Computing and Informatics, Vol. 37, 2018, 865–893, doi: 10.4149/cai_2018_4_865

BACKGROUND SUBTRACTION BASED ON PERCEPTION-CONTAINED PIECEWISE MEMORIZING FRAMEWORK

Songbo Liu

School of Computer Science, Harbin Institute of Technology
No. 92 Dazhi Street, Harbin, China
e-mail: sbliu@hit.edu.cn

Xudong Zhao

College of Information and Computer Engineering
Northeast Forestry University
No. 26 Hexing Road, Harbin, China

Xianglong Tang

School of Computer Science, Harbin Institute of Technology
No. 92 Dazhi Street, Harbin, China

Abstract. A key issue for full-time video surveillance is to search or establish a reference image of background which corresponds to current video frame. However, background that was ever in presence long time ago is enclosed or discarded due to background forgetting assumption. How to rapidly pick up or even rebuild long-term background needs to be discussed. This paper aims to present a framework for background maintenance in order to solve the problem. A piecewise memorizing framework is proposed for matching, updating and even rebuilding long-term background. Based on the metaphors of psychological selective attention theory, the framework is composed of a prior piecewise perception processor for intensity stationary test. Besides, a hierarchical memorizing mechanism constitutes the other part of the framework for overcoming the exponential forgetting of long period background appearances. Applied to Gaussian mixture model (GMM), this framework is

capable of maintaining short-term background states, identifying long period background appearances, and rapidly adjusting to new background states according to different expressions derived from the prior perception of scene intensity changes. Its effectiveness can be demonstrated by experimental results for solving various typical problems.

Keywords: Long-term background memory, piecewise stationary test, Gaussian mixture model, background subtraction, foreground detection

Mathematics Subject Classification 2010: 68-T45

1 INTRODUCTION

Background subtraction, which compares each current frame with a dynamic reference image of background derived from previous frames, forms a favorite and powerful anterior mechanism for applications such as foreground detection [1], face recognition [2], object tracking [3], etc. Prevailing methods mainly exist in three aspects. One focuses on statistical features, for instance, Gaussian mixture model (GMM) [4], a kernel density estimation [5], Bayesian modeling [6], auto-regressive model [7], etc. Other integrates structural information with local binary pattern (LBP) [8], self-organizing network [9], codebook model [10], co-occurrence matrix [11], etc. The last concentrates on the statistical and structural hybrid accompanied with a three-stage model [12], spatio-temporal information saliency [13], a statistical multi-point-pair model [14], an efficient hierarchical method [15], etc.

Despite its importance, background subtraction is far from a complete solution to the disturbances of complex environments. The problems listed mainly in [12] can be re-categorized as follows: background varying illumination including *time of day*, *light switch* and *shadows*; background periodic motion (e.g. *waving trees*); background consistency containing *camouflage*, *bootstrapping* and *foreground aperture*; and semantic feedback (e.g. *moved objects*, *sleeping person* and *waking person*, namely *ghosts*) which is not able to be handled using the background subtraction module only, as indicated in [12]. Except for background consistency, these problems seem to be probably solved once we memorize a series of full-time background appearances with one of which the current frame is matched. However, it is hard for memorizing background that was ever in presence long time ago (namely, long-term or long period background) considering the restriction of memory capacity.

On account of the storage limitation, time constraint is always taken into consideration. A background forgetting assumption is made that states often appearing in high frequency are treated as background. A forgetting factor is employed to control the contribution of earlier observations. Thus, only background appearances corresponding to recent video frames are memorized. Inevitably, states representing long period background are discarded. The strategy works as to the presence of the sta-

tionary states such as *time of day*; while, it lapses when non-stationary states such as *light switch* appear. In order to solve this problem, a local principal component analysis transformation accompanied with an adaptive learning (ALPCA) [16] is proposed to learn separate feature eigensubspaces representing different background appearances. ALPCA is effective especially when *light switch* appears, for it uses spatial information to guide switches among short-term background appearances representing different illumination conditions. Meanwhile, parameters corresponding to the forgetting factor are rapidly updated. However, it will inevitably lapse into frequent switches among recent background appearances with different illumination conditions due to absence of long-term background memory.

Instead of memorizing long-term background, prevailing methods mainly focus on concrete problems mentioned above. Narrowing down to GMM as an effective background model, there have been various improvements contributed to address these problems. A texture-contained GMM (TGMM) [17] and a spatial hierarchical GMM (HGMM) [18] are utilized for solving the problem of *light switch*. *Shadows* are eliminated using a Gaussian mixture shadow model (GMSM) [19] and TGMM. As to the problem of *ghosts*, a memorizing GMM (MGMM) [20] is applied to tackling the exponential decay [21] of GMM. Although these algorithms are competent to handle certain problems, there is a lack of a common way to simultaneously solve all these problems because of neglecting long period background memory. Even if we can break through the restriction of memory capacity, it is still difficult for us to seek the matched states of background associated with current video frame in real time due to the large amount of full-time background storage.

In order to tackle the problem about long-term background, a piecewise memorizing framework is proposed. Three major contributions can be claimed. Taking the metaphors of psychological selective attention theory into consideration, hypotheses of background subtraction indicating what to perceive and memorize are firstly proposed. Second, a prior perception-concerned recognition of non-stationary intensity change is presented based on a segmented stationarity test. Third, a memorizing hierarchy corresponding to GMM is put forward for the storage of long period background and the adaptation to rapid intensity change. Particularly, the memorizing framework accompanied with a prior perception processor based on segmented stationarity test is an adaptive multi-scale hierarchical system, which is constituted by a conventional mixture of short-term background models, spatial states memorizing a long-term background, and a sequence of global differences adapting to fast intensity change of background. Unlike these referenced methods including our previous research [22], a prior piecewise perception processor for intensity stationary test is proposed to search or establish long-term background rapidly that corresponds to current video frame. Regarding GMM as the basic background memorizing model in this paper, this framework is named as P-MGMM (piecewise memorizing GMM). The rest of the paper is organized as follows: in Section 2, we revisit GMM, point out the limitations of related methods, discuss the corresponding metaphors of cognitive psychology, and make related assumptions; in Section 3, we present a prior segmented stationarity test, which is viewed as a pre-perception of background vari-

ation; in Section 4, a memorizing hierarchy applied to GMM is proposed; the experimental results are discussed in Section 5; and conclusion is made in Section 6.

2 METAPHORS AND HYPOTHESES

In this section, we will first review GMM [4] and point out the limitations of GMM-based models. Afterwards, an analogy will be made between the attention theory of cognitive psychology and the corresponding background subtraction methods including both GMM-based and non-GMM approaches. Then, related hypotheses will be made.

