



Performance comparison of different Retinex methods on enhancement of low light cave images

Düşük ışıklı mağara görsellerinin iyileştirilmesinde farklı Retineks yöntemlerinin performans karşılaştırması

Bilgin Yazlık¹ , Egemen Yazlık^{2,*} 

¹ Nevşehir Hacı Bektaş Veli University, Computer Engineering Department, 50400, Nevşehir, Türkiye

² Nevşehir Hacı Bektaş Veli University, Alternative Energy Resources Program, 50400, Nevşehir, Türkiye

Abstract

Caves are spaces that open into the earth and are often either completely dark or receive very little light. Photographs taken inside caves are usually subject to extremely low light conditions and image quality is negatively affected. In this study, state-of-the-art Retinex-based image enhancement methods are used to enhance the quality of images taken in the cave. Enhanced images are evaluated both quantitatively and qualitatively. A cave image dataset consisting of 82 low-light images is used for quantitative analysis. The enhanced results from all methods are visually compared using three sample images from the dataset for qualitative analysis. Eleven different image quality assessment methods are applied for quantitative analysis. The results of the study show that Retinex-based methods are successful in the enhancement of low-light cave images. While BIMEF is recommended for applications requiring general image enhancement and structural consistency, MSRETINEX and LIME can be preferred in applications requiring high dynamic range and detail preservation. In cases where noise reduction and objective quality assessment metrics are at the forefront, SRIE provides advantages.

Keywords: Retinex, cave, Image enhancement, Performance, Compare

1 Introduction

Caves are in-earth cavities formed by natural geological processes or people and are of a suitable size for at least for one human cover [1]. These cavities are a great focus of interest for researchers due to their natural formations and historical importance. However, working in a cave environment is extremely difficult. Caves are generally narrow, irregular in structure, and make movement difficult due to the physical dangers they contain. In addition, natural light is either completely absent or extremely limited in caves. Because of this, photographs taken inside the cave are subject to extremely low light conditions and the quality of the images decrease.

Öz

Mağaralar dünyanın içine açılan, çoğunlukla ya tamamen ışıksız ya da çok az ışık alan mekanlardır. Mağaraların içinde çekilen fotoğraflar genellikle çok düşük ışık koşullarına maruz kalır ve görsel kaliteleri olumsuz etkilenir. Bu çalışmada, mağarada çekilen görüntülerin kalitesini iyileştirmek için güncel Retinex tabanlı görüntü iyileştirme yöntemleri kullanılmıştır. İyileştirilen görüntüler hem nicel hem de nitel olarak değerlendirilmiştir. Nicel analiz için 82 adet düşük ışık görüntüsünden oluşan bir mağara görüntü veri seti kullanılmıştır. Nitel analiz için tüm yöntemlerin iyileştirme sonuçları, veri setinden seçilen üç örnek görüntü ile görsel olarak karşılaştırılmıştır. On bir farklı görüntü kalitesi değerlendirme yöntemi nicel analiz için uygulanmıştır. Çalışmanın sonuçları, Retinex tabanlı yöntemlerin düşük ışıklı mağara görüntülerini iyileştirmede başarılı olduğunu göstermektedir. Genel görüntü iyileştirme ve yapısal tutarlılık gerektiren uygulamalar için BIMEF önerilirken, yüksek dinamik aralık ve detay korunması gereken uygulamalarda MSRETINEX ve LIME tercih edilebilir. Gürültü azaltma ve nesnel kalite değerlendirme metriklerinin ön planda olduğu durumlarda SRIE avantaj sağlamaktadır.

Anahtar Kelimeler: Retineks, mağara, Görüntü iyileştirme, Performans, Kıyaslama

Digital image enhancement methods can be used to enhance low-quality images. For example, it is possible to enhance low-light images with the help of popular photo editing software such as Photoshop. However, such computer applications require a high level of technical knowledge and advanced user experience.

One of the important methods developed for enhancing low-light images is the Retinex method [2]. Retinex aims to enhance the visibility of details in poorly illuminated regions more visible by evaluating the lighting and reflection components in an image separately, like human visual perception. Today, many different sub-methods based on Retinex theory have been developed and are used effectively to enhance image quality, especially in low-light scenes [3].

* Sorumlu yazar / Corresponding author, e-posta / e-mail: egemenyazlik@nevsehir.edu.tr (E. Yazlık)

Geliş / Received: 02.09.2025 Kabul / Accepted: 29.09.2025 Yayınlanma / Published: 15.10.2025

doi: 10.28948/ngumuh.1776861

In this study, state-of-art Retinex-based image enhancement methods are applied to low-light photographs taken in caves and the enhanced images are evaluated in detail with qualitative and quantitative analyses. Seven Retinex-based image enhancement methods are used to enhance images taken inside caves. To obtain reliable results, a data set consisting of 82 low-light images is used. Enhanced images are examined with quantitative and qualitative analysis. Three sample images are selected from the data set for qualitative analysis, and all methods were compared visually. For quantitative analysis, eleven different visual quality analysis methods are used.

2 Material and method

2.1 Cave image dataset

In this study, a unique image dataset containing a total of 82 photographs taken by the authors inside the caves is used. All images were taken in very low light conditions. All photos were scaled to 1600x1200 resolution, are in JPG format, and have an RGB color scheme. The authors have published the image dataset used in this study openly at the link on the data availability section of the article.

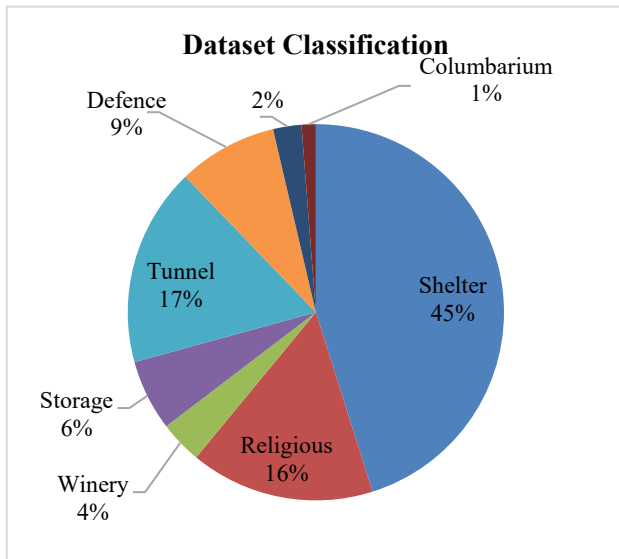


Figure 1. Classification of the cave image dataset.

The classification of the photographs in the data set according to the function of the caves in which they were taken is shown in Figure 1.

