Background Subtraction: Various Methods for Different Inputs

Nathan Johnson

Computer Science & Engineering, 301 Main Street
Columbia, South Carolina, USA 29208
johnso66@cse.sc.edu

Abstract. With computer vision the problem of speed comes up quite a bit. Images typically only contain a small amount of relevant information which needs to be processed. To help limit the information to be processed a method of pulling out the relevant information is needed. Background subtraction is an often used method to limit the information to be processed, since the background does not contain any "new" information; however, deciding whether an object is in the foreground or background remains a problem. Several methods have been proposed, but there does not exist a perfect method as of yet. Some require extra hardware, require extensive preprocessing, or only perform in certain cases. Two methods of background subtraction looked at in depth include the Gordon, et al. and Wallflower methods. The value of the method is determined not only for its ability to handle the typical problems of background subtraction, but also whether extra hardware is required.

1 Introduction

Images and video contain a cornucopia of information and to process this information takes time. Due to the redundancy of the material much of it can be ignored without changing the data to be examined more thoroughly. The task of removing this extra data is a difficult task. In this paper many different methods to reduce or remove this extra information are examined and compared.

The largest area in most videos is the background. Backgrounds are often fairly static and offer very little in the way of useful information. These methods attempt to segment the frame into the background and the foreground. Performing this also allows the possibility of inserting a virtual background without the use of a blue/green screen typically used currently.

These different methods range from a simple software change to affect previously recorded video streams to requiring addition hardware during the recording of the video stream one wishes to process. While comparing the different methods the hardware requirements are examined. The ability to use more than one method cooperatively is also reviewed.

There exists problems one must consider when trying to perform background subtraction. These problems include: lighting, shadows, gradual or sudden illumination changes, camouflage, moving objects in the background, foreground aperture, bootstrapping, and when a foreground object becomes motionless [1,2,3]. An algorithm to solve all of these problems is something still being researched, but there are partial solutions to one or more of these

2 Methods

A pixel is considered part of the foreground if it is far from the background model. In equation form this can be seen in Equation 1. P_i is the pixel in the image being processed, P_m is the pixel value in the background model, k is the threshold, and σ is the variance. This equation is a simplified form of what every algorithm to perform background subtraction attempts to perform. If a pixel is significantly different from the background model then it is considered to be part of the foreground.

$$F \equiv |P_i - P_m| > k\sigma \tag{1}$$

2.1 Classical

The classical method of removing the background from an image during preprocessing is done at the pixel level, or more precisely the luminosity of the pixel. A running average is kept on the image and any pixel matching the average within a predefined tolerance is considered to be foreground. This way of performing background subtraction has many disadvantages, but it also has the advantage of being very simple and easy to implement.

The drawbacks are mainly found in the types of problems that it cannot perform adequately; sudden and gradual illumination changes being one of the most obvious. A room with daylight streaming in or a light switch being toggled would be examples of these problems. The results of a sudden illumination change while using the classical method is the entire image to be considered foreground. A gradual change results in a similar, although less drastic, foreground image as the sudden change. Shadows in an image introduce similar problems causing an object to be considered foreground due to the slight change in luminosity.

Other situations where the classical method does not work include: foreground aperture, foreground object becomes motionless, and when background objects move slightly. Since the classical method only looks at an image at the pixel level no higher level statistical models are used to determine whether a pixel is in the foreground. In addition to not adapting to more complicated models, this method handles some of the other problems better than some of the other methods, including camouflage. Camouflage is meant as a way to fool an onlooker by appearing similar to the background, but when one only looks at a very small area, for example a 3x3 pixel area, then the same pattern will still stand out quite easily if it is positioned slightly differently than the original averaged background.

2.2 Wallflower

A method like Wallflower [1] is a much higher level algorithm for removing background information from a video stream or image. It has three main components when deciding whether a pixel is in the foreground: pixel, region, and frame levels. In developing this method a modified classical method is used named Linear Prediction.

The pixel level uses Linear Prediction [1], making the initial judgment on whether a pixel is in the foreground. Linear Prediction is very similar to the classical method except that two running averages are kept. The first average is the same as the classical method, being the running average of the true pixel values. The other is an average of the predicted values. The predicted average and the true average are both used to come up with the next prediction.

