A Robust Motion Detection Algorithm for Complex Background Using Statistical Models

Zhen Yu, Yanping Chen
Dept. of Automation
Xiamen University
Xiamen, China
yuzhen20@vip.sina.com, chen_yanping1@163.com

Abstract—**Based on the fact that most of the algorithms assume that the camera is fixed and the changing background is learned in the training period, a robust algorithm is proposed for complex background where a shaking camera, changing background and shadows are presented. It combines a new improved mixture of Gaussians model and a square neighborhood matching algorithm to eliminate shadows and reduce false positive detections caused by camera motion and changing background. Experiments results demonstrate the efficiency and accuracy of this algorithm.**

Keywords—**Motion detection, Camera motion, mixture of Gaussians, square neighborhood matching.**

I. INTRODUCTION

Interesting motion detection is the foundational work in automated video surveillance and has become a hot topic for scientists in the computer vision community for decades.

Researchers have already brought forward many algorithms for detecting moving objects, including consecutive frames subtraction [1,2,3,4], optical flow approach [7,8,9], background subtraction [10,11,12,13],etc. Recent methods represent the boundaries of moving objects by active contours [5]. This method can be adapted to the case of multiple moving objects by using the popular level-set methods [6].Among all the methods, background subtraction algorithms are the most popular, since they are quite efficient and relatively simple in computing in a static scene. However, it is not natural that the background is always static as assumed. Shaking cameras, waving trees, lighting changes are quite common in real world, yet they may cause serious problems to these algorithms. Although many of the algorithms have taken into account the common disturbing factors that occur in nature, most of them don't address the problems of shaking cameras, moving background that is not modeled in the training period and shadows. Often, additional measurements are taken to deal with these problems. But most of these measurements don't take the advantage of the detection algorithm, thus may lower the efficiency and increase computational complexity.

The rationale of background subtraction is to build a statistical model to represent the pixels of the background. Some of the most classic statistical models include the single Gaussian distribution model, multi Gaussians model, a mixture of Gaussians, the kernel density estimation and non-parametric model.

The single Gaussian distribution model store a single mean per pixel, and is more likely to creating false detection in a changing background. Although multi Gaussians model build several Gaussians for every pixel, each Gaussian of the model represents a different object of the pixel, and only one Gaussian belongs to the background. Multi Gaussians can only represent unimodal background, which is similar to the single Gaussian model to some degree. Non-parametric model calculates the average probability of the input pixel compared to its corresponding samples, so as to avoid the chosen of the parameters for each Gaussian. Experimental results show that non-parametric model has lower false alarm rate, but it is complex in computing, and is not very satisfactory for real-time detection. Here, we choose the mixture of Gaussians as the background representation. The mixture of Gaussians model was put forward by Stauffer and Grimson in 1999 in [16], and has been discussed fully from then on. Its outstanding performance has been demonstrated in many applications in moving backgrounds. The comparison of the background modeling methods is shown in TABLE I.

TABLE I. STATISTICAL MODELS AND PERFORMANCE ANALYSIS

Models	Speed	Accuracy
Single Gaussian	O(I)	Low
Multi Gaussians	O(K)	Medium
Mixture of Gaussians	O(K)	High
Kernel Density Estimation	O(N)	High
Non-parametric model	O(MN)	High

In this paper, we present a method to detect interesting moving objects using an improved mixture of Gaussians model. The model is used not only to subtract the background, but also to compensate the camera motion and to detect the changing

978-1-4244-1676-9/08 /$25.00 ©2008 IEEE

background and shadows. Since the whole process utilizes the background statistical model, the efficiency is largely improved.

II. THE FRAMEWORK OF THE PROPOSED ALGORITHM

Fig. 1 illustrates the block diagram of the algorithm. The camera motion of the input frame is compensated and background subtraction is performed to extract moving area from the image. The moving area includes interesting motions (motions of the foreground) and uninteresting motions (the moving background). Further steps are needed to eliminate shadows and uninteresting motions from the detection. To perform the algorithm in real-time and with high accuracy, we design every block carefully. Since background model is utilized in every phase, the efficiency of detection is largely improved.

First, the input image is compensated using a camera motion compensation algorithm, which estimates the motion of the camera with a square neighborhood matching method, and adjusts the whole image to make up the motion of the camera. The square neighborhood matching method will be fully described in the following section. It shows how we record the deviation of the image.

