Indonesian Journal of Electrical Engineering and Computer Science
Vol. 26, No. 3, June 2022, pp. 1589~1598
ISSN: 2502-4752, DOI: 10.11591/ijeecs.v26.i3.pp1589-1598

Overcoming camera instability problem for detecting and tracking moving objects in video using reduced data

Jalal H. Awad[1], Balsam D. Majeed[2]
[1]Department of Computer, College of Science, Mustansiriyah University, Baghdad, Iraq
[2]Department of Computer techniques Engineering, Imam Kadhim Faculty of Islamic Sciences University, Baghdad, Iraq

Article Info

Article history:

Received Oct 20, 2021
Revised Mar 28, 2022
Accepted Apr 2, 2022

Keywords:

Bayer patterns
Frame difference
Moving objects detection
Singular value decomposition
Speed up robust feature transform
Tracking

ABSTRACT

Moving objects detection is a vital field of study in various applications. Many of such applications may have to capture and process a lot of data, then such these data need to be reduced as much as possible in order to have a reasonable and suitable system for achieving the desired aims efficiently. The proposed algorithm utilizes singular value decomposition (SVD) and Bayer pattern filter for their good properties in producing very representative reduced data. This data is then handled by frame difference objects detection, which in turn is an approach that doesn't need to handle much data. The camera shaking which can be caused by a windy weather in the case of the outdoor static camera may introduce a frame difference with imprecise moving objects detection, hence frames compensation is conducted utilizing a transformation based on speed up robust feature transform (SURF) detected key points.

Corresponding Author:

Jalal H. Awad
Department of Computer, College of Science, Mustansiriyah University
Baghdad, Iraq
Email: jalalhameed@uomustansiriyah.edu.iq

1. INTRODUCTION

The first step for each video surveillance system is the moving object detection [1]. Intuitively object detection is the preliminary stage for the successive tracking operation [2]-[4]. Employing human beings is inefficient way for monitoring places because of their limited capabilities and high cost demands as compared with the automated surveillance systems [5]. When using a static camera, an outdoor environment may cause camera shaking. Sometimes, cameras are installed on mobile machines. Thus advanced algorithms should be implemented to deal with finding the moving objects with such hypothesis [6]. Video surveillance with the utilization of moving cameras requires some compensation operations before starting in the moving object detection operation [7]. It is a challenge matter to detect moving objects with mobile camera or camera shaking. Tracking of the moving objects can give good results when offering suitable object location and shape information. Many articles depend on camera motion compensation as a pre-step for moving objects detection, where each incoming frame in the video stream is registered with the corresponding background model or previous frame. Such registration helps to determine a 2D compensation transformation matrix [8]. Image registration is used to estimate the relationship between two images of the same scene with relative shift, rotation or any affine transformation between them. Thus, dealing with the problem of detecting moving objects under moving camera is more challenging matter than with static one [8].

One way to match between images or objects in these images is through the use of key points [9]. Extracting key points from images is very useful in a lot of applications; like objects detection and tracking in images. It can be utilized in images registration, where it is possible to identify the same objects across multiple images [10]. The most famous algorithms for detecting and describing these key points are scale invariant feature transform (SIFT) and speed up robust feature transform (SURF) due to their robustness, effectiveness, less time consumption and complexity [9]. SIFT has the ability to identify the same key points even with image exposing to various transformations like, scaling, and rotation. Without affecting the descriptive property of the feature [10].

SURF has less vector dimension, it has more efficient computation capabilities [9], [11]. As the SURF has the privilege of less vector dimension, this can be boosted through exploiting another approach in order to get more vector reduced dimension. Ignoring the less informative features and preserving the most important ones is a reasonable and logical way in dealing with data, this is called data reduction. Such approach belongs to the principle component analysis (PCA) family which is a data reduction approach. According to this approach, the Eigen values and their corresponding Eigen vectors are extracted for a signal (for example an image) and preserved as the most important features, with a projection of the remaining features as the less important features on these Eigen vectors space [12]. Doing this will permit for dealing with few data (Eigen vectors) and at the same time preserving good signal main features quality. Another data reduction approach is the independent components analysis (ICA) [1], [7]. As well as there is Another important data reduction tool which is singular value decomposition (SVD) that has good energy compaction and stability properties which make it widely used for many image processing applications [13], [14].

In video files, an extra data reduction can be achieved using a trick of neglecting and skipping some of the intermediate frames which haven't much importance due to the small-time interval between them, so as there isn't any considered alteration in between such these frames. Sometimes the reduction can be gotten by reducing the frame size [15]. Hence any signal processing operation, such as retrieving, classification, and matching. Can be done elegantly depending on such reduced features space.

In the consideration of video files, one frame in response to other frame may goes more than one displacement type, an affine transform for example which includes rotation, scaling, translation and Cartesian transformation when captured with handheld, shaking or moving camera [16]. Therefore, this paper is dedicated for dealing with the problem of detecting moving objects under such circumstances with utilizing compact data as much as possible.

