

A Lightweight Surveillance System for Human Re-Identification Using Traditional Image Processing Techniques

Sophan Wahyudi Nawawi^{1*}, Muhammad Irfan Jaafar², Sulaiman Sabikan³

¹Faculty of Electrical Engineering, Universiti Teknologi Malaysia, 81310 Johor Bahru, Johor, Malaysia

²Worlby Network Sdn Bhd, L1-55 Menara Shaftsbury Putrajaya, Jalan Alamanda, Presint 1 62000 Putrajaya, Malaysia

³Faculty of Electrical Engineering Technology, Universiti Teknikal Malaysia Melaka, Hang Tuah Jaya, 76100 Durian Tunggal, Melaka, Malaysia.

Corresponding author* email: e-sophan@utm.my

Available online: 30 December 2025

ABSTRACT

Surveillance systems play a pivotal role in enhancing security, especially in large-scale environments such as academic campuses. While modern re-identification systems typically employ deep learning techniques, these models demand significant computational resources and extensive labeled datasets, limiting their applicability in real-world low-resource environments. This study presents the Integrated Campus-Activities Monitoring System (ICAMS), a lightweight, real-time surveillance system for human re-identification based on traditional image processing methods. The proposed system integrates a revised Structural Similarity Index (SSIM) and a novel Luv Similarity metric within the CIELUV color space, eliminating the need for computationally expensive deep learning models. Experimental evaluations show that the system achieves satisfactory accuracy and robustness in appearance-based human tracking while maintaining low hardware requirements. These results demonstrate the potential for deploying cost-effective surveillance systems capable of real-time performance in environments with limited infrastructure.

Keywords: Human Re-Identification; SSIM; CIELUV; Luv Similarity; Surveillance Systems

1. Introduction

In modern security frameworks, real-time human monitoring through intelligent surveillance systems is crucial for ensuring safety across sensitive and public environments. Conventional surveillance systems rely heavily on manual supervision, often resulting in delayed response times and labor-intensive review processes. Although machine learning and deep neural networks have significantly advanced object recognition and tracking capabilities, their practical implementation remains constrained by high computational demand and the necessity for large labeled datasets. To bridge this gap, we propose a computationally efficient surveillance solution, the **Integrated Campus-Activities Monitoring System (ICAMS)**. ICAMS is designed for environments where resource constraints prohibit the use of GPU-intensive models. It enables human tracking and re-identification by leveraging traditional image processing algorithms—namely, an improved Structural Similarity Index (SSIM) and a newly proposed Luv Similarity measure formulated within the perceptually uniform CIELUV colour space. The core advantage of this system lies in its **low hardware dependency**, making it suitable for deployment in real-world settings such as university campuses, small institutions, and developing regions. The system not only detects and tracks individuals across multiple camera views but also triggers alerts when unauthorized movements or boundary violations occur. This paper presents the design and implementation of ICAMS, evaluates its performance on standard surveillance scenarios, and compares its effectiveness with several state-of-the-art re-identification approaches. The findings aim to contribute to the field of smart sensing and monitoring systems by offering a viable alternative to conventional deep learning-based surveillance solutions.

2. Related Work

Tracking involves estimating an object's trajectory across frames while maintaining consistent identification despite changes in appearance, motion, or scale. In surveillance contexts, tracked objects typically include humans, vehicles, or animals. Object tracking typically begins with object detection, followed by associating features across frames to maintain continuity [1].

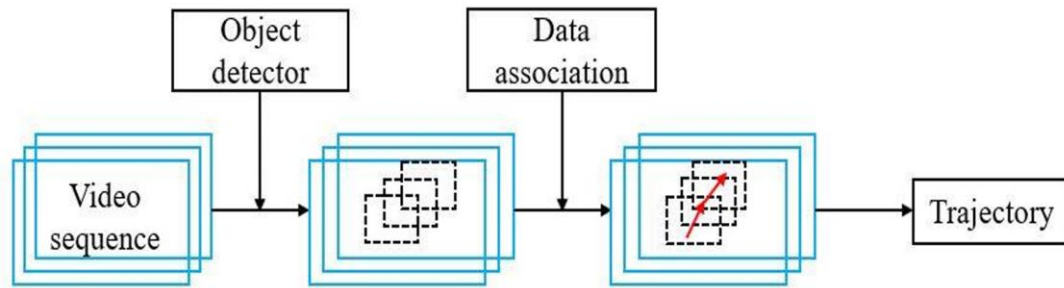


Figure 1. Basic Object tracking framework

The object tracking process begins with detecting the object of interest in each frame, typically using a bounding box. Features such as points [2][3], geometric shapes [4], colour histograms [5][6], or deep learning-based [7] embeddings are then extracted to represent the object. These features are used to associate detections across consecutive frames, ensuring consistent tracking despite changes in appearance, motion, or scale. Depending on the application, tracking may rely on a single feature or a combination to handle challenges like occlusion, deformation, or illumination changes. Motion models, such as the Kalman filter, may also be applied to predict the object's next position based on its velocity or size change [8]. Finally, a data association algorithm, like the Hungarian algorithm [9], matches detected objects to existing tracks by minimizing the feature distance, ensuring accurate and continuous tracking over time.

Re-identification involves locating a specific object with a distinct visual appearance across images or video frames using cues like visual features and position. Two main approaches exist: traditional feature-distance methods and end-to-end deep learning. Traditional methods rely on color histograms [10] or hierarchical Gaussian descriptors like GOG and ZOZ [11][12] to match appearances. Deep learning methods, especially unsupervised ones, are increasingly used due to the difficulty of manual labeling. Studies in [13], [14], and [15] apply unsupervised deep transfer learning to combine labeled and unlabeled data, while [16] introduces an autoencoder-based framework integrating detection and re-identification for improved accuracy.

Visual similarity is a key factor in associating objects across frames during tracking, complementing spatial, geometric, and colour-based features. While deep learning approaches like Siamese Neural Networks are commonly used to quantify similarity within bounding boxes, traditional methods such as the Structural Similarity Index Measure (SSIM) [17] also offer effective visual comparison. SSIM evaluates statistical correlations between images and aligns closely with human perception, making it useful for detecting visual patterns. Although SSIM is sensitive to significant transformations, its effectiveness remains in scenarios with minimal inter-frame variation, such as real-time human tracking.

The Structural Similarity Index Measure (SSIM) evolved from the Universal Quality Index (UQI) [18], developed by Wang and Bovik, which lacked the constants later introduced to prevent division by zero. Extensions like Multi-Scale SSIM (MS-SSIM) [19] evaluate contrast and structure across multiple scales, while CW-SSIM [20] uses complex wavelet coefficients to improve robustness against scaling, translation, and rotation. Further enhancements to SSIM include incorporating sharpness, Shannon entropy [21], histogram analysis, and edge detection [22][23], as well as integrating Stevens' power-law for signal adjustment [24]. Optimizations for perceptual alignment [25] and large-scale image efficiency [26] have also been proposed, demonstrating SSIM's adaptability to various image assessment challenges.

Our research improves human tracking by integrating a proposed SSIM variant into the Hungarian algorithm to address inaccuracies in the original SSIM formulation, particularly its sensitivity bias between low and high values [27]. We also develop a re-identification system that enhances tracklet association across multiple cameras. Using the CIELUV colour space, our novel colour-based approach computes image moments to effectively compare and re-identify individuals across different camera perspectives, advancing surveillance and tracking technologies.

3. Research Methodology

The methodology of our monitoring system combines online tracking and global re-identification. The online tracking system continuously monitors individuals across multiple cameras, providing real-time insights into their movements. It uses object detection, Euclidean distance, bounding box intersection over union (IoU), and appearance similarity scores to link tracks between frames on a single camera. The global re-identification system, activated as needed, associates tracklets from different cameras. It starts with background subtraction, then uses k-means clustering to identify dominant colours and measure colour similarity. Image moments extracted from these similarities help assess identity matches. Together, these systems enhance surveillance and security by accurately tracking and re-identifying individuals.



