Virtual Reality
https://doi.org/10.1007/s10055-021-00519-0

ORIGINAL ARTICLE

An improved colour binary descriptor algorithm for mobileaugmented reality

Siok Yee Tan[1] · Haslina Arshad[1] · Azizi Abdullah[1]

Received: 23 September 2020 / Accepted: 18 March 2021

Abstract

The incorporation of both virtual content and real world in augmented reality (AR) allows real-time engagement with the virtual objects. The selection of an appropriate tracking algorithm is important to optimise the performance of mobile AR applications given the limited processing capabilities and memories of mobile devices like smartphones. Tracking in AR consists of four essential components, namely detector, descriptor, matcher, and pose estimator. Since a descriptor substantially affects the overall performance of a mobile AR application, it must have short computational time and remains invariant to scale, rotation, and lighting changes. Studies have proposed Fast Retina Keypoint (FREAK) descriptor as the most suitable descriptor for mobile AR applications. Unlike other greyscale descriptors, FREAK has shorter computational time and is less likely to be affected by scale and rotation changes. However, it overlooks the vital colour space information. Focusing on enhancing the efficiency and robustness of FREAK, this study proposed the use of CRH-FREAK (RGB + HSV) descriptor and applied the vertical concatenation technique that combined all extracted keypoints vertically. The robustness of the proposed descriptors against scale, rotation, and lighting changes was verified using Mikolajczyk and Amsterdam Library of Object Images (ALOI) datasets. The developed CRH-FREAK descriptors used six colour spaces to describe the keypoints, which made them slower than the original FREAK. However, the size reduction of CRH-FREAK from 512 bits to 128 bits in this study successfully reduced the computational time to 29.49 ms, which was found comparable to the original FREAK. The improved efficiency and robustness of a 128-bit CRH-FREAK descriptor benefit the future development of mobile AR applications that remain invariant to scale, rotation, and lighting changes.

Keywords Colour descriptor · FREAK · HSV · Mobile augmented reality · RGB

1 Introduction

Mobile devices, including smartphones, are powerful tools for augmented reality (AR) (Flavián et al. 2019; Lam et al. 2020a) due to their ability to facilitate data access and processing with a large amount of computing power (Sung et al. 2016). Unlike personal computers, there is a different market for smartphones in AR applications because they are equipped with a built-in camera that is capable of running computer vision process. For an effective operation of an AR application, high-speed tracking of lower than 0.1 s is required to track and register the position of the device or user in reference to the real world (Tan et al. 2019). Apart from that, the tracking algorithm must be invariant to scale, rotation, and lighting changes (Bleser 2009; Mahieu and Tilak 2015; Satoh et al. 2001). Therefore, the selection of tracking algorithms has to be carefully made in order to achieve optimised performance of mobile AR applications, as smartphones have lower memory and processing power than personal computers.

Accordingly, the tracking process for an AR application involves four key steps (Obeidy et al. 2013). As illustrated in Fig. 1, the tracking process first involves detecting features in the greyscale images using a detector algorithm. These greyscale images are converted from the captured camera image (using the camera of the smartphone) and the reference image stored in the database image. The second step involves describing essential keypoints detected from both greyscale images using a feature descriptor. The detected keypoints from the reference image are stored in

✉ Siok Yee Tan
esther@ukm.edu.my

1 Center for Artificial Intelligence and Technology, Faculty of Information Science and Technology, Universiti Kebangsaan Malaysia, 43600 Bangi, Selangor, Malaysia

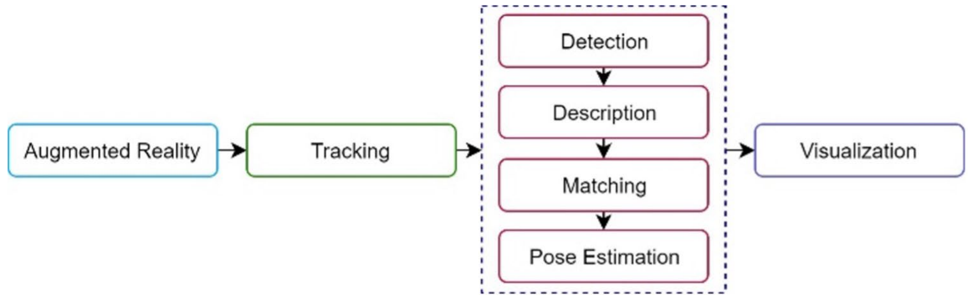


Fig. 1 Tracking process in AR application

the database. Following that, the keypoints from the input image are matched with the keypoints stored in the database. Lastly, pose estimation, which determines the position of the virtual objects to be superimposed above the input image, is conducted if the matching accuracy exceeds the threshold; otherwise, the detection process repeats. A completed tracking process produces a successfully augmented 3D virtual object above the (real-time) input image in the right orientation (Tan et al. 2019; Uchiyama and Marchand 2012).

Meanwhile, the tracking process in an AR application is generally measured in terms of two aspects; **robustness** reflects the matching accuracy of the detected keypoints between an input image and reference image with the consideration of scale, rotation, and lighting changes, whereas **efficiency** reflects the capacity to extract and match the detected keypoints between an input image and reference image in the shortest time (Tan et al. 2019). To date, the development of a robust and efficient AR application in both desktop and mobile platforms has been extensively explored (Cruz et al. 2019; Lam et al. 2020b; Majid and Majid 2018; Tan et al. 2018; Yang et al. 2019). The robustness and efficiency of mobile AR applications in the presence of various changes are clearly important and should be critically considered (Ihsan and Sehat 2013).

All tracking algorithms, including the detector, descriptor, and matcher used in a tracking process, directly affect the performance of a mobile AR application. However, feature descriptors have been the main focus of most studies (Koyasu et al. 2014; Tan et al. 2019; Zhang et al. 2018) given its direct influence on the robustness and efficiency of mobile AR applications. A feature descriptor for a mobile AR application must be able to function at high speed and remains invariant to scale, rotation, and lighting changes. Studies have proposed Fast Retina Keypoint (FREAK) descriptor as the most suitable descriptor for a mobile AR given its shorter operating time (as compared to other binary descriptors) and robustness against various changes (Tan et al. 2019). With that, the current study focused on the FREAK descriptor algorithm.

However, descriptors, including FREAK, overlook vital colour space information and can only extract features in greyscale images (Alahi et al. 2012; Bay et al. 2008; Calonder et al. 2012; Leutenegger et al. 2011; Lowe 2004; Rublee et al. 2011). Therefore, studies in other domains, such as 2D ear recognition, video concept tracking, and image recognition, have placed greater emphasis on the robustness of descriptors in the presence of lighting changes (Markatopoulou et al. 2015; Tang et al. 2019; Van De Sande et al. 2010; Zhou et al. 2011). Addressing the need to improve the robustness of these applications against lighting changes, there have been continuous efforts to develop a descriptor that makes use of RGB or HSV colour space. Colour information such as RGB and HSV may provide important information or features in the description (Hua et al. 2019; Markatopoulou et al. 2015). Moreover, the use of RGB and/or HSV colour space descriptors in underwater image recognition, 2D ear recognition, video concept tracking, and image recognition applications has demonstrated good robustness (Markatopoulou et al. 2015; Van De Sande et al. 2010; Zhou et al. 2011).

For instance, Kottman (2011) made use of RGB colour space for the Binary Robust Independent Elementary Features (BRIEF) descriptor (RGB BRIEF) and found that the descriptor increased the robustness of the image recognition application (as compared to the original BRIEF descriptor) in the presence of lighting changes. However, the robustness of the descriptor against scale and rotation changes for mobile AR applications were never tested since the descriptor was not designed for such changes. In another study, Wafy and Madbouly (2015) made use of RGB colour space for the Speeded Up Robust Features (SURF) descriptor (RGB SURF) for an image recognition application and found improved performance in the extraction of red, green, and blue images (as compared to the original SURF descriptor). This study applied the vertical concatenation technique to combine all extracted keypoints, which was considered for the current study. Meanwhile, Markatopoulou et al. (2015) combined RGB and *Opponent* in SURF and Oriented FAST and Rotated BRIEF (ORB) descriptors for video concept tracking applications. The combination of two colour spaces in this study led to improved robustness, as compared to the use of only one colour space, which was also considered for the current study. However, the use of colour space descriptors in recent studies was revealed to increase the computation time, as compared to the use of greyscale descriptors (Fan et al. 2009; Kottman 2011; Markatopoulou et al. 2015; Wu et al. 2013). Calonder et al. (2012) highlighted that reducing the size of the descriptor from 512 bits to 128 bits can speed up the computational time of the descriptor without affecting its robustness. Therefore, the current study considered lowering the size of the descriptor for shorter computational time using the BRIEF method.

This paper is organized into eight sections which include Sect. 1 as introduction. The following Sect. 2 explained the key challenges of this study. The overview of the descriptors is described in Sect. 3. Section 4 and Sect. 5 determined the development of the proposed algorithm. Comparative results of each descriptors are shown in Sect. 6, and the overall results are discussed in Sect. 7. Finally, the paper is concluded in Sect. 8.

2 Key challenges

The selection of an appropriate descriptor for the tracking process is very important given its significant influence on the performance of mobile AR applications (Bolyós 2013; Guan et al. 2012; Nguyen et al. 2016; Ufkes and Fiala 2013). Most importantly, a descriptor for a mobile AR application must have a short operating time (Azuma 1997) and remain invariant to scale, rotation, and lighting changes (Bleser 2009; Mahieu and Tilak 2015; Satoh et al. 2001). Unlike other descriptors, FREAK has been recognised as the most robust and efficient descriptor (Tan et al. 2019). However, it overlooks important colour information and describes the keypoints in greyscale (Alahi et al. 2012).

Colour information such as RGB and HSV provides important information for the description process, which can increase the robustness of the descriptor. The use of only one colour space for descriptors in numerous image recognition applications has demonstrated promising results in terms of robustness (Bernal et al. 2010; Bianco et al. 2015; Kottman 2011). Despite that, there are descriptors such as RGB SIFT and HSV SIFT that are not suitable for mobile AR applications, as such descriptors take up longer computational time (Andono et al. 2014; Zhou et al. 2011). The use of colour space causes the descriptor to perform slower than greyscale descriptors—such attribute is not suitable for mobile AR applications (Fan et al. 2009; Kottman 2011; Markatopoulou et al. 2015; Wu et al. 2013). For instance, descriptors that make use of RGB and HSV colour spaces need six colour spaces to describe the keypoints, resulting in longer computational time (as compared to descriptors that make use of greyscale). In another study, Markatopoulou et al. (2015) attempted to combine two colour spaces, namely RGB and *Opponent*, in SURF and ORB descriptors, for a video concept tracking application. Although the results showed that the use of two colour spaces led to higher robustness and better descriptive power than the use of only one colour space, the proposed RGB + *Opponent* ORB descriptor is still not suitable for mobile AR applications, as the ORB descriptor is easily affected by both scale and rotation changes (Tan et al. 2019).

Nevertheless, according to Calonder et al. (2012), reducing the size of the descriptor from 512 bits to 128 bits can significantly speed up the computational time without affecting the robustness of the descriptor. With its high-speed function and robustness against scale, rotation, and lighting changes, the proposed colour binary descriptor in the current study was expected to be a good alternative and benefit the future development of mobile AR applications.

3 Descriptors

A descriptor extracts or describes the detected keypoints. In particular, there are two types of descriptors, namely binary descriptors (e.g. FREAK, BRIEF, ORB, and Binary Robust Invariant Scalable Keypoints [BRISK]) and floating-point descriptors (e.g. SURF and Scale-Invariant Feature Transform [SIFT]). Unlike other descriptors, FREAK is regarded as the most efficient and robust descriptor (Tan et al. 2019).

3.1 FREAK descriptor

Alahi et al. (2012) proposed FREAK to improve the sampling pattern and pair selection method used by BRISK. As illustrated in Fig. 2, the FREAK sampling pattern appears like an eye retina, and each pixel generally overlaps with a higher density near the centre point. The structure of a sampling pair is similar to a coarse-to-fine approach that corresponds to a human retina model. Similar to how a human eye operates, the first sampling pair is compared to the sampling pair at the outer ring, whereas the second sampling pair is compared to the sampling pair at the inner ring. As

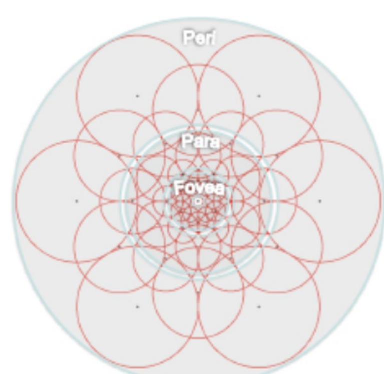


Fig. 2 FREAK sampling pattern. Source: Alahi et al. (2012)

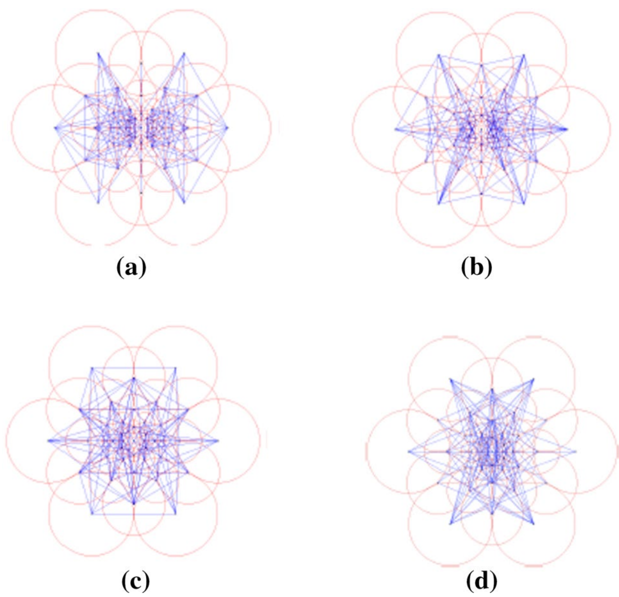


Fig. 3 Sampling pairs: **a** first cascade; **b** second cascade; **c** third cascade; **d** fourth cascade. Source: Alahi et al. (2012)

shown in Fig. 3, there are four cascades of sampling pairs. In particular, there are 128 sampling pairs (from left to right; from top to bottom) for each cascade.

