International Journal of Engineering & Technology, 7 (3.27) (2018) 407-411

International Journal of Engineering & Technology

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

Research paper

Novel Adaptive Background Segmentation Algorithm for Multiple Object Tracking

V. Ramalakshmi @ Kanthimathi[1*], M. Germanus Alex[2]

[1]Research Scholar, Department of Computer Science, Bharathiar University, Coimbatore, India.
Assistant Professor, Kamarajar Government Arts College, Surandai.
[2] Research Guide, Department of Computer Science, Bharathiar University, Coimbatore, India.
Assistant Professor, Kamarajar Government Arts College, Surandai.

Abstract

Multiple object tracking plays a vital role in many applications. The objective of this paper is to track multiple objects in all the scenes of the video sequence. In this paper, an algorithm is proposed to identify objects between scenes by dividing the scenes in the video sequence. Within each scene, objects are identified and tracked between scenes by segmenting the background adaptively. The proposed method is tested on four publicly available datasets. The experimental results substantially proved that the proposed method achieves better performance than other recent methods.

Keywords: *object tracking, background subtraction, scene detection.*

1. Introduction

Multiple object tracking is an widely learned area of research. More approaches has been introduced which recursively update the object tracks with the most recent detections. Most importantly, Kalman filtering is the most efficient method to track multiple objects [1] if the number of objects tracked is small. It is well suited for real-time applications. If the number of objects increases, identity switches become more frequent and are difficult to correct, due to the recursive nature of the method. Wu et.al tracks multiple humans using mean-shift algorithm [2] which also has the same drawback. Particle filtering overcomes some of the drawbacks of Kalman filtering by using some hypotheses [3]. This method tracks multiple hockey players [4]in the ground [5]. Similarly, the algorithm [6] tracks multiple object tracking to recover trajectories of targets using a set of observations. In [7] Probability Hypothesis Density filter is used to track multiple objects from noisy environment.

In order to increase the efficiency, some methods follow hybrid approach. The algorithm in [8] uses the hierarchical version of the same concept, while [9] uses a variant of AdaBoost to automatically learn the best criterion for linking low level tracks together. Similarly, [10] used some observations into trajectory segments using local PCA, and then links those segments based on their spatial proximity and smoothness constraints. In [11] mean-shift or particle filtering is used to generate tracklets from the detected results. It uses data association to combine the tracklets into full tracks, and to automatically estimate the best parameters for the model. Motion model and nearest neighbor is used in [12] to build tracks detected from a top mounted calibrated camera. Then these tracks are merged and split into final trajectories using heuristics based on overlap, directions and speed. Another method [13] is introduced for tracklet generation in a crowded environment. It detects multiple people and creates tracklets by applying Bayesian clustering. In contrast, the algorithm developed in [14] assumed that the track graph has already been produced and focused on linking identities in the provided track graph.

To improve robustness, most researches have been recently focused on linking detections over a larger time using various optimization schemes. In [15], graph cuts are used to extract trajectories from a batch of people obtained using homographic constraints on images from multiple cameras while [16] optimizes detections and tracking. It coupled into a Quadratic Boolean Problem. Dynamic Programming [17] can be used to track multiple detections, which solves the multi-target tracking problem. Moreover, it can be extended to enable the optimization of several trajectories simultaneously [18]. Unfortunately, it suffers from computational complexity.

In this paper, the input video sequence is divided into scenes. Within each scene, key frame is selected to identify objects. Background is segmented in each scene using the key frame selected. The identified objects are tracked in the remaining of the scene using correlation filtering. The proposed method is tested on three publicly available datasets and the experiments proved the good performance of it. The remaining of the chapter is organized as follows: Section II describes the proposed system architecture. Section III explains the proposed algorithm. Section IV demonstrates some experiments for proving the performance followed by conclusion in section V.

