Object tracking using motion flow projection for pan-tilt configuration

Luma Issa Abdul-Kreem[1], Hussam K. Abdul-Ameer[2]
[1]Department of Control and Systems Engineering, University of Technology, Iraq
[2]Department of Biomedical Engineering, Al-Khawarizmi College of Engineering, University of Baghdad, Iraq

ABSTRACT

We propose a new object tracking model for two degrees of freedom mechanism. Our model uses a reverse projection from a camera plane to a world plane. Here, the model takes advantage of optic flow technique by re-projecting the flow vectors from the image space into world space. A pan-tilt (PT) mounting system is used to verify the performance of our model and maintain the tracked object within a region of interest (ROI). This system contains two servo motors to enable a webcam rotating along PT axes. The PT rotation angles are estimated based on a rigid transformation of the the optic flow vectors in which an idealized translation matrix followed by two rotational matrices around PT axes are used. Our model was tested and evaluated using different objects with different motions. The results reveal that our model can keep the target object within a certain region in the camera view.

Corresponding Author:

Luma Issa Abdul-Kreem,
Department of Control and Systems Engineering,
University of Technology,
Baghdad, Iraq.
Email: 60021@uotechnology.edu.iq

1. INTRODUCTION

Visual object tracking is one of the interesting fields in computer vision which is intervened in different applications such as surveillance, traffic monitoring, human computer interaction, and robotics [1–4]. Object tracking, in essence, seeks to estimate motion trajectory of an object in visual scenario. This research seeks to address the following questions: how a path for a moving object can be estimated in world plane? and how this estimation can be analyzed to generate a feedback signals to enable a PT system tracking an object? A considerable amount of literature has been published on object tracking (for more discussion, see the review in [5]). The related studies to our work will be reviewed. In [2], the authors used raw pixel representation for object tracking in which colors and intensity values of an image are used to represent the object regions. Such representation is simple and effect, however, raw pixel information alone is not sufficient for robust tracking [6]. This issue led some researchers [7, 8] to include other visual features such as shape, edge and texture into the raw pixel representation.

Other studies such in [9, 10], on the other hand, used histogram representation to track an object in the visual scenario. In general, histogram extracts features based on object distribution in the visual field. Thus, the use of histogram representation might result in the loss of spatial information [6]. Several studies (see, [11, 12]) used optical flow representation for object tracking. In principle, optic flow represents displacement vectors of each pixel in the image. This representation is commonly used to capture the spatio-temporal motion information of an object in the visual scene [11]. PT cameras are widely used as a part of surveillance systems. PT camera is mounted and adjusted to certain viewing configurations. For an effective application

of object tracking, accurate calibration is required [13]. A number of authors (see, e.g., [14–17]) estimated parameters of a calibration model for PT cameras by computing both focal length and radial distortion at many different zoom settings. L. Diogo in [18], on the other hand, introduced a method for PTZ camera calibration over the camera's zoom range by minimizing re-projection errors of feature points detected in images which are captured by the camera at different orientations and zoom levels. Recently, [19] presented a method to calibrate the rotation angles of a PT camera by using only one control point and assuming that the intrinsic parameters and position of the camera are known in advance.

In the field of object recognition based on human visual system, group of studies (see [20–24]) demonstrate the importance of motion information for object recognition. The studies showed that the visual perception can detect and focus on an object based on motion features. In essence, when an object is moving across the visual field, our eyes follow the object movement. The use of PT cameras in autonomous surveillance systems is still a challenge in spite of their great potential for real applications. There is still a need for an easy and accurate technique to estimate the rotation angles of a PT system. In this paper we suggest a new model for estimating PT rotation angles. We used a reverse projection from an image space to a world space in order to calculate the PT angles. In keeping with the importance of the motion information, our model takes advantage of optic flow technique by re-projecting the flow vectors from the image space into world space. Here, PT rotation angles are estimated based on Pinhole camera model and the rigid transformation of flow vectors in which an idealized translation matrix followed by two rotational matrices around PT axis are used. Our model was evaluated using rigid motions of different objects that have different shapes such as circular, square and silhouette objects. Such objects were moved with various directions to show the robustness of the model to track and focus on the object. The results reveal that when the object moves out of the region of interest, the reverse projection of the motion vectors from the camera plane to the real-world plane can be transformed to two values that represent the required rotational angles along pan and tilt axes to keep the object within a certain view. The paper is organized along the following structure and content. Section 2, demonstrates the methodology of the tracking model using PT system. The methodology has three main stages:(i) motion detection, (ii) object tracking, and (iii) PT rotational angles calculation. Section 3, shows system set up and results. Section 4, summarizes the discussion and the proposed perspectives of this work.