Accordingly, a 512-bit FREAK consists of four cascades (128-bit per cascade). If the distance is lower than the corresponding threshold for the first cascade, FREAK proceeds to make comparisons for the subsequent cascades. In particular, the intensity between each sampling pair (P_1, P_2) is compared. Referring to Eq. 1, "1" is noted when the intensity at P_1 exceeds the intensity at P_2, while "0" is noted when the intensity at P_2 exceeds the intensity at P_1. Only 10% of the candidate features from the first cascade are retained for the comparison in the subsequent cascade. In most cases, the first cascade includes the perifoveal receptive field (the extraction of coarse information), while the fourth and final cascade includes the fovea receptive field (the extraction of fine information). A 512-bit descriptor is developed based on the sum of binary strings (Eq. 2).

$$B(p;x,y) := \begin{cases} 1: & I(p,x) < I(p,y) \\ 0: & \text{otherwise} \end{cases} \tag{1}$$

$$\sum_{1 \le i \le 512} 2^{i-1} B(p;x,\ y) \tag{2}$$

3.2 Colour descriptor

A binary descriptor describes features in greyscale images only (one channel). Most descriptors (e.g. SIFT, SURF, BRIEF, ORB, BRISK, and FREAK) omit colour information in RGB (three channels) and/or HSV (three channels) images. However, colour information such as RGB and HSV provide important information for the description process (Bernal et al. 2010; Bianco et al. 2015; Kottman 2011). Figure 4a presents an image with colour information. The same greyscale image is presented in Fig. 4b. The number "8" can be easily identified in Fig. 4a. In other words, descriptors that make use of colour information are more efficient and robust than descriptors that make use of greyscale.

3.2.1 RGB SURF

RGB SURF has been proposed for image recognition applications (Wafy and Madbouly 2015). Figure 5 presents the steps involved in the tracking process using RGB SURF. Overall, the use of RGB SURF involves detecting and describing RGB image and then matching between the two descriptors of the RGB image. In particular, RGB SURF splits the image into red, green, and blue colour spaces (I1, I2, and I3). SURF detector in RGB SURF detects important features in I1, I2, and I3, resulting in the formation of P1, P2, and P3 keypoints. These keypoints are then described by RGB SURF, which subsequently leads to the formation of vpts1, vpts2, and vpts3. The vertical concatenation technique is then used to combine P1, P2, and P3 keypoints with vpts1, vpts2, and vpts3 to generate only one P and one vpts.

Wafy and Madbouly (2015) utilised a total of 10 images from the ALOI dataset that consisted of different lighting colours (24 images), lighting directions (12 images), and lighting values (72 images). Based on the obtained results of the study, the keypoints described by RGB SURF were higher than SURF, as RGB SURF described keypoints of all three colour spaces. Besides that, the study also made use of the graffiti image from the Mikolajczyk dataset and revealed 34 matching keypoints for RGB SURF and only 17 matching keypoints for SURF. The RGB SURF descriptor that incorporated three colour spaces (red, green, and blue) using the vertical concatenation technique displayed better performance than SURF that made use of the greyscale only.

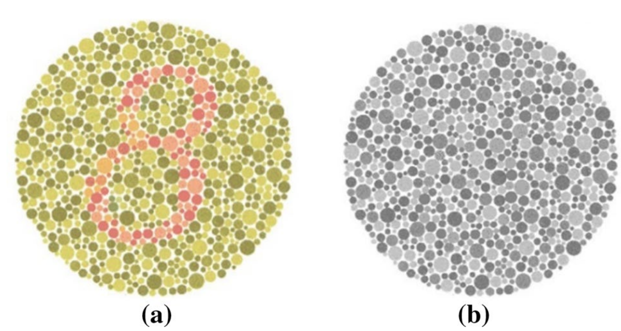


Fig. 4 Images with **a** RGB colour information and **b** greyscale information

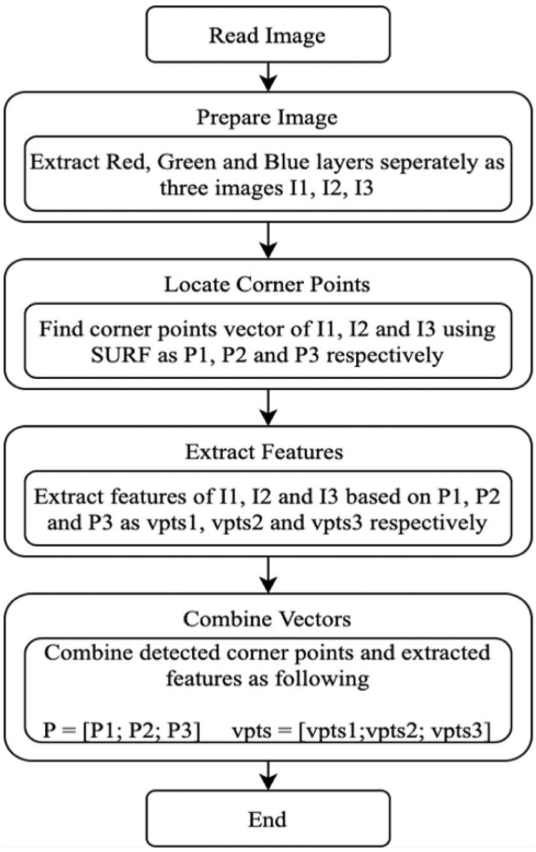


Fig. 5 Description steps used by RGB SURF. Source: Wafy and Madbouly (2015)

3.2.2 RGB SIFT and HSV SIFTRGB

SIFT and HSV SIFT have been proposed for underwater image recognition applications (Andono et al. 2014). As shown in Fig. 6, SIFT is extracted from all three channels of RGB and HSV colour spaces, which produces a 3×128 dimensions vector per descriptor to develop RGB SIFT and HSV SIFT, respectively. Based on the obtained results, HSV SIFT (9.31%) recorded the highest matching accuracy, as compared to the recorded matching accuracy of RGB SIFT (4.46%) and YCbCr SIFT (3.89%) (Andono et al. 2014). Considering that HSV colour space was not applied in the existing binary descriptor, the current study opted to apply HSV colour space for the proposed FREAK that has been widely regarded as the most suitable descriptor for mobile AR applications.

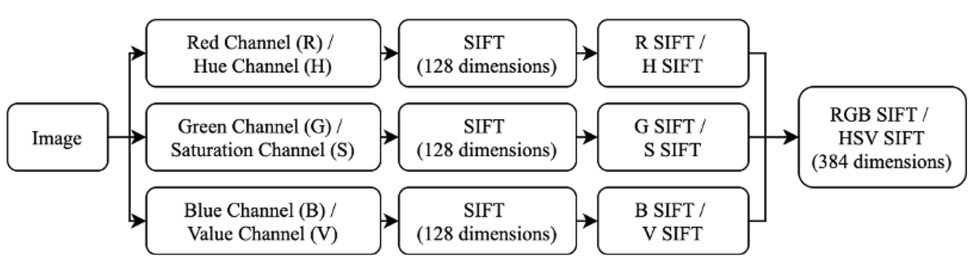
3.2.3 RGB + *Opponent* ORB

Markatopoulou et al. (2015) suggested RGB ORB, *Opponent* ORB, and RGB + *Opponent* ORB descriptors for video concept detection applications. As ORB alone overlooks colour information and only makes use of greyscale for the description of the detected images, Markatopoulou et al. (2015) incorporated RGB and *Opponent* into ORB to increase the robustness of the descriptor. There are red, green, and blue images in the RGB image; the use of ORB alone converts the RGB image into greyscale image for the description process. With the proposed RGB ORB and *Opponent* ORB, ORB is then directly applied to each channel (R, G, and B; O_1, O_2, and O_3) for each keypoint, respectively, resulting in the formation of three d-element features. These features are then concatenated into $3 * d$-element. Both RGB ORB and *Opponent* ORB are then combined to form RGB + *Opponent* ORB ($6 * d$-element).

Markatopoulou et al. (2015) tested the performance of ORB, RGB ORB, *Opponent* ORB, and RGB + *Opponent* ORB using the TRECVID 2013 Semantic Indexing (SIN) dataset (Over et al. 2013). Based on the obtained results, RGB + *Opponent* ORB (17.45%) displayed the best performance in terms of matching accuracy, as compared to ORB (11.43%), RGB ORB (13.58%), and *Opponent* ORB (12.73%). In other words, the use of two colour spaces (instead of only one colour space) and the combination of descriptors can improve the performance of the descriptor. Meanwhile, Wafy and Madbouly (2015) applied the vertical concatenation technique and proposed RGB SURF (of the same size as the generated descriptor), but found that the proposed descriptor was not suitable for mobile AR applications due to its lack of efficiency (longer computational time for the description process).

In view of the above, the current study considered combining two colour spaces (RGB and HSV) and applying the same technique (vertical concatenation technique). As for the proposed FREAK, its size was further reduced from 512 bits to 128 bits by opting for only one cascade of the descriptor. The proposed colour binary descriptor in the current study, specifically the 128-bit CRH-FREAK (RGB + HSV)

Fig. 6 Development of RGB SIFT and HSV SIFT. Source: Andono et al. (2014)



descriptor, was expected to have shorter computational time and display solid robustness against scale, rotation, and lighting (lighting colour, lighting direction, and lighting value) changes for mobile AR applications.

4 Development of 512-bit CRH-FREAK

Studies have identified FREAK as the most efficient and robust descriptor for mobile AR applications given its shortest computational time and solid robustness against scale and rotation changes (as compared to other descriptors) (Tan et al. 2019). However, this descriptor overlooks important colour information (such as RGB and HSV) and can only describe features in greyscale. The ability to make use of colour information in the description process can further improve the robustness of the descriptor (Bernal et al. 2010; Bianco et al. 2015; Kottman 2011). With that, the current study opted to make use of both RGB and HSV colour spaces in the development of a FREAK-based descriptor for mobile AR applications. A 512-bit CRH-FREAK descriptor using red, green, and blue colour spaces as well as hue, saturation, and value was first proposed for the description process.

4.1 512-bit CRH-FREAK framework

Figure 7 presents the proposed 512-bit CRH-FREAK framework. Basically, the proposed descriptor in this study incorporated a FREAK descriptor as well as RGB and HSV colour spaces (converted from both input image and reference image).

Both reference image and input image were separated into six layers, resulting in the formation of six images, respectively: red (offline tracking: IR; online tracking: IIR), green (offline tracking: IG; online tracking: IIG), blue (offline tracking: IB; online tracking: IIB), hue (offline tracking: IH; online tracking: IIH), saturation (offline tracking: IS; online tracking: IIS), and value (offline tracking: IV; online tracking: IIV).

4.2 Detector

This study used the BRISK detector to detect IR, IG, IB, IH, IS, and IV and produce red (TR), green (TG), blue (TB), hue (TH), saturation (TS), and value (TV) keypoints. All keypoints were then stored in "RedtempDetector", "GreentempDetector", "BluetempDetector", "HtempDetector", "StempDetector", and "VtempDetector", respectively.

4.3 Descriptor

Accordingly, FREAK compares the intensity of each pixel in a sampling pair, resulting in the formation of a binary string or in other words, the area around the keypoint and find each sampling pair (p_x, p_y). A binary value of "1" is set in the binary string if the intensity at the point p_y exceeds the intensity at the point p_x, and "0" is set if the comparison outcome is otherwise:

$$T(p;x,y) := \begin{cases} 1 \text{ if } I(p,x)I(p,y) \\ 0 \text{ otherwise} \end{cases} \tag{3}$$

For this study, CRH-FREAK was used to describe each keypoint (TR, TG, TB, TH, TS, and TV). Each feature was described using red, green, blue, hue, saturation, and value colour spaces. For the formation of a binary string, CRH-FREAK was used to compare the intensity of eh pixel in the red, green, blue, hue, saturation, and value colour spaces. The binary strings for CRH-FREAK were generated based on the following equations:

$$T\left(p_{\text{red}};x,y\right) := \begin{cases} 1 \text{ if } I\left(p_{\text{red}},x\right) I\left(p_{\text{red}},y\right) \\ 0 \text{ otherwise} \end{cases} \tag{4}$$

$$T\left(p_{\text{green}};x,y\right) := \begin{cases} 1 \text{ if } I\left(p_{\text{green}},x\right)I\left(p_{\text{green}},y\right) \\ 0 \text{ otherwise} \end{cases} \tag{5}$$

$$T\left(p_{\text{blue}};x,y\right) := \begin{cases} 1 \text{ if } I\left(p_{\text{blue}},x\right) I\left(p_{\text{blue}},y\right) \\ 0 \text{ otherwise} \end{cases} \tag{6}$$

$$T\left(p_{\text{hue}};x,y\right) := \begin{cases} 1 \text{ if } I\left(p_{\text{hue}},x\right) I\left(p_{\text{hue}},y\right) \\ 0 \text{ otherwise} \end{cases} \tag{7}$$

$$T\left(p_{\text{saturation}};x,y\right) := \begin{cases} 1 \text{ if } I\left(p_{\text{saturation}},x\right)I\left(p_{\text{saturation}},y\right) \\ 0 \text{ otherwise} \end{cases} \tag{8}$$

$$T\left(p_{\text{value}};x,y\right) := \begin{cases} 1 \text{ if } I\left(p_{\text{value}},x\right)I\left(p_{\text{value}},y\right) \\ 0 \text{ otherwise} \end{cases} \tag{9}$$

A 512-bit FREAK (four cascades with 128-bit per cascade) first compared the first cascade (128 bits) and proceeded to make comparisons with the subsequent cascades if the distance was found smaller than the threshold for the analysis of finer information. About 10% of the candidate features from the first cascade were retained. As illustrated in Fig. 8, CRH-FREAK described the keypoints using these four cascades. The following equations were used to add up each binary string (from Eq. 4 to Eq. 9).