2.1 GMM Revisited and the Limitations of Related Models

In pixel-wise GMM, the current intensity belongs to an existing Gaussian that represents either background or foreground. The probability of the current intensity X_t is expressed as $P(X_t) = \sum_{k=1}^{K} \omega_{k,t} \cdot \eta(X_t, \mu_{k,t}, \Sigma_{k,t})$, where η is a Gaussian probability density function. K represents the number of Gaussians. $\mu_{k,t}$, $\Sigma_{k,t}$ and $\omega_{k,t}$ refer to the current mean, the covariance matrix and the weight estimation of Gaussian k, respectively. Furthermore, the covariance matrix $\Sigma_{k,t}$ is $\Sigma_{k,t} = \sigma_{k,t} \cdot I$, assuming the red, green, and blue pixel values are independent[1]. A new pixel intensity X_{t+1} is checked against each Gaussian of the pixel-wise GMM in order to find the appropriate one to which it belongs. First, the parameters of the pixel-wise GMM are updated as follows:

$$\begin{aligned} \omega_{k,t+1} &= (1-\alpha) \cdot \omega_{k,t} + \alpha \cdot M_{k,t+1}, \\ \mu_{k,t+1} &= \mu_{k,t} + \rho \cdot (X_t - \mu_{k,t}) \cdot M_{k,t+1}, \\ \sigma^2_{k,t+1} &= \sigma^2_{k,t} + \rho \cdot \left((X_t - \mu_{k,t})^T (X_t - \mu_{k,t}) - \sigma^2_{k,t} \right) \cdot M_{k,t+1} \end{aligned} \tag{1}$$

where α and ρ denote the learning rates. Let ρ be $\rho = \alpha \cdot \eta(X_t | \mu_{k,t}, \sigma_{k,t})$. If the k^{th} Gaussian is matched, $M_{k,t+1} = 1$; otherwise, $M_{k,t+1} = 0$. Then Gaussians are re-ordered by $\omega_{k,t}/\sigma_{k,t}$. If intensity X_{t+1} is matched with none of the Gaussian models, the mean $\mu_{K,t}$, the standard deviation $\sigma_{K,t}$ and the weight estimation $\omega_{K,t}$ of the last Gaussian are to be replaced by the current intensity X_{t+1}, an initially high variance and a low prior weight, respectively. Ultimately, the weight estimation $\omega_{k,t+1}$ is re-normalized. The first b Gaussians are chosen as the background model. That is

$$B_{t+1} = \arg\min_{b} \left(\sum_{k=1}^{b} \omega_{k,t+1} > T \right) \tag{2}$$

where B_{t+1} denotes the current Gaussians representing background. T is a user-defined threshold.

[1] While this is certainly not the case, the assumption allows us to avoid a costly matrix inversion at the expense of some accuracy.

Pixel-wise GMM is especially suitable for the problem of *time of day* and *waving trees*. Yet, it also has shortcomings which is mainly presented as three aspects. Firstly, the most frequently matched state k is supposed to hold a much lower standard deviation $\sigma_{k,t}$ and a firmly high weight $\omega_{k,t}$ as expressed in Equation (1), particularly when the background appearance remains unchanged. That will make the state keep to the first place and hold the priority for matching with the new pixel intensity. Secondly, the weight $\omega_{k,t}$ of matched Gaussian is updated steadily with a fixed learning rate α. Thirdly, Gaussians are chosen to represent background according to Equation (2), which accumulates $\omega_{k,t+1}$ and is regarded as a conservative way to keep background especially when the problem such as *waving trees* occurs. Therefore, pixel-wise GMM keeps its own rhythm to fulfill parameter updating, no matter how rapidly the scene intensity varies. When background changes suddenly, the current state will be wrongly considered to be foreground (e.g. *light switch* and *moved objects*) without any doubt. Actually, these shortcomings can be rightly overcome using a fast learning strategy when immediate background change occurs, once the most frequently matched states can be memorized as a long-term background.

GMSM [19], which tackles the problem of *shadows*, develops pixel-wise GMM by supposing shadows to be associated with frequently seen states labeled as foreground. It improves the parameter updating strategy and incorporates frequently matched shadows into background. In addition, a second pixel-wise GMM is built for maintenance of shadows. Using GMSM, shadows are distinguished from foreground. However, GMSM keeps only short-term background even though it improves the parameter updating strategy of pixel-wise GMM. Moreover, it does not concern any spatial information. When global change of background (e.g. *light switch*) occurs, states corresponding to a former background before the time of change will be pushed to perform a low ranking Gaussian and ultimately discarded.

Suppose that background sudden change keeps spatial consistency. Local texture information or spatial statistic is combined with pixel-wise GMM, and that forms TGMM [17] or HGMM [18]. Some non-GMMs (e.g. LBP [8] and ALPCA [16]) are also this type. Due to consideration of spatial information, these models are effective to *light switch* especially when rapid parameter learning is performed. Yet, they are invalidated once faced with the problems such as *moved objects*, *sleeping person* and *waking person* (namely, *ghosts*), especially when foreground scale is comparative to the size of scene. Except for the storage of short-term background, it is needed to be further discussed how to memorize long-term background.

To the best of our knowledge, MGMM [20] is one of the first GMM-based models that considers memorizing long period background. Despite of its competence for local background sudden change, it still has some problems to be discussed. First, it only focuses on pixel-wise states but not spatial background simultaneously, and that makes a lapse when *light switch* occurs. Second, it needs an initial period for model learning, which excludes *bootstrap* in consideration. Besides, it might be incapable of tackling *moved objects* and *ghosts*, when uncovered background never appears during the learning period. Third, it makes an alternate learning rate for matched

states to rapidly adapt to sudden change, which also may absorb *sleeping person* into background. Fourth, it stores the longest memorized short-term background states, which currently keeps a low rank order, into its long-term background memory. That will lead to frequent switches of background states between short-term and long-term memory, especially when background keeps changing repeatedly.

In order to make a robust background subtraction model, improvements have been made to memorize long period background in our previous work [22]. Base on GMM, a framework that contains components representing not only short-term but also long-term background memory is presented. This paper makes a further effort to expound how this framework works and why it is composed of real-time state record, spatial background memory and global difference memory, in order to solve most of the problems except *camouflage* and *foreground aperture*. Besides, a prior perception processor is precisely proposed to replace the simple stationary test in [22]. Before that, statements on the origin of the piecewise memorizing framework are to be made, which is derived from the metaphors of cognitive psychology.

2.2 The Metaphors of Cognitive Psychology

In cognitive psychology, the term selective attention [23] refers to the fact that we usually focus our attention on one or a few tasks or events rather than on many. In other words, the information that people process is divided into the attended and unattended message. Selective attention theory is mainly classified into three categories: filter theory, attenuation theory and late-selection theory, as shown in Figure 1.