Thus, with the utilization of the above-mentioned approaches and tools, many attempts have been done to deal with moving camera or data reduction. Oji [10] affine scale invariant feature transform (ASIFT) is used to deal with detecting the objects in multiple images even with some transformation existence like translation, scaling, and rotation. the proposed approach gives good objects boundary determination due to the use of a region merging segmentation algorithm. While [17] Proposed an improved SURF (faster and has reduced data dimensions than SIFT) algorithm which reduces the unnecessary detected key points and thus reduce the computations. It may give best results even than a SURF supported with RANSAC algorithm. On the other hand [9] Combines both SURF and Meanshift algorithm in which a search window is placed arbitrarily in a search area, with an attempt to continuously adjust this window position based on translating its center to a new centroid representing the mean of the samples under this window [18]. Used a hybrid approach in which SURF for key points detection and improved CAMshift algorithms are combined together to get best moving objects detection and tracking results. Improved CAMshift algorithm is based on its previous version which is called the Meanshift. CAMshift tries to best fit this window size and orientation based on the samples under consideration after each Meanshift convergence. Another approach is used in [5] where it tries to use the optical flow approach in order to compensate for the camera motion, where it can be used to determine the speed and direction (velocity) of each object in consecutive frames. This approach depends on the idea of observing things through car window, where the nearby objects (moving) appear to move faster (have high scale vectors) than the faraway ones (background) objects. In this paper interest points are found first in each two consecutive frames and matching them together in order to find the optical flow vectors after frames compensation based on these matched key points. The proposed approach has limited accuracy in determining the moving objects boundaries [7]. Depended similar approach, the method works well in real time [19]. Used singular value decomposition (SVD) in order to have compact representation and hence reduced computations for the data [20]. Discussed the family of principle components analysis (PCA), robust PCA as well as SVD for their role in moving objects detection with reduced data dimension and computation.

2. MOVING OBJECT DETECTION APPROACHES

The main approaches for moving object detection are background subtraction, optical flow, features detection and frame difference [17], [21]. According to the first approach, a background model is built to mimic the static objects in the video scene. The background approach has very good ability for moving object detection as well as moderate computations. According to this approach a comparison is conducted between the current frame and the background model in order to identify their difference as moving objects [18]. This approach may fail to handle abrupt illumination change, hence the comparison of each background pixel with its corresponding current frame highly illuminated pixel leads to high difference. The other lack of this approach is its inability to deal with movement turbulence like for example moving trees or a sudden appearance or disappearance of objects. It also requires several initial frames for training to make this model which represents an extra time preventing from implementing such approach in real or low capability systems. Thus, one solution is to use a compressed form of the data [1]. Sometimes the background can be modeled as a mixture of Gaussians (MoG) by observing each pixel values distribution and then recognizing each pixel as either background or moving object pixel according to the deviation from this model. But also, such systems are vulnerable to illumination changes due for example for moving clouds or on/off light switching and also it is complex and time consuming [19].

Background model subtraction cannot work with the moving or shaking camera [22]. The optical flow approach is more complicated, hasn't good ability to deal well with noise and has high time consumptions, thus inability to be implemented in environments with real time requirements. The feature detection approach depends on corners detection and texture features but it is also sensitive to noise. In [23] different feature detection approaches were studied and used like Harris corner detection, scale invariant feature transform (SIFT) and speed up robust feature transform (SURF). For harris corner detection, a convolution window is used to detect drastic intensity change in all directions which indicate a corner (key feature) point. The window must be translated pixel by pixel, crossing the whole image rows and columns in order to detect all the corner points in the image. Feature detection can be utilized for even moving background because such points are scale invariant. With SIFT, many steps have to be done in order to identify key points. First, the addressed image has to be scaled and blurred in many scales, then difference of Gaussian (DoG) of the neighbor scales are taken, and the candidate key points are determined as minima/maxima with respect to the other local eight neighbor pixels, as well as such neighbors in the upper and lower scale levels. Then each key point has to be described with appropriate orientation and descriptor in order to be identified later even when they undergo some transformation types. SURF has a lot in common with the SIFT steps but with reduced data dimensions and hence reduced processing time. It also has high immunity against noise.

Real time systems can utilize Frame difference approach for its simplicity and low time consumptions but with low detection accuracy [17]. In this approach, two successive frames are subtracted and using their absolute difference to indicate the motion in the scene. Abrupt Illumination change which can affect other approaches hasn't any effect on the detection result due to the use of two successive frames which have very small-time interval between them, so it has well adaptability for the dynamic environment. Other main advantages for such approach is the easiness of implementation with low complexity, computation time. It also doesn't need high storage requirements due to the use of just the consecutive frames. The main considered disadvantage for this approach is its inability to detect the interior region (the overlapped moving object region which has the same intensity value in both frames) of the moving objects this problem is called the cavity phenomenon where just the moving objects contours are determined. The selection of the between frames time interval and segmentation threshold control these affections. Unfortunately, in spite of its privileges, such approach may fail when there is a camera shake [12]. In frame difference approach the previous frame for the current one can be used as a background model. The moving object speed and the utilized threshold effect the detection accuracy [18]. The motion mask in the frame difference approach is identified whenever there is a deviation from the (1):

$$B_t(k,l) = I_{t-1}(k,l) \tag{1}$$

where $B_t(k,l)$ is the previous frame at pixel location$(k,l\ )$ and time t which is considered as background. And $I_{t-1}(k,l)$ is the current frame at the same pixel location [21].