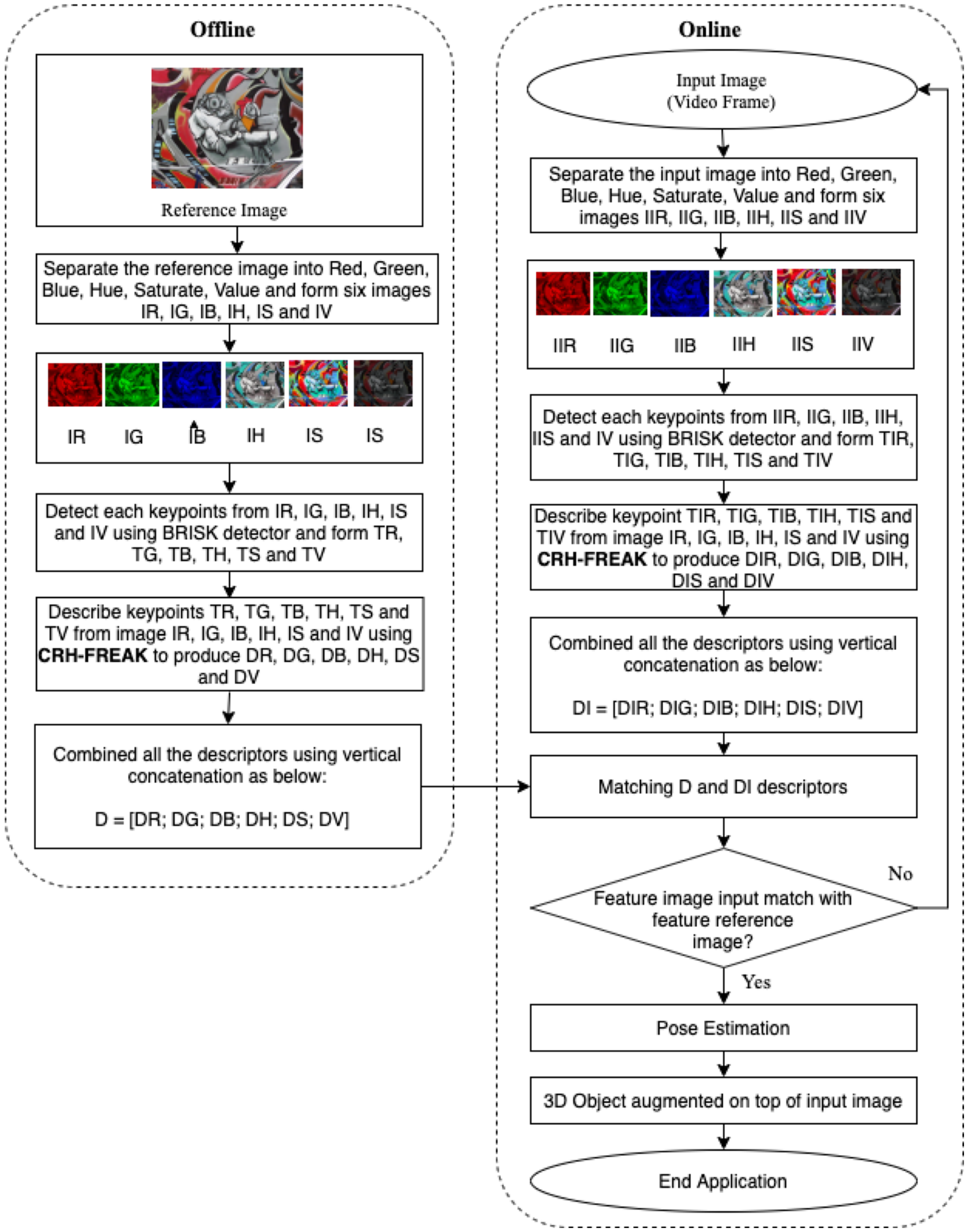


Fig. 7 The proposed CRH-FREAK framework for MAR applications

$$\sum_{1 \le i \le 512} 2^{i-1} T\left(p_{\mathrm{red}}; x_i, y_i\right), \tag{10}$$

$$\sum_{1 \le i \le 512} 2^{i-1} T\left(p_{\mathrm{green}}; x_i, y_i\right), \tag{11}$$

$$\sum_{1 \le i \le 512} 2^{i-1} T\left(p_{\mathrm{blue}}; x_i, y_i\right), \tag{12}$$

$$\sum_{1 \le i \le 512} 2^{i-1} T\left(p_{\mathrm{hue}}; x_i, y_i\right), \tag{13}$$

$$\sum_{1 \le i \le 512} 2^{i-1} T\left(p_{\mathrm{saturation}}; x_i, y_i\right), \tag{14}$$

$$\sum_{1 \le i \le 512} 2^{i-1} T\left(p_{\mathrm{value}}; x_i, y_i\right). \tag{15}$$

4.4 Vertical concatenation technique

Following the description of keypoints, the descriptors for each colour space (DR, DG, DB, DH, DS, and DV)

Fig. 8 The description of keypoints by CRH-FREAK using four cascades

were formed, which were then vertically merged using the vertical concatenation technique to form only one descriptor. For example, the use of this technique vertically combines 512-bit DR, DG, DB, DH, DS, and DV with 20 keypoints into a 512-bit descriptor with 120 keypoints. The following equations demonstrate the use of vertical concatenation technique in this study:

$$DR_{512} = [r_1, r_2, r_3, \ldots r_{20}] \tag{16}$$

$$DG_{512} = [g_1, g_2, g_3, \dots g_{20}] \tag{17}$$

$$DB_{512} = [b_1, b_2, b_3, \dots b_{20}] \tag{18}$$

$$DH_{512} = [h_1, h_2, h_3, \dots h_{20}] \tag{19}$$

$$DS_{512} = [s_1, s_2, s_3, \dots s_{20}] \tag{20}$$

$$DV_{512} = [v_1, v_2, v_3, \dots v_{20}] \tag{21}$$

$$D_{512} = vconcat[DR; DG; DB; DH; DS; DV] \tag{22}$$

$$P_{512} = [r_1, r_2, r_3, \dots r_{20}, g_1, g_2, g_3, \dots g_{20}, v_1, v_2, v_3, \dots v_{20}] \tag{23}$$

The vertical concatenation technique combines the existing descriptors without increasing the size of the descriptor. Similar steps of detecting and describing in the offline tracking process are performed for the online tracking process. In particular, TIR, TIG, TIB, TIH, TIS, and TIV keypoints were formed based on the detected features from the input images (IIR, IIG, IIB, IIH, IIS, and IIV). CRH-FREAK was then used to describe TIR, TIG, TIB, TIH, TIS, and TIV to form DIR, DIG, DIB, DIH, DIS, and DIV, which were merged using the vertical concatenation technique to develop DI descriptors. Following that, DI descriptor of the input image was matched with the D descriptor of the reference image. The tracking process was performed again if the matching percentage (between DI descriptor and D descriptor) was lower than the threshold. However, if the matching percentage exceeded the threshold, pose estimation and augmentation process ensued. A mobile AR application is deemed successful when a successfully augmented 3D virtual object above the input image is produced from a completed tracking process.

5 Size Reduction of 512-bit CRH-FREAK

As previously discussed, the sampling pattern of a 512-bit FREAK is shaped like an eye retina. The structure of the sampling pair is similar to a coarse-to-fine approach, which is compatible with the human retina model. In particular, the outer layer extracts the coarsest information, while the inner layer extracts the finest information. Alahi et al. (2012) divided the layers into four cascades (128-bit per cascade). The first cascade mainly covers the area of the perifoveal receptive field (coarse information), and the fourth cascade mainly covers the area of the fovea receptive field (fine information). Similar to FREAK, the proposed 512-bit CRH-FREAK in this study first compared only the first (128 bits) cascade and proceeded to make comparisons with the subsequent cascades if the distance is lower than the threshold. About 10% of the candidate features from the first cascade were retained. However, the features described by the CRH-FREAK is approximately six times more compared to the FREAK description. Considering that, the size of CRH-FREAK was reduced from 512 bits to 128 bits, resulting in only one cascade for the description of the detected features. In particular, four descriptors with four cascades (128-bit CRH-FREAK) were developed, namely CRH-FREAK First Cascade (CRH-FREAK (L1)), CRH-FREAK Second Cascade (CRH-FREAK (L2)), CRH-FREAK Third Cascade (CRH-FREAK (L3)), and CRH-FREAK Four Cascade (CRH-FREAK (L4)), which are discussed in the next subsections.

5.1 128-bit CRH-FREAK using one cascade

As said, four cascades of 512-bit CRH-FREAK were used to develop CRH-FREAK (L1), CRH-FREAK (L2), CRH-FREAK (L3), and CRH-FREAK (L4). The size of the proposed 512-bit CRH-FREAK in this study was reduced to 128 bits. The developed 128-bit CRH-FREAK compared the intensity of each pixel in the red, green, blue, hue, saturation, and value colour spaces to form the binary string, according to the previously discussed equation in Sect. 4.3. As shown in Fig. 9, CRH-FREAK (L1) described only 128 bits using the first cascade (mainly coarse information). The following equations were used to add up each pixel described by CRH-FREAK (L1).

$$\sum_{1 \le i \le 128} 2^{i-1} B(p_{\text{red}}; x_i, y_i), \tag{24}$$

$$\sum_{1 \le i \le 128} 2^{i-1} B(p_{\text{green}}; x_i, y_i), \tag{25}$$

$$\sum_{1 \le i \le 128} 2^{i-1} B(p_{\text{blue}}; x_i, y_i), \tag{26}$$

$$\sum_{1 \le i \le 128} 2^{i-1} B(p_{\text{hue}}; x_i, y_i), \tag{27}$$

$$\sum_{1 \le i \le 128} 2^{i-1} B(p_{\text{saturation}}; x_i, y_i), \tag{28}$$

$$\sum_{1 \le i \le 128} 2^{i-1} B(p_{\text{value}}; x_i, y_i). \tag{29}$$

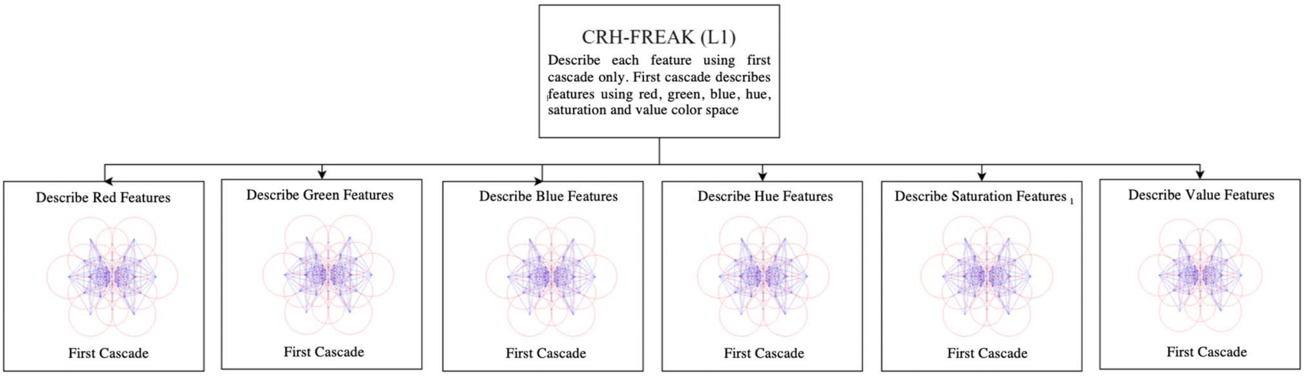


Fig. 9 CRH-FREAK (L1) using the first cascade

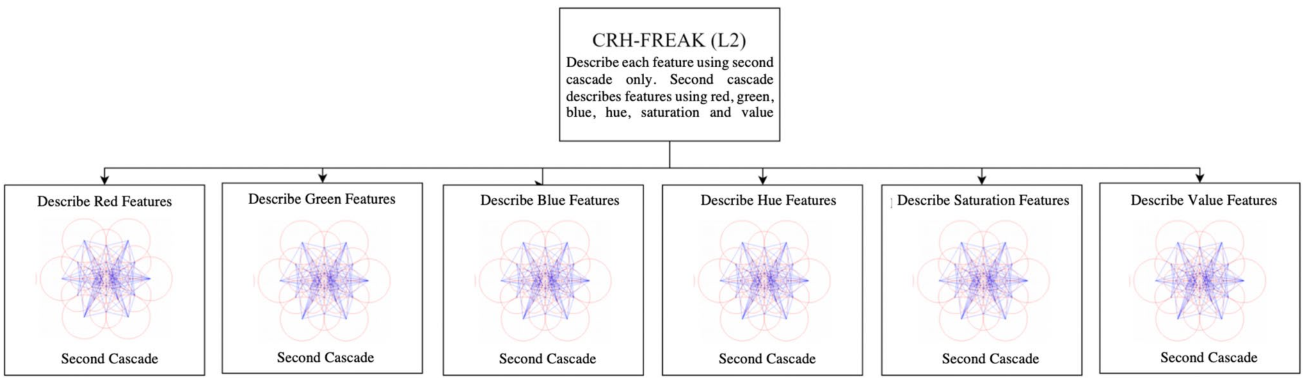


Fig. 10 CRH-FREAK (L2) using the second cascade

Likewise, the number of bits used for CRH-FREAK (L2) was 128 bits, but only the second cascade (mostly coarse information with some fine information) in the 512-bit CRH-FREAK was used. Figure 10 presents CRH-FREAK (L2) using the second cascade. The presented equations were used to add up each pixel described by CRH-FREAK (L2).

$$\sum_{129\le i\le 256} 2^{i-1}B\left(p_{\text{red}};x_i,\,y_i\right), \tag{30}$$

$$\sum_{129\le i\le 256} 2^{i-1}B\left(p_{\text{green}};x_i,\,y_i\right), \tag{31}$$

$$\sum_{129\le i\le 256} 2^{i-1}B\left(p_{\text{blue}};x_i,\,y_i\right), \tag{32}$$

$$\sum_{129\le i\le 256} 2^{i-1}B\left(p_{\text{hue}};x_i,\,y_i\right), \tag{33}$$

$$\sum_{129\le i\le 256} 2^{i-1}B\left(p_{\text{saturation}};x_i,\,y_i\right), \tag{34}$$

$$\sum_{129\le i\le 256} 2^{i-1}B\left(p_{\text{value}};x_i,\,y_i\right), \tag{35}$$

Similarly, the number of bits used to develop CRH-FREAK (L3) was also 128 bits, but only the third cascade (mostly fine information with some coarse information) in the 512-bit CRH-FREAK was used. Figure 11 illustrates CRH-FREAK (L3) using the third cascade only. The presented equations were used to add up each pixel described by CRH-FREAK (L3).

$$\sum_{257\le i\le 384} 2^{i-1}B\left(p_{\text{red}};x_i,\,y_i\right), \tag{36}$$

$$\sum_{257\le i\le 384} 2^{i-1}B\left(p_{\text{green}};x_i,\,y_i\right), \tag{37}$$

$$\sum_{257\le i\le 384} 2^{i-1}B\left(p_{\text{blue}};x_i,\,y_i\right), \tag{38}$$

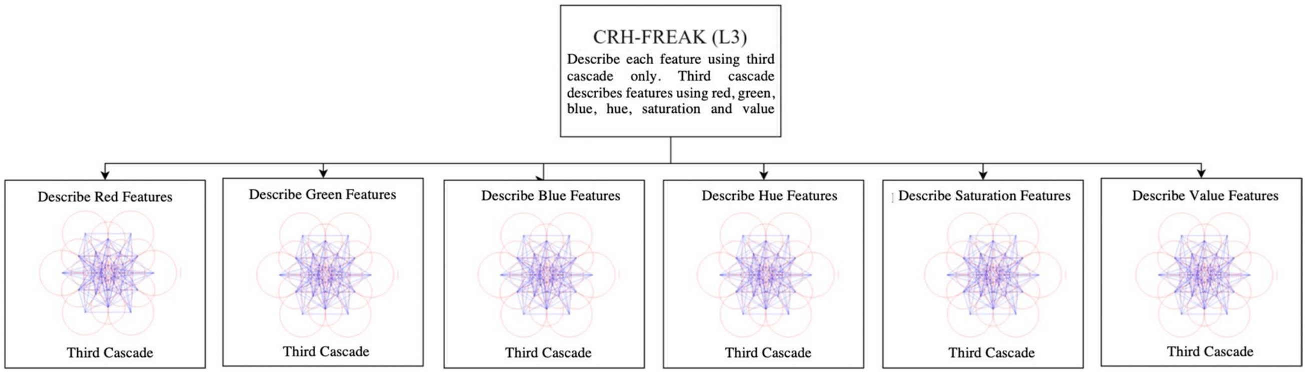


Fig. 11 CRH-FREAK (L3) using the third cascade

$$\sum_{257 \le i \le 384} 2^{i-1} B\left(p_{\text{hue}}; x_i, y_i\right), \tag{39}$$

$$\sum_{257 \le i \le 384} 2^{i-1} B\left(p_{\text{saturation}}; x_i, y_i\right), \tag{40}$$

$$\sum_{257 \le i \le 384} 2^{i-1} B\left(p_{\text{value}}; x_i, y_i\right), \tag{41}$$

