Available online at www.sciencedirect.com

ScienceDirect

Procedia Engineering 147 (2016) 637 – 642

Procedia Engineering

www.elsevier.com/locate/procedia

11th conference of the International Sports Engineering Association, ISEA 2016

Finding the optimal background subtraction algorithm for EuroHockey 2015 video

David Higham[a]*, John Kelley[a], Chris Hudson[a], Simon R. Goodwill[a]

[a]*Centre for Sports Engineering Research (CSER), Sheffield Hallam University, Broomgrove Road,Sheffield, S10 2NA, United Kingdom*

Abstract

Background subtraction is a classic step in a vision-based localization and tracking workflow. Previous studies have compared background subtraction algorithms on publicly available datasets; however comparisons were made only with manually optimized parameters. The aim of this research was to identify the optimal background subtraction algorithm for a set of field hockey videos captured at EuroHockey 2015.
Particle Swarm Optimization was applied to find the optimal background subtraction algorithm. The objective function was the F-score, i.e. the harmonic mean of precision and recall. The precision and recall were calculated using the output of the background subtraction algorithm and gold standard labeled images. The training dataset consisted of 15 x 13 second field hockey video segments. The test data consisted of 5 x 13 second field hockey video segments. The video segments were chosen to be representative of the teams present at the tournament, the times of day the matches were played and the weather conditions experienced. Each segment was 960 pixels x 540 pixels and had 10 ground truth labeled frames. Eight commonly used background subtraction algorithms were considered.
Results suggest that a background subtraction algorithm must use optimized parameters for a valid comparison of performance. Particle Swarm Optimization is an appropriate method to undertake this optimization. The optimal algorithm, Temporal Median, achieved an F-score of 0.791 on the test dataset, suggesting it generalizes to the rest of the video footage captured at EuroHockey 2015.

Peer-review under responsibility of the organizing committee of ISEA 2016

Keywords: Background Subtraction; Particle Swarm Optimisation; Field Hockey; EuroHockey 2015

1. Introduction

The localization and tracking of players in field sports, such as field hockey, is a vital tool used in the assessment of performance [1]. The data can provide both a physiological and tactical insight. Automated vision-based methods provide a non-invasive way to collect this data; important when radio frequency solutions are unsuitable or opposition data is required. Background subtraction, a method for the segmentation of the foreground in a scene, is a classic step in a vision-based workflow. An algorithm classifies each pixel in an image as either foreground or background. The development of background subtraction algorithms is an active research field and the BGS Library[2] contains some of the most popular. Previous studies have compared background subtraction algorithms on publicly available datasets[3,4]; but each algorithm has performance affecting parameters that may not be optimal. Without using the optimal values for these parameters the comparison will be biased towards one algorithm. [3] made no attempt to optimize the parameters. [4] manually optimized the algorithm parameters using grid search, however the optimal values are restricted by the discrete coarseness of the grid.

Particle Swarm Optimization (PSO) [5] is a stochastic optimization technique based upon the principles of swarm intelligence[6]. Each of the candidate solutions (particles) in the swarm moves around the search-space for a defined number of generations. The update of a particle's state from one generation to the next is based upon the swarm's best solution and the particle's best solution. If the algorithm is allowed to run for enough generations the particles of the swarm will converge. The

* Corresponding author. Tel.: +44-(0)114-225-2355;
E-mail address: d.higham@shu.ac.uk

Peer-review under responsibility of the organizing committee of ISEA 2016
doi:10.1016/j.proeng.2016.06.261

swarm's best observed solution is deemed the optimal solution. The search-space is continuous, so in contrast to grid search PSO is not limited by the need to define the resolution of the grid.

PSO has been applied to optimize the parameters of a single background subtraction algorithm[7], specifically Gaussian Mixture Model [8]. However no attempt has been made to optimize all algorithms in a set and then subsequently identify the best algorithm. In [7] the objective of the optimization is to maximize the F-score. The F-score is the harmonic mean of the precision and recall. The calculations for precision and recall are illustrated in Fig. 1.

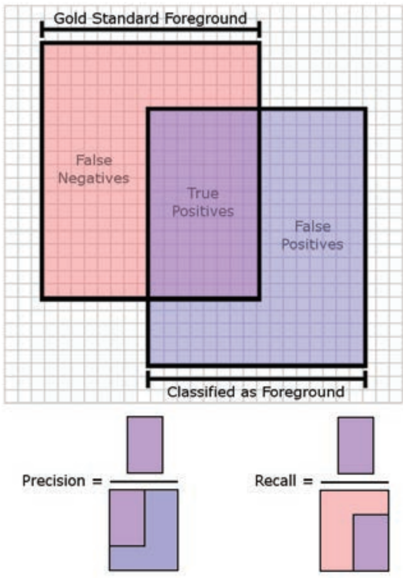

Fig. 1. Calculations for precision and recall given a gold standard foreground image and the classified output of a background subtraction algorithm.

