Cloud Computing and Data Science
https://ojs.wiserpub.com/index.php/CCDS/

UNIVERSAL WISER PUBLISHER

Research Article

A Comparative Analysis Using Silhouette Extraction Methods for Dynamic Objects in Monocular Vision

Md Rajib M Hasan[1*], **Noor H. S. Alani**[2]

[1]BICC-Cyber Security & IT Solution, Auckland, New Zealand
[2]School of Computing, Eastern Institute of Technology, Napier, New Zealand
E-mail: rajib@bicc.co.nz

Received: 8 October 2021; **Revised:** 29 November 2021; **Accepted:** 27 December 2021

Abstract: Moving or dynamic object analysis continues to be an increasingly active research field in computer vision, with many types of research investigating different methods for motion tracking, object recognition, pose estimation, or motion evaluation (e.g., in sports sciences). Many techniques are available to measure the forces and motion of people, such as force plates to measure ground reaction forces for jumping or running sports. In training and commercial solutions, the detailed motion of an athlete is captured using motion capture devices based on optical markers on the athlete's body and multiple calibrated fixed cameras around the sides of the capture volume. In some situations, it is not practical to attach any kind of marker or transducer to the athletes, or the existing machinery is being used, making it necessary to use a pure vision-based approach that relies on the natural appearance of the person or object. When a sporting event is taking place, there are opportunities for computer vision to help the referee and other personnel involved in the sports to keep track of incidents occurring, which may provide full coverage and detailed analysis of the event for sports viewers. The research aims at using computer vision methods, specially designed for monocular recording, for measuring sports activities, such as high jump, wide jump, or running. To indicate the complexity of the project: a single camera needs to understand the height at a particular distance using silhouette extraction. Moving object analysis benefits from silhouette extraction, and this has been applied to many domains, including sports activities. This paper comparatively discusses two significant techniques to extract silhouettes of a moving object (a jumping person) in monocular video data in different scenarios. The results show that the performance of silhouette extraction varies depending on the quality of the used video data.

***Keywords*:** moving objects, object tracking, visualisation, computing vision, silhouette detection

1. Introduction

The goal of computer vision is to understand the scenes (captured by a camera) in the real world. It offers implementation and design solutions for the methodological and algorithmic problems related to a specific topic [1]. At present computer vision plays a vital role in many fields related to technology [2]. The implementation of computer vision can be found in a wide range such as face recognition [3] login in a modern computer, in smartphones [4], vision-based driver assistance system [5], and audio-visual interaction with computer games [6]. Computer vision is continuously improving the quality and process control in many factories or industrial automation [7]. Computer vision

DOI: https://doi.org/10.37256/ccds.3120221201

contributes a lot in the film industry such as the creation of AVATAR and TRON (virtual worlds derived from image data, enhanced of historic video data) [1]. This is just mentioning a few application areas, which all come from various sources of recorded image or video data. In many situations, these data may need to be processed or analyzed [1].

Vision plays an important role in balance and postural control in human movement, where proprioception and vestibular function are involved. Monocular vision is seeing with only one eye at a time (see Figure 1). When both eyes are used, this would be binocular vision (to perceive 3D estimation).

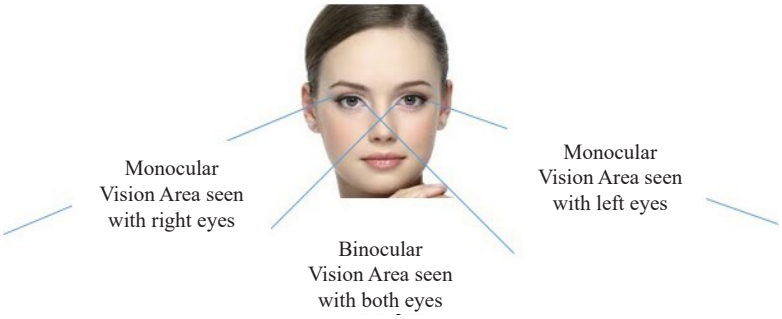

Figure 1. Angular regions a seen by one or both eyes

3D estimation and visualization of motion in a multi-camera network for sports has been proposed by [8]. The work illustrates how a low-cost camera network can be effectively used for performance analysis if the tennis ball and player tracking from a match scenario is known [9]. Extended the common techniques proposed by [10] where memory augmented deep generative models for 2D ball tracking, feature-based automatic video synchronization, and 3D estimation are described and utilized the above-mentioned techniques and improvise the overall quality of the system by developing own algorithm for prediction in case of temporally missing points in a ball's trajectory.