2. METHODOLOGY

2.1. Re-projection camera model

Our model proposes that the camera is rotated around Y_c and X_c axes as shown in Figure 1. These axes demonstrates the coordinate system of the camera model. From the previous figure, (p_w) represents the position of a 3D point in the world plane $\{w\}$ and (p_c) represents its projection in the camera plane $\{c\}$. Assume that (p_w) has moved to ($\bar{p_w}$) in the $\{w\}$ plane and ($\bar{p_c}$) defined its projection in the $\{c\}$ plane.
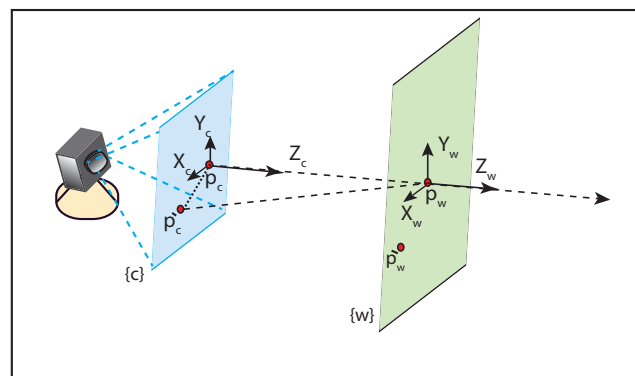

Figure 1. Point projection from real-world plane $\{w\}$ to camera plane $\{c\}$. (p_w) shows a 3D point in the world plane $\{w\}$ and (p_c) represents the point projection in the camera plane $\{c\}$. ($\bar{p_w}$) represents the next location of the point in the $\{w\}$ plane and ($\bar{p_c}$) defined its projection in the $\{c\}$ plane

We used Pinhole camera model and rigid transformation of parallelogram in the $\{w\}$ plane. The ideal Pinhole camera model transcribes the relationship between a point in 3D space $\{w\}$ and its projection in 2D space $\{c\}$ (for underlying principle of Pinhole camera model see [25]). To simplify further derivation, we define this relationship using homogeneous coordinate:

$$\mathbf{P_c} = \mathbf{M}.\mathbf{P_w} \tag{1}$$

where $\mathbf{P_c}$ $\begin{bmatrix} x_c & y_c & f & 1 \end{bmatrix}^T$ represents the projection of the 3D point $\mathbf{P_w}$ $\begin{bmatrix} x_w & y_w & z_w & 1 \end{bmatrix}^T$ on the image plane $\{c\}$ in which f denotes focal length. $\mathbf{M}$ represents a tomography matrix operation which defines two types of parameters: intrinsic and extrinsic parameters. Since most cameras have overcame the distortion and skew problems, our model focuses on the extrinsic parameters. Given that the coordinate systems of the camera reference frame is simply set at the object centroid in the world reference frame, the tomography matrix is then

$$\mathbf{M} = \mathbf{T_{P_w}^{P_c}} + \mathbf{P_w}. \tag{2}$$

where $\mathbf{T_{P_w}^{P_c}} = \begin{bmatrix} t_x & t_y & t_z & 1 \end{bmatrix}$ represents the translation vector from the center of the world space $\{w\}$ to the center of the image plane $\{c\}$. In order to keep the camera model tracking an object moves away from the center of the plane (i.e. the object moves from $\mathbf{p_w}$ to $\mathbf{\acute{p}_w}$, a rotation operation $\mathbf{R}$ is incorporated. The rotation matrix $\mathbf{R}$ is here addresses the rotation around pan-axis, $\mathbf{R_{pan}}(\alpha)$ and tilt-axis, $\mathbf{R_{tilt}}(\beta)$. Thus, the tomography matrix is defined as a sequence of matrices operations that include translation $\mathbf{T}$ and rotation $\mathbf{R}$. The tomography matrix is then define by the following equation:

$$\mathbf{M} = \mathbf{T_{P_w}^{P_c}} + \mathbf{P_w}\mathbf{R}. \tag{3}$$

$$\mathbf{R} = \mathbf{R_{pan}}(\alpha)\mathbf{R_{tilt}}(\beta) \tag{4}$$

$$\mathbf{R_{pan}}(\alpha) = \begin{bmatrix} cos(\alpha) & 0 & sin(\alpha) \\ 0 & 1 & 0 \\ -sin(\alpha) & 0 & cos\alpha \end{bmatrix} \tag{5}$$

$$\mathbf{R_{tilt}}(\beta) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & cos(\beta) & -sin(\beta) \\ 0 & sin(\beta) & cos\beta \end{bmatrix} \tag{6}$$

As a consequence, the object tracking model can be written as an idealized translation matrix followed by two rotational matrices around pan and tilt axes. The camera tracking model is defined by

$$\mathbf{\acute{P}_c} = \mathbf{T_{P_w}^{P_c}} + \mathbf{R_{pan}}(\alpha).\mathbf{R_{tilt}}(\beta).\mathbf{P_w}. \tag{7}$$

where $\mathbf{\acute{P}_c} = (\acute{x_c} \ \acute{y_c} \ f \ 1)$ represents the homogeneous coordinates of the moving object.

2.2. Motion detection

In order to detect the motion of an object in the visual scene, an optical flow algorithm is used. The optical flow depicts the motion of an object in the visual scene by two dimensional velocity vectors. These two vectors describe the velocity of pixels in a time sequence of two consequent images [26]. For the sake of simplicity, the 3D object in real time world plan is transfer into 2D object in the image plane in which the image can be defined by means of 2D dynamic brightness function $I(x, y, t)$. Given that the brightness between two consecutive image frames is constant within small space-time step (brightness constancy constraint) [27], the brightness constancy equation can be defined by

$$I(x, y, t) = I(x + \Delta x, y + \Delta y, t + \Delta t) \tag{8}$$

where $(\Delta x, \Delta y)$ represents the motion displacement for an object in x, and y directions, respectively, within short time (Δt). Given the assumption of small object motion, the brightness constancy equation 8 can be developed using Taylor series to get $I(x + \Delta x, y + \Delta y, t + \Delta t) = I(x, y, t) + \frac{dI}{dx}\Delta x + \frac{dI}{dy}\Delta y + \frac{dI}{dt}\Delta t + H.O.T.$ By neglecting higher order (HOT) term and rearranging the series [26], we get

$$\nabla I.\mathbf{v} = -I_t \tag{9}$$

where (ΔI) represents the spatial gradient of brightness intensity and $\mathbf{v}$ denotes the optical flow vectors (v_x, v_y) and I_t defines the derivative of the image intensity along time (t). Since equation 9 has two unknown parameters (aperture problem), another set of equations is needed. Several studies have been conducted to solve this equation (see for examples [27–29]). In this article, we solved equation 2 and estimated the optic flow following to the work in [30].

2.3. Calculate pan and tilt rotational angles

The rotation angle increments, $\delta\alpha$ and $\delta\beta$, are calculated relative to the current camera coordinate frame. These increments keep the object within a region of interest (ROI) in the image frame $\{c\}$. In order to account the effect of different units (points in image plane are expressed in pixels, while points in word space are represented in physical measurements (centimeters)), we introduce scale parameters k and l with unit of $\frac{cm}{pixel}$ where these parameters correspond to the change of units along x and y axis respectively. Thus, we adjust equation 7 by adding the scale parameters $\{k, l\} \in s$, as defined by

$$s.\mathbf{\acute{P}_c} = \mathbf{T^{pc}_{pw}} + \mathbf{R_{pan}}(\alpha).\mathbf{R_{tilt}}(\beta).\mathbf{P_w}. \tag{10}$$