We firstly discuss these theories and make several extracts as follows:

- Filter theory predicts that all unattended messages will be filtered out, that is to say, not processed for recognition or meaning. A filter is set to make a selection of what message to process early in the processing, typically before the meaning of the message is identified. Meanwhile, a lower working-memory capacity refers to a decreased ability to actively block the unattended message.
- As for attenuation theory, some meaningful information in unattended messages may still be available, even if it is difficult to recover. Furthermore, incoming messages are subjected to an alternative hierarchical analysis. Moreover, only a few firmed meaningful messages, together with those in a special context, have permanently lowered thresholds.
- Late-selection theory holds that all messages are routinely processed but which message to respond to is selected "late" in processing. In fact, the selection is made in the response to "stimuli", rather than due to the recognition of it. In other words, the selective filter performs after the perception model.

Using metaphor, computers can be also simulated to have "attention". In this research, prevailing background subtraction methods can be described as the result of selective attention. Applying this analogy to GMM, the "matching" step can be

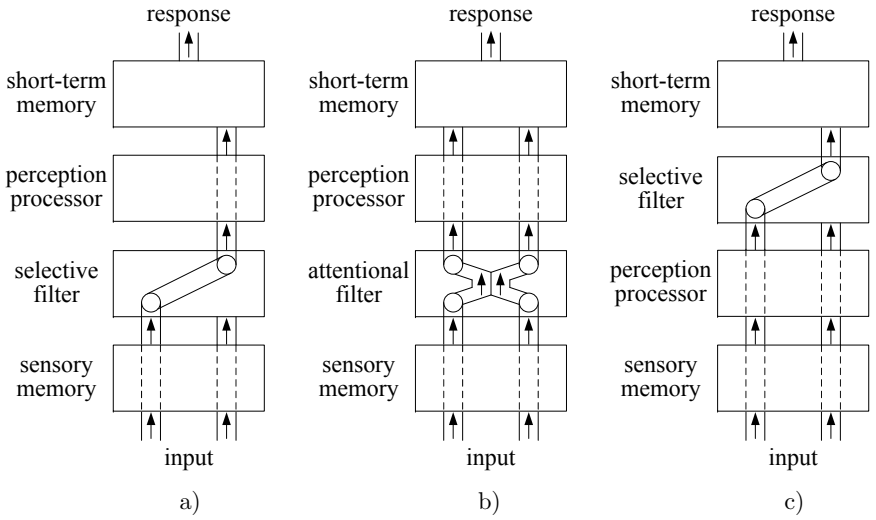


Figure 1. Illustration of selective attention theory [23]. a) Filter theory, b) attenuation theory, c) late-selection theory.

regarded as the act of paying "attention". In other words, the matched Gaussian refers to the attended message; whereas unmatched states correspond to unattended messages. Moreover, the working-memory capacity is equivalent to the number of Gaussians. Thus, prevailing background subtraction methods can be described as the result of selective attention. Correspondingly, background modeling approaches (not solely GMM-based methods) can be classified into following three categories:

Match-concerned model pays attention to matched states. Only frequently matched states are regarded as attended messages and viewed as background. A limited number of states lead to a frequent replacement of background appearances. Just like setting a selective filter before a perception processor as shown in Figure 1 a), the always un-matching state equivalent to the unattended message are discarded. In match-concerned model, states may correspond to Gaussians derived from temporal statistics (for example, GMM [4] and TGMM [17]) or histograms generated from structural information, such as LBP [8].

Meaning-concerned model concentrates on the potential background. Just like setting an attentional filter before a perception processor as illustrated in Figure 1 b), matched states together with meaningful unmatched states are memorized. In fact, the attenuation of unmatched states depends on the meaningful information they contain. Commonly, the meaning of the unmatched states is derived from a spatial consistency (e.g. HGMM [18]), from a permanently lowered threshold that enables frequently seen states labeled as foreground to be background (e.g. GMSM [19]), or even from a tag labeling a background state

that was ever in presence (e.g. MGMM [20]). In meaning-concerned model, each state may represent a pixel-wise Gaussian [19, 20], a hierarchical Gaussian [18] or an eigensubspace (e.g., ALPCA [16]).

Perception-concerned model focuses more on the identification of complex environments rather than on state discrimination. That is to say, the perception processor module is laid before the selective filter, as illustrated in Figure 1 c). In fact, even an early stage of perception processing may contribute to good state discrimination results, i.e., states representing the appearances of a long period background can be purposefully memorized. A representative perception-concerned model is P-MGMM, about which details are to be discussed in Section 3 and Section 4.

2.3 Hypotheses of Background Subtraction

According to the metaphors mentioned above, robust background models should firstly possess memorizing ability. The fact that GMM uses exponential forgetting is a case in point. However, exponential forgetting contradicts the purpose of memorizing long period background due to the constraint of memory capacity. To overcome this, a primary stage of perception should be made prior to the attention selection for memorizing. Before that, a long-term image sequence should be divided into short ones and then processed for better perception and memorization of long-term background, which is named as a "piecewise" action. Now what to perceive and memorize becomes a central issue. In order to answer these questions, we propose the following three hypotheses:

- Background modeling essentially derives from a segmented long-term sequence. According to the exponential forgetting of background appearances, long period background must be memorized.
- A prior stage of perception is needed, in accordance with perception-concerned model. That is, special attention to scene general change existing only in short-term image sequences should always be paid in advance.
- Spatial information should be included, considering scene change always meets a certain spatial consistency. The analogous status is also found in the data-driven attention model [23] in cognitive psychology.

Taking the memory capacity into account, we present a segmented stationarity test as the prior perception processor of late-selection theory [23] in order to identify different scene changes generally, and propose a memorizing mechanism applied to GMM to be the corresponding selective filter.

3 PRIOR PERCEPTION PROCESSOR

It is always necessary to detect in advance and rapidly adapt to scene change. Different intensity changes of scene exist only in short period of time. Thus, stationarity

test is regarded as a perceptional criterion when we consider pixel-wise intensity values in short-term image sequences to be time series. Inversion number test is viewed as a common stationarity test approach [24]. Hence, we constitute the inversion number of pixel-wise temporal mean values to test the stationarity of segmented time series $\{X_t\}$ and subsequently identify different intensity changes.

