

Trackerbot: A Robotic Surveillance System based on Stereo-Vision and Artificial Neural Networks

Hooman Khosravi, Ehsan Fazl-Ersi

Department of Computer Engineering

Ferdowsi University of Mashhad

Mashhad, Iran

hooman.khosravi@stu-mail.um.ac.ir, fazlersi@ferdowsi.um.ac.ir

Abstract— Master-slave camera systems are ideal for detailed surveillance of desired targets in wide scenes. In this paper, we propose a fully automatic camera system that is structurally unconstrained and independent from both intrinsic and extrinsic camera parameters for detecting activities in indoor or outdoor environments. Unlike traditional models in our system, two wide cameras are used to reach the minimum error in estimating the pan, tilt and zoom parameters (PTZ) which then would be only affected by the resolution of the wide cameras or the PTZ motor system. After an initial automatic calibration, a feed-forward artificial neural network (ANN) takes charge of controlling the PTZ unit according to the information extracted from the frames of the wide cameras.

Keywords— Intelligent surveillance; Master-slave camera system; camera control; PTZ parameters

I. INTRODUCTION

Holistic observation of the scene is needed for surveillance applications to control public or private places for detecting and tracking potential abnormal activities.

Such systems have to capture high resolution images from the point of interest, but also have to capture wide images to control the whole scene and detect points of interest. The wide angle cameras have a wide field of view that enables them to monitor all of the desired location. Their low resolution images are also ideal for real-time computation of the scene. However, due to their limited resolution, they cannot offer a detailed image. On the other hand, the pan, tilt and zoom camera, briefly called PTZ, solves this issue by adding a zoom lens. Hence, it will capture detailed images in a very narrow field of view. Therefore, it has to mechanically move in its workspace in order to observe a specific location in the desired scene. At any single time, it can only observe a part of the environment, so the system has to intelligently choose the right target. There are lots of research works on the combination of PTZ and the wide camera [1-4].

Shakeri and Zhang in [5] proposed a method based on dual camera system in which a wide camera and a PTZ is used. They have mathematically modeled camera parameters explicitly and it is possible to use their method in real-time. However, in the illustrated system, each camera calibrates manually to calculate its own intrinsic parameters followed by a stereo calibration. Hu et al. [6] introduced a master-slave

dual camera (MSDC) system which tracks the specified target. In their system, the specific target in the wide or master camera, detects and tracks using codebook [7] and Tracking-Learning-Detection (TLD) algorithms. Moreover, to calibrate cameras and control the PTZ, they have used Look Up Table (LUT) [8] method, and so, it gets possible to calculate pose parameters of PTZ camera. Chen et al. [9] used background subtraction based on Gaussian mixture model (GMM) to detect moving objects on the wide camera. Moreover, to track objects in the scene, they have used Kalman Filter. The performance of their method is limited to the structure of the place that the system is running and they cannot track fast moving objects. Neves et al. [10] introduced a method which detects the 2D position of a human head from the wide camera and then tries to track it, also, calculates the projection matrix between cameras coordinates. Li et al. [11], also proposed a structural constraint based dual camera cooperation model to calculate the PTZ parameters. In their method, cameras are fixed by structural constraints which lets them to calibrate once. Bastanir [12] calibrated cameras by assuming that the optical axis of Cameras are perpendicular to the ground and the PTZ camera is fixed to a reference; and so, he calibrated cameras by two scene points. Liao and Chen [17] introduced a system that contains two PTZ cameras. One of the PTZ units is always in a wide mode while the other one zooms on the specific target. However, using a PTZ camera in the role of the system's wide camera is financially impractical. A number of recent systems [6, 12, 17] use high resolution PTZ camera image during runtime to compute the next pan and tilt values. However, visual computation on these high resolution images are time consuming. Since numerous PTZ units exist in surveillance areas a high performance control unit is required.

