

Moving Object Detection in Highly Corrupted Noise using Analysis of Variance

Asim ur Rehman Khan¹, Muhammad Burhan Khan², Haider Mehdi³, Syed Muhammad Atif Saleem⁴

Multimedia Lab

National University of Computer and Emerging Sciences (NUCES)
Pakistan

Abstract—This paper implements three-way nested design to mark moving objects in a sequence of images. Algorithm performs object detection in the image motion analysis. The inter-frame changes (level-A) are marked as temporal contents, while the intra-frame variations identifies critical information. The spatial details are marked at two granular levels, comprising of level-B and level-C. The segmentation is performed using analysis of variance (ANOVA). This algorithm gives excellent results in situations where images are corrupted with heavy Gaussian noise $\sim N(0, 100)$. The sample images are selected in four categories: ‘baseline’, ‘dynamic background’, ‘camera jitter’, and ‘shadows’. Results are compared with previously published results on four accounts: false positive rate (FPR), false negative rate (FNR), percentage of wrong classification (PWC), and an F-measure. The qualitative and quantitative results prove that the technique out performs the previously reported results by a significant margin.

Keywords—Analysis of variance (ANOVA); image motion analysis; object detection

I. INTRODUCTION

Tracking of objects in a sequence of images has several applications in robotics and computer graphics. A few real life applications are security, traffic control, medical applications, animation, and automation. The broader research area covers several scenarios, like object detection by having more than one camera, a moving camera, changing luminance level, etc. Here, however, only the simplest situation is considered with a single camera in a fixed position. The luminance of the surrounding area is assumed to be unchanged. Any change in the gray level is attributed to the moving object in the foreground, while the non-changing pixel values are considered as regions with the stationary background. There are generally two research directions to handle this situation; first is the pixel-based approach where each pixel is identified as either the part of moving object or the stationary background area. The second approach is region-based which considers a group of pixels for identifying regions with an overall common pattern.

One of the simplest pixel-based approach for identifying moving object is by taking the absolute difference of two consecutive frames [1]. In region-based approach, the pixel interrelations are also considered [2]. Recent research proposed several techniques of object detection. The object-detection using histogram based background subtraction and fuzzy logic is examined in [3] and [4]. A generic algorithm based on real-time traffic surveillance scheme has been

proposed in [5]. Reference [6] studied the use of K-nearest neighbor clustering method to identify vehicles that come dangerously close while driving. Several researchers have considered the robust principal component analysis (RPCA) as a method for object detection [7]-[10]. A Gaussian-based model is employed to identify moving objects in the presence of atmospheric turbulence in [11]. Other researchers proposed a hierarchical modeling [12] and saliently fused sparse decomposition approaches [13]. In addition [14] explores combined shape and feature-based non-rigid object tracking algorithm for object detection. All of the above approaches essentially considered noise-free images.

The real images are generally corrupted with various levels of noise. The noise is more prominent when raw images needed to be transmitted to another location where the processing is performed. This noise may have any distribution; but for simplicity it is assumed that the noise is additive, and have Gaussian distribution. This assumption has been proved to be valid through several tests. The analysis of variance (ANOVA) exhibits strong resilience under heavy Gaussian noise [15]. This paper extends a previously published paper [16]. In the previous paper statistical comparison was made to identify edges of an object. Here, the objective is to identify both the temporal information based on inter-frame statistical analysis, as well as the spatial details based on intra-frame information. A comprehensive mathematical background of ANOVA is presented in [17]. Consequently, this paper implements three-way nested design to segment image frames having both the temporal as well as the spatial information. The suggested algorithm performed effectively in both the noise-free situations as well as in the presence of heavy Gaussian noise of $\sim N(0, 100)$. The algorithm was validated through rigorous simulations performed on standard image sequence. Section II reviews the essential mathematical background for three-way nested design, Section III presents the simulation results. Section IV concludes this study.