3. GEOMETRIC DISTORTION

Image capturing process utilizing cameras may lead to some geometric distortion in the captured images. Such distortion can be expressed mathematically through a transformation model [10]. Briefly, these before mentioned transformations and the others are also utilized in a reverse manner to derive the homography matrix to compensate the various image distortions [24].

4. SPEED UP ROBUST FEATURE TRANSFORM (SURF)

An important primitive step in any object detection approach is the features detection. Where such these features are unique and recognizable for this object [25], [26]. SURF is an approach which is invariant to scale, rotation and translation of the images or objects inside these images. It doesn't need a lot of computation time [18], [27]. It depends on Hessian matrix determinant to find the key points.

5. BAYER COLOR FILTER ARRAY (CFA)

Bayer pattern is usually used in the single sensor cameras as a color filter array (CFA). CFA is a mosaic of red, green and blue color components. Thus an interpolation operation is required to get the complement of the two remaining components for each spatial location sample in this pattern before getting the full color (RGB) image. Therefore neighbor locations in this filter will pass the corresponding color in this pattern [28]-[30].

6. SINGULAR VALUE DECOMPOSITION (SVD)

Some authors supposed that transformation domains like for example DWT, DCT or SVD are illumination invariant [31]. Singular value decomposition is an approach for matrix (for example an image) decomposition. If it assumed that the matrix to be decomposed is A, then the SVD for this matrix can be given by (2):

$$A = USV^t \tag{2}$$

where, U and V are orthogonal matrices of M*M and N*N dimension respectively. While S is a diagonal matrix with a dimension of M*N. Thus it is possible to get the U and V matrices, by utilizing the Eigen vectors of AA^t and A^tA respectively [32], [33]. Thus:

$$AA^t = USV^t(USV^t)^t = US^2U^t \tag{3}$$

$$A^tA = (USV^t)^tUSV^t = VS^2V^t \tag{4}$$

the Eigen values of S matrix are the square roots of either AA^t or A^tA singular values. The main image information are contained in the singular vector matrices [33]. U and V matrices hold the most important information (Eigen vectors) of the analyzed matrix (an image for example). While S matrix holds the less important information (Eigen values). Thus in the inverse operation of SVD (namely ISVD), the U and V matrices have the most important role than S matrix in reconstructing the image [32], [34].

7. TRACKING

Object tracking is an advanced operation in the video surveillance systems [22]. Several features can be used for image-region correlation like color, texture, intensity and histograms [31]. For each tracked moving object, its features required to be updated progressively [7].

8. PROPOSED METHOD

The block diagram for the proposed approach is given in Figure 1. The proposed approach can be well explained through dividing its operations into the following phases:

- Phase 1: reading two frames from the video sequence, considering one of them as a reference frame for comparison with the other read one as a current frame for the sake of discovering moving objects in this current frame. As there isn't real difference between the immediate consecutive frames, the selected read frames can be chosen with relative time interval separation between them by skipping some of them through the reading operation. This leads to decreased processed data and hence decreased processing time intervals. Thus, for example after reading frame '1', instead of reading frame '2 ', it may be more suitable to read frame '5'.
- Phase 2: converting the two frames into grayscales in order to reduce the data as more as possible by dealing just with one layer instead of three layers. This will be done by utilizing the Bayer pattern instead of using the conventional way for getting a grayscale image from a colored one. The well-known RGB to grayscale conversion (5) is given:

$$grayscale = 0.2989 * \text{red} + 0.5870 * \text{green} + 0.1140 * \text{blue} \tag{5}$$

While the Bayer CFA uses the (6).

$$grayscale = Bayer\ red\ channel + Bayer\ green\ channel + Bayer\ blue\ channel \tag{6}$$

