Video Segmentation via a Gaussian Switch Background Model and Higher Order Markov Random Fields

Martin Radolko and Enrico Gutzeit
Fraunhofer Institute for Computer Research IGD, 18059 Rostock, Germany
{*martin.radolko, enrico.gutzeit*}*@igd-r.fraunhofer.de*

Keywords: Image Segmentation, Background Substraction, Belief Propagation, Otsu's Method, Markov Random Fields.

Abstract: Foreground-background segmentation in videos is an important low-level task needed for many different applications in computer vision. Therefore, a great variety of different algorithms have been proposed to deal with this problem, however none can deliver satisfactory results in all circumstances. Our approach combines an efficent novel Background Substraction algorithm with a higher order Markov Random Field (MRF) which can model the spatial relations between the pixels of an image far better than a simple pairwise MRF used in most of the state of the art methods. Afterwards, a runtime optimized Belief Propagation algorithm is used to compute an enhanced segmentation based on this model. Lastly, a local between Class Variance method is combined with this to enrich the data from the Background Substraction. To evaluate the results the difficult Wallflower data set is used.

1 INTRODUCTION

Nowadays Vision Systems are used in many fields of applications such as surveillance, industrial automation, transportation or inspection. In the last decades Background Substraction has become a valuable source of low level visual information. It can detect arbitrary objects in almost any scene, as long as they are in motion. This information can afterwards be used for all kinds of high-level vision tasks.

In order to gather the aforementioned data the background for each individual scene has to be modelled. The creation of this model is not a trivial task and associated with many difficulties like changes in the lightning conditions, moving background objects (swaying trees), shadows or a changing background. To cope with all these requirements a great number of different approaches have been deployed. Some use Subspace Learning Models like LDA (Kim et al., 2007), INMF (Bucak et al., 2007) or PCA (Marghes et al., 2012) to generate a background model. Other prominent methods adopt techniques like Kalman Filters (Cinar and Principe, 2011), SVMs (Lin et al., 2002) or histograms (Zhang et al., 2009) to cope with these problems.

However, most of the current algorithms model each background pixel as a Gaussian Distribution. This is justified by the fact that the intensity of a pixel in a completely static scene will vary according to a Normal distritution $\mathcal{N}(\mu, \sigma^2)$ due to the measurement errors inherent in every camera system. With this information, a threshold per pixel can be easily created to distinguish between foreground and background.

There are approaches which use just one Normal Distribution per pixel (Wren et al., 1997), algorithms which use a Mixture of Gaussians (Stauffer and Grimson, 1999; Setiawan et al., 2006) or Gaussian-Kernel based methods (Elgammal et al., 2000) to model the background. Methods which use a Mixture of Gaussians (MoG) produce in most cases better results than the Single Gaussian (SG) algorithms, but also have some disadvantages. One is a higher memory usage and another the need to be tuned for the right amount of Gaussians.

A shared drawback of all Background Substraction approaches is that they do not incorporate the spatial informations about the scene in the model, although natural images are commonly assumed to be very smooth. To use this assumption to improve the segmentation derived from the Background Substraction different strategies have been applied. In (Toyama et al., 1999) a simple approach is used which discards all connected regions containing less than a certain amount of pixels. A more complex approach is used in (Y. Wang and Wu, 2006) where a Conditional Random Field models the neighbourhood relations of the pixels. Graph Cuts are used in (Boykov and Funka-Lea, 2006) and to represent the spatial in-

formation in all dimension in a Subspace Model a Tensor was applied (Li et al., 2008).

The most frequently utilized method for this problem is a MRF which models the relations via an undirected graph. The most likely overall state of the model (maximum a posteriori probability - MAP) will correspond with a good segmentation if the underlying data (a probability map) is adequate. Since it is NP-hard to compute the MAP for a cyclic MRF, there exists several methods to approximate the best solution. In (Xu et al., 2005) Gibbs-Sampling is combined with simulated annealing, (Sun et al., 2012) applies a Branch and Bound algorithm and (Yedidia et al., 2003) utilises a Belief Propagation method to get a good approximation of the MAP.

