Probabilistic Model-Based Background Subtraction

V. Krüger, J. Anderson, and T. Prehn

[1] Aalborg Media Lab, Aalborg University, Copenhagen
Lautrupvang 15, 2750 Ballerup
[2] Aalborg University Esbjerg, Niels Bohrs Vej 8, 6700 Esbjerg, Denmark

Abstract. In this paper we introduce a model-based background subtraction approach where first silhouettes, which model the correlations between neightboring pixels are being learned and where then Bayesian propagation over time is used to select the proper silhouette model and tracking parameters. Bayes propagation is attractive in our application as it allows to deal with uncertainties in the video data during tracking. We eploy a particle filter for density estimation. We have extensively tested our approach on suitable outdoor video data.

1 Introduction

Most vision systems work well in controlled environments, e.g., attempts to recognize humans by their face and gait has proven to be very successful in a lab environment. However, in uncontrolled environments, such as outdoor scenarios, the approaches disgracefully fail, e.g., the gait recognition drops from 99% to merely 30%. This is mainly due to low quality video data, the often small number of pixels on target and visual distractors such as shadows and strong illumination variations.

What is needed are special feature extraction techniques that are robust to outdoor distortions and that can cope with low-quality video data. One of the most common feature extraction techniques in surveillance applications is background subtraction (BGS) [5, 3, 8, 6]. BGS approaches assume a stable camera. They are able to learn a background as well as possible local image variations of it, thus generating a background model even of non-rigid background objects. During application the model is compared with novel video images and pixels are marked according to the belief that they are fitting the background model. It can be observed that BGS approaches are not able to distinguish between a foreground object and its shadow. And very often, the very same objects cause different outputs when the scenario changes: E.g. when a person walks on green grass or gray concret the outout can be severely different.

In this paper we present a Model-based Background Subtracting (MBGS) method that learns in addition to the background also the foreground model. A particle filter is used for finding and tracking the right silhouette.

In this work, without limit of generality, we consider only humans as objects and ignore objects that look different from humans. Also, we limit our discussion

F. Roli and S. Vitulano (Eds.): ICIAP 2005, LNCS 3617, pp. 180–187, 2005.

to silhouettes of humans as they deliver a fairly clothing-independent description of an individual.

The Model-based Background Subtraction System (MBGS System) consists of a learning part to learn possible foreground objects and a MBGS part, where the output of a classical BGS is verified using the previously trained foreground object knowledge.

To learn and represent the foreground knowledge (here silhouettes of humans) is non-trivial due to the absence of a suitable vector space. One possibility is to describe the data in a hierarchical manner, using a suitable metric and a suitable representation of dynamics between the silhouettes. Our approach for a hierarchical silhouette representation is inspired by [4, 13].

In the second part, we again consider the silhouettes as densities over spatial coordinates and use normalized correlation to compute the similarity between the silhouette density and the computed one in the input image. Tracking and silhouette selection is being done using Bayesian propagation over time. It can be applied directly since we are dealing with densities and it has the advantage that it considers the uncertainty in the estimation of the tracking parameters and the silhouette selection. The densities in the Bayesian propagation are approximated using an enhancement of the well-known Condensation method [7]. A similar enhancement of the Condensation method has been applied in video based face recognition [11].

The remainder of this paper is organized as follows: In Sec. 2 we introduce the learning approaches. The actual BGS method is discussed in Sec. 3. We conclude with experimental results in Sec. 4 and final remarks are in Sec. 5.

2 Learning and Representation of Foreground Objects

In order to make use of foreground model knowledge, our main idea is the following: Apply the classical BGS to a scenario that is controlled in a manner that facilitates the learning process. In our case, since we want to learn silhouettes of humans, that only humans are visible in the scene during training and that the background variations are kept as small as possible to minimize distortions. Then, we use this video data to learn the proper model knowledge.

After the application of a classical BGS, applying mean-shift tracking [1] allows to extract from the BGS output-data a sequence of small image patches containing, centered, the silhouette. This procedure is the same as the one used in [9], however, with the difference that here we do not threshold the BGS output but use probabilistic silhouettes (instead of binary ones as in [9]) which still contain for each silhouette pixel the belief of being a foreground pixel.

