast availability and FFT algorithm completion, a floating point value of the mean of the

magnitude indicates the relative sharpness of the whole image. Of course, since this value is relative, conclusions cannot be made without a reference sharpness value.



**Figure 2:** The result of calculating and displaying the Fourier Transform magnitude of an input image

Another option is to use the variance of the Laplacian.

The Laplacian of a function  $f$  at a point  $p$  is (up to a factor) the rate at which the average value of  $f$  over spheres centered at  $p$  deviates from  $f(p)$  as the radius of the sphere shrinks towards 0.

The Laplacian operator is defined as the divergence of the gradient of function  $f$ , as shown in formula (2).

$$\Delta f(x, y) = \text{div}(\text{grad}(f)) \quad (2)$$

In this definition, the gradient is the slope of steepest accent and it gives information about the point and direction of the highest accent in local maxima and, likewise, the local minima.

In case of image sharpness detection, divergence is the vector field associated with blurriness from subject motion or natural background blurriness. The calculated matrix of Laplacian is demonstrated on Figure 3.

0	1	0
1	-4	1
0	1	0

**Figure 3:** The Laplacian kernel for convolution with the source image

Using the process of convolution with the source image, it is being transformed based on the Laplacian kernel. This is used to find areas of rapid source changes (but this works if there is no noise in the image [22]).

Using the OpenCV library this method will be tested on sample images which contain both blurry areas and sharp subjects. This partial blurriness was caused by a very open aperture, which made a very distinct background separation.

```
var imageMat = Highgui.imread(image.absolutePath, Highgui.CV_LOAD_IMAGE_GRAYSCALE)
val destination = Mat()
Imgproc.Laplacian(imageMat, destination)
val median = MatOfDouble()
val std = MatOfDouble()
Core.meanStdDev(destination, median, std)
val variance = Math.pow(std.get(0, 0).get(0), 2.0)
```

To get a single floating point number representing the overall sharpness of a photo it is first needed to load the source image in grayscale, after which the basic Laplacian function should be called.

### 3.2. The alternative algorithm

Looking at the results of the two popular algorithms it can be stated that neither one is optimised enough for the problem stated at the beginning. During photo sessions most portrait photos will be taken with an open aperture which will result in the background having many areas with low frequencies [23]. If the area with naturally low frequencies is greater than the area with high frequencies it will result in the image being labelled as "not sharp", which is not necessarily

true. The question arises – how to make judgements on the photo subject sharpness if analysing the whole image is not optimal?

As a top level solution, the sharpness detection algorithm should be modified in such a way, so that only areas that are supposed to be in focus are evaluated and the background/foreground is ignored. This is the general principle of evaluating visible errors in specific areas of the photo [24]. An example of such approach can be found in the framework for measuring sharpness in natural images [25]. Also, there is research about image quality assessment based on regions of interest, which are identified as features which are highly spatially nonstationary [26].

The solution will be developed based on the limitation that the subjects that need to be in focus are human faces. This limitation, however, can be eliminated by modifying the algorithm and providing support for more subjects.

The following list describes the steps of the algorithm:

1. Find the coordinates and boxing boundaries of every face in the frame
2. Extract the focus distance from the photo
3. Calculate the distance to every face in the frame
4. Calculate the ideal depth of field
5. Select the faces that are within the intended focus plane (focus distance +/- ideal depth of field)
6. Apply the sharpness detector only for the boxes containing the selected faces
7. Calculate the average sharpness score based on the individual sharpness scores

Next, the steps of the algorithm will be described in more detail with the propositions and variants for implementation.

### **3.2.1. Finding the boundaries and coordinates of faces in the frame**

There are quite a few methods and options when it comes to face detection in images. Given an arbitrary image, the goal of face detection is to determine whether or not there are any faces in the image and, if present, return the image location and extent of each face.

Some methods for detecting faces include:

- Knowledge-Based Top-Down Methods;
- Feature invariant approaches;
- Template matching methods;
- Appearance-based methods.

Knowledge-based methods are developed with the scientific knowledge about human faces as the primary source of information. The problem with this approach is the difficulty in translating human knowledge into well-defined rules. It is very hard to find a perfect balance of strictness in these rules. If the rules are defined too strictly – the method may fail to detect faces that do not pass all rules at once. On the other hand, if the rules are too general – the method may result in false positives.

The feature-based methods are based on the fact that humans have the natural ability to recognize a face in different lighting conditions, under different angles and circumstances, meaning we are trained to detect certain features. This method builds on this fact and many variants were proposed, when first the features are extracted and analysed. One problem with these feature-based algorithms is that the image features can be severely corrupted due to illumination, noise, and occlusion.

In template matching methods a face pattern is manually predefined. When analyzing an input image, the correlation is computed for separate parts of the presumable face: the contour, the nose, eyes, mouth. The result is positive, if the mean correlation value of these components is above certain threshold. This method, however, is not ideal due to the lack of flexibility when it comes to different face shapes, poses and scale.

Lastly, the appearance-based methods are more close-to-life, than the previous ones. The templates are not generated by experts (like in template matching), rather taken from samples of actual image databases. This method relies heavily on statistical analysis and machine learning.

Appearance-based methods have many different implementations, starting like distribution-based, support vector machines, hidden Markov model, cascade classifiers, cascaded convolutional networks [27], etc.

