Automated Human Tracking using Advanced Mean Shift Algorithm

Suraj Prakash Sahoo and Samit Ari

Abstract—**For tracking human or any kind of object, colour feature based mean shift technique is widely used. This technique uses Bhattacharya coefficient to locate the object based on the maximisation of the similarity function between object model and candidate model. Traditional mean shift algorithm fails when the object having large motion, occlusion, corrupted frames etc. In addition to that, the technique is not automatic to initiate the tracking. To overcome all these problems, this paper proposes a technique which uses three additional modules to the traditional method to make it more efficient. The proposed modules uses human detection by modelling through star skeletonization, followed by block search algorithm and occlusion handling. Block search algorithm helps to supply an overlapping area to candidate model to continue the track when tracking fails due to fast motion. Occlusion handling will help to initiate the tracking after prolonged period of occlusion. The proposed method has been tested on real time data and it outperforms the conventional method effectively to overcome the mentioned problems up to large extent.**

Index Terms— **Block search, Human detection, Mean shift, Occlusion, Star model.**

I. INTRODUCTION

SUCCESS of various available tracking algorithms mostly depends on the background features and movement of object in the video sequence. Problems which reduce the tracking efficiency are illumination changes, background clutter, and partial or full occlusions. A practical real time efficient tracking system should overcome all the problems and provide efficient performance by decreasing false alarm rate. Tracking is of two types: probabilistic and deterministic. In probabilistic model the system and its measurements remain uncertain (Kalman filtering, particle filter etc) but, when occlusion is there, it works better than deterministic methods. Mean shift tracking belongs to the second category where a model will be compared with the current frame to find the most promising region. The traditional mean shift tracker is unable to start the tracking automatically and could not handle the fast motion of the object. Similarly, while handling the occlusion, it shows poor result [1] in comparison to probabilistic method.

Since mean shift technique uses colour features [2, 3] for its operation, it is simple and robust to target deformation and scale variance (shape and size). To increase the efficiency of traditional mean shift algorithm, various methods are reported from time to time. A robust object tracking approach using mean shift is presented in [4] using a weighted histogram and eliminating the background effect. Multi-fragment representations of the target and candidate models have been implemented to improve the robustness of tracking especially to partial occlusion in [5]. Using STAGE, object recovery can be implemented up to some extent as in [6]. An Adaptive Tracking window has been used to adopt scale and orientation change in [7].

When prolonged period of occlusion is there and in those period motional characteristics of the object (velocity, acceleration) changes abruptly, success rate of above mentioned works [1-7] decreases up to large extent. Similarly, automatic tracking of human is also a challenge in this field. Therefore, proposed method selects a relatively easier deterministic approach (mean shift) and adds automated tracking by human detection, block search and occlusion handling capability to make it a better approach for tracking. The proposed method has been applied to real time videos to ensure the efficiency of the algorithm.

The remainder of this paper is organized as follows. We introduce what is simple mean shift tracking and some of its basic problems in Section II. Section III briefly presents the proposed method which describes background subtraction, human detection, block matching and occlusion handling. The performance of the proposed method is evaluated in Section IV, conclusions and future works are given in Section V.

II. PROBLEMS WITH MEAN SHIFT TRACKING

Object model can be formed by selecting a particular region and calculating the colour histogram q_u [2] of that model by :

$$q_u = C\sum_{i=1}^{n} k(\| \mathrm{x}_i^* \|^2)\delta[\mathrm{b}(\mathrm{x}_i^*) - \mathrm{u}] \qquad (1)$$

where, x_i^*, i=1,…,n represents the pixel locations inside the selected region, b (x_i^*) is the bin number (1,...,m) associated with the colour at the pixel of normalized location x, δ is the Kronecker delta function and C is the constant normalization. Similarly the candidate model at the same position in current frame and its colour histogram $p_u(y)$ [2] can be found by (2).

Suraj Prakash Sahoo is student of Signal and Image Processing Group, Department of Electronics and Communication, National Institute of Technology, Rourkela (e-mail: surajprakashsahoo@gmail.com)

Dr. Samit Ari is the assistant professor in Department of Electronics and Communication, National Institute of Technology, Rourkela (e-mail: Samit.ari@gmail.com)

$$p_u(\mathrm{y}) = \mathrm{C}_h \sum_{i=1}^{n_h} k\left[\left\|\frac{y-x_i}{h}\right\|^2\right]\delta[\mathrm{b}(\mathrm{x}_i)-\mathrm{u}] \qquad (2)$$

The pixel locations of the object's condition region, centred at y in same frame, are denoted by $\{x_i\}$, $i=1,\ldots,n_h$. k is the kernel function with radius h.

