International Journal of Engineering & Technology, 7 (4.19) (2018) 705-710

International Journal of Engineering & Technology

Website: www.sciencepubco.com/index.php/IJET

Research paper

Camera Motion Estimation based on Phase Correlation

Abdulkadhem Abdulkareem Abdulkadhem[1*], Tawfiq A. Al-Assadi[2]

[1,2]College of Information Technology, University of Babylon, Babylon, Iraq.
**Corresponding Author Email: kazum2006k@yahoo.com*

Abstract

In this paper, we introduce a new style for a relative localization estimation and trajectory determination of a camera sensor based on a vision in GPS-denied environments. The input to the system is video film taken from a camera placed on the vehicle as forward facing camera. The output of the system is a trajectory (path) of camera movement .The proposed framework consists of many main steps, the first one extracts the FFT of two consecutive frames of video. The next step is to find the entry-wise product of frequency domain of frames. The third step is extracting the FFT inverse of entry-wise product. Next, the system finds the location of a maximum peak that represents the translation motion between two frames of video. The proposed system is faster than traditional methods that depend on spatial features and system have done without any external information of camera calibration.

Keywords: *Pose estimation, phase correlation, visual odometry , camera motion, trajectory extraction.*

1. Introduction

The estimation of position for vehicles depend on cameras is now a hot study subject in robotic and computer vision fields, with limited application in the GPS-denied environments. The visual information are used for ego-motion extraction which offerings numerous problems, for example the search features, association of data (feature correlation), inhomogeneous features distribution in the image, and so forth.[1]
A large portion of the outside localization strategies utilizes Global Positioning joined with other Navigation Systems for correct positioning error. Be that as it may, GPS might be missing or turned out to be less compelling when just a small number of satellites are accessible in urban, remote areas or inside the water. In spite of the fact that the algorithm like Visual odometry (VO) and Simultaneous Localization and Mapping (SLAM) are utilized in outdoor localization and they need complex kinds of hardware and high computational power. Commercially available video cameras offer additional information than other sensors. Hence visual odometry based localization can be successfully utilized as few a low cost localization alternative. Currently the image processing techniques will be expanded and the technologies of video capturing support visual odometry localization with less computational overhead.[2,3] The scenario of the proposed system is using monocular camera (single) setups as a Forward-Facing Camera (FFC) of a vehicle or car Fig(1). The goal of the proposed system is to extract the trajectory of the vehicle based on video film only without any other information.

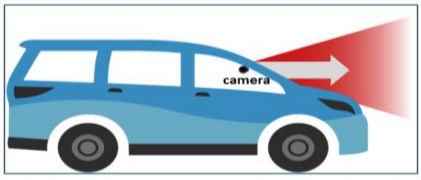


Fig. 1: The forward-facing camera

Most ego-motion methods depend on the corresponding the extracting features from the spatial domain of frames of video. In these methods, extract important features such as corner points from the first frame and tracking points to the second frame for extracting the rotation matrix and translation vector between frames and then estimate the trajectory movement. The drawback of these methods are slow and can not be done without other parameters such as (intrinsic, camera, extrinsic camera, focal length, the principal point, distortion coefficients, ground truth locations, camera calibration). Thus, our proposed system focus on extracted features from the frequency domain of frames to estimate the camera motion between frames and it is fast than traditional methods and without need any other external parameters. This leads to a complex calculation involving lots of information and it's sensitive to a small distinction among video frames. The methods based on the frequency domain transforms the image from the spatial domain to the frequency domain, and get the relationships of rotation, translation, and zoom over Fourier transformation (like the phase-correlation, Walsh transform). [4]
In the organization of paper, section 2, Background, gives a thorough explanation of the problems with visual odometry that represent the traditional related works for motion estimation based on vision and also explain the phase correlation method. Section 3 introduce the proposed system, which describes how the system presented. Section 4, Experimental Results, gives an explanation about the output trajectory of camera motion. In section 5, Conclusions, where the writers reflect on the used.