2. Proposed System Architecture

The proposed system architecture is shown in Fig. 1. The video sequence consists of more relevant information within consecutive frames. The idea behind the proposed method is to use this characteristic of the video sequence. Instead of tracking the object in the full video, the object is tracked within each scene using correlation filter.

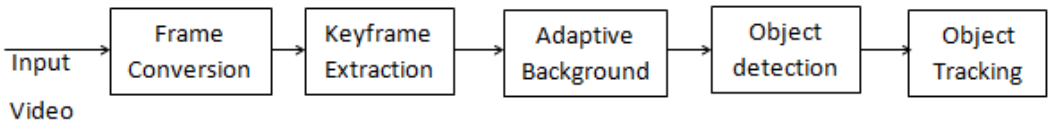

Fig. 1: Proposed system architecture

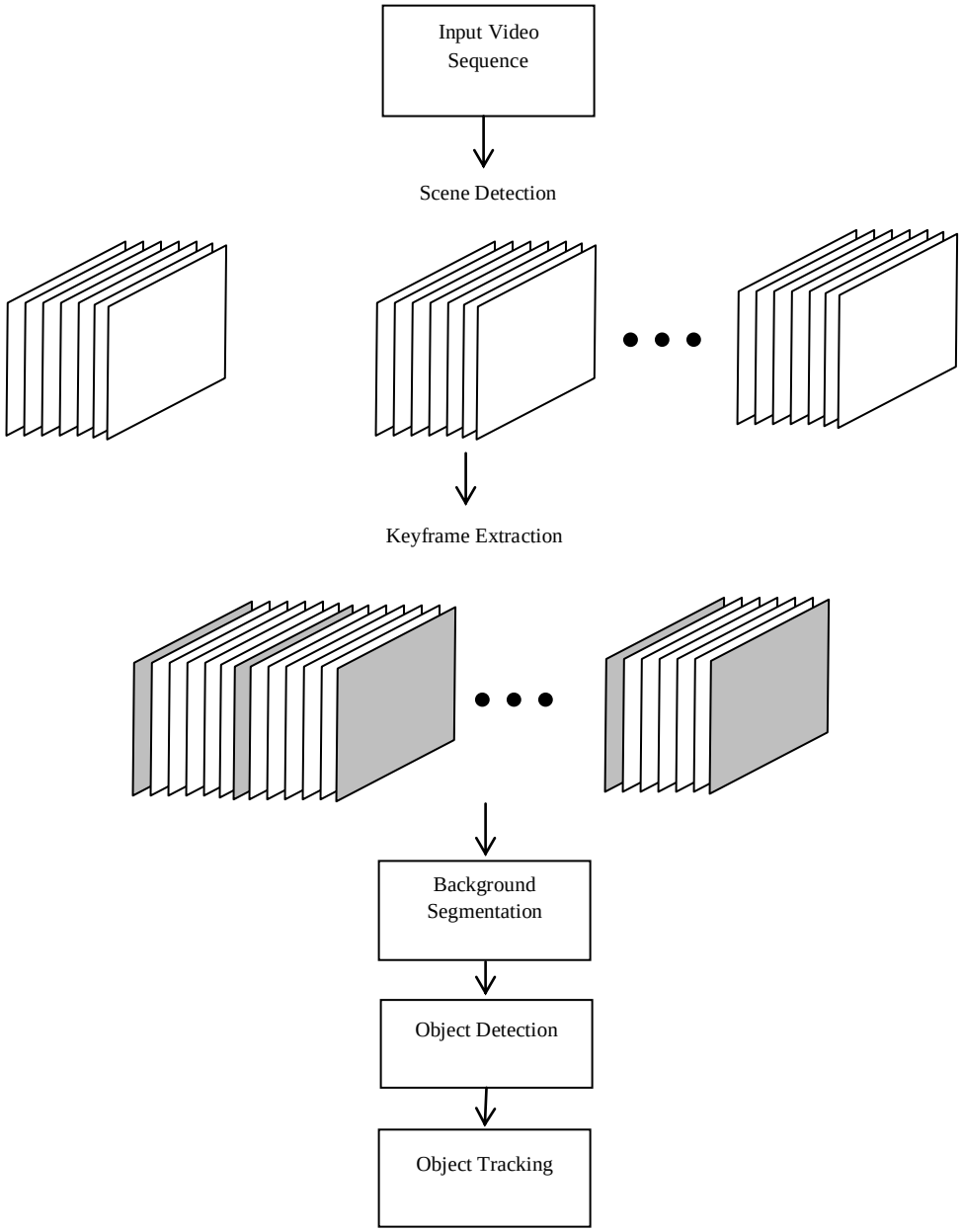

Fig. 2: Flow of the proposed method

3. Scene Change Detection

The proposed method uses scene change detection algorithm which is discussed in detail. The scene is identified using Pearson Correlation Coefficient (PCC). After the scene is identified, the first frame is selected as keyframe. Using the keyframe, the background is subtracted for each other frame in the scene. The objects are identified and tracked in each scene. For this object tracking, the standard algorithm is used.

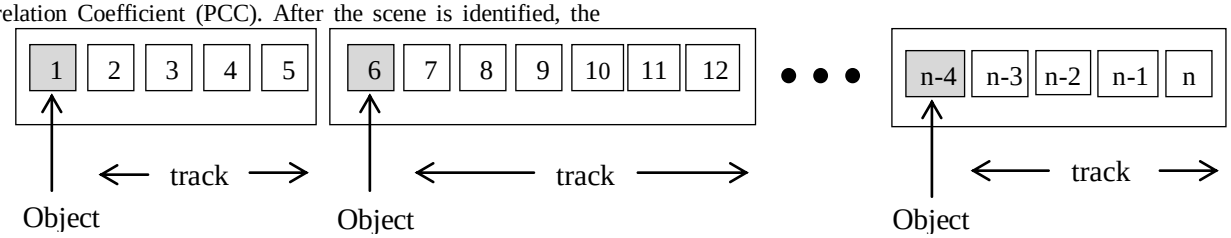

Fig. 3: Depiction of scene detection (Frame marked as gray is the keyframe in each scene)

In any video sequence, scene may change between frames. But all the changes are not visually identified. The logic of frame comparison is illustrated in Fig. 3. For evaluating the similarity of frame and its prediction, Pearson Correlation Coefficient is chosen. Pearson correlation