[Downloaded from www.aece.ro on Thursday, July 03, 2025 at 18:58:40 (UTC) by 172.69.130.199. Redistribution subject to AECE license or copyright.]

Digital Video Stabilization Verification Based on Genetic Algorithm Template Matching

Milos PAVLOVIC[1,2], Zoran BANJAC[2], Branko KOVACEVIC[1]
[1]*School of Electrical Engineering, University of Belgrade, 11000, Belgrade, Serbia*
[2]*Vlatacom Institute, Blvd. Milutina Milankovica 5, 11070, Belgrade, Serbia*
milos.pavlovic@vlatacom.com

***Abstract*—Having high precision ground-truth data is a very important factor for the development and evaluation of computer vision algorithms such as digital video stabilization. However, generating this data is time consuming and cost intensive work, requiring a lot of manual effort. In this paper we both propose a way to automatically generate a large amount of accurate data for digital video stabilization verification and provide a comprehensive dataset of video sequences taken from multi-sensor imaging system with different types of disturbances. A novel method for generating verification data is based on genetic algorithm template matching. Paper provides quantitative analysis together with the visual assessment of digital video stabilization performance.**

***Index Terms*—formal verification, genetic algorithms, image motion analysis, multispectral imaging, pattern matching.**

I. INTRODUCTION

In the field of video surveillance, multi-sensor systems are very popular and used in various applications, where in addition to the most commonly used cameras that work in the visible-light range, thermal cameras are increasingly used. In addition to high-resolution cameras, modern video surveillance solutions have an increasing need for complex video processing algorithms, such as video stabilization, image quality enhancement, and video tracking. Hand-held cameras, cameras mounted on vehicles or ships, as well as those on pan-tilt platforms exposed to windy conditions, give a video stream with shakings due to various unwanted movements. In such cases, video stabilization is an inevitable component that allows uninterrupted observation of the recorded scene. One solution to this problem is hardware solution. It involves the use of a motion sensor (e.g. a gyroscope sensor) [1-4] and based on the data from such a sensor, the optical system compensates the effect of unwanted camera motion. This solution is very efficient, but significantly increases the cost of the system. Another solution is software-based video signal processing or digital video stabilization [5-9].

In digital video stabilization algorithms, the basic tasks are estimation the unwanted motion (shaking) of the camera and compensation of that motion. The solution must be robust, so that the conditions of different illumination, geometric transformation, image blurring would not affect the accuracy of the assessment of the camera's motion. The result of poor evaluation of camera motion parameters is the retention of unwanted motion in stabilized video sequences and reduced stabilization performance [10].

In addition to selection of a video stabilization algorithm, the performance evaluation of the algorithm is important for adjusting and comparing different methods. A lot of methods have been proposed for video stabilization verification [11-13]. However, it is a matter of objectively assessing the performance of the video sequence stabilization. In the case of a lack of reference video sequence for comparing stabilization performance, one of the very efficient methods for estimating motion within video frames can be the method based on template matching [14].

Generally, template matching is a method for identifying features in a source image that match the image patch called the template image [14]. In template matching approach for verification of video stabilization, the position of the defined image patch (template) on the first frame of the video sequence in the following frames of the video sequence is determined by a pixel-wise comparison of the image with a given template that contains the desired pattern. In the basic approach, the template image defined on the first video sequence frame slides over the source frame of video sequence and at each position the degree of similarity between the template and source image over the template area is estimated calculating a grey-scale correlation measure using pixel intensities.

To calculate the degree of similarity, the normalized cross correlation [15-16] is used in template matching. Although a good technique for template matching, for exhaustively search in whole image (similarity calculation is performed at every position in source image to find the best template matches), normalized cross correlation can be very time consuming and requires a lot of computing resources since it is based on summation and multiplication operations. Therefore, fast algorithm to computing normalized cross-correlation from the spatial domain using the precomputed convolutional integrals over the search image window has been proposed [17-18].

The position of the template image defined in the first frame, in the unstabilized video sequence due to disturbances can be located at each pixel in the source frame. If the nature of disturbance is known in advance, then local searches can be performed around the position where the template was defined in the first frame. Local search will lead to a faster search but will be more susceptible to errors in the case of higher intensity disturbance.

To overcome problems of local search and speed up the process of exhaustive search in the whole image, the paper presents a method of objective assessment of image movement in a video sequence based on a genetic algorithm template matching [19-20]. The paper defines the appropriate fitness function for optimization in genetic algorithm, and in that way, genetic algorithm template

Digital Object Identifier 10.4316/AECE.2022.02007

1582-7445 © 2022 AECE

[Downloaded from www.aece.ro on Thursday, July 03, 2025 at 18:58:40 (UTC) by 172.69.130.199. Redistribution subject to AECE license or copyright.]

matching performs the evaluation of digital video stabilization performance. The method is applied to the video sequences in which the object being recorded is static.

The paper is organized as follows. Section II presents the general structure of genetic algorithms and proposes the method for search in an image based on genetic algorithm template matching. Section III describes the algorithm for digital video stabilization. Section IV gives a description of the experimental work, a description of the created video database using the multi-sensor imaging system, details of the implemented genetic algorithm and digital video stabilization algorithm. The results of the implemented genetic algorithm template matching for verifying the operation of the digital video stabilization algorithm are presented in Section V. Section VI lists conclusions and indicates directions for future work in this research area.

II. Genetic Algorithm Template Matching

In template matching approach, an image patch (template) can be defined in the first frame of the video sequence, and then its position in the following frames of the video sequence can be tracked and compared with the position in the first frame (Fig. 1). The template is defined in the region of the image with many edges and as discriminatory features as possible compared to the rest of the image.

Figure 1. Selected image template to be tracked in video sequence frames

The normalized cross correlation between grayscale image I and template I_t used in template matching algorithms is given by:

$$r(u,v)=\frac{\sum_{x,y}(I(x,y)-\overline{I(u,v)})(I_t(x-u,y-v)-\overline{I}_t)}{\sqrt{\sum_{x,y}(I(x,y)-\overline{I(u,v)})^2\sum_{x,y}(I_t(x-u,y-v)-\overline{I}_t)^2}} \quad (1)$$

$\overline{I(u,v)}$ and $\overline{I}_t$ denote the mean value of image I in the region coincident with the template image and mean value of the template image, respectively.