2.2 Retinex based image enhancement methods

The Retinex theory was developed by Edwin Land to simulate how the Human Visual System perceives color and luminance. From its early calculations, this theory has developed over time to a form that links individual neurons in the cerebral cortex, lateral geniculate nucleus, and primate retina to center/surround spatial opposition processes [4]. Retinex theory explains that by creating maps of lightness in different wavelength bands, color sensations that are strongly correlated with reflectance arise independently of the illuminating [2]. Following Land's work, Hurlbert has

thoroughly studied this new form of Retinex and other lightness theories, defining a common mathematical explanation. However, this explanation has the limitation that it cannot fully account for scene-specific reflections. In “monochromatic” scenes, where there is a dominant color in the scene, the “gray world” hypothesis assumed by Retinex—the average reflectance in the three spectral bands being equal—loses its validity. This causes Retinex to achieve an unrealistic, forced gray balance in such scenes. Hurlbert approached the aperture problem from an artificial neural network perspective, seeking a learning-based solution, and showed that the resulting solution again has the form of a center/surround spatial contrast. This finding suggests that center/surround mechanisms may be a universal solution for relative reflectance estimation under variable illumination conditions. However, the Human Visual System evaluates not only physical reflectance but also context-dependent relative reflectance, since surfaces in shadow do not retain their perceived brightness when directly illuminated [5], [6]. Moore et al. have addressed the Retinex problem as a natural application area on analogy Very Large-Scale Integration (VLSI) resistive networks. Their results have shown that color representation is directly dependent on the scene content. The common theoretical approach in all these studies has been to perform spatial operations separately in each spectral band and to prevent any interaction between spectral and spatial operations. This method provides strong global color constancy throughout the scene [7]. According to the Retinex theory, when considered from a human perspective, an image $I(x, y)$ can be separated into the illumination component $L(x, y)$ and the reflectance component $R(x, y)$ [8].

$$I(x, y) = R(x, y)L(x, y) \quad (1)$$

While the $L(x, y)$ value in Equation (1) is the illumination value, the $R(x, y)$ value is the reflection value that depends on the characteristics of the object or objects in the image [9]. Figure 2 shows an illustration representing how Retinex theory works in principle. It is known that the Retinex algorithm can be carried out on technical computer programs such as MATLAB.

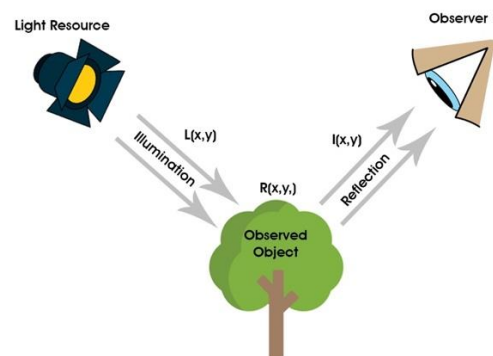


Figure 2. Principal Illustration of Retinex Theory

2.2.1 Multi-exposure fusion framework method inspired by the human visual system (BIMEF)

BIMEF is a method that aims to enhance images obtained in low-light conditions by taking inspiration from the principles of the Human Visual System. This framework consists of four basic components. The first component, Multi-Exposure Sampler, determines the number of synthetic images required for image fusion and the exposure ratio of each. The second component, Multi-Exposure Generator, generates multiple exposure images using a camera response model using the defined exposure ratios. The third component, Multi-Exposure Evaluator, calculates the weight maps to be used during image fusion based on illumination estimation techniques. In the final stage, the Multi-Exposure Combiner component combines the synthetic images created with the input image according to the weight maps and produces the final enhanced image that minimizes contrast and luminance distortions. This method significantly reduces the problems of over or under enhancement of contrast, which are frequently encountered in existing methods, by providing correct exposure especially in low-light regions [10], [11].

2.2.2 Light enhancement of low-light video content (DONG)

Low-Light Video Enhancement method known as DONG, proposes an unsupervised Retinex-based decomposition strategy and scene-level continuity constraint to ensure spatiotemporal consistency in low-light videos. Furthermore, consistent illumination and reflection features are achieved across different frames by using a dual-structure enhancement network and inter-frame interaction mechanism. Extensive experiments show that the proposed method surpasses existing methods and offers a new level of state-of-the-art performance [12].

Although the DONG method was developed for videos; for a single image, the method is applied by decomposing the low-light input into illumination and reflection components based on the Retinex model. Unlike video, where temporal consistency is required, only spatial constraints are imposed to ensure smooth yet edge-preserving illumination estimation. The illumination map is then enhanced through nonlinear adjustment to brighten dark regions, while the reflection component is preserved to maintain the original textures and colors. Finally, enhanced illumination and reflection are recombined, producing a visually enhanced low-light image. In essence, a photo can be regarded as a time-independent single frame of a video, so apart from the absence of temporal consistency, the technique remains technically the same.

2.2.3 Illumination map estimation-based method (LIME)

LIME is a simple but effective method to enhance visibility in low-light images. In the method, the illumination of each pixel is estimated by the maximum value in the RGB channels, and this initial map is refined using structural priors [13].

2.2.4 Fusion based low light enhancement method (MF)

MF method performs the fusion of information obtained from different exposures, illumination levels or processed variations to enhance the quality of low-light images. Unlike traditional methods, instead of directly changing the global illumination level, Feature-based Lighting (FBL) techniques selectively combine the data with the highest information content in local regions. In this way, both detail preservation is ensured and enhanced images that are perceived as more natural visually are obtained [14].

2.2.5 Multi scale retinex with color restoration (MSR)

MSR is an image processing method that aims to enhance both the detail and color accuracy in an image by imitating human visual perception. To enhance local contrast and adjust the light-dark balance, MSR normalizes the effects of illumination using Gaussian filters of varying sizes. Even in low light, this produces images that are more balanced, realistic, and detailed [15].

2.2.6 Naturalness preservation priority method (NPE)

The NPE method was developed as a technique to enhance details while preserving the natural appearance of images. This method enhances visual quality by adding appropriate brightness in low-light areas while keeping the illumination component of the image intact, thus providing enhancement without disrupting the natural appearance [16].

2.2.7 Weighted variational model for reflection and illumination estimation (SRIE)

SRIE is developed to estimate the reflectance and illumination components in an image. Instead of the traditional logarithmic transformation, this model preserves the reflectance components with more detail by using a weighted variational approach that provides a better priori representation. It also provides some degree of suppression on the noise and is resolved by an alternative minimization scheme. These features make the SRIE model more effective and superior compared to traditional variational models [17].

2.3 Image quality assessment metrics

2.3.1 Average gradient (AVG GRD)

Average Gradient is an important measure of image quality and visibility that is directly related to visual perception. According to Weber-Fechner's law, human sensitivity to brightness is limited, particularly in low-light situations [18]. Average gradient is calculated as follows:

$$AG = \frac{1}{(H-1)(W-1)} \sum_x \sum_y \frac{|G(x,y)|}{\sqrt{2}} \quad (2)$$

The image's size is denoted by $H \times W$ in this equation, and its gradient vector is denoted by $G(x,y)$ in Equation (2). A more successful image in terms of light enhancement is indicated by a greater average gradient value [19].