II. THREE-WAY NESTED DESIGN

A three-way nested design comprises of levels A, B, and C such that the level-C is nested within level-B and the level-B is nested through level-A. This is given as A, B(A), and C(B). The level-A is based on inter-frame temporal information, whereas the level-B and level-C identify intra-frame pixel variations highlighting the important features. The level-B identifies regions with slowly moving objects, and the level-C

identifies motion at a smaller granular level. Fig. 1 explains the distribution of pixels into various levels. Following parametric model has been used:

$$\Omega: y_{ijkl} = \mu + a_i + b_{ij} + c_{ijk} + e_{ijkl} \quad (1)$$

where μ represents general mean of pixels. The inter-frame effect (factor-A) is represented by $a_i, i = 1, \dots, I$, the intra-frame feature (factor-B) is represented by $b_{ij}, j = 1, \dots, J_i$, and the sub-intra-frame feature (factor-C) is represented by $c_{ijk}, k = 1, \dots, K_{ij}$. Here, the value of I is equal to 2, while the values of J_i and K_{ij} are equal to 4. The parameters, a_i, b_{ij} , and c_{ijk} are unknown, having fixed-effect. The model, in essence, assumes no randomness in any of these parameters. All variations are included in the error, $e_{ijkl}, l = 1, \dots, L_{ijk}$. The value of L_{ijk} is 4. The error is assumed to have uncorrelated, Gaussian distribution $\sim N(0, \sigma^2 I)$.

The objective is to test the hypothesis H_Ω for the presence of a particular effect against the null hypothesis H_0 , where null hypothesis confirms the absence of that particular effect. The test is first performed for inter-frame feature a_i . In case the null hypothesis is rejected, then there is a sufficient justification to assume that a significant moving area is present within the mask region. Next, the selected image window is tested for the presence of sufficient variation at two gray levels B and C. This test is performed on single frame for intra-frame feature extraction.

Under the Ω -model of (1), the net effect is the sum of mean μ , various parameters a_i, b_{ij}, c_{ijk} , and error e_{ijkl} . The effect of mean and parameters is combined in η_{ijk} as,

$$\eta_{ijk} = \mu + a_i + b_{ij} + c_{ijk} \quad (2)$$

We impose the side conditions,

$$\begin{aligned} \sum_i u_i a_i &= 0 & \sum_j v_{ij} b_{ij} &= 0 \text{ for all } i \\ \sum_k w_{ijk} c_{ijk} &= 0 \text{ for all } i, j \end{aligned} \quad (3)$$

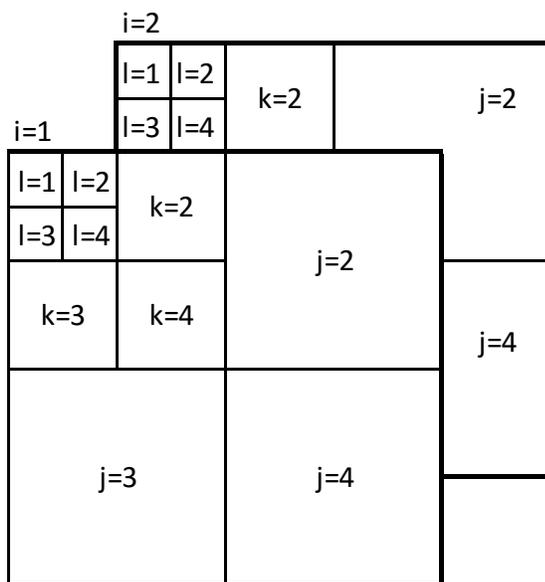


Fig. 1. Mask used for Effects A, B(A), and C(B).

where $\{u_i\}, \{v_{ij}\}, \{w_{ijk}\}$ are nonnegative weights such that

$$\sum_i u_i = 1; \sum_j v_{ij} = 1 \text{ for all } i; \sum_k w_{ijk} = 1 \text{ for all } i, j \quad (4)$$

The mean-square error is defined as Λ ,

$$\Lambda = \sum_i \sum_j \sum_k \sum_l^{L_{ijk}} (y_{ijkl} - \eta_{ijk})^2 \quad (5)$$

The objective is to minimize the parameter Λ with respect to each of the desired effects. To derive the sums of squares of various effects we consider the following hypothesis:

$$H_A : \text{all } a_i = 0; H_B : \text{all } b_{ij} = 0; H_C : \text{all } c_{ijk} = 0 \quad (6)$$