Now relation between these two colour probabilities can be calculated through similarity function $\rho_1(y)$ [8] or Bhattacharyya coefficient [9, 10] $\rho_2(y)$ by following (3) and (4).

$$\rho_1(\mathrm{y}) \equiv \rho[p(\mathrm{y}),\mathrm{q}] = \sum_{u=1}^{m}\sqrt{q_u / p_u(\mathrm{y})} \qquad (3)$$

$$\rho_2(\mathrm{y}) \equiv \rho[p(\mathrm{y}),\mathrm{q}] = \sum_{u=1}^{m}\sqrt{p_u(\mathrm{y})\mathrm{q}_u} \qquad (4)$$

Now, the patch for the candidate model will be shifted in the direction of mean shift in an iterative way to maximize the similarity parameter. When similarity parameter will be maximized, that patch will be the new position for the object in the current frame.

A. Manual or Semi Automatic Technique

The traditional mean shift algorithm is a manual procedure in which a moving patch, which has to be tracked in successive frames, is selected manually. In case of surveillance the tracking should be automatic and target oriented. The moving pixels in a frame are detected by simple background subtraction techniques but, we can't say whether that moving object is a human or not. Therefore, traditional method is not a fully automated human tracking technique.

B. Fast motion of object

To work with mean shift algorithm it should be kept in mind that the motion of the object should not be very large. Because if the patch containing object will not have any overlapping area in the current frame with respect to previous frame, mean shift algorithm fails.

C. Prolonged Occlusion

In addition to the above problem if the object is having occlusion or there is sudden change of illumination, the algorithm fails as the similarity function becomes considerably low. Probabilistic approach can solve the occlusion to some extent [2, 11]. But, if the object is going for a prolonged period of occlusion and in that period the information about the object can't be estimated, tracking fails.

III. PROPOSED FRAMEWORK

Flow chart of the proposed automated human tracking using advanced mean shift algorithm is shown in Fig. 1. The proposed method helps the traditional method to start the tracking automatically by providing a patch containing human. Background of the frame will be ignored to get the moving objects by background subtraction. Human detection procedure detects the movable human and selects the patch for mean shift tracking. During the tracking procedure, if similarity parameter goes below 0.05, block matching technique will be initiated which helps to continue the tracking if it has been stopped due to fast motion. If block matching fails to make the tracking continue, it will be concluded that occlusion is there and so, occlusion handling will be continued. In this technique, background subtraction and human detection will be done in an iterative way till the human is not redetected after occlusion. After detection of human, the patch will be compared with the previous patch containing human before occlusion through similarity parameter. If similarity parameter will be found more than 0.1, tracking restarts from that patch by mean shift algorithm, otherwise process ends.