2. Background

In this section we explain the visual odometery that represent the traditional ego-motion estimation that depend on the spatial features of video frames. The second sub section explain the phase correlation that is one method of image registration

Copyright © 2018 Authors. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

2.1 Visual odometry

A special situation of Structure From Motion (SFM) is visual odometry (VO) [5,6]. An image series (frames) in this method to discover the motion of important points feature between frames. The developments are utilized in arrange to calculate the motion and pose of the camera, in this way giving the position. VO and SLAM are a fundamental tasks in an space where a certain set of standard techniques have advanced all through the recent 2 to 3 decades of intensive research such as descriptor-based matching, complex feature descriptors and feature detectors, RANdom Sample, Consensus (RANSAC)-supported, Perspective-n-Point (PnP). These techniques can be used for extract the relative pose between video frames [7].
The general pipeline of visual odometry methods illustrates as follow:
1- Read video film.
2- Take two consecutive frames I1, I2.
3- Extract important feature points by using one of the corner detection methods such as The Scale-Invariant Feature Transform (SIFT), (Features from Accelerated Segment Test (FAST), Speeded Up Robust Features (SURF), Harrise.[8]
4- Discovery the corresponding features in another image which can be completed by either using matching feature or tracking feature (by using the Lucas-Kanade pyramidal optical flow algorithm).
5- Finding the essential matrix E or Fundamental matrix between two matching points of two frames
6- Decompose E into three Matrices by using Singular Value Decomposition (SVD).
7- Finding the rotation matrix R and translation vector t from three matrices of SVD. There are two methods to extract R and t. The first one is computing four possible solutions, one only is true and extracted by using the triangulation method and others are wrong. The four possible solutions are explained in equations (1-4):

$R1=UW'V^t$ (1)

$R2=UWV^t$ (2)

$t1=+u[3]$ (3)

$t2=-u[3]$ (4)

The second method for finding R and t as follow equations:

$R= UWV'S$ (5)

$t= UWV'S$ (6)

where U represents the left matrix of SVD, V represents the right matrix of SVD, S represents the singular matrix of SVD and W represents a 3x3 matrix and equal to
[[0 -1 0];[1 0 0]; [0 0 1]].
8- Estimate location based on t vector.[9]
The method above extract features on spatial domain and cannot give us a true trajectory without using external parameters of a video camera or at least we know some world 3D-points between two frames to estimate true R and t by using the bundle adjustment method. An alternative to this methodology is to use spectral features, or frequency-based features, and to determine correspondences in the frequency domain. Another problem it's slow compared with features extraction in the frequency methods. Our proposed system depends on the frequency domain by using phase correlation that is an approach used the Fast Fourier Transform (FFT) as a frequency domain.

2.2. Phase correlation

Image registration is a critical mission in image processing to overlap at least two images. Registration techniques can be separated into the following classes: some algorithms witch utilization image pixel values straightforwardly, e.g., correlation strategies; some algorithms witch utilization the frequency domain, e.g., fast Fourier transform (FFT) techniques; some algorithms that utilization low level features, for example, edges and corners, e.g., features-based techniques ; and methods that utilization high-level features, for example, recognized (parts of)objects, or relations between features, e.g., graph-theoretic techniques.
The idea behind this registration strategy is based on the Fourier shift property, which states that a shift in the coordinate frames is transformed to the Fourier domain also as a linear phase contrast.[3]
In this correspondence, we introduce an expansion of the phase correlation procedure for registration of image automatically, which is described by its insensitivity to scaling, rotation, translation, and noise and additionally by its low cost computational.[10]
The phase correlation is a technique of cross-correlation depend on the Fourier transform; phase correlation calculation is very quickly because of the wide optimizations that fast Fourier transform (FFT) algorithms allow. In its most straight forward implementation, phase correlation are used to register collected two images witch vary by a relative translation and can be extended to changes in scaling and rotation.[11]
Phase correlation method consists of the following steps:
1- Given 2 images (frames of video) $\boldsymbol{ga}$ *and* $\boldsymbol{gb}$.
2- Calculating the discrete 2D-Fourier Transform of both images: $\boldsymbol{Ga}=\boldsymbol{F}\{\boldsymbol{ga}\},\boldsymbol{Gb}=\boldsymbol{F}\{\boldsymbol{gb}\}$.
3- Calculating the cross-power spectrum by multiplying the first Fourier transform and the complex conjugate of the second Fourier transform and normalizing the product element-wise.

$$\boldsymbol{R} = \frac{\boldsymbol{G_a}\circ\boldsymbol{G_b^*}}{|\boldsymbol{G_a}\circ\boldsymbol{G_b^*}|} \qquad (7)$$

4- Apply the inverse Fourier transform to acquire the normalized cross-correlation.

$$\boldsymbol{r}=\boldsymbol{F}-\boldsymbol{1}\{\boldsymbol{R}\} \qquad (8)$$

5- Determine the peak location in the r.[12]

$$(\boldsymbol{\Delta x}, \boldsymbol{\Delta y}) = \boldsymbol{argmax}_{(x,y)}\{\boldsymbol{r}\} \qquad (9)$$

