# Adaptive model for foreground extraction in adverse lighting conditions

Stewart Greenhill, Svetha Venkatesh, and Geoff West

Department of Computing, Curtin University of Technology
GPO Box U1987, Perth 6845, Western Australia
{stewartg, svetha, geoff}@cs.curtin.edu.au

**Abstract.** Background elimination models are widely used in motion tracking systems. Our aim is to develop a system that performs reliably under adverse lighting conditions. In particular, this includes indoor scenes lit partly or entirely by diffuse natural light. We present a modified "median value" model in which the detection threshold adapts to global changes in illumination. The responses of several models are compared, demonstrating the effectiveness of the new model.

## 1  Introduction

The focus of this paper is on developing a robust foreground extraction module to be used in conjunction with tracking systems for tracking people in their home. These indoor environments do not have the controlled indoor conditions of office spaces or laboratories. By making long term recordings within a home we note that consideration must be given to the following factors:

– Illumination is spatially variable. A significant amount of the lighting is natural light, diffusing into the space via windows, doors, and skylights. The light sources are often very bright (ie. sunlight), although the overall level of illumination may be low. There is thus a large dynamic range in intensity which can lead to saturation in the images.
– Light sources may not be overhead. Most windows are close to ground level meaning that many objects are lit from the side. Objects (eg. people) may obscure light sources, casting broad shadows that are disconnected from the obscuring object.
– Illumination is temporally variable. Over the course of a day lighting is influenced by changes in external conditions due to clouds, shadows, and reflections. It is influenced by internal events such as the opening and closing of doors, windows and curtains, and the switching on and off of internal lighting.

A first step in motion analysis is to model the normal variation of the background of a scene. Two current approaches are to use a mixture of Gaussian distributions [1] or the median value over a short time window [2]. Both approaches have problems in situations where illumination changes rapidly. Adaptation in the mixture of Gaussians model is determined by a learning rate. To perform well under varying illumination, the number of distributions $K$ and the learning rate $\alpha$ must be adapted to match the time scale of the input, which is not known *a priori*. The median value approach uses a single

global threshold for foreground extraction. This is insufficient to deal with relatively long-term trends in the illumination level. An efficient alternative the Gaussian mixture model is to use clusters with varying weights [3] but these are subject to the same considerations with respect to learning rate.

This paper seeks to address the issues with these existing background elimination models. We use a median value model with an adaptive threshold to improve the quality of extracted foreground images. As part of this investigation a motion tracking system was implemented, and the performance of different background models was compared. This paper presents some results from this process, and describes an improved background elimination algorithm.

## 2   Background Elimination

The Gaussian mixture (GM) model [1] maintains a number of Gaussian distributions $K$ for each pixel. These are characterised by a mean $\mu$, variance $\sigma$ and weight $\omega$ which are adjusted over time as new data becomes available. The rate at which the models adapt is determined by $\alpha$, the *learning rate*. Small values of $\alpha$ make the model adapt slowly, favouring historical evidence over new evidence. Large values of $\alpha$ cause rapid adaptation, but can also introduce additional problems. If the learning rate is too high, the Gaussian models become too specific (ie. $\sigma$ becomes small) too quickly and the background model becomes unstable as models are continually invalidated by new evidence. In practice it is necessary to balance $\alpha$ and $K$ to cover the expected variation in the background.

The median value (MV) model [2] maintains a set $S$ of $N$ samples for each pixel. The background model is the value of the pixel that minimises the distance to all other pixels according to the *L-inf distance* in the RGB colour space:

$$Distance(a, b) = max(|a.c - b.c|) \quad c = R, G, B$$

A threshold $T_L$ is used to identify which image pixels are different from the background value, according to the same distance measure. This defines the image foreground. As stated, this technique has a relatively short term memory. Within roughly $N/2$ samples, any previous evidence is replaced by new evidence. That is, in response to a step change in the input, the median value will have shifted to the new state. Cyclic changes are handled if their period is less than $N$. Typically, $N$ is chosen to be a small number (eg. 8). The complexity of updating the background model is $O(N^2)$, but once this is done, frames can be classified in constant time. Therefore, it is usual to subsample the original input (eg. to one in 10 frames). In contrast, the cost of updating a GM model is $O(K)$ but the proportionality constant is higher due to the more complex calculation. It is uncommon to use more than 4 or 5 distributions in real-time applications.