Xu et al. presented a novel approach for automatic sports video semantic event detection [11] based on analysis and alignment of webcast text and broadcast video and head pose in sport [12]. In this study, webcast text has been analysed to cluster and detect text events in an unsupervised way using probabilistic Latent Semantic Analysis (pLSA). Based on the detected text event and video structure analysis, a Conditional Random Field Model (CRFM) has been employed to align text event and video event by detecting event moment and event boundary in the video. The incorporation of webcast text into sports video analysis significantly facilitates sports video semantic event detection. The experiments have been conducted on 33 hours of soccer and basketball games for webcast analysis, broadcast video analysis, and text/video semantic alignment [11].

Multimarket tracking has been a difficult problem of broad interest for years in computer vision. Surveillance is perhaps the most common scenario for multi-target tracking, but team sports is another popular domain that has a wide range of applications in the strategy analysis, automated broadcasting, and content-based retrieval. Recent work in pedestrian tracking has demonstrated promising results by formulating multi-target tracking in terms of data association using deep learning models [13-16]. Support vector machine technology is used as a statistical learning model to embrace region and pixel models [17]. However, there is limited information on the used dataset or qualitative analysis to interpret the results.

The paper is structured as follows. Section 2 reviews human motion detection by monocular vision, section 3 reviews the existing problem in object detection using monocular vision Section 4 presents the proposed and the compared methods. Section 5 provides the experiment result and a comparative discussion. A comparative discussion presented in Section 6. Section 7 concludes.

2. Monocular vision and human motion

In the monocular vision (as shown in Figure 1), both eyes are used separately which affects how the brain perceives

its surroundings by decreasing the available visual area. This is an impairing peripheral vision on one side of the body which compromises depth perception. These are the major contributors to the role of vision in balance [18].

In the monocular vision, the field view is increased, while depth perception is limited [19], but the fact that 20 percent of the visual field is now effectively considered as a blind spot. This could increase to 40% if the unaffected eye tries to look in the side of the affected direction since it will be obstructed by the bridge of the nose (see Figure 2) [20]. Thus, the range and scope in monocular vision are different compared to binocular vision. The major difference between monocular vision and binocular vision can be noticed in the line of sight of the two eyes. With monocular vision, only one eye can take part in the visual field presented. With binocular vision, both eyes are used to take in the visual field. Monocular vision does not overlap vision fields because the eyes are located (left and right) besides the nose [21]. In computer vision, when one camera is used it is considered monocular vision and when two cameras are employed it is considered binocular vision.

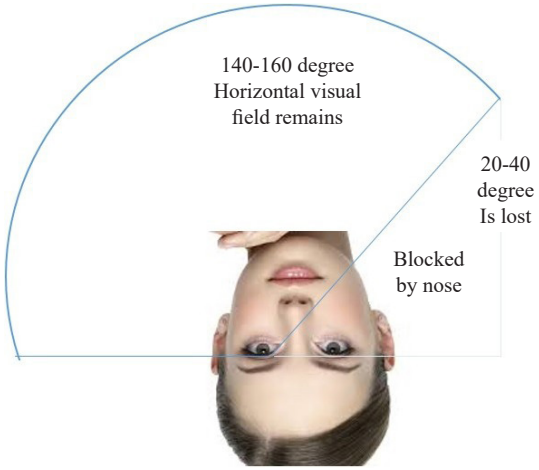

Figure 2. Graphic illustration of the visual field of a person with monocular vision

Visual measurement of the racket trajectory in spinning ball-striking for table tennis players has been proposed by [22]. Robotics table tennis has been fascinating among researchers all over the world, as it is a challenging task in both machine vision and control fields. However, a robust vision system and precise control of the manipulator required in the task are only the basic parts to fulfill playing ping-pong with people [22]. Over the past decades, some robotic systems could already compete against a human being for hundreds of rounds. Acosta et al. developed a low-cost ping-pong player using a monocular vision system [23]. Nevertheless, the aforementioned robotic systems apply only to the non-spinning case. Returning the spinning ball by a robot remains an international puzzle. Rotational measurement of the flying ball, prediction of the spinning ball succeeding trajectory, and strategy planning of striking the spinning ball, are all crucial to return a spinning ball. Among these factors, rotational measuring is a prerequisite.

Several studies already exist on rotational detection. In [24], a method for detecting the rotation using a ball marked with some feature points was provided and the rebound phenomenon between a ball and the table/racket rubber was also studied therein. High-speed multiple cameras with the capturing rate of 900 frames per second (f/s) or 1200 f/s were also needed to capture the image of the ball with marks [25-26]. The approaches with excellent performance have two common characteristics. First, the experimental expenses are largely due to the high-speed cameras with capture rates over 1000 f/s. Second, it requires a stricter experimental condition. Strong light sources and marked balls are needed in the experiments for high-speed cameras.

Wu et al. proposed a method to calculate the rotational velocity with the first several measured positions of the flying ball on the trajectory [27]. Ivan et al. proposed a new fast and robust Kinematic method [28] based on monocular vision is proposed to extract the feature sequence of rackets pose during the process of striking. Considering the complexity of the working environment and fast movement of the racket, a novel image processing technique such as a

corner extraction algorithm with good robustness and high efficiency is presented, which mainly consists of two parts, i.e., target segmentation and line detection. Then PnP-based algorithm is adopted to get the rough pose of the racket, which is then optimized by an orthogonal iteration algorithm to guarantee the orthogonality of the racket's orientation matrix [29-30].