3. The Proposed System

The proposed system as illustrated in Fig (2)" consists of many steps.
The video is indeed the collection of images, each image is called the frame, displayed in a fast-paced enough so that the human eye can be aware of the continuity of its content. The image processing techniques can be applied to each frame. The contents of the sequential two frames are generally closely correlated [13]. The first step of the proposed system is the input video film file of a camera in an urban environment (street), the camera puts on a car as forward facing camera and capture a stream of images, let the images collection taken at the discrete time instants k be denoted by $I_{0:n} = \{I_0, \ldots, I_n\}$. The next operation is converting this video frames from 24-bit color RGB(Red, Green, Blue) to gray-scale form. The others steps are explained as follow:

3.1. Preprocessing and Filtering Process

The preprocessing and filtering process involved noise removal, smoothing, and enhancement the frames of video by using low pass filtering, median filtering , or high pass filterin.

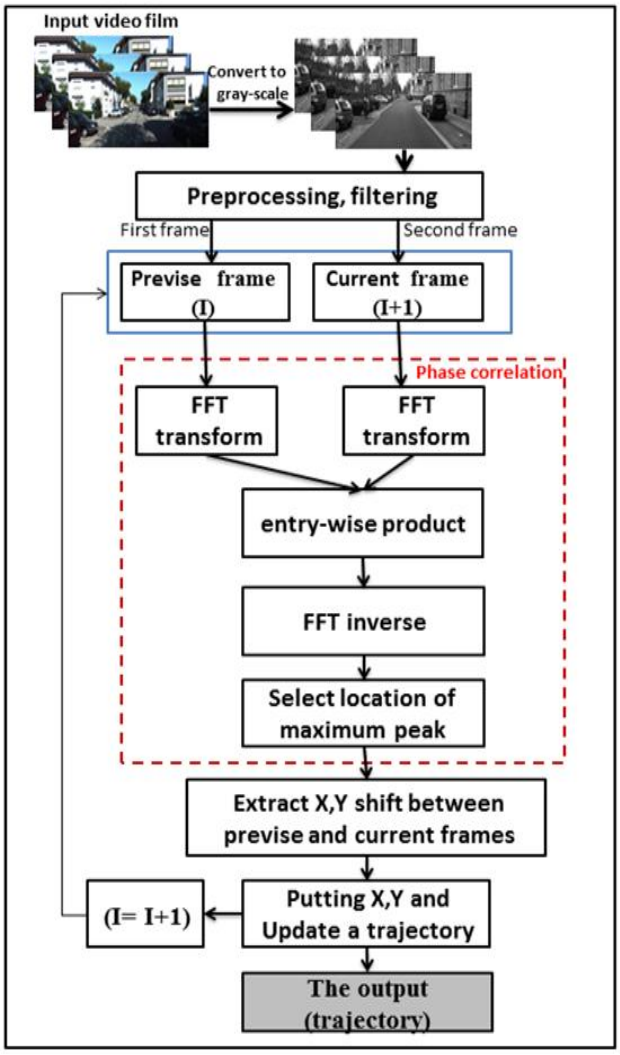


Fig. 2: The block diagram of the proposed system

3.2. Taking Two Frames from Video

The next step of the proposed system is taking two consecutive frames (pair) from the video as a previous and current frame. The initial case set the first frame of video (I_0) to the previous frame and the second frame of video (I_1) to the current frame. In the next loop, the previous frame will be (I_1) and the current frame will be (I_2) and so on. Fig(3) illustrate this operation.

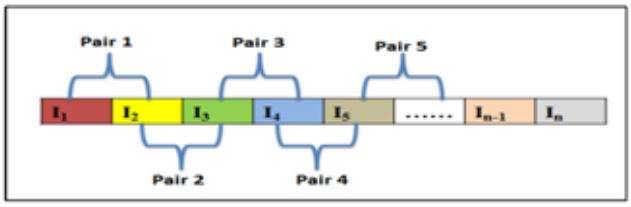


Fig. 3: The pairs of frames that taking sequentially from the stream of video

3.4. Fourier Transformations

Fourier transform is a standout amongst the most huge apparatuses which have been extensively utilized not just to understanding the nature of the images and its formation yet in addition for processing the images. It has been possible to examine the images by utilizing Fourier transform as an arrangement of spatial sinusoids in various directions, every sinusoid having a specific frequency. The two-dimensional discrete Fourier transform of a two-dimensional signal f (x , y) of dimension M x N with integer indices x and y running from 0 to M - 1 and 0 to N - 1, is represented by

$$F(u,v) = \frac{1}{MN}\sum_{x=0}^{M-1}\sum_{y=0}^{N-1} f(x,y)\exp\left[-j2\pi\left(\frac{ux}{M}+\frac{vy}{N}\right)\right]. \qquad (10)$$