Lastly, the number of bits used to develop CRH-FREAK (L4) in this study was also 128-bit. However, only the fourth cascade (mainly fine information) in the 512-bit CRH-FREAK was used. Figure 12 shows CRH-FREAK (L4) using the fourth cascade only. The subsequent equations were used to add up each pixel described by CRH-FREAK (L4).

$$\sum_{385 \le i \le 512} 2^{i-1} B\left(p_{\text{merah}}; x_i, y_i\right), \tag{42}$$

$$\sum_{385 \le i \le 512} 2^{i-1} B\left(p_{\text{hijau}}; x_i, y_i\right), \tag{43}$$

$$\sum_{385 \le i \le 512} 2^{i-1} B\left(p_{\text{biru}}; x_i, y_i\right), \tag{44}$$

$$\sum_{385 \le i \le 512} 2^{i-1} B\left(p_{\text{rona}}; x_i, y_i\right), \tag{45}$$

$$\sum_{385 \le i \le 512} 2^{i-1} B\left(p_{\text{penepuan}}; x_i, y_i\right), \tag{46}$$

$$\sum_{385 \le i \le 512} 2^{i-1} B\left(p_{\text{nilai}}; x_i, y_i\right). \tag{47}$$

6 Results

This section presents the efficiency and robustness results of the original FREAK, CRH-FREAK (L1), CRH-FREAK (L2), CRH-FREAK (L3), CRH-FREAK (L4), and 512-bit CRH-FREAK descriptors for mobile AR applications. The efficiency of all descriptors was tested in terms

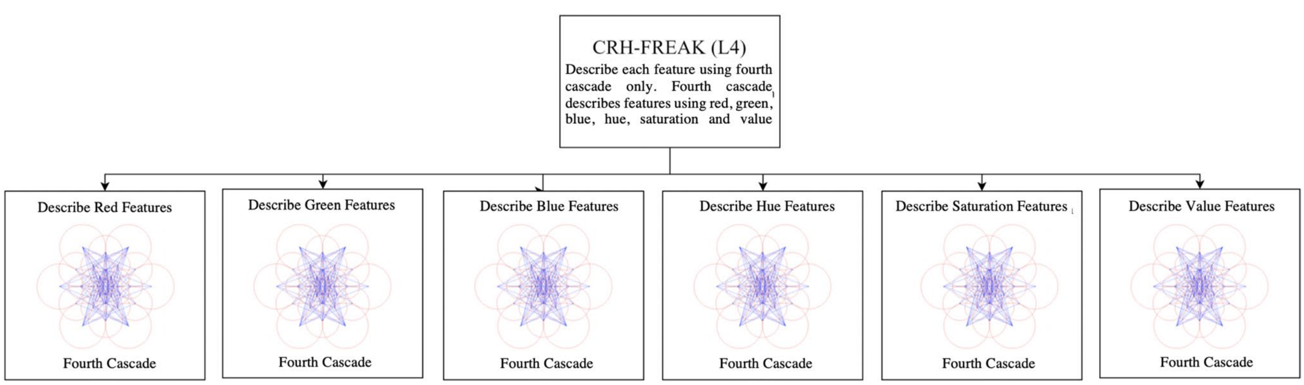


Fig. 12 CRH-FREAK (L4) using the fourth cascade

of computational time. Meanwhile, the robustness of the descriptors against scale, rotation, and lighting (lighting colour, lighting direction, and lighting value) changes was evaluated. The ALOI dataset Geusebroek et al. (2005) was used to evaluate the robustness of the descriptors against the changes in lighting colour and lighting direction. Meanwhile, the Mikolajczyk dataset (Mikolajczyk 2005) was used to evaluate the robustness of the descriptors against the changes in scale, rotation, and lighting value.

6.1 Computational time

One of the most important aspects for mobile AR applications involves the computational time required to complete the tracking process. For a smooth real-time operation, the tracking process must be completed within a short amount of time. The computational time to perform the tracking process can be calculated by combining the computational time of each process, from the start of the video frame (input image), feature detection, feature description, and feature matching to pose estimation and 3D object augmentation. The following equation was used in this study to calculate the computational time for each process:

$$f(x) = T_t\left(T_e - T_s\right) \tag{48}$$

where T_s denotes the initial time of a process; T_e denotes the final time of a process; T_t denotes the total computational time of a process; $f(x)$ denotes the computational time to perform the tracking process (from the start of the video frame to the 3D object augmentation).

In particular, the computational time for the detection process, description process, and matching process were recorded at every 500 keypoints. This particular step was repeated for 50 times for every process (i.e. detection process, description process, and matching process). Table 1 presents the computational time needed to perform each process for each descriptor in this study.

The obtained results revealed that FREAK (29.1 ms), CRH-FREAK (L1) (29.4 ms), CRH-FREAK (L2) (29.6 ms), CRH-FREAK (L3) (29.6 ms), and CRH-FREAK (L4) (29.5 ms) used almost similar computational time to complete the tracking process. Furthermore, all four 128-bit CRH-FREAK descriptors successfully lowered the computational time by 3.5 times (as compared to the computational time recorded by the 512-bit CRH-FREAK). Besides that, the computational time recorded by any cascade of 128-bit CRH-FREAK descriptors was almost similar to that of the original FREAK descriptor.

6.1.1 One-way ANOVA test

One-way ANOVA test was performed to determine whether the descriptors under study display significant differences in the computational time:

H0 = There are no significant differences in the computational time among FREAK, CRH-FREAK (L1), CRH-FREAK (L2), CRH-FREAK (L3), CRH-FREAK (L4), and 512-bit CRH-FREAK.
H1 = There are significant differences in the computational time among FREAK, CRH-FREAK (L1), CRH-FREAK (L2), CRH-FREAK (L3), CRH-FREAK (L4), and 512-bit CRH-FREAK.

The obtained results revealed p-value which less than 0.05 (support H1). In other words, FREAK, CRH-FREAK (L1), CRH-FREAK (L2), CRH-FREAK (L3), CRH-FREAK (L4), and 512-bit CRH-FREAK displayed significant differences in the computational time.

Table 1 Computational time required to perform each tracking process

Descriptors	FREAK-Time (ms)	CRH-FREAK(L1) Time (ms)	CRH-FREAK(L2) Time (ms)	CRH-FREAK(L3) Time (ms)	CRH-FREAK(L4) Time (ms)	512-bit CRH-FREAKTime (ms)
Tracking Process						
Video frame shot using camera	1.3	1.3	1.3	1.3	1.3	1.3
Convert input image to greyscale or colour scale	2.3	2.3	2.3	2.3	2.3	2.3
Detection (500 keypoints)	13.8	13.9	14.0	14.1	14.1	71.4
Description (500 keypoints)	4.3	4.5	4.6	4.5	4.4	21.4
Matching (500 keypoints)	1.9	1.9	1.9	1.9	1.9	1.9
Pose estimation	4.1	4.1	4.1	4.1	4.1	4.1
3D object augmentation	1.4	1.4	1.4	1.4	1.4	1.4
Total computational time (ms)	29.1	29.4	29.6	29.6	29.5	103.8

Following that, multiple comparison analysis was conducted to determine whether there is significant difference in computational time between two descriptors:

H0 = There is no significant difference in computational time between the two descriptors.
H1 = There is significant difference in computational time between the two descriptors.

The obtained results are tabulated in Table 2, which revealed p-values of more than 0.05 (reject H1) for all comparisons, except for the comparisons with 512-bit CRH-FREAK. This may be due to the (longer) computational time used by 512-bit CRH-FREAK to complete the tracking process, as compared to the other descriptors under study. Meanwhile, FREAK and 128-bit CRH-FREAK (any cascade) did not display significant differences in computational time. In other words, the proposed 128-bit CRH-FREAK that incorporated six colour spaces (RGB and HSV) and used only one cascade in this study was comparable to the original FREAK and even successfully reduced the computational time to complete the overall tracking process.

6.1.2 Mean test

Mean test was also conducted for this study to determine which descriptor uses the shortest computational time to complete the overall tracking process. The results of mean test for each descriptor under study are tabulated in Table 3.All descriptors in this study recorded low standard deviation. This implies that the recorded computational time (from 50 tests) for all descriptors were almost similar to the average computational time itself. Besides that, FREAK (29.125 ms), colour spaces still recordedCRH-FREAK (L1) (29.354 ms), CRH-FREAK (L2) (29.586 ms), CRH-FREAK (L3) (29.612 ms), and CRH-FREAK (L4) (29.493 ms) recorded lower average computational time than 512-bit CRH-FREAK (103.779 ms), which may be attributed to the high number of bits used by 512-bit CRH-FREAK to describe the keypoints. CRH-FREAK (L1), CRH-FREAK (L2), CRH-FREAK (L3), and CRH-FREAK (L4) only used 128 bits.

FREAK and all four 128-bit CRH-FREAK descriptors were able to function within a short amount of time (no significant differences between both types of descriptors). This study proved that a 128-bit CRH-FREAK descriptor that makes use of a single cascade and colour spaces can describe both coarse and fine information within the same amount of time to the original FREAK descriptor that makes use of only greyscale.

Table 2 Results of multiple comparison analysis for computational time

Multiple Comparison
Independent Variable: Computational Time

Descriptor	Descriptor	Significant value
FREAK	CRH-FREAK (L1)	$\rho > 0.05$
	CRH-FREAK (L2)	$\rho > 0.05$
	CRH-FREAK (L3)	$\rho > 0.05$
	CRH-FREAK (L4)	$\rho > 0.05$
	512-bit CRH-FREAK	ρ ***<0.05***
CRH-FREAK (L1)	FREAK	$\rho > 0.05$
	CRH-FREAK (L2)	$\rho > 0.05$
	CRH-FREAK (L3)	$\rho > 0.05$
	CRH-FREAK (L4)	$\rho > 0.05$
	512-bit CRH-FREAK	ρ ***<0.05***
CRH-FREAK (L2)	FREAK	$\rho > 0.05$
	CRH-FREAK (L1)	$\rho > 0.05$
	CRH-FREAK (L3)	$\rho > 0.05$
	CRH-FREAK (L4)	$\rho > 0.05$
	512-bit CRH-FREAK	ρ ***<0.05***
CRH-FREAK (L3)	FREAK	$\rho > 0.05$
	CRH-FREAK (L1)	$\rho > 0.05$
	CRH-FREAK (L2)	$\rho > 0.05$
	CRH-FREAK (L4)	$\rho > 0.05$
	512-bit CRH-FREAK	ρ ***<0.05***
CRH-FREAK (L4)	FREAK	$\rho > 0.05$
	CRH-FREAK (L1)	$\rho > 0.05$
	CRH-FREAK (L2)	$\rho > 0.05$
	CRH-FREAK (L3)	$\rho > 0.05$
	512-bit CRH-FREAK	ρ ***<0.05***
512-bit CRH-FREAK	FREAK	ρ ***<0.05***
	CRH-FREAK (L1)	ρ ***<0.05***
	CRH-FREAK (L2)	ρ ***<0.05***
	CRH-FREAK (L3)	ρ ***<0.05***
	CRH-FREAK (L4)	ρ ***<0.05***

Table 3 Mean and standard deviation of the computational time used by each descriptor

Descriptor	Mean (ms)	Standard Deviation
FREAK	29.125	0.156
CRH-FREAK (L1)	29.354	0.190
CRH-FREAK (L2)	29.586	0.194
CRH-FREAK (L3)	29.612	0.188
CRH-FREAK (L4)	29.493	0.234
512-bit CRH-FREAK	103.779	0.169

6.2 Scale invariance

The robustness of the original FREAK, CRH-FREAK (L1), CRH-FREAK (L2), CRH-FREAK (L3), CRH-FREAK (L4), and 512-bit CRH-FREAK descriptors against scale changes was evaluated using the Mikolajczyk dataset. For this study, the configuration of scale invariance test was conducted based on the following equation:

$$\text{Accuracy Percentage} = \frac{\text{Number of Correct Matching}(n)}{\text{Number of Matching}(N)} * 100\% \tag{49}$$

In particular, this study made use of each image with the same scale changes repeatedly for 25 times. With that, each image in the same state was used for a total of 50 times (two images; boat image and bark image × 25 times). Figure 13 presents the obtained accuracy results (using the recorded mean value of each test) for each descriptor in this study.

6.2.1 One-way ANOVA test

One-way ANOVA test was performed to determine whether the accuracy of the descriptors under study is significantly different in the presence of scale changes:

H_0 = The accuracy of FREAK, CRH-FREAK (L1), CRH-FREAK (L2), CRH-FREAK (L3), CRH-FREAK (L4), and 512-bit CRH-FREAK is not significantly different in the presence of scale changes.

H_1 = The accuracy of FREAK, CRH-FREAK (L1), CRH-FREAK (L2), CRH-FREAK (L3), CRH-FREAK (L4), and 512-bit CRH-FREAK is significantly different in the presence of scale changes.

The obtained results revealed p-value which more than 0.05 (reject H_1). In other words, the accuracy of FREAK, CRH-FREAK (L1), CRH-FREAK (L2), CRH-FREAK (L3), CRH-FREAK (L4), and 512-bit CRH-FREAK was not significantly different in the presence of scale changes. As there were no significance differences in accuracy, multiple comparison analysis was not required.

6.2.2 Mean test

Mean test was conducted to determine which descriptor has the highest accuracy in the presence of scale changes. In this case, a high mean value suggests high accuracy. Table 4 presents the results of mean test for each descriptor under study.