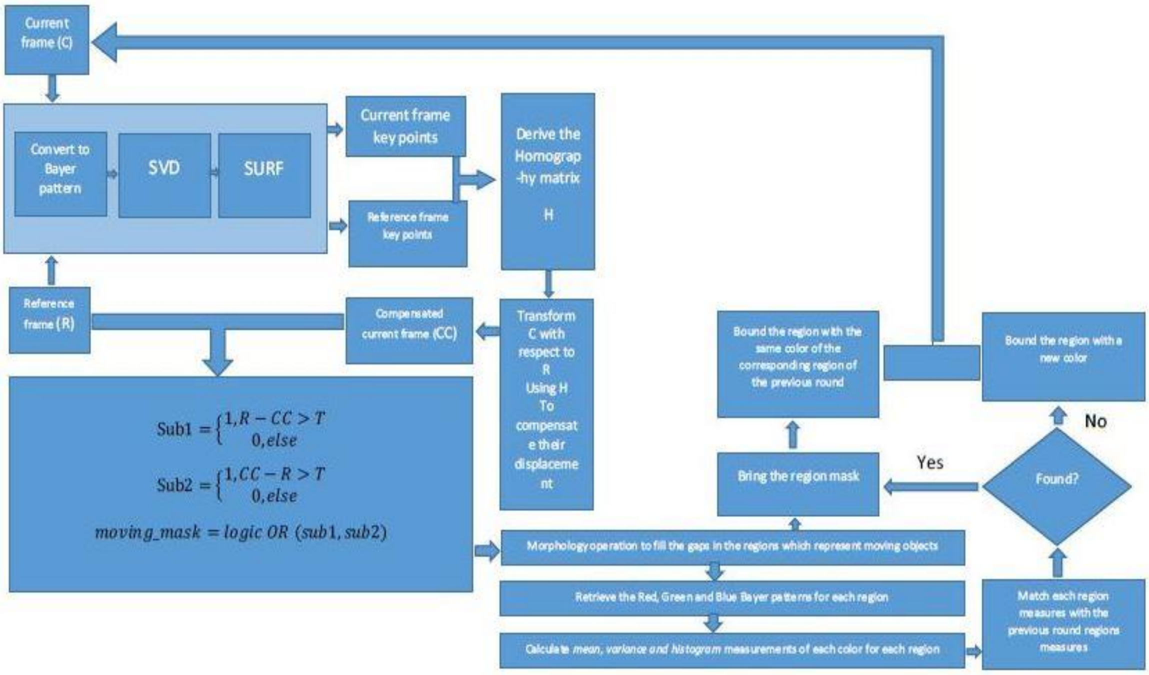

Figure 1. Block diagram of the proposed system

The reason for using this latter approach is its ability to retrieve the color information in contrast to the conventional way which loses this privilege. In the proposed approach this color information is of high importance in the process of moving objects tracking across the successive frames of the video sequence, where some measurements for them are depended for comparisons.

– Phase 3: extracting the principle components of the resulted grayscale frame through the use of the singular value decomposition (SVD) which decomposes the frame into its Eigen vectors that are ordered from left to right according to their importance (data reduction). Thus, specific number of such vectors can be used in an inverse operation of the SVD to get an estimation for this frame. This acts as a data reduction approach which in turn contributes in the reduction of the time complexity.
– Phase 4: in order to compensate any camera motion that may be occurred during video capturing, it is necessary to find the key points in both the reference and the current frames accompanied with these key points' descriptors for the matching process. After finding the corresponding matched key points in both frames, a registration process has to be done using at least four (sufficient for describing an affine transformation) pairs of such key points in the aim of deriving homography matrix coefficients which describe the transformation that is occurred in the coordinates of one frame in reference to the other one coordinates.
– Phase 5: now using the derived homography matrix H, it is easy to compensate the casual undesired motion of a frame by applying the transformation that is described by the H coefficients.
– Phase 6: according to the frame difference approach in order to find the moving objects, the two frames are subtracted from each other. Due to the drawback of this approach in extracting the whole moving object, in this paper the reverse difference is also applied to increase the extracted moving area boundaries. The two differences are conducted according to (7) and (8).

$$\text{Sub1} = \begin{cases} 1, frm1 - frm2 > T \\ 0, else \end{cases} \tag{7}$$

$$\text{Sub2} = \begin{cases} 1, frm2 - frm1 > T \\ 0, else \end{cases} \tag{8}$$

After finding this net difference, a binarization according to a threshold and then binary OR operations between the two differences are used to get the preliminary moving objects mask. This is followed by dilation morphology process to fill the gaps and connect unconnected regions of the moving object mask.

– Phase 7: each connected region is considered as a moving object, hence its color information should be retrieved in order to be considered for comparison with the next rounds' objects for tracking. Then the mean, variance and the histogram for each moving object pixels are calculated and saved for the comparison purpose with the moving objects of the current frame of the next round. The moving objects of the two consecutive rounds with the minimum absolute difference are considered as the same. The remaining moving objects with no-matches should be rounded with bounding boxes of new colors and save their metrics in companying with their gained colors for the subsequent round.

SURF has three steps which are feature extraction, feature description and matching [21], [22]. One approach for extracting key points is by using the scale space extremum concept. According to mathematics terminology, extremum means the maximum or minimum value according to application. This approach uses Gaussian function in order to convolve with the intended image to have a blurred image version. Of course, this help to reduce fine details and noise effects as the first step in the process of key point detection. Gaussian function is given in the (9):

$$G(x, y, \sigma) = \frac{1}{2\pi\sigma^2} e^{-(x^2+y^2)/2\sigma^2} \tag{9}$$

The next step in the aim of finding the key points is to down sample the original image by a factor of 2 through taking the intensity average of every 2*2 pixels of the original image to represent one pixel in the next octave (down sampled original image). This is repeated for a predefined number of octaves [23].