3. Research problem

Recognizing object movement especially a human movement in a video has become a common research interest in computer vision and also in machine learning [31]. It is more difficult to extract the silhouette from the actual moving objects in a background scene where it is very complex [32] with several camera motions [33].

Extracting meaningful human motion information from video sequences is of interest for applications like intelligent human-computer interfaces, biometrics, video browsing, and indexing, virtual reality, or video surveillance. A robust human motion perception system has to necessarily deal with incomplete, ambiguous, and noisy measurements. Fundamentally, these difficulties persist irrespective of how many cameras are used [34].

Motion segmentation in video sequences is known to be a significant and difficult problem, which aims at detecting regions corresponding to moving objects [35] such as people in sports scenes. Background subtraction is a particularly popular method for motion segmentation, especially in situations with a relatively static background [36]. It attempts to detect moving regions in an image by the difference between the current image and a reference background image in a pixel-by-pixel fashion [37]. However, it is extremely sensitive to changes in dynamic scenes due to lighting and extraneous events [38].

In this study, several videos have been fed into different computer vision algorithms such as frame difference in Background subtraction, multi-layer background subtraction, and statistic frame difference in background subtraction to obtain the most accurate silhouette of the moving object. This study first obtained the video data (captured at 720 × 1280 pixels by a single camera: Samsung J5) for the sports motion. In the first video sample, the girl is jumping indoor where the lighting condition was poor and the ground was comprised of elasticity; no other object movement occurred in this situation (see Figure 3).

In the second video sample, a person is skipping and another one is jumping outdoor where the lighting condition was better compared to the first video. In both cases the camera pixels were similar. In this case, some other object movement in the background has been detected (see Figure 3).

Figure 3. Sample motion: Sports activity (indoor and outdoor)

The performance of silhouette extraction may differ due to the frame rate per sec, frame width, frame length, data rate in a camera during video shooting besides the camera resolution. The existing computer vision is proven to be good to detect the vertical movement of an object. However, when it involves sports the horizontal movement also counts as an important factor such as in a jump both the human body and clothes move upward while during walking (vertical

movement) it may not be a considerable factor in the performance of object detection. Not only that in a jump scenario if the sport participant is a girl then the hair may move up which may affect the performance of silhouette extraction.

4. Research method

Background Subtraction is a process to detect a movement or significant differences inside of the video frame, when compared to a reference, and to remove all the non-significant components (background) [39]. In this study, we have employed frame difference and multilayer background subtraction to extract silhouette. For this research, a methodology has been designed. The graphical representation depicts the research design and steps to be followed (Figure 4).

Moving object analysis benefits from silhouette extraction. At present, this study comparatively discusses two significant techniques to extract silhouettes of a moving object (here a jumping person) in monocular video data in different scenarios. The results show that the performance of silhouette extraction varies in dependency on the quality of used video data. To detect the moving object especially in sports scenarios with monocular vision, RGB mean-shift segmentation, several background subtractions, and foreground subtraction methods.