All descriptors in this study recorded low standard deviation. This implies that the recorded mean values of accuracy (from 50 tests) for all descriptors were almost similar to the average mean value of the accuracy itself. Based on the results, 512-bit CRH-FREAK recorded the highest accuracy (92.64%) in the presence of scale changes, which was followed by CRH-FREAK (L4) (91.38%), CRH-FREAK (L3) (91.28%), CRH-FREAK (L2) (90.94%), and CRH-FREAK

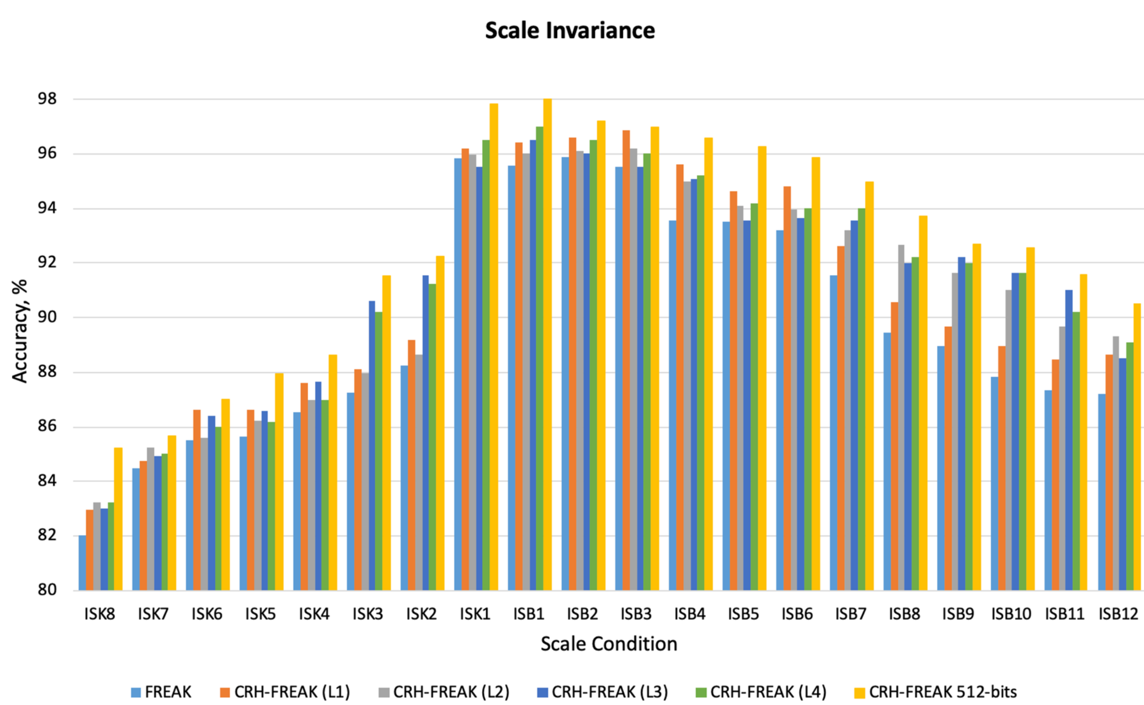


Fig. 13 Accuracy results for each descriptor in the presence of scale changes

Table 4 Mean and standard deviation of the overall tracking accuracy in the presence of scale changes for each descriptor

Descriptor	Mean	Standard Deviation
FREAK	89.760	3.851
CRH-FREAK (L1)	90.795	3.930
CRH-FREAK (L2)	90.937	3.742
CRH-FREAK (L3)	91.276	3.535
CRH-FREAK (L4)	91.377	3.789
512-bit CRH-FREAK	92.645	3.720

(L1) (90.80%). Meanwhile, FREAK recorded the lowest accuracy (89.76%).

This study successfully proved the robustness of all four 128-bit CRH-FREAK descriptors against scale changes. Furthermore, the accuracy of these 128-bit CRH-FREAK descriptors was not significantly different from the accuracy of the original FREAK and 512-bit CRH-FREAK descriptors despite the size reduction of these descriptors (from 512-bit to 128-bit).

6.3 Rotation invariance

The robustness of the original FREAK, CRH-FREAK (L1), CRH-FREAK (L2), CRH-FREAK (L3), CRH-FREAK (L4), and 512-bit CRH-FREAK descriptors against rotation changes was evaluated using the boat image and bark image from the Mikolajczyk dataset. For this study, the configuration of rotation invariance test was conducted using Eq. 49 (Tan et al. 2019).

In particular, this study made use of each image with the same rotation changes repeatedly for 25 times. With that, each image in the same state was used for a total of 50 times (two images; boat image and bark image × 25 times). Figure 14 presents the obtained accuracy results (using the recorded mean value of each test) for each descriptor in this study.

Overall, all 128-bit CRH-FREAK descriptors in this study displayed relatively high accuracy in the presence of rotation changes. The accuracy of all four 128-bit CRH-FREAK descriptors was higher than the accuracy of the original FREAK descriptor, but lower than the accuracy of the 512-bit CRH-FREAK descriptor. For instance, 512-bit CRH-FREAK recorded the highest accuracy (92.54%) in $I_{R_{12}}$ condition, which was followed by CRH-FREAK (L1) (90.00%), CRH-FREAK (L2) (90.77%), CRH-FREAK (L3) (90.99%), CRH-FREAK (L4) (91.00%), and lastly, FREAK (87.01%). In other words, the developed 128-bit CRH-FREAK descriptors in this study were less accurate than the 512-bit CRH-FREAK descriptor. This can be explained by the size reduction of the descriptors, which causes the omission of certain information during the description process.

6.3.1 One-way ANOVA test

One-way ANOVA testOne-way ANOVA test was performed to determine whether the accuracy of the descriptors under

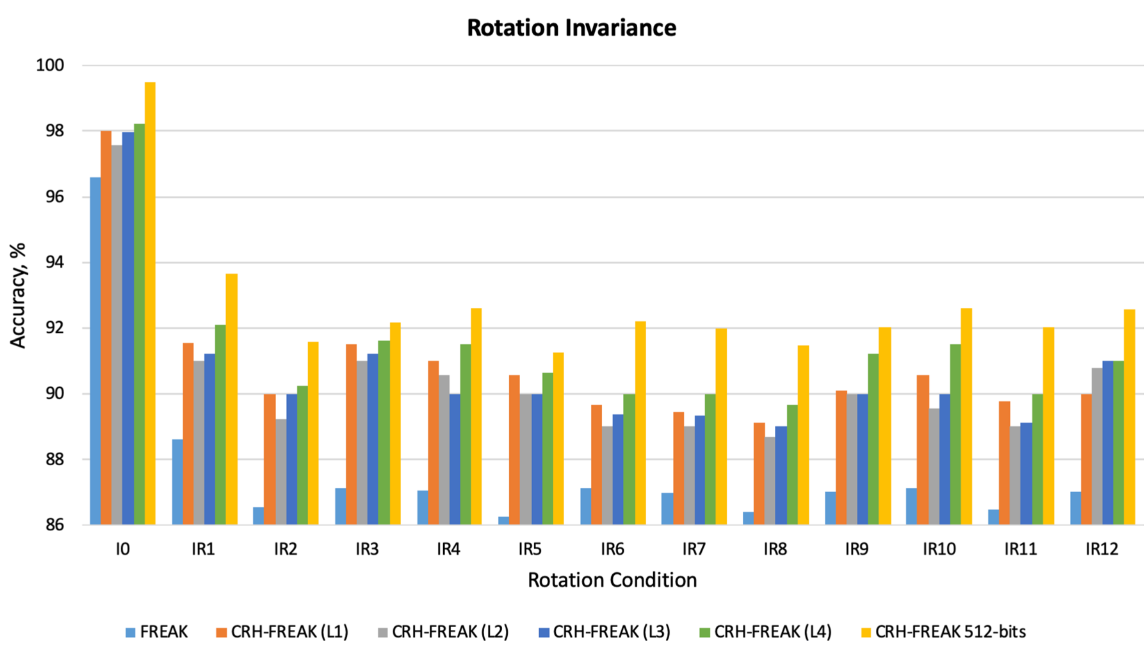


Fig. 14 Accuracy results for each descriptor in the presence of rotation changes

study is significantly different in the presence of rotation changes:

H_0 = The accuracy of FREAK, CRH-FREAK (L1), CRH-FREAK (L2), CRH-FREAK (L3), CRH-FREAK (L4), and 512-bit CRH-FREAK is not significantly different in the presence of rotation changes.
H_1 = The accuracy of FREAK, CRH-FREAK (L1), CRH-FREAK (L2), CRH-FREAK (L3), CRH-FREAK (L4), and 512-bit CRH-FREAK is significantly different in the presence of rotation changes.

The obtained results revealed p-value which less than 0.05 (reject H0). In other words, the accuracy of FREAK, CRH-FREAK (L1), CRH-FREAK (L2), CRH-FREAK (L3), CRH-FREAK (L4), and 512-bit CRH-FREAK was significantly different in the presence of rotation changes.

Following that, multiple comparison analysis was conducted to determine whether the difference in accuracy of the two descriptors is significant in the presence of rotation changes:

H_0 = The difference in accuracy of the two descriptors is not significant in the presence of rotation changes.
H_1 = The difference in accuracy of the two descriptors is significant in the presence of rotation changes.

The obtained results are tabulated in Table 5, which revealed that the difference in accuracy of the two descriptors (in the following comparison pair list) was not significant (*p*-value of more than 0.05) in the presence of rotation changes: (1) CRH-FREAK (L1) and CRH-FREAK (L2); (2) CRH-FREAK (L1) and CRH-FREAK (L3); (3) CRH-FREAK (L1) and CRH-FREAK (L4); (4) CRH-FREAK (L1) and 512-bit CRH-FREAK; (5) CRH-FREAK (L2) and CRH-FREAK (L3); (6) CRH-FREAK (L2) and CRH-FREAK (L4); (7) CRH-FREAK (L2) and 512-bit CRH-FREAK; (8) CRH-FREAK (L3) and CRH-FREAK (L4); (9) CRH-FREAK (L3) and 512-bit CRH-FREAK; (10) CRH-FREAK (L4) and 512-bit CRH-FREAK. This study demonstrated that 128-bit CRH-FREAK descriptors that made use of RGB and HSV colour spaces displayed high accuracy that was not significantly different from the 512-bit CRH-FREAK descriptor.

On the other hand, the differences in accuracy of FREAK from CRH-FREAK (L1), CRH-FREAK (L2), CRH-FREAK (L3), CRH-FREAK (L4), and 512-bit CRH-FREAK were significantly different (*p*-value of less than 0.05) in the presence of rotation changes. This study demonstrated that the accuracy for FREAK was relatively lower than the other descriptors that made use of RGB and HSV colour spaces.

6.3.2 Mean test

Mean test was conducted to determine which descriptor has the highest accuracy in the presence of rotation changes. In this case, a higher mean value suggests higher accuracy. Table 6 presents the results of mean test for each descriptor under study.

The low standard deviations recorded by all descriptors in this study imply that the recorded mean values of accuracy (from 50 tests) were almost similar to the average mean value of the accuracy itself. Based on the results, 512-bit CRH-FREAK recorded the highest accuracy (92.74%) in the presence of rotation changes, which was followed by CRH-FREAK (L4) (91.36%), CRH-FREAK (L1) (90.86%), CRH-FREAK (L3) (90.63%), CRH-FREAK (L2) (90.41%), and lastly, FREAK (87.71%).

This study successfully proved the robustness of all four 128-bit CRH-FREAK descriptors against rotation changes. Although these descriptors displayed lower accuracy than the 512-bit CRH-FREAK descriptor, the differences in accuracy were not significantly different (based on the results of multiple comparison analysis).

6.4 Lighting colour invariance

The robustness of the original FREAK, CRH-FREAK (L1), CRH-FREAK (L2), CRH-FREAK (L3), CRH-FREAK (L4), and 512-bit CRH-FREAK descriptors against lighting colour changes was evaluated using the first image (a bear toy) to the 50th image (a bean) from the ALOI dataset. For this study, the configuration of lighting colour invariance test was conducted using Eq. 49 (Tan et al. 2019).

In particular, this study made use of each image with the same lighting colour changes, but with different objects. With that, there were 50 accuracy results of the same lighting colour changes for each descriptor. Figure 15 presents the obtained accuracy results (using the recorded mean value of each test) for each descriptor in this study.

Overall, all 128-bit CRH-FREAK descriptors in this study displayed relatively high accuracy in the presence of lighting colour changes. The accuracy of all four 128-bit CRH-FREAK descriptors was higher than the accuracy of the original FREAK descriptor, but lower than the accuracy of the 512-bit CRH-FREAK descriptor. For instance, 512-bit CRH-FREAK recorded the highest accuracy (90.00%) in 2975 K lighting temperature condition, which was followed by CRH-FREAK (L1) (89.20%), CRH-FREAK (L2) (88.20%), CRH-FREAK (L3) (89.27%), CRH-FREAK (L4) (88.90%), and lastly, FREAK (85.04%). In other words, the developed 128-bit CRH-FREAK descriptors in this study were less accurate than the 512-bit CRH-FREAK descriptor.

Table 5 Results of multiple comparison analysis for the accuracy of descriptors in the presence of rotation changes

Multiple Comparison		
Independent Variable: Rotation Changes		
Descriptor	Descriptor	Significant Value
FREAK	CRH-FREAK (L1)	***ρ < 0.05***
	CRH-FREAK (L2)	***ρ < 0.05***
	CRH-FREAK (L3)	***ρ < 0.05***
	CRH-FREAK (L4)	***ρ < 0.05***
	512-bit CRH-FREAK	***ρ < 0.05***
CRH-FREAK (L1)	FREAK	***ρ < 0.05***
	CRH-FREAK (L2)	ρ > 0.05
	CRH-FREAK (L3)	ρ > 0.05
	CRH-FREAK (L4)	ρ > 0.05
	512-bit CRH-FREAK	ρ > 0.05
CRH-FREAK (L2)	FREAK	***ρ < 0.05***
	CRH-FREAK (L1)	ρ > 0.05
	CRH-FREAK (L3)	ρ > 0.05
	CRH-FREAK (L4)	ρ > 0.05
	512-bit CRH-FREAK	ρ > 0.05
CRH-FREAK (L3)	FREAK	***ρ < 0.05***
	CRH-FREAK (L1)	ρ > 0.05
	CRH-FREAK (L2)	ρ > 0.05
	CRH-FREAK (L4)	ρ > 0.05
	512-bit CRH-FREAK	ρ > 0.05
CRH-FREAK (L4)	FREAK	***ρ < 0.05***
	CRH-FREAK (L1)	ρ > 0.05
	CRH-FREAK (L2)	ρ > 0.05
	CRH-FREAK (L3)	ρ > 0.05
	512-bit CRH-FREAK	ρ > 0.05
512-bit CRH-FREAK	FREAK	***ρ < 0.05***
	CRH-FREAK (L1)	ρ > 0.05
	CRH-FREAK (L2)	ρ > 0.05
	CRH-FREAK (L3)	ρ > 0.05
	CRH-FREAK (L4)	ρ > 0.05

Table 6 Mean and standard deviation of the overall tracking accuracy in the presence of rotation changes for each descriptor

Descriptors	Mean	Standard deviation
FREAK	87.705	2.328
CRH-FREAK (L1)	90.863	1.871
CRH-FREAK (L2)	90.408	1.898
CRH-FREAK (L3)	90.628	1.922
CRH-FREAK (L4)	91.355	1.797
512-bit CRH-FREAK	92.735	1.722

This can be explained by the size reduction of the descriptors, which causes the omission of certain information during the description process.