The regions Ω_A, Ω_B , and Ω_C corresponds to the intersection of the overall hypothesis space Ω with each of the specific hypothesis sub-spaces identifying various effects represented by parameters, a_i, b_{ij} , and c_{ijk} . n_{ijkl} represents each pixel. The value of L_{ijk}, n_{ij}, n_i , and n are equal to 4, 16, 64, and 128, respectively. The weights are found from,

$$w_{ijk} = L_{ijk}/n_{ij}; v_{ij} = n_{ij}/n_i; u_i = n_i/n \quad (7)$$

The dot notation in the subscript signifies aggregation or averaging over the index represented by the dot, like $y_{ijk.} = \sum_l y_{ijkl}$. A bar on $y_{ijk.}$ represents weighed sum, as $\bar{y}_{ijk.} = y_{ijk.}/n$. The least square (LS) estimate of parameter space, η_{ijk} is given by $\hat{\eta}_{ijk} = \hat{\mu} + \hat{a}_i + \hat{b}_{ij} + \hat{c}_{ijk}$, where $\hat{\mu}, \hat{a}_i, \hat{b}_{ij}$ and \hat{c}_{ijk} are LS estimates of the corresponding parameters, which are μ, a_i, b_{ij} and c_{ijk} . By adding the LS estimate $\hat{\eta}_{ijk}$ and subtracting the LS estimate of various parameters as given above, (5) is rearranged as,

$$\Lambda = \sum_i \sum_j \sum_k \sum_l [(y_{ijkl} - \hat{\eta}_{ijk}) + (\hat{\mu} - \mu) + (\hat{a}_i - a_i) + (\hat{b}_{ij} - b_{ij}) + (\hat{c}_{ijk} - c_{ijk})]^2 \quad (8)$$

By solving the above equation and also using the side conditions, the operator Λ is given by,

$$\begin{aligned} \Lambda = & \sum_i \sum_j \sum_k \sum_l (y_{ijkl} - \hat{\eta}_{ijk})^2 + n (\hat{\mu} - \mu)^2 + \\ & \sum_i n_i (\hat{a}_i - a_i)^2 + \sum_i \sum_j n_{ij} (\hat{b}_{ij} - b_{ij})^2 + \\ & \sum_i \sum_j \sum_k L_{ijk} (\hat{c}_{ijk} - c_{ijk})^2 \end{aligned} \quad (9)$$

TABLE. I. SUM OF SQUARE OF VARIOUS EFFECTS

SSA	=	$\sum_i n_i \bar{y}_{i...}^2 - n \bar{y}^2$
SSB	=	$\sum_i \sum_j n_{ij} \bar{y}_{ij.}^2 - \sum_i n_i \bar{y}_{i...}^2$
SSC	=	$\sum_i \sum_j \sum_k L_{ijk} \bar{y}_{ijk.}^2 - \sum_i \sum_j n_{ij} \bar{y}_{ij.}^2$
SSE	=	$\sum_i \sum_j \sum_k \sum_l y_{ijkl}^2 - \sum_i \sum_j \sum_k L_{ijk} \bar{y}_{ijk.}^2$
SST	=	$\sum_i \sum_j \sum_k \sum_l y_{ijkl}^2 - n \bar{y}^2$
SST	=	SSA + SSB + SSC + SSE

All cross-product terms become zero due to the selection of weights (4) in the side conditions (3) as illustrated by the last two terms [17],

$$\begin{aligned} \sum_i \sum_j \sum_k \sum_l (\hat{b}_{ij} - b_{ij})(\hat{c}_{ijk} - c_{ijk}) &= \\ \sum_i \sum_j (\hat{b}_{ij} - b_{ij}) \sum_k L_{ijk} (\hat{c}_{ijk} - c_{ijk}) &= 0 \end{aligned} \quad (10)$$

Similarly, other cross-product terms also vanish. In (9), the first term on the right is $\sum_i \sum_j \sum_k \sum_l (y_{ijkl} - \hat{\eta}_{ijkl})^2$ which is the sum of squares of error (SSE). The total sum of square (SST) is equal to the sum of squares due to each effect and the sum of square of error. This, in essence, is the partitioning of the SST into sums of squares of various effects and the sum of square of error (SSE) as given in Table I. The corresponding degrees of freedom (d.f.) are outlined in Table II.