Taking into account the advantages and disadvantages of the abovementioned face detection methods, it was decided that appearance-based methods are well suited for the task. When implementing the improved sharpness detection algorithm, cascade classifiers can be used. OpenCV contains pre-trained open classifiers that can be freely downloaded. To retrieve face coordinates using the OpenCV library one can use the CascadeClassifier.detectMultiScale function. The arguments for this function are: the input image (in gray scale), scaleFactor and minNeighbours. The scaleFactor specifies how much the image size is reduced with each scale. minNeighbours specifies how many neighbors each candidate rectangle should have to retain it. These parameters will be fine-tuned during the development process.

Of course, everything that involves image processing and subject recognition will always take up some valuable processing time, so there is another way of extracting the faces in the frame. The other method relies heavily on the camera's integrated ability to detect faces in real time and write this information to the metadata. Of course, the algorithm should not be strictly reliant on using this optional metadata, but definitely should utilize it if available – that would dramatically improve the performance of the method.

As a result of image scanning with classifiers or face extraction using metadata, the coordinates and boundaries of all faces in the frame will be retrieved for further analysis.

### 3.2.2. Extracting the focus distance and other useful parameters from the photo

Focusing plane is the image sensor of the camera. Focusing distance is the distance from the focusing plane to the subject.

When using automatic lenses (lenses with autofocus, that do not require manual input to put the subject in critical sharp focus) there are complex algorithms running in order to produce optimal results in terms of focus. The camera constantly monitors the image, understands the distance to the subject in real time and adjusts the focus motors accordingly. To achieve the autofocus, the camera lens moves to a position where the clearest image is obtained. The maximum clarity is measured from the histograms of the images on which a filter with the role of highlighting the edges was initially applied. Knowing the focal length of the lens, the distances can be found from the lens equation [28]. So, we can see that this is clearly possible and cameras perform these calculations all the time. Moreover, cameras can utilize different focusing algorithms [29] for accuracy. The distance that the focus motor travelled in the lens before the photo was taken corresponds to the estimated focusing distance which is needed for the modified sharpness detection algorithm.

It is possible to retrieve the calculated focusing distance by looking at the photo metadata. EXIF (Exchangeable Image File Format) is a standard that allows adding information (metadata) to photos and videos. This format is quite flexible, meaning that all users can modify it by adding new data entries with original names. This means that not all camera bodies from different manufacturers produce the same metadata. Having that said, many cameras contain valuable information that can be used for the purpose of this research. By viewing the metadata using an EXIF reader we can find the following relevant information (Table 1).

**Table 1**  
**Partial, most relevant information extracted from the test photo metadata**

Parameter Name	Value
Lens Spec	FE 70-200mm F4 G OSS
Min Focal Length	70.0 mm
Max Focal Length	200.0 mm
Focal Length	159.0 mm (35 mm equivalent: 159.0 mm)

Focus Location	6000 4000 2709 1291
Lens Zoom Position	76%
Aperture	4.0
Focus Distance 2	3.46 m
Circle Of Confusion	0.030 mm
Field Of View	12.9 deg
Hyperfocal Distance	210.35 m
Faces Detected	3
Face 1 Position	789 2344 1256 1256
Face 2 Position	1521 3506 1575 1575
Face 3 Position	413 1688 844 844
Circle Of Confusion	0.030 mm

---

The parameter "Focus Distance 2" implies the calculated distance (in meters) to the focused subject. This calculation was performed by the camera by analysing the rotation angle of the focus motor when the edges of the image were in focus. This parameter can be used as a ready solution and will be very useful for further calculations. There is no other reliable way of retrieving/calculating the focus distance of a photo, especially when the photo is not guaranteed to be in critical focus initially. This is why the method heavily relies on this meta parameter.

Some other useful parameters include the number of faces detected and the detected faces positions. This is a "luxury" to have these parameters in the metadata and only modern cameras provide such information. Heavily relying on the presence of these parameters would limit the method to work only for modern high-tech camera bodies. So, the method will utilise these shortcuts in metadata if they are present, but still have the ability to detect faces' boundaries described in the section above.

### 3.2.3. Calculating the distance to every face in the frame

It is required to find distances to all faces in the frame because the method needs to determine which faces in particular need to be in focus. The "need to be in focus" criteria is described in the next point.

At this step the focus distance, as well as all face boundaries (and thus, sizes) are already known. To find the distance to the human face on the photo it is necessary to first estimate the height of the faces in the real life.

To find the real-world height of each face in the frame, two methods can be used:

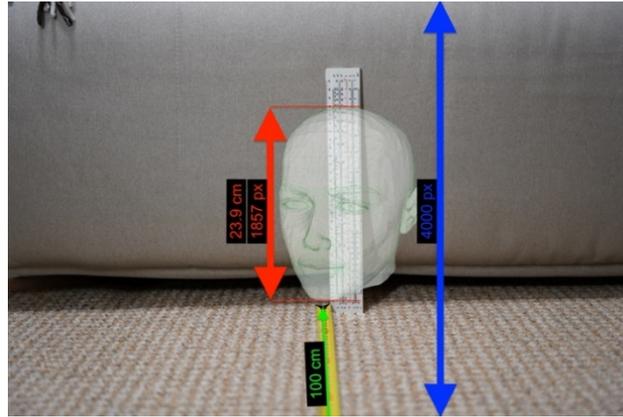
- the average statistical face height value (21.8cm -23.9 cm);
- calculated from a proportionality formula based on the distance between the eyes and the mouth.

The second option is very complicated, involves extra operations like edge detection to find the location of eyes and mouth, depends on the pose (will not work for side portraits) and does not guarantee accuracy greatly more than the first option. This is why the first option of assuming that the height of the face is somewhere between 21.8cm and 23.9cm will be used during the implementation.