The aim of this research was to identify the optimal background subtraction algorithm for the video dataset using Particle Swarm Optimization. The set of videos was captured at the 2015 EuroHockey Championships. It was shown that this optimal algorithm can then be applied to other videos from the same tournament to achieve similar segmentation results.

2. Method

2.1. Optimization Video Dataset

15 video segments were used for the training dataset. A further five video segments formed the test dataset. Each segment was the starting 13 second sequence, at 25 frames per second, of a different quarter from EuroHockey 2015. The videos in the training and test datasets were chosen to be representative of: 1) the teams present at the tournament; 2) the times of day matches were played; and 3) the weather conditions experienced. The original quarters were captured at a resolution of 3840 x 2160 (4K) using a Sony FDR-AX1 fitted with a 0.3x wide angle lens, however the video segments included in the dataset were of the resolution 960 x 540. This decision was made to decrease the execution time of the optimization. A previous unpublished study showed a strong correlation ($r(24) = 0.96$, $p < 0.01$) between the optimal F-score for an algorithm at 4K and at 960 x 540. The optimal parameters themselves were not consistent across resolutions. The frames of the video were resized using average pixel interpolation.

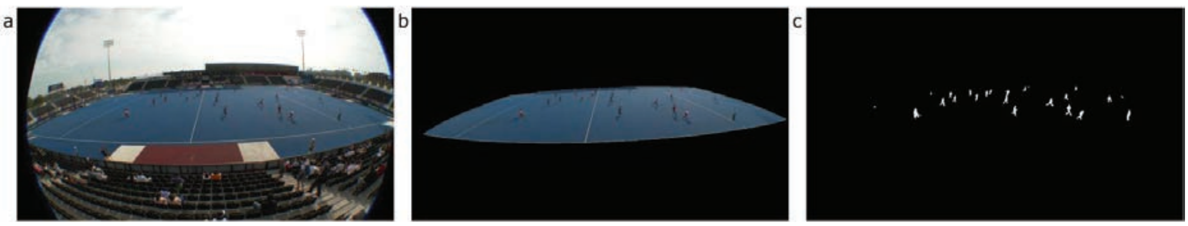

Fig 2. (a) Initial frame; (b) Frame masked to the pitch area; (c) Player pixel mask.

Each video contained 10 comparison frames; those frames for which the gold standard foreground had been manually labelled. For each video the first 100 frames were a dedicated learning period, necessary for some of the algorithms. The comparison set comprised of the 101st frame and every subsequent 25th frame up to the 326th frame. Therefore there were 150 frames in the training set and 50 frames in the test set. As illustrated in Fig 2b, each comparison frame was masked to the pitch area. This eliminated foreground classifications outside the area of interest. The player pixels were then labelled as in Fig 2c. A data access statement for the dataset can be found at the end of this article.

Table 1. Background subtraction algorithms included in the study. For each algorithm the parameters to be optimized are listed with their default values and range of valid values.