The sums of squares of various factors and the error have a chi-square distribution χ_r^2 where ‘r’ is the d.f. The mean square of factors A, B, and C and the error are equal to the sum of square divided by their corresponding degrees of freedom.

$$\begin{aligned} MSA &= \frac{SSA}{(I-1)} & MSB &= \frac{SSB}{\sum_i (J_i - 1)} \\ MSC &= \frac{SSC}{\sum_i \sum_j (K_{ij} - 1)} & MSE &= \frac{SSE}{\sum_i \sum_j \sum_k (L_{ijk} - 1)} \end{aligned} \quad (11)$$

The various effects are tested through the F-test using the corresponding hypotheses, where the interest is in testing the null hypotheses against the alternate hypotheses. As mentioned, the significance of effects A, B, and C is found by partitioning the total space H_Ω into sup-spaces for individual effect, corresponding to $\Omega_A, \Omega_B,$ and Ω_C . Mathematically this is similar to finding the ratio of mean square of a particular effect {e.g. MSA, MSB, or MSC} with respect to the mean square of error (MSE) as given in Table III.

TABLE II. DEGREES OF FREEDOM

Effect	Degree of Freedom (d.f.)
A	$(I - 1) = 2 - 1 = 1$
B(A)	$\sum_i (J_i - 1) = 2(3) = 6$
C(B)	$\sum_i \sum_j (K_{ij} - 1) = 2(4)(3) = 24$
Error	$\sum_i \sum_j \sum_k (L_{ijk} - 1) = 2(4)(4)(3) = 96$

TABLE III. THRESHOLDS FOR HYPOTHESIS TESTING

Effect	Space for various effects	Threshold
A	$\Omega_A = H_A \cap H_\Omega$	$f_A = MSA/MSE$
B	$\Omega_B = H_B \cap H_\Omega$	$f_B = MSB/MSE$
C	$\Omega_C = H_C \cap H_\Omega$	$f_C = MSC/MSE$

The mean square of A (MSA) corresponds to chi-square distribution, χ_{I-1}^2 with (I-1) degree of freedom. The mean square of error (MSE) also corresponds to a chi-square distribution χ_p^2 with ‘p’ degree of freedom. Here, ‘p’ is equal to $\sum_i \sum_j \sum_k (L_{ijk} - 1)$. The ratio of two chi-square distributions is equal to an F-distribution, F_{α, v_1, v_2} where α is the upper $100\alpha\%$ point. Statistically, α is defined as $P\{F_{v_1, v_2} > F_{\alpha, v_1, v_2}\} = \alpha$ where v_1 and v_2 are the degrees of freedom of MSA, and MSE respectively. In this particular case, v_1 is equal to (I-1) and v_2 is equal to $\sum_i \sum_j \sum_k (L_{ijk} - 1)$. The null hypothesis, H_0 is rejected against the alternate, H_A if $f_A > F_{\alpha, (I-1), IJK(L-1)}$. The test is repeated for effect-B by comparing ratio $f_B > F_{\alpha, I(J-1), IJK(L-1)}$. Similarly, the threshold $f_C > F_{\alpha, IJ(K-1), IJK(L-1)}$ is tested.

III. SIMULATION RESULTS

The test was performed on two consecutive image frames selected in four categories: baseline, dynamic background, camera jitter, and shadows. It was assumed that the frames were captured at a small time interval with sufficient overlapping area. A mask size of 8 x 8 pixels was used for the statistical comparison. It was expected that the masks were sufficiently overlapped to preserve regions of objects having some kind of motion.

The template for three-way nested design is shown in Fig. 1. Level-A compared inter-frame pixels and identified regions with some kind of motion. The other two levels, B(A) and C(B) provided intra-frame pixel analysis, using mask sizes of 4 x 4 and 2 x 2 pixels, respectively. Furthermore, regions with both the temporal as well as the spatial information were marked. The threshold for the three levels, $f_A, f_B,$ and f_C , were equal to 8.49, 3.49, and 2.29 respectively with the value of ‘ α ’ selected as 0.001 for f_A and 0.005 for both f_B and f_C [16].