2.3.2 Information entropy (INF ENT)

Information Entropy is a measure of randomness or uncertainty in a signal or image. It quantifies how much

information is conveyed: the higher the transmitted information, the better the image quality. The reduction in uncertainty after imaging reflects how much useful information the image provides. This concept, based on Shannon's entropy [20], is a simple and unified assessment of image noise and blur. The entropy for the input is represented by $H(x)$ and that for the output by $H(y)$, assuming that x and y are two random variables that correspond to an input variable and an output variable, respectively. The joint entropy, $H(x, y)$, in this instance is defined as follows:

$$H(x, y) = H(x) + H_x(y) = H(y) + H_y(x) \quad (3)$$

where conditional entropies are represented by $H_x(y)$ and $H_y(x)$. They are, respectively, the input's entropy when the output is known and the output's entropy when the input is known. We can calculate $T(x; y)$, in this case as follows [21]:

$$\begin{aligned} T(x; y) &= H(x) - H_y(x) = H(y) - H_x(y) \\ &= H(x) + H(y) - H(x, y) \end{aligned} \quad (4)$$

2.3.3 Variance

Variance is used as a statistical method to evaluate the quality measurements. It is used to show whether the hypothesized component such as the image class factor or the image compression type factor can explain the variation in the data. Finding the image quality measurements that most reliably and discriminately determine the artifacts of noise, blur, and compression are the result of variance. The one-way variance equation is written as follows [22]:

$$F = \frac{MS_b}{MS_w} = \frac{SS_b/(k-1)}{SS_w/(N-k)} \quad (5)$$

Here, the F value is a statistical test value that shows the ratio of the variance between groups to the variance within groups. SS_b is the sum of squares between groups, SS_w is the sum of squares within groups, MS is Mean Square, k is the number of groups and N is the total number of samples.

2.3.4 Visual Information Fidelity (VIF)

VIF is information theory-based metric for evaluating image quality. The reference image serves as the signal from the statistical source that is modelled to represent natural settings. A channel model is used to represent Human Visual System (HVS). Degraded images (e.g. compressed or blurred) follow a degradation channel before passing through the Human Visual System. In these two scenarios, VIF computes the mutual information between the Human Visual System input and output and compares them to determine the image quality. This situation is illustrated in the Figure 3 [23].

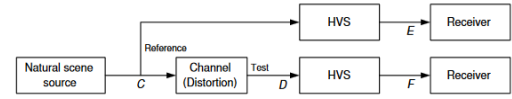


Figure 3. The mutual information ratio between C and E and C and F compares the amounts of detectable information from the reference and distorted signals.

$$VIF = \frac{\sum_{j=1}^M I(C_j; F_j)}{\sum_{j=1}^M I(C_j; E_j)} \quad (6)$$

In the Equation (6) C_j is reference sub band (wavelet subspace), E_j is information extracted from the reference sub band by Human Visual System, F_j is corresponding sub band in the distorted image and $I(C_j; F_j)$ is mutual information.

2.3.5 Gradient Magnitude Similarity Deviation (GMSD)

GMSD is an effective image quality assessment metric that evaluates the quality difference between the reference and the degraded image via the Gradient Magnitude Similarity (GMS). After calculating the gradient magnitude similarity for each pixel, GMSD estimates the overall image quality by taking the standard deviation of this similarity map. This method provides both high accuracy and is computationally fast, making it suitable for applications that process large volumes of visual data. The image quality assessment metric that applies mean pooling to the GMS map is called Gradient Magnitude Similarity Mean (GMSM) [24].

$$GMSM = \frac{1}{N} \sum_{i=1}^N GMS(i) \quad (7)$$

In Equation (7), $GMS(i)$ is Gradient Magnitude Similarity at i th pixel and N is the total number of pixels. GMSD equation is written as follows:

$$GMSD = \sqrt{\frac{1}{N} \sum_{i=1}^N (GMS(i) - GMSM)^2} \quad (8)$$

2.3.6 Image Mean Squared Error (IMMSE)

The IMMSE metric calculates the Mean Squared Error (MSE) between two numeric arrays (usually images) of the same size and data type, X and Y . The resulting MSE value quantifies the similarity between the two images; a lower MSE value indicates that the images are more similar. IMMSE is used particularly in image processing applications, for example to evaluate the effectiveness of compression, denoising or filtering operations [25].

2.3.7 Structural Similarity Index (SSIM)

SSIM is a perception-based image quality assessment metric developed by Wang et al. in 2004. This method evaluates image distortion not directly by pixel differences but by structural information change. Since the Human

Visual System is highly sensitive to structural information, SSIM measures the perceptual similarity between two images by considering brightness, contrast, and structure components. Structural information contains more information about visual objects that are composed of strongly connected or spatially close pixels. Luminance masking refers to a situation where distortions are less obvious at image edges, while contrast masking refers to a situation where distortions are less visible in textures. SSIM estimates the perceived image quality by measuring the similarity between the original and recovered images [26]. For two images x and y , SSIM is defined as:

$$SSIM(x, y) = [l(x, y)]^\alpha [c(x, y)]^\beta [s(x, y)]^\gamma \quad (9)$$

In Equation (9), l represents luminance, c represents contrast, s represents structure, and α , β , and γ represent positive constants.

2.3.8 Multi-Scale Structural Similarity Index (MULTISIM)

MULTISIM is a perception-based image quality assessment metric. It is an upgraded version of the Structural Similarity Index (SSIM) model that assesses image distortion as a change in structural data. Like SSIM, it considers the effects of brightness and contrast masking. MS-SSIM assesses the multi-scale similarity of the original and degraded images by examining brightness, contrast, and structural changes at various image sizes. Thus, it provides a more accurate quality assessment that is closer to human visual perception [27]. MS-SSIM is expressed as:

$$MS-SSIM(x, y) = [l_M(x, y)]^{\alpha M} \cdot \prod_{j=1}^M [c_j(x, y)]^{\beta j} [s_j(x, y)]^{\gamma j} \quad (10)$$

Here $l_M(x, y)$ is luminance similarity at lowest resolution, $c_j(x, y)$ is contrast similarity at j th scale, $s_j(x, y)$ is structural similarity at the j th scale, αM , βj , γj are weighting coefficients of each component and M is the number of scales.

2.3.9 Peak Signal-to-Noise Ratio (PSNR)

PSNR is a metric used to calculate the ratio of the maximum possible signal power compared to the distortion noise that affects the quality of an image. This ratio between two images is expressed in decibels. PSNR is calculated on a logarithmic scale because signals have a very wide dynamic range. Dynamic range represents the difference between the largest and smallest possible values of the signal, and this range may vary depending on the image quality. PSNR is the most widely used quality measurement technique to evaluate the reconstruction quality, especially for lossy image compression codecs. Here, the signal is considered as the original data, and the noise is considered as the error caused by compression or distortion. PSNR provides an approximate estimate of the reconstruction quality after compression, close to human perception [28]. PSNR is as follows:

$$PSNR = 10 \log_{10}(\text{peakval}^2)/MSE \quad (11)$$

In Equation (11) "peakval" represents the highest pixel value in the image data.

2.3.10 The Natural Image Quality Evaluator (NIQE)

NIQE is a reference-free metric for evaluating image quality. It makes it possible to measure an image's perceived quality without comparing it to the original. NIQE estimates image quality by focusing on the analysis of structural features based on the Human Visual System. This method uses statistical features to determine whether an image conforms to its natural structure [29].