With the knowledge of the object movement in the image plane, it is possible to track the object in the word space by back-projecting the 2D point in the image plane to a 3D point in the world plane. Here, we calculate the increments in two parts, the pan angle increment $\delta\alpha \in [0^o, 180^o]$ and then tilt angle increment $\delta\beta \in [-90^o, 90^o]$, where they are derived from equation 10. For simplicity, we utilize camera projection with plane world z=0 rather than a general 3D case. The rotational angles is defined by

$$\alpha = cos^-1(k.\acute{x_c} - t_x)/x_w \tag{11}$$

$$\beta = cos^-1(l.\acute{y_c} - t_y)/y_w \tag{12}$$

3. SYSTEM SET UP AND RESULTS

In the following, we demonstrate the capability of the suggested model for object tracking by using a set of test motions. The camera was mounted on a pan-tilt system in which camera surface is aligned parallel and centred to the object. The block diagram of the suggeted model can be shown in Figure 2. The system was placed 27 cm away from the moving object. The object motions were recorded using webcam camera with a resolution of (480×640) in which real world view $\{w\}$ is about (16.5×12) cm. Due to the low resolution of the camera, it is challenging to estimate object motion in the visual scene. As mention in sec.(2.3.), the rotational range of the motor along pan is $[0^0, 180^0]$ and $[-90^0, 90^0]$ along tilt axes.
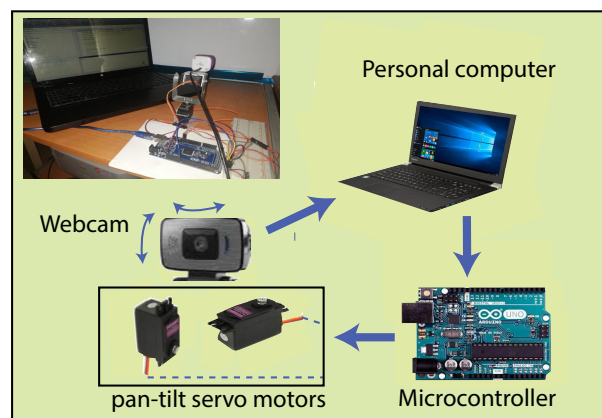

Figure 2. The block diagram of suggested model. The microcontroller (arduino MEGA) controls two servo motors (S3300). The servo motors drive the camera motion along pan-tilt axes

The tracking model keeps the object within the camera fovea which represents the region of interest ROI. Here, we select a window with (252, 233) that is placed in the center of the camera view. In order to keep the target object within ROI, the camera is moved along pan-tilt axes via the PT system. In our model, the pan-tilt angles are calculated based on equations 11 and 12 respectively. Since $x_w = t_z sin(\alpha)$ and $\acute{x}_c = u$, equation 11 can be defined by $\alpha = tan^{-1}(ku/t_z)$. Equation 12, on the other hand, can be written as $\beta = tan^{-1}(lv/t_z)$ due to the fact that $y_w = t_z sin(\beta)$ and $\acute{y}_c = v$. Based on camera resolution and view dimensions along x and y directions, the values of the scale parameters k and l are 16.5(cm)/480(pixel) and 12(cm)/480(pixel), respectively. In order to calculate the optic flow vectors (u, v), we utilized the algorithm that introduced by [30].