The maximum value of $r(u,v)$ indicates a position where the template $\overline{I}_t$ best matches the image I.

For an image of dimensions M x N and an image template of dimensions p x r (p < M and r < N), it is necessary to perform (M-p) x (N-r) calculations of the correlation between the template image and other parts of the image. For Full HD images (1920x1080 pixels) the process of correlation calculations can take a very long time. In order to accelerate this process by achieving the same results, a genetic algorithm is applied in this paper.

The genetic algorithm belongs to the class of evolutionary optimization algorithms. The algorithm simulates the process of natural selection, using techniques such as inheritance, mutation, selection and crossover [21-24]. At the beginning, genetic algorithm creates a population of individuals (solutions for the defined problem) randomly or by applying some heuristics, and for each individual calculates a cost or fitness value. The code of each individual represents the set of its genes. The coding of an individual depends on the nature of the problem. It can be binary, integer, real or represented by some other set of symbols. Following a natural selection techniques, based on the fitness values, a subset of the population is selected as the parents. A new generation of individuals is obtained by crossing these parents, with mutations in some genes occurring from time to time in some individuals. The new generation is used to replace members of the original population. Also, a new cost value is calculated for each individual in new generation. This process of choosing and using the fittest individuals in the population to produce new generations is then repeated. It is expected that the change of generation will result in a population of individuals that yield better approximations to an optimal solution.

In order to apply the genetic algorithm to the defined problem, the position of the center of the defined template image I_t is encoded by a chromosome of length L. The center of the template I_t of dimension p x r in image I of dimension M x N can take a value in the range [p/2, M-p/2] x [r/2, N-r/2]. The initial population is created with a uniform distribution in a defined space.

A. Fitness Function

The fitness function f for finding the position of the selected template in the image is based on the normalized correlation coefficient $corr$ between I_t and a part of the image I covered by the template – I_p:

$$f = corr(I_p I_t) \quad (2)$$

As the defined fitness function (2) takes value from the range [-1,1] due to the properties of the correlation coefficient, while genetic algorithms are mainly developed to maximize the positive fitness function, the given criterion function (2) needs to be modified.

As the problem that occurs, in this case, is a negative value of the fitness function, the modification is performed as follows:

$$g=\begin{cases} f-C_{\min}, & f>C_{\min} \\ 0, & f\le C_{\min} \end{cases} \quad (3)$$

The coefficient C_{min} represents the minimum value of the original fitness function, and with the value of -1, the modified fitness function has the following form:

$$f = corr\left(I_p, I_t\right)+1 \quad (4)$$

In order to avoid the problem that the local maximum is much higher than the mean value of the fitness function, and thus makes it more difficult to find the global maximum, scaling the fitness function is another modification that has been made [25].

Since, in the selection process, the probability of selecting an individual is directly proportional to the value of its fitness function, it follows that individuals with higher fitness have a more significant advantage in being selected for the reproduction. However, due to the dynamics of the fitness function, too much preference may be given to fitter individuals (as measured by a fitness function), and worse ones are given almost no chance. If there is an individual in the population with a value of the fitness function significantly higher than the rest of the population, due to its

[Downloaded from www.aece.ro on Thursday, July 03, 2025 at 18:58:40 (UTC) by 172.69.130.199. Redistribution subject to AECE license or copyright.]

dominance, there may be a complete loss of the ability to search for other, potentially good parts of space. For this reason, it is essential to allow passing into the next generation even individuals with low fitness function, which may evolve into significantly better individuals in other parts of the search space and thus discover a better local maximum.

A new fitness function g is defined as:

$$g = af + b \tag{5}$$

where, f is defined in (4), with conditions:

$$\overline{g} = \overline{f}, \quad g_{\max} = C\overline{f} \tag{6}$$

Here $\overline{f}$ and $\overline{g}$ represent the mean values of the functions f and g, respectively, g_{max} is the maximum value of function g, and C is a constant that can control the expected number of the best individuals selected to participate in the next generation.

The values of the coefficients a and b are obtained from (5) with the stated conditions (6) and amounts:

$$a = \frac{\overline{f}(1-C)}{\overline{f} - f_{\max}}, \quad b = \frac{\overline{f}(C\overline{f} - f_{\max})}{\overline{f} - f_{\max}} \tag{7}$$

where, f_{max} is the maximum value of the fitness function (4).

The constant C is taken from the range (1.5, 2), and for the problem in this paper, experimentally, C = 1.7 was adopted. As the maximum value of the fitness function after the first modification in (4) is f_{max} = 2 and the mean value is $\overline{f}$ = 1, the obtained values for the coefficients a and b are 0.7 and 0.3, respectively. The fitness function after scaling has the form:

$$f = 0.7 \cdot (corr(I_p, I_t) + 1) + 0.3 \tag{8}$$

B. Selection and Crossover

For a defined difference between successive generations (generation gap) G ∈ (0.8, 1), from the previous generation to the next is directly passed [(1-G) x N] individuals (N is the total number of individuals in population), while the rest is taken for crossover. It is also important that the number of individuals taken for crossover is even.

In order to ensure that the quality of the next generation is at least as good as the previous one, an elitist technique is applied. Applying elitism increases selective pressure and thus reduces the genetic drift. Elitism is implemented in that one individual with the best fitness function is transferred from each generation. Other individuals are passed using roulette wheels selection [26].

In this way, the value of the fitness function directly influences the selection of an individual for reproduction in the roulette wheel method, where greater preference is given to individuals with a higher fitness function, but a chance is also given to those individuals with a lower value of the fitness function.

The selection of individuals for crossover is also performed using the roulette wheel method. Crossover is performed with probability from the range (0.6, 0.9). The method of crossover with one intersection point is applied [27].