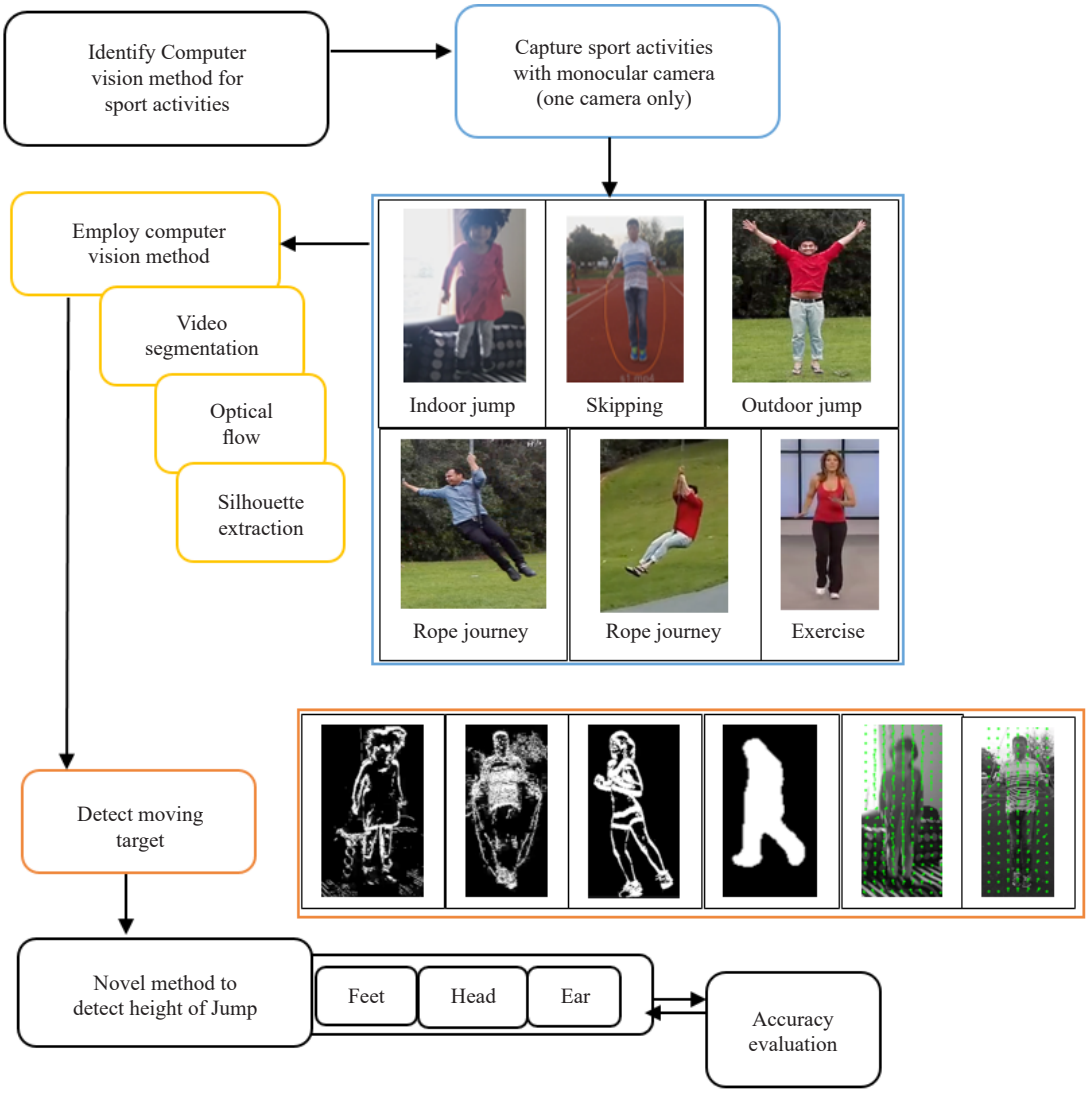

Figure 4. Research method

4.1 *Frame difference in background subtraction*

Frame differences are computed by finding the difference between consecutive frames. In this phase, the first step is to read the videos then convert them to frames. We compute the frame difference F in a time interval i.e. between F_i and F_{i+1}, represented by

$$F = F_{i+1} - F_i \tag{1}$$

In the video sequence of a total of 584 frames in 19s, F_0 to F_{301} is in the elastic ground plane with t_0 to t_{10s}. The silhouette extraction could not perform better due to elasticity in the ground plane. While the performance is better in F_{302} to the end of the video due to no elasticity on the ground plane.

In another outdoor sports video of a total frame of 689 in Figure 5 & 6 depict that the performance may vary when the outdoor environment is noisy (such as a tree is shaking due to windy). In other scenarios, if the outdoor is not much noisy the background subtraction method frame difference for the human movement performs better (see Figure 7).

Figure 5. Sample in sports: Jump indoor

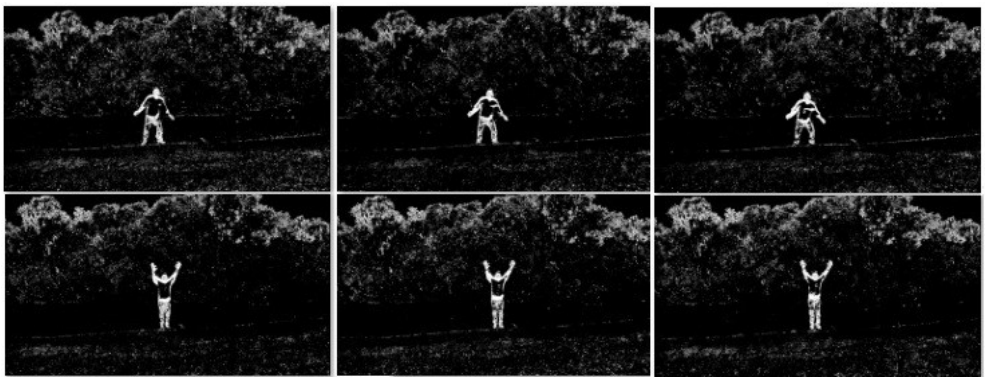

Figure 6. Sample in sports: Jump outdoor

Figure 7. Sample in sports: Skipping

4.2 *Multi-layer background subtraction*

We applied a robust multi-layer background subtraction technique [40] which takes advantage of local texture features represented by Local Binary Patterns (LBP) and photometric invariant color measurements in RGB color space. In texture and color features we introduce the local binary pattern that is to model texture and the photometric invariant color measurements, which are combined for background modeling and foreground detection. Hence, the performance is better here (see Figure 8).