3.1 Inversion Number Test

We firstly cut $\{X_t\}(t = 1, 2, \ldots, N)$ into l segmented sequences, each is at a length of M ($N = lM$). The i^{th} segmented sequence is denoted by $\{X_{i,j}\}$ ($i = 1, 2, \ldots, l$; $j = 1, 2, \ldots, M$), where $X_{i,j} = X_{(i-1)M+j}$. The mean value of each segmented sequence is expressed as follows:

$$\mu_i = \frac{1}{M} \sum_{j=1}^{M} X_{i,j}. \tag{3}$$

Meanwhile, we select r pieces of $\{X_{i,j}\}$ to constitute a sub-sequence $\{X_{i,h_j}\}$, where $X_{i,h_j} = X_{i-r+h,j} = X_{(i-r-1+h)M+j}$ ($i = 1, 2, \ldots, l$; $h = 1, 2, \ldots, r$; $j = 1, 2, \ldots, M$). Each two neighboring sub-sequences, for instance $\{X_{i,h_j}\}$ and $\{X_{i+1,h_j}\}$, share a $(r-1)-r^{\text{th}}$ overlap. Let X_{i,h_j} ($i = -r+1, -r+2, \ldots, 0$) be $X_{i,h_j} = X_{1,h_j}$ for convenience. Correspondingly, we get a sub-sequence $\{\mu_{i,h}\}$ from the mean sequence $\{\mu_i\}$, where $\mu_{i,h} = \mu_{i-r+h}$ ($i = 1, 2, \ldots, l$; $h = 1, 2, \ldots, r$).

As to each segmented sequence in a sub-sequence $\{X_{i,h_j}\}$, a reverse order of $\mu_{i,h}$ ($h = 1, 2, \ldots, r-1$) is defined, when there is $\mu_{i,j} < \mu_{i,h}$ ($j = h+1, h+2, \ldots, r$). An inversion number $A_{\mu_{i,h}}$ of $\mu_{i,h}$ is obtained after traversing all the $\mu_{i,j}$ from $\mu_{i,h}$ to $\mu_{i,r}$, i.e., $A_{\mu_{i,h}}$ records the counts when $\mu_{i,h} > \mu_{i,j}$ ($h < j$). The sum of $A_{\mu_{i,h}}$ in $\{\mu_{i,h}\}$ is defined as follows:

$$A_{\mu_i} = \sum_{h=1}^{r-1} A_{\mu_{i,h}}. \tag{4}$$

We define a random variable $I_{h,j}$ that represents whether or not the ordered pair $\langle \mu_h, \mu_j \rangle$ in $\{\mu_{i,h}\}$ is reverse order. That is

$$I_{h,j} = \begin{cases} 1, & \text{if } \mu_{i,h} > \mu_{i,j}, \\ 0, & \text{others}, \end{cases} \qquad \forall h < j. \tag{5}$$

Thus, the inversion number is expressed as

$$A_{\mu_{i,h}} = \sum_{j=h+1}^{r} I_{h,j}. \tag{6}$$

Assuming the sub-sequence $\{X_{i,h_j}\}$ to be stationary, the random variable $I_{h,j}$ is equiprobable. That is,

$$P(\mu_{i,j} > \mu_{i,h}) = P(\mu_{i,j} \leq \mu_{i,h}) = \frac{1}{2} \tag{7}$$

where $h = 1, 2, \ldots, r$ and $j = h, h+1, \ldots, r$. From Equations (5) and (7) we obtain

$$E(I_{h,j}) = \frac{1}{2} \times 1 + \frac{1}{2} \times 0 = \frac{1}{2}. \tag{8}$$

Then we evaluate the expectation and the variance of the inversion number sum A_{μ_i} in $\{\mu_i\}$. The expectation $E(A_{\mu_i})$ is derived from Equations (4) and (6):

$$E(A_{\mu_i}) = E\left(\sum_{h=1}^{r-1} A_{\mu_{i,h}}\right) = \sum_{h=1}^{r-1} \sum_{j=h+1}^{r} E(I_{h,j}). \tag{9}$$

When we substitute Equation (8) in Equation (9), the expectation $E(A_{\mu_i})$ is expressed as follows:

$$E(A_{\mu_i}) = \frac{1}{2} \sum_{h=1}^{r-1} (r-h) = \frac{r(r-1)}{4}. \tag{10}$$

On account of the independence of $A_{\mu_{i,h}}$ and $A_{\mu_{i,j}}$ $\forall h, j \in (1, r)$ where $h \neq j$, we calculate the expression of the variance $Var(A_{\mu_i})$ from Equation (4). That is

$$Var(A_{\mu_i}) = \sum_{h=1}^{r-1} \left(E\left(A_{\mu_{i,h}}\right)^2 - E^2\left(A_{\mu_{i,h}}\right) \right). \tag{11}$$

Substitute the $A_{\mu_{i,h}}$ by Equation (6), and we have

$$Var\left(A_{\mu_i}\right) = \sum_{h=1}^{r-1} \left(\sum_{j=h+1}^{r} E(I_{h,j})^2 + 2C_{r-h}^2 E\left(I_{h,j} I_{h,k}\right)_{j \neq k} - \left(\sum_{j=h+1}^{r} E\left(I_{h,j}\right) \right)^2 \right). \tag{12}$$

Once Equation (8) is introduced in, $Var(A_{\mu_i})$ in Equation (12) can be simplified as follows:

$$Var(A_{\mu_i}) = \frac{r(r-1)(2r-1)}{24}. \tag{13}$$

3.2 Identification of Intensity Changes

The sum of the inversion number of each sub-sequence $\{\mu_{i,h}\}$, which is derived from Equation (4), forms a corresponding sequence as $\{A_{\mu_i}\}$ $(i = 1, 2, \ldots, l)$. Due to the histogram count of A_{μ_i} to all the pixels in a selected area, it is reasonable to assume that A_{μ_i} is a normal stochastic variable, which is expressed as follows:

$$u_{\mu_i} = \frac{|A_{\mu_i} + 0.5 - E(A_{\mu_i})|}{\sqrt{Var(A_{\mu_i})}}. \tag{14}$$

When $|u_{\mu_i}| \leq 1$, we believe that there is no significant difference between $X_{i,j}$ and $X_{i,k}$ $\forall j \neq k$. That is, $\{X_{i,j}\}$ is a stationary sequence. Due to the spatial hypothesis proposed in Section 2.3, we identify different intensity changes in each sub-sequence $\{X_{i,h_j}\}$ by calculating u_{μ_i} $(i = 1, 2, \ldots, l)$ from Equation (14). That is

$$\left| \max_{y \in \Gamma} (u_{\mu_i}(y)) \right| \leq 1 \tag{15}$$

where Γ represents the selected area. $\max_{y\in\Gamma}(u_{\mu_i}(y))$ denotes the max value of u_{μ_i} in Γ. If Inequation (15) holds, the last segmented sequence $\{X_{i,j}\}$ of the current subsequence $\{X_{i,h_j}\}$ contains gradual intensity change; otherwise, there is fast intensity change.