6.4.1 One-way ANOVA test

One-way ANOVA test was performed to determine whether the accuracy of the descriptors under study is significantly different in the presence of lighting colour changes:

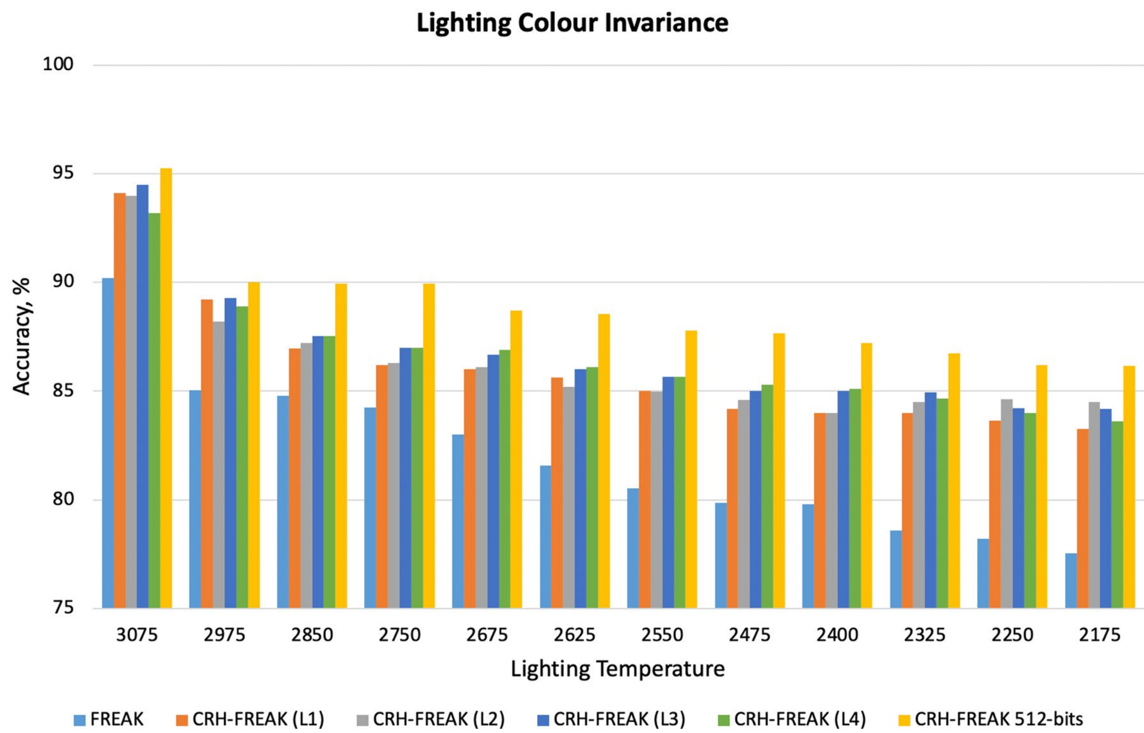


Fig. 15 Accuracy results for each descriptor in the presence of lighting colour changes

> H_0 = The accuracy of FREAK, CRH-FREAK (L1), CRH-FREAK (L2), CRH-FREAK (L3), CRH-FREAK (L4), and 512-bit CRH-FREAK is not significantly different in the presence of lighting colour changes.
> H_1 = The accuracy of FREAK, CRH-FREAK (L1), CRH-FREAK (L2), CRH-FREAK (L3), CRH-FREAK (L4), and 512-bit CRH-FREAK is significantly different in the presence of lighting colour changes.

The obtained results revealed p-value which less than 0.05 (reject H0). In other words, the accuracy of FREAK, CRH-FREAK (L1), CRH-FREAK (L2), CRH-FREAK (L3), CRH-FREAK (L4), and 512-bit CRH-FREAK was significantly different in the presence of lighting colour changes.

Following that, multiple comparison analysis was conducted to determine whether the difference in accuracy of the two descriptors is significant in the presence of lighting colour changes:

> H_0 = The difference in accuracy of the two descriptors is not significant in the presence of lighting colour changes.
> H_1 = The difference in accuracy of the two descriptors is significant in the presence of lighting colour changes.

The obtained results are presented in Table 7, which revealed that the difference in accuracy of the two descriptors (in the following comparison pair list) was significant (*p*-value of less than 0.05) in the presence of lighting colour changes: (1) FREAK and CRH-FREAK (L1); (2) FREAK and CRH-FREAK (L2); (3) FREAK and CRH-FREAK (L3); (4) FREAK and CRH-FREAK (L4); (5) FREAK and 512-bit CRH-FREAK. This study demonstrated that the original FREAK descriptor was less accurate than all CRH-FREAK descriptors that made use of RGB and HSV colour spaces.

On the other hand, the results revealed that the difference in accuracy of the two descriptors (in the following comparison pair list) was not significant (*p*-value of more than 0.05) in the presence of lighting colour changes: (1) CRH-FREAK (L1) and CRH-FREAK (L2); (2) CRH-FREAK (L1) and CRH-FREAK (L3); (3) CRH-FREAK (L1) and CRH-FREAK (L4); (4) CRH-FREAK (L1) and 512-bit CRH-FREAK; (5) CRH-FREAK (L2) and CRH-FREAK (L3); (6) CRH-FREAK (L2) and CRH-FREAK (L4); (7) CRH-FREAK (L2) and 512-bit CRH-FREAK; (8) CRH-FREAK (L3) and CRH-FREAK (L4); (9) CRH-FREAK (L3) and 512-bit CRH-FREAK; (10) CRH-FREAK (L4) and 512-bit CRH-FREAK. Basically, 128-bit CRH-FREAK and the 512-bit CRH-FREAK descriptors, which made use of RGB and HSV colour spaces, in this study were more accurate than the original FREAK descriptor that made use of greyscale only. In addition, CRH-FREAK (L1), CRH-FREAK (L2), CRH-FREAK (L3), and CRH-FREAK (L4) descriptors displayed high accuracy that was not significantly different from the 512-bit CRH-FREAK descriptor.

Table 7 Results of multiple comparison analysis for the accuracy of descriptors in the presence of lighting colour changes

Multiple Comparison		
Independent Variable: Lighting Colour Changes		
Descriptor	Descriptor	Significant Value
FREAK	CRH-FREAK (L1)	ρ ***< 0.05***
	CRH-FREAK (L2)	ρ ***< 0.05***
	CRH-FREAK (L3)	ρ ***< 0.05***
	CRH-FREAK (L4)	ρ ***< 0.05***
	512-bit CRH-FREAK	ρ ***< 0.05***
CRH-FREAK (L1)	FREAK	ρ ***< 0.05***
	CRH-FREAK (L2)	$\rho > 0.05$
	CRH-FREAK (L3)	$\rho > 0.05$
	CRH-FREAK (L4)	$\rho > 0.05$
	512-bit CRH-FREAK	$\rho > 0.05$
CRH-FREAK (L2)	FREAK	ρ ***< 0.05***
	CRH-FREAK (L1)	$\rho > 0.05$
	CRH-FREAK (L3)	$\rho > 0.05$
	CRH-FREAK (L4)	$\rho > 0.05$
	512-bit CRH-FREAK	$\rho > 0.05$
CRH-FREAK (L3)	FREAK	ρ ***< 0.05***
	CRH-FREAK (L1)	$\rho > 0.05$
	CRH-FREAK (L2)	$\rho > 0.05$
	CRH-FREAK (L4)	$\rho > 0.05$
	512-bit CRH-FREAK	$\rho > 0.05$
CRH-FREAK (L4)	FREAK	ρ ***< 0.05***
	CRH-FREAK (L1)	$\rho > 0.05$
	CRH-FREAK (L2)	$\rho > 0.05$
	CRH-FREAK (L3)	$\rho > 0.05$
	512-bit CRH-FREAK	$\rho > 0.05$
512-bit CRH-FREAK	FREAK	ρ ***< 0.05***
	CRH-FREAK (L1)	$\rho > 0.05$
	CRH-FREAK (L2)	$\rho > 0.05$
	CRH-FREAK (L3)	$\rho > 0.05$
	CRH-FREAK (L4)	$\rho > 0.05$

Table 8 Mean and standard deviation of the overall tracking accuracy in the presence of lighting colour changes for each descriptor

Descriptor	Mean	Standard deviation
FREAK	81.947	3.271
CRH-FREAK (L1)	86.014	2.649
CRH-FREAK (L2)	86.187	2.359
CRH-FREAK (L3)	86.664	2.477
CRH-FREAK (L4)	86.496	2.200
512-bit CRH-FREAK	88.677	2.096

6.4.2 Mean test

Mean test was conducted to determine which descriptor has the highest accuracy in the presence of lighting colour changes. In this case, a higher mean value suggests higher accuracy. Table 8 presents the results of mean test for each descriptor under study.

The low standard deviations recorded by all descriptors in this study imply that the recorded mean values of accuracy (from 50 tests) were almost similar to the average mean value of the accuracy itself. Based on the results, 512-bit CRH-FREAK recorded the highest accuracy (88.68%) in the presence of lighting colour changes, which was followed by CRH-FREAK (L3) (86.66%), CRH-FREAK (L4) (86.50%), CRH-FREAK (L2) (86.19%), CRH-FREAK (L1) (86.01%), and lastly, FREAK (81.95%).

This study successfully proved the robustness of all four 128-bit CRH-FREAK descriptors against lighting colour changes. Furthermore, the accuracy of these descriptors was not significantly different from the accuracy of 512-bit CRH-FREAK descriptor.

6.5 Lighting direction invariance

The robustness of the original FREAK, CRH-FREAK (L1), CRH-FREAK (L2), CRH-FREAK (L3), CRH-FREAK (L4), and 512-bit CRH-FREAK descriptors against lighting direction changes was evaluated using the first image (a bear toy) to the 50th image (a bean) from the ALOI dataset. For this study, the configuration of lighting direction invariance test was conducted using Eq. 49 (Tan et al. 2019).

In particular, this study made use of each image with the same lighting direction changes, but with different objects. With that, there were 50 accuracy results of the same lighting direction changes for each descriptor. Figure 16 presents the obtained accuracy results (using the recorded mean value of each test) for each descriptor in this study.

The obtained results revealed that all 128-bit CRH-FREAK descriptors in this study recorded relatively high accuracy in the presence of lighting direction changes. The accuracy of all four descriptors was higher than the accuracy of the original FREAK descriptor, but lower than

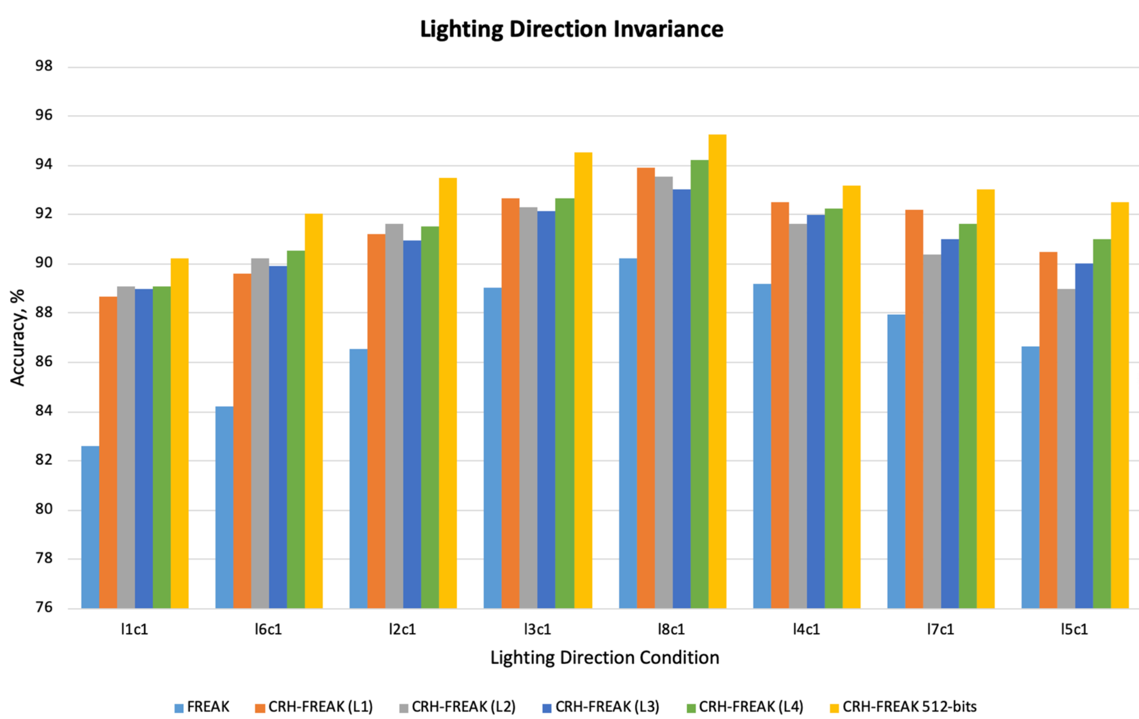


Fig. 16 Accuracy results for each descriptor in the presence of lighting direction changes

the accuracy of the 512-bit CRH-FREAK descriptor. For instance, 512-bit CRH-FREAK recorded the highest accuracy (95.25%) in 18c1 lighting direction condition, which was followed by CRH-FREAK (L1) (95.25%), CRH-FREAK (L1) (93.92%), CRH-FREAK (L2) (93.52%), CRH-FREAK (L3) (93.00%), CRH-FREAK (L4) (94.21%), and lastly, FREAK (90.20%). In other words, the developed 128-bit CRH-FREAK descriptors in this study were less accurate than the 512-bit CRH-FREAK. This can be explained by the size reduction of the descriptors, which causes the omission of certain information during the description process.

6.5.1 One-way ANOVA test

One-way ANOVA test was conducted to determine whether the accuracy of the descriptors under study is significantly different in the presence of lighting direction changes:

> H_0 = The accuracy of FREAK, CRH-FREAK (L1), CRH-FREAK (L2), CRH-FREAK (L3), CRH-FREAK (L4), and 512-bit CRH-FREAK is not significantly different in the presence of lighting direction changes.
> H_1 = The accuracy of FREAK, CRH-FREAK (L1), CRH-FREAK (L2), CRH-FREAK (L3), CRH-FREAK (L4), and 512-bit CRH-FREAK is significantly different in the presence of lighting direction changes.

The obtained results revealed p-value which less than 0.05 (reject H0). In other words, the accuracy of FREAK, CRH-FREAK (L1), CRH-FREAK (L2), CRH-FREAK (L3), CRH-FREAK (L4), and 512-bit CRH-FREAK was significantly different in the presence of lighting direction changes.

Following that, multiple comparison analysis was conducted to determine whether the difference in accuracy of the two descriptors is significant in the presence of lighting direction changes:

> H_0 = The difference in accuracy of the two descriptors is not significant in the presence of lighting direction changes.
> H_1 = The difference in accuracy of the two descriptors is significant in the presence of lighting direction changes.

The obtained results are tabulated in Table 9, which revealed that the difference in accuracy of the two descriptors (in the following comparison pair list) was not significant (*p*-value of more than 0.05) in the presence of lighting direction changes: (1) CRH-FREAK (L1) and CRH-FREAK (L2); (2) CRH-FREAK (L1) and CRH-FREAK (L3); (3) CRH-FREAK (L1) and CRH- FREAK (L4); (4) CRH-FREAK (L1) and 512-bit CRH-FREAK; (5) CRH-FREAK (L2) and CRH-FREAK (L3); (6) CRH-FREAK (L2) and CRH-FREAK (L4); (7) CRH-FREAK (L2) and 512-bit CRH-FREAK; (8) CRH-FREAK (L3) and CRH-FREAK (L4); (9) CRH-FREAK (L3) and 512-bit CRH-FREAK; (10) CRH-FREAK (L4) and 512-bit CRH-FREAK. This study demonstrated that the 128-bit CRH-FREAK and 512-bit CRH-FREAK descriptors that made use of RGB and HSV colour spaces were more accurate that the original FREAK descriptor that made use of greyscale only.