When the real world height of all faces in the frame is known, the real world distance to the faces can be determined in two ways:

- predefined proportions at 1meter distance method (see Figure 4, photo by Dolhanenko O.);
- using an alleged "focused face" from the focus point coordinates as a reference.

The first method can be described as follows: practically or mathematically find the percentage height (from the frame height) of a 23.9 cm high object at a distance of 1 meter at a focal length of 50mm (which gives 1.0 times magnification) on a 35mm full frame sensor.



**Figure 4:** The result of measuring a 23.9 cm high object at a distance of 1 meter with specific settings

As seen from the photo, the height of the face on the image is 1857px (46.4%) on a horizontal photo with 4000\*6000 resolution. On a vertical photo of the same resolution, the facial height is 30.95% of the full height. This gives a point of reference which can be proportionally scaled based on the given (non-full frame) sensor size or/and focal length.

To convert the reference measurement to a different focal length on a different format camera sensor, the following formula can be used (see formula 3):

$$h = H * \frac{f}{50}, \quad (3)$$

where h is the new reference height of the face, f is the 35mm equivalent focal length of the selected lens, H is the initial reference height of the face (taken via 50mm full frame lens from 1 meter).

When the reference height of the 23.9 cm face is calculated taking into account the sensor format and the focal length, the distance D to a particular face in the frame can be calculated by comparing the height of the face to the reference height and determining proportionally the distance to the sensor (see formula 4).

$$D = \frac{h}{g} \quad (4)$$

where e D is the distance (in meters) to the face in the frame, g is the height of the face of interest in pixels, h is the reference height of the face.

The method described above is simplified. In future work the accuracy can be improved by adding support for lens distortions and other factors that may influence the measurements.

The second method heavily relies on metadata. The coordinates of the focus point in the meta data indicates the precise area of focus interest on the photo. Since the coordinates and boxes of all faces are available at this point, the face within the area of interest can be selected as the reference. Since the distance to this face is already available, once again, from the metadata, we can calculate the proportional distance to other faces in the frame by using the (4) formula.

The second method can be reliable if the image does not have stacked, slightly shifted faces, but since this is not a guarantee, the first method can be preferred. However, if the conditions are ideal, the second method will produce more precise results, since it uses more non-estimated measurements. There are other methods of object distance extraction by using reference targets [30], however, such methods are not suitable for general photography.

### 3.2.4. Calculating the ideal depth of field

Depth of field (DOF) is the distance between the nearest and the farthest objects that are in acceptably sharp focus in an image.

To understand whether or not some subject at a specific distance is supposed to be sharp in the photo (based on the camera and lens settings) it is required to first calculate the depth of field

using the captured camera and lens settings. The formulas for calculating the distance to the front plane of the focus area and the distance to the back plane of the focus area are as shown in formula (5):

$$R_1 = \frac{R * f^2}{f^2 - K * f * z + K * R * z}; \quad (5)$$

$$R_2 = \frac{R * f^2}{f^2 + K * f * z - K * R * z}''$$

where  $R_1$  is distance to the front edge of the critical focus plane,  $R$  is distance of focus (can be retrieved from metadata),  $R_2$  is distance to the back edge of the critical focus plane,  $f$  is the focal length of the lens in meters (can be retrieved from metadata),  $K$  is the f-stop of the lens (can be retrieved from metadata),  $Z$  is the Circle Of Confusion (can be retrieved from metadata).

By subtracting  $R_1$  from  $R_2$  the depth of field can be found. However, it will be easier to use the raw  $R_1$  and  $R_2$  values for further calculations.

### 3.2.5. Applying sharpness detection only for faces required to be in focus

To understand which faces to check for sharpness it is first required to understand the criteria of expected portrait framing depth (EPFD).

The EPFD is a generalized assumption of the maximum intended distance between the subjects when taking photos of multiple rows of people at once. For example, a photographer is taking a group portrait photo with 10 people, which are placed in two rows. In this case, the expected portrait framing depth is within 0.8 meters, since there are two rows and each one is about 0.3 meters wide +- a margin in between. For such portrait framing depth, the photographer needs to set a particular aperture (f-stop), so that the depth of field is equal to or more than the framing depth.

The EPFD is a very subjective parameter and cannot be calculated from analyzing the photo. That is because it is infinitely hard for a computer to determine if the photo of a group of people is intended to be shot with an open aperture (to have only front subjects in focus), or it was intended to have both rows of people in focus and the low aperture value was chosen by mistake.

The subjective nature of this parameter leaves this part of the method to be fine-tuned by the end user, selecting one of two options on a collection of photos before the sharpness detection: single/couple styled photos or intended multi-row group photos.

For the second option, where a group photo (composed in two or more rows) is selected, the faces can be classified as "intended to be sharp" by the following sequence of formulas:

Given the measured light value (EV) from the metadata, it is required to calculate the maximum aperture, using which the EV value would be the same. The bigger the aperture value, the greater the depth of field is, meaning more faces need to be in critical sharp focus. The photographer could make the mistake of setting a low aperture, which results in only one row of people being in focus. This is why the maximum aperture value needs to be determined and the photo needs to be analysed as if these "ideal" settings were set in camera.

$$EV = \log_2 \left( \frac{100 * K^2}{I * S} \right), \quad (6)$$

where EV is the exposure value;  $K$  is the f-stop value,  $I$  is the ISO.