After the initial choice of each pixel a more refined judgment is then made at the region level. This takes into account neighboring pixel information to help discern the status of a pixel. This level helps solve the foreground aperture problem. With the foreground aperture problem the outside of an object is able to be detected as foreground, but the interior may not be high textured and therefore is not able to be discerned from its original position, so it is marked as background. With the region level the notice of the outside border being part of the foreground is noticed and morphological operations are performed to attempt to fill in the inside since an object with a boundary that is in the foreground is more than likely also foreground.

The final level considers the entire frame and attempts to match predefined models to the change in the pixels to see if any match. If a model matches the changes in the current frame then the decision of a pixels foreground can be adjusted to deal with the change. An example of one of the models includes one that checks to see if all pixels change in intensity by a uniform amount, within a certain tolerance. An image matching this model would signify that there existed a sudden illumination change, so consider that change in pixel value to not be of any consequence when considering whether something is in the foreground.

The results of running the mean & threshold (classical method), Wallflower, and Linear Prediction algorithms are a series of test images representing one of the problems with previous background subtraction methods can be seen in Figure 1. These results show the advantages of using these modified algorithms and the number of problems that they solve. The algorithms are not perfect and sometimes the mean & threshold still provides a better result.

The first problem addressed is a background object is moved, but it is still considered to be within the background. All three algorithms adapted to this change in the background model appropriately and still consider the moved object part of the background. The only difference between the three algorithms is the running time, but since the Wallflower algorithm is not meant as a real-time solution, the running time is not an issue.

The next two problems addressed deal with changes in lighting, gradual and sudden. The mean & threshold performs very poorly in this method, because it cannot adapt quick enough for the gradual change and believes that everything is foreground in the sudden case. Linear Prediction performs slightly better, but still not completely correct. The gradual case shows that it does adapt quickly enough for slower changes, but the sudden change still causes most of the image to be considered foreground. Actually the only part considered background is actually the section that should be foreground. The Wallflower algorithm truly shines in these cases. In the time of day it is able to adapt the same way Linear Prediction adapts, but in the sudden change the Wallflower method lets the frame level take control and adjusts the input to match with this new found problem and thus is able to correctly identify the foreground.

The last three cases provide varying results. The waving trees and bootstrapping problems do not perform too well in any of the three. The Wallflower and Linear Prediction methods perform slightly better, but nothing that indicates a dramatic, more accurate result. Camouflage is one case where the mean & threshold performs the best, since it does not try to perform more advanced comparisons with the image. The foreground aperture is the last case where the Wallflower method actually performs much better than the other two. This is due to the region level and the morphological processing that is performed, to help fill in the missing segment inside the outline.

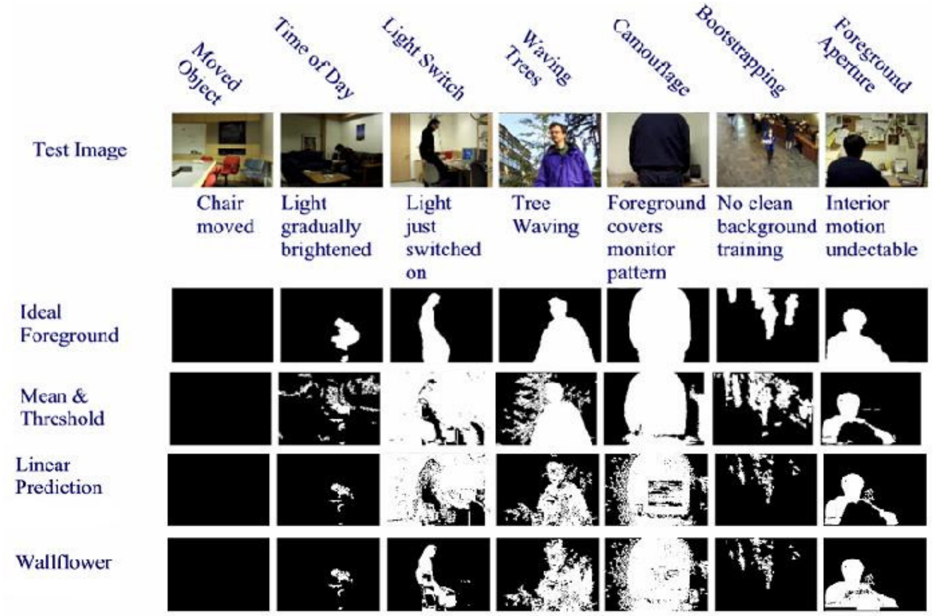

Figure 1: Wallflower & Linear Prediction results against the classical problems

2.3 Gordon, et al.

The Gordon, et al. method [2] is the first method that is evaluated that required additional hardware for the input. This requirement is due to the need of depth information in the video stream. With the depth information several of the problems can be solved including lighting and shadows, bootstrapping, and when a foreground object becomes motionless.