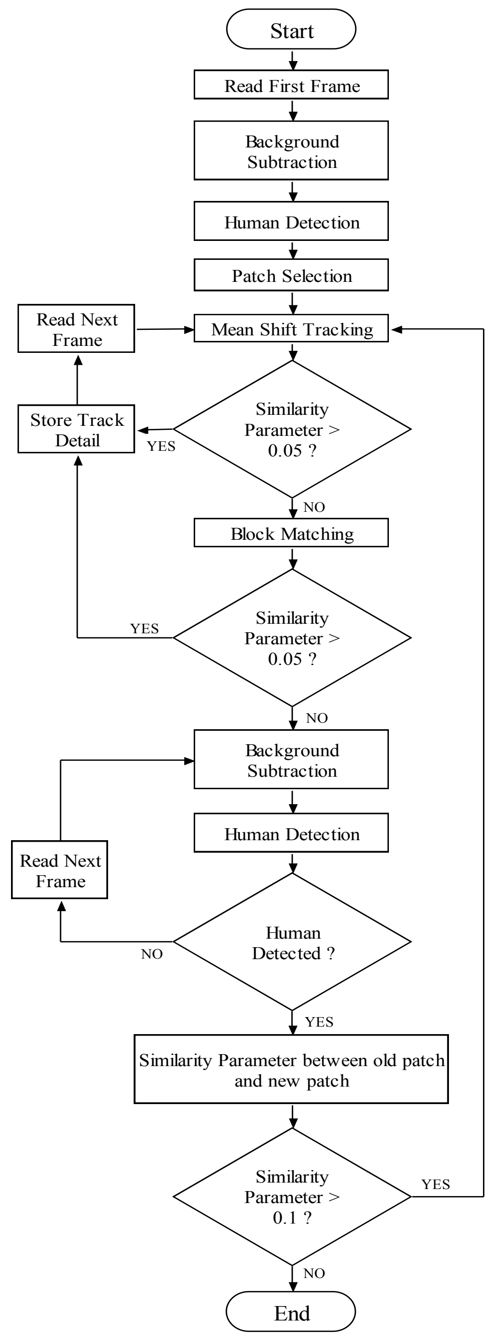


Fig. 1. Flow chart for Advanced automated mean shift tracking

Details about the proposed techniques are explained as follows:

A. Background Subtraction

Simple background subtraction [12] is used for the proposed method. The first frame is considered as background. All the frames of the video sequences are subtracted from the background frame. A threshold value is set to subtract background from the moving object based on the binary hypothesis. To discard noise from the background subtracted binary frame, morphological operations like erosion, dilation, and hole filling are used. Now, the frame having moving objects will be put into human detection procedure to ensure the presence of human.

B. Human Detection

Human modeling is requied for human detection and it is necessary to choose the patch automatically to start the tracking procedure.

It can be done by star skeletonisation method as in [13] to detect the human and separate it from all other moving objects. The method is based on distance function which means distance of boundary points from the centroid in clockwise direction. After getting the distance function, the curve will be smoothened by median filtering and then local maxima will be calculated. The points, where the maxima will be found, will be joined with the centroid to get a star like structure. By analyzing the star skeleton, it can be concluded whether that moving body is human or not. Skeleton can be analyzed like distance of head and legs from centroid are relatively same, angle of legs with the line joining centroid to head is greater than 150 degree in maximum cases.

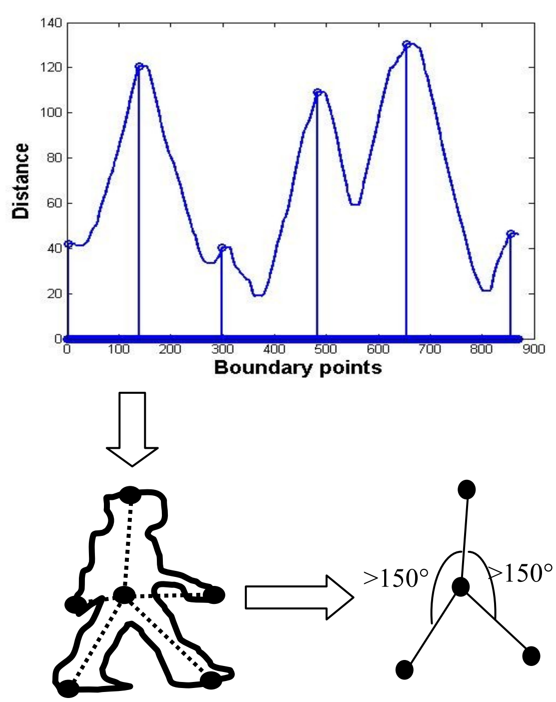


Fig. 2. Star skeletonisation method for Human modelling

C. Block Matching Technique

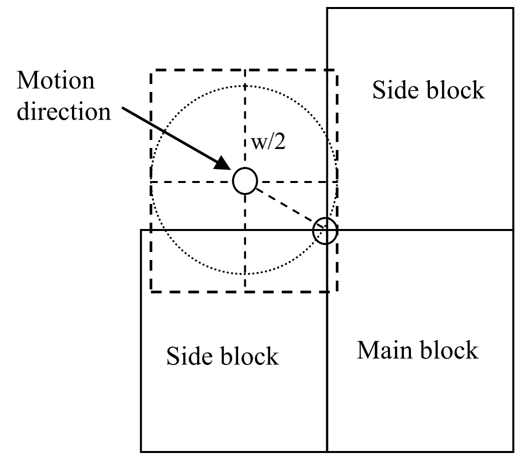


Fig. 3. Forward blocks for block matching in the motion direction

Sometimes the object may have fast motion for which mean shift tracker may not get an overlapping area to continue. At this moment similarity function goes below certain threshold and so tracking will be stopped. This is the time when block matching procedure will be initialized. Since block matching is a simple and relatively fast process, it affects the execution time negligibly and it can be called whenever tracking fails to ensure whether it is due to fast motion or not.

