This full text paper was peer-reviewed at the direction of IEEE Instrumentation and Measurement Society prior to the acceptance and publication.

Reliability of Bench-Mark Datasets for Crowd Analytic Surveillance

Mohamed Abul Hassan, Aamir Saeed Malik[1], Walter Nicolas, Ibrahima Faye, Nadira Nordin
Centre for Intelligent Signal and Imaging Research (CISIR),
Universiti Teknologi PETRONAS, Malaysia,
[1]aamir_saeed@petronas.com.my

***Abstract*— This paper presents an evaluation on the reliability of the bench-mark datasets for outdoor crowd analytic surveillance systems. The credibility of the databases are assessed based on their diverseness to yield challenges of dynamic environments. The main object of this paper is to assess the challenges imposed by the databases for sudden illumination variance and effect of wavering trees. Two bench-mark databases, PETS 2010 and OTCBVS, along with our proposed dataset are evaluated using the three most popular background modelling algorithms in crowd analytic surveillance; Approximate Median Method, Gaussian Mixture Model and Codebook. The diverseness of these databases are assessed, with respect to the performance of the basic algorithms using qualitatively and quantitatively. Eventually the reliability of these bench-mark databases for outdoor crowed analytic surveillance is assessed.**

Keywords*— Dynamic Background; Crowd Analytic Surveillance; PETS2010; OTCBVS; Wavering Trees

I. Introduction

This visual surveillance for crowd analytics has become one of the most attractive research areas in the fields of computer vision and pattern recognition. The goal of visual surveillance system is to extract information from video footages collected by the visual surveillance cameras [1]. The availability of low cost sensors, processors and the need for safety and security at public sectors are the main reasons for this emerging interest in research for visual surveillance systems. Visual surveillance systems generally consist of four main steps; image acquisition, image pre-processing, background modelling and behaviour understanding [2].

Therefore making background modelling a key step towards behaviour understanding applications of crowd analytic surveillance systems. Background modelling, comprises of foreground/background segregation to provide object measurements for behaviour understanding applications [3, 4]. Basic background modelling methods would operate by performing a basic subtraction between two adjacent frames. Nevertheless, background modelling at dynamic environments requires more diversity to adapt to various scenarios challenged by the environment.

Some of the challenges background modelling methods has to overcome while operating at dynamic environments are noise due to poor image quality, camera jitter, bootstrapping, camouflage, foreground aperture, moving background objects, waking foreground objects, wavering trees, water ripples, water surface and illumination variance [5, 6]. Researches in the past have developed many background modelling methods, out of which Approximate Median Method (AMM) [7], Gaussian Mixture Model (GMM) [8], Kernel Density Estimation (KDE) [9], K Mean Clustering (KMC) [10] and Code Book (CB) [11] are the most commonly used methods for crowd analytic surveillance systems.

These background modelling methods have overcome most issues such as; poor image quality, camera jitter, bootstrapping, camouflage, foreground aperture, moving background objects, waking foreground objects and wavering trees. However, issues related to illumination variance, shadows, moving background objects, waking foreground objects have persisted as shown in Fig. 1. Most of recent research attention has attempted to solve the issues of illumination variance and shadows elimination.

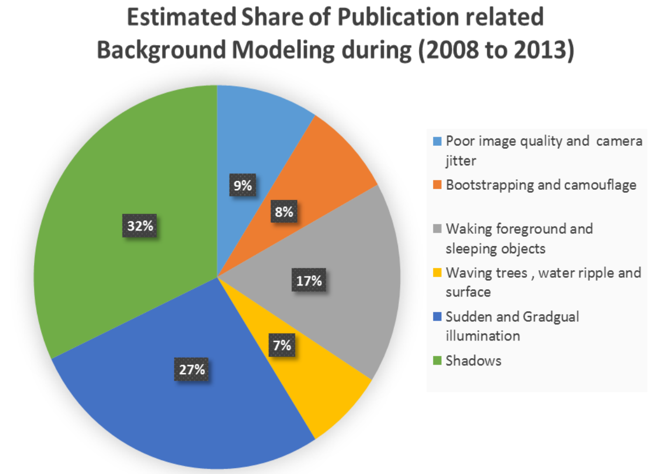

Fig. 1. Estimated share of publications related to background modeling for the past 5 years