2.3.11 Perception-Based Image Quality Evaluator (PIQE)

PIQE is a reference-free image quality assessment metric. It evaluates the quality of an image by analysing perceptual features based on the Human Visual System. This method detects distortions based on inherent structural features in images and calculates a quality score for each image. It relies on the local features of each block in the image to assess the quality. It performs local block-based assessment and divides the image into small parts and measures the quality [30].

3 Results and discussion

3.1 Quantitative analysis

The dataset of 82 cave images has been enhanced using Retinex-based algorithms, with parameters that are listed in Table 1.

Table 1. Methods and selected parameters

Method	Parameter and value
BIMEF	Illumination map coefficient (λ)=1 Enhancement degree (μ)=0,5
DONG	Number of frame (T)= 1 Batch size= 4
LIME	Refinement parameter (λ) = 0,8
MF	Weight: Laplacian contrast + chromaticity
MSR	Gaussian scales: $\sigma = \{15, 80, 250\}$
NPE	Difference parameter (ϵ)= 12 Brightness = 0,7
SRIE	Weighted variational regularization (α) = 0,02–0,2 Noise suppression parameter (β) = 0,001–0,01

Enhanced images are tested with quality assessment metrics and the results are shown with average values which are shown in Table 2.

In this study, the available "original" image is already degraded by soot, meaning that it cannot serve as a true ground-truth reference. Consequently, full-reference metrics GMSD, IMMSE, MULTISIM, PSNR, SSIM and VIF do not provide an absolute measure of perceptual quality but rather quantify the degree of similarity to the degraded input. A

higher PSNR or SSIM score in this context indicates that the enhancement has preserved the characteristics of the sooty image, whereas a lower score suggests stronger deviations from the degraded reference. This interpretation creates a paradox: enhancements that increase the visibility of details and enhance perceptual quality may appear as “losses” with respect to the original degraded image. Therefore, PSNR and SSIM should be understood not as direct indicators of restoration success, but as auxiliary measures that reflect how much the enhanced image diverges from the input.

According to the quantitative analysis results in Table 2 for all images, the performances of different image enhancement methods in various metrics are comparatively presented. The BIMEF method has achieved the best results in PIQE, GMSD, MULTISIM and SSIM metrics, demonstrating significant superiority over other methods. This shows that BIMEF is an effective method in enhancing image quality, reducing structural distortions and preserving multi-scale similarity. The LIME method has shown the highest performance in AVG GRD, INF ENT and VARIANCE metrics, and has been especially successful in dynamic range and detail preservation metrics such as average gradient, information entropy and variance in the image. These results show that LIME is a strong option in terms of contrast and local detail enhancement. The SRIE method stands out in NIQE, IMMSE and PSNR metrics. This performance demonstrates that SRIE is effective in areas such as natural image quality estimation, error minimization, and signal-to-noise ratio optimization. Especially the high values in PSNR support that the method is successful in noise suppression and image restoration.

The first and second column of Table 2 presents the NIQE and PIQE scores of Retinex-based image enhancement methods in comparison. Lower scores in both NIQE and PIQE represent higher perceived visual quality. Among the applied enhancement methods, SRIE outperformed with a NIQE score of 2.75 and BIMEF outperformed all other methods with a PIQE score of 24.698 and provided the lowest quality loss among the enhanced images. SRIE method is a very close second to BIMEF with a value of 24.702. In Figure 4 and 5, the radar graphic representation of the methods according to their NIQE and PIQE scores is presented and the clear performance superiority of BIMEF is also visually indicated.

In the third column of Table 2, Average Gradient scores are presented comparatively. Average Gradient quantitatively measures the level of detail and sharpness of an image, with higher values indicating better detail preservation and increased visual acuity. According to the results, the LIME method obtained the highest value among all methods with Average Gradient score of 55.93, thus showing the most successful enhancement performance in terms of detail preservation and sharpness. LIME is followed by MS-RETINEX (50.59) and DONG (46.67) methods, respectively. In Figure 6, Average Gradient scores are presented on the radar graph, and the superior detail preservation performance of LIME method is also visually indicated.

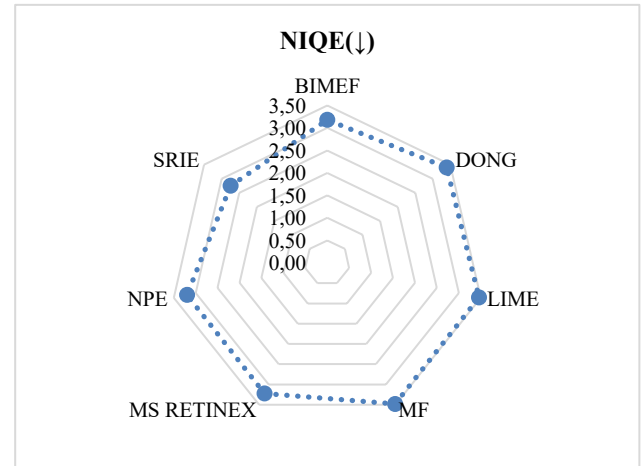


Figure 4. NIQE Analysis Results

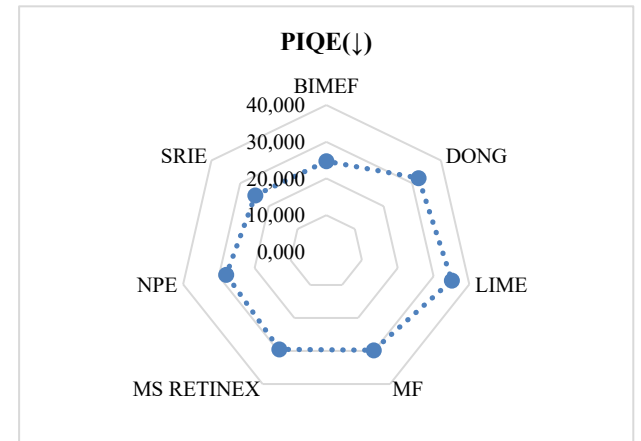


Figure 5. PIQE Analysis Results

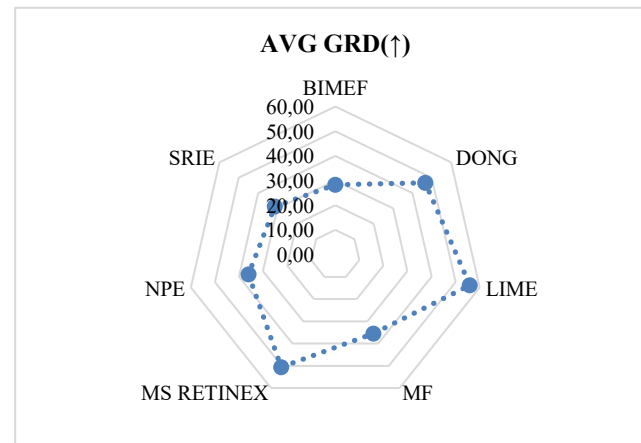


Figure 6. Average Gradient Analysis Results

The Information Entropy values used in the study are shown in the fourth column of Table 2. Information Entropy measures the amount and variety of information in an image; therefore, high entropy values generally indicate images that contain more detail and richer texture information. According to the evaluation results, the LIME method provided the highest information entropy with an entropy value of 7.46 and performed better than other methods in this

respect. This result shows that the LIME method can both preserve details better and reduce loss of information in low-light images. This shows that the enhancement methods not only increase the illumination but also enrich the information in the image content. LIME proved to be an effective method in increasing visual quality in low-light conditions by providing the best results in terms of information content. In Figure 7, Information Entropy scores are presented on the radar graph.