C. Mutation

The mutation was performed on [N x L x p_m] bits of the whole generation of individuals, where p_m is the mutation probability from the range [0.1%, 0.4%]. An inverse mutation is applied. The mutation is implemented by first selecting one of the N individuals in the population at random and then randomly selecting the position of the bit to be inverted in it. The procedure is repeated until the [N x L x *pm*] bits are inverted.

D. Termination

As a criterion for termination of the algorithm, it is implemented that the algorithm is terminated in case one of the two conditions is met. The first is related to the value of the fitness function of the best individual in the population. The algorithm terminates if the value of the fitness function of the best individual is greater than 90% of the maximum value of the fitness function, f, (8). It has been empirically determined that if this condition is met, the position of the template is found in the image with high accuracy. The second condition is related to the maximum number of iterations of the genetic algorithm. If the number of iterations of the genetic algorithm reaches the maximum number of 50 iterations determined by the simulations, the algorithm is terminated, and the position of the individual with the best fitness function at that time is taken as the found template position.

III. Digital Video Stabilization

The goal of digital video stabilization is to improve video quality by removing the effects of the unwanted camera movements while preserving the motion of moving objects in the video sequence. Video stabilization is a key step in image preprocessing before performing higher-level video processing algorithms such as object tracking. For example, the accuracy of object trajectory prediction can be significantly reduced in the case of unstabilized video sequences.

Fig. 2 shows a block diagram of the adopted video stabilization algorithm. The input frame comes from a visible-light or thermal camera, and then is converted to a grayscale format to improve the efficiency of the algorithm without compromising the quality of the results.

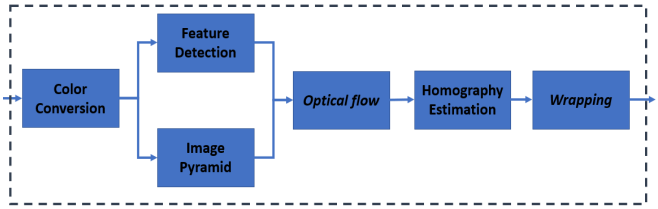


Figure 2. Digital video stabilization block diagram.

Motion estimation and unwanted motion filtering present two main parts of the video stabilization algorithms.

As this task can be very difficult, in some cases even impossible, there must be some simplifications. The 2D transformation between pixels of consecutive frames of a video sequence is estimated. It is very important that these methods are robust enough to deal with situations when moving objects are in the foreground in the scene. This is achieved by using the Harris features [28] and the Optical flow pyramid Lucas Kanade algorithm [29] with RANSAC (Random Sample Consensus) [30] for outliers detection. The output of this part of the algorithm is a set of

[Downloaded from www.aece.ro on Thursday, July 03, 2025 at 18:58:40 (UTC) by 172.69.130.199. Redistribution subject to AECE license or copyright.]

transformations that describe the effect of unwanted camera movement during the video sequence.

Compensation of unwanted camera motion effects is then performed in the next step by applying a homography matrix - a global transformation matrix that contains the motion parameters estimated in the previous step. The matrix of affine transformations is implemented in the paper [31]. When there is intentional motion in the image, as well as the unwanted motion that needs to be compensated, Kalman filtering [32] provides a good solution. Instead of modeling the motion of the camera, a cumulative transformation of image parameters is used with the affine homography model [32]. The unwanted camera motion is the difference between global motion and estimated wanted motion parameters with Kalman filter. The state-space model is given in [32], where some parameters in matrix which describes zoom, rotation and dolly motion of the camera, as well as in vector which describes panning and tracking motion are modeled as motion with constant velocity, while others have the simple dynamic model.

IV. Experimental Work and Results

Before applying the proposed method for verification of the video stabilization algorithm, the performance of the verification method presented in this paper needs to be evaluated. A database of manually labeled thermal and visible-light video sequences with shaking was created to determine the parameters of the genetic algorithm and evaluate performance on ground-truth data. The results are presented by objective metrics as well as graphically.

A. Database

The database made for this paper consists of 10 video sequences - 5 recorded with a visible-light camera (FULL HD resolution: 1920x1080 pixels) and 5 with a thermal camera (Medium Wave InfraRed (MWIR): 8–14μm wavelength, resolution: 640x480 pixels) using Vlatacom electro-optical long-range surveillance system [33]. The first frame of each video sequence is shown in Table I.

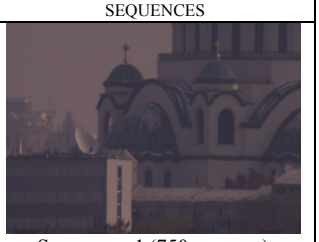
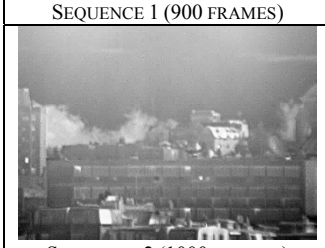
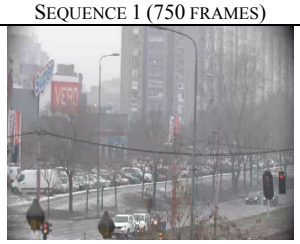
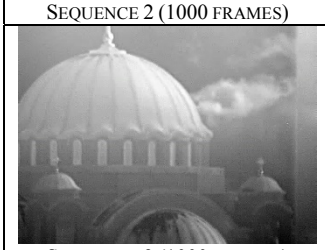
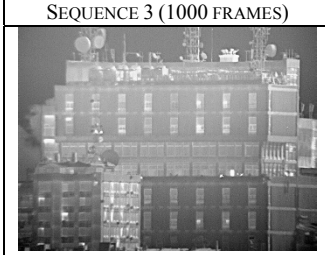
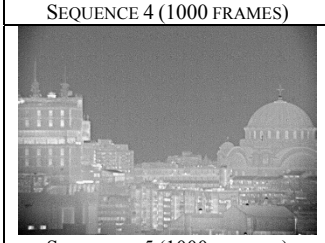
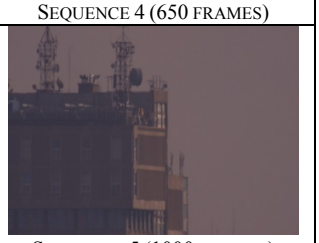
Each video sequence is recorded with a fixed Field of View (FOV) and pan-tilt position in azimuth and elevation, which are set at the beginning of the sequence recording. During recording, the system is placed on a shaking platform to simulate the shaking conditions (most commonly caused by wind) that occur on the platforms where the system is mounted in real-life applications.