The SAKBOT system [2] improves the stability of the MV background model by incorporating *adaptivity* and *object-level reasoning*. A background model $B_t$ is maintained separate to the statistical background model $B_t^s$. Foreground regions are classified as moving object, shadow, ghost, or ghost-shadow. The background model is
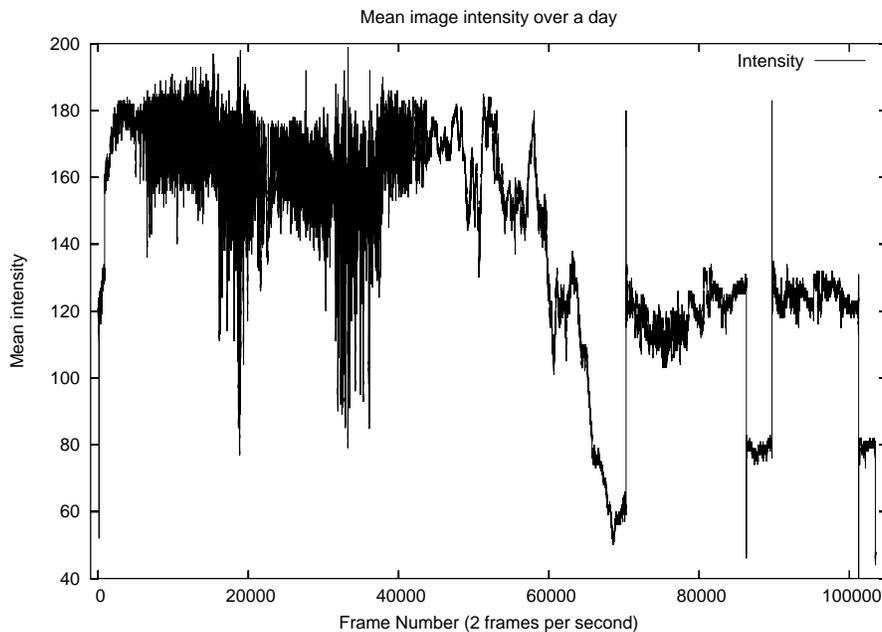
Mean image intensity over a day



**Fig. 1.** Mean illumination for an indoor scene over approximately 14.5 hours.

assigned the statistical background value $B_t^s$ for background, ghost, and ghost-shadow regions, and the previous background value $B_{t-1}$ for moving objects and their shadows. In this way, objects moving through the scene do not disrupt the background model. An adaptivity factor is included, adding $B_t$ to the set $S$ but weighting the distances used for the median function by a factor $\omega_b$.

## 3  Response of models to illumination changes

Both GM and MV models perform well in conditions of near constant lighting. However, rapid illumination changes produce a large disparity between the current pixel values and the background model. As a result, large areas of the image become classified as foreground. This interferes with object segmentation and tracking.

The problem arises from two sources. Firstly, under natural lighting external changes (due to shadows, clouds) cause significant changes in indoor illumination. The magnitude of the change is generally greater in the corresponding outdoor scene. Secondly, most cameras have an internal gain or aperture level that is adjusted to normalise the overall brightness of the image. The presence of a temporary bright object can cause the camera to suddenly change gain level.

Figure 1 shows the mean illumination for an indoor scene over approximately 14.5 hours. Mean illumination $\bar{I}$ is defined as the average over all image pixels of:

$$\sqrt{r^2 + b^2 + g^2}/\sqrt{3}$$

where $r$, $g$, and $b$ are the values of the red, green and blue image channels. Most objects in the scene are small compared to the image size, so object motion generally has a small effect on mean illumination.

Over the course of a day $\bar{I}$ is influenced by changes in external conditions due to clouds, shadows, and reflections. It is influenced by internal events such as the opening and closing of doors, windows and curtains, and the switching on and off of internal lighting. In addition, since lighting is not always overhead people can temporarily obscure light sources (eg. by walking in front of windows).