Figure 8. Sample Sports: Jump (Multi-Layer)

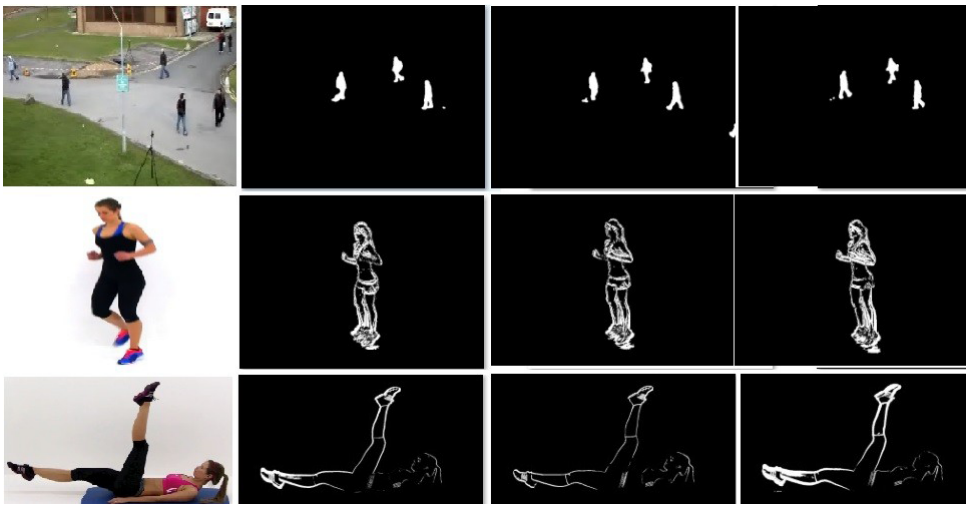

Figure 9. Silhouette extraction

5. Silhouette extraction

The silhouette is an image region defined by the circumscribing occlusion boundary, needed, for example, for human pose understanding [41]. Accurate detection and the elimination of moving cast shadows are still a challenge besides numerous efforts in this area with substantial achievements. In this study, we have employed the multilayer background subtraction method to extract the human silhouette extraction from several videos such as in Figure 9 first row explains several pedestrians were walking, 2nd and 3rd rows explain the sports activity such as workout.

6. Discussion

In this study we have used six video sequences, i. e. jump (indoor), jump (outdoor), skipping (outdoor), walk (outdoor), exercise (indoor), exercise workout (outdoor). We have noticed that when we employ frame difference for the indoor video data the performance is better when there the plane is free from elasticity. On the other hand, the same method performed better when the lighting condition is better. In both cases, the frame difference algorithm suffers for either camera is shaking or noise occurs.

However, Figure 10 shows that multilayer background subtraction outperforms mostly when the video quality is better, the frame rate is not more than 30 fps, the frame width, and height is equivalent to 720 p video resolution, and when no camera movement has occurred. Figure 10 depicts the performance between frame difference and multilayer background subtraction. The performance frame difference performs better than multilayer which is 67.69% while the frame rate in both methods was nearly the same which is 585. This is because in multilayer segmentation to the background cannot be removed completely.

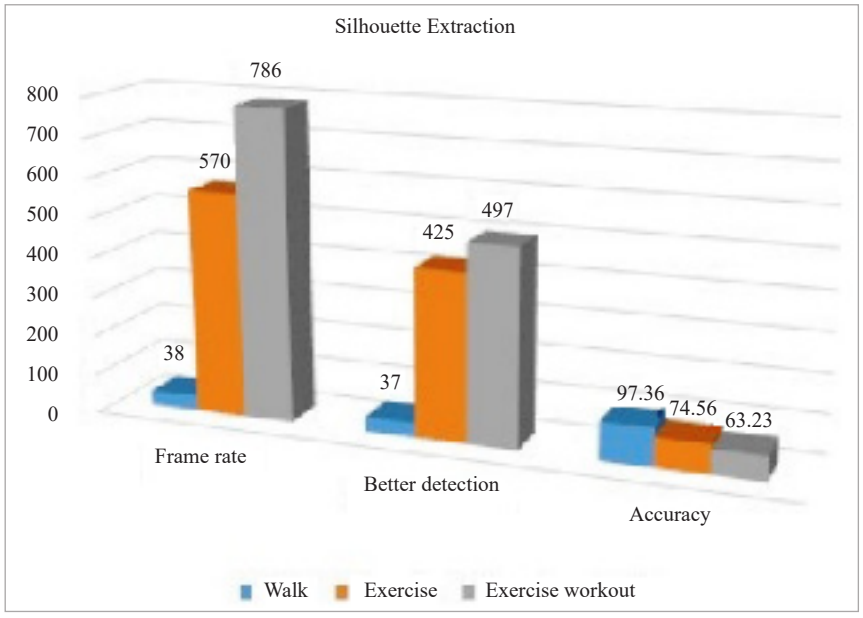

Figure 10. Silhouette extraction performance implemented on three sports activities (Walk, Exercise, and Exercise workout) using Frame difference method versus Multilayer