In the following, recording details of thermal and visible-light video sequences are presented. FOV is horizontal FOV in degrees, and distance is measured to the object in the center of the image using a laser range finder. Thermal sequences are recorded with the following parameters: Sequence 1: FOV - 1,014, distance – 17000 m; Sequence 2: FOV - 0.669, distance – 4040 m; Sequence 3: FOV - 0.942, distance – 4590 m; Sequence 4: FOV - 0.857, distance – 4165 m; Sequence 5: FOV - 3,024, distance – 4330 m. Visible-light sequences are recorded with the following parameters: Sequence 1: FOV: 1.016, distance – 4590 m, Sequence 2: FOV - 7.130, distance – 310 m; Sequence 3: FOV - 12,671, distance – 170 m; Sequence 4: FOV - 0.904, distance – 4040 m; Sequence 5: FOV - 0.730, distance – 4160 m.

The created dataset contains video sequences that can be found in real-life applications, from a situation with blur in the image (thermal sequence 1), through situations with a large clutter (thermal sequence 2 and visible-light sequences 2 and 3), to the case where features are clearly expressed and the image quality is very good (thermal sequences 3 and 4 and visible-light sequence 1) and situations with a large presence of a uniform background in the image (thermal sequences 5 and visible-light sequences 4 and 5).

For each video sequence in the database, all frames of the video sequence are manually labeled by marking the point on the first frame and then marking the same point on the same object in the following frames in the video sequence to estimate the motion in the video sequence (camera shaking) caused by different disturbances. This data is further used to select the parameters of the genetic algorithm and to validate the implemented verification algorithm.

Table I. Video Sequences Database

Thermal video sequences	Visible-light video sequences
 Sequence 1 (900 frames)	 Sequence 1 (750 frames)
 Sequence 2 (1000 frames)	 Sequence 2 (800 frames)
 Sequence 3 (1000 frames)	 Sequence 3 (630 frames)
 Sequence 4 (1000 frames)	 Sequence 4 (650 frames)
 Sequence 5 (1000 frames)	 Sequence 5 (1000 frames)

B. Genetic Algorithm Implementation

One video sequence from the thermal and the one

[Downloaded from www.aece.ro on Thursday, July 03, 2025 at 18:58:40 (UTC) by 172.69.130.199. Redistribution subject to AECE license or copyright.]

sequence from the visible-light camera were used for the determination of the genetic algorithm parameters. Sequence 5 was used from the thermal camera, and sequence 1 was used from the visible-light camera due to good image quality and the possibility of selecting the image template which is significantly different in relation to the rest of the image, thus reducing the impact of the image quality and image content on the algorithm performance. The following parameters of the genetic algorithm were adopted by experiments on these two video sequences:

- Number of individuals in the population: $N = 120$
- Chromosome length: $L = 24$
- Generation gap: $G = 0.8$
- Crossover probability: $p_c = 0.9$
- Mutation probability: $p_m = 0.005$

The size of the selected image patch is 150x150 pixels, so the search area is 1770x930 in the visible-light image, while in the thermal image, it is 490x330.

The number of iterations for the execution of the algorithm should represent a balance between performance and algorithm processing time. Experimentally, the number of iterations was set to 50, while for the earlier termination, the condition is that the best individual in the population has a value of the fitness function greater than 90% of the maximum fitness function value.

With the adopted parameters, the algorithm is applied to other video sequences to show performance. The algorithm is implemented in the MATLAB® software package.

C. Results and Discussion

For the selected video sequences for determination of the genetic algorithm parameters, a comparison of the ground truth displacement - manually labeled and estimated using a genetic algorithm template matching is provided. The graphics in Fig. 3 and Fig. 4 show the accuracy of finding the segmented template in the frames of the visible-light video sequence 1. Accuracy is expressed as the percentage of video sequence frames in which the selected template is found in the horizontal or vertical direction with the error lower than the given threshold in pixels. It can be seen that the accuracy with 0 pixels error is up to 30% in the vertical direction (in 30% of the frames of the entire video sequence, in the vertical direction, the template was found with 0 pixels error), and with an error of up to 10 pixels the accuracy is maximum, which means there was no frame in which the selected template in the vertical direction was found with an error greater than 10 pixels. In the horizontal direction, the accuracy with 0 pixels error is up to 18% and the maximum accuracy is reached with an error up to 11 pixels. As the image resolution is 1920x1080 pixels, the biggest error of 10 pixels is 0.92% of the image height and 0.57% of the image width with the biggest error of 11 pixels in the horizontal direction.

The graphs in Fig 5 and Fig. 6 show the accuracy of finding the selected image template in the frames of the thermal video sequence 5. It can be seen that for the image resolution of 640x480 pixels, maximum accuracy in the horizontal direction is reached with an error up to 8 pixels (1.25% of image width) and with 0 pixels error accuracy is up to 10%. In vertical direction maximum accuracy is reached with the error up to 9 pixels (1.87% of image height), while with 0 pixels error, accuracy is up to 20%.