In the developmental phase of an algorithm, bench mark datasets are tested to evaluate their credibility. For the application of crowd analytic surveillance, the computer vision community has formulated many benchmark data bases which would deduce the above mentioned issues. PETS2010 [12], PETS2006 [13], WallFlower datasets [14], ATON dataset [15], AVSS 07 [16], and OTCBVS [17] are some of the largely used as bench-mark datasets.

In this paper we will assess the credibility of the bench-mark datasets along with our proposed dataset. The rest of the paper is organized as follows; Section II describing

978-1-4799-6144-6/15/$31.00 ©2015 IEEE

the bench-mark datasets, section III describing the proposed dataset, experimental results and analysis are presented in section IV and finally the finding of this study is drawn in to conclusion in section V.

II. Benchmark Dataset

On evaluation, the credibility of background modelling algorithms, are tested on bench-mark datasets to provide a general baseline to compare the performance and effectiveness of the algorithms. Here we have discussed two publicly used bench-mark datasets, PETS2010 and OTCVS. We have excluded other databases since they did not meet the criteria to access background modelling algorithms for dynamic environments.

A. PETS2010

The PETS 2010 dataset is especially developed for visual surveillance research on crowd behaviour, in which the data sub set S0, is used for evaluating background modelling algorithms. This data sub set contains three types of crowd sequences to evaluate background models. Firstly background; here the background model for multiple camera views is obtained. The frames may contain people or other moving objects. Secondly city centre; this dataset includes random walking crowd flow resembling a sparse crowd. Lastly, regular flow; here the dataset includes a regular walking pace crowd flow which resembles a dense crowd behaviour.

Fig. 2. Sample frames resembling various background modeling challenges; (a) and (b) resemble single object detection under gradual illumination variance. (c) Resembles a high illumination variation condition corresponding to the frame (b). (d) and (e) resemble dense crowd motion under normal weather conditions while yielding the effects of camouflage and bootstrapping. (f) resembles sparse crowd behavior under heavy illumination variance and wavering trees. (g) and (h) resemble dense crowd behavior under camera jitter effect. (i) yielding sudden illumination variance and shadow effect on sparse crowd scenarios.

Each of the above mentioned crowd sequences are captured by multiple camera views during various times of the day. The dataset is captured at outdoors, with a natural background scene of trees, grass and buildings. The recording at various times of the day, clearly indicates the variation of weather and lighting conditions. The sample frames resembling various background modeling challenges for this data base is shown in Fig. 2.

The significance of these datasets, apart from testing the models on different crowd behavior, tests the models on detecting small objects from pixel range of 10×35 to 20×70, which challenges the accuracy of segmentation. The unsynchronized nature of the dataset adds the effect of sudden and gradual illumination variance continuously at every 30 to 40 frames. The natural lighting condition also deduces a strong shadow backdrop of foreground and background objects, such as buildings and trees. The dataset overcomes all the above mentioned challenges, except for water ripples, and water surface.

B. OTCBVS

The OSU color-thermal database was compiled on inspecting the busy pathways of Ohio State University. The database consists of 6 sequences obtained from 3 different locations at the campus. The dataset comprised of sparse crowd scenes which provided a random motion. The database focuses on the issues of accurate segmentation of small objects i.e. human silhouettes, sudden and gradual illumination variance and shadows.

The entire crowd dataset consisted of 17089 sample frames. The frames were acquired from a long range camera view, which provided a border aspect of the background filled with the Ohio State University building pathway. This correspondingly increases the challenge of accurate object segmentation, camera jitter, gradual and sudden illumination foreground and background object shadows and shadows of the clouds in the sky. Sample images of various sequences are shown in Fig.3.