As is shown in Fig. 1 many of the previous works use a pre-calculated estimator based on prior knowledge. As we can

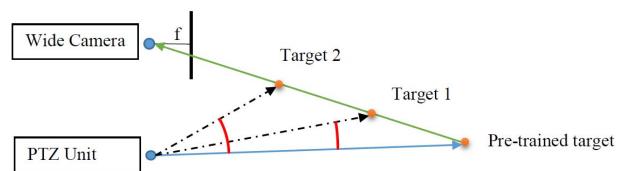


Fig. 1. Single wide camera system is calibrated at a specific distance. Pan values error at other depths are shown in red arcs.

see in Fig 1, the single wide camera, was trained at a particular distant. As a result, PTZ camera is pointing directly at the desired target. But, as the target gets closer to the camera, it can be clearly seen that, PTZ starts to drift from the target.

The rest of this paper is organized as follows: first, an overview of the proposed system is given and then an offline, accurate, automatic and unconstraint calibration process is introduced based on machine learning approaches. Finally, at the experiments section, performance and accuracy of the proposed method is tested in real and simulated scenarios.

II. SYSTEM OVERVIEW

Structural properties and procedures of the system are described in this section. As shown in Fig. 2, our system consist of a central computer-controlled PTZ camera which is capable of capturing detailed images of the objects in its field of view. The PTZ camera is made by two servo motors for pan and tilt, forming two revolute joints and a camera as the end effector. The two servo motors are connected to a Pololu servo driver; this driver accepts position commands through serial communication. Two wide cameras (fitted with “super wide 0.4x” lenses) are placed beside the PTZ unit. The first wide camera, is placed above the main camera and the other is located at the left side of it. This structure is optimal, however, any camera placement can be chosen for the proposed



Fig. 2. The proposed camera system consist of two wide cameras and a central PTZ unit.

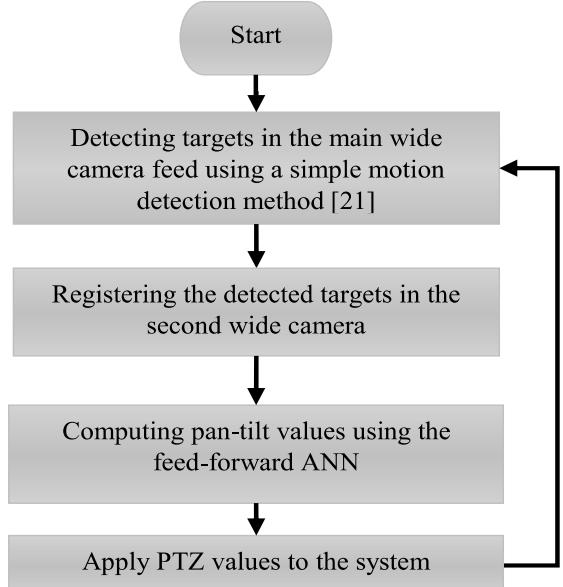


Fig. 3. Operation process of the proposed system.

calibration method since there are two wide cameras. Our system uses the wide angle cameras feeds as its input, in another word it has a Hand to eye configuration.

At the beginning, the system should be calibrated. The PTZ unit automatically moves around in its workspace and takes images from the wide cameras and the PTZ unit, and calibrates itself using machine learning approaches. After calibration, the system starts to operate as described in Fig. 3.

III. CALIBRATION

Stereo-wide PTZ camera systems can suffer from inaccuracy if there are issues in parts or the structure of the system. Hence, in the proposed method the system will calibrate itself after it has been set up. In order to establish correspondence between the position information of the target in wide cameras images and the pan-tilt values a feed-forward ANN is used (Fig. 4). A feed-forward ANN with just one hidden layer is proved to be a global approximation method [13].