4 HIERARCHICAL MEMORIZING MECHANISM

Hypotheses proposed in Section 2.3 have answered what to perceive and memorize in perception-concerned background model. In order to tackle different scenarios of background such as various illumination conditions, long period background must be memorized. Considering the fast detection and rapid adaptation to scene change, image sequences or video frames are to be segmented into sub-sequences for prior perception of pixel-wise intensity change. It has been solved how to perceive the rapid intensity change that occurs in segmented image sequences in Section 3, let us now consider how to memorize long period background appearances and how to track the rapid intensity change of background.

4.1 P-MGMM Framework

Due to the hypothesis of spatial consistency proposed in Section 2.3, long period background memory must contain models of different scales. Supposing that a pixel-wise mode is regarded as a state, a spatial combination of background state models is identified to be a version, which has been specified in HGMM [18]. Once we perceive current sub-sequences to be non-stationary, parameters of matched pixel-wise states need to be rapidly updated. Therefore, we need to adjust learning rates expressed in Equation (1) as high as possible, which inevitably accelerates background forgetting. This traditional exponential forgetting way makes real-time state record. In order to avoid discarding long period background, additional states at each pixel together with the corresponding versions are utilized to memorize different scales of long-term background. Put another way, spatial background memory containing additional states and versions which represent long period background is needed. Once rapid intensity changes are detected, it is apparently not enough to match current frame with states or versions in either real-time state record or spatial background memory only, especially when new illumination condition appears in scene. Therefore, global difference memory is presented, which can be viewed as the first order difference of spatial background memory.

Narrowing down to GMM, modules of real-time state record, spatial background memory and global difference memory constitute a piecewise memorizing hierarchy, which together with prior perception processor forms P-MGMM framework. As illustrated in Figure 2 a), states, versions and differences are expressed as S, V and D, respectively. When labeled with '*', they represent background appearances. Real-time state record corresponds to a conventional mixture of states. Spatial background memory for long period background storage consists of states and versions.

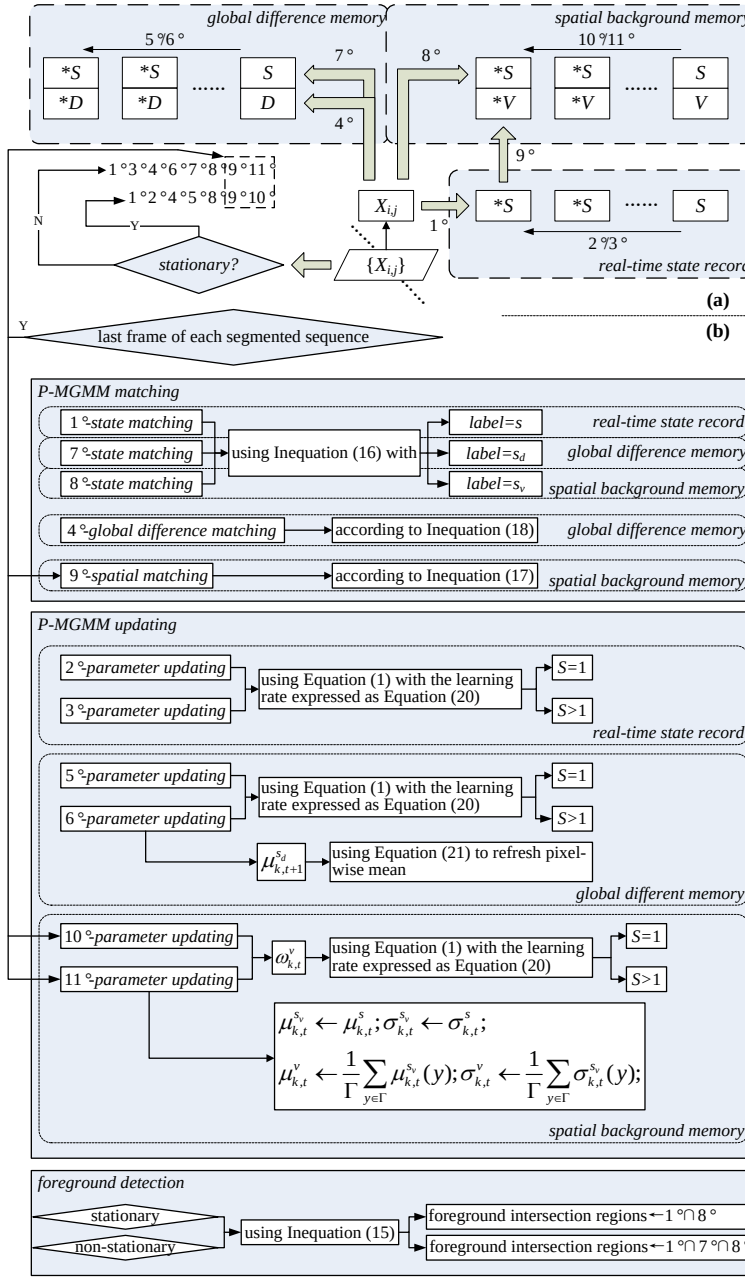


Figure 2. Sketch map of perception-contained piecewise memorizing framework, i.e. P-MGMM for short. a) P-MGMM framework, b) P-MGMM maintenance algorithm from step 1° to step 10° or 11°.

Together with spatial differences between current intensity values and the means of k states in real-time state record, additional pixel-wise states constitute global difference memory, which is competent to adapt to rapid variations of background. According to P-MGMM, several considerable improvements are able to be made. Concrete improvements are to be introduced following the maintenance steps of P-MGMM framework, as illustrated in Figure 2 b).

4.2 P-MGMM Maintenance

First of all, P-MGMM matching is considered. Different hierarchical matches are defined as pixel intensity values within D times the standard deviations of a distribution, respectively. Then we have:

$$\left|X_t^s(y) - \mu_{k,t}^{label}(y)\right| \leq D \cdot \sigma_{k,t}^{label}(y), \quad label \in \{s, s_v, s_d\}, \tag{16}$$

$$\left|\frac{1}{\Gamma}\sum_{y\in\Gamma}\mu_{1,t}^{s}(y) - \mu_{k,t}^{v}\right| \leq D \cdot \sigma_{k,t}^{v}, \tag{17}$$

$$\left|\frac{1}{\Gamma}\sum_{y\in\Gamma}(X_t^s(y) - \mu_{k,t}^{s}(y)) - \mu_{k,t}^{d}\right| \leq D \cdot \sigma_{k,t}^{d} \tag{18}$$

where $D = 2.5$. Superscripts s, s_v and s_d correspond to state parameters in real-time state record, spatial background memory and global difference memory, respectively. Accordingly, superscripts v and d represent version parameters in spatial background memory and difference parameters in global difference memory. Inequation (16) expresses three state matchings at different parts of the framework. $X_t^s(y)$ represents the current intensity value at location y. $\mu_{k,t}^{label}(y)$ and $\sigma_{k,t}^{label}(y)$ denote the mean and deviation of the k^{th} state in real-time state record, spatial background memory and global difference memory with *label* varying from s to s_v and s_d. As to spatial matching expressed in Inequation (17), spatial mean of the first state in real-time state record is calculated and compared with k means of versions in spatial background memory, when reaching the last frame of each segmented sequence. Besides, a global difference matching is made by comparing spatial differences of current intensity values and the means of k states in real-time state record with k means of differences in global difference memory, as expressed in Inequation (18). Note that state matchings at different parts of the framework (i.e., step 1°, 7° and 8°) expressed in Inequation (16) are applied to foreground detection. Version and difference matching expressed in Inequation (17) and (18) are only used for parameter updating.