coefficient is widely used to measure the similarity of two frames for cut detection. The value of Pearson Correlation Coefficient can fall between 0 (no correlation) and 1 (perfect correlation). Correlations above 0.80 are considered as really high and lowest values will be determined as cuts. The Pearson correlation coefficient for 2D signals like video sequences is expressed as follows.

$$PCC = \frac{\sum_{i=1}^{M}\sum_{j=1}^{N}(f(i,j)-f^m)(f_p(i,j)-f_p^m)}{\sqrt{\sum_{i=1}^{M}\sum_{j=1}^{N}(f(i,j)-f^m)^2(f_p(i,j)-f_p^m)^2}} \quad (1)$$

where f, $f_p$represent pixel intensities of the current frame and the P-frame respectively. M, N are the number of rows and columns in the frame. f^m, f_p^m are mean pixel intensities of the current frame and the P- frame.

Initially the first frame (A frame) which is the keyframe compared with the next frame (B frame) in the video sequence. If the correlation between them is low, scene cut is detected and the size of the scene is fixed. The next frame (again it is considered as A frame) is chosen as keyframe for the next scene and the flow continues as previously explained. If the correlation is high, then the next frame is compared with A frame. This process is repeated till end of the video sequence is reached. Figure 4 illustrates the process by which the GOP is fixed adaptively. At each step, the correlation factor given in Eq. (1) is computed.