Building on this, a new method composed of three parts is developed. First, a new Background Substraction algorithm will be proposed which uses exactly two Gaussians and thereby eliminates most of the disadvantages of the MoG approaches, but nonetheless generates state of the art results. Secondly, a higher order MRF with a variable neighbourhood is created to model the spatial relations of the objects in the scene. A Belief Propagation algorithm is used to derive a suitable approximation of the MAP. Lastly, a component was added which influences the segmentation to maximize the between-class variance. The basic idea was first proposed in (Otsu, 1979) and is widely appreciated since then (Liao et al., 2011; Huang et al., 2001). Here, the approach is used in a combination with the Belief Propagation algorithm so that the between class variance is optimized during the iterations as well.

2 HIGHER ORDER MARKOV RANDOM FIELDS

The MRF is a well established and widely used statistical model which can describe the dependencies between various random variables. It originates from the works of Ising on ferromagnetism (Ising, 1925) but was since extended and adopted to a large number of different problems.

A small example of a MRF represented as a graph can be seen in Figure 1. The random variables are depicted as circles and the edges show the dependencies between them. This easy and graphical way of modeling the relations between random variables can be useful in a great variety of applications, one example are the spatial relations between pixels in an image. In this case every pixel is represented by one random variable for which the state is unknown and which has dependencies on all neighbouring pixels.

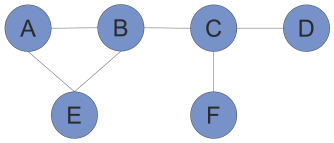

Figure 1: A graphical depiction of a small MRF.

Thereby, the state of a random variable could indicate if the corresponding pixel is in the foreground of the image, denote the optical flow at that point or any other information which can be deduced from the data.

A crucial point for this is the selected neighbourhood system. A small system like the von Neumann neighbourhood might be unable to model all the complexity of the relations and bigger systems will soon create models which are unmanageable. For computer vision algorithms the von Neumann neighbourhood is almost always chosen because it will create a pairwise MRF. These MRF have the advantage that they only have cliques of one or two nodes which makes the computation much easier.

The computational difficulties derive from the fact that in every clique all members will be influenced from all the others. To deduce an approximate solution of the MAP a Belief Propagation algorithm will compute messages from every clique to all of its members. This means that in Figure 1 node E will receive one message depending only on A (E and A are a clique), one depending only on B but also a third message depending on both of them (A, E and B are a clique). For bigger neighbourhoods each node can be in tens or even hundreds of different cliques which will make the computation and storage of all the messages nearly impossible.

In the Moore neighbourhood, which is the next bigger neighbourhood system occasionally used for images, every node is already a member in 24 cliques. Also, it has to be noted that there is not just one message from every clique to each member but one message for every possible state the whole clique can be in. To reduce this heavy computational load there will be two techniques introduced in section 3.2 which will make it possible to compute good approximations of the MAP even for advanced neighbourhood systems.

Until now the proposed MRF only models the spatial relationship of the pixels. To generate good segmentations the information given by the image also have to be included into the model. In the proposed method this will be a value generated by the Background Substraction method denoting the probability of a pixel being in the background. To include this information, a second node with a fixed/known

state will be created for every pixel. This new node is called an evidence node because it represents given information in the model. The others are called hidden nodes since they indicate an unknown state of the system which shall be deduced. Every evidence node will influence only the corresponding hidden node and in a way that it will more likely attain the state favoured by the given data. An example of a MRF model with a Moore Neighbourhood can be seen in Figure 2.

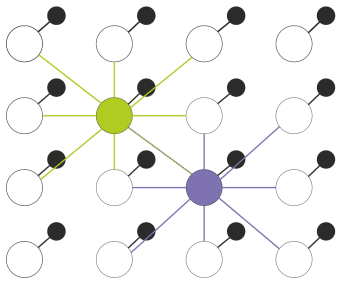

Figure 2: The evidence nodes are drawn as small dark circles. Each is connected with one edge to the corresponding hidden node. For two hidden nodes the edges to the other neighbouring hidden nodes are also drawn.

3 OUR APPROACH

Our approach consists of three main steps.

- First, a novel efficient Gaussian Switch Model is used to create an approximation of the background which is then subtracted from every new frame to create a first segmentation.
- A MRF is used in the second step to model the spatial relations between the pixels. Unlike many other approaches, a MRF of higher order is used here to reproduce the relations between different pixels in the MRF more precisely.
- In the last step, a local version of Otsu's Method is added to optimize the data generated by the Background Substraction. These changes will influence the Belief Propagation, so that the resulting segmentation has a high local between Class Variance which will further improve the outcomes.