Second, a mixture of Gaussians is used to model the background, with a new relatively independent Gaussian added to the model to detect shadows. We choose the mixture of Gaussians as the background model because it deals with the moving background quite well, and is relatively simple in computing. Background subtraction is performed to extract moving area from the input image in this phase.

An area of moving objects is obtained in the second phase, which contains interesting motions and uninteresting motions. The third stage of the algorithm is to eliminate the uninteresting motions, which is important for an accurate detection. Since many algorithms have incorporated in them the method to address most false detections, here we only focus on the false caused by unmodelled moving background, i.e., some moving background which is not incorporated in the background model.

At the end of the algorithm, we have to update the background model. The updating method is almost the same as that for the mixture of Gaussians model, only with some improvement since shadow Gaussian is added. Detailed description of the improvement can be found in the following subsections.

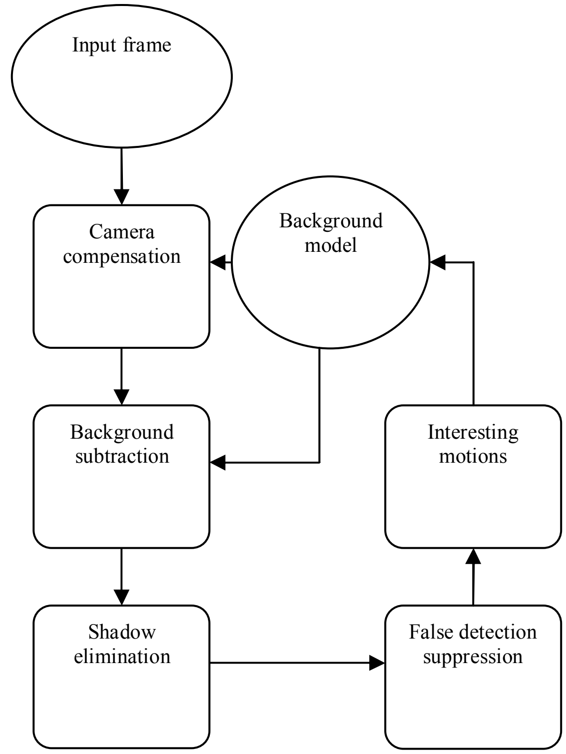

Figure 1. Block diagram of the proposed algorithm

A. *Square neighborhood matching algorithm*

This algorithm is used typically to identify the displacement of a certain pixel. It matches the pixel to its neighborhood points in the same or another frame. By identifying the matching points, we can easily calculate the displacement of the image or the pixels.

First, we define a neighbor area for each pixel, normally a square of N×N pixels for the pixel in the center of the square is taken. For each square, we establish a coordinate with the center pixel in the square as the origin, and define X-component, Y-component, and XYL-component, XYR-component for the line going through the origin and 45 degrees away from the Y-component on the left and on the right separately (XYL-component for the left and XYR-component for the right), as shown in Fig. 2. For the given pixel in the center of the square, all the pixels that are on the lines of the four components in the image are matched, by computing the matching probability. The match here is defined the same as what it is when matching a pixel to its background model. Since no additional model or equations is used to do the matching work, we lower the complexity of calculation by reusing the background model, and thus making the algorithm more real-time and efficient. The locations of the pixels that match the center pixel are recorded for further use. Theoretically, estimating the motion of a pixel, we can obtain the motion displacement of the image. But in practice, we often choose more pixels to calculate to reduce errors. How large the number of the pixels should be chosen to calculate depends on the sensitivity requirement of the detection.

When more than one pixel are used to match their neighbor points, and each pixel has more than one matching pixels, the displacement of the motion can be calculated by computing the number of matching pixels for each location in the corresponding squares. In other words, for each location in the square (e.g. the positive point 2 in the Y-component of the square), we calculate the times when the given pixel matches to the pixel at this location and then choose the one with the most matching times as the displacement of the motion.