Fig. 7 shows the calibration process of the system. The PTZ unit, will move within its' motors workspace in a regular static pattern with several pre-defined stops in the trajectory (Fig. 6) while maintaining a fixed zooming value. At each specific pan and tilt value, controller will capture images from the PTZ camera and both of the wide cameras. As illustrated in Fig. 5, Surf features (Speeded up robust features) [14] will be

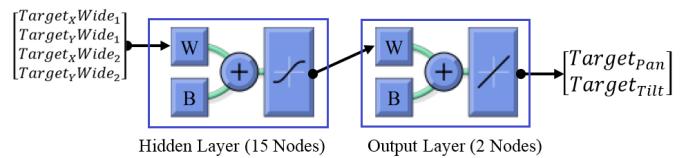


Fig. 4. The feed-forward ANN with a single hidden layer takes the position information of the target in both wide cameras and computes the desired pan and tilt values.

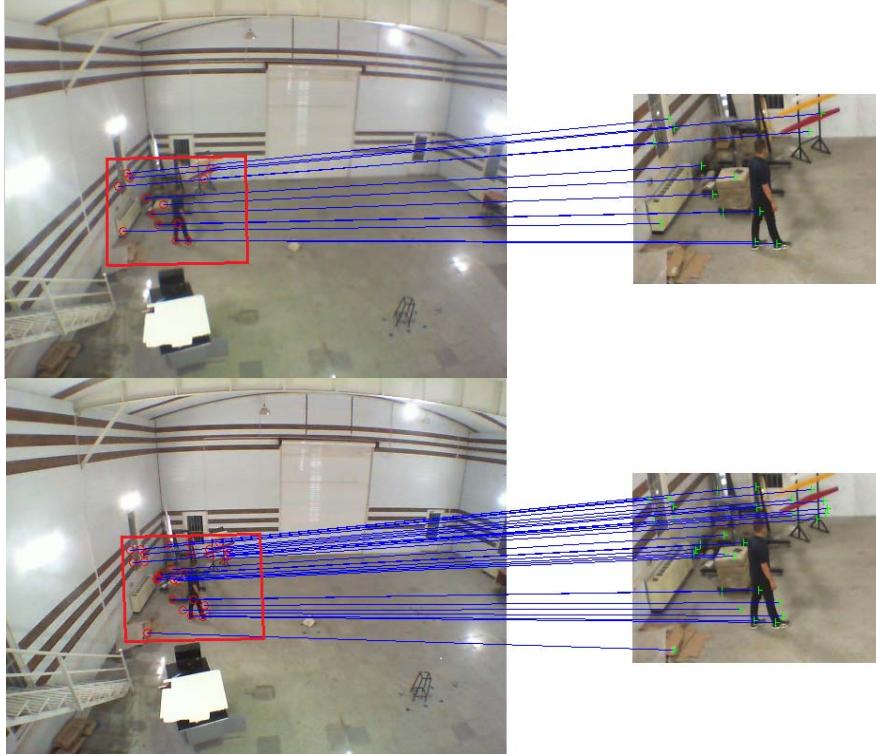


Fig. 5. A sample from the PTZ is registered in the both wide cameras using SURF features matching and affine geometric transformation.

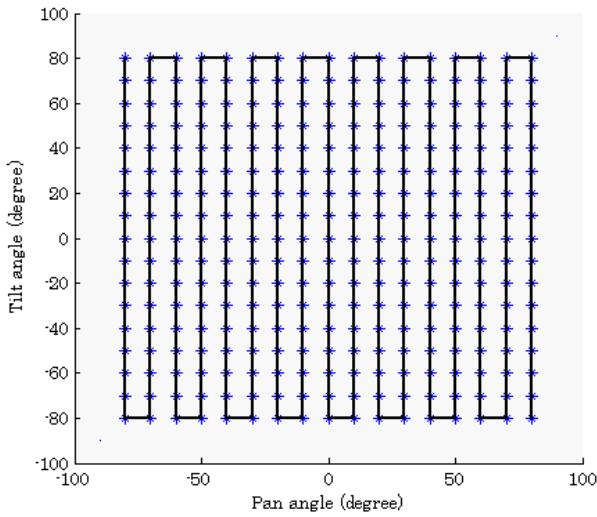


Fig. 6. PTZ unit will execute the above trajectory for the purpose of initial calibration. A constant zooming factor is used for the pattern. Denser patterns can be utilized for more complicated camera systems.