In block matching technique, motion direction vector will be calculated first which can be done from the centroids of previous two frames. After getting the direction, a shifted center will be chosen as in Fig. 3 and by referencing that point, three blocks will be chosen in the direction of motion. The shifted center position is $w/2$ pixels away from the original centroid position, where w is the width of the patch containing the object.

Now similarity functions will be calculated between object block and three new blocks. There are three blocks: one main block and two side blocks. Generally main block is the block, which may contain the shifted object in most cases. If the object has changed direction of motion suddenly, then one of the side blocks will contain the object. The block which will have maximum similarity with the object model will be upgraded as candidate model to continue the mean shift algorithm.

D. Occlusion Handling

If there is full occlusion for a long time, it is difficult to get the track after occlusion by traditional deterministic and Probabilistic methods. It may happen that the object will halt its motion during occlusion for which probabilistic trackers fail to estimate the position of the object. Sometimes it may happen that the object before to the occlusion and after the occlusion may be different i.e. a new object of similar type will come out. The proposed method so applied which tracks the object after the prolonged occlusion irrespective of motional characteristics variation during prolonged occlusion.

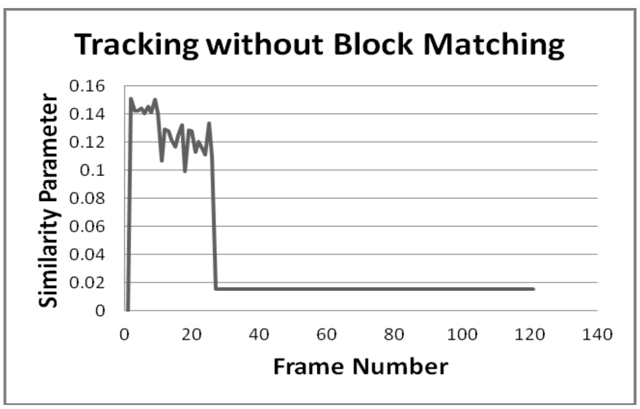
In this approach when the object is being occluded, tracking will stop till the object is not coming out of that occlusion. Tracking is not needed when the object is going for occlusion because during that period we can't get information about the object. The important thing is that after occlusion the tracking for that object should start automatically. It is done by putting the object into detection procedure to conclude whether that is a human object or not. After detection the object model will be compared with the object model before occlusion by similarity parameter to know whether the same object is coming out of the occlusion or not and if same object is detected, then tracking continues through mean shift procedure.

IV. Results And Discussion

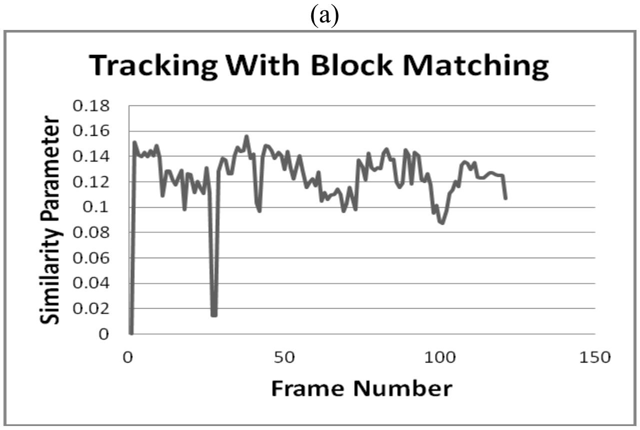
To conduct the experiment, real time videos have been collected by a panasonic lumix dmc lz-30 camera with the dimensions 1280× 720, 30fps. Two types of various videos have been taken: one, where objects fails to have overlapping area in next frame and another which is having prolonged occlusion. Test1, Test2, Test3, Test4 are four test videos having different length of prolonged occlusion.

A. Block Matching:

Tracking fails for fast moving objects where there is no overlapping patch for mean shift tracking.