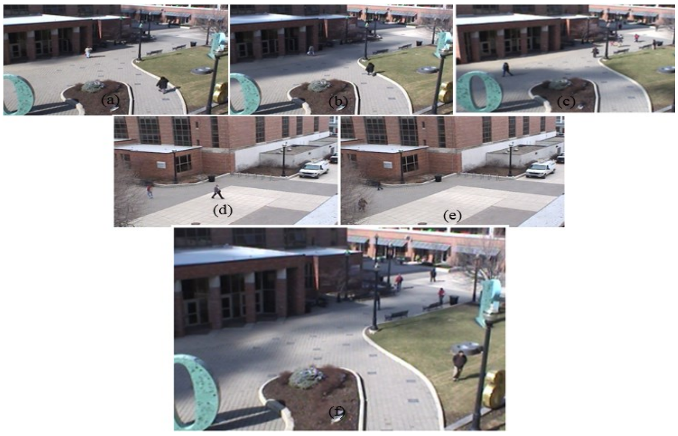

Fig. 3. Sample frames resembling various background modeling challenges; (a) resembles single object detection under gradual illumination variance yielding foreground and background object shadows. (b) and (c) yield the challenge of sudden illumination change due to shadows of clouds in the sky. (d) and (e) shows the issue of boots strap and gradual illumination variance. (f) is effected by sudden illumination, shadows of clouds and camouflage.

C. Crowd Analytic Dataset (CAD)

The proposed dataset was compiled at the bridgeway of the CISIR research center (See Fig. 4). We have compiled 10 scenarios of crowd behavior, including sparse and dense crowd motion for scenarios such as bottleneck, departure, lane, arch/ ring and blocking. The scenarios were captured using 5 camera views providing the database with 90,000 sample frames.

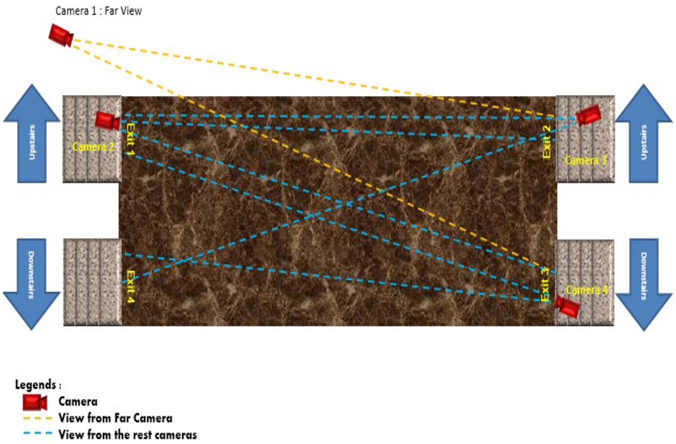

Fig. 4. Description of the camera placement to compile the database. Five camera views are used and the field of view for each camera is illustrated by dotted lines.

The data set specializes in providing a bench mark for crowd surveillance related application. Here, the dataset was created mainly to resolve the challenges of illumination variance and wavering trees. The illumination variance was continuous in the background of the scene, where different parts of the background were affected. The back drop of this dataset is combined with buildings, trees and mild textures of grass. The database was challenged in detecting objects from scale of 10×25 to 25×60. The mild motion of tree leaves in the background, given the issue of wavering trees and natural variations of lighting conditions, steered the issue of illumination variance and shadows (See Fig. 5). Sample frames of various scenarios for different camera views are shown in Fig. 6.

Fig. 5. Description of the scene of the dataset. Here different regions are separated by ovals and arrows. The green arrow denotes the bridge way of the CISIR research center. The red ovals denote the areas in the backdrop of the scene which are more prone to yield the effect of illumination variance. The yellow rectangle shows the area where, object detection in depth is challenged. The double headed brown arrows show the trees which addresses the challenge of wavering trees.

Fig. 6. Sample frames resembling various background modeling challenges; (a) to (c) are for single object detection for challenges of sleeping foreground objects, camouflage, depth object detection and gradual illumination. Sample frames (d) to (f) resemble random motion of sparse crowds, heavy illumination variance and different motion scenarios. Frames (g) to (i) are sample frames of dense crowd, performing arch ring, bottleneck and lane motion scenarios under heavy illumination variance in the background of multiple camera views.