The number of complex multiplications and additions to compute the Discrete Fourier transform (DFT) is $O(N^2)$. we can approve a divide-and-conquer method to decrease the computational complexity of the method to $O(N \log N)$. This method is generally well-known as Fast Fourier Transform (FFT). [14]

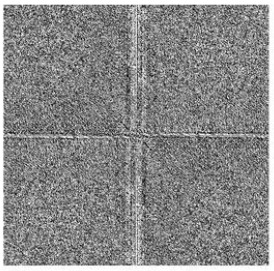
In this step, the proposed system takes two consecutive frames from the above step and apply 2D-FFT transformation for each frame to convert the domain of the video frames from spatial to the frequency domain. The fig(4) illustrated an example of how to convert block of size (8x8) of lean image and whole image from spatial representation to the frequency representation by using 2D-FFT transform.



Lena gray image of size 512x512

	1	2	3	4	5	6	7	8
1	108	81	52	41	44	52	53	55
2	109	83	52	38	42	49	57	56
3	110	81	52	38	42	51	52	55
4	107	80	50	37	42	49	51	55
5	109	82	49	35	39	47	51	54
6	108	81	51	35	40	42	52	52
7	106	83	51	32	37	45	50	53
8	103	82	51	31	37	42	48	53

Block of size 8x8



FFT transform of Whole Lena image

	1	2	3	4	5	6	7	8
1	3755	835.4	361	239	255	239	361	835
2	20.7	5.707	-3.4	-1.05	-5.5	-4.39	-13	-24
3	14	-6.31	8	-9.59	-4	16.3	2	-12
4	19.3	-10.9	2.78	4.29	1.54	8.26	9.36	-26
5	25	-6.31	9	16.3	-15	16.3	9	-6.3
6	19.3	-25.6	9.36	8.26	1.54	4.29	2.78	-11
7	14	-12.4	2	16.3	-4	-9.59	8	-6.3
8	20.7	-24.3	-13	-4.39	-5.5	-1.05	-3.4	5.71

FFT transform of 8x8 block

Fig. 4: The FFT transform of the whole image and block with size (8*8)

3.5. The Entry-wise product (cross-power spectrum)

In arithmetic, the Entry-wise product(otherwise called the Hadamard) is a binary task that takes two matrices of similar dimensions and creates another matrix where every component in the location (i,j) is the product of elements in the location (i,j) of the above original two matrices. For instance, the Entry-wise product for a 3×3 matrix (A) with a 3×3 matrix (B) is:[15]

$$\begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} \circ \begin{bmatrix} b_{11} & b_{12} & b_{13} \\ b_{21} & b_{22} & b_{23} \\ b_{31} & b_{32} & b_{33} \end{bmatrix} = \begin{bmatrix} a_{11}b_{11} & a_{12}b_{12} & a_{13}b_{13} \\ a_{21}b_{21} & a_{22}b_{22} & a_{23}b_{23} \\ a_{31}b_{31} & a_{32}b_{32} & a_{33}b_{33} \end{bmatrix}$$

In this step, the proposed system takes two matrices from the above step (the first Fourier transform of the first frame and the complex conjugate of the second Fourier transform of the second frame) and apply the Hadamard product in the equation (7) for them to produce a new matrix with the same size of original frames from above step.

3.6. The FFT inverse

The equivalent two-dimensional inverse DFT is

$$f(x,y) = \sum_{u=0}^{M-1}\sum_{v=0}^{N-1} F(u,v)\exp\left[j2\pi\left(\frac{ux}{M}+\frac{vy}{N}\right)\right]. \quad (11)$$

Now, in this step, Appling 2D-FFT inverse transform for a matrix of above step to produce the phase correlation between two sequential frames, fig (5) show this operation.