SURF had been developed so as to provide enhanced time complexity and compactness of features vector dimensions as compared with scale invariant feature transform (SIFT). In order to find key points, the derivative of Gaussian is used. SIFT uses less time consumption approximation for such derivative by utilizing the difference of Gaussian (DoG) in order to avoid the complex derivation operations. SURF goes an extra enhancement step in this trend by utilizing approximated box filters to simulate the DoG effect. Thus instead of being forced to make multiple levels of the image pyramid, an approximated box filter pyramid is derived based on an initial Gaussian sigma value to mimic the DoG effect. SURF also adopts the idea of integral image to summarize the Gaussian averaging steps into just one step (thus getting time complexity of O(1) instead of O(n^2)), and at the same way applying all the box filters in the pyramid simultaneously instead of successive approach as in the DoG. Another recognizable SURF feature is the Lablacian sign (the trace of the Hessian matrix) which is added as a boost feature with each key point vector to support the key point discriminative information. Wavelet is used in order to find the gradient in both x and y directions and using them to determine the dominant orientation of each key point. This is done through using a Gaussian filter centered at the key point to give it the most importance than its surrounding neighbors. SURF depends on Hessian matrix determinant (Eigen values multiplication) and trace (Eigen values addition) to find the key points:

$$H(x, \sigma) = \begin{matrix} L_{xx}(X, \sigma) & L_{xy}(X, \sigma) \\ L_{xy}(X, \sigma) & L_{yy}(X, \sigma) \end{matrix} \tag{10}$$

where, H is the Hessian matrix. $L_{xx}(x, \sigma)$ is the convolution of the image with the Gaussian partial second derivative in the point X with respect to x, while $L_{yy}(X, \sigma)$ is with the Gaussian partial second derivative in the point X with respect to y, and lastly $L_{xy}(X, \sigma)$ is with the Gaussian partial first derivative in the point X with respect to x and then with respect to y.

The third step is to find the key points descriptors. Here wavelet transform is used in order to provide feature description. Such features must be invariant to scale, shift, rotation, affine transformation and intensity, they must also have the repeatability, stability and low computation requirements [21], [22]. thus even in the case of camera instability and hence frames extreme sever of different kinds of such mentioned transformations, it is possible to turn back the transformed (distorted) frame into its adequate pose through utilizing these descriptors to find the corresponding (peer) points in both the transformed and the non-transformed frame in order to depend these points values as basis for forming some equations for the purpose of estimating an inverse transformation to refix the distorted frame pose.

Analyzing (decomposing) the image into its U, V, and S components makes things easier to deal just with the most important information (eigen vectors), thus as these vectors are ordered descending according to their importance, it is easy to select the most important ones to reconstruct later an approximation of the original image, hence in such way it is possible to control the level of the processed information quantity. In this case SVD has an important role in reducing the required data processing demands, which is one of the main aims in this research. The pseudo code for the proposed algorithm is given in Figure 2.

Algorithm: the pseudo code of moving object detection and tracking

1: read two consecutive frames, Reference Frame (*RF*) and Current Frame (*CF*).

2: convert *RF* and *CF* into grayscale using the Bayer pattern such that:

- The *red channel* = red intensities in the pattern locations, other locations=0
- The *green channel* = green intensities in the pattern locations, other locations=0
- The *blue channel* = blue intensities in the pattern locations, other locations=0
- $grayscale\ (Grayscale\ Reference\ Frame(\boldsymbol{GRF})or$
- $Grayscale\ \ Current\ Frame(\boldsymbol{GCF})) =$
  $Bayer\ red\ channel$(Reference Frame($\boldsymbol{RF}$)or Current Frame($\boldsymbol{CF}$) +
  $Bayer\ green\ channel(\boldsymbol{RFor\ CF}) + Bayer\ blue\ channel(\boldsymbol{RFor\ CF})$.

3: get [*U S V*] = SVD ($\boldsymbol{GRFor\ GCF}$) through the following:

- Compute its transpose $(\boldsymbol{GRFor\ GCF})^t$ and $(\boldsymbol{GRFor\ GCF})^t(\boldsymbol{GRFor\ GCF})$.
- Determine the eigenvalues of $(\boldsymbol{GRFor\ GCF})^t(\boldsymbol{GRFor\ GCF})$ and sort these in descending order, in the absolute sense. Square roots these to obtain the singular values of $(\boldsymbol{GRFor\ GCF})$.
- Construct diagonal matrix S by placing singular values in descending order along its diagonal. Compute its inverse, S^{-1}.
- . Use the ordered eigenvalues from step 2 and compute the eigenvectors of $(\boldsymbol{GRFor\ GCF})^t(\boldsymbol{GRFor\ GCF})$. Place these eigenvectors along the columns of V and compute its transpose, V^T.
- Compute U as U = $(\boldsymbol{GRFor\ GCF})VS^{-1}$.

4: rebuild $\boldsymbol{GRFand\ GCF}$ ← Inverse SVD $(\boldsymbol{GRFor\ GCF})$ = ***U S V***^{T} for just *30* vectors of ***U, S*** and ***V***.

5: detect *SURF* Features($\boldsymbol{GRFand\ GCF}$).

6: match *SURF* Features($\boldsymbol{GRFand\ GCF}$), then estimate Geometric Transform (*TR*) between them.

7: warp $\boldsymbol{GCF}$ in reference to $\boldsymbol{GRF}$ or vice versa ← *TR* ($\boldsymbol{GRFor\ GCF}$).