3.1 The Gaussian Switch Model

As justified previously, Gaussian distributions are used to model the colour of each pixel. Instead of using a batch method, which will save the last n pictures and generate a background model from these, the Gaussians are updated with every new frame to save memory and computing power. Thereby, the algorithm gives the new samples automatically a higher weight than old samples and thus even improves the results in comparison to the batch method. For every Gaussian, the mean μ and variance σ^2 have to be stored. The mean is initiated with the pixel value taken from the first frame of the video stream and the variance is set to a predefined value. Afterwards, they are updated in the following way

$$\mu^{t+1} = \alpha\,\mu^t + (1-\alpha)\,v^t, \tag{1}$$

$$(\sigma^{t+1})^2 = \alpha\,\sigma^t + (1-\alpha)(\mu^t - v^t)^2. \tag{2}$$

The variable $\alpha \in [0,1]$ controls the update rate and v^t is the pixel value taken from the t-th frame. In the rest of the paper the index t will be omitted because all variables are taken from the same frame.

With these formulas, the Gaussian distribution of a background pixel can be modelled very precise and efficient. Nevertheless, one problem is that the distribution becomes erroneously when a foreground object is visible. To overcome this problem, the background Gaussian $\mathcal{N}(\mu^{bg},(\sigma^{bg})^2)$ is created which will only get updated when the new pixel is classified as background. This results in a much more resilient background model but has two inherent problems. The first issue are objects in motion visible in the first frame because the real background in this area would never get included into the model. The other problem are foreground objects that become background (e.g. a car that parks), because they will never get included into the background model. To eliminate these errors an overall Gaussian $\mathcal{N}(\mu^{og},(\sigma^{og})^2)$ is needed which will be updated with every new frame.

If this overall Gaussian has a small variance but a different mean than the background Gaussian, a foreground object was visible and immobile for a long period of time. This foreground object should therefore become a part of the background model. To achieve this, the mean and variance of the background Gaussian are switched to the overall Gaussian's values. Now, one intensity value per pixel can be modelled in this way but most videostreams today are not in grayscale and consequently have three colour values for each pixel. To use these additional data a special colour space is applied which normalises the different intensities in respect to the illumination (Li et al., 2008). Let R, G and B be the given values for a single pixel in the standard RGB colour space, then these will be transformed into the three new image channels

$$I = R+B+G,$$
$$\tilde{R} = R/I,$$
$$\tilde{B} = B/I.$$

Afterwards the intensity I is normalized, so that all values are in the range of $[0,1]$. The colour information stored in $\tilde{R}$ and $\tilde{B}$ are normalised with the intensity and will thus not be altered by small or medium changes in the lighnting conditions. This is used to avoid the detection of shadows as foreground.

For each of these three values an independent Gaussian Switch Model (GSM) has to be applied. To decide whether a pixel matches the background model the values for each channel will be separately compared to the corresponding GSM. Let p_R be the new $\tilde{R}$ value for a pixel, it is classified as matching the background model if the following inequality is satisfied:

$$(p_R - \mu_R^{bg})^2 < \max(\beta \cdot (\sigma_R^{bg})^2, 0.001). \quad (3)$$

The maximum is used because the variance could approach near zero values, especially since only matching values are included into the background model. The parameter β can control the range of values which are still classified as "matching the model". This use of the variance provided the algorithm with a pixel-wise adaptive threshold.

To get a decision for a single pixel as a whole a voting procedure is chosen. If for a single pixel equation (3) is satisfied for atleast two channels, then the pixel is marked as background, otherwise as foreground. Thereby, the colour information can overrule the brightness information and hence shadows should not be detected as foreground.

In (Toyama et al., 1999) a method is proposed which takes global changes in the lightning condition into account, for example when a cloud is blocking the sun and makes the whole scene darker. These events often result in the classification of almost the whole scene as foreground and afterwards it takes the model a long time to adapt to the new conditions. To improve this behaviour, the algorithm will check in every new frame if more than 75% of the pixels are classified as foreground, if only the intensity channel is taken into consideration. Should this be the case the update rate α is set to 0.5 to increase the adaption speed of the model drastically.

3.2 The Markov Random Field

In the next step the result from the Background Substraction algorithm will be optimized by correcting small areas with false detections to get contiguous foreground and background regions. To achieve this the MRF described in section 2 is used. It can model the spatial relations between single pixels and hence forces the segmentation to be locally coherent.