extracted from the images. After matching these features using nearest-neighbor ratio method [18], the outliers will be removed using affine geometric transformation [15]. Using the computed affine transformation, position and the zoom factor of the target in the wide cameras images will be computed [16]. When all of the samples are collected, the feed forward neural network trains using the x, y position of the wide

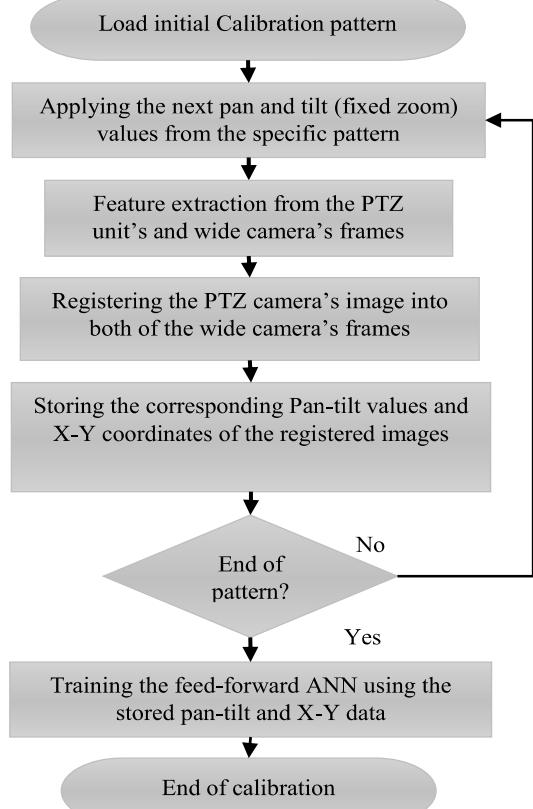


Fig. 7. Calibration process of the proposed system.

cameras and correspondent parameters to set the view point of the PTZ camera on the specific target detected on the wide cameras. Also zoom parameter will be computed according to size of the target in the wide camera's field of view and its focal length.

IV. EXPERIMENTS

A. Real scenarios

The system is placed in an outdoor situation and a person, as a moving object, moves from one point to another. The system is calibrated and running according to the procedures in Fig. 3. and Fig. 7. As can be seen in Fig. 8, the target is detected by a simple motion detector [21] on the primary wide camera. Afterward, the selected target is registered in the

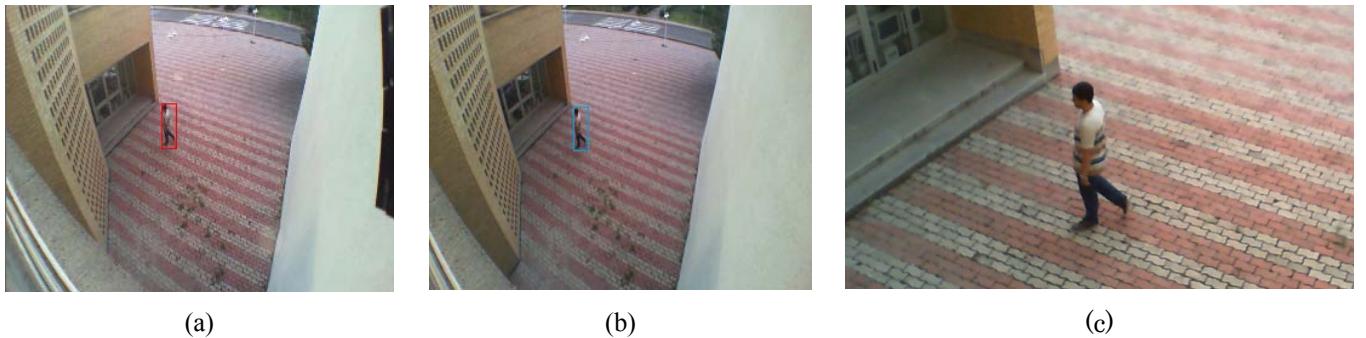


Fig. 8. An outdoor experiment with a moving target.(a) Target is located in the primary wide camera's image. (b) The selected target registers in the second wide camera image using feature based template matching. (c) shows the result of the outdoor experiment, PTZ is pointed at the target with 100% overhead.