III. Experiment Result

The experimentation phase was carried out qualitatively and quantitatively using Precision and Recall [18]. The databases were tested with three popular background modelling methods such as approximate median method (AMM) [7], Mixture of Gaussian method (MoG) [19] and Code book (CB) [20]. These algorithms were implemented on Intel Core i7 processor with NVIDIA GeForce GT650 4GB Graphic card and a 4GB DDR3 RAM. The algorithms were tested for different scenarios on PETS2010, OTCBVS and the proposed dataset (See Fig. 7). The final results of the algorithm were addressed based on the binary map of the extracted foreground.

The obtained binary map was evaluated, with respect to the ground truth result for the corresponding frame. The objective of this study is to evaluate the credibility of the datasets for yielding the challenges of illumination variance and effect of wavering trees. Specific datasets of each database were selected, so that the algorithms would be challenged by these issues. In PETS 2010, we used the camera views 1, 2 and

4 of the dataset S0. In OTCBVS, we used the Dataset 03; sequence 1 and 4. For the proposed dataset, we have used sequence 4 and 9 for camera view 1. The detail description of all the datasets are mentioned in Table. 1.

TABLE I. DESCRIPTION OF THE DATASETS

Image sequence	Database	No. of frames	Size	Challenges	Description
Sequence 1	OTCBVS	1507	320×240	BS,SI,GI	Dataset 03: OSU Color-Thermal Database
Sequence 2	OTCBVS	1054	320×240	GI,BS,S	Dataset 03: OSU Color-Thermal Database
Sequence 3	PETS2010	841	768×576	BS,GI,CF	Dataset of S0, City Center, view 4,
Sequence 4	PETS2010	841	768×576	SI,GI, WT, S	Dataset of S0, City Center, view 2,
Sequence 5	PETS2010	841	768×576	SI,GI, WT, S	Dataset of S0, City Center, view 1,
Sequence 6	CAD	1870	640×480	SI,GI, CF, WT	CISIR bridge way, sequence 4 view 1
Sequence 7	CAD	780	640×480	SI,GI, WT, S	CISIR bridge way, sequence 9 view 1