Figure 2. Monitoring system with Online tracking and global re-identification

3.1 Online Tracking

Online tracking enables real-time monitoring of individuals or objects within a single camera view, ensuring consistent identification across frames despite variations in posture, motion, or scale. The system performs object detection in each frame, extracting key features such as the coordinates, dimensions, and the image segment of the detected object. These features are then compared with those from the preceding frame. In the event of track loss, the previous features are preserved for a few frames, allowing for the potential reappearance of the object. If no correlation is found, a new track ID is assigned.

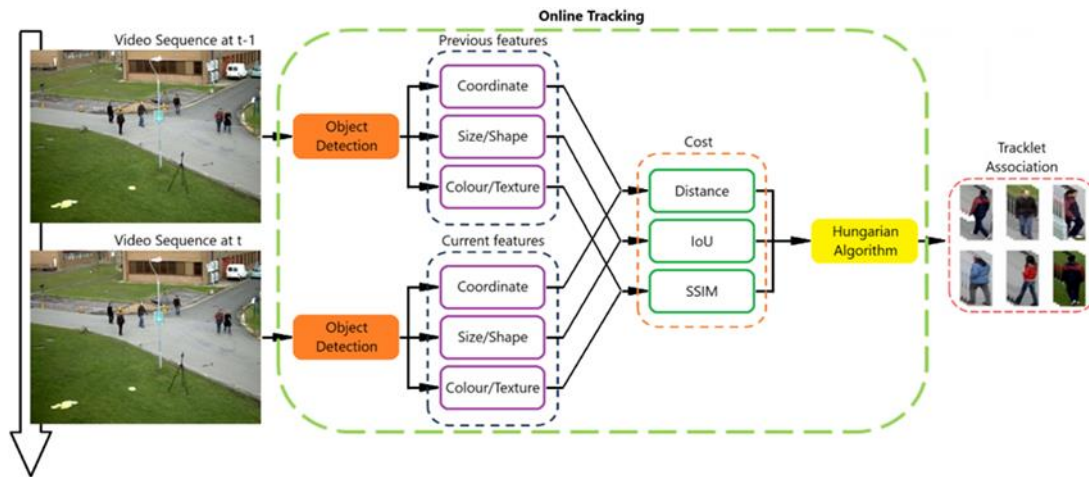


Figure 3. Online tracking

To improve track association, the system uses Euclidean distance, Intersection over Union (IoU), and object size changes as cost metrics. However, basic tracking struggles with occlusion and collisions. To overcome these limitations, the Structural Similarity Index (SSIM) is employed to assess pattern similarity in the detection region, helping to reassociate lost tracks and address appearance collisions. The Hungarian algorithm is then used to optimally associate costs from distance, IoU, and SSIM scores between consecutive frames, enhancing tracking accuracy.

3.2 Proposed SSIM

The proposed modification to SSIM involves replacing original equation of the luminance and contrast components and substituting it with a more linear approach, such as an absolute error equation or a simple relative difference for both luminance and contrast. This change aims to align luminance and contrast with the Weber-Fechner law [28], which asserts that the human perception of changes in stimuli (ΔI) is proportional to the background stimuli (I) at the rate of (k) or also referred as Weber contrast. In other words, the human sensory system perceives changes in relation to the background stimuli.

$$k = \frac{\Delta I}{I} \quad (1)$$

The suggested approach for the luminance equation involves using the relative difference in the local mean between two images, with the highest intensity serving as the background stimulus. Since this result is negatively correlated with human perception, 1 must be subtracted from the relative difference to invert it. The equation is as follows:

$$l(x, y) = 1 - \frac{|\mu_x - \mu_y|}{(2^n - 1)} \quad (2)$$

Contrast refers to the rate or gradient of intensity changes across an image. While contrast is more complex for humans to assess compared to intensity, most people can still distinguish differences in contrast between images. The proposed equation for contrast involves a scale that compares low and high standard deviations, and the formula is as follows:

$$c(x, y) = \begin{cases} \left| \frac{\sigma_x + k}{\sigma_y + k} \right|, & \sigma_x < \sigma_y \\ \left| \frac{\sigma_y + k}{\sigma_x + k} \right|, & \sigma_x \geq \sigma_y \end{cases} \quad (3)$$

While the Structure components of SSIM remain the same where it is the measurement of the correlation between 2 standard deviations of image x and y similar to Pearson's correlation coefficient with some constant C_2 :

$$s(x, y) = \frac{(\sigma_{xy} + C_3)}{(\sigma_x \sigma_y + C_3)} \quad (4)$$

This study adopts the same structural equation as the original SSIM because it effectively compares textures between images, eliminating the need for modification. The proposed SSIM is also a product of $l(x, y)$, $c(x, y)$ and $s(x, y)$ weighted by the functions α , β , and γ , which result in:

$$SSIM(x, y) = [l(x, y)^\alpha \cdot c(x, y)^\beta \cdot s(x, y)^\gamma] \quad (5)$$

3.3 Proposed Luv Similarity

CIELUV is a color space defined by the International Commission on Illumination (CIE) to model how humans perceive color [29]. It's often preferred in contexts like computer screens, lighting, or projection, where colors are mixed using light (additive color mixing), as opposed to pigment mixing of CIELAB. An experiment was conducted to examine how colours appear under varying brightness levels across different colour spaces. A piece of paper with matte red, green, blue, and white areas was equally illuminated with gradually increasing white light. Images were captured from darkness to full illumination, and the pixel values were plotted in RGB, CIELAB, and CIELUV colour spaces to observe how each represents changes in colour and brightness, highlighting how human colour perception adjusts to lighting variations.

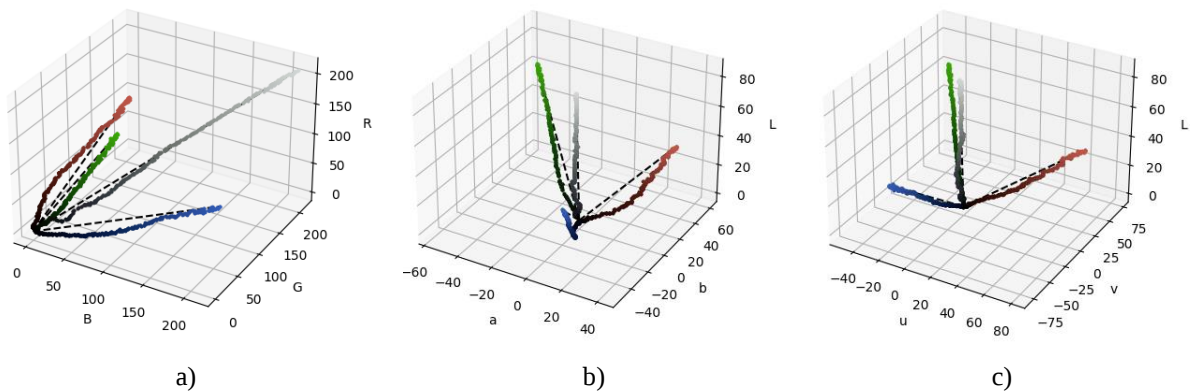


Figure 4. The rate of change of intensity in colour space, a) RGB b) CIELAB c) CIELUV

Table 1. Angular distance of each colour with regression line in degree

Colour	RGB	CIELAB	CIELUV
Red	12.9413	17.0596	5.4376
Green	7.7355	8.1904	3.8013
Blue	12.0131	17.8055	4.3646
White	3.3414	10.7039	8.8851
Mean distance	9.0078	13.4398	5.6221

Figure 4 shows that as brightness increases, all colors in the RGB, CIELAB, and CIELUV spaces move outward from the origin toward their respective color regions. Regression lines from the origin to each color's final point represent the ideal path of consistent color across varying brightness levels. Table 1. presents the angular distances between each color point and its regression line, revealing that CIELUV has the smallest mean angular distance among the three spaces. This indicates that CIELUV maintains the most linear relationship between color and brightness, making it particularly effective for comparing colors at different intensities.