Fig. 9. A predefined target is located in the main wide camera (b). Afterwards, the target is registered in the second wide camera using the matched SURF feature and geometric transformation estimation (a).



Fig. 10. Result of the proposed method (a) in comparison with the single camera system (b).

second wide camera as well. Pan-Tilt parameters are calculated by giving the X-Y coordinates of the target in both wide cameras to the trained ANN and then are sent to the PTZ system via serial communication. As Fig 8 (c) shows the PTZ unit is pointing at the target correctly. For a more broad view, 100% overhead is chosen for small targets.

In another real world experiment, using the presented system in Fig. 2 we compared the proposed method with the single wide camera system. Fig 9 (b) shows a target in the FOV of the main wide camera which is located on the left side of the PTZ unit. The single wide camera method only uses the target's position in the main camera, however, our method find the target in the other wide camera as well (Fig.9 (a)). As shown in Fig 10(b), since the mono wide system was pre-trained with samples at the end of room it was not able to calculate the correct pan and tilt values. In contrast, the proposed method using the position information of the target from both of the wide cameras pointed the PTZ camera to the target correctly (Fig. 10 (a)).

B. Simulation

To further validate the proposed calibration method, the suggested camera system is digitally simulated and to represent targets in real environments, uniformly random samples are generated for training and testing the method. Fig 11, shows the simulated setup, as can be seen, the two wide cameras are modeled using the pinhole camera geometry model and consist of an image plane and a focal point. However, since we do not use the image feed of the PTZ camera and our only objective is to aim it at the correct direction, we modeled the PTZ unit using a three dimensional vector.

In the training phase a generated sample will form a line with both of the wide cameras' focal points. Interception of these lines and the image planes are calculated. In other word, the exact pixels that the target will appear in the wide cameras images are computed. This information along the relative angle between the target and the PTZ unit is calculated for each sample, afterwards, the feed forward ANN (Fig. 4) is trained using these data.

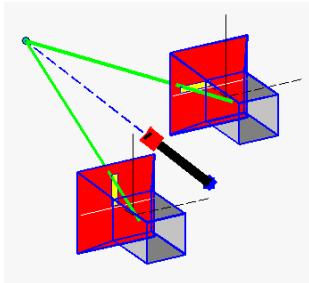


Fig. 11. The simulated camera system and a generated sample. The sample(blue dot) forms two lines with the wide cameras' focal points(green lines).The interception point of these lines and the image planes (red planes) are calculated (yellow rectangles show the X and Y values of the interception points). Also, the PTZ unit is represented by the black vector and is pointing at the sample.

In the testing phase, a server will generate random samples in common field of view of the both wide cameras, then the server will calculate the position of the target in both of the wide cameras' images (pixel information). Using this information, the suggested method will calculate the estimated pan and tilt values using the feed forward ANN, finally, the output of the ANN will be compared with the actual needed pan and tilt values to measure the error of the method.

To test the performance of the proposed method, it is compared with four other methods in terms of absolute and relative error of pan and tilt values in different depths of targets. Relative errors is calculated by dividing the absolute error by the magnitude of the measured value. Samples were generated in distances from 6 up to 100 meters in the range of angles between -35 to 35 degrees relative to the camera systems.

Fig. 12 shows the pan estimation accuracy of different methods. As can be clearly seen, the proposed method outperforms the other systems by a long margin regarding the absolute and relative error of the system. In terms of absolute error, other methods converge to around 1 degrees, moreover, their performance in distances between 6 to 20 meters drops noticeably. However, our method reaches to one thousandths of a degree at the 100 meters and although its performance drops relatively in the range of 6 to 20 meters its accuracy is still about three thousandths of a degree at 6 meters. At this error magnitude the wide cameras should be at least over 600 mega pixel for the pan estimation error to be the bottle neck of the system.