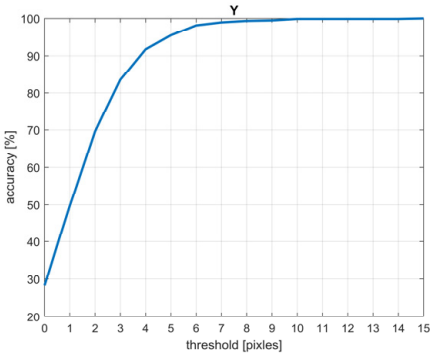


Figure 3. The accuracy of finding a segmented template in visible-light video sequence1 in the vertical direction

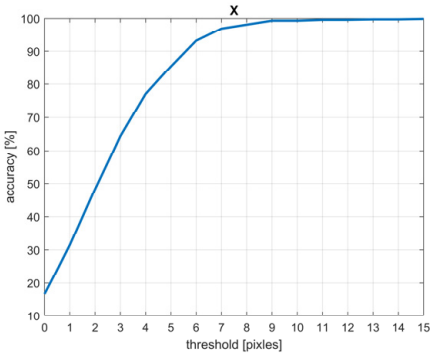


Figure 4. The accuracy of finding a segmented template in visible-light video sequence 1 in the horizontal direction

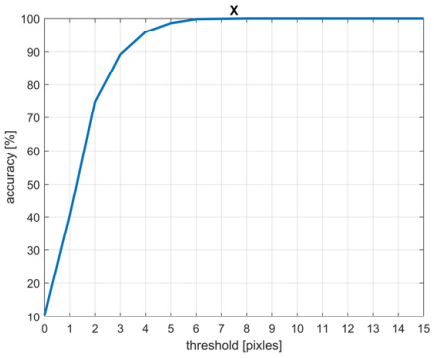


Figure 5. The accuracy of finding a segmented template in thermal video sequence 5 in the horizontal direction

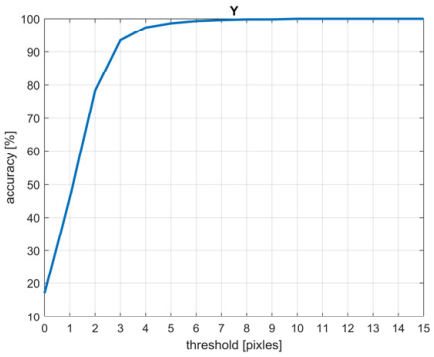


Figure 6. The accuracy of finding a segmented template in thermal video sequence 5 in the vertical direction

Table II and Table III show the performance comparison of the genetic algorithm template matching based image search and exhaustive image search based on the normalized cross correlation coefficient in relation to maximum error (maximum displacement in pixels from the labeled position in the entire video sequence), mean error and standard

[Downloaded from www.aece.ro on Thursday, July 03, 2025 at 18:58:40 (UTC) by 172.69.130.199. Redistribution subject to AECE license or copyright.]

deviation, in the horizontal and vertical direction respectively, for all video sequences in the database.

TABLE II. MAXIMUM (XMAX), MEAN (XMEAN) AND STANDARD DEVIATION (XSTD) ERROR IN VIDEO SEQUENCES IN HORIZONTAL DIRECTION IN PIXELS

Video sequence	**Genetic algorithm template matching**			**Exhaustive search**		
	Xmax	Xmean	Xstd	Xmax	Xmean	Xstd
Visible 1	11	3.41	2.36	9	3.24	2.01
Visible 2	6	1.74	1.35	6	1.29	0.93
Visible 3	13	2.59	2.40	6	1.60	1.29
Visible 4	13	3.23	2.56	11	2.01	1.58
Visible 5	10	2.47	1.80	7	2.15	1.51
Thermal 1	14	3.66	2.47	11	3.55	2.34
Thermal 2	7	2.68	1.42	6	2.53	1.25
Thermal 3	11	1.69	1.51	4	1.32	0.98
Thermal 4	5	1.19	1.01	6	1.12	0.96
Thermal 5	8	1.90	1.27	7	1.95	1.25

TABLE III. MAXIMUM (YMAX), MEAN (YMEAN) AND STANDARD DEVIATION (YSTD) ERROR IN VIDEO SEQUENCES IN VERTICAL DIRECTION IN PIXELS

Video sequence	Genetic algorithm template matching			Exhaustive search		
	Ymax	Ymean	Ystd	Ymax	Ymean	Ystd
Visible 1	10	2.34	1.86	9	1.96	1.43
Visible 2	9	1.67	1.48	5	1.17	0.92
Visible 3	9	1.58	1.31	6	1.25	0.96
Visible 4	9	1.97	1.54	6	1.56	1.17
Visible 5	8	1.96	1.59	7	1.82	1.35
Thermal 1	14	2.62	2.01	10	2.43	1.96
Thermal 2	5	0.80	0.78	3	0.67	0.66
Thermal 3	11	2.73	1.90	5	2.04	1.13
Thermal 4	5	1.25	0.97	6	1.28	0.98
Thermal 5	9	1.82	1.33	9	1.54	1.02

Accurate estimation along the vertical direction (Y-axis) is especially important because the disturbance occurs especially in that direction due to the nature of the pan-tilt positioner on which the camera system is mounted. Accuracy comparison of the genetic algorithm template matching based image search and exhaustive image search based on the normalized cross correlation coefficient is shown in Fig. 7 for thermal video sequence 1, as an example of the significant presence of blur in the image. In Fig. 8 and Fig. 9, accuracy comparison is presented for situations with large clutter in thermal sequence 2 and visible-light sequence 3, respectively. In Fig. 10 accuracy comparison is shown for visible-light video sequence 5 in which there is a uniform background. Here, accuracy is also expressed as the percentage of video sequence frames in which the selected template is found in the vertical direction with the error lower than the given threshold in pixels. Results presented in Table II and III and graphics in Fig. 7 - 10 show that the accuracy in the motion estimation in a video sequence of the genetic algorithm template matching approach reaches the accuracy of the exhaustive search based on the correlation coefficient. On the other hand, for the visible-light image resolution of 1920x1080, thermal image resolution of 640x480 pixels, and template size of 150x150 pixels, the number of correlation calculations is 1646100 and 161700, while for the genetic algorithm template matching with 120 individuals in the population in the worst case of the 50 iterations, the number of correlation calculations is 6000. In that way, the number of correlation calculations is reduced for 99.635% for visible-light image, and 96.289% for thermal image. These results show the great advantage of the proposed genetic algorithm template matching approach for motion estimation and data generation for digital video stabilization verification.