Given the maximum acceptable ISO value by the photographer (the greater the ISO the more noise there is in the photo, thus it is down to personal photographer preference), the maximum value of the aperture to achieve the same light value can be determined as follows:

$$K_{max} = \sqrt{\frac{2^{EV} * I_{max} * S_{max}}{100}} \quad (7)$$

where  $K_{max}$  is the maximum f-stop number for the pre-defined EV and ISO,  $I_{max}$  is the maximum ISO value, acceptable by the photographer,  $S_{max}$  is the maximum duration of the shutter speed.

When operating a camera, the photographer may or may not use automatic settings. This is why in the formula above the shutter speed is not taken from the metadata, rather calculated as well. The maximum acceptable shutter speed can be calculated as follows:

$$S_{max} = \frac{1}{Fl} \quad (8)$$

where Fl is the focal length used for the photo, taken from the metadata.

However, there is an informal rule, stating, that when shooting portraits, one should not select shutter speeds below 0.008 of a second. This may also be an external fine-tuning setting available for the end user.

Next, the maximum depth of field needs to be calculated. To achieve this, the formulas (5) can be used, inserting the maximum aperture f-stop value ( $K_{max}$ ). As a result,  $R_1$  (distance to the front edge of the critical focus plane) and  $R_2$  (distance to the back edge of the critical focus plane) with the "ideal" camera settings can be found.

Given the calculated distance to the subjects face, the criteria of in focus intention can be formulated as follows: the face of a subject was intended to be in critical sharp focus, if the distance D satisfies the condition:

$$D \leq R_2 \quad (9)$$

### 3.2.6. Sharpness detection only for regions containing the selected faces

At this stage of the algorithm, the most valuable information is already achieved – the understanding of which faces in the frame were most likely intended to be in focus. Having this information, the sharpness detection algorithm of choice can be applied these regions exclusively. This will result in very accurate results, as the background and other subjects will not be evaluated.

It is worth mentioning, that having this information, a whole window of possibilities for photo categorizing opens up. Not only the sharpness can be evaluated, but also the open state of the eyes (open/half-open/closed), preferred facial expressions of focused subjects, even the poses can be classified as appealing and not.

## 4. Experiment

The purpose of the following experiment is to demonstrate relations of the Laplacian and FFT sharpness evaluation algorithms to different factors of the photo. Revealing these dependencies on factors will lead to conclusions about the steps needed to be taken in order to improve the effectiveness of the algorithms.

In this experiment the following set of images will be used:

1. A reference representation of a completely blurry photo
2. A reference representation of a very sharp photo in all parts (no background blurriness)
3. A photos which contain sharp and blurred parts due to the aperture effect

The first and second images are used to determine the dramatic maximum and minimum sharpness results both algorithms provide.

All other 7 test photos contain human subjects and should be the same resolution and have identical lighting conditions for optimal evaluation results. There should not be any colour or sharpness corrections done before the experiment. All photos from the experiment are similar to figure 1 and contain same subjects and same scenes but in different configurations.

The following Table 2 verbally describes the photos that were used for the experiment.

The first photo was made with a very closed aperture (f22) and has extremely "busy" foreground, which leads to extreme sharpness results by the Laplacian method. The second photo is fully out-of-focus and results in very low value results by both algorithms. No conclusions can be made from these results yet.

**Table 2**  
**Description of the test images**

Image name	Verbal description
1.JPG	A sharp photo containing one in-focus subject which takes up 24.7% of the frame
2.JPG	A sharp photo containing one in-focus subject which takes up 37.5% of the frame
3.JPG	A sharp photo containing two in-focus subjects, high aperture (background blurriness is high), collectively taking up 38.38% of the frame.
4.JPG	A sharp photo containing one in-focus and one out-of-focus subject, collectively taking up 36.09% of the frame
5.JPG	A sharp photo containing two in-focus subjects, low aperture (background blurriness is low), taking up 33.13% of the frame
6.JPG	A sharp photo containing one in-focus and one in-motion subject, collectively taking up 37.87% of the frame
7.JPG	A sharp photo containing one in-focus and one out-of-focus subject, collectively taking up 57.24% of the frame
8.JPG	A very "busy" and sharp photo, shot with a closed aperture (f22)
9.JPG	A completely blurry, out-of-focus photo of the same scene as in 8.JPG

The following Table 3 represents the results of the main 7 photos analysis (provided in the Laplacian, FFT columns), alongside additional calculated parameters that will be needed for dependencies analysis. It should be stated, that the Laplacian and FFT methods produce values that do not relate to each other and thus should not be compared directly.

**Table 3**  
**Parameters and results of the experiment with the main photos**

Image Name	Laplacian	FFT	Background quantity (%)	Total subjects size from photo (%)	ff-stop value
1.JPG	33.84	20.89	75.22	24.78	4
2.JPG	77.03	13.39	62.475	37.525	4
3.JPG	60.63	40.95	61.62	38.38	20
4.JPG	23.97	25.53	63.91	36.09	4
5.JPG	36.76	35.02	66.87	33.13	4
6.JPG	44.06	32.12	62.129	37.871	4
7.JPG	39.57	21.24	42.76	57.24	2.8
8.JPG	647.4	64.066	7.1	92.9	22
9.JPG	3.71	-6.80	100	0	2.8

The "background quantity" is the percentage value of how much of the photo is taken up by the background. The "total subject size" from photo is the opposite value, describing how much area of a photo is taken up by the subjects.

The next phase of the experiment includes utilizing the parts of the improved algorithm that was described in the previous section. Within this experiment the dependencies which lead to unstable results were determined. One of the main parts of the described algorithm is face detection. The following test was conducted on a set of images that were previously cut to contain only faces. Images that contained two subjects are represented by two individual cut images.