Fig. 13 illustrates the tilt accuracy comparison of different methods. The proposed method absolute tilt error is about a hundred to a thousand times smaller than the other methods throughout its target's depth range. Since the height of the objects are limited at far distances in the FOV of the cameras, the magnitude of their tilt angle relative to the camera system gradually converges to zero as they get further away from the system. As a result, the relative error of the tilt angles diverges as the depth increases. However, as can be seen the proposed method maintains much less relative error in comparison to other methods.

V. CONCLUSION

The presented system, based on ANN, calibrates automatically in different situations without any structural constraint or need for intrinsic or extrinsic calibration of the cameras. As a moving object detects, having specific position of detected object, ANN computes the commands that should be send to the servo motors of the PTZ unit to find and track the object. Finally, examination results on real and simulated environments show that the proposed method works perfectly in different conditions.

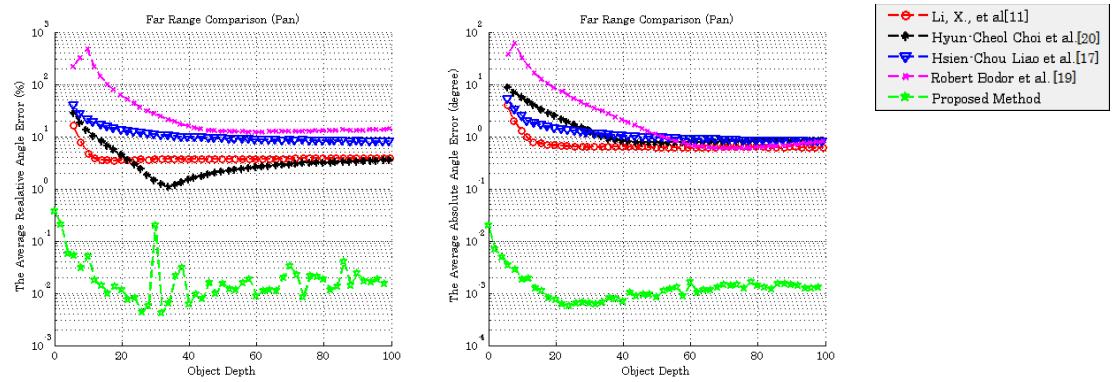


Fig. 12. Absolute and relative angle error of pan value in distances from 6 to 100 meters.

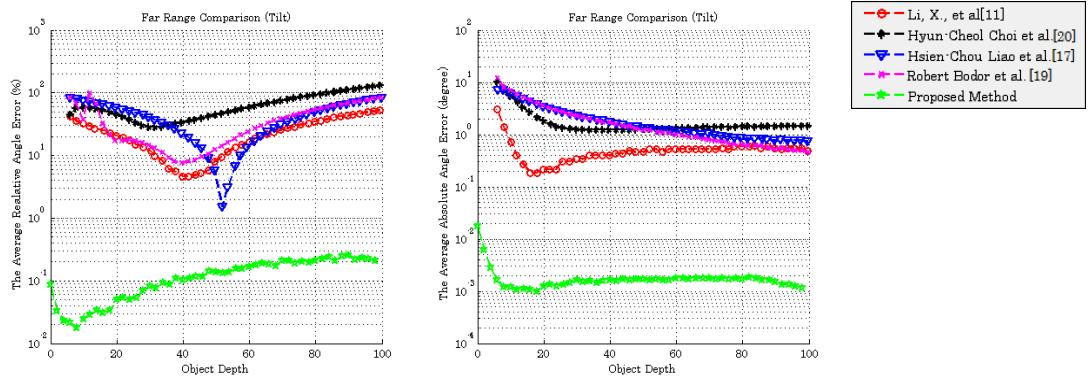


Fig. 13. Absolute and relative angle error of tilt value in distances from 6 to 100 meters