We consider real-time state record as a short-term background memory. Like classical GMM [4] which is a match-concerned model, Gaussians in this part only serve as a real-time record of matched states. Put another way, there are only pixel-level state models that exist in real-time state record. Inevitably, the always matched state k is supposed to hold a much lower standard deviation $\sigma_{k,t}$, particularly when

the background appearance remains unchanged. Therefore, a fixed lower bound of standard deviation σ_{low} for each state is used. Besides, we re-ordered Gaussians only by $\omega_{k,t}$, while taking the over-learning of the above-mentioned standard deviation into consideration. Then each short-term state in sorted Gaussians is regarded as background, if its weight value is large enough. That is:

$$B_{t+1} = \{\omega_{k,t+1} > T \mid k = 1 \dots K\} \tag{19}$$

where B_{t+1} and T keep the same meaning, as has been explained in Equation (2). Last but foremost, we define the learning rate as follows:

$$\alpha_t = \alpha \cdot S \tag{20}$$

where the parameter $S > 1$ when there is rapid intensity change, and $S = 1$ in the other case (i.e., step 2° or 3°). Parameters of real-time state record are updated using Equation (1), once state matching in real-time state record corresponding to step 1° is performed.

After parameters in real-time state record are refreshed, global difference matching regarded as step 4° is made using Inequation (18). Once variations in scene are perceived considering current segmented image sequence to be non-stationary, sequential background appearances derived from continuous intensity changes emerge. States and versions in either real-time state record or spatial background memory are incompetent at accommodating new background appearances in time. Therefore, global difference memory is devoted to adapting to the continuity of background intensities. Global difference memory consists of GMMs representing global differences and pixel-wise states, respectively. Following the parameter updating strategy in real-time state record, the current mean $\mu^d_{k,t}$, the standard deviation $\sigma^d_{k,t}$ and the weight estimation $\omega^d_{k,t}$ of the matched global difference are updated using Equation (1) with the learning rate α_t switching between different intensity changes as expressed in Equation (20), i.e. step 5° or 6°. According to the rightly matched global difference, corresponding states are selected for parameter updating. In fact, only states referring to the inversion of foreground detection results derived from both real-time state record and spatial background memory are needed to be refreshed. The standard deviation $\sigma^{s_d}_{k,t}$ of selected state k is updated using Equation (1). As to pixel-wise mean $\mu^{s_d}_{k,t}$, it is refreshed using Equation (1) when the intensity values appear stationary in segmented sequences. Once there is obvious intensity change in scene, the replacement of $\mu^{s_d}_{k,t}$ is expressed as follows:

$$\mu^{s_d}_{k,t+1} = \mu^{s_d}_{k,t} + \mu^d_{k,t}. \tag{21}$$

If no global match exists, the last global difference and pixel-wise state need to be re-initialized.

Parameter updating in real-time state record and global difference memory reinforce foreground detection results. Considering the stationarity of segmented image

sequences, pixels representing foreground are labeled using either step 1° or an intersection of step 1° and 7°. In order to differ foreground from long period background appearance, state matching is also performed in spatial background memory using Equation (16) and considered as step 8°. Unlike MGMM [20], P-MGMM memorizes the whole of the first states in real-time state record as the latest background appearances once at the end of each segmented image sequence. That is, the most probable background of real-time state record is preserved. Concretely, the most probable background appearance B_t^1 representing the spatial mean of the first states in real-time state record is checked against each global Gaussian of background version in spatial background memory (i.e., step 9°), as expressed in Inequation (17). Nearest neighboring version matching is considered rather than order-first state matching. As to stationary sub-sequences, parameters of states and versions in spatial background memory are updated using Equation (1), as step 10°. When rapid intensity changes appear in scene, the weight estimation of the matched version $\omega_{k,t}^v$ and that of the corresponding states are updated using Equation (1) with its learning rate expressed in Equation (20). The mean $\mu_{k,t}^{s_v}$ and the standard deviation $\sigma_{k,t}^{s_v}$ of pixel-wise states corresponding to the matched version are derived from the first Gaussian in the real-time state record. Accordingly, the mean $\mu_{k,t}^v$ and the standard deviation $\sigma_{k,t}^v$ of the matched version are obtained from the spatial average of $\mu_{k,t}^{s_v}$ and $\sigma_{k,t}^{s_v}$. And that is step 11°. If no match exists, the last states in spatial background memory obtained from the descending order of $\omega_{k,t}^v$ should be replaced by B_t^1 at the end of each segmented sequence.

5 EXPERIMENTAL RESULTS AND ANALYSIS

In this section, the reason for introducing prior perception processor and hierarchical memorizing mechanism will be described. Using P-MGMM, experimental results of foreground detection and corresponding analysis are to be made. Matlab R2012b is selected as the experimental platform.

5.1 Reason for Introducing P-MGMM Using Experiment

First, an illustration why we propose three hypotheses of background subtraction is made. Let us examine a long-term image sequence from PETS01[2] (see Figure 3 a)). In Figure 3 b), a long-term intensity sequence labeled with line '-' is sampled at a fixed pixel of dataset PETS01-D3-T-C1. Stable background appearances frequently emerge; moreover, it can be seen that obvious intensity changes indicated in Figure 3 b) actually exist in short-term sequences. Therefore, a 'piecewise' action is made, i.e., we divide the long-term sequence into short-term ones of equal length with the endpoints labeled with '*'. The short-term sequences obtained are naturally categorized into two types. One represents a stationary sequence without any apparent intensity change. The other attends to the obvious intensity change with

[2] `http://ftp.pets.rdg.ac.uk/PETS2001`

great difference. The same process is in progress at the pixels of the selected area depicted in Figure 3 a), so that a certain spatial correlation is found. In light of these observations, we propose three hypotheses of background subtraction mentioned in Section 2.3.