(a)



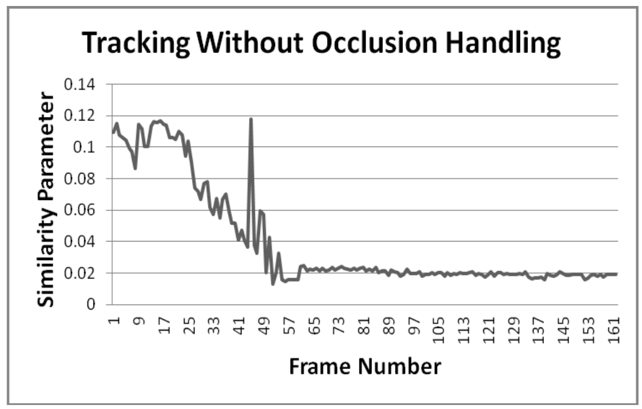
(b)

Fig. 4. Similarity parameter vs. frame number curves for (a) normal mean shift tracking (b) mean shift with block matching

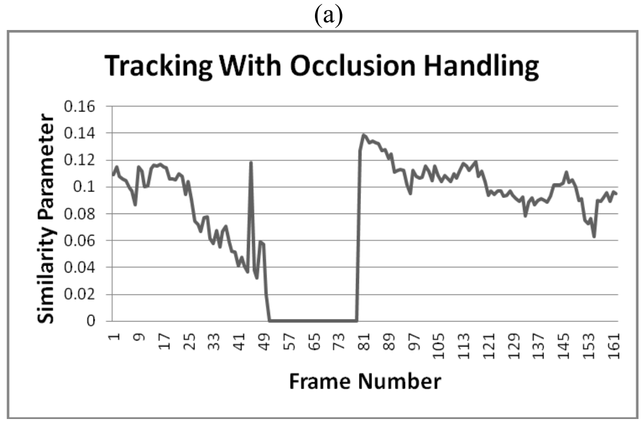
In Fig. 4(a), it is clearly shown that the similarity parameter goes below 0.02 and remains there for further frames which concludes the failure of traditional technique. Block matching technique had supplied an overlapping area to continue the track whose success has been reflected in Fig. 4(b) where similarity parameter has become considerably high.

B. Occlusion Handling

The Fig. 5 is for test video Test1 in which occlusion occurs at 158th frame and so similarity parameter has gone below 0.02 and remains at that level for further frames as shown in Fig. 4(a). Proposed algorithm stops the tracking when similarity parameter goes below 0.05 as shown in Fig. 5(b) where tracking has been stopped for occlusion from 160-189th frame. After that, again tracking starts by background subtraction and human detection and it continues till the end of the database as in Fig. 5(b).



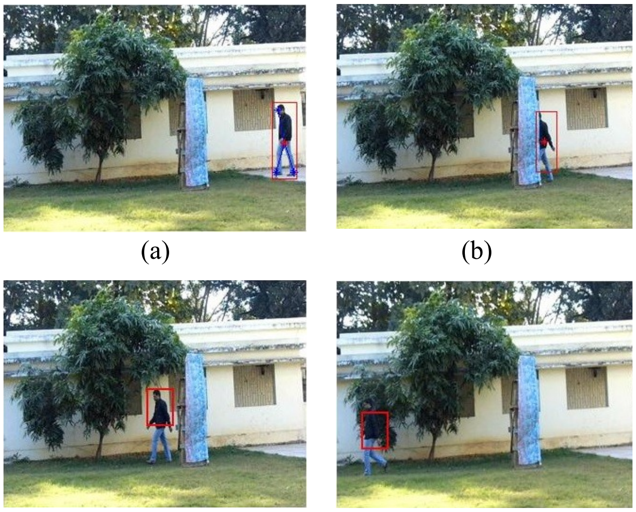
(a)



(b)

Fig. 5. Similarity parameter vs. frame number curves for (a) tracking without Occlusion Handling (b) tracking with Occlusion Handling. (curves have been shown excluding first 110 frames)

The real time frames of test video Test1 has been shown in Fig. 6, in which the red colour rectangle shows the success of tracking the human in occlusion condition. The human detection has been indicated in Fig. 6(a) and the starting of tracking, Fig. 6(b) is the last frame before tracking fails due to prolonged occlusion, Fig. 6(c) shows the continuation of tracking by proposed method after occlusion, and the Fig. 6(d) is an arbitrary frame which shows still the tracking is continuing after occlusion.



(a) (b) (c) (d)

Fig. 6. Human tracked frames of the video. (a) Human detected and start of tracking, (b) Tracked frame before occlusion, (c) Tracked frame after occlusion, (d) Tracked frame toward end of frame.