The neighbourhood system is the most important part of a MRF. Here a Generalized Moore Neighbourhood (GMN) is used to ensure the homogeneity of the MRF and because it can be easily changed in size. The normal Moore Neighbourhood is shown in Figure 2 for two different nodes. For a node N this neighbourhood system is defined by a three times three square of nodes with N in the center of it. All nodes of the square are then neighbours of N. The first order GMN uses a 5×5 square instead of a 3×3 one, the second order GMN then enlarges this to a 7×7 square and so forth. Hereby, the number of cliques will increase radically with the order of the GMN.

As input the MRF needs in principle two values for each pixel, one should indicate the probability of the pixel for being in the foreground and the other the probability of being in the background. These data is needed for the evidence nodes, so that they can represent the result of the Background Substraction algorithm. Here just one value w_i is used, which is the probabilty that the pixel i belongs to the background. As this value is already normalisied the other probability can be set to $1 - w_i$. To calculate w_i the voting algorithm is used again. At the beginning w_i is set to 0.9 and then for each of the three channels which does not match the background model the value is lowered.

For the two colour channels 0.3 are subtracted in each case and for the intensity channel 0.2 is deducted. By this means the probability will always be above 50% if at least two of the channels favour the background and less than 50% otherwise. The colour channels get an higher weight because a change in the colour is a better indicator for the pixel being foreground than a change in the intensity.

To eventually compute an approximation of the MAP a cost function has to be defined which measures how good a segmentation matches the MRF model. A function $D(i)$ is needed for the evidence nodes and shall describe how good a certain state of the corresponding hidden node matches the data delivered from the Background Substraction. In our case the function for the hidden node i is defined as follow:

$$D(i) = \begin{cases} w_i, & i \text{ is background} \\ 1 - w_i, & i \text{ is foreground} \end{cases} \quad (4)$$

A second set of functions is necessary for the cliques of hidden nodes. They are named C_k, the k indicating the size of the clique. These functions should represent the spatial relationship between the pixels and hence there can be different functions for all possible clique sizes and spatial arrangements. However, in this case the beneficial homogenous structure of the MRF can be exploited. Since all hidden nodes have the same neighbourhood and a coherent segmentation shall be created everywhere, the functions C_k can be the same for all nodes. A small exception are the boundaries, there the neighbourhood is smaller

and consequently the number of cliques decreases. Nonetheless, for the remaining cliques the standard functions can be applied.

More problematic is the fact that different function values have to be calculated for every single clique a node is part of, this increases the complexity of the model dramatically. To reduce the computational load only one clique size was chosen to contribute to the energy function. This simplification will drastically decrease the number of messages which have to be send later in the Belief Propgation algorithm.

Another way to lessen the computational burden is the usage of a simple energy function for the remaining cliques. Due to the fact that large homogenous fore- or background regions are presumed, the functions C_k should favour the cases where all nodes in the clique have the same state. This can be achieved by returning 0 energy for this case and a positiv static value for all other cases. Equation (5) gives an example for the case $k = 4$, there h_1 to h_4 are the states of the four different hidden nodes contained in the clique.

$$C_4(h_1,h_2,h_3,h_4) = \begin{cases} 0, & h_1 = h_2 = h_3 = h_4 \\ 1, & \text{elsewise} \end{cases} \quad (5)$$

These simplifications are required, since without them it would not be feasible to use more than the Moore Neighbourhood on any picture with a reasonable resolution. A comparison of the effects of different neighbourhood systems and clique sizes can be seen in Figure 3 in the result section.

To eventually compute the MAP of the MRF the respective graph was first converted to a factor graph and then a loopy max-product Belief Propagation was applied. For details see (Yedidia et al., 2003) and (Felzenszwalb and Huttenlocher, 2004).

3.3 Using the Between Class Variance

In (Otsu, 1979) a method is described to segment a picture into two classes by maximizing the variance between the two classes. This results in an useful segmentation only under very specific circumstances. Namely, when the objects of interest are all similar among them and different from the background in colour and brightness. It is obvious that this assumption cannot be made in real life images.

However, in most cases this assumption will hold if only a small area around a pixel is taken into account. To use this to improve the segmentation created in the first two steps, a local between Class Variance is introduced and coupled with the Belief Propagation to enrich the data taken from the Background Substraction in each iteration.