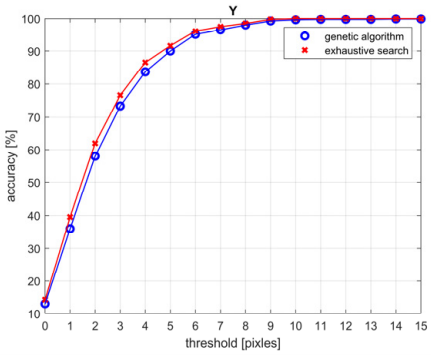


Figure 7. Accuracy comparison of finding a segmented template in the vertical direction in thermal video sequence 1 with blur

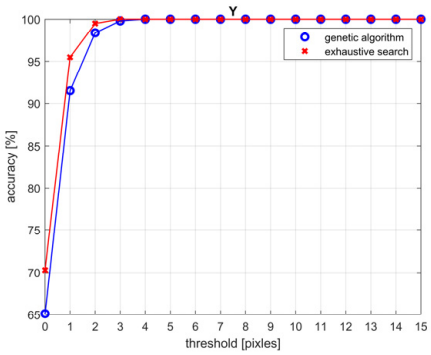


Figure 8. Accuracy comparison of finding a segmented template in the vertical direction in thermal video sequence 2 with clutter

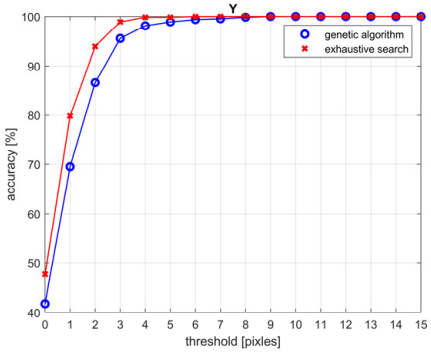


Figure 9. Accuracy comparison of finding a segmented template in the vertical direction in visible-light video sequence 3 with clutter

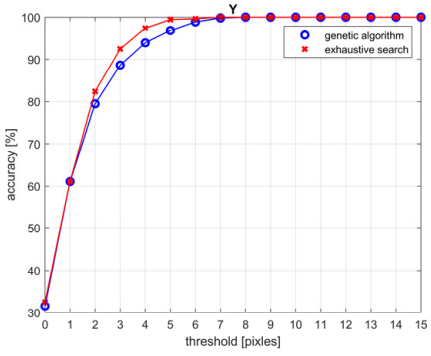


Figure 10 Accuracy comparison of finding a segmented template in vertical direction in visible-light video sequence 5 with large uniform background

V. DIGITAL VIDEO STABILIZATION VERIFICATION

The video stabilization algorithm described in section III is applied to the video sequences from the database with the disturbance. The application of this algorithm very successfully eliminates the disturbance, which can be clearly

[Downloaded from www.aece.ro on Thursday, July 03, 2025 at 18:58:40 (UTC) by 172.69.130.199. Redistribution subject to AECE license or copyright.]

confirmed by subjective assessment observing the original and stabilized video sequence. As shown in section IV - C, the implemented genetic algorithm for image search can very accurately estimate the motion (disturbance) of the video sequence and reduce the number of calculations compared to the exhaustive search template matching in the whole image. This section presents an objective measure of the success of the stabilization algorithm using the implemented genetic algorithm template matching. To demonstrate the results of the stabilization algorithm two, sequences from the thermal camera are selected: 4 and 5, as well as two sequences from the visible-light camera: 2 and 3. Fig. 11 shows a comparison of the original visible-light sequence 2 with the disturbance and the stabilized sequence in the vertical direction, and in Fig. 12 comparison for sequence 3.

From the given graphics, the effect of the applied video stabilization is clearly visible - the disturbance has been largely eliminated, which is especially pronounced and significant in the vertical direction in which the disturbance acted.

As the parameter for "smoothing" the path is set to 100 in the stabilization algorithm, this effect can be clearly seen in every 100 frames as a step on the graphics of the stabilized sequence, and then in the next 100 frames the graphics are smoother. Although with a shift of every 100 frames, this shift is of low intensity, with a much lower frequency compared to the disturbance in the original sequence, so the video sequence can be observed without interruption. Once again, it has been confirmed that the implemented genetic algorithm template matching can be used very efficiently to verify digital video stabilization because it clearly displays all the effects and changes that occur in the video sequence.

Fig. 13 and Fig. 14 show a comparison of the original thermal video sequences with the disturbance and the stabilized sequences in the vertical direction.

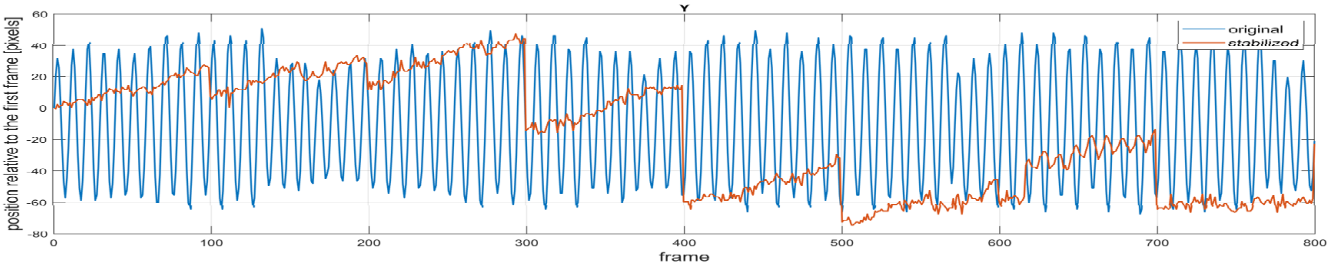


Figure 11. Motion estimation in vertical direction for visible-light video sequence 2: original (blue) and stabilized (red)