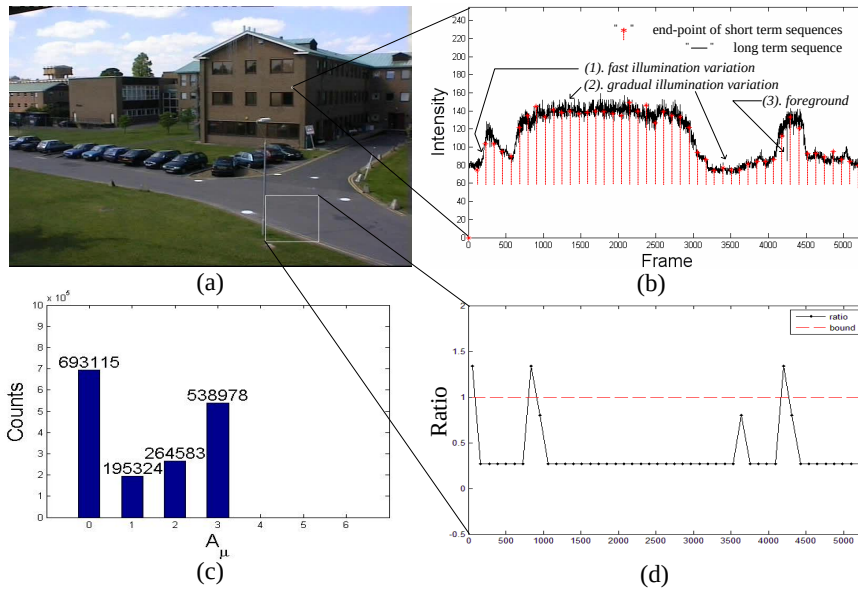


Figure 3. Illustration of the statistical characteristics derived from PET01. a) A long term image sequence with varying illumination, b) a time sequence of the selected pixel, c) a histogram of A_μ derived from the selected area, d) a segmented stationarity test result in each sub-sequence $\{X_{i,h_j}\}$ $(i = 1, 2, \ldots, 47;\ h = 1, 2, \ldots, 4;\ j = 1, 2, \ldots, 112)$ of the selected area.

The proposed hypotheses have indicated the importance of piecewise memorizing framework with a primary stage of perception. Therefore, the reason for introducing prior perception processor is further illustrated. The sum of the inversion number of each sub-sequence $\{\mu_{i,h}\}$, which is derived from Equation (4), forms a corresponding sequence as $\{A_{\mu_i}\}$ $(i = 1, 2, \ldots, l)$. We calculate A_{μ_i} in the selected area of Figure 3 a) to establish soundness of segmented stationarity test. In Figure 3 c), a histogram of A_{μ_i} is counted to all the pixels in the selected area. Let $r = 4$. Thus, $E(A_{\mu_i}) = 3$, as calculated by Equation (10). Due to the histogram count of A_{μ_i}, it is reasonable to assume that A_{μ_i} is a normal stochastic variable, as expressed in Equation (14).

We also identify different intensity changes in each sub-sequence $\{X_{i,h_j}\}$ of the selected area in dataset PETS01-D3-T-C1 by testing Inequation (15) to demonstrate its effectiveness, as shown in Figure 3 d). In order to indicate the reasonable structure

Figure 4. Long period background appearances of different video clips derived from PETS01-D3-T-C1. Each row depicts sorted background versions of the selected area in different clips, as shown in Figure 3. The 1st row corresponds to the 1st video clip. The 2nd row represents background states of the 8th video clip under fast illumination change. The last row illustrates long period background states kept in the 47th video clip. It is obvious that fast intensity change interferes with accurately memorizing long period background.

of our P-MGMM containing a spatial background memory for a long period memory, we experiment with the selected area illustrated in Figure 3 a). The memorized long period background states of different segmented image sequences are shown in Figure 4.

5.2 P-MGMM for Foreground Detection

In order to demonstrate the reasonability and the effectiveness of P-MGMM to foreground detection under complex environments, we tested and compared P-MGMM with seven statistical or structural background subtraction methods (including ALPCA [16], LBP [8], GMM [4], TGMM [17], HGMM [18], GMSM [19] and MGMM [20]) using 12 datasets from four databases. Each long-term video is simply segmented to ensure that each video clip has almost the same length, because of the observation that the detection results of foreground are independent of different l and M values. All the parameters of P-MGMM are user-settable. Besides, the optimal parameter values of comparative approaches for each dataset are selected. Experiments in [22] are repeated due to adding a prior perception processor in this work, as illustrated in Figure 5, Figures 6, 7 and 10. Let r and c be the row and column size of a frame. Therefore, the time and memory complexity of P-MGMM can be expressed as $O(Mrc)$ and $O(Klrc)$, respectively. The overall memory usage for processing each long-term video can be calculated by $bytes \times rc(KM + 2Kl)$, where

bytes represents the bytes needed for storage of each pixel. In addition, we listed the processing time of P-MGMM and several comparative background subtraction methods on F & GLC[3], as shown in Table 1.

Background Model	Average Processing Time (second/frame)	
	Gradual Illumination Change	Fast Illumination Change
ALPCA [16]	3.6102	4.6426
TGMM [17]	0.6927	0.9545
HGMM [18]	0.8367	1.1354
MGMM [20]	0.0825	0.1037
P-MGMM	0.0992	0.1250

Table 1. Average processing time of a single frame on F&GLC

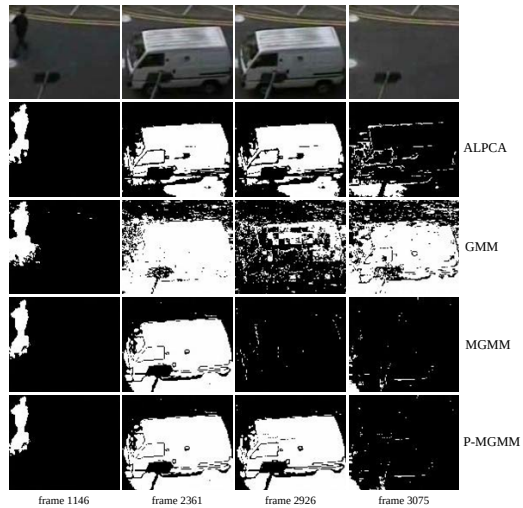


Figure 5. Experimental results on PETS01-D1-T-C2 ($l = 27$ and $M = 119$) containing *ghosts*. Each column depicts the detection results of the chosen area at different frames in different segmented video clips. Parameters of P-MGMM for PETS01-D1-T-C2 are given as follows: $K = 3$, $T = 0.34$, $\sigma_{low} = 2$, $\alpha = 0.001$, and $S = 3$ when the sub-sequence is non-stationary.