8: make the frame difference:

$$\boldsymbol{Sub1} = \begin{cases} 1, \boldsymbol{GCF} - \boldsymbol{GRF} > Threshold \\ \quad 0, else \end{cases}$$

$$\boldsymbol{Sub2} = \begin{cases} 1, \boldsymbol{GRF} - \boldsymbol{GCF} > Threshold \\ \quad 0, else \end{cases}$$

9: Moving objects *mask*= dilation morphology (binary ($\boldsymbol{Sub1}$) Boolean OR binary ($\boldsymbol{Sub2}$)).

10: for each connected component pixels in the *mask* and round *R*:

- Retrieve the Bayer pattern *red, green* and *blue* channels.
- Calculate *mean, variance* and *histogram* for each channel.
- *Similarity* = abs (abs ($mean_R - mean_{R+1}$) + abs ($variance_R - variance_{R+1}$) + abs($histogram_R - histogram_{R+1}$)).
- Identify each connected components with minimum *Similarity* as the same moving object in the successive frames.

Figure 2. The proposed algorithm pseudo code

9. EXPERIMENT RESULTS AND DISCUSSION

9.1. Grayscale and bayer patterns

The use of grayscale one layer versions of the processed frames leads to a lot of reduction of processing time as compared to the colored of 3 layers versions. This is usually done through the utilization of the standard RGB to gray scale formula. So the one layer grayscale version according to this standard formula for a colored pixel of red, green and blue intensity values of 32, 230, and 96 respectively is 155.5188. Coincidently the grayscale version for other colored pixel with of red, green and blue intensity values 8, 255, and 24 respectively is 154.8122 which is approximately equal to the previous grayscale pixel intensity value. As a consequence it isn’t possible to know what were the original colored channels values that may lead to the same grayscale values, and as there isn’t a reverse operation to get back from grayscale to colored pixels, then the use of this approach in getting the grayscale version will indeed causes a lot of degradation in operations encompass matching processes such as tracking of moving objects in a sequence of video frames. The mitigation to this problem is through the use of the Bayer pattern (utilizing (6)) to get the one layer frame version while preserving the ability to getting and handling the three channels information separately for a better matching process than that of the standard formula.

9.2. Two levels of data reduction

This paper goes far by implementing the SVD to get just the most important vectors of the resulted frame from the previous step to be processed without losing frames' worth mentioned information. In order to test if the resulted frame is qualified for the various objectives. Usually in image processing operations and after conducting some modifications on it, an image to be accepted for the further processing should have a PSNR of at least 28 db. Table 1 shows the calculated values of these two metrics for the proposed approach resulted frames. As these values are in the accepted range, it is possible to utilize the frames confidently.

Table 1. PSNR and SSIM metrics values for the proposed approach

Argument1	Argument2	PSNR	SSIM
Grayscale of frame j	Bayer version of frame j	30.3049	0.7682
Bayer version of frame j	Reconstructed fame j using just 60 vectors	33.5876	0.9104

Global or local frame pixel values can be translated into another more informative form which is known as feature space, such features are stored as vectors that is considered as a reduced frame representation. Color histograms and color moments (mean, variance, and standard deviation) are the most common used color features. Color features are invariant to scaling, rotation and translation of the scenes. Similarity between frames can be calculated by measuring the similarity between their feature vector representations using any of the well-known similarity measures (for example sum of squared difference or sum of absolute difference). Color histogram represents the image pixels distribution. The histogram bins number depends on the pixel depth such that for image with pixels' depth of n, the histogram bins number be 2^n with color range from 0 to 2^n-1 [34].

In order to verify the proposed algorithm accuracy, two of the most commonly used measures are used which are precision and recall. Where precision is the retrieved fraction of relevant from the total retrieved items. Recall is defined as the retrieved relevant fraction from the total relevant. Where 'precision' and 'recall' are given by:

$$Precision = \frac{TP}{TP+FP} \tag{11}$$

$$Recall = \frac{TP}{TP+FN} \tag{12}$$

where, TP, FP, and FN are the true positive, false positive, and false negative respectively. Table 2 shows these two parameters for the proposed algorithm as compared to some related ones. From this table, it is obvious that our proposed approach hasn't enough precise retrieval as we hope, but it has the ability to retrieve the most relevant items efficiently.

Table 2. Detection results evaluation indicators values of foreground detection methods

Detection Method	Precision	Recall
Proposed	≈0.70	≈0.9
[17]	≈0.77	≈0.7
[18]	≈0.84	≈0.65
[19]	≈0.72	≈0.44

10. CONCLUSIONS

A fast as grayscale processing and a rich of information as a full colored image channels utilization, the proposed method in this paper gets these privileges by utilizing the Bayer pattern for converting any color image into just one layer just like a grayscale with preservation of the color information for each of the three channels (RGB). Thus a reduced processing time as it is required to handle just single layer and a well matching information as it is preserved the three colored channels information into this single layer. Further data reduction and hence reduced time consumption is achieved by utilizing the most important principle components resulted from the SVD and ignoring the other less important components. The method also treated the undesired camera motion using the SURF technique in order to get the corresponding key points of two consecutive frames for the purpose of registering one frame in according to the other in order to know the transformation that happened between them as a consequence of this undesired motion. A reverse transformation for the derived one has to be conducted in an aim to compensate the effect of this motion. The compensation leads to precise moving objects detection when conducting frames difference, which is a situation that couldn't happen without such compensation.