Algorithm	Summary	Parameter	Description	Default	Min	Max
Frame Difference	Absolute difference with previous frame. Pixel is foreground if difference is greater than a threshold.	threshold	Pixel foreground if greater than value	15	0	255
Static Frame Difference	Absolute difference with first frame. Pixel is foreground if difference is greater than a threshold.	threshold	Pixel foreground if greater than value	15	0	255
Weighted Moving Mean	Absolute difference with weighted mean frame. Pixel is foreground if difference is greater than a threshold.	threshold	Pixel foreground if greater than value	15	0	255
Temporal Median [9]	Absolute difference with median frame. Pixel is foreground if difference is greater than threshold	threshold	Pixel foreground if greater than value	30	0	255
		samplingRate	How often background resampled	5	1	20
		historySize	Number of frames in sample	16	1	70
		weight	Amount of influence given to previous samples	5	1	20
Average Gaussian [10]	Per pixel background represented by a running average Gaussian. Pixel is foreground if it does not fall within a defined distance of the Gaussian.	threshold	Pixel foreground if greater than this many variances from mean	12.25	0.25	100
		alpha	Adaptive filter update rate	0.005	0	1
		learningFrames	Number of frames for initialization	30	0	24
Gaussian Mixture Model [8]	Per pixel background represented by a mixture of Gaussians. Pixel is foreground if does not fall within a defined distance of one of the Gaussians. Background Gaussians updated as more evidence becomes available.	threshold	Pixel foreground if greater than this many variances from background mean	12.25	0.25	100
		backgroundRatio	Ratio of Gaussians that account for the background	0.75	0	1
		alpha	Adaptive filter update rate	0.001	0	1
		numGaussians	Maximum number of Gaussians	3	1	5
Gaussian Mixture Model (Zivkovic) [11]	Extension of Gaussian Mixture Model. Uses an exponentially decaying envelope to limit the influence of old data.	alpha	Parameter that defines the exponentially decaying envelope	0.05	0	1
		backgroundRatio	Ratio of Gaussians that account for the background	0.9	0	1
		threshold	Pixel foreground if greater than this many variances from background mean	9	0.25	100
		fVarInit	Variance of new Gaussian model	15	0	100
		fCT	Complexity reduction parameter	0.05	0	1
Dominant Color	Similar to [12]. Assumes the pitch is the dominant color in the scene and everything else is a player. Absolute difference between pixel and peak value. Pixel is foreground if difference in each channel is greater than threshold.	thresholdChannel1	Threshold on channel 1 of colorSpace	-	0	255
		thresholdChannel2	Threshold on channel 2 of colorSpace	-	0	255
		thresholdChannel3	Threshold on channel 3 of colorSpace	-	0	255
		numberBinsChannel1	Number of histogram bins on channel 1 of colorSpace	-	1	255
		numberBinsChannel2	Number of histogram bins on channel 2 of colorSpace	-	1	255
		numberBinsChannel3	Number of histogram bins on channel 3 of colorSpace	-	1	255
		colorSpace	Color Space that the algorithm works in. 0 = RGB, 1 = HSV, 2 = Lab, 3 = YCrCb	-	0	3

2.2. Background Subtraction Algorithms

Eight background subtraction algorithms were included in the study. These eight algorithms were chosen based upon their common use in previous research. The algorithms along with the parameters to be optimized are displayed in Table 1. For all algorithms but Dominant Color the BGS Library implementation was used.

2.3 Optimization Procedure

The parameters for each background subtraction algorithm were optimized once using PSO. The optimizations were run on a HP Z230 Workstations (Intel® Core^{TM} i7-4790 processor - 3.6 GHz, 16 GB RAM). For each background subtraction algorithm, the number of particles and behavior parameters were based upon the number of parameters and taken from literature [13]. Each parameter's initial state was randomized within the search space. The PSO ran for 100 generations or until convergence. If a parameter state is given as a normalized vector, convergence was defined as 60% of the particles within the population having a best observed state within 0.1 of the global best state [14]. No attempt was made to optimize the execution time of the algorithm. The F-score for a parameter state was calculated using the mean precision and recall across the 150 labeled training frames.

Following a pilot study, it was deemed that cross-validation was impractical. The average execution time of a single parameter state evaluation for the Temporal Median algorithm was approximate real-time (190 seconds). Assuming the optimization does not converge, then over 20000 parameters states would be evaluated, making multiple executions unfeasible.

2.4 Test Procedure

The test dataset was evaluated using both the default parameter state and the optimal parameter state. The F-score for a parameter state was calculated using the mean precision and recall across the labeled 50 test frames.

3. Results

Fig. 3 illustrates for each algorithm the F-score for the default parameters and the F-score for the optimal parameters. The default parameters for Weighted Moving Mean did not return an F-score and no defaults were defined for Dominant Color so both are omitted.