To organize this data we use, similar to [4], a combination of tree structuring and k-means clustering. We use a top down approach: The first level is the root of the hierarchy which contains all the exemplars. Then the second level is constructed by using a the k-means clustering to cluster the exemplars from the root. The third level is constructed by clustering each cluster from the second level, again, using k-means, see Fig. 1 for an example. The k-means clustering uses the Kullback-Leibler divergence measure which measures.

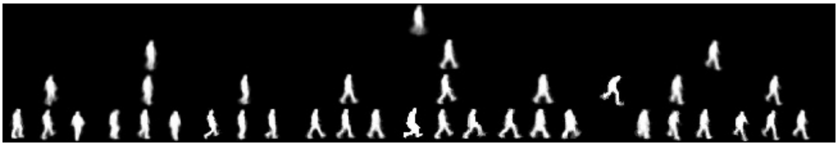

Fig. 1. An example of our clustering approach: 30 exemplars with K=3 and the algorithm stops after reaching 3 levels

Once the tree is constructed, we generate a Markov transition matrix: Assuming that the change over time from one silhouette to a next one can be understood as a first order Markov process, the Markov transition matrix M_{ij} describes the transition probability of silhouette s_j following after silhouette s_i at level l in the hierarchy. During MBGS application particle filtering [7,10,2] will be used to find the proper silhouette (see Sec. 3). The propagation of silhouettes over time is non-trivial, as silhouette do not form a vector space. However, what is sufficient, is a (not necessarily symmetric) metric space, i.e., given a silhouette s_i, all silhouettes are needed that are close according to a given metric. In the tree structure similar silhouettes are clustered which facilitates the propagation process. The Markov transition matrix M_{ij} on the other hand describes directly the transition likelihoods between clusters.

3 Applying Background Subtraction and Recognizing Foreground Objects

The MBGS system is built as an extension to a pixel based BGS approach. It uses foreground models to define likely correlations between neighbored pixels in the output $P(\mathbf{x})$ of the BGS application.

Each pixel in the image $P(\mathbf{x})$ contains a value in the range $[0,1]$, where 1 indicates the highest probability of a pixel being a foreground pixel. A model in the hierarchy can be chosen and deformed according to a 4-D vector

$$\theta = [i, s, x, y], \tag{1}$$

where x and y denote the position of the silhouette in the image P, s its scale, and i is a natural number that refers to a silhouette in the hierarchy.

We use normalized correlation to compute the distance between a model silhouette, parameterized according to a deformation vector θ_t and the appropriate region of interest in the BGS image $P_t(\mathbf{x})$, appropriately normalized.

In order to find at each time-step t the most likely θ_t in the image $P_t(\mathbf{x})$, we use Bayesian propagation over time

$$\begin{aligned} p(\theta_t|P_1, P_2, \ldots, P_t) &\equiv p_t(\alpha_t, i_t) \\ &= \sum_{i_{t-1}} \int_{\alpha_{t-1}} p(P_t|\alpha_t, i_t) \\ &\qquad p(\alpha_t, i_t|\alpha_{t-1}, i_{t-1}) p_{t-1}(\alpha_{t-1}, i_{t-1}) \end{aligned} \tag{2}$$

with $\alpha_t = [s, x, y]_t$. Capital "P_t" denotes the probability images while little "p" denotes density functions. We approximate the posteriori density $p(\theta_t|P_1, P_2, \ldots, P_t)$ with a sequential Monte Carlo method [2, 7, 10, 12]. Using Bayesian propagation allows to take into account the uncertainty in the estimated parameter. Monte Carlo methods use random samples for the approximation of a density function. Our MBGS system uses separate sample sets for each object in the input image. A new sample set is constructed every time a new object in the video image matches sufficiently well.

As the diffusion density $p(\alpha_t, i_t|\alpha_{t-1}, i_{t-1})$ in Eq. 2 we use the Brownian motion model due to the absence of a better one. For the propagation of the position and scale parameters, x, y, and s, this is straight forward. The propagation of the silhouette is, however, not straight forward. Inspired by [11] we use the variable i to reference a silhouette in the silhouette database. By considering the joint distribution of the silhouette id with the tracking parameter we are able to view the tracking and the recognition as a single problem. By marginalizing over the geometric parameters $\alpha = (x \;\; y)$,

$$p(i_t|Z_1, \ldots, Z_t) = \int_{\alpha_t} p(\alpha_t, i_t|Z_1, \ldots, Z_t) \;\; . \tag{3}$$

we can estimate the likelihood of each silhouette at any time. In [11] where the parameter i is constant, it is sufficient for the recogntion to wait until all particles of the Monte Carlo Markov Chain have converged to the same identity. This is equivalent to minimizing the uncertainty, i.e., the entropy.