A pixel's luminosity is generally the only information available about a pixel in video streams, but introducing the additional hardware to record depth each pixel contains information about the pixel's color (Red, Green, and Blue) and the depth (Z). With this depth information a decision can be made without relying on the luminosity, which in turn prevents an object to be marked as foreground, or background, from shadows or reflections in the frame. A pixel's depth is used only if the pixel has a valid depth, otherwise the method falls back on classical methods for background subtraction. A depth is considered valid at a pixel if there is enough information in the image to accurately determine the pixel's depth. An area where this method fails includes areas without a high texture.

Using only the range or luminosity will often produce similar results, an image in the foreground with holes in the object where part of it is considered to be a section of the background. With both forms of background subtraction, range and luminosity, the resulting image is much better with many of the gaps in the image are filled in since the reason for these bad areas is usually due to the way the information is not sufficient to determine the end result. This can be seen in Figure 2.

A pronouncement using the range may still be able to be done even if the range is invalid in the background model. If the background model has an invalid background range, r_m, and the pixel in the input image, r_i, is valid and smoothly connected to areas that have valid background ranges then a range decision can be made using the neighboring regions. The equation for determining the foreground value using range data can be seen in Equation 2.

$$F_r \equiv Valid(r_i) \wedge (\nabla r_i < G) \wedge \neg(Valid(r_m) \wedge (|r_i - r_m| < k\sigma) \qquad (2)$$

If r_i is invalid or r_m is invalid and not smoothly connected to other valid regions, then the classical color methods are performed. Since there may still exist areas of shadow or reflections that change the luminosity, this method puts in two ratios, one for shadows and one for reflections. The reason for two separate ratios is due to the fact that shadows affect the intensity of the light much more than reflection from a nearby object. The equation for calculating the value for the foreground using color can be seen in [4].

$$YValid(Y) \equiv Y > Y_{\min} \qquad (3)$$

$$\begin{aligned} F_c \equiv\ & (YValid(Y_m) \wedge YValid(Y_i) \wedge \Delta color > c\sigma)) \vee (YValid(Y_m) \wedge \\ & ((\frac{Y_i}{Y_m} < shad) \vee (\frac{Y_i}{Y_m} > reflect)) \vee (\neg YValid(Y_m) \wedge (Y_i > \alpha Y_{\min})) \end{aligned} \qquad (4)$$

Examples of how this algorithm improves the traditional methods of segmenting based only on color, or even some of the newer range based methods is shown in Figure 2 below. The top left image in each group is the current background model and the top right image is the current image being processed. The second row shows the depth information provided for each image. The white regions in the depth images represent a pixel that has an invalid depth.

The first set of images (on the left) displays the result of using luminosity only based segmentation on the bottom left. The missing areas occur from areas in the foreground object that match the background in luminosity. Combining it with the depth information the missing areas are filled in since it can be seen that the depth information is valid at those points and the method does not consider the color of the pixels before marking it as foreground.

The second set of images (on the right) displays the result of using range only based segmentation. Due to the lack of high texture objects, a significant portion of the background model and the current image have invalid depths. The lack of valid depths causes missing areas in the result, however, since the difference in color exists; falling back on the traditional methods allows these areas to be filled in.

Both of these examples show cases where the combination provides excellent results, however, there can still be errors in the results. If the depth information is insufficient and the color of the object matches the background then there will exist missing regions that cannot be determined. Another problem with this method is the addition of a "halo" affect when using range information.

Figure 2: Results using only color (bottom left), range & color (bottom middle, left), only range (bottom middle, right), and range & color (bottom right)

2.4 Fast, Lighting Independent

The method developed by Ivanov, Bobick, and Liu [3] is an example of a very restrictive, but effective in the one problem it is meant to solve. The purpose of their method is to provide a way to remove the background from an image in an environment where the lighting is constantly changing and may be erratic, for example a theatre where the lights are used to convey meaning. In a setting such as this the more conventional methods of background subtraction fail.