This study leverages the linear behavior of color points in the CIELUV space to develop a color similarity formula that reduces the impact of external brightness when comparing a test image to a reference image. Both Euclidean and angular similarity measures are derived from an inverted form of the Weber contrast expression, which defines contrast as the perceived change relative to a background stimulus [30]. The proposed Luv Similarity (S_{luv}) composed of two similarity metric which is Euclidean similarity (S_{euc}) and angular similarity (S_{ang}).

$$S_{euc} = 1 - \frac{\Delta E_{uv}}{d_{ref}} \quad (6)$$

$$S_{ang} = 1 - \frac{\theta_{uv}}{a_{ref}} \quad (7)$$

$$S_{luv} = S_{euc} \times S_{ang} \quad (8)$$

Where the Euclidean or Delta-E, ΔE_{uv} , of Luv Similarity can be defined as:

$$\Delta E_{uv} = \sqrt{(L^*_1 - L^*_2)^2 + (u^*_1 - u^*_2)^2 + (v^*_1 - v^*_2)^2} \quad (9)$$

And the formulated polynomial reference distance, d_{ref} , expression is articulated as follows:

$$d_{ref} = 183.319349 - 83.319381L^* + 30.504766u^* + 25.438910v^* - 120.527103L^*u^*v^* \quad (10)$$

The angular distance, θ_{uv} , can be calculated using the smallest angle between the 2 hues:

$$k_C = \tanh\left(\frac{\sqrt{u^{*2}_1 + v^{*2}_1}}{1.6}\right) \times \tanh\left(\frac{\sqrt{u^{*2}_2 + v^{*2}_2}}{1.6}\right) \quad (11)$$

$$\Delta h = \text{atan2}(v^*_1, u^*_1) - \text{atan2}(v^*_2, u^*_2) \quad (12)$$

$$\theta_{uv} = k_C \times \Delta h \quad (13)$$

And the polynomial angular reference, a_{ref} , is as follows:

$$a_{ref} = 1.12522175 - 0.47827603h_{uv} + 0.04969178h_{uv}^2 + 0.04356436h_{uv}^3 \quad (14)$$

3.4 Global Re-identification

Re-identification in computer vision focuses on matching the visual features of objects across one or more surveillance cameras to maintain continuous tracking over time. It is especially useful for resolving issues like occlusion or temporary tracking loss. By linking location-time data through multiple detections, re-identification enables the construction of uninterrupted motion paths. The goal of this study is to develop a re-identification system suitable for industrial use, optimized to run on mid-range hardware for practical deployment.

Traditional re-identification approaches involve complex combinations of human attributes using multiple feature detections and bag-of-words models. Due to this complexity, recent works have shifted toward neural networks ([31], [32], [33], [34]) to handle the re-identification process, often performing the computations offline or concurrently with video feeds. However, neural networks demand substantial training data and computational resources [35], making them costly compared to classical image processing methods. To enhance efficiency, this study limits the use of neural networks to object detection and applies conventional techniques for the re-identification phase.

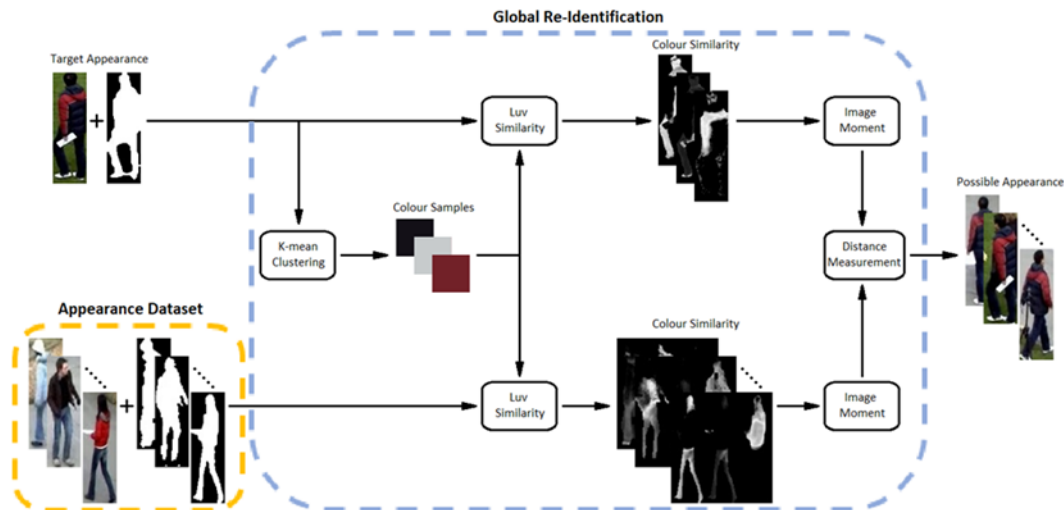


Figure 5. Proposed appearance recognition architecture

As shown in Figure 5, the proposed system separates tracking from re-identification to reduce computational load, reserving appearance-based searches for specific queries. The system leverages background subtraction to isolate subjects, then applies k-means clustering to extract colour features. A custom Luv-based similarity metric is used to assess appearance resemblance, followed by comparing image moments of colour similarity distributions. Similar colour moments indicate a likely match, allowing the system to accurately determine the time and location of an individual's appearance in the surveillance footage.

The integrated human tracking and re-identification system combines real-time monitoring with cross-camera identification to enhance surveillance capabilities. The online tracking system links object detections across video frames using features like location, size, and appearance, utilizing a modified Structural Similarity Index (SSIM) aligned with Weber's law to improve accuracy and reduce computational cost. Meanwhile, the global re-identification system matches appearances by comparing the weighted average of color features across time and camera views. To account for varying lighting conditions, it introduces Luv-Similarity, a color distance metric based on the CIELUV color space that maintains hue consistency while minimizing brightness effects.

4. Result and Discussion

The results of the proposed SSIM and Luv-Similarity formulations are presented, comparing their performance with alternative methods. The evaluation includes a detailed comparison of the proposed SSIM with various image assessment techniques to determine how well it aligns with human-perceived image quality, enhancing the tracking system's ability to accurately identify individuals based on their appearance. Additionally, Luv-Similarity is assessed against commercially available color distance measurements, focusing on its effectiveness in distinguishing colors under varying lighting conditions and its immunity to brightness fluctuations, which is crucial for re-identification systems in dynamic environments.

Benchmarking experiments also compare the online tracking system with traditional point-based prediction tracking and deep learning-based tracking methods, evaluating accuracy and real-time processing (FPS) in simulated CCTV footage scenarios. The global re-identification system is compared to traditional color histogram-based and Siamese network-based methods, with key metrics such as precision, recall, accuracy, and latency being measured. The research concludes by addressing the initial research question and summarizing the findings and the success in meeting the research objectives.

4.1 Image Quality Assessment

A comprehensive comparison is conducted between the proposed SSIM and several existing SSIM variants, including the original SSIM, Multi-Scale SSIM (MS-SSIM), and Complex Wavelet SSIM (CW-SSIM), as well as widely recognized metrics such as Mean Squared Error (MSE) and Peak Signal-to-Noise Ratio (PSNR). The analysis also includes alternative image quality assessment methods, namely Erreur Relative Globale Adimensionnelle de Synthèse (ERGAS) [36], Spatial Correlation Coefficient (SCC) [37], and Visual Information Fidelity (VIF) [38], as detailed in [39]. The aim of this evaluation is to assess the performance of the proposed SSIM in measuring image similarity relative to human perception. For the experiment, the correlation between these methods and human subjective scoring is analyzed, while also considering the processing speed of each approach. The datasets used in the study include 29 high-

resolution color images compressed using JPEG and JPEG2000 at different compression ratios, resulting in 233 and 227 total images, respectively [40].