The Variance values presented in the fifth column of Table 2 represent a measure of the distribution of pixel values for the outputs of each image enhancement method. Higher Variance values generally represent more distinct details and better visual perception quality. According to the results obtained, the LIME method achieved the highest variance score with a value of 4 087.20 and provided the highest contrast and detail diversity compared to the other methods. This shows that the LIME method more effectively reveals the details in the image. LIME significantly contributed to the enhanced image quality in terms of both information content and visual contrast. In Figure 8, Variance scores are presented on the radar graph.

The GMSD metric is used to measure the structural fidelity between the enhanced and reference images. Lower GMSD values indicate higher structural similarity and less distortion. As presented in the sixth column of Table 2, the BIMEF method showed superior performance in preserving gradient-based structural information under low-light enhancement conditions, achieving the lowest GMSD score of 0,08. In Figure 9, GMSD scores are presented on the radar graph.

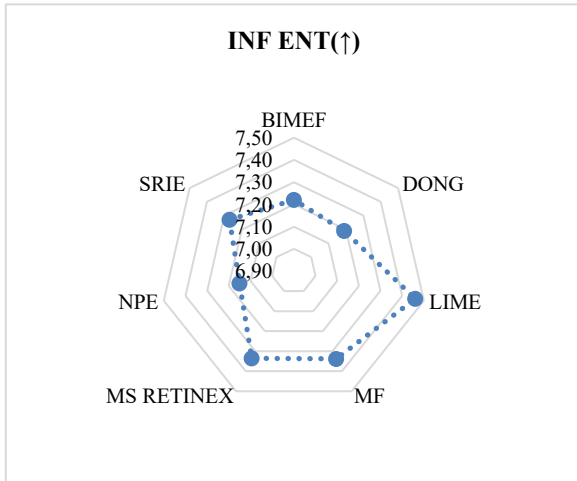


Figure 7. Information Entropy Analysis Results

The IMMSE evaluates the pixel-wise error between the enhanced and reference images, where lower values indicate better reconstruction accuracy. As shown in the seventh column of Table 2, the SRIE method outperformed all other methods by achieving the lowest IMMSE score of 1 838.25. This result confirms the ability of SRIE to minimize reconstruction errors during low-light image enhancement. In Figure 10, IMMSE scores are presented on the radar graph.

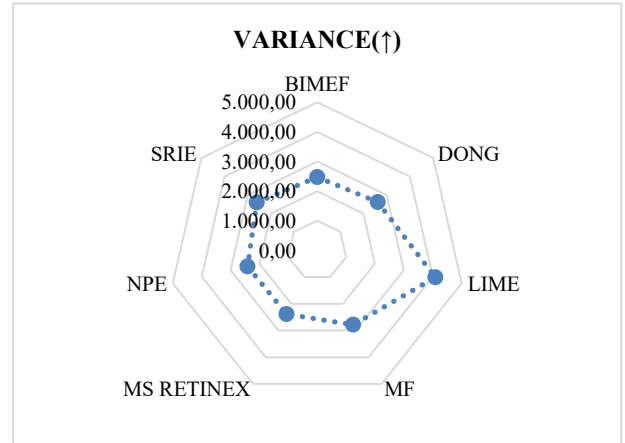


Figure 8. Variance Analysis Results

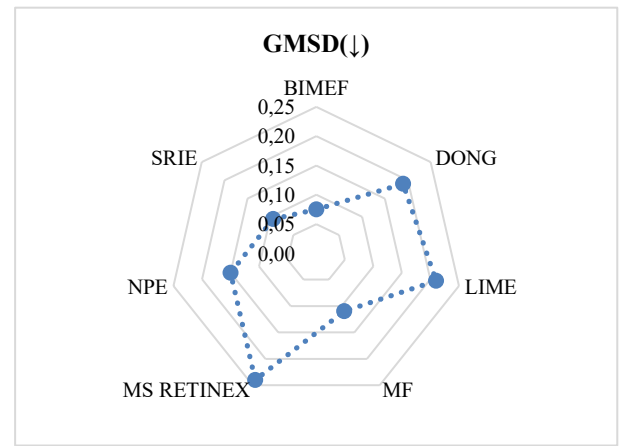


Figure 9. GMSD Analysis Results

The MULTISIM measures the perceptual similarity between the enhanced and reference images on multiple scales. Higher MULTISIM values indicate better structural preservation and perceptual quality. As shown in the eighth column of Table 2, the BIMEF method produced a better result than the other methods by obtaining a score of 0.91. This result highlights the effectiveness of BIMEF in preserving structural fidelity and visual quality during low-light image enhancement. In Figure 11, MULTISIM scores are presented on the radar graph.

PSNR is a widely used analysis method to evaluate the quality of enhanced images. Higher PSNR values indicate better image fidelity and lower distortion level. As shown in the ninth column of Table 2, the SRIE method achieved the highest PSNR score of 15.91. This result shows the success of the SRIE method in preserving image quality and minimizing noise distortion in the low-illumination image enhancement process. In Figure 12, PSNR scores are presented on the radar graph.

Table 2. Quantitative Analysis Results.

	NIQE (↓)	PIQE (↓)	AVG GRD(↑)	INF ENT(↑)	VARIANCE (↑)	GMSD (↓)	IMMSE (↓)	MULTISIM (↑)	PSNR (↑)	SSIM (↑)	VIF (↑)
BIMEF	3.18	24.70	28.29	7.22	2 473.70	0.08	2 020.23	0.91	15.34	0.66	1.80
DONG	3.40	32.19	46.67	7.19	2 618.10	0.19	5 869.80	0.70	10.89	0.44	2.87
LIME	3.46	35.12	55.93	7.46	4 087.20	0.21	10 508.29	0.64	8.04	0.37	6.62
MF	3.48	29.77	35.51	7.34	2 786.91	0.11	3 125.16	0.84	13.60	0.56	2.47
MS RETINEX	3.22	29.45	50.59	7.34	2 383.39	0.24	14 361.82	0.66	6.83	0.36	6.83
NPE	3.20	27.93	36.07	7.15	2 410.70	0.15	3 146.63	0.81	13.99	0.55	3.01
SRIE	2.75	24.70	31.35	7.27	2 604.54	0.09	1 838.25	0.87	15.91	0.63	2.23
ORG	2.92	17.50	19.18	6.81	2 296.83	0.00	0.00	1.00	65.535.00	1.00	1.00

ORG: Original source image to be enhanced. Bold and italic numbers show the best performance per criteria.