ACKNOWLEDGEMENTS

It is our pleasure to express our appreciation and thanks for Computer Science Department, College of Science, University of Mustansiriyah, Baghdad, Iraq and Department of Computer techniques Engineering, Imam Kadhim Faculty of University Islamic Sciences for the valuable assistance and encouragement to accomplish this research.

REFERENCES

[1] B. Kang, W. P. Zhu, and J. Yan, "Object detection oriented video reconstruction using compressed sensing," *EURASIP Journal on Advances in Signal Processing,* vol. 2015, no. 1, 2015, doi: 10.1186/s13634-015-0194-1.
[2] J. H. Awad and B. D. Majeed, "Moving objects detection based on frequency domain," *Baghdad Science Journal*, vol. 17, no. 2, 2020, doi: 10.21123/bsj.2020.17.2.0556.
[3] L. Lin, W. Lin, and S. Huang, "Group object detection and tracking by combining RPCA and fractal analysis," *Soft Computing,* vol. 22, no. 1, pp. 231-242, 2016, doi: 10.1007/s00500-016-2329-1.
[4] Y. Zhou and B. W. K. Ling, "Detecting moving objects via the low-rank representation," *Signal, Image and Video Processing,* vol. 13, no. 8, pp. 1593-1601, 2019, doi: 10.1007/s11760-019-01503-7.
[5] O. M. Sincan, V. B. Ajabshir, H. Y. Keles, and S. Tosun, "Moving object detection by a mounted moving camera," presented at the IEEE EUROCON 2015 - *International Conference on Computer as a Tool (EUROCON),* Salamanca, Spain, 02 November 2015. [Online]. Available: https://ieeexplore.ieee.org/document/7313714?reload=true.
[6] Y. Yu, L. Kurnianggoro, and K. H. Jo, "Moving Object detection for a moving camera based on global motion compensation and adaptive background model," *International Journal of Control, Automation and Systems,* vol. 17, no. 7, pp. 1866-1874, 2019, doi: 10.1007/s12555-018-0234-3.
[7] M. Yazdi and T. Bouwmans, "New trends on moving object detection in video images captured by a moving camera: A survey," *Computer Science Review,* vol. 28, pp. 157-177, 6 Mar 2018, doi: 10.1016/j.cosrev.2018.03.001.
[8] W. Zhang, X. Sun, and Q. Yu, "Moving object detection under a moving camera via background orientation reconstruction," *Sensors (Basel),* vol. 20, no. 11, May 2020, doi: 10.3390/s20113103.
[9] A. Pareeka and N. Arorab, "Re-projected SURF features based mean-shift algorithm for visual tracking," in *International Conference on Computational Intelligence and Data Science (ICCIDS)*, 2019, pp. 1553-1560.
[10] R. Oji, "An automatic algorithm for object recognition and detection based on asift keypoints," *Signal & Image Processing : An International Journal,* vol. 3, no. 5, pp. 29-39, 2012, doi: 10.5121/sipij.2012.3503.
[11] W. He, T. Yamashita, H. Lu, and S. Lao, "SURF Tracking," in *12th International Conference on Computer Vision (ICCV)*, 2009, IEEE, pp. 1586-1592.
[12] J. Ju and J. Xing, "Moving object detection based on smoothing three frame difference method fused with RPCA," *Multimedia Tools and Applications,* vol. 78, no. 21, pp. 29937-29951, 2018, doi: 10.1007/s11042-018-6710-1.
[13] C. Kumar, A. K. Singh, and P. Kumar, "Dual watermarking: An approach for securing digital documents," *Multimedia Tools and Applications,* vol. 79, no. 11-12, pp. 7339-7354, 2019, doi: 10.1007/s11042-019-08314-5.
[14] J.-Y. Wu, W.-L. Huang, W.-M. Xia-Hou, W.-P. Zou, and L.-H. Gong, "Imperceptible digital watermarking scheme combining 4-level discrete wavelet transform with singular value decomposition," *Multimedia Tools and Applications,* vol. 79, no. 31-32, pp. 22727-22747, 2020, doi: 10.1007/s11042-020-08987-3.
[15] K. Kalirajan and M. Sudha, "Moving object detection for video surveillance," *ScientificWorldJournal,* vol. 2015, p. 907469, 2015, doi: 10.1155/2015/907469.
[16] Y. Chen, R. Zhang, and L. Shang, "A novel method of object detection from a moving camera based on image matching and frame coupling," *PLoS One,* vol. 9, no. 10, p. e109809, 2014, doi: 10.1371/journal.pone.0109809.
[17] E. Dong, B. Han, H. Jian, J. Tong, and Z. Wang, "Moving target detection based on improved Gaussian mixture model considering camera motion," *Multimedia Tools and Applications,* vol. 79, no. 11-12, pp. 7005-7020, 2019, doi: 10.1007/s11042-019-08534-9.
[18] S. Joshi, S. Gujarathi, and A. Mirge, "moving object tracking method using improved camshift with surf algorithm," *International Journal of Advances in Science Engineering and Technology,* vol. 2, no. 2, pp. 14-18, April 2014. [Online]. Available: http://www.iraj.in/journal/journal_file/journal_pdf/6-46-139763132914-18.pdf.
[19] M. M. Jlassi, A. Douik, and H. Messaoud, "Objects detection by singular value decomposition technique in hybrid color space: application to football images," *International Journal of Computers Communications & Control,* vol. 5, no. 2, 2010, doi: 10.15837/ijccc.2010.2.2474.
[20] S. E. Ebadi, "Approximated robust principal component analysis for improved general scene background subtraction," *arXiv*, 23 May 2016.
[21] P. R. Karthikeyan, P. Sakthivel, and T. S. Karthik, "Comparative study of illumination-invariant foreground detection," *The Journal of Supercomputing,* vol. 76, no. 4, pp. 2289-2301, 2018, doi: 10.1007/s11227-018-2488-1.
[22] W. Kim, "Background subtraction with variable illumination in outdoor scenes," *Multimedia Tools and Applications,* vol. 77, no. 15, pp. 19439-19454, 2017, doi: 10.1007/s11042-017-5410-6.
[23] K. Lal and K. M. Arif, "Feature extraction for moving object detection in a non-stationary background," presented at the *IEEE/ASME International Conference on Mechatronic and Embedded Systems and Applications (MESA)*, 2016.
[24] W. Witwit, Y. Zhao, K. Jenkins, and S. Addepalli, "Global motion based video super-resolution reconstruction using discrete wavelet transform," *Multimedia Tools and Applications,* vol. 77, no. 20, pp. 27641-27660, 2018, doi: 10.1007/s11042-018-5941-5.
[25] S. Dhivya, J. Sangeetha, and B. Sudhakar, "Copy-move forgery detection using SURF feature extraction and SVM supervised learning technique," *Soft Computing,* vol. 24, no. 19, pp. 14429-14440, 2020, doi: 10.1007/s00500-020-04795-x.