MATLAB routine ‘bwperim’ using an 8-level connected neighborhood was used to identify the contours. A preprocessing stage reduced the original 256 gray shades using (12). I_{in} is the input image, I_1 is an intermediate image with smaller range of gray shades, I is the image used in the algorithm and ‘ α ’ is a scaling factor; its value is arbitrarily selected as 40.

$$I_1 = \frac{1}{\alpha} I_{in} \quad \text{and} \quad I = \alpha \times I_1 \quad (12)$$

The sample images were taken from the CDnet database [18]. These images have been tested previously in [19]. Fig. 2 shows first frame of an image, the noisy image by the addition of Gaussian noise $\sim N(0, 100)$, the processed clean image, processed noisy image, and the contour of processed noisy image.

An overall observation is that the algorithm performed better in case of noisy images. Further, the algorithm performed exceptionally well in case of ‘dynamic background’. It was found that the algorithm was extremely sensitive to the slightest gray level change in area comprising of roads.

By reducing the granular level using (12), the number of quantization levels are reduced, thus significantly improving the F-measure. The quantitative analysis include false positive rate (FPR), false negative rate (FNR), percentage of wrong classification (PWC), and F-measure. The first three parameters were expected to have smaller value, while the overall quality measure, F-measure was expected to be closer to 1. The result of this algorithm is compared with previously published results for clean images. To the best of our knowledge, the results for noisy images are not available. It was observed that higher value of PWC in case of ‘shadow’

were due to the fact that the shadow constituted a larger area and the algorithm had mistakenly identified this area as the moving object. The algorithm was tested using MATLAB 2014, on a core i7 processor. The processing time was roughly equal to 2.5 seconds. A better result is expected by using faster computers employing parallel processing. By looking at the F-measure, it is clear that the algorithm is able to detect moving objects in both the noise-free as well as the noisy images. Comparison of proposed algorithm with recently published results is shown in Table IV.



Fig. 2. Original and Processed Images in Four Categories.

TABLE IV. COMPARISON OF PROPOSED ALGORITHM WITH RECENTLY PUBLISHED RESULTS

S.No.	Image Category	Noise Level	Previous Results [19]				Proposed Algorithm			
			FPR	FNR	PWC	F-Measure	FPR	FNR	PWC	F-Measure
1.	Baseline	no noise	0.003	0.13	0.9	0.87	0.545	0.0004	1.526	0.986
		w/ noise	-	-	-	-	0.293	0.001	0.896	0.943
2.	Dynamic Background	no noise	0.009	0.20	1.2	0.65	0.033	0.0347	3.418	0.919
		w/ noise	-	-	-	-	0.096	0.0073	3.280	0.982
3.	Camera Jitter	no noise	0.018	0.27	2.9	0.69	0.553	0.0029	4.259	0.962
		w/ noise	-	-	-	-	0.161	0.0043	1.647	0.949
4.	Shadows	no noise	0.011	0.17	2.1	0.78	0.399	0.0183	13.018	0.956
		w/ noise	-	-	-	-	0.231	0.0873	12.937	0.789

IV. CONCLUSION

This paper outlines a novel approach employing a three-way nested design using the analysis of variance (ANOVA). The sample images comprised of only two frames recorded within a short time interval with sufficient overlapping of pixels. The pixels identified to contain some kind of motion were combined to become part of the moving object. The temporal part was identified by statistical analysis of pixels at the two image frames. Spatial information at two granular levels were used to identify important details. The simulations were formed first in noise-free situation, and then in the presence of heavy Gaussian noise $\sim N(0, 100)$. The algorithm response was tested using the qualitative as well as the quantitative methods. The qualitative response was based on images taken in four categories: baseline, dynamic background, camera jitter, and images containing shadows. The quantitative analysis was based on false positive rate (FPR), false negative rate (FNR), percentage of wrong classification (PWC), and F-measure. The results were compared with a previously published results of clean images. The proposed algorithm performed better compared to the previous results. The proposed algorithm performed excellent in case of noisy images. In our view, the technique has great potential. The limitation of this approach is relatively longer processing time, and therefore not appropriate for real-time applications. One promising future extension is the analysis of covariance (ANCOVA).