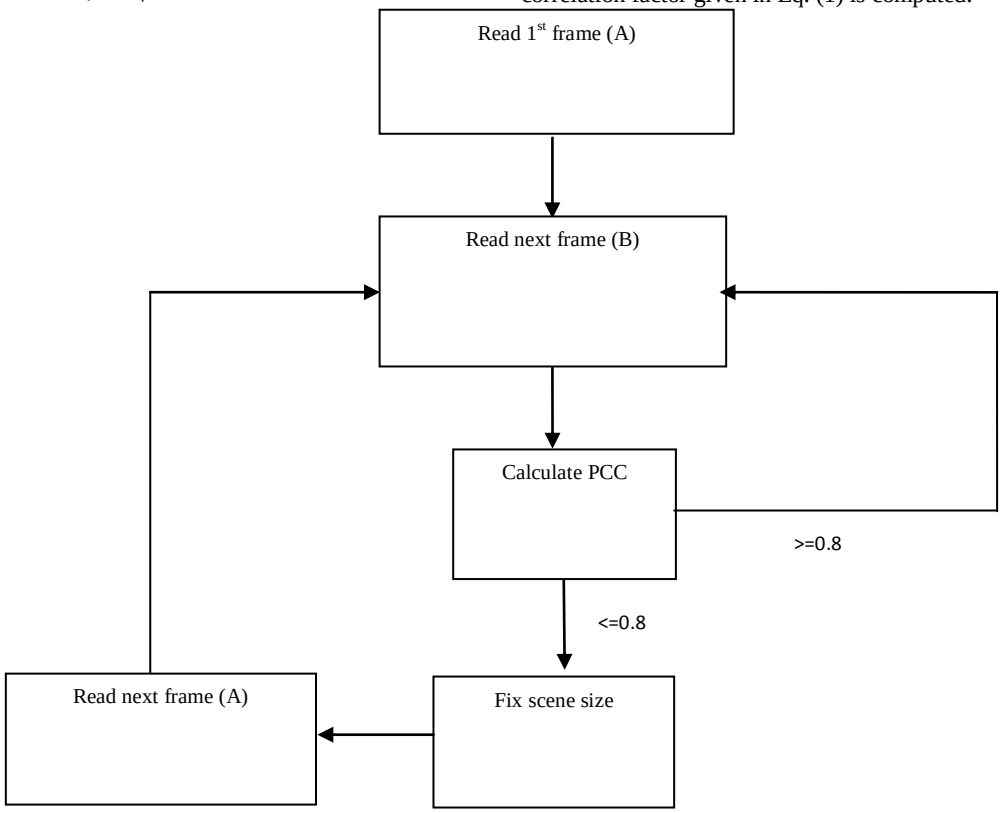

Fig. 4: Flow chart of scene change detection

The background is subtracted in each scene with the selected keyframe. The object is identified in each keyframe which is tracked in each background subtracted frame of the scene using correlation filter.

4. Experimental Results

The proposed method is tested on a variety of challenging video datasets: TUD Campus, TUD Crossing, TUD Stadtmitte and PETS2009 S2-L1 [19]. They are commonly used video datasets and they are very challenging for several reasons. They include outdoor environment where lighting conditions are not controlled. In PETS2009 video, the video covers large area, so people look very small when they are far from the camera making their tracking more challenging. In TUD dataset, targets have a similar size and they walk with similar speeds. However, targets are frequently occluding each other (heavy inter-object occlusion) and are occluded by static objects. To obtain the detections, we use the detections originally provided with the videos [19]. Video sequence details of each datasets are given in Table 1.

Table 1: Details of Dataset

Sequence	# frames	Persons	Resolution
TUD-CAMPUS	71	Up to 6	640x480
TUD-CROSSING	201	Up to 8	640x480
TUD-STADTMITTE	179	Up to 8	640x480
PETS2009-S2-L1	795	Up to 10	768x576

Two performance metrics are used to evaluate the performance of the proposed MOT system as used in [20]. They are explained as follows.