Our model was tested and evaluated using rigid motions of different directions. Such motions are generated from different shapes of objects that moved with different directions. With no loss of generality, we used simplified objects to keep the computational complexity and the preprocessing as simple as possible. Figure 3 shows moving objects (circle, square, tiger, bird) with four directions, down, up, right and diagonal (up-left), respectively. As a consequence to these motions, the objects become outside the ROI as shown in the second column of Figure 3. The third column of the figure shows the camera view $\{c\}$ after camera moving along pan and tilt axes. As an extended results, our model was tested with silhouette motion for a human that moves to the left direction, see Figure 4.The results reveal that when the object moves out of the ROI, the reverse projection of the motion vectors from the camera plane to the real world plane can be transformed to two values that represent the required rotational angles along pan and tilt axes to keep the object within ROI.
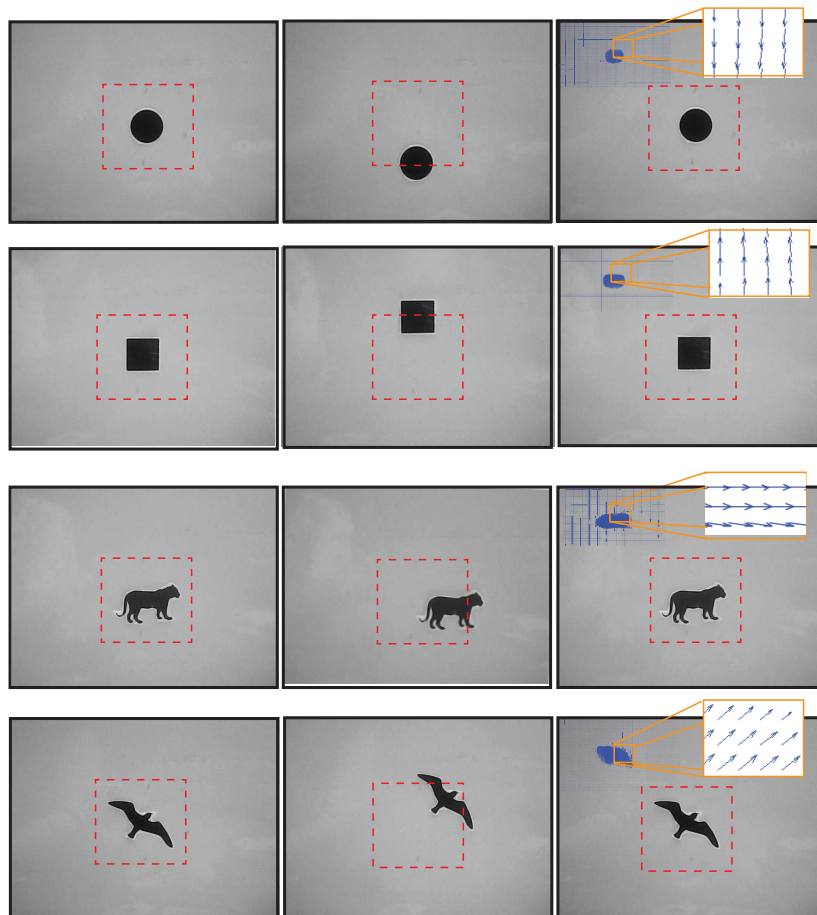

Figure 3. The first column, shows the target object in which the camera fovea is surrounded by dashed line. The second column, demonstrates the object movement in a particular direction, while the third column, shows the optic flow of the object motion and the camera view after the pan-tilt rotation
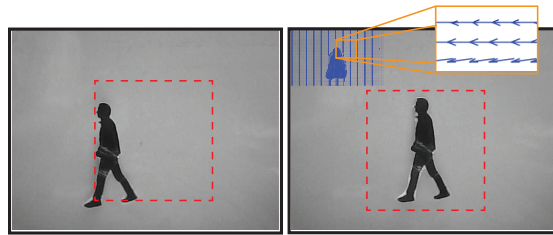
Figure 4. The first column shows the object moves out of the ROI demonstrates the man movement to the left
direction. Second column shows the camera view after the pan-tilt rotation

4.   DISCUSSION

In this paper, we introduce a model for object tracking using motion vectors and PT system. The model utilizes the pinhole camera algorithm and rigid transformation from camera plane to the world plane. Our tracking model keeps the target object within the fovea of the camera view, i.e., ROI. Here, we used pan-tilt system that controls the rotation of the camera view. The angles of the pan-tilt axes are derived from the reverse projection of the optic flow.

In order to account the different units (pixels in the camera plane and centimeter in real world plane) we incorporate a scale parameter k and l in the mathematical expression of α and β along pan and tilt axes, respectively. Optical flow technique is found to be more robust and efficient for object tracking due to higher detection accuracy (see a recent review in [31]). This prompted several researchers (see for e.g., [12, 32]) to use optic flow for object tracking. Our model differs from others approaches in term of using reverse projection from the camera plane $\{c\}$ to the world plane $\{w\}$ for estimating the pan-tilt rotational angles. These estimated angles are fed to two motors that control camera moving along PT axes. In our model, the camera keeps the target object within the fovea of the camera view which we called it region of interest (ROI). Here, we utilized the the rigid transformation of flow vectors in which an idealized translation matrix followed by two rotational matrices around PT axis are used.

To evaluate the performance of our model, we performed a set of rigid motions using different objects. These motions are generated from different shapes of objects in which circular and square objects are used. Such objects are moved to upward and downward directions, respectively. In addition, our model was probed with silhouette objects which are bird, tiger and human that move in diagonal,right and left directions, respectively. The results reveal that our model can keep tracking the object within a certain view.