To test the performance of silhouette extraction we have employed three different data sets here which are namely walking, Exercise and exercise workout (see Figure 11). We adopt the pixel-based measurements as employed in [42] with the proposed and implemented accuracy calculation as follows:

$$\text{Accuracy} = \frac{TP + TN}{TP + FP + TN + FN} \tag{2}$$

Where TP notation in Eq. (2) represents the total number of true positive pixels, TN stands for True Negative as it represents the total number of pixels showing in silhouette but it belongs to background. FN represents the total number of false-negative pixels, and FP represents false positive pixels. For the sake of clarity, we treated shadow areas in silhouette as False Positive (FP) and missing areas as False Negative (FN).

We noticed that the accuracy is better (97.36% for walk data) when the total video length is small with a smaller number of frames. The second-highest performer is exercise video data (74.56%) as noted the number of frames is less than exercise workout which is 570 dues to shorter video length while the lowest performance is 63.23% for exercise workout which contains 786 number frames.

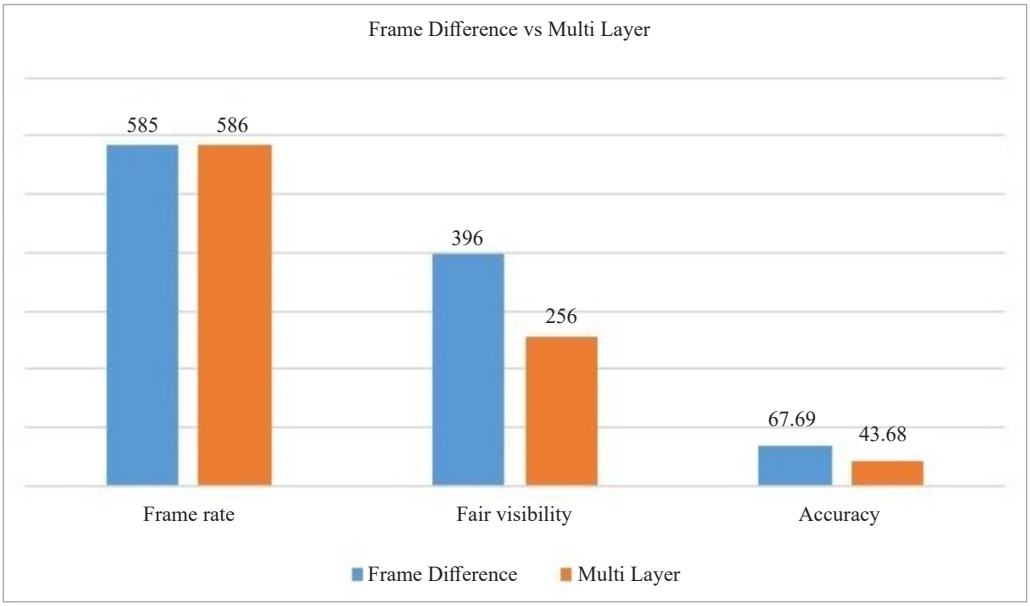

Figure 11. Test performance of frame difference versus multilayer method using three criteria (frame rate detection, fair visibility, and accuracy)

7. Conclusion

We presented a comparative performance evaluation of silhouette extraction using frame difference and multilayer. As preprocessing we applied the video stabilization technique. Afterward, we applied frame detection with background subtraction and multilayer background subtraction for extracting the silhouette from moving objects. The two methods are then systematically compared on various jumping person sequences. Experiments performed to show the performance of silhouette extraction vary due to the quality of the video (i.e. lighting condition, frame rate, data rate). For the video sequence of exercise, it performed nearly well. The indoor sports video with poor lighting conditions for jump worked better in a stable ground plane compared to the elastic ground plane. In the other video of skipping it was difficult to observe the performance due to the background object's movement, even the lighting condition was better. Where multilayer background subtraction offers the best performance regardless of the data rate, frame rate while the lighting condition is better and the video is stabilized. One solution to improve the multilayer frame difference is to investigate the usage of Self-Organizing Background Subtraction (SOBS) or Convolutional Neural Networks (CNN) focusing on the domain of sport activities.

Acknowledgment

The authors thank Reinhard Klette and Enrico Haem-merle (both AUT) for discussions and hints, Ting Yen Chen (Academica Sinica, Taiwan) for providing the used mean-shift code for the RGB feature space, and Taeba, ASM Hemayet Karim, and Monirul Alam Chowdhury for being our models for sport motion capture.

Conflict of interest statement

The authors declare that there is no conflict of interest.