Similarly, the Table 1 shows the success of tracking for four different videos having occlusion problem in which traditional method fails to track. From the table it is clearly understood that tracking stops after 2-5 frames of starting of occlusion and again restarts after 5-10 frames after occlusion. It is due to the degradation of similarity parameter value till threshold and proper human detection after occlusion respectively.

TABLE I. PERFORMANCE OF PROPOSED METHOD

Test videos	**Test1**	**Test2**	**Test3**	**Test4**
Number of frames	300	495	300	345
Occluded frames	158-184	134-180	123-175	163-210
Failed at frame	160	137	127	167
Track continued by proposed method at frame	189	185	185	220
Tracked till frame	272	465	295	290

Limitations of the technique are that it is unable to track during occlusion. It needs an efficient background subtraction for human detection. Tracking cannot be started during partial occlusion.

V. CONCLUSION

This paper presents an efficient approach to track single human by advanced mean shift tracker. The human position and patch to start tracking is selected by human detection through star skeletonisation. Separate databases are used to study the fast motion and occlusion problems. Block search algorithm has provided overlapping patch to continue the track in case of fast motion. Occlusion handling is successfully applied and executed to prolonged period of occlusion. Results show that the proposed method is efficient for real time tracking where the tracking starts automatically and handles the obstacles like fast motion and prolonged occlusion. The method can be applied to any real time environment and crowded areas for surveillance.

REFERENCES

[1] Wen J, Gao X, Yuan Y, Tao D, JieLi, “Incremental tensor biased discriminant analysis: a new color-based visual tracking method”, *Journal of Neurocomputing*, pp. 827–839, 2010

[2] A. H. Mazinan, A. Amir-Latifi, “Applying mean shift, motion information and Kalman filtering approaches to object tracking,” *ISA Transactions,* Vol. 51, Issue 3, pp. 485-497, 2012

[3] Zivkovic Z, Kröse B., “An EM-like algorithm for color-histogram-based object tracking”. *IEEE Conference on Computer Vision and Pattern Recognition*, Vol. 1, pp. 798-803, 2004

[4] Z. Wen and Z. Cai, “A Robust Object Tracking Approach using Mean Shift”, *3rd International Conference on Natural Computation*, Vol. 2, pp. 170-174, 2007.

[5] J. Jeyakar, R. V. Babu and R. K. Ramakrishnan, “Robust Object Tracking using Kernels and Background Information,” *IEEE International Conference on Image Processing*, Vol. 5, pp. 49-52, 2007.

[6] F. Tung, J. S. Zelek, and D. Clausi, “Efficient target recovery using stage for mean-shift tracking”, In *Canadian Conf. Computer and Robot Vision*, pp. 16-22, 2009.

[7] M Zhao, L Wang, J Han, “An Adaptive Tracking window Based on Mean-Shift Target Tracking Algorithm”, in *Chinese Automation Congress*, pp. 348 – 352, 2013

[8] H Zhi-qiang, L. Xiang, Y. Wang-sheng, L. Wu, Huang An-qi, “Mean shift tracking Algorithm with improved Background weighted Histogram”, in *5th International Conference on Intelligent Systems Design and Engineering Applications*, pp. 597 – 602, 2014

[9] Comaniciu D, Ramesh V, Meer P. , “Real-time tracking of non-rigid objects using mean shift”, *IEEE Conference on Computer Vision and Pattern Recognition*, vol. 2, pp. 142–149, 2000.

[10] Chen J J, An G C, Zhang S F, et al., “A mean shift algorithm based on modified Parzen window for small target tracking”, *IEEE International Conference on Acoustics, Speech and Signal Processing*, pp. 1166–1169, 2010

[11] A. Ali and K. Terada, “A framework for human tracking using kalman filter and fast mean shift algorithms,” *IEEE Intl. Conf. Computer Vision Workshops*, pp. 1028-1033, 2009.

[12] Y. Benezeth, P. Jodoin, B. Emile, H. Laurent, and C. Rosenberger, “Review and evaluation of commonly-implemented background subtraction algorithms”, *In IEEE 19th International Conference on Pattern Recognition*, pp. 1-4, 2008.

[13] N. Zarka, Z. Alhalah and R. Deeb. "Real-Time Human Motion Detection and Tracking," *3rd International Conference on ICTTA*, pp. 1-6, 2008.