(a) **Multiple Object Tracking Accuracy (MOTA)**: It is one of the widely used evaluation metrics for multiple object tracking applications. It is defined as below.

$$MOTA = 1 - \frac{\sum_t(FP(t)+FN(t)+ID(t))}{\sum_t N_{GT}(t)} \quad (2)$$

(b) **Multiple Object Tracking Precision (MOTP)**: The MOTP is defined as below.

(c)

$$MOTP = \sum_{t,i}\frac{\bar{d}\left(\mathcal{G}T_i^t\mathcal{H}_{g(i)}^t\right)}{\sum_t m_t} \quad (3)$$

The proposed method is compared with recent state-of-the-art MOT algorithms. Among the compared approaches, a first

category studied MOT with the aim of improving detection responses using model-free tracker [19, 21], a second category aimed to improve the data association technique [22-23], and a third category aimed to improve the appearance model [24 and 25]. The results are obtained from the authors' papers. Table 2 and 3 show the results of the proposed method and its performance comparison with the recent methods.[26]

Table 2: Comparison of the Proposed Method in TUD-CAMPUS and TUD-CROSSING Datasets

Dataset	TUD-CAMPUS		TUD-CROSSING	
Measure	**MOTA (%)**	**MOTP (%)**	**MOTA (%)**	**MOTP (%)**
Proposed	79.8	69.5	78.5	67.42
[Riahi, 2014]	72	74	72	76
[Breitenstein, 2011]	73	67	84	71

Table 3: Comparison of the proposed method in TUD-STADTMITTE and PETS2009-S2L1 datasets

Dataset	TUD-STADTMITTE		PETS2009-S2-L1	
Measure	**MOTA (%)**	**MOTP (%)**	**MOTA (%)**	**MOTP (%)**
Proposed	68.12	58.65	85.22	67.56
[Andriyenko, 2011]	60.5	66	80	76
[Milan, 2014]	71	65.5	90	80

Fig 4 and 5 shows the bar chart showing performance comparison of the proposed method with the recent methods .

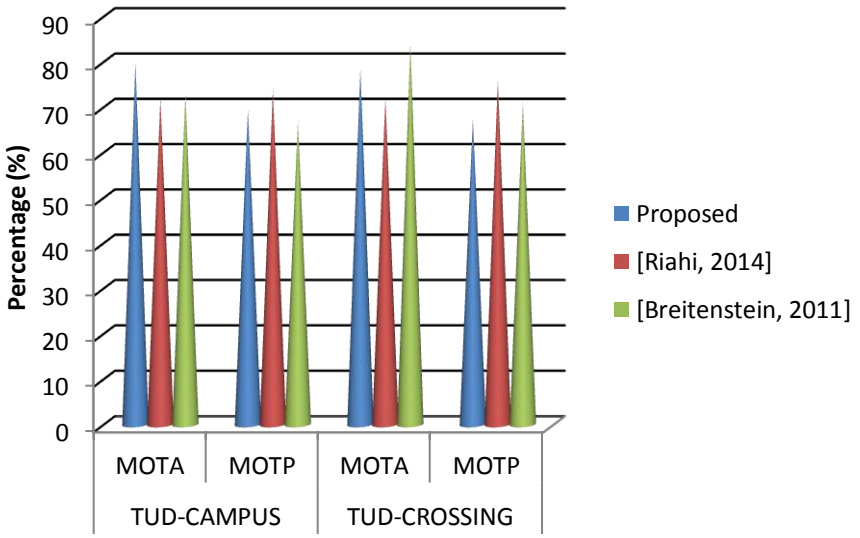

Fig. 4: Bar chart displaying performance comparison of the proposed method with recent method in TUD-CAMPUS and TUD-CROSSING datasets