REFERENCES

- [1] Sato, K., et al. CAD-based object tracking with distributed monocular camera for security monitoring. in CAD-Based Vision Workshop, 1994., Proceedings of the 1994 Second. 1994. IEEE.
- [2] Meyer, M., M. Höller, and T. Ohmacht. New options in video surveillance applications using multiple views of a scene. in Security Technology, 1998. Proceedings., 32nd Annual 1998 International Carnahan Conference on. 1998. IEEE.
- [3] Choi, H.-C., U. Park, and A.K. Jain. Ptz camera assisted face acquisition, tracking & recognition. in Biometrics: Theory Applications and Systems (BTAS), 2010 Fourth IEEE International Conference on. 2010. IEEE.
- [4] Lain, B., S. Ghidoni, and E. Menegatti. Robust object tracking in a dual camera sensor. in Human System Interactions (HSI), 2011 4th International Conference on. 2011. IEEE.
- [5] Shakeri, M. and H. Zhang. Cooperative Targeting: Detection and Tracking of Small Objects with a Dual Camera System. in Field and Service Robotics. 2015. Springer.
- [6] Hu, J., S. Hu, and Z. Sun. A real time dual-camera surveillance system based on tracking-learning-detection algorithm. in Control and Decision Conference (CCDC), 2013 25th Chinese. 2013. IEEE.
- [7] Kim, K., et al., Real-time foreground-background segmentation using codebook model. Real-time imaging, 2005. 11(3): p. 172-185.
- [8] Senior, A.W., A. Hampapur, and M. Lu. Acquiring multi-scale images by pan-tilt-zoom control and automatic multi-camera calibration. in Application of Computer Vision, 2005. WACV/MOTIONS'05 Volume 1. Seventh IEEE Workshops on. 2005. IEEE.
- [9] Chen, P., et al. A Dual Camera System for Moving Objects Close-up Tracking. in Proceedings of the 2015 Chinese Intelligent Systems Conference. 2016. Springer.
- [10] Neves, J., et al., A calibration algorithm for multi-camera visual surveillance systems based on single-view metrology, in Pattern Recognition and Image Analysis. 2015, Springer. p. 552-559.
- [11] Li, X., et al., A Structural Constraint Based Dual Camera Model, in Pattern Recognition. 2014, Springer. p. 293-304.
- [12] Bastanlar, Y., A simplified two-view geometry based external calibration method for omnidirectional and PTZ camera pairs. Pattern Recognition Letters, 2016. 71: p. 1-7.
- [13] Hornik, K., Approximation capabilities of multilayer feedforward networks. Neural networks, 1991. 4(2): p. 251-257.
- [14] Bay, H., T. Tuytelaars, and L. Van Gool, Surf: Speeded up robust features, in Computer vision-ECCV 2006. 2006, Springer. p. 404-417.
- [15] Torr, P.H. and A. Zisserman, MLESAC: A new robust estimator with application to estimating image geometry. Computer Vision and Image Understanding, 2000. 78(1): p. 138-156.
- [16] Hartley, R. and A. Zisserman, Multiple view geometry in computer vision. 2003: Cambridge university press.
- [17] Liao, H.C., Chen, W.Y.: Eagle-Eye: A dual-PTZ-Camera system for target tracking in a large open area. Information Technology and Control 39(3), 227-235 (2010)
- [18] David Lowe, "Distinctive image features from scale-invariant keypoints," International Journal of Computer Vision, 60, 2 (2004) 23.
- [19] Bodor, R., Morlok, R., Papanikolopoulos, N.: Dual-camera system for multi-level activity recognition. Intelligent Robots and Systems (IROS) 1, 643-648 (2004)
- [20] Choi, H.C., Park, U., Jain, A.K., et al.: Face Tracking and Recognition at a Distance: A coaxial & concentric PTZ Camera System. IEEE TCSVT 8(10), 1665-1677 (2011)
- [21] Cutler, Ross, and Larry S. Davis. "Robust real-time periodic motion detection, analysis, and applications." IEEE Transactions on Pattern Analysis and Machine Intelligence 22.8 (2000): 781-796.