On the other hand, the results revealed that the difference in accuracy of the two descriptors (in the following comparison pair list) was significant (*p*-value of less than 0.05) in the presence of lighting direction changes: (1) FREAK and CRH-FREAK (L1); (2) FREAK and CRH-FREAK (L2); (3) FREAK and CRH-FREAK (L3); (4) FREAK and CRH-FREAK (L4); (5) FREAK and 512-bit CRH-FREAK. The accuracy of the original FREAK descriptor was relatively lower than the developed 128-bit CRH-FREAK descriptors that made use of RGB and HSV colour spaces. In addition, CRH-FREAK (L1), CRH-FREAK (L2), CRH-FREAK (L3), and CRH-FREAK (L4) descriptors displayed high accuracy that was not significantly different from the 512-bit CRH-FREAK descriptor.

6.5.2 Mean test

Mean test was conducted to determine which descriptor has the highest accuracy in the presence of lighting direction changes. In this case, a higher mean suggests higher accuracy. Table 10 presents the results of mean test for each descriptor under study.

The low standard deviations recorded by all descriptors in this study imply that the recorded mean values of accuracy (from 50 tests) were almost similar to the average mean value of the accuracy itself. Based on the results, 512-bit CRH-FREAK recorded the highest accuracy (93.03%) in the presence of lighting direction changes, which was followed by CRH-FREAK (L4) (91.60%), CRH-FREAK (L1) (91.40%), CRH-FREAK (L3) (91.00%), CRH-FREAK (L2) (90.96%), and lastly, FREAK (87.04%).

This study successfully proved the robustness of all four 128-bit CRH-FREAK descriptors against lighting direction changes. Although the accuracy of these descriptors was lower than the accuracy of 512-bit CRH-FREAK descriptor, the differences in accuracy were not significantly different.

6.6 Lighting value invariance

The robustness of the original FREAK, CRH-FREAK (L1), CRH-FREAK (L2), CRH-FREAK (L3), CRH-FREAK (L4), and 512-bit CRH-FREAK descriptors against lighting value changes was evaluated using the Leuven's image from the Mikolajczyk dataset. For this study, the configuration of

Table 9 Results of multiple comparison analysis for the accuracy of descriptors in the presence of lighting direction changes

Multiple Comparison		
Independent Variable: Lighting Direction Changes		
Descriptor	Descriptor	Significant value Significant value
FREAK	CRH-FREAK (L1)	ρ ***< 0.05***
	CRH-FREAK (L2)	ρ ***< 0.05***
	CRH-FREAK (L3)	ρ ***< 0.05***
	CRH-FREAK (L4)	ρ ***< 0.05***
	512-bit CRH-FREAK	ρ ***< 0.05***
CRH-FREAK (L1)	FREAK	ρ ***< 0.05***
	CRH-FREAK (L2)	$\rho > 0.05$
	CRH-FREAK (L3)	$\rho > 0.05$
	CRH-FREAK (L4)	$\rho > 0.05$
	512-bit CRH-FREAK	$\rho > 0.05$
CRH-FREAK (L2)	FREAK	ρ ***< 0.05***
	CRH-FREAK (L1)	$\rho > 0.05$
	CRH-FREAK (L3)	$\rho > 0.05$
	CRH-FREAK (L4)	$\rho > 0.05$
	512-bit CRH-FREAK	$\rho > 0.05$
CRH-FREAK (L3)	FREAK	ρ ***< 0.05***
	CRH-FREAK (L1)	$\rho > 0.05$
	CRH-FREAK (L2)	$\rho > 0.05$
	CRH-FREAK (L4)	$\rho > 0.05$
	512-bit CRH-FREAK	$\rho > 0.05$
CRH-FREAK (L4)	FREAK	ρ ***< 0.05***
	CRH-FREAK (L1)	$\rho > 0.05$
	CRH-FREAK (L2)	$\rho > 0.05$
	CRH-FREAK (L3)	$\rho > 0.05$
	512-bit CRH-FREAK	$\rho > 0.05$
512-bit CRH-FREAK	FREAK	ρ ***< 0.05***
	CRH-FREAK (L1)	$\rho > 0.05$
	CRH-FREAK (L2)	$\rho > 0.05$
	CRH-FREAK (L3)	$\rho > 0.05$
	CRH-FREAK (L4)	$\rho > 0.05$

Table 10 Mean and standard deviation of the overall tracking accuracy in the presence of lighting direction changes for each descriptor

Descriptor	Mean	Standard deviation
FREAK	87.042	2.203
CRH-FREAK (L1)	91.404	1.345
CRH-FREAK (L2)	90.960	1.177
CRH-FREAK (L3)	90.995	1.036
CRH-FREAK (L4)	91.603	1.159
512-bit CRH-FREAK	93.029	1.139

lighting value invariance test was conducted using Eq. 49 (Tan et al. 2019).

In particular, this study made use of Leuven's image with the I_{DR_2} condition, which was repeatedly tested for 50 times using the same descriptor. With that, there were 50 accuracy results of the same lighting value changes for each descriptor. Figure 17 presents the obtained accuracy results (using the recorded mean value of each test) for each descriptor in this study.

The obtained results revealed that all 128-bit CRH-FREAK descriptors in this study recorded relatively high accuracy in the presence of lighting value changes. The accuracy of these four descriptors was higher than the accuracy of the original FREAK descriptor, but lower than the accuracy of 512-bit CRH-FREAK descriptor. For instance, 512-bit CRH-FREAK recorded the highest accuracy (86.65%) in $I_{BR_{10}}$ lighting value condition, which was followed by CRH-FREAK (L1) (83.12%), CRH-FREAK (L2) (82.65%), CRH- FREAK (L3) (83.65%), CRH-FREAK (L4) (83.95%), and lastly, FREAK (64.29%). In other words, the developed 128-bit CRH-FREAK descriptors in this study were less accurate than the 512-bit CRH-FREAK. This can be explained by the size reduction of the descriptors, which causes the omission of certain information during the description process.

For the FREAK descriptor, a 3D virtual cube was not augmented over the input image when the input image was in a very bright condition (I_{BR_9} and $I_{BR_{10}}$) or a very dark condition (I_{DR_8}, I_{DR_9}, and $I_{DR_{10}}$). This can be linked to the low accuracy obtained by the FREAK descriptor as the lighting value of the input image increased (I_{BR_9} and $I_{BR_{10}}$) or decreased (I_{DR_8}, I_{DR_9}, and $I_{DR_{10}}$). However, as for the CRH-FREAK (L4) descriptor, a 3D virtual cube was augmented over the input image when the input image went through different lighting values, except for the $I_{DR_{10}}$ lighting value condition. This can be attributed to the dark condition ($I_{DR_{10}}$ lighting value condition), where a large proportion of the features in the input image cannot be displayed. Clearly, the accuracy of 128-bit CRH-FREAK descriptors was higher than the accuracy of the original FREAK descriptor in the presence of lighting value changes.

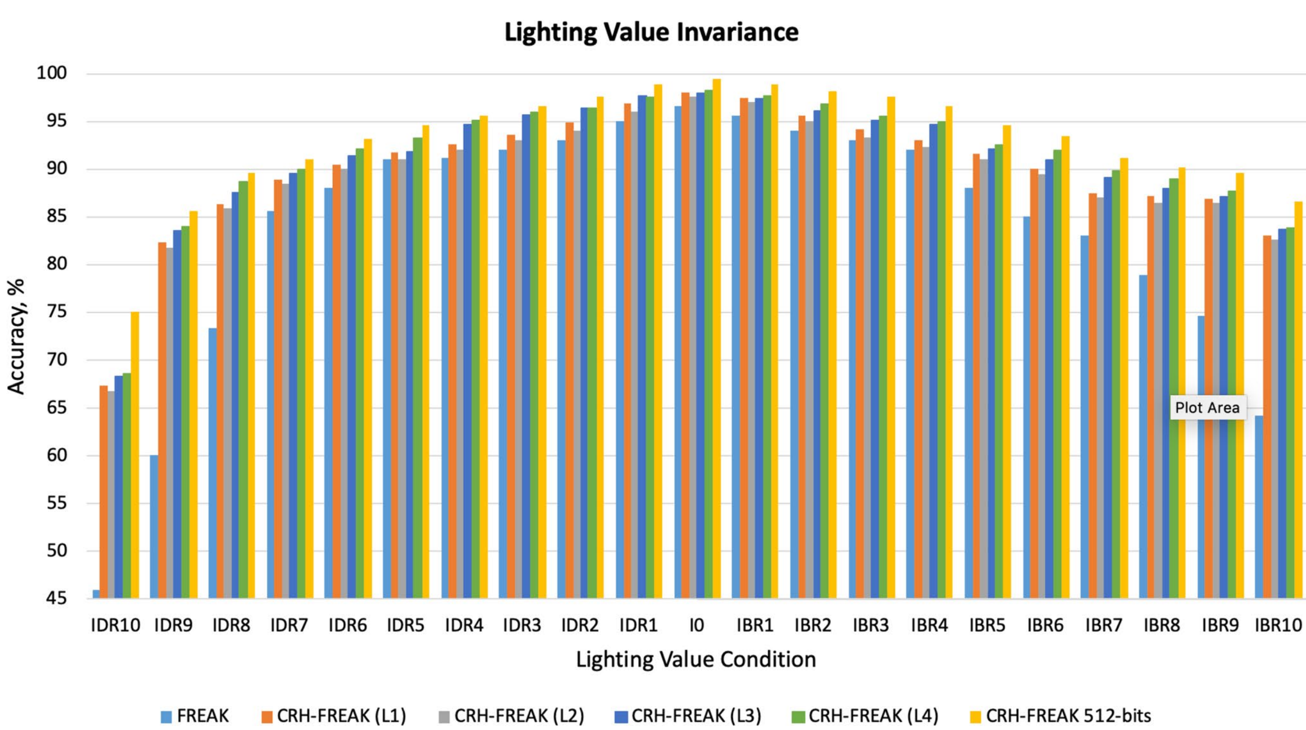


Fig. 17 Accuracy results for each descriptor in the presence of lighting value changes

6.6.1 One-way ANOVA test

One-way ANOVA test was conducted to determine whether the accuracy of the descriptors under study is significantly different in the presence of lighting value changes:

H_0 = The accuracy of FREAK, CRH-FREAK (L1), CRH-FREAK (L2), CRH-FREAK (L3), CRH-FREAK (L4), and 512-bit CRH-FREAK is not significantly different in the presence of lighting value changes.
H_1 = The accuracy of FREAK, CRH-FREAK (L1), CRH-FREAK (L2), CRH-FREAK (L3), CRH-FREAK (L4), and 512-bit CRH-FREAK is significantly different in the presence of lighting value changes.

The obtained results revealed p-value which less than 0.05 (reject H0). In other words, the accuracy of FREAK, CRH-FREAK (L1), CRH-FREAK (L2), CRH-FREAK (L3), CRH-FREAK (L4), and 512-bit CRH-FREAK was significantly different in the presence of lighting value changes.

Following that, multiple comparison analysis was conducted to determine whether the difference in accuracy of the two descriptors is significant in the presence of lighting value changes:

H_0 = The difference in accuracy of the two descriptors is not significant in the presence of lighting value changes.
H_1 = The difference in accuracy of the two descriptors is significant in the presence of lighting value changes.

The obtained results are tabulated in Table 11, which revealed that the difference in accuracy of the two descriptors (in the following comparison pair list) was significant (*p*-value of less than 0.05): (1) FREAK and CRH-FREAK (L1); (2) FREAK and CRH-FREAK (L2); (3) FREAK and CRH-FREAK (L3); (4) FREAK and CRH-FREAK (L4); (5) FREAK and 512-bit CRH-FREAK. This study demonstrated that all CRH-FREAK descriptors that made use of colour spaces were more accurate than the original FREAK descriptor that made use of greyscale only.

On the other hand, the results revealed that the difference in accuracy of the two descriptors (in the following comparison pair list) was not significant (*p*-value of more than 0.05) in the presence of lighting value changes: (1) CRH-FREAK (L1) and CRH-FREAK (L2); (2) CRH-FREAK (L1) and CRH-FREAK (L3); (3) CRH-FREAK (L1) and CRH-FREAK (L4); (4) CRH-FREAK (L1) and 512-bit CRH-FREAK; (5) CRH-FREAK (L2) and CRH-FREAK (L3); (6) CRH-FREAK (L2) and CRH-FREAK (L4); (7) CRH-FREAK (L2) and 512-bit CRH-FREAK; (8) CRH-FREAK (L3) and CRH-FREAK (L4); (9) CRH-FREAK (L3) and 512-bit CRH-FREAK; (10) CRH-FREAK (L4) and 512-bit CRH-FREAK. All four 128-bit CRH-FREAK and 512-bit CRH-FREAK descriptors in this study demonstrated high and significant accuracy, as compared to the original FREAK descriptor that made use of greyscale for the description process. Moreover, the insignificant differences in the accuracy of 128-bit CRH-FREAK and 512-bit CRH-FREAK descriptors imply that the accuracy obtained by CRH-FREAK (L1), CRH-FREAK (L2), CRH-FREAK (L3), and CRH-FREAK (L4) was comparably high as the accuracy of 512-bit CRH-FREAK.

Table 11 Results of multiple comparison analysis for the accuracy of descriptors in the presence of lighting value changes

Multiple Comparison		
Independent Variable: Lighting Value Changes		
Descriptor	Descriptor	Significant value
FREAK	CRH-FREAK (L1)	ρ ***<0.05***
	CRH-FREAK (L2)	ρ ***<0.05***
	CRH-FREAK (L3)	ρ ***<0.05***
	CRH-FREAK (L4)	ρ ***<0.05***
	512-bit CRH-FREAK	ρ ***<0.05***
CRH-FREAK (L1)	FREAK	ρ ***<0.05***
	CRH-FREAK (L2)	$\rho > 0.05$
	CRH-FREAK (L3)	$\rho > 0.05$
	CRH-FREAK (L4)	$\rho > 0.05$
	512-bit CRH-FREAK	$\rho > 0.05$
CRH-FREAK (L2)	FREAK	ρ ***<0.05***
	CRH-FREAK (L1)	$\rho > 0.05$
	CRH-FREAK (L3)	$\rho > 0.05$
	CRH-FREAK (L4)	$\rho > 0.05$
	512-bit CRH-FREAK	$\rho > 0.05$
CRH-FREAK (L3)	FREAK	ρ ***<0.05***
	CRH-FREAK (L1)	$\rho > 0.05$
	CRH-FREAK (L2)	$\rho > 0.05$
	CRH-FREAK (L4)	$\rho > 0.05$
	512-bit CRH-FREAK	$\rho > 0.05$
CRH-FREAK (L4)	FREAK	ρ ***<0.05***
	CRH-FREAK (L1)	$\rho > 0.05$
	CRH-FREAK (L2)	$\rho > 0.05$
	CRH-FREAK (L3)	$\rho > 0.05$
	512-bit CRH-FREAK	$\rho > 0.05$
512-bit CRH-FREAK	FREAK	ρ ***<0.05***
	CRH-FREAK (L1)	$\rho > 0.05$
	CRH-FREAK (L2)	$\rho > 0.05$
	CRH-FREAK (L3)	$\rho > 0.05$
	CRH-FREAK (L4)	$\rho > 0.05$

6.6.2 Mean test

Mean test was conducted to determine which descriptor has the highest accuracy in the presence of lighting value changes. The following Table 12 presents the results of mean test for each descriptor in this study.