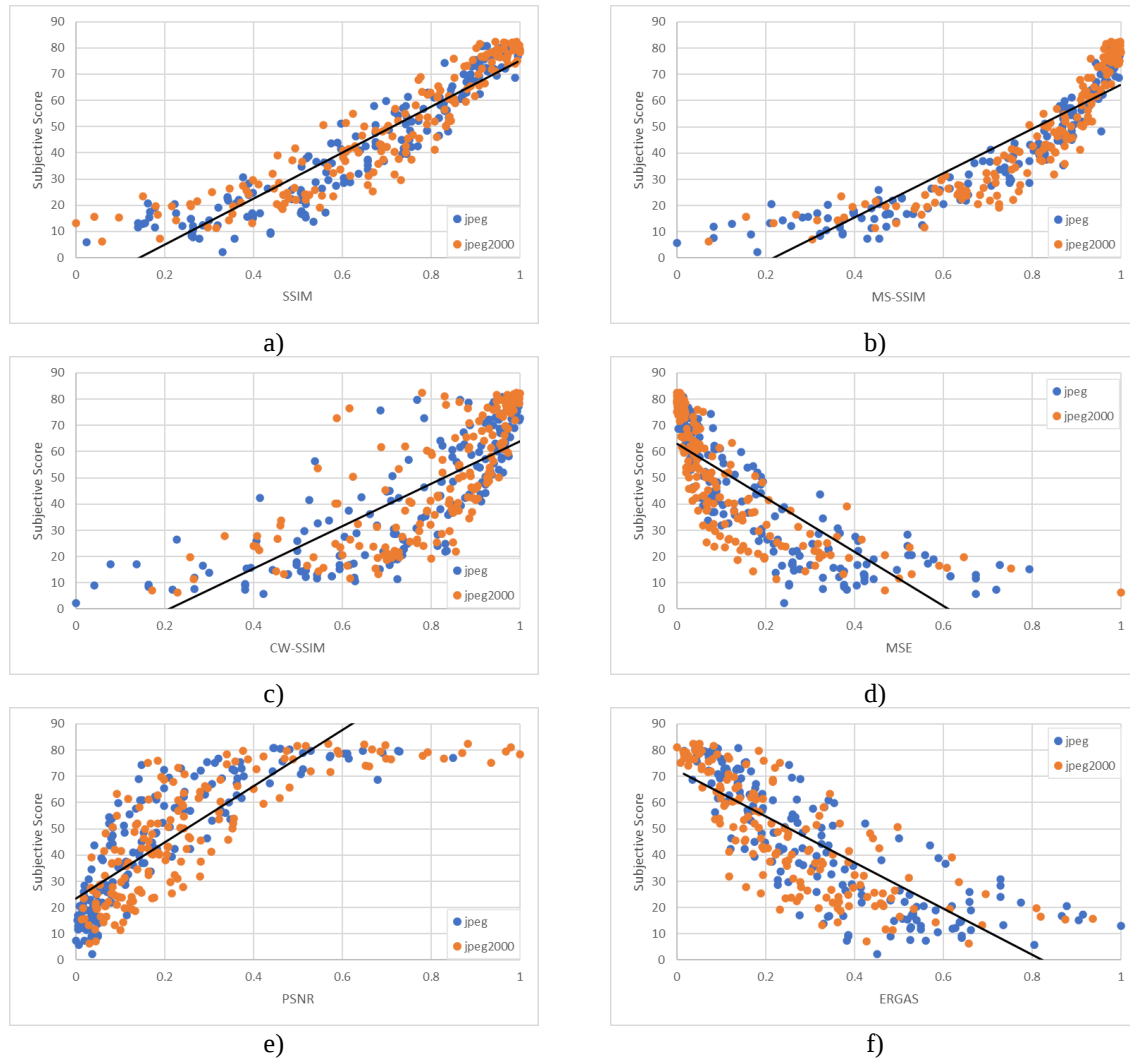


Figure 6. Scatter plot of mean subjective score versus model prediction. a) SSIM. b) MS-SSIM. c) CW-SSIM. d) MSE. e) PSNR. f) ERGAS (cont'd)

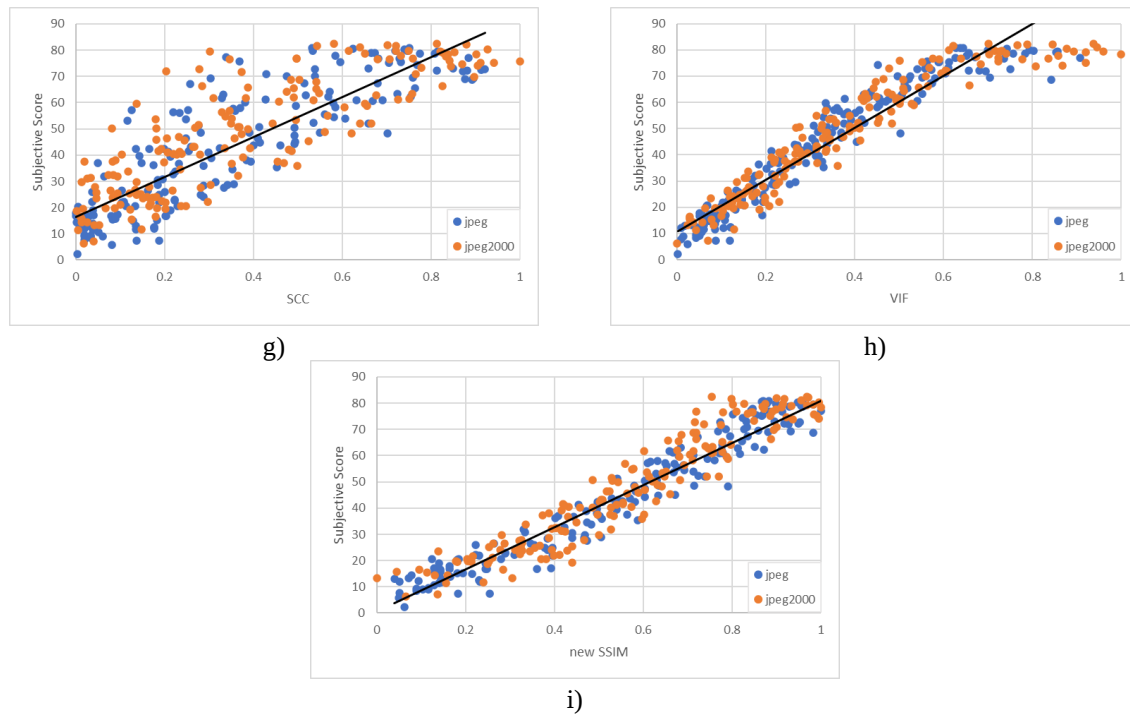


Figure 6. Scatter plot of mean subjective score versus model prediction. g) SCC. h) VIF. i) Proposed SSIM

Figure 6 shows a scatter graph depicting the relationship between each model and subjective scores. The results are normalized for each model to focus on the distribution pattern, and linear regression lines are plotted for both JPEG (blue) and JPEG2000 (orange) compression to identify the best fit. MSE and ERGAS exhibit a negative correlation, reflecting error-based assessment methods, while CW-SSIM, MSE, PSNR, ERGAS, and SCC show significant deviations from the regression line, indicating high variance and inconsistency. In contrast, SSIM, MS-SSIM, and VIF exhibit more consistent trends, although their distributions are non-linear, with SSIM and MS-SSIM concentrated at higher values and VIF at lower values. The proposed SSIM shows a linear and consistent distribution relative to human subjective scores.