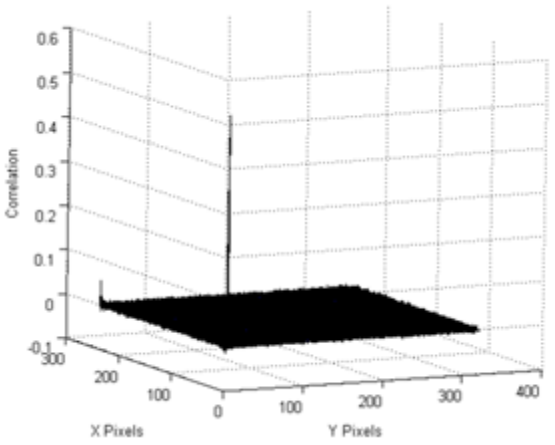


Fig. 5: The Phase correlation between two images. The maximum Peak represents the location of translation.

3.7. Selecting the Maximum Peak

From the figure above the system selects the location of the maximum peak that represents the shift translation between two sequential frames of video.

3.8. Extracting the Shift Translation

In this step, Subtract the X coordinate of the peak point of the above step from the X coordinate of the center location of the current frame. Subtract the Y coordinate of the peak point of the above step from the Y coordinate of the center location of the current frame. The rustles represent the shift translation between the previous frame and the current frame.

3.9. Updating the trajectory of the camera moving

The first point or the starting point of a trajectory in the location (0,0). The next movement of a trajectory computed as the following equation

$$X=X+ (Shift_X_t - Shift_X_{t+1}) \quad (12)$$

$$Y=Y+ Abs(Shift_Y_t - Shift_Y_{t+1}) \quad (13)$$

Where ($Shift_X_t$, $Shift_Y_t$) corresponding to the translation between the frame (I) and frame (I-1), and (Shift_Xt+1, Shift_Yt+1) corresponding to the translation between the frame (I) and frame (I+1).
The next operation is to plot the point of new (X,Y) location of a trajectory. The last operation is taking the next pair of consecutive frames when the previous frame equal to the current frame and takes the next frame (I=I+1) and go to step B. the system is continued until to reach the last consecutive frames of video (I_{n-1}, I_n) and in this step the system is ready to display the camera movement trajectory of vehicle as output.

4. Experimental Results

Experiments of the proposed strategy did to demonstrate the efficiency. The proposed method has been simulated using vb.net program on the Windows 7 platform on Intel Core i5 2.5 GHz with 4 GB of main memory. Experiments of the proposed method in this paper are performed on two sets of sequential frames of video film taken from the Kitti vision benchmark[16]. The frame dimensions are (1267x387) pixels. The type of camera used in the proposed system is a monocular camera (the single camera). The first set sequence frames explain the forward camera movements as illustrated below in fig(6).



Fig. 6: The first set of frames with forwarding camera motion.

The starting point of a trajectory in the location (0,0). The translation shift (phase correlation) between the frames of video are computed and update the location of the current movement depends on the previous point of the trajectory. The translation between the frames represents the location of the maximum peak of phase correlation that illustrated in the table(1).

Table 1: The Shift Translation and Locations of a Trajectory for the First Set of Frames

Pairs frame	Shift translation		Location of trajectory	
	X	Y	X	Y
1-2	1.5	0.5	1.5	0.5
2-3	0.5	-0.5	0.5	1.5
3-4	1.5	0.5	1.5	2.5
4-5	0.5	0.5	0.5	2.5
5-6	1.5	2.5	0.5	4.5

Fig (7) illustrates the phase correlation maps between the frames of the first set (forward moving) and fig (8) shows the corresponding trajectory movement.