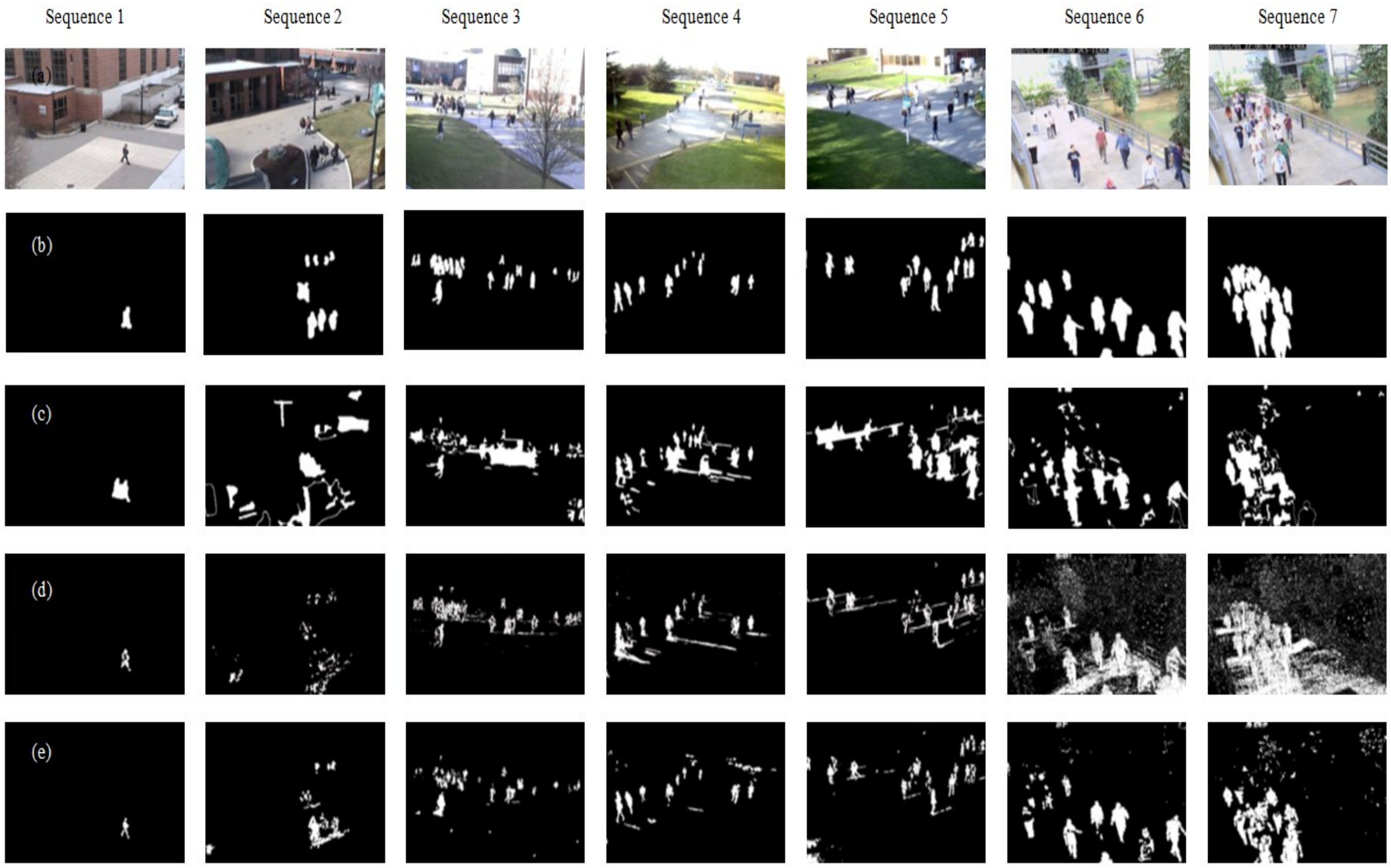

Fig. 7. Sample frames of the extracted binary silhouettes for background modeling algorithms for different sequences. (a) is the original frame; (b) ground truth image for the original frame; (c) binary silhouette for AMM; (d) binary silhouette for MoG and (e) extracted binary silhouette for CB,

The three most often used background modelling algorithms were experimented for the 7 sequences compiled from 3 databases. The main object of the experiment was to determine the effect of sudden illumination variance and wavering trees. The sequences variably challenged the background modelling algorithms under different environmental circumstances. Sequence 1 and 2 developed by OTCBVS, yielded the issue of sudden illumination variance where majority of the algorithms performed satisfactory by achieving a high precision recall (see Fig. 8). The sequences 3, 4 and 5 of PETS 2010, contrarily challenged the algorithms with far more diverse effects of illumination variance, while yielding a great threat of shadows. As shown in the sample frames in Fig. 7, the algorithms comparatively failed to compensate for the effect of sudden illumination and resulted in producing a poor recall rate due to the higher rate of true negative pixel extraction.

The sequences 3 and 4 also show the effect of wavering trees, while examining the binary silhouette of the foreground. It is clearly seen that the issue of wavering trees didn't challenge the algorithms in extracting false foreground pixels. The results obtained from the sequences of our proposed dataset differed greatly. However, this resulted in a poor recall rate compared to other datasets. The algorithms clearly were challenged by the diverse environment in the background and failed to extract true positive foreground pixels with the same efficiency shown in the other sequences.