Based on the results, 512-bit CRH-FREAK recorded the highest accuracy (93.09%) in the presence of lighting value changes, which was followed by CRH-FREAK (L4) (91.43%), CRH-FREAK (L3) (90.90%), CRH-FREAK (L1) (90.02%), CRH-FREAK (L2) (89.43%), and lastly, FREAK (83.87%). FREAK recorded a very high standard deviation, which suggests its lack of robustness against lighting value changes, especially when the lighting value increased to a very bright condition or decreased to a very dark condition. Meanwhile, all four 128-bit CRH-FREAK and 512-bit CRH-FREAK descriptors recorded low standard deviation and demonstrated their robustness against similar lighting value changes. Furthermore, the accuracy of all four 128-bit CRH-FREAK descriptors was comparably high as the accuracy of 512-bit CRH-FREAK in this study.

7 Discussion

Table 13 presents the performance ratings of FREAK, CRH-FREAK (L1), CRH-FREAK (L2), CRH-FREAK (L3), CRH-FREAK (L4), and 512-bit CRH-FREAK descriptors in terms of efficiency (computational time) and robustness against scale, rotation, and lighting (lighting colour, lighting direction, and lighting value) changes. The presented values were based on the obtained mean test results. As for the performance ratings of descriptors on computational time, descriptor that recorded the shortest computational time (lowest mean value) was rated as "1", whereas descriptor that recorded the longest computational time (highest mean value) was rated as "6". Meanwhile, as for the performance ratings of descriptors on robustness, descriptor that recorded the highest accuracy (highest mean value) was rated as "1", whereas descriptor that recorded the lowest accuracy (lowest mean value) was rated as "6". Ideally, a mobile AR application requires a descriptor that records the shortest computational time and solid robustness against scale, rotation, and lighting changes.

Table 12 Mean and standard deviation of the overall tracking accuracy in the presence of lighting value changes

Descriptor	Mean	Standard deviation
FREAK	83.866	13.383
CRH-FREAK (L1)	90.020	6.842
CRH-FREAK (L2)	89.425	6.796
CRH-FREAK (L3)	90.898	6.783
CRH-FREAK (L4)	91.428	6.744
512-bit CRH-FREAK	93.088	5.815

Referring to Table 13, the overall performance of 512-bit CRH-FREAK was the best among all descriptors in this study. This was followed by CRH-FREAK (L4), CRH-FREAK (L3), CRH-FREAK (L1), CRH-FREAK (L2), and lastly, FREAK. Despite its solid robustness, 512-bit CRH-FREAK recorded the lowest rating for computational time (the least efficient). In addition, its efficiency in terms of computational time was significantly different from the performance of other descriptors in the same aspect. On the other hand, FREAK recorded the lowest overall performance ratings, but its rating for computational time was the highest (the most efficient). Nevertheless, FREAK and all four 128-bit CRH-FREAK in this study did not display significant differences in computational time.

Meanwhile, the overall performance of CRH-FREAK (L1), CRH-FREAK (L2), CRH-FREAK (L3), and CRH-FREAK (L4) were found moderate. Moreover, there were no significant differences between these descriptors and 512-bit CRH-FREAK in terms of scale, rotation, lighting colour, lighting direction, and lighting value changes. Basically, all four 128-bit CRH-FREAK descriptors in this study performed as well as the 512-bit CRH-FREAK descriptor.

Apart from that, all 128-bit CRH-FREAK descriptors that made use of only one cascade were outperformed.

Table 13 Performance ratings of each descriptor in terms of efficiency and robustness

Descriptors	FREAK	CRH-FREAK (L1)	CRH-FREAK (L2)	CRH-FREAK (L3)	CRH-FREAK (L4)	512-bit CRH-FREAK
Aspects						
Computational time	1.0	2.0	4.0	5.0	3.0	6.0
Scale changes	6.0	5.0	4.0	3.0	2.0	1.0
Rotation changes	6.0	3.0	5.0	4.0	2.0	1.0
Lighting colour changes	6.0	5.0	4.0	2.0	3.0	1.0
Lighting direction changes	6.0	3.0	5.0	4.0	2.0	1.0
Lighting value changes	6.0	4.0	5.0	3.0	2.0	1.0
Overall performance	5.2	3.7	4.5	3.5	2.3	1.8
Performance ratings	6	4	5	3	2	1

The obtained results revealed that 128-bit CRH-FREAK descriptors that described fine information (cascade 3 and cascade 4), specifically CRH-FREAK (L3) and CRH-FREAK (L4), achieved higher robustness than 128-bit CRH-FREAK descriptors that described coarse information (cascade 1 and cascade 2), specifically CRH-FREAK (L1) and CRH-FREAK (L2). This can be explained by the fact that fine (detailed) information are more in-depth than coarse information. In particular, CRH-FREAK (L4) that described fine information was ranked first. Its computational time was as short as the original FREAK descriptor, and its robustness against scale, rotation, and lighting changes were comparable to the 512-bit CRH-FREAK descriptor. With that, CRH-FREAK (L4) was identified as the most appropriate descriptor for mobile AR applications.

8 Conclusions

AR has recently gained immense popularity in research given its significant impact on our daily lives. Technological advancements have shifted AR from its initial platforms via desktops or personal computers to lower-end mobile devices, such as smartphones. Focusing on the fundamental needs of mobile AR applications, this study attempted to improve the overall tracking process through the development of an efficient and robust FREAK-based descriptor. A comprehensive review of literature on the key components of tracking process was conducted to provide a better understanding on the key challenges of mobile AR applications. Some of the identified key challenges of mobile AR applications include real-time functionality and robustness against various changes, such as scale, rotation, lighting colour, lighting direction, and lighting value changes, which were addressed in this study.

For this study, 512-bit CRH-FREAK descriptor that made use of six colour spaces was first proposed. However, this descriptor required longer computational time to describe features, which was higher than the computational time required by the original FREAK descriptor (that made use of greyscale only) by 3.7 times. Therefore, this study opted to reduce the number of bits for the proposed CRH-FREAK from 512-bit to 128-bit. As a result, all 128-bit CRH-FREAK descriptors recorded shorter computational time. Moreover, FREAK and 128-bit CRH-FREAK did not display significant differences in computational time.

This study successfully proved the efficiency (computational time) and robustness of 128-bit CRH-FREAK descriptor against scale, rotation, and lighting (lighting colour, lighting direction, and lighting value) changes. Considering the significance of the tracking performance in mobile AR applications, the obtained results and findings of this study on descriptors were deemed significant in the AR domain. This study was expected to present a new direction of image recognition element in the fields of image recognition and computer vision.

Acknowledgement This work was supported by the Ministry of Higher Education Malaysia (FRGS/1/2018/ICT01/UKM/02/5).

Declaration

Conflict of interest The authors have no conflicts of interest to declare.

References

Alahi A, Ortiz R, Vandergheynst P FREAK: Fast retina keypoint. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2012. IEEE, pp 510–517

Andono PN, Pramunendar RA, Supriyanto C, Hariadi M, Kondo K, Less DN (2014) An Evaluation of Color SIFT Descriptors for Underwater Images Proceedings of the Fourth IIEEJ International Workshop on Image Electronics and Visual Computing

Azuma RT (1997) A survey of augmented reality. Presence Teleoperators Virtual Environ 6:355–385

Bay H, Ess A, Tuytelaars T, Van Gool L (2008) Speeded-up robust features (SURF) computer vision and image understanding, vol 110, pp 346–359. doi:https://doi.org/10.1016/j.cviu.2007.09.014

Bernal J, Vilarino F, Sánchez J (2010) Feature detectors and feature descriptors: where we are now

Bianco S, Mazzini D, Pau D, Schettini R (2015) Local detectors and compact descriptors for visual search: a quantitative comparison. Digital Signal Process 44:1–13

Bleser G (2009) Towards visual-inertial slam for mobile augmented reality. der Technischen Universität Kaiserslautern

Bolyós C (2013) Scarlet–Real Time Mobile Augmented Reality Library. Paper presented at the Proceedings of CESCG 2013: The 17th Central European Seminar on Computer Graphics, Smolenice, Slovakia, 28–30 April 2013

Calonder M, Lepetit V, Özuysal M, Trzcinski T, Strecha C, Fua P (2012) BRIEF: Computing a local binary descriptor very fast IEEE Transactions on Pattern Analysis and Machine Intelligence, vol 34, pp 1281–1298. doi:https://doi.org/10.1109/TPAMI.2011.222

Cruz E, Orts-Escolano S, Gomez-Donoso F, Rizo C, Rangel JC, Mora H, Cazorla M (2019) An augmented reality application for improving shopping experience in large retail stores. Virtual Reality 23:281–291

Fan P, Men A, Chen M, Yang B Color-SURF: A surf descriptor with local kernel color histograms. In: 2009 IEEE International Conference on Network Infrastructure and Digital Content, 2009. IEEE, pp 726–730

Flavián C, Ibáñez-Sánchez S, Orús C (2019) The impact of virtual, augmented and mixed reality technologies on the customer experience. J Bus Res 100:547–560

Geusebroek JM, Burghouts GJ, Smeulders AWM (2005) The Amsterdam library of object images. Int J Comput Vis. https://doi.org/10.1023/B:VISI.0000042993.50813.60

Guan W, You S, Newmann U (2012) Efficient matchings and mobile augmented reality ACM transactions on multimedia computing, communications, and applications (TOMM). Vol 8, p 47

Hua J-Z, Liu G-H, Song S-X (2019) Content-based image retrieval using color volume histograms. Int J Pattern Recognit Artif Intell 33:1940010

Ihsan R, Sehat U (2013) A survey on augmented reality challenges and tracking acta graphica znanstveni časopis za tiskarstvo i grafičke komunikacije, vol 24, pp 29–46

Kottman M (2011) The color-BRIEF feature descriptor. in: spring conference on computer graphics SCCG, pp 28–30

Koyasu H, Nozaki K, Maekawa H (2014) Evaluation of image feature descriptors for marker-less AR applications advances in visual computing

Lam MC, Sadik MJ, Elias NF (2020a) The effect of paper-based manual and stereoscopic-based mobile augmented reality systems on knowledge retention. Virtual Reality. https://doi.org/10.1007/s10055-020-00451-9

Lam MC et al (2020b) Interactive augmented reality with natural action for chemistry experiment learning. TEM J 9:351

Leutenegger S, Chli M, Siegwart RY (2011) BRISK: Binary Robust invariant scalable keypoints, pp 2548–2555. doi:https://doi.org/10.1109/ICCV.2011.6126542

Lowe G (2004) SIFT—The scale invariant feature transform. Int J 2:91–110

Mahieu R, Tilak H (2015) Real-time mobile augmented reality using markerless subject tracking. Standford University Department of Electrical Engineering, Standford

Majid NA (2018) Augmented reality to promote guided discovery learning for STEM learning. Int J Adv Sci Eng Inf Technol 8:1494–1500

Markatopoulou F, Pittaras N, Papadopoulou O, Mezaris V, Patras I (2015) A study on the use of a binary local descriptor and color extensions of local descriptors for video concept detection. In: MultiMedia Modeling, Springer, pp 282–293

Mikolajczyk KSC (2005) A performance evaluation of local descriptors. IEEE Trans Pattern Anal Mach Intell 27:1615–1630. https://doi.org/10.1109/TPAMI.2005.188

Nguyen TV, Tan D, Mirza B, Sepulveda (2016) J MARIM: mobile augmented reality for interactive manuals. In: Proceedings of the 2016 ACM on multimedia conference, 2016. ACM, pp 689–690

Obeidy WK, Arshad H, Chowdhury SA, Parhizkar B, Huang J (2013) Increasing the tracking efficiency of mobile augmented reality. In: International visual informatics conference, 2013. pp 447–457. doi:https://doi.org/10.1007/978-3-642-41939-3

Over P et al (2013) Trecvid 2013—an overview of the goals, tasks, data, evaluation mechanisms and metrics. In: Proceedings of TRECVID 2013, NIST USA, 2013 2013. pp 1–45

Rublee E, Rabaud V, Konolige K, Bradski G (2011) ORB: An efficient alternative to SIFT or SURF. pp 2564–2571. doi:https://doi.org/10.1109/ICCV.2011.6126544

Satoh K, Anabuki M, Yamamoto H, Tamura H (2001) A hybrid registration method for outdoor augmented reality. In: IEEE and ACM international symposium on Proceedings of augmented reality, IEEE, pp 67–76

Sung Y-T, Chang K-E, Liu T-C (2016) The effects of integrating mobile devices with teaching and learning on students' learning performance: a meta-analysis and research synthesis. Comput Edu 94:252–275

Tan SY, Arshad H, Abdullah A (2018) An efficient and robust mobile augmented reality application . Int J Adv Sci Eng Inf Technol 8:1672–1672. https://doi.org/10.18517/ijaseit.8.4-2.6810

Tan SY, Arshad H, Abdullah A (2019) Distinctive accuracy measurement of binary descriptors in mobile augmented reality. PLoS ONE, p 14

Tang G, Liu Z, Xiong J (2019) Distinctive image features from illumination and scale invariant keypoints. Multimedia Tools Appl 78:23415–23442

Uchiyama H, Marchand E (2012) Object detection and pose tracking for augmented reality: recent approaches foundation in computer vision, pp 1–8

Ufkes A, Fiala M A markerless augmented reality system for mobile devices. In: 2013 International conference on computer and robot vision (CRV), IEEE, pp 226–233

Van De Sande KE, Gevers T, Snoek CG (2010) Evaluating color descriptors for object and scene recognition. IEEE Trans Pattern Anal Mach Intell 32:1582–1596

Wafy M, Madbouly A (2015) Increase Efficiency of SURF using RGB Color Space. Int J Adv Comput Sci Appl (IJACSA) 6:81–85

Wu J, Cui Z, Sheng VS, Zhao P, Su D, Gong S (2013) A Comparative Study of SIFT and its Variants Measurement. Sci Rev 13:122–131

Yang C-K, Chen Y-H, Chuang T-J, Shankhwar K, Smith S (2019) An augmented reality-based training system with a natural user interface for manual milling operations. Virtual Reality, pp 1–13

Zhang L, Gao G, Zhou C, Cui Z, Wang L (2018) An Efficient Feature Extraction Scheme for Mobile Anti-Shake in Augmented Reality. Tehnički vjesnik 25:1119–1124

Zhou J, Cadavid S, Abdel-Mottaleb M (2011) Exploiting color sift features for 2d ear recognition. In: 2011 18th IEEE international conference on image processing, 2011. IEEE, pp 553–556