[26] J. Pan, W. Chen, and W. Peng, "A new moving objects detection method baased on improved SURF algorithm," in *25th Chinese Control and Decision Conference (CCDC), 2013*, 25-27 May 2013, pp. 901-906, doi: 10.1109/CCDC.2013.6561051.
[27] M. L. D. L. Calleja, T. Nagai, M. Attamimi, M. N. Miyatake, and H. P. Meana, "Object detection using SURF and superpixels," *Journal of Software Engineering and Applications,* vol. 06, no. 09, pp. 511-518, 2013, doi: 10.4236/jsea.2013.69061.
[28] R. Lukac, K. N. Plataniotis, and D. Hatzinakos, "Color image zooming on the Bayer pattern," *IEEE Transactions on Circuits and Systems for Video Technology,* vol. 15, no. 11, pp. 1475-1492, 2005, doi: 10.1109/tcsvt.2005.856923.
[29] A. Lukin and D. Kubasov, "High-quality algorithm for bayer pattern interpolation," *Programming and Computer Software,* vol. 30, no. 6, pp. 347-358, 2004, doi: 10.1023/B:PACS.0000049512.71861.eb.
[30] H. S. Malvar, H. Li-wei, and R. Cutler, "High-quality linear interpolation for demosaicing of Bayer-patterned color images," presented at the 2004 *IEEE International Conference on Acoustics*, Speech, and Signal Processing, 2004.
[31] M. Russell, J. J. Zou, and G. Fang, "An evaluation of moving shadow detection techniques," *Computational Visual Media,* vol. 2, no. 3, pp. 195-217, 2016, doi: 10.1007/s41095-016-0058-0.
[32] T. K. Araghi, A. A. Manaf, and S. K. Araghi, "A secure blind discrete wavelet transform based watermarking scheme using two-level singular value decomposition," *Expert Systems with Applications,* vol. 112, pp. 208-228, 2018, doi: 10.1016/j.eswa.2018.06.024.
[33] P. Zheng and Y. Zhang, "A robust image watermarking scheme in hybrid transform domains resisting to rotation attacks," *Multimedia Tools and Applications,* vol. 79, no. 25-26, pp. 18343-18365, 2020, doi: 10.1007/s11042-019-08490-4.
[34] S. R. Kodituwakku and S. Selvarajah, "Comparison of color features for image retrieval," *Indian Journal of Computer Science and Engineering,* vol. 1, no. 3, pp. 207-211, October 2010.

BIOGRAPHIES OF AUTHORS

Jalal H. Awad was graduated from al Rafidain University College and Had B.Sc. degree in the Department of Information System Ranked 1 among the graduated students, he had the M.Sc. degree in Computer Science, College of Science, Baghdad University, his Ph.D. granted from Babylon University in Computer Science, College of Science. He can be contacted at email: jalalhameed@uomustansiriyah.edu.iq.

Balsam D. Majeed was graduated from Mustansiriyah University and Had B.Sc. degree in the Department of Computer Science, College of Science, she had the M.Sc. degree in Computer Science, College of Science, Amirkabir University of Technology, and she is studying for the Ph.D. degree in Computer Engineering at Ferdowsi University of Mashhad. She can be contacted at email: balsamdhyia83@gmail.com.