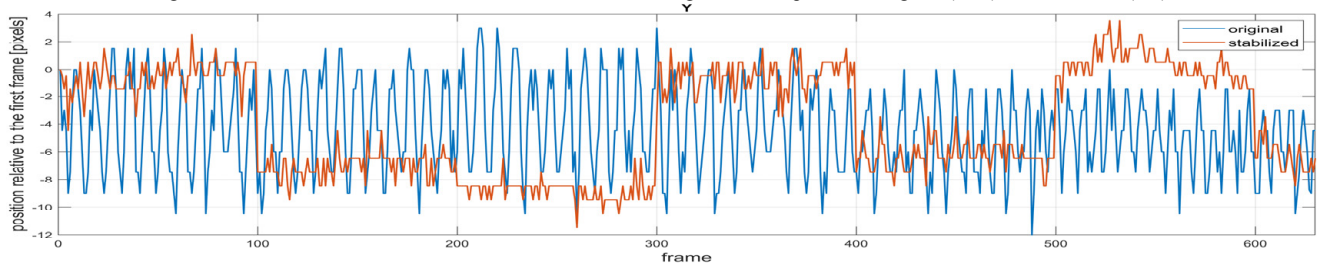


Figure 12. Motion estimation in vertical direction for visible-light video sequence 3: original (blue) and stabilized (red)

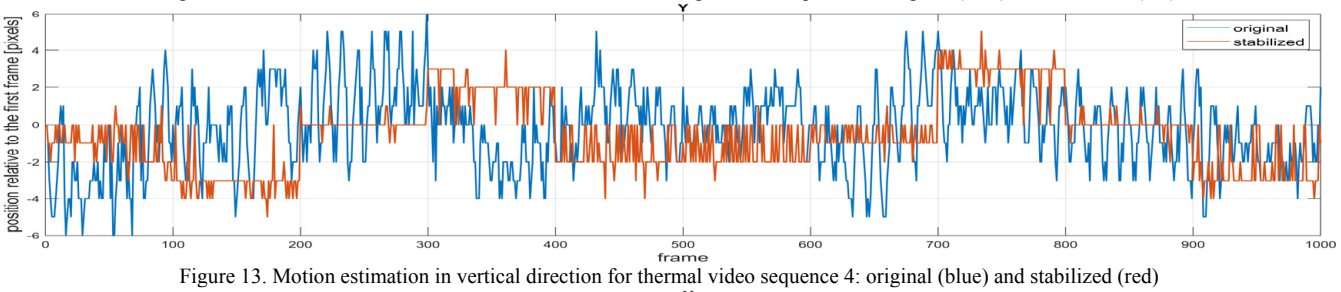


Figure 13. Motion estimation in vertical direction for thermal video sequence 4: original (blue) and stabilized (red)

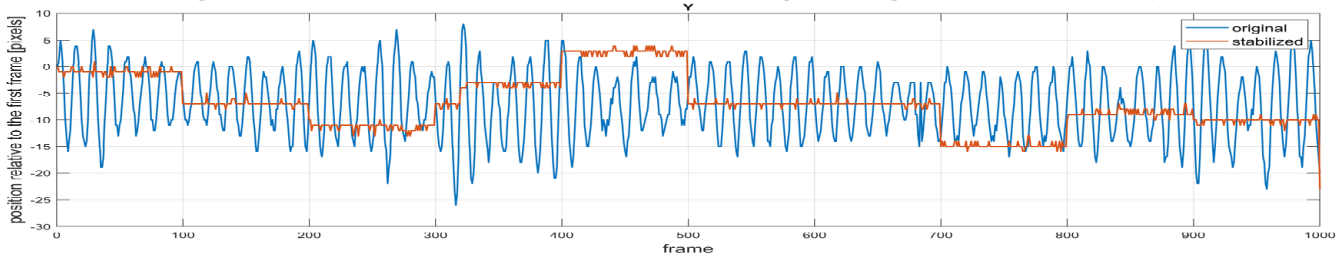


Figure 14. Motion estimation in vertical direction for thermal video sequence 5: original (blue) and stabilized (red)

VI. Conclusion

This paper proposed a method for template matching search in image based on a genetic algorithm, very successful in verification of the digital video stabilization without reference data, when only distorted video sequences are available. The proposed method significantly speeds up the verification process compared to the exhaustive search based on the normalized cross correlation. Experimental results of motion estimation on labeled video sequences show the high accuracy of the proposed method in estimating sensor shaking. As such, applied to verify digital

[Downloaded from www.aece.ro on Thursday, July 03, 2025 at 18:58:40 (UTC) by 172.69.130.199. Redistribution subject to AECE license or copyright.]

video stabilization in the same images gives results consistent with the subjective assessment of the stabilization measure by the human eye and clearly displays all the effects and changes that occur in the video sequence.

For the genetic algorithm template matching verification method, the paper proposes a new fitness function of the genetic algorithm based on the modification of the normalized cross correlation between the image and tracked template. The genetic algorithm in this paper manipulates a 120 binary strings length of 24 in parallel, searching for a larger number of local maxima in the image space. By applying operators and mechanisms of genetic algorithms, including crossover using the roulette wheel method with a probability of 0.9 and inverse mutation with a probability of 0.005, the information on found local maxima is exchanged among individuals of the population to recognize the global one among the found local maxima. The set parameters of the genetic algorithm can remain stable for digital video stabilization verification. This is confirmed by very accurate results in motion estimation on the database of real-life visible-light and thermal video sequences, created in different recording conditions with different nature of disturbances. The search for the image template was performed with raw pixels. This method can be sensitive to changes in image scaling and changes in illumination conditions, so the plan for further research is to use standard image features (HOG, SIFT, SURF) as well as deep features that should be less sensitive to these problems and further improve the performance of the algorithm.