Both GM and MV models eventually adapt to changed lighting conditions. The GM model generally has a slower response time, but maintains multiple models of background state so can be more stable under repetitive changes. The MV model responds relatively quickly (roughly $N/2$ frames), but both models produce disrupted images while adaptation takes place. As shown in the figure the fluctuation in illumination may be 20 to 50% of the mean value. The time scale of the fluctuations varies from a few seconds to a few minutes.

## 4 Adaptive median value model

To improve image quality of the MV model during adaptation, we employed a correction to the detection threshold $T_L$ that adapts in response to global illumination changes. The system works as follows:

The mean illumination $\bar{I}$ is computed for each image. The original model [2] maintains a history $S$ of $N$ images. In addition, we maintain a history $H$ of $\bar{I}$ for the previous $M = N/2$ images. The difference $D$ between the largest and smallest value in $H$ approximates the worst-case disparity between the current and historical median value of $\bar{I}$. This value $D$ is scaled by a correction factor $cf$ and added to $T_L$ to correct for differences due to shifts in mean illumination. This causes the detection threshold $T_L$ to increase during periods of rapid change by an amount that is proportional to the difference in mean illumination.

Figure 2 shows the performance of various background models in response to natural changes in illumination. In this scene, no objects are moving, so any foreground pixels are due to artifacts of the background model. A measure of the disruption to the image is the proportion of the total image area detected as foreground. Ideally, this should be small. The figure shows values obtained by four background models working on the same image sequence. The best performance is obtained by the MV model with $cf = 2$. This peaks at only 6% (around frame 200 and 900), whereas all other models peak at between 85 and 90%. The "fast" GM model ($\alpha = 0.1$) adapts well to rapid changes, and is significantly better than the "slow" GM model ($\alpha = 0.01$) over long disturbances. Having less "inertia", the MV models respond better to long disturbances.

Figure 3 shows the response to sudden illumination changes. In this experiment an overhead light is turned on and off and a bright object (a white tray) is carried across the room, inducing a step change in the camera gain of about 5%. There is also some variation in natural lighting. Foreground pixels are classified as object or shadow. Shadow pixels are not included in the foreground area, so the overall performance of all models is better than the unclassified foreground values. Again, the adaptive MV model
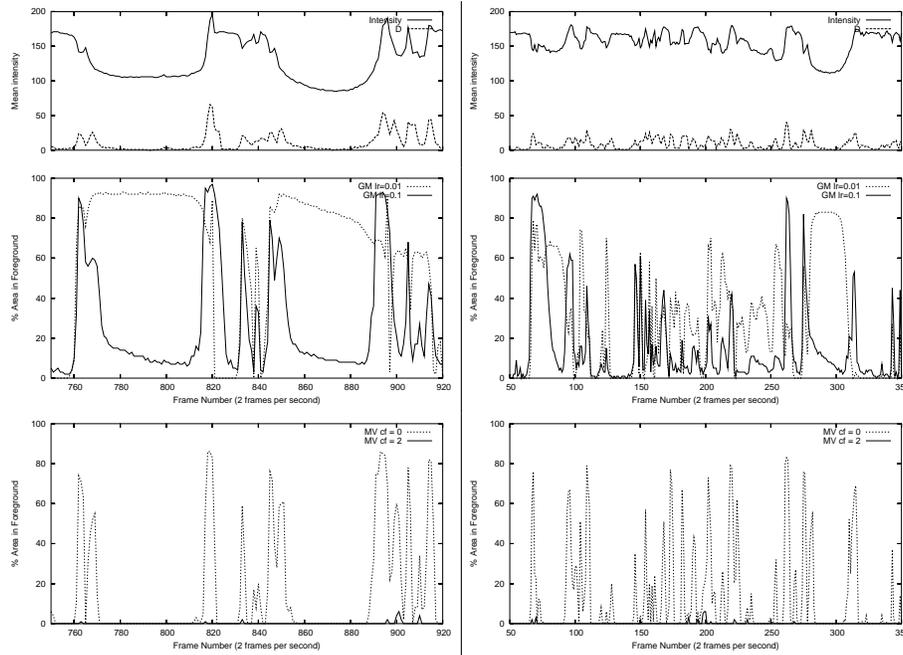
**Fig. 2.** Background elimination under natural illumination changes. Columns show two different time periods. Rows show intensity $\bar{I}$ and correction $D$ (top), GM performance (centre) and MV performance (bottom). Scene has no moving objects. Foreground is all non-background pixels.