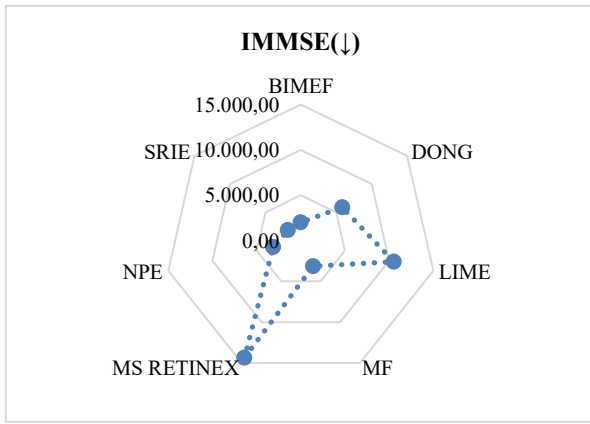


Figure 10. IMMSE Analysis Results

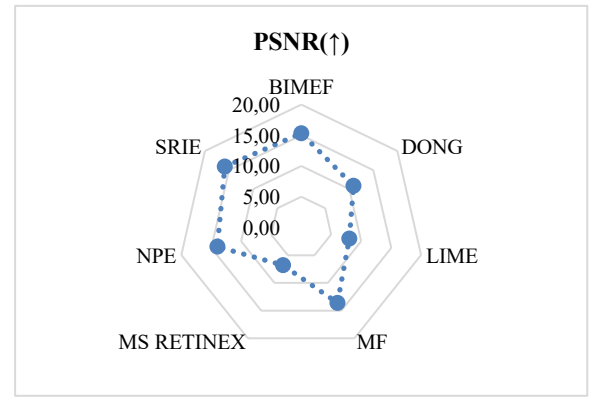


Figure 12. PSNR Analysis Results

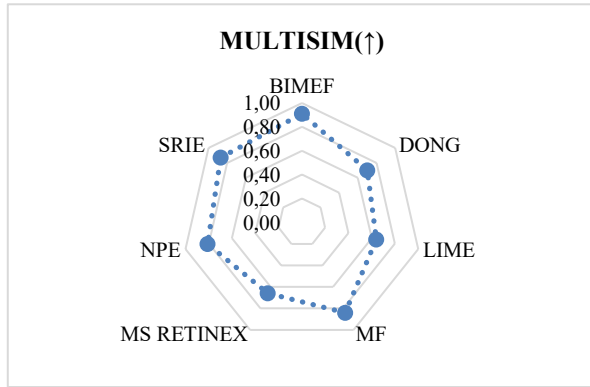


Figure 11. Multi-SIM Analysis Results

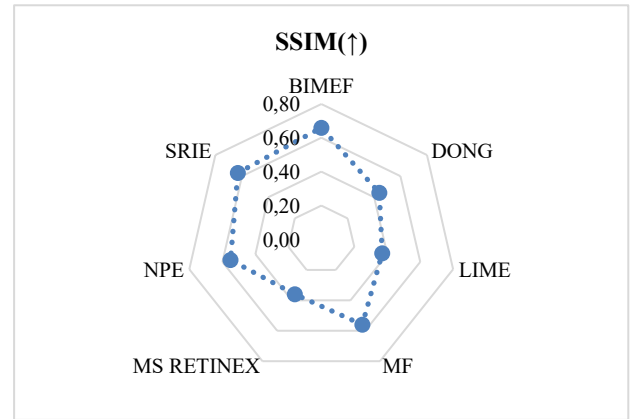


Figure 13. SSIM Analysis Results

SSIM measures the structural similarity between the enhanced image and the reference image. Higher SSIM values indicate better structural preservation and higher perceptual quality. As shown in the tenth column of Table 2, the BIMEF method exhibited the best performance among other enhancement methods with an SSIM score of 0.66. This result shows that BIMEF effectively preserves structural integrity while enhancing low-illumination images. In Figure 13, SSIM scores are presented on the radar graph.

The VIF metric measures the visual information preservation between the enhanced image and the reference image. Higher VIF values indicate more effective preservation of perceptual information. According to the results presented in the eleventh column of Table 2, the MS Retinex method achieved the highest VIF score of 6.83, showing the most successful performance among other enhancement methods. This result reveals that the MS Retinex method exhibits superior effectiveness in preserving perceptual information integrity in low-illumination images. In Figure 14, VIF scores are presented on the radar graph.

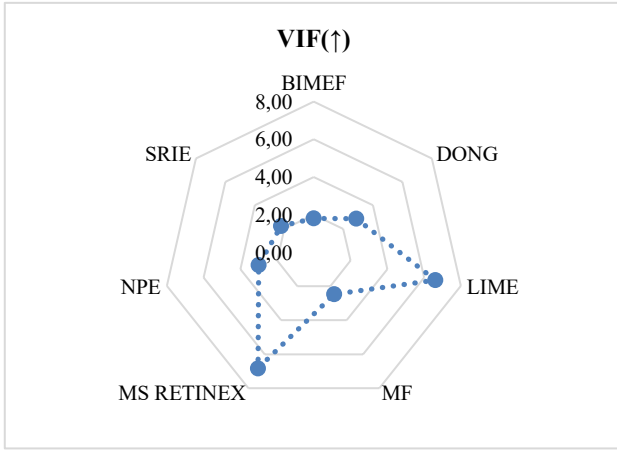


Figure 14. VIF Analysis Results

3.2 Qualitative analysis

Two different enhancement results are presented for detailed qualitative evaluation with two images selected from a dataset of 82 images. **Figure 15** shows the apse of a rock-carved church. Details not visible in the original image (15.a) due to low light have become visible in the enhanced image (15.b).



(a)



(b)

Figure 15. Image No. 56 a) Original image b) MultiScale Retinex Output

Figure 16 shows a caver working in a cave in front of a lock stone. In the enhanced image, details of the lock stone behind the caver are clearer. Furthermore, the hole in the middle of the lock stone became visible at the enhanced image.



(a)



(b)

Figure 16. Image No. 81 a) Original image b) NPE Output

For sample individual analysis of images, we have selected three samples out of 82 images. **Figure 17** shows all output results as enhanced images of the three selected images in the image dataset, enhanced by Retinex-based methods. According to the results of the quality analysis of the images in **Figure 17** with subjective analysis, the BIMEF method provides the most satisfactory results in terms of detail preservation, color accuracy and lighting balance among the low-light image enhancement methods. LIME, which is especially superior in terms of the clarity of scene details and natural color reproduction, stands out in terms of visual quality. Similarly, the MSRETINEX method offers a high-quality visualization with natural color distribution and contrast balance. LIME, on the other hand, exhibits a successful performance with its low noise and balanced lighting features. In contrast, the DONG and SRIE methods cause loss of detail due to excessive brightness or low contrast and produce weaker results in terms of visual

perception. In general, the BIMEF algorithm emerges as the most successful enhancement approach in terms of both visual quality and structural integrity.