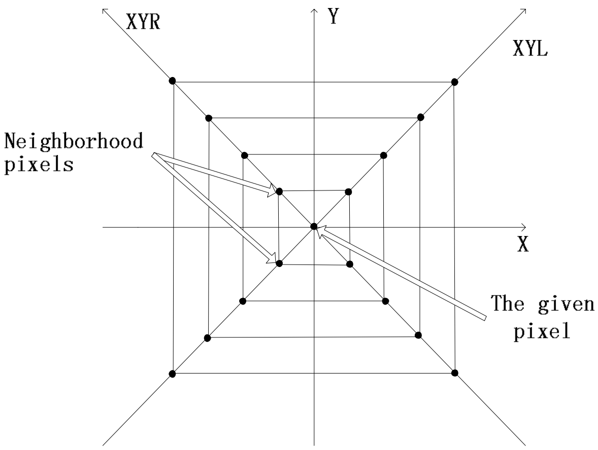

Figure 2. Square neighborhood matching model

Camera motion compensation is very important in that a small motion of the camera may cause errors in the whole image and even corrupt the whole algorithm. To secure the accurate detection of salient motions, camera motion must be compensated before detecting. Here we assume that camera motion is very small and the camera moves in the same direction during two consecutive frames [15]. Under this assumption, we use the square neighborhood algorithm to estimate the camera motion.

To compensate the camera motion, first we divide the images into blocks with the block size a 10×10 pixels square and choose 100 of them for square neighborhood matching (The 100 blocks should scatter in every corner of the image). We count the number of points for each corresponding locations in squares and record the location that has most matching points as the motion displacement, and adjust the image wholly to compensate the displacement at last.

B. *Background subtraction*

Since the mixture of Gaussians model has demonstrated its outstanding performance in many applications in moving backgrounds, we choose it as our basic background model. In addition, some improvements are added to it to suppress shadows. For an input frame, background subtraction is first performed to get a crude motion area.

In a mixture of Gaussians Model, a pixel's intensity is modeled by a mixture of K Gaussian distributions, where K is a small number from 3 to 5, to model variations in the background like tree branch motion and similar small motion in background. The probability that a certain pixel has intensity X_t at time t is estimated as [14]:

$$P(X_t)=\sum_{i=1}^{K}\frac{w_{i,t}}{(2\pi)^{\frac{d}{2}}\left|\Sigma_{i,t}\right|^{\frac{1}{2}}}e^{-\frac{1}{2}(X_t-\mu_t)^T\sum^{-1}(X_t-\mu_t)} \qquad (1)$$

Where $w_{i,t}$ is the weight, $\mu_{i,t}$ is the mean value, $\sum_{i,t}$ is the covariance matrix, all of which are the corresponding values of the X_t Gaussian at time t. The K distributions are ordered based on $w_{i,t}/\sigma_{i,t}^{2}$, and the first B distributions are selected as the background model. Compute the probability of the current pixel value with the equation (1), if $P(X_t)>T$, where T is the threshold, which is usually set to 2.5 standard deviations of the distribution, then pixel X_t is considered as the background pixel, and we say that pixel X_t matches the background model, else as the foreground pixel.

If none of the K distributions match the current pixel value, the least probable distribution is replaced with a distribution with the current value as its mean value, an initially high variance, and low prior weight. Otherwise, update the parameters of the distribution which matches the input pixel with a certain updating algorithm.

As shadows can be easily classified as foreground points, for they significantly differ from background and move along with the objects that cast them, they often create problems in detecting motions. Many measures have been taken to deal with shadows, but most of them are independent of the background model, and require extra calculations. In this paper, we propose a new method to suppress shadows by adding a Gaussian to the background model. As shadows are relatively stable and independent of the objects that cast them, a Gaussian is sufficient for accuracy requirement.

A shadow Gaussian is initialized with the pixel value as its mean value and an initially high variance. It is relatively independent of the K Gaussians representing the background. A current pixel value is matched to the K Gaussians first to decide whether it is a background point or not. If none of the K Gaussians matches the pixel, the shadow Gaussian is used to classify the shadow from the foreground. Matching algorithm for the shadow Gaussian is the same as that of the background Gaussians.

C. *False detection suppression*

In changing backgrounds, two sources of factors may cause false positives to the detection. One is the motion of the camera, which has been addressed in previous section in this paper. The other is the accidental changing background, which occurs frequently but gets little attention from most algorithms. The background model takes into account the moving background and deal with it very well by building different Gaussians representing different states of the background pixel, yet there are some cases that the model doesn't address, like the unmodelled moving background, that is, some small movements in the scene background that are not represented in the background model. This can occur, for example, if a sea wave or a tree branch moves further than it did during model generation [15].