A segmentation has to be known first to compute the between Class Variance. For this reason, one is created after every iteration of the Belief Propagation algorithm and then for every pixel the average colours of all background and foreground pixels, but only in a small square-sized area around this pixel, are calculated. Now the background probability w_i will be increased if the colour of current pixel is closer to the background average colour and lowered elsewise.

To be more precise, let $c^{bg} = (c_I^{bg}, c_R^{bg}, c_B^{bg})$ and c^{fg} be the local average colour of the background respectively foreground and c_i the colour of the current pixel. If $\|c^{bg} - c_i\|_2 < \|c^{fg} - c_i\|_2$ the between class variance will increase when the current pixel is classified as background. If the inequality does not hold, the pixel should be classified as foreground to increase the local between class variance.

However, the classification of the current pixel will not be directly changed according to the between Class Variance. Instead, the data obtained from the Background Substraction is altered to reinforce a classification which increases the local between Class Variance in the following iterations. For the pixel i this is done by computing the value

$$v_i = \gamma \cdot (\|c^{fg} - c_i\|_2 - \|c^{bg} - c_i\|_2) \quad (6)$$

and changing the probability w_i accordingly. Gamma is a factor which controls the impact the Class Variance will have on the data from the Background Substraction. Now the values p_i^{bg} and p_i^{fg} can be calculated which represent the new foreground and background probabilities,

$$p_i^{bg} = \max(w_i + v_i, 0.001)$$
$$p_i^{fg} = \max((1 - w_i) - v_i, 0.001).$$

The maximum is required to avoid negativ probabilities. In the next step these values have to be normalized so that $p_i^{bg} + p_i^{fg} = 1$ and afterwards the value w_i can be replaced by p_i^{bg}.

This process improves the results but is also computational expensive, especially the calculation of the average values c^{bg} and c^{fg} for every pixel in every iteration for every frame. To reduce the runtime, Integral Images (first introduced by (Viola and Jones, 2004)) are used to efficiently compute these values in a static time, independent of the size of the square-sized area over which these values are averaged. For every channel two Integral Images have to be created, one for the background pixels and one which only adds up the foreground pixels. After calculating these integral images the actual averages can be obtained by a simple operation consisting only of three additions.