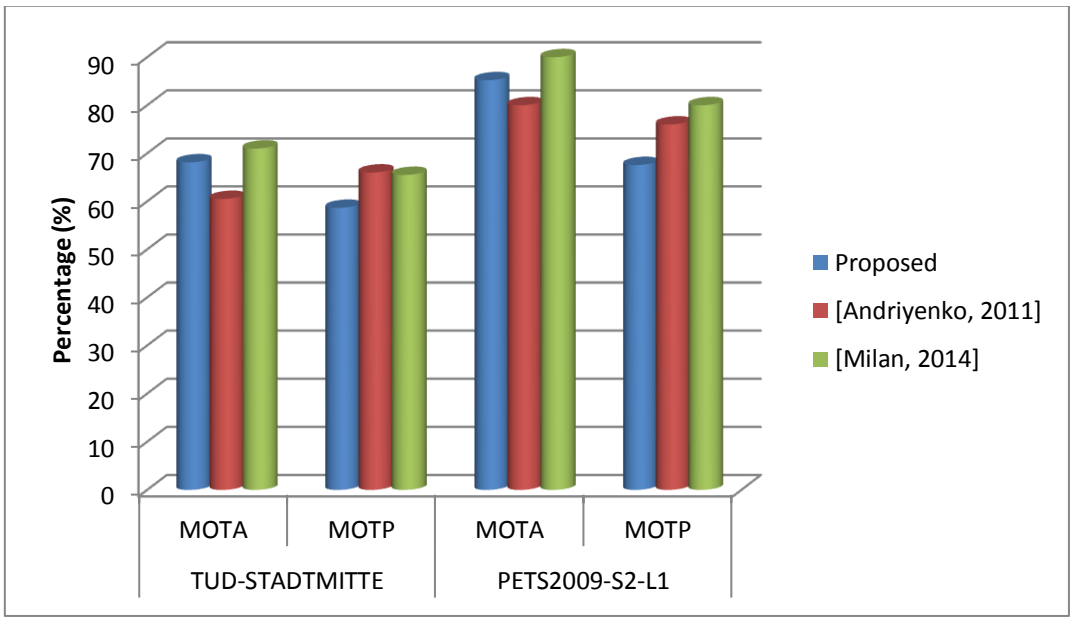

Fig. 5: Bar chart displaying performance comparison of the proposed method with recent method in TUD-CAMPUS and TUD-CROSSING datasets

From Table 2, it is clear that the proposed method achieves higher MOTA and lesser MOTP than other recent methods.

5. Conclusion

Combining frame-by-frame detections to estimate the most likely trajectories of an unknown number of targets, including their entrances and departures to and from the scene, is one of the most difficult components of a multi-object tracking algorithm. It is obtained by dividing the video into scenes. The objects are identified and tracked within each scene. The resulting algorithm is far simpler than current state-of-the-art algorithms. The proposed method obtains a better performance than a state-of-the-

art method on difficult datasets. In future, various scene change detection algorithms can be used.

References

[1] Black J, Ellis T & Rosin P, "Multi-View Image Surveillance and Tracking", *IEEE Workshop on Motion and Video Computing*, (2002).

[2] Wu B & Nevatia R, "Tracking of Multiple, Partially Occluded Humans Based on Static Body Part Detection", *Conference on Computer Vision and Pattern Recognition*, (2006), pp.951–958.

[3] Vermaak J, Doucet A & Perez P, "Maintaining Multimodality Through Mixture Tracking", *International Conference on Computer Vision*, (2003), pp.1110–1116.

[4] Okuma K, Taleghani A, de Freitas N, Little J & Lowe D, "A Boosted Particle Filter: Multi target Detection and Tracking", *European Conference on Computer Vision*, (2004).

[5] Du W & Piater J, "Multi-Camera People Tracking by Collaborative Particle Filters and Principal Axis-Based Integration", *Asian Conference on Computer Vision*, (2007), pp.365–374.

[6] Yu Q, Medioni G & Cohen I, "Multiple Target Tracking Using Spatio-Temporal Markov Chain Monte Carlo Data Association", *International Conference on Computer Vision*, (2007).