($cf = 2$) shows least disruption to the image. This time the "slow" GM model outperforms the "fast" GM model, illustrating that the performance of the GM model depends on the match between model parameters and the input. In this scene there are moving objects which is an additional source of perturbations of the background model.

Figure 4 shows the effect of a rapid change in illumination (here, a shift in camera gain) on the background model. This occurs at frame 363 in the sequence shown in Figure 3. Since the scene becomes rapidly darker, there is a large disparity between the current frame and the background model. The background regions are shown in white in the bottom 4 image rows. Foreground regions are classified as shadow (grey) or object (black). An object moves from right to left, leaving a "ghost" in its initial location. Importantly, changing the MV foreground threshold $T_L$ does not adversely affect the ability of the system to track objects moving in the scene. Again, the corrected MV model shows the least disruption to image quality. In this case, a decrease in illumination levels can be handled by normal shadow reduction techniques [4]. For an increase in illumination, the shadow regions shown here would not be distinguishable from foreground objects.
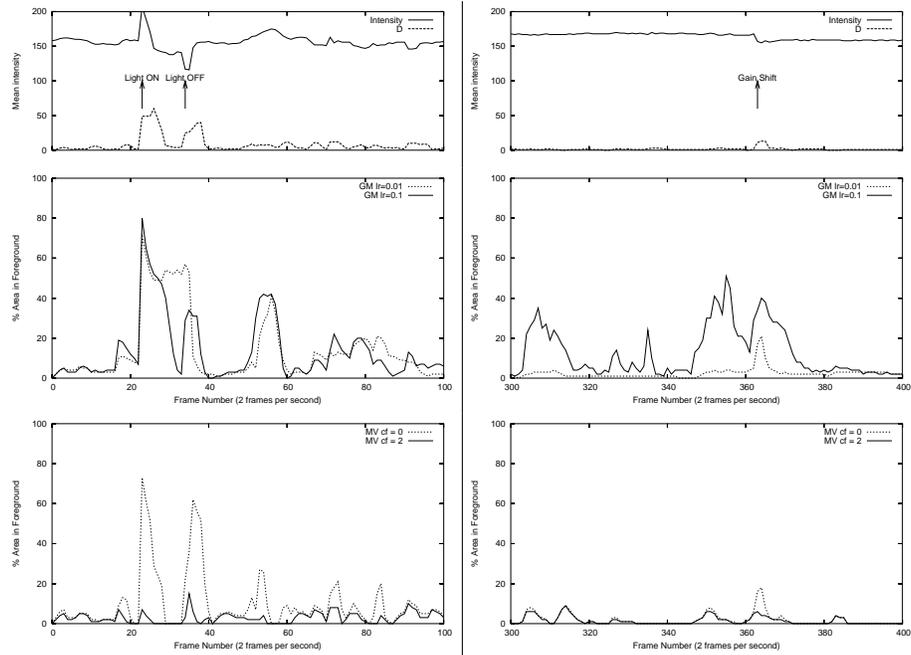
**Fig. 3.** Background elimination under imposed illumination changes. Left column includes lighting change. Right column shows response to camera gain shift (frame 363). Rows show intensity $\overline{I}$ and correction $D$ (top), GM performance (centre) and MV performance (bottom). Scene includes moving objects. Foreground areas exclude "shadow" pixels. See Figure 4 for classified images.

## 5 Conclusion and Future Work

This paper describes the response of various background elimination models to adverse, real-world lighting effects. Models are compared according to their ability to reject false foreground objects under rapid illumination changes. An adaptive median value model (MV) is described, which consistently performs better than for the uncorrected MV model, and Gaussian mixture (GM) models.

The MV model is attractive because it is computationally less intensive than GM models. MV has a global foreground threshold ($T_L$) which is easily adapted to lighting changes, as we have presented here. It may be possible to similarly improve GM models by perturbing the distribution means, although it is less clear over what time scales this might be possible. The full SAKBOT model [2] augments the MV statistical model with feedback. As such, it becomes vulnerable to rapid changes, and we have not yet studied the extent of this problem.