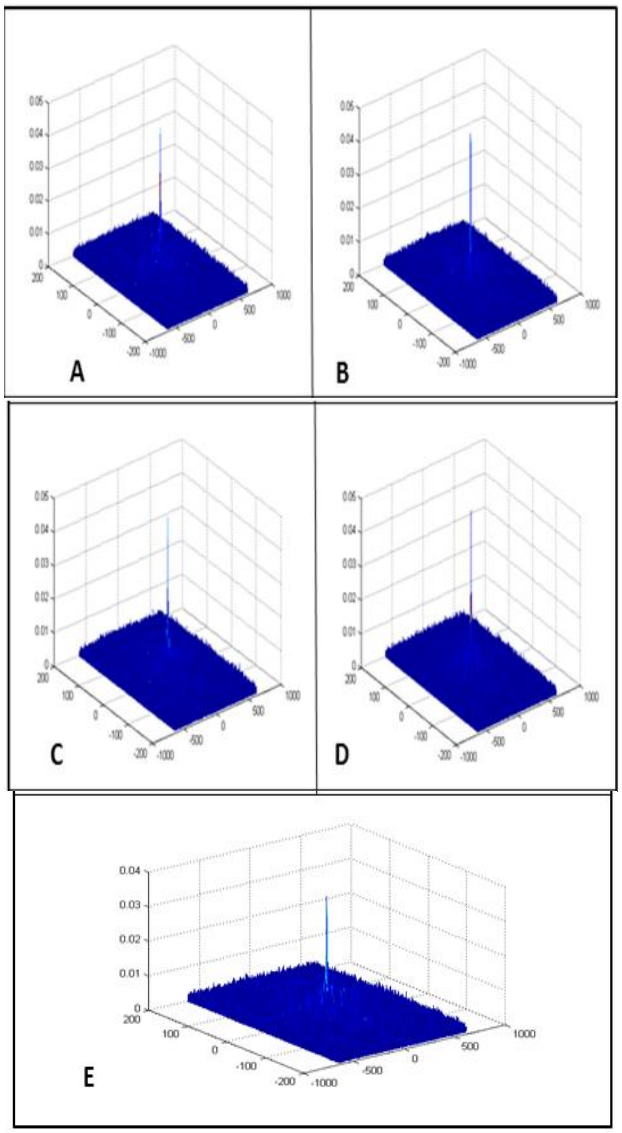


Fig. 7: The phase correlation maps between the frames video of the first set (forward moving). **A**. corresponding the phase correlation between frames(1-2), **B**. phase correlation between frames (2-3), **C**. represent phase correlation between frames (3-4), **D**. represent the phase correlation between frames (4-5) and **E.** represent the phase correlation between frames (5-6).

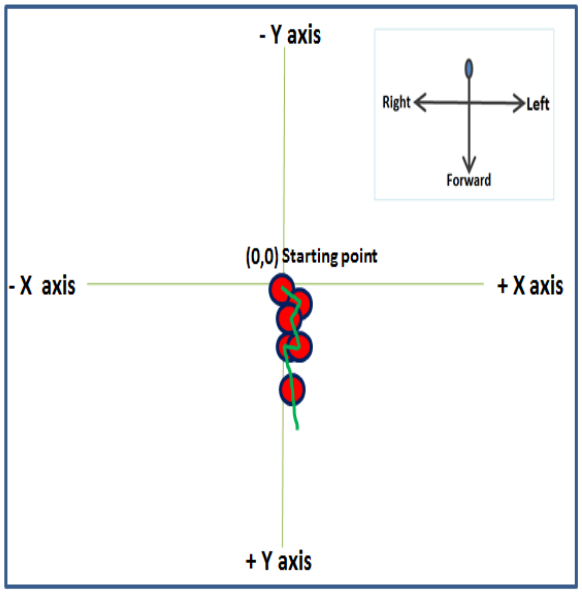


Fig. 8: The locations of a trajectory for the first set of frames video

We notice the change in both the X and Y axis of trajectory are very small, thus the camera motion trajectory will be straightforward with small distortions on left or right.

The other set of sequence frames are illustrated in fig(9). We notice an abrupt change in camera movement to the right side.