References

[1] Klette R. *Concise Computer Vision*. Berlin, German: Springer; 2014.
[2] Sarfaraz M. *Computer Vision and Image Processing in Intelligent Systems and Multimedia Technologies*. Hershey: IGI Global; 2014.
[3] Chandra S, Thomas P. Face recognition from one sample per person. *International Journal of Scientific Research in Computer Science Applications and Management Studies*. 2014; 3: 832-434.
[4] Han S, Nandakumar R, Philipose M, Krishnamurthy A, Washington D. Glimpse Data: Towards continuous vision-based personal analytics. *WPA '14: Proceedings of the 2014 workshop on physical analytics*. New York: Association for Computing Machinery; 2014. p. 31-36.
[5] Klette R. Vision-based driver assistance. *Wiley Encyclopedia Electrical Electronics Engineering*. Wiley Online Library; 2015. Available from: doi: 10.1002/047134608X.W8272.
[6] Tung T, Gomez R, Kawahara T, Matsuyama T. Group dynamic and multi-modal interaction modelling using a smart digital signage. *European Conference on Computer Vision*. Berlin, German: Springer; 2012. p. 362-371. Available from: doi: 10.1007/978-3-642-33863-2_36.
[7] Miller G. *A first blueprint for machine vision: Look, record, then perfect*. Available from: https://possibility.teledyneimaging.com/first-blueprint-machine-vision-look-record-perfect/ [Accessed 3rd July 2021].
[8] Xu Y, Li YJ, Weng XS, Kitani K. Wide-baseline multi-camera calibration using person re-identification. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Piscataway, NJ, USA: IEEE; 2021. p. 13134-13143.
[9] Melo AG, Pinto MF, Marcato ALM, Biundini IZ, Rocha NMS. Low-cost trajectory-based ball detection for impact indication and recording. *Journal of Control, Automation and Electrical Systems*. 2021; 32: 367-377. Available from: doi: 10.1007/s40313-020-00677-7.
[10] Tharindu F, Simon D, Sridha S, Clinton F. Memory augmented deep generative models for forecasting the next shot location in tennis. *IEEE Transactions on Knowledge and Data Engineering*. 2020; 32(9): 1785-1797. Available from: doi: 10.1109/TKDE.2019.2911507.
[11] Xu J, Brubaker SC, Mullin MD, Rehg JM. Fast asymmetric learning for cascade face detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 2008; 30(3): 369-382. Available from: doi: 10.1109/TPAMI.2007.1181.
[12] Kian AR, Matiolanski A. Crowd density estimation based on face detection under significant occlusions and head pose variations. *International Conference on Multimedia Communications, Services and Security*. Berlin, German: Springer; 2020. p. 209-222.
[13] Mangi S. Multi-target tracking for video surveillance using deep affinity network: a brief review. arXiv e-prints [Preprint]. 2021. Available from: https://ui.adsabs.harvard.edu/abs/2021arXiv211015674N/abstract [Accessed 16th September 2021].
[14] Kong L. Precise positioning and calibration of sports teaching movements based on artificial intelligence deep learning technology. *Design Engineering*. 2020; 2020(2): 444-459. Available from: doi: 10.17762/de.vi.213.
[15] Duan C. Deep learning-based multitarget motion shadow rejection and accurate tracking for sports video. *Complexity*. 2021; 2021: 5973531. Available from: doi: 10.1155/2021/5973531.
[16] Yu K. Visual mapping of target tracing methods based on CiteSpace bibliometrics. *2021 IEEE International*