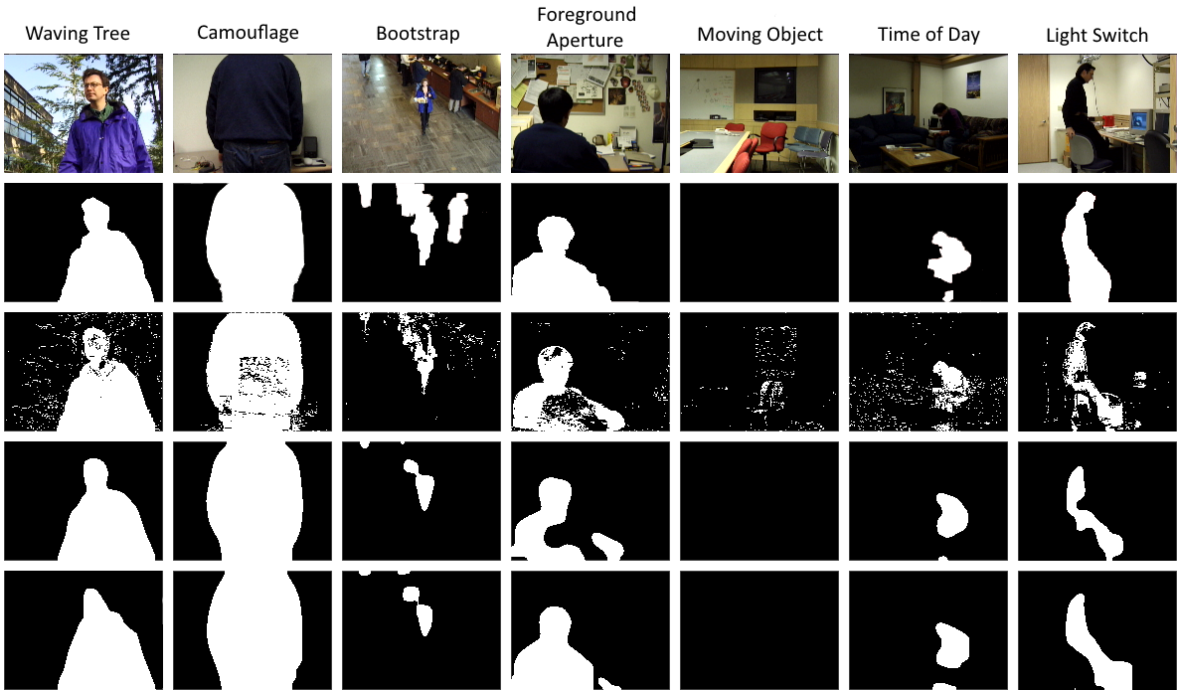

Figure 4: This figure depicts the results the proposed approach generated for the Wallflower data set. The first row shows the images from the videos, the second row the corresponding ground-truth data, the next the results after the GSM algorithm, the fourth row the results when the GSM is combined with the MRF (without the Between Class Variance) and the last row shows the results of the final algorithm.

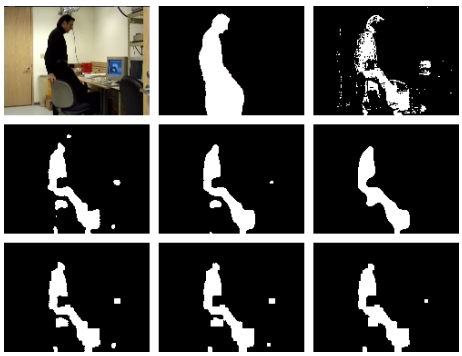

Figure 3: The first row shows the original picture, the ground-truth data and the result after the GSM. In the second row the results after the smallest clique Belief Propagation algorithm with the Moore Neighbourhood, first and second order GMN are depicted. The last row shows the same for the maximal clique Belief Propagation.

4 RESULTS

To compare our algorithm with existing methods the Wallflower data set (Toyama et al., 1999) is chosen because various different methods were already evaluated with it. Each of the seven examples depicts one unique problem of video segmentation which can be assessed with the provided ground-truth data.

As only one size of cliques is taken into consideration by the Belief Propagation, a comparison between two algorithms which use different clique sizes is made. One would only factor in the maximal cliques and another only the smallest cliques (C_2). Both algorithms achieved considerable improvements of the segmentation delivered by the Background Substraction and the extracted foreground regions are in general smoother if the minimal clique method is used (see Figure 3). For the final results the second order GMN and the algorithm which used the smallest cliques are applied.

The results for all video sequences and the three stages of our approach with one set of parameters are shown in Figure 4. It can be seen that the GSM alone generates good segmentations but still has many single false detections. A good example of this is the segmentation of the Moving Object scene.

In general, these can be very good eliminated with the MRF approach and only areas in which whole objects were not or falsly detected remain as errors (see Bootstrap or Foreground Aperture). As the results with the MRF are already quite good, the range of improvement for the between Class Variance addition to the Belief Propagation is limited. Nonetheless, it enhances the overall results considerably, mainly by expanding the foreground area if the surrounding pix-

Table 1: The results of different algorithms for the Wallflower data set. Each row shows the number of wrongly classified pixels for one approach separated in false positives and false negatives.

Algorithm		**MO**	**ToD**	**LS**	**WT**	**C**	**B**	**FA**	**Total**
Single Gaussian	FN	0	949	1857	3110	4101	2215	3464	
(Wren et al., 1997)	FP	0	535	15123	357	2040	92	1290	35133
Mixture of Gaussian (MoG)	FN	0	1008	1633	1323	398	1874	2442	
(Stauffer and Grimson, 1999)	FP	0	20	14169	341	3098	217	530	27053
Kernel Density Estimation	FN	0	1298	760	170	238	1755	2413	
(Elgammal et al., 2000)	FP	0	125	14153	589	3392	993	624	26450
MoG with PSO	FN	0	807	1716	43	2386	1551	2392	
(White and Shah, 2007)	FN	0	6	772	1689	1463	519	572	13916
MoG in improved HLS Color Space	FN	0	379	1146	31	188	1647	2327	
(Setiawan et al., 2006)	FP	0	99	2298	270	467	333	554	9739
MoG with MRF	FN	0	47	204	15	16	1060	34	
(Schindler and Wang, 2006)	FP	0	402	546	311	467	102	604	3808
Gaussian Switch Model (GSM)	FN	0	457	1636	244	829	1708	1567	
this paper	FP	466	641	543	736	164	166	561	9718
GSM with Belief Propagation (BP)	FN	0	394	1789	113	208	2064	1686	
this paper	FP	0	40	289	156	13	3	414	7169
GSM with improved BP	FN	0	321	1383	174	246	2081	469	
this paper	FP	0	199	695	356	66	0	92	6092
Independent Component Analysis	FN	0	1199	1557	3372	3054	2560	2721	
(Tsai and Lai, 2009)	FP	0	0	210	148	43	16	428	15308
Nonnegativ Matrix Factorization	FN	0	1282	2822	4525	1491	1734	2438	
(Bucak and Gunsel, 2009)	FP	0	159	389	7	114	2080	12	17053
Wallflower	FN	0	961	947	877	229	2025	320	
(Toyama et al., 1999)	FP	0	25	375	1999	2706	365	649	11478