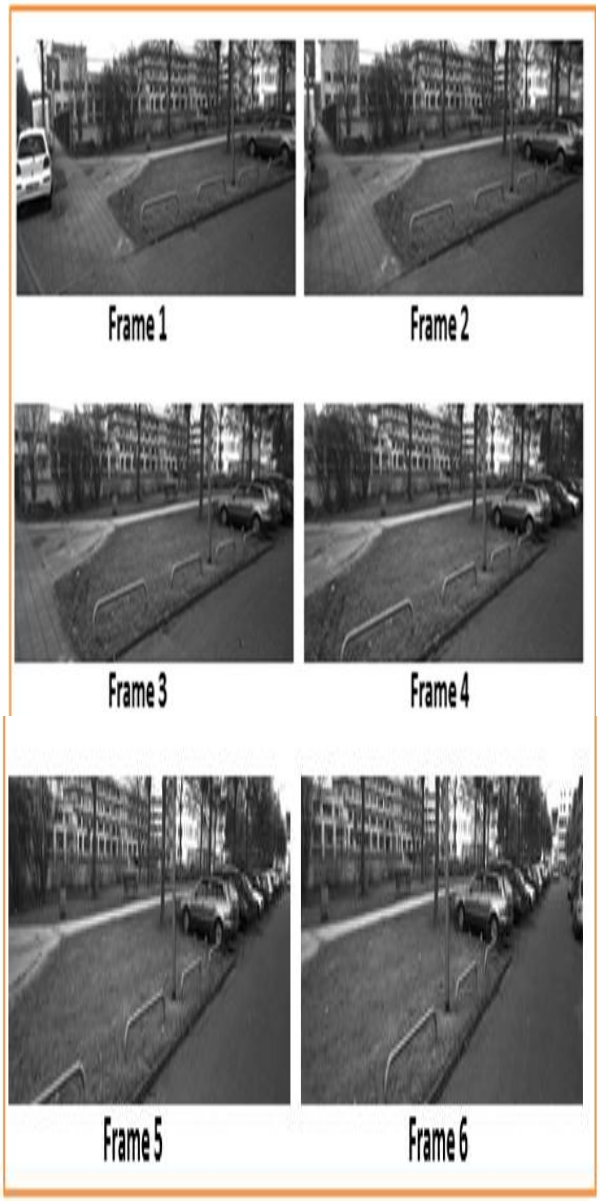


Fig. 9: The second set of frames with an abrupt change of camera motion direction

The translation between the frames represents the location of the maximum peak of phase correlation that illustrated in the table (2).

Table 2: The Shift Translation and Locations of a Trajectory for the Second Set of Frames Video

Pairs frame	Shift translation		Location of trajectory	
	X	Y	X axis	Y axis
1-2	-27	-0.5	-27	0.5
2-3	-56	-3.5	-56	3.5
3-4	-75.5	-2.5	-67	4.5
4-5	-81.5	-1.5	-74	5.5
5-6	-82.5	-1.5	-80	5.5

Fig (10) illustrates the phase correlation maps between the frames of the second set (an abrupt change in direction) and fig (11) shows corresponding trajectory movement.

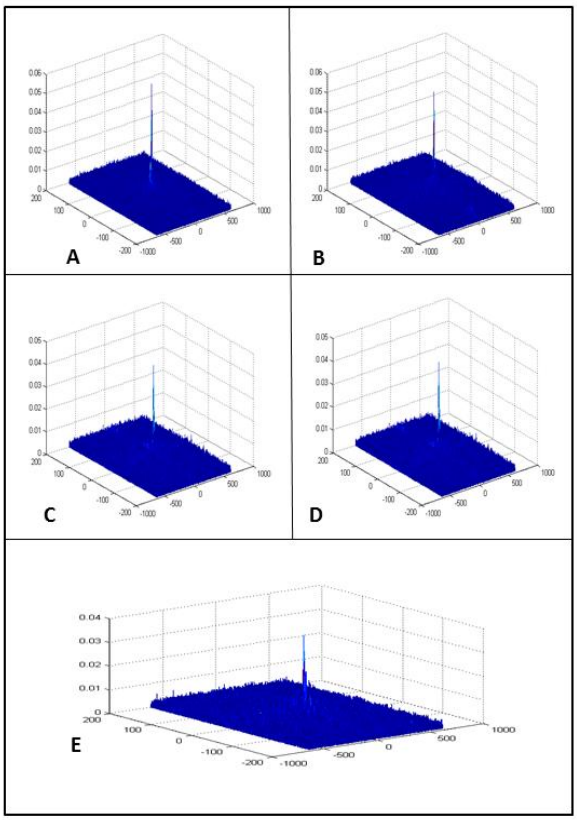


Fig. 10: The phase correlation maps between the frames video of the second set (an abrupt change in direction moving). **A.** corresponding the phase correlation between frames(1-2), **B.** phase correlation between frames (2-3), **C.** represent phase correlation between frames (3-4) , **D.** represent the phase correlation between frames (4-5) and **E.** represent the phase correlation between frames (5-6)

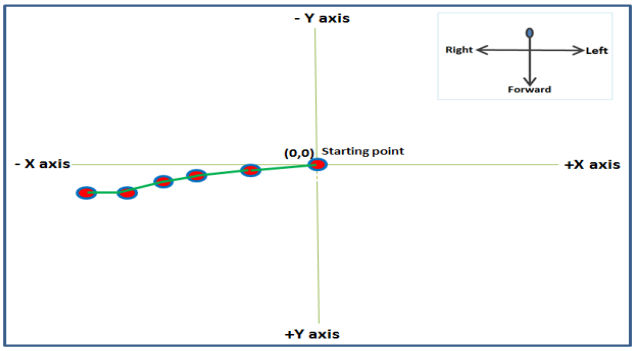


Fig. 11: The locations of a trajectory for the second set of frames video

We notice the change of both the X and Y axis of trajectory are very large, thus the camera motion trajectory will change in direction to the right or to the left movement. The important axis of paper is the X-axis. When the absolute value of X is large that indicate there is a high change in direction moving. If the value of X is a negative value, thus the change will be in the right direction displacement, and if the value of X is a positive value, then the change will be in the left change displacement.