In this problem setup, however, the correct silhouette parameter i is not constant but changes as the person walks. Therefore, we have to consider the two following issues: (1) We have to find a likelihood measure $P(i_t|i_{t-1})$ for the propagation step: given a silhouette i_{t-1}, what are the likelihoods for the other silhouettes. (2) We have to define an approach that allowes to approximate the density p_t with an MCMC technique. This is complicated because one wants the particles to clearly converge to the correct silhouette at each time step and at the same time wants enough diffusion in the propagation step to allow silhouette changes.

Issue (1) was partially solved in the learning process (see Sec. 2) where silhouettes were clustered and the Markov transition matrix $M(i, j)$ was computed which represents the transition from silhouette i to silhouette j.Then,

- the likelihood for selecting a silhouette from a certain silhouette cluster in the hierarchy is computed from the Markov transition matrix M by marginalizing over the silhouettes in that particular cluster.
- Within a cluster, the new silhouette is then chosen randomly.

The reason for marginalizing over the clusters is because our training data is too little so that the Markov transition matrix M appeares to be specific to the training videos.

In order to approach issue (2), we diffuse only 50% of the particles at each time step with respect to the silhouette number. Our experiments have shown that if the diffusion of the silhouette number was larger then there was often no clear convergence to one particluar silhouette. This, however, is needed to assure recognition of the correct silhouette at each time-step.

4 Experiments

In this section we present qualitative and quantitative results obtained from experiments with our MBGS implementation. The experiments clearly show the potentials of an effective MBGS approach. The purpose of the experiments was to verify that the drawbacks of the classical BGS approach, which were mentioned in section 1, can be remedied with MBGS. More specifically the MBGS system verifies the following: (1) Because shadows are not part of the model information provided, these will be classified as background by the implemented MBGS approach. In fact, most non-model object types will be classified as background, and therefore MBGS allows for effective object type filtering. (2) The output presented from the MBGS does not vary, even when the scenario changes significantly. If a model is presented as output, it is always presented intact. The object behind a silhouette is therefore always determinable.

Qualitative verification is done by comparing the AAU MBGS system with two previously developed pixel-based BGS approaches. One is the non-parametric approach developed at Maryland (UMD BGS) [3]. The other (AUE BGS), which utilizes alternative image noise filtering, has previously been developed at AUE.

Figure 2 shows a scenario, with a pedestrian walking behind trees, thereby at times being occluded. The output of two pixel-based approaches is shown in the lower part of the figure. Notice that the shadow cast by the pedestrian is classified as foreground by these pixel-based BGS approaches. Since the MBGS system operates by inserting a model as foreground, this problem is effectively resolved. Figure 3 shows the same scenario, in a frame where the pedestrian is heavily occluded. The occlusion causes the pedestrian to more or less disappear with pixel based approaches. This happens because the occlusion divides the pedestrian silhouette into separate smaller parts, which are then removed by the applied image filters. The scenario presented in figure 4, shows two pedestrians walking towards each other, thereby crossing behind lamp posts and a statue. When processing this scenario, a combination of image filtering and the background variation, renders the silhouettes of the pixel-based approaches unidentifiable. Also both pixel-based approaches severely distorts the silhouettes of the pedestrians. By only inspecting the pixel-based results, it is hard to tell that the foreground objects are actually pedestrians.