Table 2. Human Correlation and Latency of each model.

Image Quality Assessment	Human Correlation	Latency
SSIM	0.926297575	52.74674
MS-SSIM	0.894578	1471.707
CW-SSIM	0.766724	2162.616
MSE	-0.785227	22.33088
PSNR	0.798379	25.0241
ERGAS	-0.806139	35.68067
SCC	0.843961	698.7607
VIF	0.946512	2913.831
Proposed SSIM	0.960861	59.49695

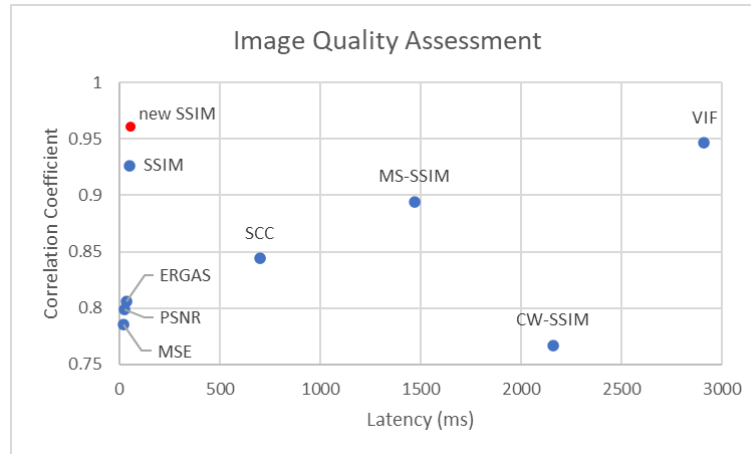


Figure 7. The image quality assessment of each model

Table 2 shows the correlation coefficients between image assessment methods and human subjective scores, alongside their corresponding latency in milliseconds. The correlation coefficient measures how closely each model aligns with human perception, with higher values indicating better alignment. Latency reflects the computational complexity of the algorithm, with lower values indicating faster performance. Figure 7. visualizes the relationship between correlation and latency for each method. CW-SSIM, MSE, PSNR, ERGAS, and SCC show moderate correlation coefficients between 0.766724 and 0.843961, while the proposed SSIM, original SSIM, MS-SSIM, and VIF have very strong correlations above 0.894578, indicating better alignment with human perception. However, MS-SSIM, CW-SSIM, SCC, and VIF have poor latency, exceeding 1000 milliseconds, while the proposed SSIM, original SSIM, MSE, PSNR, and ERGAS perform faster, with latency between 22.33 and 59.5 milliseconds. The proposed SSIM outperforms others with the highest correlation coefficient and relatively low latency, making it well-suited for real-time applications like fast object tracking systems.

4.2 Tracking System

In this study, we benchmarked our proposed online tracking system against established systems like SORT [41] and DeepSORT [42]. SORT is a traditional approach that uses data association techniques and a Kalman filter for state estimation, while DeepSORT enhances SORT by integrating deep learning to improve the association process. All systems tested follow a two-stage process, with detection and tracking happening independently, using the YOLOv6-Nano model [43] for detection. The evaluation focused on tracking precision, recall, mean Average Precision (mAP), and Frames Per Second (FPS), with a correct track prediction defined by an Intersection over Union (IoU) greater than 0.5. Precision, recall, and mAP measure accuracy and consistency, while FPS assesses computational efficiency. The benchmarking was conducted using the PETS09 sample from the MOT15 dataset [44], containing 795 frames of surveillance footage.

Table 3. Precision, recall, mAP and FPS of each online tracking model

Online Tracking	Precision	Recall	mAP	FPS
SORT	0.975984508	0.548148374	0.53754941	12.73
DeepSORT	0.85818151	0.683788203	0.618688053	3.01
Tracking+SSIM	0.845935008	0.713357298	0.641054745	11.23
Tracking+new SSIM	0.86661092	0.75536197	0.67885317	11.15

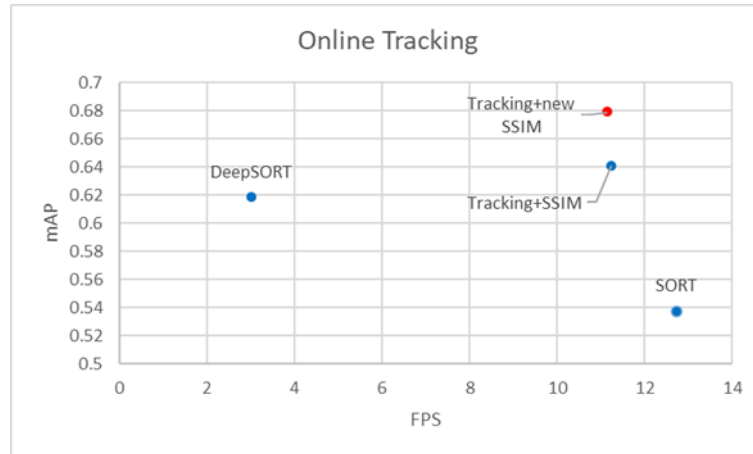


Figure 8. mAP vs FPS of each Online Tracking model

Table 3 presents the performance breakdown of various tracking models, highlighting their precision, recall, mAP, and FPS metrics. Figure 8. illustrates the trade-off between mAP and FPS for each model. SORT achieves high precision, meaning it correctly predicts detections most of the time, but its recall is lower, indicating some true positive detections are missed. Its mAP score is moderate, and it operates at 12.73 FPS. DeepSORT offers a better balance between precision and recall, with higher recall and mAP compared to SORT, though its FPS drops to 3.01 due to its deep learning-based approach.

The proposed tracking system using SSIM and new SSIM outperforms both SORT and DeepSORT in recall and mAP, achieving a strong balance between precision and recall, with significantly higher recall than DeepSORT. It also maintains high mAP scores, surpassing both SORT and DeepSORT. The system operates efficiently with frame rates of 11.23 FPS (SSIM) and 11.15 FPS (new SSIM), making it suitable for real-time processing. The new SSIM is recommended for better accuracy in analyzing appearance patterns, making the proposed tracking system a promising solution with a good performance in accuracy and at a high speed for object tracking.

4.3 Re-Identification System

The global re-identification system was evaluated using the Market-1501 dataset [45], which includes 1501 identities captured across six cameras and was selected for its similarity to Malaysian environments in terms of lighting and clothing. Evaluation metrics included the Confusion Matrix [46], ROC curve [47], F1 score, and Precision-Recall curve to measure the system's accuracy. A key component of the system is the colour proportion of each element, which serves as the primary identifier for re-identification. Using CIELUV clustering and Luv-Similarity, the system compares the target image to the dataset based on colour distribution, further analyzing the result using image moments to characterize the appearance.