The experiment conditions and results are displayed in table 3. The test images are extracted faces from the frame (see Figure 5, photos by O. Dolhanenko) and the test results are stated in Table 4.

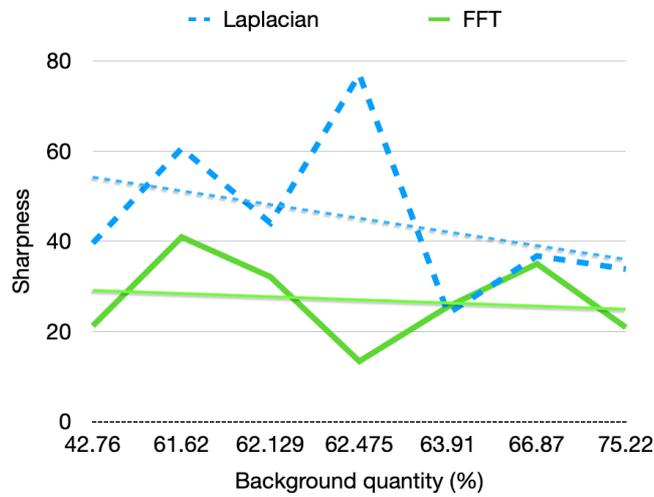


**Figure 5:** Some examples of the test images used for the experiment

The methods were implemented using Kotlin and launched on the JVM, running on MacOS with Core-I9 CPU

## 5. Results

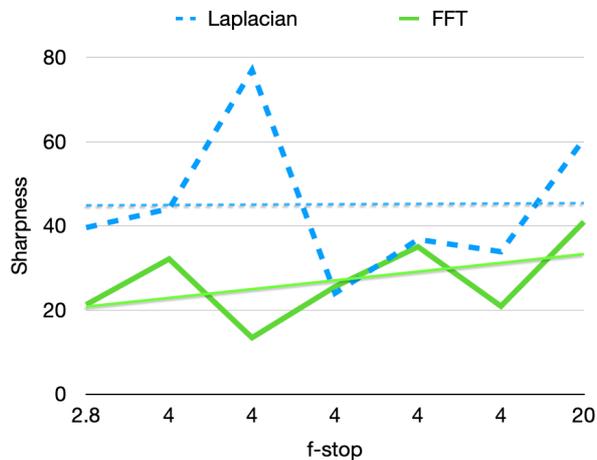
By plotting the sharpness result values obtained during the experiment against the background quantity, the following chart can be achieved (see Figure 6)



**Figure 6:** Results of both algorithms plotted against the background quantity

As seen from the chart, a clear linear trendline describes the relation of the quantity of the background to the calculated sharpness (both in Laplacian and FFT methods). This means that the more background there is visible on a photo (even if the subject is in critical sharp focus), the less likely the photo will be classified as sharp.

By plotting the sharpness result values obtained during the experiment against the f-stop values (see Figure 7), the following chart can be achieved.



**Figure 7:** Results of both algorithms plotted against the background quantity

The relation on this chart is not as obvious as the previous one, since the data set is rather small, but even here a visible linear trendline for both Laplacian and DFT methods is noticeable. This trend implies that the higher the f-stop value, the more likely the photo will be classified as sharp.

Having analyzed both relations it can be stated, that the result of DFT and Laplacian methods are dependent on the scene and camera settings, which means that they are not reliable for generic automatic photo filtering. Furthermore, comparing the average Laplacian sharpness value for subject photos (45.12) to the average sharpness value for the 2 reference photos (325.54) reveals a major difference. This difference demonstrates the source of errors that can appear by selecting an incorrect reference value.

**Table 4**  
**Results of the experiment for face-only evaluation**

Image(s) name	Expectation (1-sharp 0-blurry)	Laplacian	Laplacian avg	Laplacian normalized	Laplacian time (ms)	FFT	FFT avg	FFT normalized	FFT time (ms)
10-2.JPG	0	7.1056	23.2832	0.0898	7	-4.2154	7.1930	0.2517	132
10.JPG		39.4608			10	18.6013			155
4-2.JPG	0.2	7.0557	78.0437	0.3010	12	2.1674	12.1821	0.4263	233
4.JPG		149.0316			10	22.1968			96
3-2.JPG	0.3	45.4098	92.2698	0.3559	5	13.8346	18.2867	0.6400	75
3.JPG		139.1298			13	22.7388			201
7-2.JPG	0.4	6.5819	72.2417	0.2786	18	9.8695	18.1252	0.6343	353
7.JPG		137.9015			40	26.3808			589
6-2.JPG	0.4	373.5180	203.0986	0.7834	11	26.3782	19.9107	0.6968	104
6.JPG		32.6791			7	13.4432			102
5-2.JPG	0.7	80.4655	213.9522	0.8252	6	17.0262	21.8001	0.7630	50
5.JPG		347.4389			8	26.5740			133
1.JPG	0.8	221.8970	221.8970	0.8559	52	26.2262	26.2262	0.9179	277
2.JPG	1	259.2589	259.2589	1.0000	49	28.5732	28.5732	1.0000	879