els have a very similar colour.

An assessment of the results can be seen in Table 1 where the number of pixels which were wrongly classified are shown for all videos and many different approaches. The upper part shows various methods which are all using Gaussians for modelling the background. In the middle the results of our approach are illustrated and in the lower part are some methods which use completlely different principles (non-Gaussian) to model the background.

As the GSM method uses Gaussians for the background modeling it should be primarily compared to the Single Gaussian or MoG approaches shown at the top of the table. Although the GSM uses only two Gaussians and is therefore not as memory consuming as the MoG methods (which normally are tuned to five or more Gaussians), it still does perform better than almost all of them. Only one method could generate better results than the GSM algorithm when it was combined with the MRF and Otsu's Method.

5 CONCLUSION

A new and efficient way to model the background of a video with Gaussians is proposed and linked with a novel voting mechanism. The updated model is subtracted from every new frame and with a pixelwise adaptive threshold a segmentation can be created. In a second step the segmentation was improved by applying a higher order MRF on the generated data. There several adaptations were applied to obtain a manageable model even with advanced neighbourhood systems. Lastly, the Belief Propagation algorithm used to solve the MRF was extended by a process which would change the underlying probability map based on a local version of Otsu's method.

The benefit of the approaches using the MoG method should theoretically be mainly in cases like the Waving Tree video, where background objects are constantly in motion but do not become foreground. In this case the different backgrounds per pixel can be specifically modelled by the different Gaussians of the MoG. In practice, the GSM algorithm performs equally good there, although it only models one background per pixel.

The adaptions made to the MRF to simplify the calculation of the MAP made it possible to use this model on a today's standard PC and still achieve substantial and reliable improvements of the segmentation. Furthermore, it is demonstrated that the local version of Otsu's Method can alter the segmentation in a way that it aligns with the borders of the objects in the given scene. Overall this new approach delivers state of the art results in this well-studied subject.

REFERENCES

Boykov, Y. and Funka-Lea, G. (2006). Graph cuts and efficient n-d image segmentation. *International Journal of Computer Vision*, 70:109–131.

Bucak, S., Gunsel, B., and Guersoy, O. (2007). Incremental nonnegative matrix factorization for background modeling in surveillance video. In *Signal Processing and Communications Applications, 2007. SIU 2007. IEEE 15th*, pages 1–4.

Bucak, S. S. and Gunsel, B. (2009). Incremental subspace learning via non-negative matrix factorization. *Pattern Recogn.*, 42(5):788–797.

Cinar, G. and Principe, J. (2011). Adaptive background estimation using an information theoretic cost for hidden state estimation. In *Neural Networks (IJCNN), The 2011 International Joint Conference on*, pages 489–494.

Elgammal, A. M., Harwood, D., and Davis, L. S. (2000). Non-parametric model for background subtraction. In *Proceedings of the 6th European Conference on Computer Vision-Part II*, ECCV '00, pages 751–767, London, UK, UK. Springer-Verlag.

Felzenszwalb, P. and Huttenlocher, D. (2004). Efficient belief propagation for early vision. In *Computer Vision and Pattern Recognition, 2004. CVPR 2004. Proceedings of the 2004 IEEE Computer Society Conference on*, volume 1, pages I–261–I–268 Vol.1.

Huang, D.-Y., Lin, T.-W., and Hu, W. C. (2001). Automatic multilevel thresholding based on two-stage otsu's method with cluster determination by valley estimation. *Journal of Information Science and Engineering*, 17:713–727.