Our model can be extended and improved by adding additional preprocessing techniques that enrich the model to deal with more sophisticated biological motions. The biological motions represent a complex and articulated motions which consist different speeds and directions. The model could be boosted with an algorithm that segregates the body motion from the limbs motion (see [33] for a discussion underling the limbs and body motions). As a consequence the tracking model could be depended on the body motion without taking into account the movement of the limbs.

5.   CONCLUSION

We introduce a new model of PT configuration to control the movement of the surveillance camera. The PT angles of the camera movement are estimated using Pinhole camera model and the projection of the motion vectors in which these vectors are transformed from the image space into a world space. The motion vectors are estimated using the optic flow technique in which these vectors are transformed from the camera plane into the real-world plane using an idealized translation matrix followed by two rotational matrices around PT axis. Our model was tested and evaluated using different objects that have different shapes (such as circular, square and silhouette objects) where these objects are moved in different directions. This work has shown that when an object moves out of the region of interest, the projection of the motion vectors can transcribe into two values. These values are transform into PT angles that control the movement of the surveillance camera. Building upon, The transformation of the motion vectors enables the surveillance camera moving relative to the movement of the target object. As a consequence, the camera preserves the target object within the fovea of the camera view.

REFERENCES

[1] V. Ramalakshmi, Kanthimathi, and M. G. Alex, "A survey on visual object tracking: Datasets, methods and metrics," *International Research Journal of Engineering and Technology (IRJET)*, vol. 11, no. 3, pp. 2395–0056, 2016.

[2] I. Haritaoglu and M. Flickner, "Detection and tracking of shopping groups in stores," *IEEE International Conference on Computer Vision and Pattern Recognition*, vol. 1, pp. 431–438, 2001.

[3] S. A. Mohand, N. Bouguila, and D. Ziou, "A robust video foreground segmentation by using generalized gaussian mixture modeling," *Fourth Canadian Conference on Computer and Robot Vision (CRV '07)*, pp. 503-509, 2007.

[4] A. Yilmaz, O. Javed, and M. Shah, "Object tracking: A survey," *ACM Computing Surveys*, vol. 38, no. 4, pp. 1–45, 2006.

[5] S. R. Balaji and S. Karthikeyan, "A survey on moving object tracking using image processing," *11 th International Conference on Intelligent Systems and Control (ISCO) At: Coimbatore, Tamilnadu, India*, pp. 469-474, 2017.

[6] L. Xi, H. Weiming, S. Chunhua, Z. Zhongfei, D. Anthony, and V. Anton, "A survey of appearance models in visual object tracking," *ACM Transactions on Intelligent Systems and Technology*, vol. 4, no. 4, pp. 1-48, 2013.

[7] H. Wang, D. Suter, K. Schindler, and C. Shen, "Adaptive object tracking based on an effective appearance filter," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 29, no. 9, pp. 1661–1667, 2007.

[8] M. S. Allili and D. Ziou, "Object of interest segmentation and tracking by using feature selection and active contours," *IEEE International Conference on Computer Vision and Pattern Recognition*, pp. 1–8, 2007.

[9] Q. Zhao, Z. Yang, and H. Tao, "Differential earth mover's distance with its applications to visual trackingr," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 32, no. 2, p. 274–287, 2010.

[10] D. Comaniciu, V. Ramesh, and P. Meer, "Kernel-based object tracking," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 25, no. 5, pp. 564–577, 2003.

[11] Y. Ma and C. Cluster, "An object tracking algorithm based on optical flow and temporal–spatial context," *Springer Science+Business Media, LLC, part of Springer Nature*, pp. 1-9, 2017.

[12] K. Kale, S. Pawar, S. Pawar, and P. Dhulekar, "Moving object tracking using optical flow and motion vector estimation," *2015 4th International Conference on Reliability, Infocom Technologies and Optimization (ICRITO) (Trends and Future Directions)*, pp. 1–8, 2007.

[13] N. S. Sinha and M. Pollefeys, "Pan-tilt–zoom camera calibration and high-resolution mosaic generation," *Computer Vision and Image Understanding*, vol. 103, no. 3, pp. 170–183, 2006.