5. Conclusion and Future Works

An efficient camera motion estimation system with a single forward-looking camera was effectively built and proved by analyzing the image shift from frame to frame. This work established a very different way of estimating the motion based on the phase correlation that demonstrates a very fast method of estimating the motion, finding similar accuracy to feature correlation. The main advantages of using spectral features as in this proposal are its robustness in low-quality features, works well in noisy environments, faster than spatial features correlation and without needs any external parameters of a camera, the proposed system needs only the video file to estimate the trajectory. In future work, we will normalize the trajectory extracted and compare the results with measurements from a GPS system. We are currently in the process of acquiring such data

References

[1] Gastón Araguás and etl., “Chapter 2 Monocular Pose Estimation for an Unmanned Aerial Vehicle Using Spectral Features”, book, Springer International Publishing Switzerland, 2017, doi: 10.1007/978-3-319-44735-3_2.

[2] Lasitha Piyathilaka and etl., “An Experimental Study on Using Visual Odometry for Short-run Self Localization of Field Robot”, IEEE, 2010 Fifth International Conference on Information and Automation for Sustainability, pp. 150-155, 2010.

[3] D. KnuthMerwan Birem and etl., “Visual odometry based on the Fourier transform using a monocular ground-facing camera”, Springer-Verlag GmbH Germany , J Real-Time Image Proc, SPECIAL ISSUE PAPER, 2017, doi: 10.1007/s11554-017-0706-3 .

[4] Fan YANG, Linlin WEI, Zhiwei ZHANG and Hongmei TANG, “Image Mosaic Based on Phase Correlation and Harris Operator”, Journal of Computational Information Systems vol. 8, no. 6 , 2012. IEEE Trans. Antennas Propagat., doi:10.4316/ieee.1959.3422.

[5] Scaramuzza D. and Fraundorfer F. “Visual Odometry: Part I - The First 30 Years and Fundamentals”. IEEE Robotics and Automation Magazine, vol. 18, no. 4.2011.

[6] Fraundorfer F. and Scaramuzza D. “Visual Odometry: Part II - Matching, Robustness, and Applications”. IEEE Robotics and Automation Magazine, vol. 19, no.2,2012.

[7] Nolang Fanani and etl., “Predictive monocular odometry (PMO): What is possible without RANSAC and multiframe bundle adjustment?”, Image and Vision Computing (2017), doi: 10.1016/j.imavis.2017.08.002.

[8] Dr. I sra'a Hadil and Hikmat Z. Neima, “Robust Video Shot Importance Measurement Based on SIFT and Optical Flow”, International Journal of Pure and Applied Mathematics, vol. 119, no. 15. 2018.

[9] HENRIK BERG and etl., “Visual Odometry for Road Vehicles Using a Monocular Camera”, Master thises, Department of Signals and Systems ,Chalmers University Of Technology, Gothenburg, Sweden 2016.

[10] B. Srinivasa Reddy and B. N. Chatterji, “An FFT-Based Technique for Translation, Rotation, and Scale-Invariant Image Registration”, IEEE TRANSACTIONS ON IMAGE PROCESSING, vol. 5, no. 8, 1996.

[11] Bogdan Kwolek,” Visual Odometry Based on Gabor Filters and Sparse Bundle Adjustment”, Faculty of Electrical and Computer Engineering, Rzesz´ow University of Technology, W. Pola 2, 35-959 Rzesz´ow, Poland.

[12] Ricardo Ramirez, “Fourier Techniques and Monocular Vision for Simplistic and Low-Cost Visual Odometry in Mobile Robots”, Research Experience for Undergraduates , South Dakota School of Mines and Technology,2016.

[13] Israa Hadi Ali and Sarah Abdul Rizah Abd, “Proposed New Method of Enhancement Object Trajectory Based on Typical Trajectory”, AL-Bahir Quarterly Adjudicated Journal for Natural and Engineering Research and Studies, vol. 4, no. 7, 2016.

[14] Scott E. Umbaugh, “Digital Image Processing and Analysis: Human and Computer Vision Applications with CVIPtools”,Book, Second Edition, 2010.

[15] https://en.wikipedia.org/wiki/Hadamard_product_(matrices) , Accessed 6/7/2018.

[16] Andreas Geiger and etl., “Vision meets Robotics: The KITTI Dataset”, International Journal of Robotics Research (IJRR),2013.