Conference on Electronic Technology, Communication and Information (ICETCI). Piscataway, NJ, USA: IEEE; 2021. p. 607-616.
[17] Deepak R, Shanmugapriya S, AbdulHaleem SL. Comparative analysis of motion detection methods or video surveillance systems. *Proceedings of the Third International Symposium*. Oluvil, Sri lanka: South Eastern University of Sri lanka; 2013. p. 71-80.
[18] Berela J. Use of monocular and binocular visual cues for postural control in children. *Journal of Vision*. 2011; 11(12): 1-8. Available from: doi: 10.1167/11.12.10.
[19] Wolfe M. *E-Study for: Sensation and Perception*. Boston, USA: Allyn & Bacon; 2013. p. 124-126.
[20] Pretorius N. *An in-depth look at monocular vision and ocular prostheses*. Available from: https://eyeternus.wordpress.com/2013/11/11/an-in-depth-look-at-monocular-vision-ocular-prostheses/ [Accessed 12th September 2021].
[21] Alston C. *Monocular Vision: Definition & Explanation*. Available from: https://study.com/academy/lesson/monocular-vision-definition-lesson-quiz.html [Accessed 12th September 2021].
[22] Chen G, Xu D. Visual measurement of the racket trajectory in spinning ball striking for table tennis player. *IEEE Transactions on Instrumental and Measurement*. 2013; 62(11): 2901-2911. Available from: doi: 10.1109/TIM.2013.2265471.
[23] Acosta L, Rodrigo JJ, Mendez JA, Marichal GN, Sigut M. Ping-pong player prototype. *IEEE Robotics & Automation Magazine*. 2003; 10(4): 44-52. Available from: doi: 10.1109/MRA.2003.1256297.
[24] Lin H-I, Zhang GY, Huang YC. Ball tracking and trajectory prediction for table-tennis robots. *Sensors*. 2020; 20(2): 333. Available from: doi: 10.3390/s20020333.
[25] Nakashima A, Ogawa Y, Kobayashi Y, Hayakawa Y. Modeling of rebound phenomenon of a rigid ball with friction and elastic effects. *Proceedings of the 2010 American Control Conference*. Piscataway, NJ, USA: IEEE; 2010. p. 1410-1415.
[26] Wen HJ, Xiao ZY, Li YT, Xu Y. A vision system for shot tracking and thrown distance measurement. *2020 7th International Conference on Information Science and Control Engineering (ICISCE)*. Piscataway, NJ, USA: IEEE; 2020. p. 1647-1651.
[27] Wu E, Hideki K. Futurepong: Real-time table tennis trajectory forecasting using pose prediction network. *Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems*. New York, USA: Association for Computing Machinery; 2020. p. 1-8.
[28] Ivan ML, Sandro B, Rocco DM, Gu YD, Julien SB, Silvia F, Matteo C. Kinematic analysis of the racket position during the table tennis top spin forehand stroke. *Applied Sciences*. 2021; 11(11): 5178. Available from: doi: 10.3390/app11115178.
[29] Chen G, Xu D. Visual measurement of the racket trajectory in spinning ball striking for table tennis player. *IEEE Transactions on Instrumentation and Measurement*. 2013; 62(11): 2901-2911. Available from: doi: 10.1109/TIM.2013.2265471.
[30] Gao YP, Jonas T, Julian K, Andreas Z. Markerless racket pose detection and stroke classification based on stereo vision for table tennis robots. *2019 Third IEEE International Conference on Robotic Computing (IRC)*. Piscataway, NJ, USA: IEEE; 2019. p. 189-196.
[31] Brendel W, Todorovic S. Learning spatiotemporal graphs of human activities. *13th International Conference on Computer Vision (ICCV)*. Piscataway, NJ, USA: IEEE; 2011. p. 778-785.
[32] Wang H, Klaser A, Schmid C, Liu CL. Dense trajectories and motion boundary descriptors for action recognition. *International Journal of Computer Vision*. 2013; 103: 60-79. Available from: doi: 10.1007/s11263-012-0594-8.
[33] Harris C, Stephens M. A combined corner and edge detector. *Proceedings of the 4th Alvey Vision Conference*. Sheffield, UK: University of Sheffield Printing Office. 1988. p. 147-151.
[34] Rosenhahn B, Klette R, Metaxas D. *Human Motion*. Berlin, German: Springer; 2008.
[35] Kimmel R, Klette R, Sugimoto A. Computer vision. *10th Asian Conference on Computer Vision*. Queenstown, Berlin, German: Springer; 2010. p. 15-26.
[36] Kim H, Ryu D, Park J. Smoke detection using GMM and adaboost. *International Journal of Computer and Communication Engineering*. 2014; 3(2): 123-126. Available from: doi: 10.7763/IJCCE.2014.V3.305.
[37] Meghanathan N, Nagamalai D, Chaki N. Advances in computing and information technology. *Proceedings of the Second International Conference on Advances in Computing and Information Technology (ACITY)*. Berlin, German: Springer; 2012. p. 750-754.
[38] Apolloni B, Basis S, Esposito A, Morabito FC. *Neural Nets and Surroundings*. Berlin, German: Springer; 2012. Available from: https://link.springer.com/book/10.1007/978-3-642-35467-0 [Accessed 5th October 2021].

[39] Vinay DR, Kumar NL. Object tracking using background subtraction algorithm. *International Journal of Engineering Research and General Science*. 2015; 3(1): 237-243.
[40] Yao J, Odobez JM. Multi-Layer background subtraction based on color and texture. *2007 IEEE Conference on Computer Vision and Pattern Recognition*. Piscataway, NJ, USA: IEEE; 2007. p. 1-8.
[41] Rosenhahn B, Kersting U, Powell K, Klette R, Klette G, Seidel HP. A system for articulated tracking incorporating a clothing model. *Machine Vision and Applications*. 2006; 18: 25-40. Available from: doi: 10.1007/s00138-006-0046-y.
[42] Yao Q, Sankoh H, Sabirin H, Naito S. Accurate silhouette extraction of multiple moving objects for free viewpoint sports video synthesis. *IEEE 17th International Workshop on Multimedia Signal Processing (MMSP)*. Piscataway, NJ, USA: IEEE; 2015. p. 1-6.