Table 3 shows the objective analysis results of the three sample images in Figure 17. The numerical evaluation metrics in Table 3 reveal the all-performance metrics of eight different low-illumination image enhancement methods. The BIMEF method emerges especially in structural and perceptual quality metrics. It is seen that this method provides superiority with PSNR (highest in two images) and SSIM (highest in all images) values. Among the average results in Table 2 for the PSNR value, the best result was obtained by SRIE. This indicates that different methods can show different levels of success in processing individual images separately. In addition, GMSD (lowest in all scenes) and MULTISIM (highest in all scenes) results show that BIMEF can preserve scenes in general without distorting local details. By obtaining IMMSE with BIMEF at the lowest value twice, the error rate is minimized. In NIQE,

BIMEF twice and SRIE, once with the lowest value, provided successful results in natural image perception. In terms of PIQE metrics, the fact that MF, SRIE and BIMEF produce low values shows that these three algorithms are effective in reducing perceptual distortion.

SRIE, PSNR and IMMSE metrics give the best results once each, and it shows competitive performance especially in structural accuracy. In addition, the low values in NIQE and PIQE metrics show that SRIE offers results compatible with the human visual system. MSRETINEX and LIME emerge especially in AVG GRD, VARIANCE and VIF metrics, which are detail and contrast measures. LIME gives the highest result in both variance and VIF values in two scenes, and it is seen that it preserves the level of detail well in these scenes. MF method gives the lowest value in PIQE in one scene and offers a successful example in terms of perceptual quality. However, its performance in other metrics remains lower.

Table 3. Quantitative analysis results of sample images from data set (Image No: 56, 65, 81).

Image	AVG GRD(↑)	INF ENT(↑)	GMSD (↓)	IMMSE (↓)	MULTISI M(↑)	NIQE (↓)	PIQE (↓)	PSNR (↑)	SSIM (↑)	VARIANC E (↑)	VIF (↑)
Image No: 56 (BIMEF)	18.59	6.72	0.02	2 140.79	0.97	2.16	7.24	14.83	0.82	772.74	1.63
Image No: 56 (DONG)	43.40	6.73	0.21	7 785.42	0.63	3.30	10.29	9.22	0.51	926.78	4.14
Image No: 56 (LIME)	38.58	7.08	0.13	12 447.47	0.74	2.56	8.61	7.18	0.45	1 445.17	6.34
Image No: 56 (MF)	26.08	6.79	0.06	3 804.07	0.91	2.51	7.22	12.33	0.67	897.25	2.71
Image No: 56 (MS Retinex)	56.31	7.53	0.19	8 744.81	0.58	2.84	10.73	8.71	0.42	2 491.33	13.63
Image No: 56 (NPE)	33.71	7.10	0.12	3 372.36	0.79	2.38	8.37	12.85	0.57	1 519.62	4.67
Image No: 56 (SRIE)	23.38	6.84	0.06	2 856.47	0.91	2.23	9.20	13.57	0.72	948.80	2.66
Image No: 65 (BIMEF)	21.49	7.18	0.06	2 301.39	0.95	4.73	32.27	14.51	0.73	1 518.58	1.88
Image No: 65 (DONG)	42.71	7.08	0.20	8 012.41	0.72	4.29	52.05	9.09	0.45	1 817.97	3.00
Image No: 65 (LIME)	46.69	7.50	0.19	13 107.44	0.70	5.06	54.62	6.96	0.38	3 067.23	5.88
Image No: 65 (MF)	31.11	7.27	0.11	4 021.45	0.87	5.39	48.77	12.09	0.59	1 811.06	2.56
Image No: 65 (MS Retinex)	37.56	6.92	0.21	12 935.80	0.79	4.53	41.62	7.01	0.46	1 660.37	4.75
Image No: 65 (NPE)	31.37	7.06	0.14	4 401.85	0.85	4.74	45.54	11.69	0.57	1 511.63	2.65
Image No: 65 (SRIE)	26.21	7.33	0.09	2 772.64	0.90	3.84	31.07	13.70	0.67	1 908.99	2.53
Image No: 81 (BIMEF)	21.52	7.40	0.06	1 863.98	0.93	2.36	9.34	15.43	0.65	2 591.84	1.26
Image No: 81 (DONG)	42.47	7.40	0.21	6 702.68	0.64	2.85	22.70	9.87	0.38	2 974.09	1.96
Image No: 81 (LIME)	45.22	7.76	0.20	11 141.49	0.62	2.95	24.35	7.66	0.32	4 728.70	3.47
Image No: 81 (MF)	29.27	7.61	0.11	3 748.66	0.84	2.60	19.96	12.39	0.52	3 302.75	1.57
Image No: 81 (MS Retinex)	38.21	7.43	0.25	14 498.19	0.71	2.74	16.63	6.52	0.36	2 607.42	2.71
Image No: 81 (NPE)	28.98	7.13	0.16	3 226.00	0.79	2.78	15.03	13.04	0.51	2 236.63	1.64
Image No: 81 (SRIE)	25.24	7.55	0.09	1 687.24	0.87	2.14	10.44	15.86	0.64	2 841.63	1.75

	ORG	BIMEF	DONG	LIME	MF	MS RETINEX	NPE	SRIE
(a) Image No: 56								
(b) Image No: 65								
(c) Image No: 81								

Figure 17. Qualitative Analysis of Sample Images from Data Set (Image No: 56, 65, 81)

4 Conclusion

Retinex-based image enhancement methods provide effective enhancement on low-light images. However, the performance of Retinex-based methods on dark cave images has not been specifically addressed to date. In this study, for the first time, dark cave images were enhanced with state-of-the-art Retinex methods, and the obtained results were analyzed both quantitatively and qualitatively. According to the results, the BIMEF algorithm stands out in terms of overall success by consistently showing the highest performance in basic metrics such as SSIM, GMSD, MULTISIM, artifact measures (NIQE, PIQE) and PSNR. While MSRETINEX and LIME offer a strong image enhancement profile especially in metrics representing detail level and contrast, SRIE draws attention by providing structural accuracy and perceptual quality together in some images. While BIMEF stands out as the most balanced option in multi-dimensional quality needs, MSRETINEX and LIME can be preferred in applications where details are important.

As a result, BIMEF is recommended for cave images requiring general image enhancement and structural consistency; LIME can be preferred in cave images requiring high dynamic range and detail preservation. SRIE provides advantages in cases where noise reduction and objective quality assessment metrics are at the forefront. These results show that the method selection should be made depending on the application goals and evaluation criteria. Retinex-based image enhancement can also be used to enhance low-light underwater images and sooty murals.

Conflict of interest

There is no conflict of interest within the scope of this study.