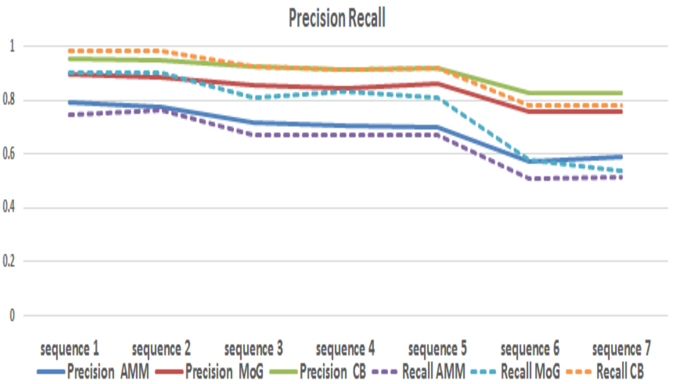

Fig. 8. Quantified Precision Recall results of three background modeling algorithms for seven different sequences

Analysing qualitatively through the extracted foreground binary mask. The poor recall rate was produced largely due to the effect of wavering trees and the effect of illumination variance was similar to that of PETS 2010. The challenge of wavering trees along with the illumination variance was the main issue yielded by our proposed dataset. The results obtained from PETS 2010 and our proposed datasets were highly contradictory. Therefore, we illustrate this matter by observing the pixel movement of the trees of sequences 4 and 6 (See Fig. 9). This deduced the reason of the effect of wavering trees. The continuous motion of the trees by a few pixels to the left and to the right of the frame caused these background modelling algorithms which are based on motion variation for extracting foreground objects, to extract these trees.

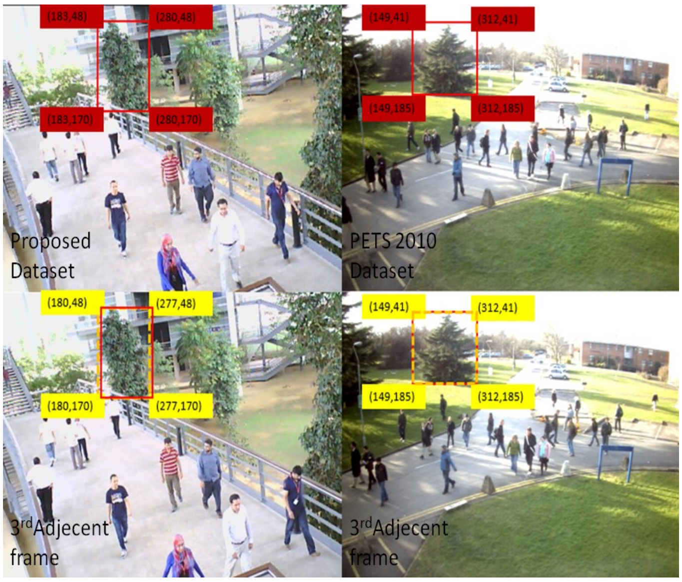

Fig. 9. Illustrates the effect of the wavering tree scenario where frame of the proposed dataset and the PETS 2010 are compared. The 3rd adjacent frame from each dataset is used to compare the motion variation of the pixel in the background. For the PETS 2010 it's clear from the rectangular boxes that there is no motion of the trees from the adjacent frames. However, the tree in the adjacent frames of the proposed dataset has shown a deviation of 3 pixels to the left yielding a motion variation of the background.

IV. Conclusion

In this paper we have proposed a dataset for background modelling for the evaluation of background modelling algorithms in diverse dynamic environments. Here we investigated two popular databases PETS2010 and OTCBVS along with our proposed dataset. We evaluated the reliability of these databases for the three most often used background modelling algorithms (i.e. AMM, MoG, CB). The results showed that the algorithms failed to compensate for the effect of the wavering trees in our proposed datasets, meanwhile efficiently compensating the effect in other databases. The databases were further investigated in comparison to our proposed dataset. They were able to justify a reason for the variation of the results from the existing datasets to the proposed dataset.

Acknowledgment

This work is supported by PRGS grant (Ref: 1/12/TK02/UTP/02/05), Ministry of Education (MoE), Malaysia

References

[1] X. Wang, "Intelligent multi-camera video surveillance: A review," Pattern Recognition Letters, vol. 34, pp. 3-19, 2013.