The adaptive MV model tends to suffer when rapid illumination changes are spatially non-uniform. One possible solution to this problem is to compute separate corrections for sub-regions of the original image. This is an area for future work.
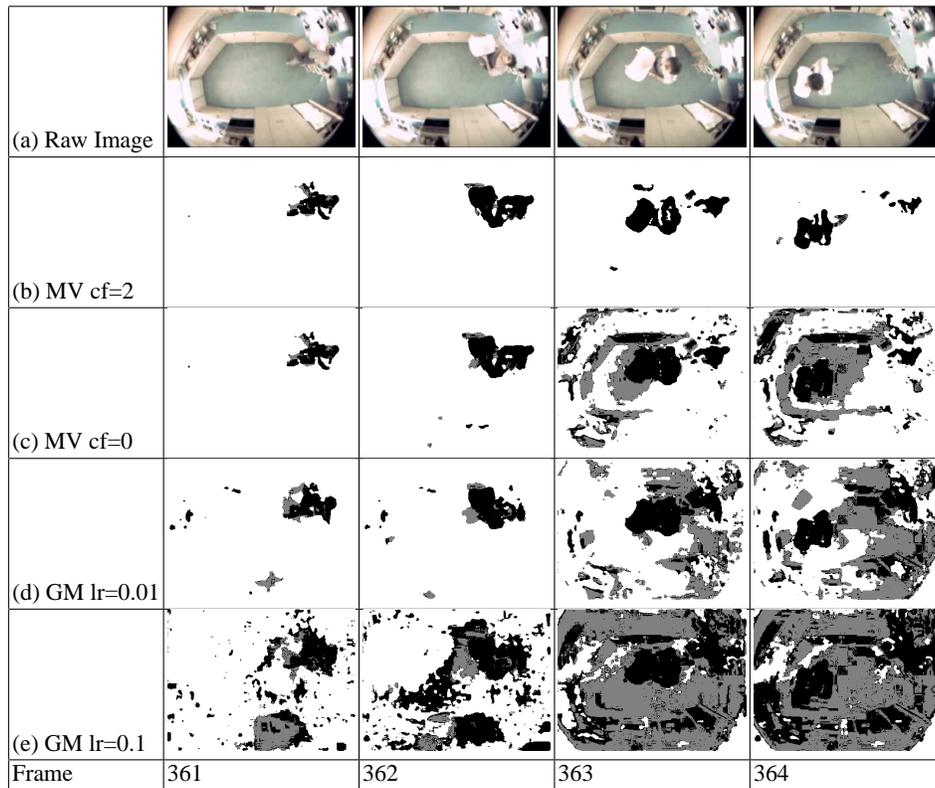
| | | | | |
|---|---|---|---|---|
| (a) Raw Image | | | | |
| (b) MV cf=2 | | | | |
| (c) MV cf=0 | | | | |
| (d) GM lr=0.01 | | | | |
| (e) GM lr=0.1 | | | | |
| Frame | 361 | 362 | 363 | 364 |

**Fig. 4.** Foreground (black) and shadow (grey) regions resulting from a camera gain shift. Note that a significant proportion of foreground pixels are removed by shadow reduction.

# References

1. Stauffer, C., Grimson, W.E.L.: Learning patterns of activity using real-time tracking. IEEE Transactions on Pattern Analysis and Machine Intelligence **22** (2000) 747–757
2. Cucchiara, R., Grana, C., Piccardi, M., Prati, A.: Detecting moving object, ghosts and shadows in video streams. IEEE Transactions on Pattern Analysis and Machine Intelligence **25** (2003) 76–81
3. Butler, D., Sridharan, S., V. Michael Bove, J.: Real-time adaptive background segmentation. In: Proceedings International Conference on Acoustics, Speech and Signal Processing (ICASSP 2003). (2003)
4. Prati, A., Mikic, I., Cucchiara, R., Trivedi, M.M.: Analysis and detection of shadows in video streams: A comparative evaluation. In: IEEE Computer Vision and Pattern Recognition Conference, Hawaii (2001)