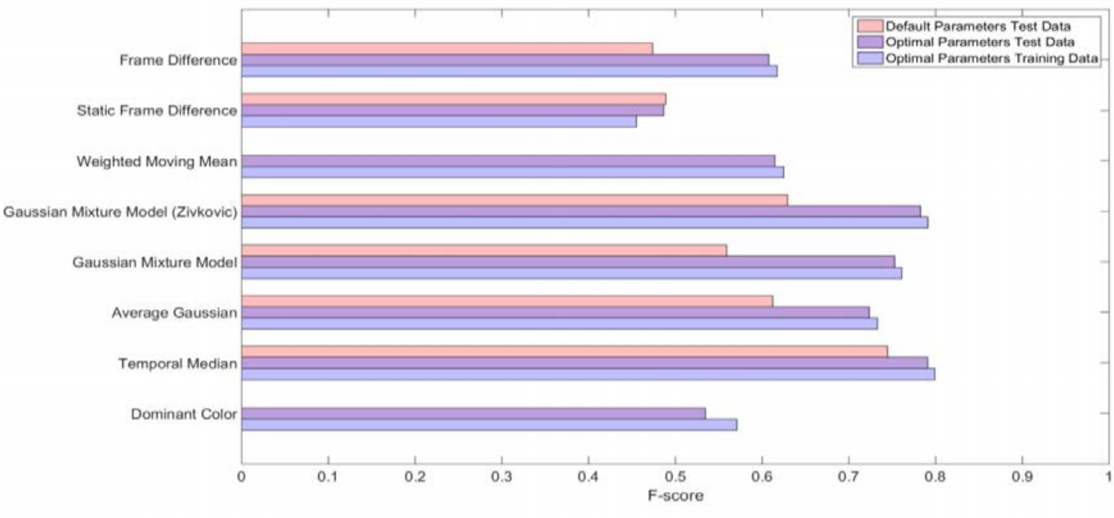

Fig. 3. For each background subtraction algorithm the observed F-score for the default parameters on the test set and the optimal parameters on the test set and training set. The default values for Weighted Moving Mean did not return an F-score. Dominant Color does not list default parameters.

Table 2 lists the parameters for the optimal algorithm, Temporal Median.

Table 2. Optimal observed value for each parameter of the Temporal Median.

Algorithm	Parameter	Optimal Value
Temporal Median	threshold	22
Temporal Median	samplingRate	10
Temporal Median	historySize	65
Temporal Median	weight	8

4. Discussion

The aim of this study was to identify the optimal background subtraction algorithm to segment video captured at EuroHockey 2015. While the dataset is different, the principle of the study is similar to that in [3]. They reviewed the performance of the algorithms using the default parameter values defined by the algorithms' authors. Fig. 3 illustrates, for the test dataset, the F-scores for the default parameter values are considerably different to the F-scores for the optimal values. These results suggest that the optimal parameters must be used for a valid comparison between the algorithms. Without optimization a comparison is biased towards certain algorithms. For example, when using the default parameters, Gaussian Mixture Model performed relatively poorly compared to the best algorithm, Temporal Median (Gaussian Mixture Model – 0.559 vs Temporal Median – 0.745). However, when using the optimal parameters it performed much better (Gaussian Mixture Model – 0.761 vs Temporal Median – 0.791). Further to this, for Moving Mean Average the default parameters failed to return an F-score as they did not classify any pixels as foreground in one of the test frames.

Having established the optimal parameters a valid comparison can be made between algorithms. Fig. 3 shows the relatively simple Temporal Median achieved the highest observed F-score (0.791). The algorithm regularly resamples the entire background, providing an up-to-date background model with which to compare. Similarly Frame Difference resamples the entire background every frame. However, as only one frame is used, the background is susceptible to noise, which will decrease the recall and as such the F-score. Frame Difference does nevertheless perform comparatively to Moving Mean Average despite being a simpler algorithm. Static Frame Difference performs the worst of all the algorithms. The player's positions in the initial frame appear as ghosts in the output for subsequent frames. These ghosts decrease the precision and as a result the F-score.

In general the algorithms that modelled the background using a Gaussian distribution outperformed those that modelled the background as a single value. This is due to the variance across previous frames being used to determine a dynamic background threshold. Further to this, Gaussian Mixture Model and Gaussian Mixture Model (Zivkovic), which both permitted multiple Gaussians, outperformed Average Gaussian. This suggests that the variance of background could not be modelled sufficiently by a single Gaussian. The similarity between Gaussian Mixture Model and Gaussian Mixture Model (Zivkovic) should also be expected due to the latter being an extension of the former.

Fig. 3 also displays the F-score for the optimal parameters on the training dataset. For all algorithms but Static Frame Difference, a small drop in F-score was observed between the performance on the training dataset and the test dataset. This could be a result of overfitting of the parameters to the training set or a result of differences in the distributions of the training set and testing set. However it does suggest that the optimal parameters can be generalized to other videos from EuroHockey 2015.

5. Future Work

Execution time is a key consideration in time critical applications. An algorithm's parameters may affect the execution time. Therefore execution time should be considered as a factor in future optimizations. This may be achieved by weighting F-scores by execution time or by setting a maximum permitted execution time.