In a quantitative evaluation we have investigated the correctness of the particle method in matching the correct silhouette. When the MBGS is started, the particles are evenly distributed and the system needed usually 20-50 frames to find a sufficiently good approximation of the true density. Then, the selected silhouette is rather random. After 50 frames, the silhouette with the maximum likelihood is the correct one in $\approx 98\%$ of the cases. In $\approx 20\%$ of the cases the

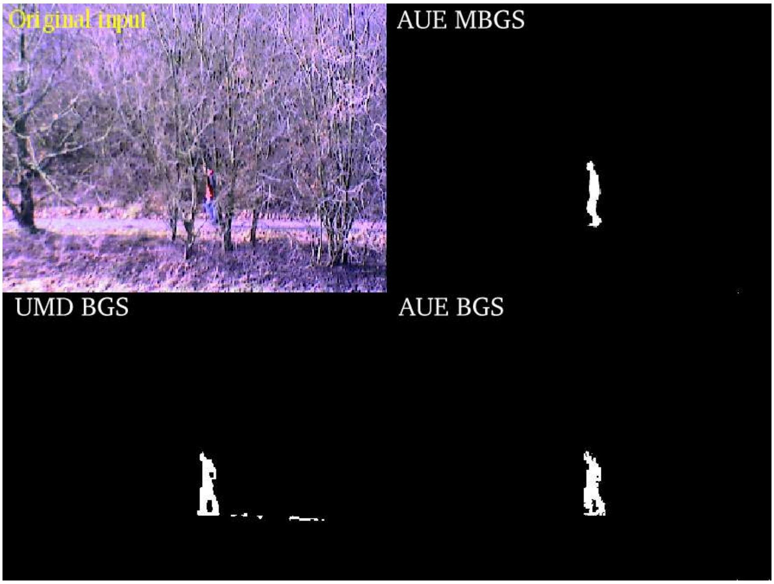

Fig. 2. BGS approach comparison of shadow issue

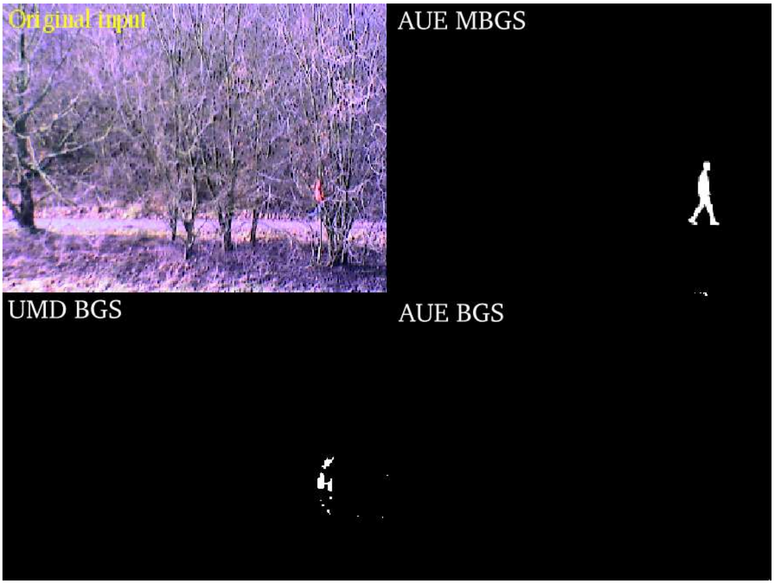

Fig. 3. BGS approach comparison of heavily occlusion

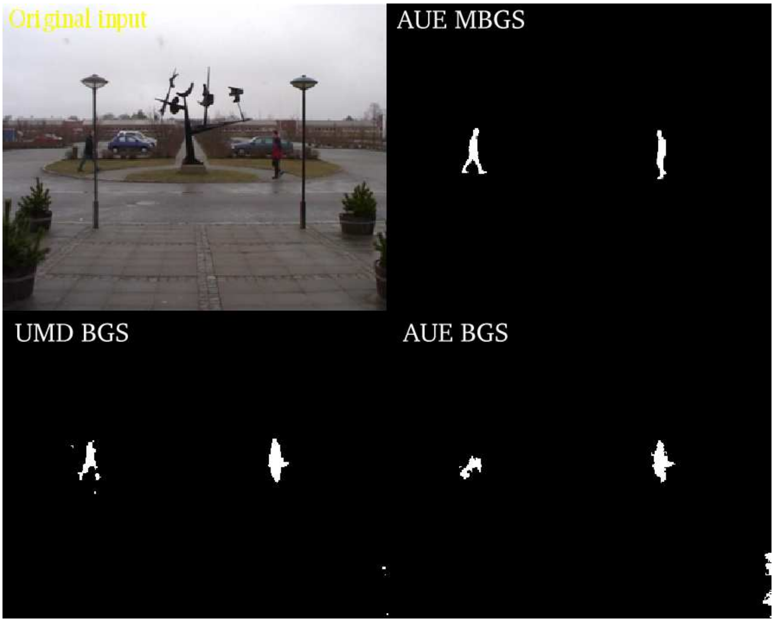

Fig. 4. BGS approach comparison of low contrast issue

ML silhouette was incorrect when e.g. a bush was largely occluding the legs. However, recovery time was within 5 frames. In case of partial occlusion of the entire body through, e.g. small trees, reliability degraded between 1% (slight occlusion) to 10% (considerable occlusion), The silhouette was incorrect in ≈ 59% of the cases where the legs were fully occluded, e.g. by a car. In the videos the individual was in average 70 px. high. Reliability increased with more pixels on target.

The system has been tested on a 2 GHz Pentium under Linux. In videos of size 320 × 240 pixels with only a single person to track, the system runs, with 350 particles, with ≈ 50 ms/frame: ≈ 25ms/frame were used by the classical BGS, ≈ 25ms/frame were used by the matching.

5 Conclusion

The presented model-based background subtraction system combines the classical background subtraction with model knowledge of foreground objects. The application of model knowledge is not applied on a binary BGS image but on the "likelihood image", i.e. an image where each pixel value represents a confidence of belonging either to the foreground or background. A key issue in this study is the clustering of the silhouettes and the temporal diffusion step in the Bayesian propagation.

In the above application we have chosen silhouettes of humans, but we belive that this choice is without limit of generality since even different object types fit into the tree structure.