REFERENCES

- [1] K. K. Hati, P. K. Sa, and B. Majhi, "Intensity range based background subtraction for effective object detection," *IEEE Signal Process. Lett.*, vol. 20, no. 8, pp. 759-762, Aug. 2013.
- [2] P. Chiranjeevi and S. Sengupta, "Detection of moving objects using multi-channel kernel fuzzy correlogram based background subtraction," *IEEE Trans. Cybernetics*, vol. 44, no. 6, pp. 870-881, Jun. 2014.
- [3] D. K. Panda and S. Meher, "Detection of moving objects using fuzzy color difference histogram based background subtraction," *IEEE Signal Process. Lett.*, vol. 23, no. 1, pp. 45-49, Jan. 2016.
- [4] Z. Wang, K. Liao, J. Xiong, and Q. Zhang, "Moving object detection based on temporal information," *IEEE Signal Process. Lett.*, vol. 21, no. 11, pp. 1403-1407, Nov. 2014.
- [5] G. Lee, R. Mallipeddi, G. Jang, and M. Lee, "A genetic algorithm-based moving object detection for real-time traffic surveillance," *IEEE Signal Process. Lett.*, vol. 22, no. 10, pp. 1619-1622, Oct. 2015.
- [6] J. Park, J. H. Yoon, M. Park, and K. Yoon, "Dynamic point clustering with line constraints for moving object detection in DAS," *IEEE Signal Process. Lett.*, vol. 21, no. 10, pp. 1255-1259, Oct. 2014.
- [7] L. Zhu, Y. Hao, and Y. Song, " $L_{1/2}$ norm and spatial continuity regularized low-rank approximation for moving object detection in dynamic background," *IEEE Signal Process. Lett.*, vol. 25, no. 1, pp. 15-19, Jan. 2018.
- [8] X. Zhou, C. Yang, and W. Yu, "Moving object detection by detecting contiguous outliers in the low-rank representation," *IEEE Trans. Patt. Anal. Mach. Intell.*, vol. 35, no. 3, pp. 597-610, Mar. 2013.
- [9] F. Shang, J. Cheng, Y. Liu, Z. Luo, and Z. Lin, "Bilinear factor matrix norm minimization for robust PCA: algorithms and applications," *IEEE Trans. Patt. Anal. Mach. Intell.*, vol. 40, no. 9, pp. 2066-2080, Sep. 2018.
- [10] S. Javed, A. Mahmood, S. Al-Maadeed, T. Bouwmans, and S. K. Jung, "Moving object detection in complex scene using spatiotemporal structured-sparse RPCA," *IEEE Trans. Image Process.*, vol. 28, no. 2, pp. 1007-1022, Feb. 2019.
- [11] O. Oreifej, X. Li, and M. Shah, "Simultaneous video stabilization and moving object detection in turbulence," *IEEE Trans. Patt. Anal. Mach. Intell.*, vol. 35, no. 2, pp. 450-462, Feb. 2013.
- [12] L. Li, Q. Hu, and X. Li, "Moving object detection in video via hierarchical modeling and alternating optimization," *IEEE Trans. Image Process.*, vol. 28, no. 4, pp. 2021-2036, Apr. 2019.
- [13] W. Hu, Y. Yang, W. Zhang, and Y. Xie, "Moving object detection using tensor-based low-rank and saliently fused-sparse decomposition," *IEEE Trans. Image Process.*, vol. 26, no. 2, pp. 724-737, Feb. 2017.
- [14] T. Kim, S. Lee, and J. Paik, "Combined shape and feature-based video analysis and its application to non-rigid object tracking," *IET Image Process.*, vol. 5, iss. 1, pp. 87-100, Feb. 2011.
- [15] J. F. Y. Cheung, M. C. Wicks, C. Genello, and L. Kurz, "A statistical theory for optimal detection of moving objects in variable corrupted noise," *IEEE Trans. Image Process.*, vol. 8, no. 12, pp. 1772-1787, Dec. 1999.
- [16] A. U. R. Khan, S. M. A. Saleem, H. Mehdi, "Detection of edges using two-way nested design," *International Journal of Advanced Computer Science and Applications (IJACSA)*, vol. 8, no. 3, pp. 136-144, March 2017.
- [17] H. Scheffe, "The Analysis of Variance," New York, Wiley, 1959.
- [18] Image Database, CDnet: <http://www.changedetection.net/>
- [19] N. Goyette, P. Jodoin, F. Porikli, et. al., "A novel video dataset for change detection benchmarking," *IEEE Trans. on Image Processing*, vol. 23, no. 11, pp. 4663-4679, 2014.