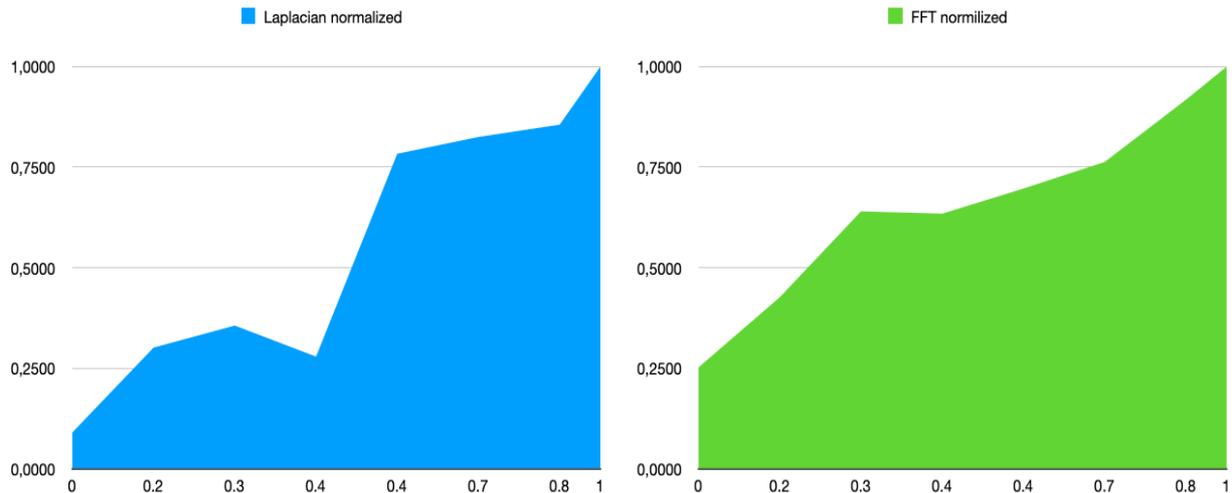
The "Image(s) name" is the name of the analyzed photo. If the photo contained multiple subjects, it was split into separate images and visually merged in the table.

The "Expectation" is a non-bias subjective rating of image quality and usability given by the photographer, where 0 is unusable (blurry) and 1 – usable (sharp). The "Laplacian", "FFT" are actual results of sharpness evaluation. The "Laplacian normalized" and "FFT normalized" are conversions of actual results to a 0-1 scale, where 1 is the maximum value taken from the algorithm output. This way the normalized values of different algorithms are relatable to the "Expectation" values. The "time" columns represent the time taken to process the image in milliseconds.

After the data was collected, the normalized columns were analyzed for visual trends. To fulfil the first objective of the experiment, the sorted input values trend (the expectation) must meet the trend of the sorted normalized output values. By viewing the plotted results (see Figure 8) it can be stated, that both algorithms generally fulfil this requirement.

Only one data point is either unnaturally classified as sharp (by both algorithms), or the subjective usability parameter was defined not accurately.

The second objective of the experiment was to determine which algorithm is more stable and should be used in the actual implementation. Comparing two charts (see Figure 8) a more linear trend line is received from the FFT algorithm, rather than the Laplacian. Moreover, the first method produces very unbalanced results.



**Figure 8:** The result of the second experiment with sharpness expectations

The third and final objective was to determine which algorithm works faster. By comparing the average speed of both algorithms – 17.7ms for Laplacian and 241.25ms for the FFT, it can be concluded that the FFT method is 14 times slower.

## 6. Discussions

Two methods were analyzed and compared – FFT (Fast Fourier Transform) and the Variance of Laplacian for image sharpness evaluation. During the first experiment it was determined that both algorithms produce results that are highly dependent on the photo scene, composition and camera settings – photos with large quantities of background (produced by a low f-stop value) were classified as blurry.

An improved algorithm was proposed which is based on subject detection [31, 32] and further sharpness evaluation exclusively of the subject box. This way all dependencies were eliminated and more natural evaluation results were obtained.

The improved algorithm uses metadata of the photo to simplify calculations, if the relevant data "shortcuts" are available. In some cases, the metadata may contain the detected faces coordinates, which greatly optimize the performance of the algorithm. The method is built on the principle of data analysis and does not require much user input to function properly. The only parameter that cannot be achieved automatically is the intended photography style – whether the user was shooting groups of people to get all in focus, or it was intended to focus on only one-two subjects in the frame, while isolating the rest of the background. Another parameter that the user may input is the maximum acceptable ISO value. This parameter has a default value but is very subjective and so depends on the user preference.

The goals of the experiment were to identify whether the improved algorithm results meet the expectation and to determine which of the sharpness evaluation methods work best for the task. As a result it was found that the improved algorithm meets the expectations and works best with the FFT sharpness evaluation method as it produces more linear trending results. The Laplacian method, however, is 14 times faster and may be preferred for very large datasets, where accuracy is not as important as time efficiency.

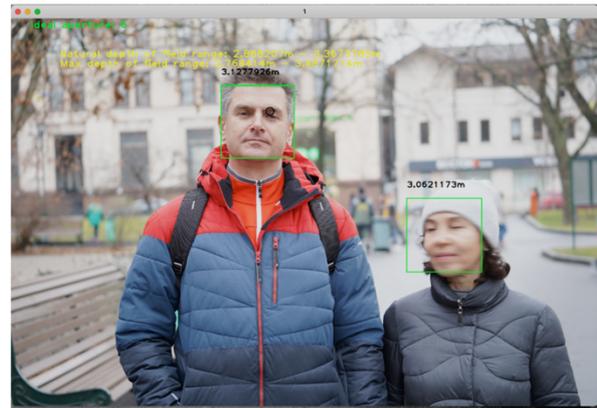
The results of the research were used to implement a software solution prototype for automated image files sorting by subject sharpness (see Figure 9).