[14] R. Collins and Y. Tsin, "Calibration of an outdoor active camera system," *In Proceedings of IEEE Computer Society, CVPR99*, vol. 1, pp. 528-534, 1999.

[15] R. G. Willson, "Modeling and calibration of automated zoom lenses," *Carnegie Mellon University, PhD thesis*, 1994.

[16] K. Pankaj and D. Anthony, "Real time target tracking with pan tilt zoom camera," *IEEE International Conference on Digital Image Computing: Techniques and Applications, Melbourne, VIC, Australia*, pp. 492-497, 2009.

[17] V. Diogo, C. N. Jacinto, and G. Jose, "Assessing control modalities designed for pan-tilt surveillance cameras," *15th portuguese Conference on pattern recognition*, 2009.

[18] L. Diogo, "Target tracking with pan-tilt-zoom cameras," *M.Sc. Thesis, Instituto Superior Technico*, 2011.

[19] L. Yunting, Z. Jun, W. Hu, and J. Tian, "Method for pan-tilt camera calibrationusing single control point," *Journal of the Optical Society of America*, vol. 32, no. 1, pp. 156–163, 2015.

[20] A. Casile and M. A. Giese, "Critical features for the recognition of biological motion," *Journal of Vision*, vol. 5, no. 4, pp. 6-6, 2005.

[21] L. I. Abdul-Kreem and H. Neumann, "Estimating visual motion using an event-based artificial retina," *Series of CCIS Communications in Computer and Information Science. Computer Vision, Imaging and Computer Graphics Theory and Applications. Springer International Publishing Switzerland*, pp. 396-415, 2016.

[22] L. I. Abdul Kreem and H. Neumann, "Neural mechanisms of cortical motion computation based on a neuromorphics sensory system," *PLOS ONE*, vol. 10, no. 11, pp. 1-33, 2015.

[23] L. I. Abdul-Kreem and H. Neumann, "Bioinspired model for motion estimation using address event

representation," *10th International Conference on computer vision theory and application, VISIGRAPP*, pp. 335-346, 2015.

[24] L. I. Abdul-Kreem, "Computational architecture of a visual model for biological motions segregation," *Network: Computation in Neural Systems*, vol. 30, no. 1-4, pp. 58-78, 2019.

[25] P. Sturm, S. Ramalingam, J. Tardif, S. Gasparini, and J. Barreto, "Camera models and fundamental concepts used in geometric computer vision," *Found Trends Comput Graph Vis*, vol. 6, no. 1-2, pp.1-183, 2011.

[26] S. Aslani and H. Mahdavi-Nasab, "Optical flow based moving object detection and tracking for traffic surveillance," *International Journal of Electrical and Computer Engineering*, vol. 7, no. 9, pp.1252-1256, 2013.

[27] B. Horn and B. Schunck, "Determining optical flow," *Artif. Intell.*, vol. 17, no. 1-3, pp. 185-203, 1981.

[28] J.Barron, D.Fleet, andS.S.Beauchemin, "Performance of optical flow techniques," *Int. J. Comput. Vision*, vol. 12, no. 1, pp. 43-77, 1994.

[29] A. Bruhn, J. Weickert, and C. Sschnorr, "Lucas/kanade meets horn/schunck: combining local and global optic flow methods," *Int. J. Comput. Vision*, vol. 61, no. 3, pp. 211-231, 2005.

[30] T. Brox, A. Bruhn, N. Papenberg, and J. Weickert, "High accuracy optical flow estimation based on a theory for warping," *Proc. European Conf. Computer Vision*, pp. 25-36, 2004.

[31] A. Agarwal, S. Gupta, D. Kumar, and D. K. Singh, "Review of optical flow technique for moving object detection," *Conference: 2016 2nd International Conference on Contemporary Computing and Informatics (IC3I)*, pp. 409-413, 2016.

[32] D. Pathak, R. Girshick, P. Dollár, T. Darrell, and B. Hariharan, "Learning features by watching objects move," *IEEE Transactions on Industrial Electronics*, pp. 2701-2710, 2017.

[33] L. I. Abdul-Kreem, "Human action recognition based on motion velocity," *Proc. SCEE2018; 3rd ieee scientific conference of electrical engineering Baghdad/ Iraq*, pp. 45-50, 2018.