As noted previously, Parameter states at different resolutions were not consistent; therefore the best observed parameter state may not be optimal with the original 4K footage. The best performing algorithms must be optimized again using a 4K dataset.

This work used the F-score, the harmonic mean of the precision and recall, as its objective function. This gave equal influence to both the precision and recall. In practice, computer vision techniques can be used to eliminate false positives but not false negatives. Consequently better segmentation results may be achieved by biasing to reduce the number of false negatives and as such increase the recall. This could be achieved using the F_β measure that weights recall to "attach β times as much importance to recall as precision" [15].

The test set contained variance in the time of day and weather conditions. Further analysis will be undertaken to assess if the best background subtraction algorithm is dependent upon these factors.

6. Conclusions

- The parameters of a background subtraction algorithm should be optimized before a valid comparison between algorithms is made.
- Particle Swarm Optimization is a suitable tool for the optimization of background subtraction algorithm parameters.
- Temporal Median was the optimal algorithm for a training set from EuroHockey 2015. It also generalizes well to other videos from EuroHockey 2015. No statement can be made upon its generalization to other field sport datasets.

Acknowledgements

The authors would like to thank the English Institute of Sport and the Engineering and Physical Sciences Research Council (ESPRC) for funding this research. The EPSRC CASE award number is EP/K504543/1. They would also like to thank England Hockey for providing filming access at EuroHockey 2015.

Data Availability Statement

Due to ethical concerns, not all supporting data can be made openly available. Further information about the data and conditions for access are available at the Sheffield Hallam University research data archive: http://doi.org/10.17032/shu-160002

References

[1] Leser R, Baca A, Ogris G. Local positioning systems in (game) sports. Sensors 2011;11:9778–97.
[2] Sobral A, Bouwmans T. BGS Library: A Library Framework for Algorithm's Evaluation in Foreground/Background Segmentation. In: Bouwmans T, Porikli F, Höferlin B, Vacavant A, editors. Background Modeling and Foreground Detection for Video Surveillance. Chapman and Hall; 2014. p. 23-1-23-16.
[3] Benezeth Y, Jodoin P-M, Emile B, Laurent H, Rosenberger C. Comparative study of background subtraction algorithms. J Electron Imaging 2010;19:033003.
[4] Sobral A, Vacavant A. A comprehensive review of background subtraction algorithms evaluated with synthetic and real videos. Comput Vis Image Underst 2014;122:4–21.
[5] Kennedy J, Eberhart R. Particle swarm optimization. IEEE Int. Conf. Part. swarm Optim 1995;4:1942–8.
[6] Beni G, Wang J. Swarm Intelligence in Cellular Robotic Systems. In: Dario P, Sandini G, Aebischer P, editors. Robots and Biological Systems: Towards a New Bionics? NATO ASI Series Vol 102; 1993. p. 703–712.
[7] White B, Shah M. Automatically Tuning Background Subtraction Parameters Using Particle Swarm Optimization. Multimed Expo, 2007 IEEE Int Conf 2007;1826–9.
[8] Stauffer C, Grimson WEL. Adaptive background mixture models for real-time tracking. Proc 1999 IEEE Comput Soc Conf Comput Vis Pattern Recognit Cat No PR00149 1999;2:246–52.
[9] Cucchiara R, Grana C, Piccardi M, Prati A. Detecting objects, shadows and ghosts in video streams by exploiting color and motion information. Proc. 11th Int. Conf. Image Anal. Process 2001;360–5.
[10] Wren CR, Azarbayejani A, Darrell T, Pentland AP. Pfinder: Real-Time Tracking of the Human Body. IEEE Trans Pattern Anal Mach Intell 1997;19:780–5.
[11] Zivkovic Z. Improved adaptive Gaussian mixture model for background subtraction. Proc 17th Int Conf Pattern Recognition, 2004;2:28–31.
[12] Seo Y, Choi S, Kim H, Hong K-S. Where are the ball and players? Soccer game analysis with color-based tracking and image mosaick. Image Anal. Process 1997;1311:196–203.
[13] Erik M, Pedersen H, Pedersen MEH. Good parameters for particle swarm optimization. Tech Rep HL1001, Hvass Lab 2010;HL1001:1–12.
[14] Zielinski K, Laur R. Stopping criteria for a constrained single-objective particle swarm optimization algorithm. Inform 2007;31:51–9.
[15] Van Rijsbergen CJ. Information Retrieval, 2nd ed. Massachussets: Butterworth-Heinemann Newton; 1979.