Since the object that occupies a new pixel, which was not part of the background model before, is a part of the background, it is very likely that the object can be matched with a high probability to the pixels within its neighborhood [15], under the assumption that the object has only a small motion between two consecutive frames. So, by matching the pixel to its neighbor pixels, we can easily decide if the pixel is the background point or the foreground.

The match of the pixel to its neighborhood is the same as the current pixel matches to the background, but with a lower-pass threshold, because the probability that the pixel is a part of the background distribution at its original place is expected to be very high. We search the matching pixel in its neighborhood using the method similar to the square neighborhood matching algorithm, which defines a neighborhood area and search the matching pixel on the lines of the four components. If the pixel matches the background pixels with a high probability, then it is classified as the background points, else as the foreground.

D. Background Updating

As to update the model, we choose the selective updating mechanism. Update the K Gaussians and the shadow Gaussian separately. When the current pixel matches to one of the K Gaussians, we update the matching Gaussian using the updating algorithm described above, leaving other Gaussians including the shadow Gaussian unchanged. Otherwise, we match the current pixel to the shadow Gaussian. If the pixel is classified as the shadow point, update the shadow Gaussian by amending its parameters using the algorithm we use to update the matching Gaussian, with the K Gaussians unchanged. It may be that none of the Gaussians matches to the pixel. On this case, we temporarily consider the input pixel as foreground point and replace the least probable distribution of the K Gaussians with a new distribution with the current value as its mean value, an initially high variance, and low prior weight. By this way, we can eliminate shadows from the foreground detection without taking extra measures.

III. EXPERIMENTAL RESULTS

In this section, the effectiveness of the proposed method to detect interesting motion in complex background is demonstrated for a variety of real environments with disturbing factors, like a shaking camera, sea waves, branches, shadows. The video sequences used in this paper are downloaded from the Internet, and the average speed for these sequences is about 54 fps in 1GB Pentium III machines.

Fig. 3 shows the results of experiments with a shaking camera. The two images on the top of the figure are two consecutive frames of the video sequences used in this experiment. We can see that the camera moves in some degree between these two frames. The left bottom image is the experimental result of the original right image using the normal Gaussian mixture model while the right bottom image is the result of the right image using the proposed method. Note that the camera motion is well compensated with our algorithm.

Figure 3. Experimental results of an image from a sequence taken with a shaking camera

Fig. 4 shows the results of experiments with shadows. The image on the top of the Fig. 4 is a frame of the video sequences used in this experiment. The left bottom image is the experimental result of the original right image using the normal Gaussian mixture model without extra measures to deal with shadows, while the right bottom image is the result of the right image using the proposed method. From the results, we can see that our algorithm works very well with shadows. It eliminates shadows from the foreground with high accuracy.

Figure 4. Experimental results of an image with shadows from a sequence

Fig. 5 is the results of experiments with some unmodelled moving background. Here we give two experimental results, one with sea waves and the other with swaying branches to show that the application of this algorithm is very wide. Experiment 1 is the result of the image with sea waves, while the experiment 2 with swaying branches. Similar to the arrangement of Fig. 4, the two experiments' images are arranged like this. On the top of the figure, it is a frame of the video sequences used in this experiment. We can see that the sea waves and the tree branches move in some way that we can't predict. The left bottom image is the experimental result of the original image using the normal Gaussian mixture model, while the right bottom image is the result of the right image using the proposed method. Note that the disturbing factors are well addressed with our algorithm.

Figure 5. Experimental results of images with unmodelled moving background from sequences

Fig. 6 is the experimental results of images with all the disturbing factors from a sequence. This result demonstrates the capacity and the robustness of the proposed algorithm to handle all the distracting factors in the background. The effectiveness and accuracy of the detection indicates that our algorithm is robust, accurate, efficient, and real-time, which are the desirable requirements of a good detection algorithm.

Figure 6. Experimental results of images with all the distracting factors in the background from sequences

IV. CONCLUSIONS

Based on an improved mixture of Gaussians model, and a square neighborhood matching algorithm for motion estimation, we proposed a new algorithm for detecting interesting motion in a moving background, where the shaking camera, the swaying branches, moving sea waves, and shadows are presented. Compared with the other similar methods, the main improvement of the proposed algorithm is that it utilizes the background model in the whole process, including background subtraction, shadow detection, camera motion compensation, unmodelled changing background detection, and the test on the standard data sets also demonstrated that it has a more outstanding performance than the present methods.