[2] M. Szczodrak, J. Kotus, K. Kopaczewski, K. Lopatka, A. Czyzewski, and H. Krawczyk, "Behavior Analysis and Dynamic Crowd Management in Video Surveillance System," in Database and Expert Systems Applications (DEXA), 2011 22nd International Workshop on, 2011, pp. 371-375.

[3] O. J. Alper Yilmaz, Mubarak Shah, "Object Tracking: A Survey," ACM Comput. Surv, vol. 38, p. 45, Dec. 2006 2006.

[4] M A Hassan, Aamir S Malik, Ibrahima Faye, Nicolas Walter, Waqas Rasheed, Nadira Binti Nordin, "Foreground Extraction for Real- time Crowd Analytics in Surveillance System," presented at the 2014 IEEE 18th International Symposium on Consumer Electronics (ISCE 2014), 2014.

[5] T. Bouwmans, "Recent Advanced Statistical Background Modeling for Foreground Detection - A Systematic Survey," 2011.

[6] M A Hassan, Aamir S Malik, I. Faye, N. Walter, T Mahmood "Mixture of Gaussian Based Background Modelling for Crowd Tracking Using Multiple Cameras," presented at the The 5th International Conference On Intelligent & Advanced Systems (ICIAS 2014), 2014.

[7] F. El Baf, T. Bouwmans, and B. Vachon, "Comparison of Background Subtraction Methods for a Multimedia Application," in Systems, Signals and Image Processing, 2007 and 6th EURASIP Conference focused on Speech and Image Processing, Multimedia Communications and Services. 14th International Workshop on, 2007, pp. 385-388.

[8] C. Stauffer and W. E. L. Grimson, "Adaptive background mixture models for real-time tracking," in Computer Vision and Pattern Recognition, 1999. IEEE Computer Society Conference on., 1999, p. 252 Vol. 2.

[9] D. H. A. Elgammal, and L.S. Davis, , "Non- parametric model for background subtraction," Proc. ECCV 2000, pp. 751-767, 2000.

[10] D. Butler, S. Sridharan, and V. M. Bove, Jr., "Real-time adaptive background segmentation," in Acoustics, Speech, and Signal Processing, 2003. Proceedings. (ICASSP '03). 2003 IEEE International Conference on, 2003, pp. III-349-52 vol.3.

[11] Kyungnam Kima, Thanarat H. Chalidabhongseb, David Harwooda, Larry Davis, "Real-time foreground–background segmentation using codebook model," Real-Time Imaging, p. 14, 2006.
[12] J. Ferryman and A. Ellis, "PETS2010: Dataset and Challenge," in Advanced Video and Signal Based Surveillance (AVSS), 2010 Seventh IEEE International Conference on, 2010, pp. 143-150.
[13] "PETS," [online] 2006, http://pets2006.net/ (Accessed: 14 May 2014).
[14] "Wall Flower. Dataset," [online] 2008, http://research.microsoft.com/users/jckrumm/WallFlower/TestImages.htm (Accessed: 14 May 2014).
[15] "ATON Dataset," [online] 2007, http://cvrr.ucsd.edu/aton/shadow/ (Accessed: 14 May 2014).
[16] "AVSS," [online] 2007, http://www.eecs.qmul.ac.uk/~andrea/avss2007.html (Accessed: 14 May 2014).
[17] "Ohio State Univeristy Visual and Termal Camera Surveillacne Dataset,"[online], 2011, http://www.cse.ohio-state.edu/otcbvs-bench/ (Accessed: 14 May 2014).
[18] D. M. W, “Evaluation: From Precision, Recall and F-Factor to ROC, Informedness, Markedness & Correlation," Journal of Machine Learning Technologies, vol. 2, p. 6, 2011.
[19] L. Xiaochun, Z. Tao, and F. Dan, "Robust compositional method for background subtraction," in Control Automation Robotics & Vision (ICARCV), 2012 12th International Conference on, 2012, pp. 1419-1424.
[20] T. Malathi and M. K. Bhuyan, "Multiple camera-based codebooks for object detection under sudden illumination change," in Communications and Signal Processing (ICCSP), 2013 International Conference on, 2013, pp. 310-314.