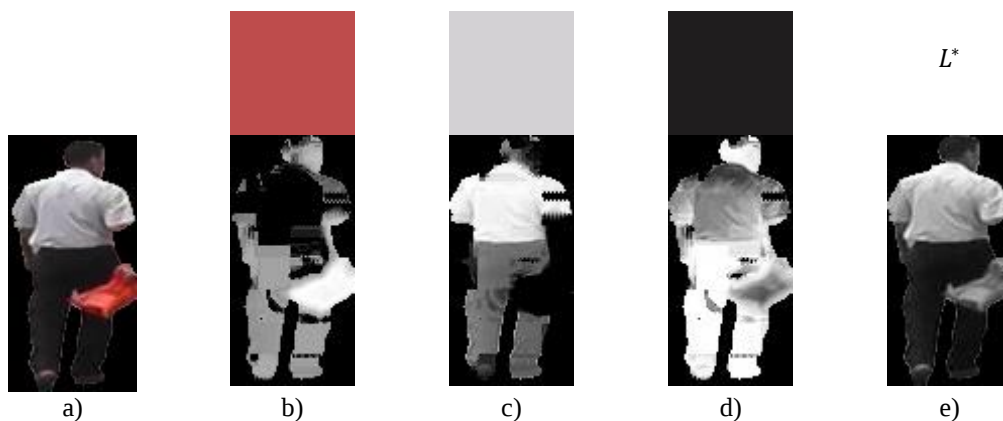


Figure 9. Input image and colour distribution channels. a) Input masked image. b) 1st colour channel. c) 2nd colour channel. d) 3rd colour channel. e) 4th colour channel (lightness).

Table 4. Example of colour distribution and its moments

	Colour Distribution Channel			
Image Moment				
μ	0.40584159	0.47837186	0.73900217	0.3627581
σ	0.30421361	0.30151013	0.21812187	0.25913531
μ_y	0.57992072	0.39907478	0.52449287	0.35496744
σ_y	0.17275193	0.18268655	0.20082745	0.15699626

Figure 9 b), c), d), and e) illustrate the distribution of individual colour elements in the target image, demonstrating that the algorithm effectively segments and assigns appropriate weights to the colours while also highlighting their spatial positions. Additionally, the lightness component, L^* , is incorporated into the colour distribution channels to address grayscale insensitivity and enhance the system's overall accuracy.

The colour distribution of an appearance can be simplified into statistical values using image moments, where the first (μ) and second (σ) order moments describe the mean and spread of colour similarity. Table 4. presents normalized moment values for overall and vertical (y-direction) data, showing that lower μ indicates smaller areas of similarity, and lower σ reflects more concentrated similarity. Similarly, μ_y and σ_y provide insights into vertical distribution. Appearances with more colour groups produce more moment variables, improving re-identification accuracy. To identify a target, its moment values are compared against those in the dataset using Euclidean distance and Cosine similarity, ensuring that shared appearances yield nearly identical colour distributions.

Table 5. Distance between the sample appearance and the target appearance (Figure 9.)

	Sample Appearance			
Similarity Scores				
d_{EC}	0.750185782	0.538106357	0.511760531	0.555320593

Table 5 presents the colour distribution similarity scores, d_{EC} , derived from both Euclidean and Cosine similarity metrics, comparing the target appearance (Figure 9.) with selected dataset samples. The findings reveal that samples with a high visual resemblance to the target yield the highest similarity scores, up to 0.75, while less similar appearances score lower, typically between 0.51 and 0.55, indicating notable differences in colour distribution.

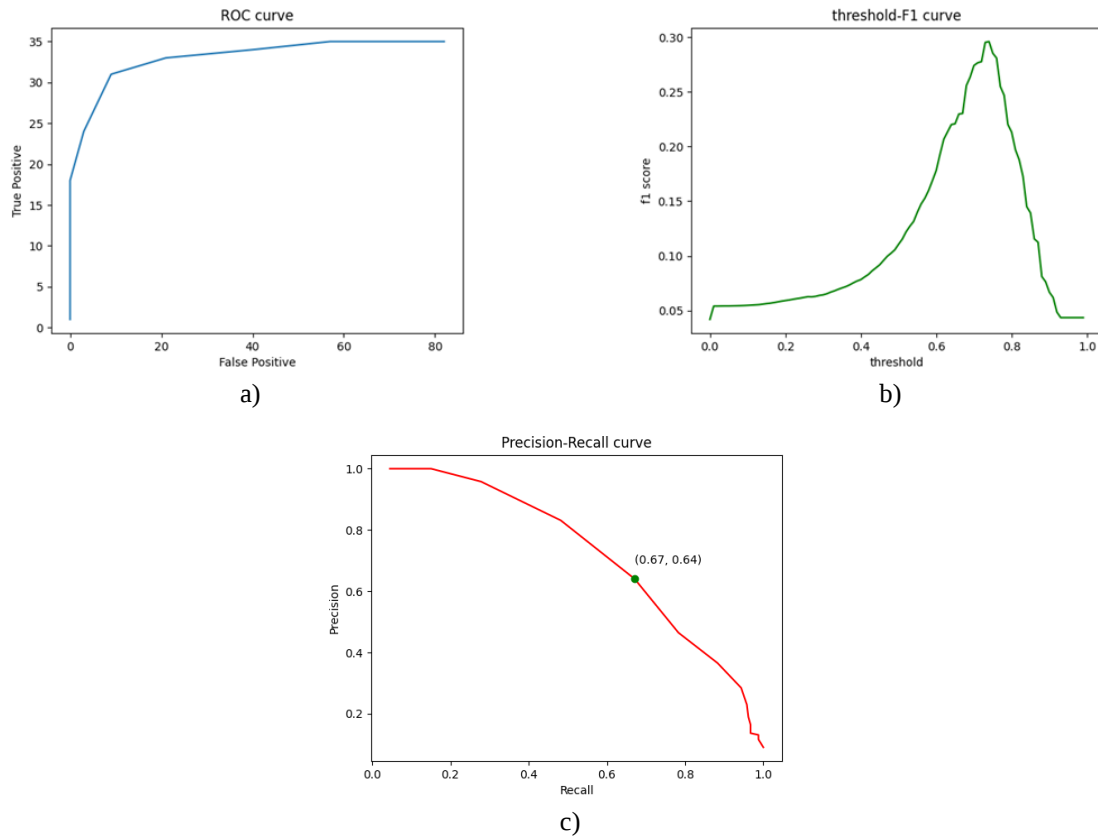


Figure 10. a) ROC curve b) Threshold-F1 curve c) Precision-Recall curve

The receiver operating characteristic (ROC) curve evaluates a binary classification system's ability to distinguish between classes by plotting the true positive rate against the false positive rate across various thresholds. As shown in Figure 10. a), the ROC curve for a single target appearance in our re-identification system bends toward the top-left corner, indicating effective performance and non-random classification. To complement this, precision-recall curves and F1 scores were also analyzed for multiple appearances, as the ROC curve alone offers limited insight into model performance. Precision and recall were derived from the Confusion Matrix, and F1 scores were calculated at different thresholds to identify the optimal range for accurate re-identification.

Figure 10. b) shows that the F1 score of the tested appearance peaks at a threshold of around 0.71, which was found to provide the best re-identification results across most appearances in the dataset. Thresholds above this value make the system overly sensitive, increasing false positives, while lower thresholds result in more matches but with reduced accuracy. Figure 10. c) presents the trade-off between precision and recall, highlighting the optimal balance point at 0.64 precision and 0.67 recall. This means the system correctly identifies 67% of the relevant appearances and that 64% of the results are accurate, though some may include visually similar identities due to the limitations of colour-based recognition.

To evaluate overall performance, Average Precision (AP) is used to condense the precision-recall curve into a single value by integrating the area under the curve. In this study, the mean Average Precision (mAP) calculated across multiple identities is approximately 0.6849, indicating solid system performance in handling appearance-based re-identification. The system also demonstrates strong computational efficiency, processing over 1500 appearances (totaling 50.5 MB) in 15.7 seconds. This translates to a speed of approximately 95.5 appearances per second or 3.2 MB per second, confirming the system's suitability for real-time or near-real-time applications.

4.4 Comparison of Re-Identification

In this evaluation, the proposed global re-identification system is benchmarked against two other models: a neural network-based Siamese model [48] and a traditional approach using dominant colour histograms. Deploying publicly available deep learning-based re-identification models locally proved challenging due to dependencies and training data requirements. As a result, the comparison is limited to one deep learning model and one traditional method. The Siamese network was designed to learn discriminative human features for comparison, while the traditional method relied on colour histograms segmented by human body parts. All models were tested using the Market-1501 dataset, focusing on

identifying five specific identities. Performance was assessed using precision-recall curves, mean Average Precision (mAP), and the time required to identify all targets.