[7] Maggio E, Taj M & Cavallaro A, "Efficient Multi-Target Visual Tracking Using Random Finite Sets", *IEEE Transactions On Circuits And Systems For Video Technology*, Vol.18, No.8, (2008), pp.1016–1027.

[8] Huang C, Wu B & Nevatia R, "Robust Object Tracking by Hierarchical Association of Detection Responses", *European Conference on Computer Vision*, (2008), pp.788–801.

[9] Li Y, Huang C & Nevatia R, "Learning to Associate: Hybrid boosted Multi-Target Tracker for Crowded Scene", *conference on Computer Vision and Pattern Recognition*, (2009).

[10] Beleznai C, Fruhstuck B & Bischof H, "Multiple Object Tracking Using Local Pca", *International Conference on Image Processing*, (2006).

[11] Ge W & Collins RT, "Multi-target data association by tracklets with unsupervised parameter estimation", *British Machine Vision Conference*, (2008).

[12] Eshel R & Moses Y, "Homography Based Multiple Camera Detection and Tracking of People in a Dense Crowd", *Conference on Computer Vision and Pattern Recognition*, (2008).

[13] Brostow GJ & Cipolla R, "Unsupervised Bayesian Detection of Independent Motion in Crowds", *Conference on Computer Vision and Pattern Recognition*, (2006), pp.594–601.

[14] Nillius P, Sullivan J & Carlsson S, "Multi-Target Tracking - Linking Identities Using Bayesian Network Inference", *Conference on Computer Vision and Pattern Recognition*, (2006), pp.2187–2194.

[15] Khan S & Shah M, "Tracking Multiple Occluding People by Localizing on Multiple Scene Planes", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol.31, No.3, (2009), pp.505–519.

[16] Leibe B, Schindler K & Gool LV, "Coupled Detection and Trajectory Estimation for Multi-Object Tracking", *International Conference on Computer Vision*, (2007).

[17] Bellman RE, *Dynamic Programming*, Princeton University Press, (1957).

[18] Wolf J, Viterbi A & Dixon G, "Finding the Best Set of K Paths Through a TrellisWith Application to Multi-target Tracking", *IEEE Transactions on Aerospace and Electronic Systems*, Vol.25, No.2, (1989), pp.287–296.

[19] Milan A, Schindler K & Roth S, "Detection-and trajectory-level exclusion in multiple object tracking", *IEEE CVPR*, (2013), pp. 3682–3689.

[20] Wu Y, Lim J & Yang MH, "Online object tracking: A benchmark", *IEEE Conference on Computer vision and pattern recognition (CVPR)*, (2013), pp.2411-2418.

[21] Breitenstein MD, Reichlin F, Leibe B, Koller-Meier E & Van Gool L, "Online multi person tracking-by-detection from a single, un calibrated camera", *IEEE Transactions on Pattern Analysis and Machine Intelligence(PAMI)*, Vol.33, No.9, (2011), pp.1820–1833.

[22] Segal AV & Reid I, "Latent data association: Bayesian model selection for multi-target tracking", *IEEE ICCV*, (2013), pp.2904–2911.

[23] Andriyenko A & Schindler K, "Multi-target tracking by continuous energy minimization", *IEEE CVPR*, (2011), pp.1265–1272.

[24] Riahi D & Bilodeau GA, "Multiple feature fusion in the dempster-shafer framework for multi-object tracking", *IEEE Computer and Robot Vision (CRV)*, (2014), pp.313–320.

[25] Z Yesembayeva (2018). Determination of the pedagogical conditions for forming the readiness of future primary school teachers, Opción, Año 33. 475-499

[26] G Mussabekova, S Chakanova, A Boranbayeva, A Utebayeva, K Kazybaeva, K Alshynbaev (2018). Structural conceptual model of forming readiness for innovative activity of future teachers in general education school. Opción, Año 33. 217-240