The resulting software was configured to operate with JPG files, built using primarily Kotlin and OpenCV, deployed and run on a MacOS environment.

DSC05740.JPG	28 November 2021, 13:13
DSC05741.JPG	28 November 2021, 13:14
DSC05742.JPG	28 November 2021, 13:14
DSC05743.JPG	28 November 2021, 13:14
DSC05744.JPG	28 November 2021, 13:14
DSC05745.JPG	28 November 2021, 13:14
DSC05746.JPG	28 November 2021, 13:15
DSC05747.JPG	28 November 2021, 13:15
DSC05748.JPG	28 November 2021, 13:15

The results arrived. Processing took:5892ms  
Ranked by DFT (sharpest first):

[/Users/admin/Desktop/dataset/DSC05737.JPG](#) : 109.05292862986634,  
[/Users/admin/Desktop/dataset/DSC05738.JPG](#) : 106.11267161722266,  
[/Users/admin/Desktop/dataset/DSC05740.JPG](#) : 104.13476823248199,  
[/Users/admin/Desktop/dataset/DSC05743.JPG](#) : 102.62237987471754,  
[/Users/admin/Desktop/dataset/DSC05741.JPG](#) : 101.18909463781522,



**Figure 9:** A software solution for automatic sorting based on subject sharpness

## 7. Conclusions

The purpose of this research was to analyze the existing methods of image sharpness evaluation and suggest improvements for the use case of sharpness evaluation of photos of people. The goal end result is an automated photo sorting solution, which classifies images from sharp and usable to blurry and unusable.

The comparison of FFT and Variance of Laplacian methods for image sharpness evaluation showed that their accuracy is affected by the photo's scene, composition, and camera settings, leading to misclassification of detailed background photos as blurry.

An enhanced algorithm leveraging subject detection for focused sharpness evaluation showed significant improvements by removing biases and yielding more accurate results. It utilizes photo metadata to streamline processes, especially when such metadata includes coordinates of detected faces, enhancing performance efficiency. The method minimizes user input, requiring only the photography style and an optional maximum acceptable ISO value to adjust for personal preference. Final testing confirmed the algorithm's effectiveness, particularly in conjunction with the FFT method for its linear results, though the faster Laplacian method may be chosen for large datasets where speed trumps precision.

The results of the research and experiments have been used to implement a prototype of an automation system for sorting photos by subject sharpness, detecting unwanted motion within the image, subject blurriness.

## References

- [1] N. Wadhwa, Synthetic depth-of-field with a single-camera mobile phone, *ACM Transactions on Graphics* 37.4 (2018) 64. doi:10.1145/3197517.3201329.
- [2] M.-J. Chen, A. C. Bovik, No-reference image blur assessment using multiscale gradient, *J Image Video Proc.* 2011, 3 (2011). doi:10.1186/1687-5281-2011-3.
- [3] H. R. Sheikh, Image information and visual quality, *IEEE Transactions on Image Processing.* 15.2 (2006): 430-444. doi: 10.1109/tip.2005.859378.
- [4] W. Zhang, K. Ma, J. Yan, D. Deng and Z. Wang, Blind Image Quality Assessment Using a Deep Bilinear Convolutional Neural Network, *IEEE Transactions on Circuits and Systems for Video Technology* 30.1 (2020) 36-47, 2020. doi:10.1109/TCSVT.2018.2886771
- [5] S. Chattopadhyay, Image Sharpness Assessment Based On local Phase Coherence. *NeuroQuantology* 5.20 (2022). doi: 0.48047/nq.2022.20.5.nq22800.
- [6] R. Hassen, Z. Wang and M. M. A. Salama, Image Sharpness Assessment Based on Local Phase Coherence, in *IEEE Transactions on Image Processing*, 22.7 (2013), pp. 2798-2810, doi:10.1109/TIP.2013.2251643.