Since the authors of the Siamese network did not provide a pretrained model, it was trained on the Market-1501 dataset before evaluation. Both the proposed method and the colour histogram approach require background subtraction using foreground masks, which are not available in the Market-1501 dataset. To overcome this, a human segmentation module [49] was used to generate the necessary masks. This comprehensive experimental setup ensures a fair comparison across all models and provides meaningful insight into the effectiveness and efficiency of the proposed system relative to existing alternatives.

Table 6. Precision, recall, mAP and duration of each Re-identification model

Re-Identification	Precision	Recall	mAP	Duration
Siamese Networks	0.53	0.81	0.31346265	134.169832
Colour Histogram	0.74	0.5	0.56897762	1.05179834
Proposed	0.64	0.67	0.68492539	15.7278854

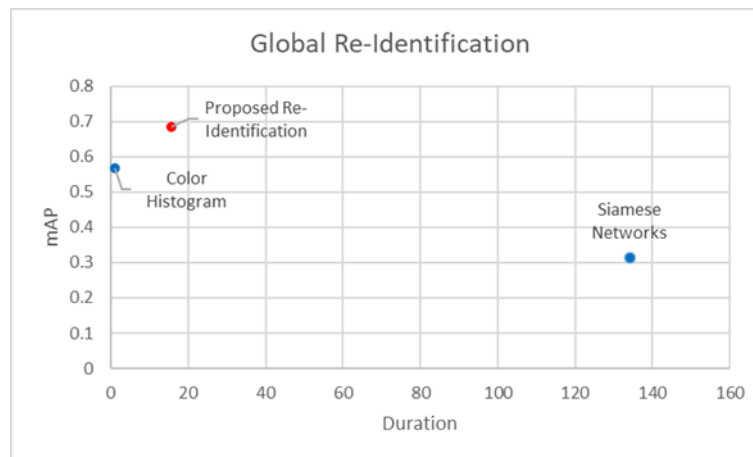


Figure 11. mAP vs duration of each Re-identification model

Table 6 and Figure 11 compare the performance of three re-identification systems in terms of precision, recall, mean Average Precision (mAP), and processing time. The Siamese Network model shows high recall but low precision, suggesting it detects many true identities but also includes more false positives, with a relatively low mAP and a high processing time of 134.17 seconds. The Colour Histogram model has higher precision and a faster runtime of just 1.05 seconds but lower recall, indicating it misses some targets. In contrast, the proposed re-identification system achieves a strong balance between precision and recall, the highest mAP, and a moderate processing time of 15.73 seconds, suggesting superior accuracy and efficiency compared to both alternatives.

5. Conclusion

The original SSIM was designed to align with Weber's law, suggesting that people notice changes relative to background brightness. However, later studies [27] found that SSIM's luminance component is too sensitive at low values and not responsive enough at high values, which doesn't match human perception. While many improvements have been made to SSIM, most still use the same flawed math for luminance and contrast. Replacing these with relative difference equations solves this issue, making the new SSIM better reflect how humans see changes [50]. This improved version not only matches human ratings more closely but also runs fast, making it ideal for real-time tasks in computer vision, like image recognition or object detection. It could also be used as a loss function in AI models that generate images, helping them produce more realistic results.

This research introduces special formulas to measure color similarity while minimizing the effect of brightness changes, assuming that the color appearance of an object remains constant. The perceptually uniform CIELAB space was chosen due to its more linear color movement in different lighting conditions. The proposed reference distance [51] normalizes the gradient distance in CIELAB, and angular distance adjusts the hue distance based on psychological theories of color perception. The results show that Luv-Similarity excels in maintaining color consistency under varying brightness and accurately distinguishes between different colors, making it ideal for color-based re-identification in environments with fluctuating lighting. It could also have practical applications in industries or medical image analysis where color patterns are essential for diagnosis, observation, or quality control.

Our aims are to develop an intelligent monitoring system for tracking and re-identifying individuals in surveillance footage based on their attributes. The system uses a traditional computer vision approach to reduce computational load while automatically alerting authorities to suspicious activities. Key components include object tracking, background subtraction, image moments, and the Hungarian algorithm. The system integrates an online human tracking feature, which monitors individuals across different camera perspectives, and a global re-identification system that compares colour patterns using the proposed Luv-Similarity method, which reduces lighting effects. The proposed SSIM enhances image quality analysis by addressing biases in the original formulation, offering improved accuracy and high-speed performance. The results demonstrate that the proposed system achieves high accuracy in tracking and re-identification, outperforming both traditional and deep learning-based methods while maintaining computational efficiency.

Acknowledgment

The authors would like to thank the Ministry of Higher Education and a very special thanks go to the Fundamental Research Grant Scheme that has been supported this research (project no. - R.J130000.7851.5F586).

References

- [1] A. Yilmaz, O. Javed, and M. Shah, 'Object tracking: A survey', *Acm computing surveys (CSUR)*, vol. 38, no. 4, pp. 13-es, 2006.
- [2] C. J. Veenman, M. J. T. Reinders, and E. Backer, 'Resolving motion correspondence for densely moving points', *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 23, no. 1, pp. 54–72, 2001.
- [3] S. Becker, R. Hug, W. Hübner, and M. Arens, 'Center point-based feature representation for tracking', in *Artificial Intelligence for Security and Defence Applications*, 2023, vol. 12742, pp. 101–109.
- [4] 'Kernel-based object tracking', *IEEE Transactions on pattern analysis and machine intelligence*, vol. 25, no. 5, pp. 564–577, 2003.
- [5] Q. Zhao and H. Tao, 'Object tracking using color correlogram', in *2005 IEEE International Workshop on Visual Surveillance and Performance Evaluation of Tracking and Surveillance*, 2005, pp. 263–270.
- [6] W. Pei and X. Lu, 'Moving object tracking in satellite videos by Kernelized Correlation Filter based on Color-Name features and Kalman Prediction', *Wireless Communications and Mobile Computing*, vol. 2022, pp. 1–16, 2022.
- [7] N. Wojke, A. Bewley, and D. Paulus, 'Simple Online and Realtime Tracking with a Deep Association Metric', *arXiv [cs.CV]*. 2017.
- [8] P. R. Gunjal, B. R. Gunjal, H. A. Shinde, S. M. Vanam, and S. S. Aher, 'Moving object tracking using kalman filter', in *2018 International Conference On Advances in Communication and Computing Technology (ICACCT)*, 2018, pp. 544–547.
- [9] H. W. Kuhn, 'The Hungarian method for the assignment problem', *Naval research logistics quarterly*, vol. 2, no. 1–2, pp. 83–97, 1955.
- [10] K. Jang, S. Han, and I. Kim, 'Person re-identification based on color histogram and spatial configuration of dominant color regions', *arXiv preprint arXiv:1411. 3410*, 2014.
- [11] T. Matsukawa, T. Okabe, E. Suzuki, and Y. Sato, "Hierarchical gaussian descriptors with application to person re-identification," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 42, no. 9, pp. 2179–2194, 2020.
- [12] W. Huang and L. Mei, "The combination of features extracted from different parts for person re-identification," *2018 2nd International Conference on Robotics and Automation Sciences (ICRAS)*, 2018.
- [13] M. Li, X. Zhu, and S. Gong, "Unsupervised tracklet person re-identification," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 42, no. 7, pp. 1770–1782, 2020.
- [14] C. Zhao, Z. Zhang, J. Yan, and Y. Yan, "Local-global feature for video-based one-shot person re-identification," *ICASSP 2020 - 2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2020.
- [15] H. Chen, Y. Wang, Y. Shi, K. Yan, M. Geng, Y. Tian, and T. Xiang, "Deep Transfer Learning for person re-identification," *2018 IEEE Fourth International Conference on Multimedia Big Data (BigMM)*, 2018.
- [16] Y. Zhang, C. Wang, X. Wang, W. Zeng, and W. Liu, 'Fairmot: On the fairness of detection and re-identification in multiple object tracking', *International Journal of Computer Vision*, vol. 129, pp. 3069–3087, 2021.
- [17] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE Transactions on Image Processing*, vol. 13, no. 4, pp. 600–612, 2004.
- [18] Z. Wang and A. C. Bovik, 'A universal image quality index', *IEEE Signal Processing Letters*, vol. 9, no. 3, pp. 81–84, 2002.
- [19] Z. Wang, E. P. Simoncelli, and A. C. Bovik, 'Multiscale structural similarity for image quality assessment', in *The Thirty-Seventh Asilomar Conference on Signals, Systems & Computers*, 2003, 2003, vol. 2, pp. 1398–1402.