REFERENCES

[1] Fu-Yuan Hu, Yan-Ning Zhang, Lan Yao, "An effective detection algorithm for moving object with complex background," Machine Learning and Cybernetics, 2005. Proceedings of 2005 International Conference on Volume 8, 18-21 Aug. 2005 Page(s):5011 - 5015 Vol. 8

[2] Durus M., Ercil A., "Robust Vehicle Detection Algorithm," Signal Processing and Communications Applications, 2007. SIU 2007. IEEE 15th 11-13 June 2007 Page(s):1 – 4

[3] Yeon-sung Choi, Piao Zaijun, Sun-woo Kim, Tae-hun Kim, Chun-bae Park, "Motion Information Detection Technique Using Weighted Subtraction Image and Motion Vector," Hybrid Information Technology, 2006. ICHIT '06. Vol1. International Conference on Volume 1, Nov. 2006 Page(s):263 – 269

[4] Ho M.A.T., Yamada Y., Umetani Y., "An HMM-based temporal difference learning with model-updating capability for visual tracking of human communicational behaviors," Automatic Face and Gesture Recognition, 2002. Proceedings. Fifth IEEE International Conference on 20-21 May 2002 Page(s):163 – 168

[5] N. Paragios and R. Deriche, "Geodesic active contour and level sets for the detection and tracking of moving objects,"IEEE Trans. Pattern An alysis and Machine Intelligence, vol. 22, no. 3, pp. 266–280, March 2000.

[6] A. R. Mansouri and J. Konrad, "Motion detection and estimation-multiple motion segmentation with level sets," IEEE Trans. on Image Processing, vol. 12, no. 2, pp. 201–220, 2003.

[7] Roth S.,Black M.J., "On the spatial statistics of optical flow," Computer Vision, 2005. ICCV 2005. Tenth IEEE International Conference on Volume 1, 17-21 Oct. 2005 Page(s):42 - 49 Vol. 1

[8] Stocker A.A., "An improved 2D optical flow sensor for motion segmentation," Circuits and Systems, 2002. ISCAS 2002. IEEE International Symposium on Volume 2, 26-29 May 2002 Page(s):II-332 - II-335 vol.2

[9] Pathirana P.N., Lim A.E.K., Carminati J., Premaratne M., "Simultaneous estimation of optical flow and object state, A modified approach to optical flow calculation," Networking, Sensing and Control, 2007 IEEE International Conference on 15-17 April 2007 Page(s):634 - 638

[10] Piccardi M.,"Background subtraction techniques a review," Systems, Man and Cybernetics, 2004 IEEE International Conference onVolume 4, 10-13 Oct. 2004 Page(s):3099 - 3104 vol.4

[11] Thongkamwitoon T., Aramvith S., Chalidabhongse T.H., "An adaptive real-time background subtraction and moving shadows detection," Multimedia and Expo, 2004. ICME '04. 2004 IEEE International Conference on Volume 2, 30-30 June 2004 Page(s):1459 - 1462 Vol.2

[12] Monnet A., Mittal A., Paragios N., Visvanathan Ramesh, "Background modeling and subtraction of dynamic scenes," Computer Vision, 2003. Proceedings. Ninth IEEE International Conference on 13-16 Oct. 2003 Page(s):1305 - 1312 vol.2

[13] Paolo Spagnolo, Tiziana D'Orazio, Marco Leo, and Arcangelo Distante, "Advances in Background Updating and Shadow Removing for Motion Detection Algorithms," Computer Analysis of Images and Patterns, Volume 3691/2005 Page(s): 398-406

[14] N. Friedman and S. Russell, "Image Segmentation in Video Sequences: A Probabilistic Approach," Proc. Conf. Uncertainty in Artificial Intelligence, pp. 175-181, 1997. C

[15] Ahmed Elgammal, David Harwood, Larry Davis, "Non-parametric Model for Background Subtraction" 6th European Conference on Computer Vision. Dublin, Ireland, June/July 2000

[16] Stauffer C., Grimson, W.E.L., "Adaptive background mixture models for real-time tracking," Computer Vision and Pattern Recognition, 1999. IEEE Computer Society Conference on. Volume 2, 23-25 June 1999 Page(s):