REFERENCES

[1] P. Milanović, I. Popadić, B. Kovačević, "Gyroscope-based video stabilization for electro-optical long-range surveillance systems," Sensors, 21(18), 6219, 2021. doi: 10.3390/s21186219

[2] S. Bell, A. Troccoli, K. Pulli, "A non-linear filter for gyroscope-based video stabilization," in European Conference on Computer Vision, pp. 294-308, September 2014. doi:10.1007/978-3-319-10593-2_20

[3] C. Jia, Z. Sinno, B.L. Evans, "Real-time 3D rotation smoothing for video stabilization," in Proc. of the 49th Asilomar Conference on Signals, Systems and Computers, Pacific Grove, CA, USA, 8–11 November 2015. doi:10.1109/ACSSC.2014.7094532

[4] S. Liu, H. Li, Z. Wang, J. Wang, S. Zhu, B. Zeng, "DeepOIS: Gyroscope-guided deep optical image stabilizer compensation," IEEE Transactions on Circuits and Systems for Video Technology, 2021. doi:10.1109/TCSVT.2021.3103281

[5] W. Guilluy, L. Oudre, A. Beghdadi, "Video stabilization: overview, challenges and perspectives," Signal Processing: Image Communication, 90,116015, 2021. doi:10.1016/j.image.2020.116015

[6] J. Dong, H. Liu, "Video stabilization for strict real-time applications," IEEE Transactions on Circuits and Systems for Video Technology, 27(4), pp. 716-724, 2016. doi:10.1109/TCSVT.2016.2589860

[7] X. Cheng, Q. Hao, M. Xie, "A comprehensive motion estimation technique for the improvement of EIS methods based on the SURF algorithm and Kalman filter," Sensors 2016, 16, 486. doi: 10.3390/s16040486

[8] S. Jen, I. Yoon, J. Jang, S. Yang, J. Kim, J. Paik, "Robust video stabilizatio using particle keypoint update and L1-optimized camera path," Sensors 2017, 17(2), 337. doi:10.3390/s17020337

[9] R. Wu, Z. Xu, J. Zhang, L. Zhang, "Robust global motion estimation for video stabilization based on improved k-means clustering and superpixel," Sensors 2021, 21(7), 2505. doi:10.3390/s21072505

[10] O. Manish, K. P. Biswas, "Video stabilization using maximally stable extremal region features," Multimedia Tools and Applications, vol. 68, no. 3, pp. 947–968, 2014. doi:10.1007/s11042-012-1095-z

[11] M. R. Souza, H. Pedrini, "Digital video stabilization: algorithms and evaluation," Anais Estendidos da XXXII Conference on Graphics, Patterns and Images, pp. 35-41, October 2019

[12] E. Altinisik, H. T. Sencar, "Source camera verification for strongly stabilized videos," IEEE Transactions on Information Forensics and Security, 16, pp. 643-657, 2020. doi:10.1109/TIFS.2020.3016830

[13] N. Vlahovic, B. Stojanovic, S. Stankovic, K.O. Al Ali, "Dehazing algorithms influence on video stabilization performance," 8th International Scientific Conference on Defensive Technologies OTEH 2018

[14] M. Seul, L. O'Gorman, M. J. Sammon, "Practical algorithms for image analysis: description, examples and code," Cambridge University Press, Cambridge, 2000

[15] R. C. Gonzalez, R. E. Woods, "Digital image processing, fourth edition," Pearson, 2018

[16] K. Briechle, U. D. Hanebeck, "Template matching using fast normalized cross correlation," Optical Pattern Recognition XII, vol. 4387, pp. 95-102, International Society for Optics and Photonics, March, 2001. doi:10.1117/12.421129

[17] J. P. Lewis, "Fast template matching", Vision Interface, pp. 120-123, 1995

[18] J. P. Lewis, "Fast normalized cross correlation," Industrial Light and Magic, 1995

[19] X. H. Zhang, Q. Liu, M. Li, "A template matching method based on genetic algorithm optimization," In Applied Mechanics and Materials, vol. 220, pp. 1298-1302, Trans Tech Publications Ltd, 2012. doi:10.4028/www.scientific.net/AMM.220-223.1298

[20] L. Si, X. Hu, B. Liu, "Image matching algorithm based on the pattern recognition genetic algorithm," Computational Intelligence and Neuroscience, 2022. doi:10.1155/2022/7760437

[21] J. H. Holland, "Adaptation in natural and artificial systems," University of Michigan Press, Ann Arbor, 1975

[22] S. Katoch, S. S. Chauhan, V. Kumar, "A review on genetic algorithm: past, present, and future," Multimedia Tools and Applications, 80(5), 8091-8126, 2021. doi:10.1007/s11042-020-10139-6

[23] S. Mirjalili, "Genetic algorithm," Evolutionary algorithms and neural networks, pp. 43-55, Springer, Cham, 2019

[24] A. Lambora, K. Gupta, K. Chopra, "Genetic algorithm - a literature review," International Conference on Machine Learning, Big Data, Cloud and Parallel Computing (COMITCon), pp. 380-384, 2019. doi:10.1109/COMITCon.2019.8862255

[25] V. Kreinovich, C. Quintana, O. Fuentes, "Genetic algorithms: what fitness scaling is optimal?," Cybernetics and Systems, 24(1), 9-26, 1993. doi:10.1080/01969729308961696

[26] K. Jebari, M. Madiafi, "Selection methods for genetic algorithms," International Journal of Emerging Sciences, 3(4), 333-344, 2013

[27] G. K. Soon, T. T. Guan, C. K. On, R. Alfred, P. Anthony, "A comparison on the performance of crossover techniques in video game," 2013 IEEE international conference on control system, computing and engineering, pp. 493-498, IEEE, 2013

[28] V. Vasco, A. Glover, C. Bartolozzi, "Fast event-based Harris corner detection exploiting the advantages of event-driven cameras," IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), Daejeon, Korea, 2016. doi:10.1109/IROS.2016.7759610

[29] J. Bouguet, "Pyramidal implementation of the affine Lucas Kanade feature tracker description of the algorithm," Intel Corporation, vol. 5, pp. 1-10, 2001

[30] T. K. W. Y. Sunglok Choi, "Robust video stabilization to outlier motion using adaptive RANSAC," in Proc. of the International Conference on Intelligent Robots and Systems, 2009

[31] E. Dubrofsky, "Homography Estimation," Diplomovápráce, The University of British Columbia, Vancouver, 2009

[32] N. Vlahović, M. Stojanović, M. Stanković, S. Stanković, "Adaptive video stabilization using Kalman filtering," in Proc. of the 5th International Conference on Electrical, Electronics and Computing Engineering, Palić, Serbia, 2018

[33] N. Latinović, I. Popadić, B. Tomić, A. Simić, P. Milanović, S. Nijemčević, M. Veinović, "Signal processing platform for long range multi-spectral electro-optical systems," Sensors, 22(3), 1294, 2022. doi:10.3390/s22031294