Similarity ratio (iThenticate): %7

Data availability: The cave image dataset used in this study can be downloaded from this address:

<https://doi.org/10.57760/sciencedb.28326>

References

- [1] D. Ford and P. Williams, Karst Hydrogeology and Geomorphology. John Wiley and Sons, 2007. doi: 10.1002/9781118684986.
- [2] E. H. Land and J. J. McCann, Lightness and retinex theory. Journal of the Optical Society of America, 61, 1–11, 1971. doi: 10.1364/JOSA.61.000001.
- [3] T. Gao and P. Tao, A comprehensive review of low-light image enhancement techniques. International Conference on Image Processing, Computer Vision and Machine Learning, 158–170, 2024. doi: 10.1109/ICICML63543.2024.10957894.
- [4] E. H. Land, An alternative technique for the computation of the designator in the retinex theory of color vision. Proceedings of the National Academy of Sciences, 83, 3078–3080, 1986. doi: 10.1073/pnas.83.10.3078.
- [5] A. Hurlbert, Formal connections between lightness algorithms. Journal of the Optical Society of America, 1684–1693, 1986. doi: 10.1364/JOSAA.3.001684.
- [6] Z. Rahman, D. J. Jobson, and G. A. Woodell, Retinex processing for automatic image enhancement. Journal of Electronic Imaging, 13, 100–110, 2004. doi: 10.1117/1.1636183.
- [7] A. Moore, J. Allman, G. Fox and R. Goodman, A VLSI Neural Network for Color Constancy, Advances in Neural Information Processing Systems 3, Morgan Kaufmann, 1990.
- [8] B. Petro, C. Sbert, and J. M. Morel, Multiscale retinex. Image Processing Online, 71–88, 2014. doi: 10.5201/ipol.2014.107.
- [9] F. Katırcıoğlu, Düşük-ışıklı renkli görüntülerin iyileştirilmesinde kullanılan retineks algoritmalarının karşılaştırmalı analizi. Mühendislik Bilimleri ve Araştırmaları Dergisi, 3, 188–206, 2021. doi: 10.46387/bjesr.955356.
- [10] Z. Ying, G. Li and W. Gao, A bio-inspired multi-exposure fusion framework for low-light image enhancement. arXiv, 2017. doi: 10.48550/arXiv.1711.00591.
- [11] H. Fu, W. Zheng, X. Meng, X. Wang, C. Wang, and H. Ma, You do not need additional priors or regularizers in retinex-based low-light image enhancement. IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 18125–18134, Vancouver, BC, Canada, 2023. doi: 10.1109/CVPR52729.2023.01738.
- [12] X. Xu, K. Zhou, T. Hu, R. Wang, and H. Bao, Low-light video enhancement via spatial-temporal consistent illumination and reflection decomposition. arXiv, 2024. doi: 10.48550/arXiv.2405.15660.
- [13] X. Guo, Y. Li, and H. Ling, LIME: Low-light image enhancement via illumination map estimation. IEEE Transactions on Image Processing, 26, 982–993, 2017. doi: 10.1109/TIP.2016.2639450.
- [14] X. Zhou, J. Guo, H. Liu, and C. Wang, A fusion-based and multi-layer method for low light image enhancement. 2023 IEEE International Conference on Acoustics, Speech and Signal Processing, pp. 1–5, Rhodes Island, Greece, 2023. doi: 10.1109/ICASSP49357.2023.10096454.
- [15] Z. Mahmood, N. Muhammad, N. Bibi, Y. M. Malik, and N. Ahmed, Human visual enhancement using multi scale retinex. Informatics in Medicine Unlocked, 13, 9–20, 2018. doi: 10.1016/j.imu.2018.09.001.
- [16] P. Joshi and S. Prakash, Image enhancement with naturalness preservation. The Visual Computer, 36, 71–83, 2020. doi: 10.1007/s00371-018-1587-6.
- [17] X. Fu, D. Zeng, Y. Huang, X. P. Zhang and X. Ding, A weighted variational model for simultaneous reflectance and illumination estimation. IEEE Conference on Computer Vision and Pattern Recognition, pp. 2782–2790. Las Vegas, NV, USA, 2016. doi: 10.1109/CVPR.2016.304.
- [18] Y. H. Liu, K. F. Yang and H. M. Yan, No-reference image quality assessment method based on visual parameters. Journal of Electronic Science and Technology, 17, 171–184, 2019. doi: 10.11989/JEST.1674-862X.70927091.

- [19] T. Stathaki, Image Fusion: Algorithms and Applications. Elsevier, 2011.
- [20] E. Shannon, A mathematical theory of communication. Bell System Technical Journal, 27, 379–423, 1948. doi: [10.1002/j.1538-7305.1948.tb01338.x](https://doi.org/10.1002/j.1538-7305.1948.tb01338.x).
- [21] D. Y. Tsai, Y. Lee and E. Matsuyama, Information entropy measure for evaluation of image quality. Journal of Digital Imaging, 21, 338–347, 2008. doi: [10.1007/s10278-007-9044-5](https://doi.org/10.1007/s10278-007-9044-5).
- [22] I. Avciabas, B. Sankur and K. Sayood, Statistical evaluation of image quality measures. Journal of Electronic Imaging, 11, 206–223, 2002. doi: [10.1117/1.1455011](https://doi.org/10.1117/1.1455011).
- [23] H. Alan and C. Bovik, A visual information fidelity approach to video quality assessment. The First International Workshop on Video Processing And Quality Metrics for Consumer Electronics, 2117–2128, 2005. https://live.ece.utexas.edu/publications/2005/hrs_vidq_ual_vpqm2005.pdf.
- [24] 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, 23, 684–695, 2014. doi: [10.1109/TIP.2013.2293423](https://doi.org/10.1109/TIP.2013.2293423).
- [25] Mathworks. Mean-Squared Error User's Guide (r2014b). Retrieved October 6, 2025 from <https://www.mathworks.com/help/images/ref/immse.html>.
- [26] Z. Wang, A. Bovik, H. Sheikh and E. Simoncelli, Image quality assessment: from error visibility to structural similarity. IEEE Transactions on Image Processing, 13, 600–612, 2004. doi: [10.1109/TIP.2003.819861](https://doi.org/10.1109/TIP.2003.819861).
- [27] Z. Wang, E. Simoncelli and A. Bovik, Multiscale structural similarity for image quality assessment. The Thirty-Seventh Asilomar Conference on Signals, Systems & Computers, pp. 1398-1402, Pacific Grove, CA, USA, 2003. doi: [10.1109/ACSSC.2003.1292216](https://doi.org/10.1109/ACSSC.2003.1292216).
- [28] U. Sara, M. Akter and M. Uddin, Image quality assessment through FSIM, SSIM, MSE and PSNR—a comparative study. Journal of Computer and Communications, 7, 8-18, 2019. doi: [10.4236/jcc.2019.73002](https://doi.org/10.4236/jcc.2019.73002).
- [29] A. Mittal, R. Soundararajan and A. Bovik, Making a 'completely blind' image quality analyzer. IEEE Signal Processing Letters, 20, 209–212, 2013. doi: [10.1109/LSP.2012.2227726](https://doi.org/10.1109/LSP.2012.2227726).
- [30] N. Venkatanath, D. Praneeth, M. Chandrasekhar, S. Channappayya and S. Medasani, Blind image quality evaluation using perception-based features. Twenty First National Conference on Communications, pp. 1–6, Mumbai, India, 2015. doi: [10.1109/NCC.2015.7084843](https://doi.org/10.1109/NCC.2015.7084843).