Like the Gordon, et al. method [2] a second camera is required to record the depth information. The Fast, Lighting independent method [3] uses the depth information alone to segment the foreground. The other requirement of this method is the need to know at any given point the current state of the background. In the theatre setting this is not an unreasonable requirement, but in general cases this cannot be done since this requirement assumes a closed world model, which does not exist in the general case.

3 Comparisons

3.2 Wallflower and Gordon, et al.

These two methods, although they approach the problems in different ways, the Wallflower [1] using statistical models to explain and judge a frame and the Gordon, et al. [2] using extra information, depth, to make the same judgment. Due to these differences in approach the results they produce are very different. The Gordon, et al. can only handle the gradual or sudden illumination change if there is a valid depth, but in bootstrapping and cases where you never start with a clean background image to produce your background model Gordon, et al. is able to produce an adequate model whereas Wallflower will contain many errors until it is able to obtain a clean background image for the model.

These two methods are not vastly different, however, and could be used in conjunction. If one were to combine the two methods by using the Gordon et al. method as the initial judgment, then in cases where one needed to fall back on the traditional methods the Wallflower method is used. This change would not require drastic changes in either algorithm and would allow for the benefits of both algorithms without introducing any new problems that the joined Wallflower-Gordon, et al. method cannot solve.

3.2 Fast, Lighting Independent

The Fast, Lighting Independent algorithm [3], due to its constraints, cannot be easily combined with any other algorithm and therefore is used by itself. In the one case it was designed for, it performs adequately, considering the conditions. Occasionally a poor judgment is made or an object will contain missing segments that were mistakenly considered to be background, but when only using the one method this is difficult to avoid. This method is an example of how there exists many different methods being researched currently that solve specific problems, but a general solution is still out of grasp.

4 Modifications

The focus of this section is on the Wallflower [1] and the Gordon, et al. [2] methods since they are more general than the Fast, Lighting Independent method [3]. The modifications suggested include both hardware and software changes and the reasons for these modifications. Some of the changes are meant to provide better results, while others are considered to provide a different hardware requirement.

As mentioned earlier, the Gordon, et al. and Wallflower methods can be combined. This combination would slow down the speed, but would drastically increase the output of the images and the problem set that is able to be solved by the new modified method. This is a fair trade off, considering this method can be used on previously recorded video streams.

Focusing on the Gordon, et al.'s requirement of a secondary camera to provide the depth information we can change this requirement to one camera, using a new kind of camera developed by Philips, the Fluid Focus [4]. This lens uses two materials that do not mix and a small electric current to change the concavity of the lens. With this lens and the knowledge of the focal length at certain voltages we can use a single camera switching back and forth rapidly and ignore all blurred images. Although this does present a new problem; appropriately segmenting the blurred areas of the image that are meant to be ignored without discarding data to be processed. The benefit of this change is the required footprint of the apparatus. This would be a good setup for robots that can not be very large, but could use the background subtraction provided by the Gordon, et al. method.

4 Conclusion

The methods discussed in this paper are all methods that are mainly for preprocessing purposes. The speeds of the algorithms vary, but are not meant as a complete real-time solution. Real-time background subtraction methods do exist, but are just as specialized as the Fast, Lighting Independent method [3]. The real-time methods, like [5], do not perform as well as [1], [2], or even [3], but they do limit the data set that needs to be processed, and that provides one of the goals of background subtraction.

As far as which method is the best, this is a matter of the problem you are trying to solve. In regards to a general case, each method provides something the others do not, or has an additional requirement that the others do not. The more information one is able to record into the video stream the more methods that are possible to perform.

References

1. Toyama, Kentaro, Krumm, John, Brumitt, Barry, Meyers, Brian. "Wallflower: Principles and Practice of Background Maintenance," IEEE Conference on Computer Vision, 1999.

2. Gordon, G., Darrell, T., Harville, M., Woodfill, J. "Background estimation and removal based on range and color," in IEEE Computer Society Conference on Computer Vision and Pattern Recognition. Fort Collins, CO, June 1999.

3. Ivanov, Yuri, Bobick, Aaron, Liu, John. "Fast Lighting Independent Background Subtraction," in International Journal of Computer Vision, 37(2), 199-207, 2000.

4. Philips Research, Press Release on FLuidFocus found on http://www.research.philips.com/newscenter/archive/2004/fluidfocus.html, March 3, 2004.

5. Martins, Fernando C. M., Nickerson, Brian R., Bostrom, Vareck, and Hzra Rajeeb. "Implementation of a Real-time Foreground/Background Segmentation System on the Intel Architecture."