Ising, E. (1925). Beitrag zur Theorie des Ferromagnetismus. *Zeitschrift für Physik*, 31(1):253–258.

Kim, T.-K., Wong, K.-Y. K., Stenger, B., Kittler, J., and Cipolla, R. (2007). Incremental linear discriminant analysis using sufficient spanning set approximations. In *Computer Vision and Pattern Recognition, 2007. CVPR '07. IEEE Conference on*, pages 1–8.

Li, X., Hu, W., Zhang, Z., and Zhang, X. (2008). Robust foreground segmentation based on two effective background models. In *Proceedings of the 1st ACM International Conference on Multimedia Information Retrieval*, MIR '08, pages 223–228.

Liao, sheng Chen, T., and choo Chung, P. (2011). A fast algorithm for multilevel thresholding. *International Journal of Innovative Computing, Information and Control*, 7(10):5631–5644.

Lin, H.-H., Liu, T.-L., and Chuang, J.-H. (2002). A probabilistic svm approach for background scene initialization. In *Image Processing. 2002. Proceedings. 2002 International Conference on*, volume 3, pages 893–896 vol.3.

Marghes, T., B., and R., V. (2012). Background modeling and foreground detection via a reconstructive and discriminative subspace learning approach. In *Proceedings of the 2012 International Conferecne on Image Processing, Computer Vision and Patternrecognition*, pages 106–113.

Otsu, N. (1979). A threshold selection method from gray-level histograms. *Systems, Man and Cybernetics,IEEE Transactions on*, 9(1):62–66.

Schindler, K. and Wang, H. (2006). Smooth foreground-background segmentation for video processing. In *Proceedings of the 7th Asian Conference on Computer Vision - Volume Part II*, ACCV'06, pages 581–590.

Setiawan, N. A., Seok-Ju, H., Jang-Woon, K., and Chil-Woo, L. (2006). Gaussian mixture model in improved hls color space for human silhouette extraction. In *Proceedings of the 16th International Conference on Advances in Artificial Reality and Tele-Existence*, ICAT'06, pages 732–741.

Stauffer, C. and Grimson, W. (1999). Adaptive background mixture models for real-time tracking. In *Proceedings 1999 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Vol. Two*, pages 246–252. IEEE Computer Society Press.

Sun, M., Telaprolu, M., Lee, H., and Savarese, S. (2012). Efficient and exact map-mrf inference using branch and bound. In *Proceedings of the Fifteenth International Conference on Artificial Intelligence and Statistics (AISTATS-12)*, volume 22, pages 1134–1142.

Toyama, K., Krumm, J., Brumitt, B., and Meyers, B. (1999). Wallflower: Principles and practice of background maintenance. In *Seventh International Conference on Computer Vision*, pages 255–261. IEEE Computer Society Press.

Tsai, D. and Lai, C. (2009). Independent component analysis-based background subtraction for indoor surveillance. In *IEEE Trans Image Proc IP 2009*, volume 18, pages 158–167.

Viola, P. and Jones, M. (2004). Robust real-time face detection. *International Journal of Computer Vision*, 57(2):137–154.

White, B. and Shah, M. (2007). Automatically tuning background subtraction parameters using particle swarm optimization. In *Multimedia and Expo, 2007 IEEE International Conference on*, pages 1826–1829.

Wren, C., Azarbayejani, A., Darrell, T., and Pentland, A. (1997). Pfinder: Real-time tracking of the human body. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19:780–785.

Xu, W., Zhou, Y., Gong, Y., and Tao, H. (2005). Background modeling using time dependent markov random field with image pyramid. In *Proceedings of the IEEE Workshop on Motion and Video Computing (WACV/MOTION'05) - Volume 2 - Volume 02*.

Y. Wang, K.-F. L. and Wu, J.-K. (2006). A dynamic conditional random field model for foreground and shadow segmentation. *IEEE Trans. Pattern Analysis and Machine Intelligence (TPAMI)*, 28:279–289.

Yedidia, J. S., Freeman, W. T., and Weiss, Y. (2003). Exploring artificial intelligence in the new millennium. chapter Understanding Belief Propagation and Its Generalizations, pages 239–269.

Zhang, S., Yao, H., and Liu, S. (2009). Dynamic background subtraction based on local dependency histogram. *International Journal of Pattern Recognition and Artificial Intelligence*, 23(07):1397–1419.