- [20] M. P. Sapat, Z. Wang, S. Gupta, A. C. Bovik, and M. K. Markey, 'Complex wavelet structural similarity: A new image similarity index', *IEEE transactions on image processing*, vol. 18, no. 11, pp. 2385–2401, 2009.
- [21] T. M. Cover, J. A. Thomas, and Others, 'Entropy, relative entropy and mutual information', *Elements of information theory*, vol. 2, no. 1, pp. 12–13, 1991.
- [22] D. Lee and S. Lim, "Improved structural similarity metric for the visible quality measurement of images," *Journal of Electronic Imaging*, vol. 25, no. 6, p. 063015, 2016.
- [23] M. A. Aljanabi, Z. M. Hussain, N. A. Shnain, and S. F. Lu, "Design of a hybrid measure for image similarity: A statistical, algebraic, and information-theoretic approach," *European Journal of Remote Sensing*, vol. 52, no. sup4, pp. 2–15, 2019.
- [24] R. Hassen and E. Steinbach, "HSSIM: An objective haptic quality assessment measure for force-feedback signals," *2018 Tenth International Conference on Quality of Multimedia Experience (QoMEX)*, 2018.
- [25] I. Bakurov, M. Buzzelli, M. Castelli, R. Schettini, and L. Vanneschi, "Parameters optimization of the structural similarity index," *London Imaging Meeting*, vol. 1, no. 1, pp. 19–23, 2020.
- [26] A. K. Venkataramanan, C. Wu, A. C. Bovik, I. Katsavounidis, and Z. Shahid, "A hitchhiker's guide to structural similarity," *IEEE Access*, vol. 9, pp. 28872–28896, 2021.
- [27] J. Nilsson and T. Akenine-Möller, 'Understanding ssim', *arXiv preprint arXiv:2006.13846*, 2020.
- [28] A. J. Hudspeth, T. M. Jessell, E. R. Kandel, J. H. Schwartz, and S. A. Siegelbaum, *Principles of neural science*. McGraw-Hill, Health Professions Division, 2013.
- [29] J. Schwiegerling, "Field guide to visual and ophthalmic optics", 2004.
- [30] A. J. Hudspeth, T. M. Jessell, E. R. Kandel, J. H. Schwartz, and S. A. Siegelbaum, *Principles of neural science*. McGraw-Hill, Health Professions Division, 2013.
- [31] M. Li, X. Zhu, and S. Gong, "Unsupervised tracklet person re-identification," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 42, no. 7, pp. 1770–1782, 2020.
- [32] C. Zhao, Z. Zhang, J. Yan, and Y. Yan, "Local-global feature for video-based one-shot person re-identification," *ICASSP 2020 - 2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2020.
- [33] H. Chen, Y. Wang, Y. Shi, K. Yan, M. Geng, Y. Tian, and T. Xiang, "Deep Transfer Learning for person re-identification," *2018 IEEE Fourth International Conference on Multimedia Big Data (BigMM)*, 2018.
- [34] Y. Zhang, C. Wang, X. Wang, W. Zeng, and W. Liu, 'Fairmot: On the fairness of detection and re-identification in multiple object tracking', *International Journal of Computer Vision*, vol. 129, pp. 3069–3087, 2021.
- [35] B. Zohuri and M. Moghaddam, 'Deep Learning Limitations and Flaws', *Modern Approaches on Material Science*, vol. 2, 01 2020.
- [36] L. Wald, 'Quality of high resolution synthesised images: Is there a simple criterion?', in *Third conference "Fusion of Earth data: merging point measurements, raster maps and remotely sensed images"*, 2000, pp. 99–103.
- [37] D. L. C. J. Zhou and J. A. Silander, 'A wavelet transform method to merge Landsat TM and SPOT panchromatic data', *International Journal of Remote Sensing*, vol. 19, no. 4, pp. 743–757, 1998.
- [38] H. R. Sheikh and A. C. Bovik, 'Image information and visual quality', *IEEE Transactions on Image Processing*, vol. 15, no. 2, pp. 430–444, 2006.
- [39] A. Khalel, "Andrewekhalel/Sewar: All image quality metrics you need in one package,," *GitHub*, <https://github.com/andrewekhalel/sewar> (accessed Nov. 28, 2023).
- [40] "Laboratory for Image & Video Engineering," *ut*. [Online]. Available: <http://live.ece.utexas.edu/research/quality/>. [Accessed: 12-Feb-2023].
- [41] A. Bewley, Z. Ge, L. Ott, F. Ramos, and B. Upcroft, 'Simple online and realtime tracking', in *2016 IEEE International Conference on Image Processing (ICIP)*, 2016.
- [42] N. Wojke, A. Bewley, and D. Paulus, 'Simple Online and Realtime Tracking with a Deep Association Metric', *arXiv [cs.CV]*. 2017.
- [43] C. Li et al., 'YOLOv6: A single-stage object detection framework for industrial applications', *arXiv preprint arXiv:2209.02976*, 2022.
- [44] J. Ferryman and A. Shahrokni, 'PETS2009: Dataset and challenge', in *2009 Twelfth IEEE International Workshop on Performance Evaluation of Tracking and Surveillance*, 2009, pp. 1–6.
- [45] "Papers with code - market-1501 dataset," *Dataset | Papers With Code*. [Online]. Available: <https://paperswithcode.com/dataset/market-1501>. [Accessed: 12-Feb-2023].
- [46] K. M. Ting, 'Confusion Matrix', in *Encyclopedia of Machine Learning and Data Mining*, 2010.
- [47] A. P. Bradley, 'The use of the area under the ROC curve in the evaluation of machine learning algorithms', *Pattern Recognition*, vol. 30, no. 7, pp. 1145–1159, 1997.
- [48] S. Wang, S. Wu, L. Duan, C. Yu, Y. Sun, and J. Dong, 'Person Re-Identification with Deep Features and Transfer Learning', in *2017 IEEE International Conference on Computational Science and Engineering (CSE) and IEEE International Conference on Embedded and Ubiquitous Computing (EUC)*, 2017, vol. 1, pp. 704–707.
- [49] X. Chen et al., 'Robust Human Matting via Semantic Guidance', *arXiv [cs.CV]*. 2022.

- [50] M. I. Jaafar, S. W. Nawawi, and R. A. Rahim, 'Improving measurement bias of structural similarity index (ssim) using absolute difference equation', *Applications of Modelling and Simulation*, vol. 6, pp. 10–19, 2022.
- [51] M. I. Jaafar, S. W. B. Nawawi, and R. A. Rahim, 'Relative Measure of Digital Colour by Utilizing the Linearity of Cieluv Space', Available at SSRN 4365843.