- [7] Z. Liu, H. Hong, Z. Gan, J. Wang, Y. Chen, An Improved Method for Evaluating Image Sharpness Based on Edge Information, *Applied Sciences* 12.13 (2022):6712. doi:10.3390/app12136712
- [8] Y. Zhang et al., Image sharpness evaluation method based on normal gradient feature, in *Proceeding of the 3rd International Symposium on Robotics & Intelligent Manufacturing Technology (ISRIMT)*, Changzhou, China, 2021, pp. 308-314, doi:10.1109/ISRIMT53730.2021.9596808.
- [9] Yan, X.Y., Lei, J., Zhao, Z.: Multidirectional gradient neighborhood-weighted image sharpness evaluation algorithm. *Math. Problems Engin.* 7864024 (2020). doi: 10.1155/2020/7864024
- [10] M. Zhu, L. Yu, Z.Wang, Z. Ke, C. Zhi, Review: A Survey on Objective Evaluation of Image Sharpness, *Applied Sciences* 13.4 (2023): 2652. doi:10.3390/app13042652.
- [11] S. Bielietsov, I. Ruban, K. Smelyakov and D. Sumtsov, Network technology for transmission of visual information, *Selected Papers of the XVIII International Scientific and Practical Conference "Information Technologies and Security" (ITS 2018)*, Kyiv, Ukraine, November 27, 2018. In *CEUR Workshop Proceedings, Vol-2318, 2018*, pp. 160-175.
- [12] K. Smelyakov, A. Chupryna, D. Sandrkin and M. Kolisnyk, Search by Image Engine for Big Data Warehouse, in: *Proceedings of the 2020 IEEE Open Conference of Electrical, Electronic and Information Sciences (eStream)*, Vilnius, Lithuania, 2020, pp. 1-4, doi:10.1109/eStream50540.2020.9108782.
- [13] K. Smelyakov, D. Karachevtsev, D. Kulemza, Y. Samoilenko, O. Patlan and A. Chupryna, Effectiveness of Preprocessing Algorithms for Natural Language Processing Applications, in: *Proceedings of the 2020 IEEE International Conference on Problems of Infocommunications. Science and Technology (PIC S&T)*, Kharkiv, Ukraine, 2020, pp. 187-191, doi:10.1109/PICST51311.2020.9467919.
- [14] Jiye Qian, Hengjun Zhao, Jin Fu, Wei Song, Jide Qian, Q. Xiao. "No-reference image sharpness assessment via difference quotients." *Journal of Electronic Imaging*, 28.1 (2019). doi: 10.1117/1.JEI.28.1.013032.
- [15] Z. Duan, G. Li and G. Fan, An Effective Sharpness Assessment Method For Shallow Depth-Of-Field Images, in *Proceeding of the IEEE International Conference on Image Processing (ICIP)'21*, Anchorage, AK, USA, 2021, pp. 1449-1453, doi:10.1109/ICIP42928.2021.9506498
- [16] L. Her and X. Yang, Research of Image Sharpness Assessment Algorithm for Autofocus, in *Proceeding of the 2019 IEEE 4th International Conference on Image, Vision and Computing (ICIVC)*, Xiamen, China, 2019, pp. 93-98, doi: 10.1109/ICIVC47709.2019.8980980.
- [17] I. Kyrychenko, G. Tereshchenko, G. Proniuk, N. Geseleva, Predicate Clustering Method and its Application in the System of Artificial Intelligence, in: *Proceedings of the 7th International Conference on Computational Linguistics and Intelligent Systems (COLINS-2023)*, 3396 (2023), pp. 395 - 406.
- [18] Z. Han, Y. Liu, R. Xie, G. Zhai, Image Quality Assessment for Realistic Zoom Photos. *Sensors* 2023 (2023) 4724. doi:10.3390/s23104724.
- [19] T. Sieberth. "Automatic detection of blurred images in UAV image sets." *ISPRS Journal of Photogrammetry and Remote Sensing*. 122 (2016). doi: 10.1016/j.isprsjprs.2016.09.010.
- [20] E. Fry. "Bridging the Gap Between Imaging Performance and Image Quality Measures." *Electronic Imaging*. 12 (2018). doi:10.2352/issn.2470-1173.2018.12.iqsp-231.
- [21] S. Triantaphillidou. "Contrast sensitivity in images of natural." *Signal Processing: Image Communication*. 75 (2019): 67. doi:10.1016/j.image.2019.03.002.
- [22] O. van Zwanenberg, S. Triantaphillidou, R. Jenkin and A. Psarrou, Edge Detection Techniques for Quantifying Spatial Imaging System Performance and Image Quality, in: *Proceedings of the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, Long Beach, CA, USA, 2019 pp. 1871-1879. doi: 10.1109/cvprw.2019.00238.
- [23] P. Marziliano, Perceptual blur and ringing metrics: application to JPEG2000, *Signal Processing: Image Communication*. 19.2 (2004): 167. doi:10.1016/j.image.2003.08.003.
- [24] Z. Wang, Image Quality Assessment: From Error Visibility to Structural, *IEEE Transactions on Image Processing* 13.4 (2004): 601. doi:10.1109/tip.2003.819861.

- [25] M. Nuutinen, A framework for measuring sharpness in natural images captured by digital cameras based on reference image and local areas, *EURASIP Journal on Image and Video Processing* (2012). doi: 10.1186/1687-5281-2012-8.
- [26] W. Yang, Method of image quality assessment based on region of interest, *Journal of Computer Applications* 28.5 (2008): 1310–1312. doi:10.3724/sp.j.1087.2008.01310.
- [27] R. Qi, R. -S. Jia, Q. -C. Mao, H. -M. Sun and L. -Q. Zuo, Face Detection Method Based on Cascaded Convolutional Networks, *IEEE Access*. 7 (2019): pp. 110740-110748. doi:10.1109/access.2019.2934563.
- [28] H.Eugen. Video distance measurement based on focus, in: *Proceedings of the 2020 International Conference and Exposition on Electrical And Power Engineering (EPE)*, Iasi, Romania, 2020, pp. 65. doi:10.1109/epe50722.2020.9305593.
- [29] S. Pertuz, D. Puig, M. A. Garcia., Analysis of focus measure operators for shape-from-focus, *Pattern Recognition* 46.5 (2013): 1415–1432. doi:10.1016/j.patcog.2012.11.011.
- [30] Z. Xu, L. Wang, J. Wang, A Method for Distance Measurement of Moving Objects in a Monocular Image, in: *Proceedings of the 3rd International Conference on Signal and Image Processing (ICSIP)*, Shenzhen, China, 2018, pp. 245-249, doi:10.1109/SIPROCESS.2018.8600495.
- [31] A. Sergiyenko, P. Serhienko, M. Orlova, Local Feature Extraction in Images, *Information, Computing and Intelligent systems* 2 (2021). doi: 10.20535/2708-4930.2.2021.244191.
- [32] Z. Huo, Y. Zhang, H. Liu, J. Wang, X. Liu, J. Zhang. Improved covariant local feature detector, *Pattern Recognition Letters* 135 (2020). doi:10.1016/j.patrec.2020.03.027.