The presented experiments were carried out with only a single individual in the database. We have experimented also with different individuals (and thus varying silhouettes), but the output was instable w.f.t. the choice if the individual. This is under further investigation and the use of our approach for gait recognition is future research.

References

1. Dorin Comaniciu, Visvanathan Ramesh, and Peter Meer. Real-time tracking of non-rigid objects using mean shift. In *Proc. IEEE Conf. on Computer Vision and Pattern Recognition*, volume 2, pages 142–149, Hilton Head Island, SC, June 13-15, 2000.
2. A. Doucet, S. Godsill, and C. Andrieu. On sequential monte carlo sampling methods for bayesian filtering. *Statistics and Computing*, 10:197–209, 2000.
3. A. Elgammal and L. Davis. Probabilistic framework for segmenting people under occlusion. In *ICCV*, ICCV01, 2001.
4. D. Gavrila and V. Philomin. Real-time object detection for "smart" vehicles. In *Proc. Int. Conf. on Computer Vision*, pages 87–93, Korfu, Greece, 1999.
5. I. Haritaoglu, D. Harwood, and L. Davis. W4s: A real-time system for detection and tracking people in 2.5 D. In *Proc. European Conf. on Computer Vision*, Freiburg, Germany, June 1-5, 1998.
6. T. Horprasert, D. Harwood, and L.S. Davis. A statistical approach for real-time robust background subtraction and shadow detection. In *Proceedings of IEEE ICCV'99 FRAME-RATE Workshop*, 1999.
7. M. Isard and A. Blake. Condensation – conditional density propagation for visual tracking. *Int. J. of Computer Vision*, 29:5–28, 1998.
8. Yuri A. Ivanov, Aaron F. Bobick, and John Liu. Fast lighting independent background subtraction. *Int. J. of Computer Vision*, 37(2):199–207, 2000.
9. A. Kale, A. Sundaresan, A.N. Rjagopalan, N. Cuntoor, A.R. Chowdhury, V. Krnger, and R. Chellappa. Identification of humans using gait. *IEEE Trans. Image Processing*, 9:1163–1173, 2004.
10. G. Kitagawa. Monta carlo filter and smoother for non-gaussian nonlinear state space models. *J. Computational and Graphical Statistics*, 5:1–25, 1996.
11. V. Krueger and S. Zhou. Exemplar-based face recognition from video. In *Proc. European Conf. on Computer Vision*, Copenhagen, Denmark, June 27-31, 2002.
12. J.S. Liu and R. Chen. Sequential monte carlo for dynamic systems. *Journal of the American Statistical Association*, 93:1031–1041, 1998.
13. K. Toyama and A. Blake. Probabilistic tracking in a metric space. In *Proc. Int. Conf. on Computer Vision*, volume 2, pages 50–59, Vancouver, Canada, 9-12 July, 2001.