Ghosts. PETS01-D1-T-C2 is included for solving the problem of *ghosts* resulting from the exponential decay of long period background. Approaches such as ALPCA and MGMM keep the ability in memorizing background. That

[3] F & GLC is a dataset made by the second author of this paper in Harbin Institute of Technology where he did his doctorate.

is, these methods are immune from the appearances of *ghosts* when the occluded object used to be there still moves away. Therefore, we detect objects appearing in the selected area of PETS01-D1-T-C2 using ALPCA, GMM, MGMM and P-MGMM. Corresponding detection results display in rows, as illustrated in Figure 5. Still foreground is misclassified into background using GMM and MGMM (see column 3) due to absence of spatial consideration. GMM also performs poorly when the van is moving away (see column 4). As to ALPCA, there are several uncovered background pixels regarded as foreground because of the large scale of foreground compared to that of the scenario (see column 4). Due to its spatial background memory, it is observed that P-MGMM works well on resisting the object absorption from background and *ghosts*.

Light switch. Here we select PETS01-D3-T-C1 and PETS01-D3-T-C2 which contains *light switch* and *waving trees*. Methods including ALPCA, TGMM, HGMM and MGMM are viewed to be effective on solving the problem of *light switch*. First we perform P-MGMM and other algorithms on PETS01-D1-T-C1 and find that P-MGMM works better than most of the other methods except for some foreground apertures (see Figure 6). This phenomenon derives from using global difference memory, which can be overcome using erosion and dilation operations. Then we apply these algorithms to PETS01-D1-T-C2. The experimental results corresponding to ALPCA, GMM, TGMM, HGMM, MGMM and P-MGMM display in rows, as illustrated in Figure 7. It is observed that ALPCA is impervious to fast illumination change. However, it can be seen from Figure 7 that *waving trees* do interfere with the detection results due to frequent switches of components. Obviously, GMM, TGMM, HGMM and MGMM are incompetent at handling complex background illumination variations. P-MGMM generally performs better than the other comparative methods on discrimination between background varying illumination and foreground. Anyway, P-MGMM also shows its imperfections in *camouflage* and *foreground aperture*.

In order to present quantitative detection results, we select models adaptive to fast illumination change (i.e., ALPCA, TGMM, HGMM, MGMM and P-MGMM), and make comparisons on F&GLC. The experimental results are shown in Figure 8. ALPCA produces false positives when pedestrians of large scale pass by, because it is restricted by the number of learning frames and the scale of illumination scenarios compared to that of pedestrians. Background illumination is falsely considered to be foreground in TGMM confronted with weak texture features, although every parameter set of its regional texture measure is tried to operate on complex illumination variations. HGMM, which is only adaptive to ever-present fast illumination change, is to lapse when new global fast illumination change appears. In MGMM, initially memorizing step of limited pixel-wise states over training time is rendered ineffective by complex varying intensities, especially when new state with low learning rate appears in testing. With a much higher precision, P-MGMM performs better than

Figure 6. Experimental results on PETS01-D3-T-C1 under complex illumination variations with $l = 47$, $M = 102$. Each column illustrates the detection results of the chosen area at different frames in different video clips. Parameters of P-MGMM for PETS01-D3-T-C1 are given as follows: $K = 3$, $T = 0.34$, $\sigma_{low} = 1$, $\alpha = 0.005$, and $S = 3$ when the sub-sequence is non-stationary.

most of the other methods on discrimination between fast illumination change and foreground, for its instant identification of illumination changes based on a prior segmented stationarity test, rapid adaptation to intensity change of background and its capacity of memorizing long period background memory. On the other hand, more foreground *aperture* appears in the detection results of P-MGMM with a lower recall because of the similarity between foreground and long period background. However, this problem can be solved using morphological methods.

Shadows. CVRR-IR[4] are selected as an evaluation of shadow removal. Approaches such as TGMM and GMSM are viewed as effective methods for shadow removal. TGMM introduces an intensity integration at gray level. As to GMSM, a limitation of high appearance frequency is adopted in color space. In fact, elaborate experimental results indicate that TGMM and GMSM only works on weak shadow. Considering the real-time need of processing time, the gray-level intensity integration of TGMM is embedded into our P-MGMM. An experiment

[4] `http://cvrr.ucsd.edu/aton/shadow/`

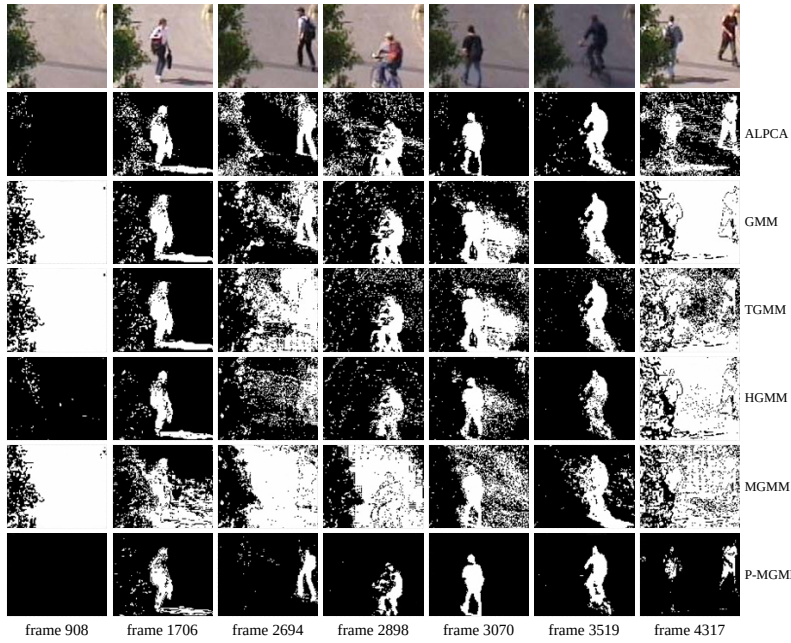


Figure 7. Experimental results on PETS01-D3-T-C2 accompanied with waving trees under complex illumination changes ($l = 53$, $M = 101$). Each column illustrates the detection results of the chosen area at different frames in different video clips. Parameters of P-MGMM for PETS01-D3-T-C2 are given as follows: $K = 5$, $T = 0.21$, $\sigma_{low} = 1$, $\alpha = 0.005$, and $S = 3$ when the sub-sequence is non-stationary.

is made on foreground detection of CVRR-S-I. The experimental results are illustrated in Figure 9. It is observed that P-MGMM keeps the same detection results as TGMM.

Wallflowers. In order to test our P-MGMM on problems besides *ghosts*, *light switch* and *shadows* and show its effectiveness with various disturbances of complex environments, qualitative and quantitative experiments are made on Wallflower[5] which includes seven sequences representing typical problems (i.e., LS, TD, WT, C, B, FA, MO). Together with P-MGMM, comparative algorithms are implemented. Qualitative and quantitative results are listed in Figure 10 and Table 2, which demonstrate the effectiveness of P-MGMM on foreground detection.

[5] `http://research.microsoft.com/en-us/um/people/jckrumm/wallflower/testimages.htm`

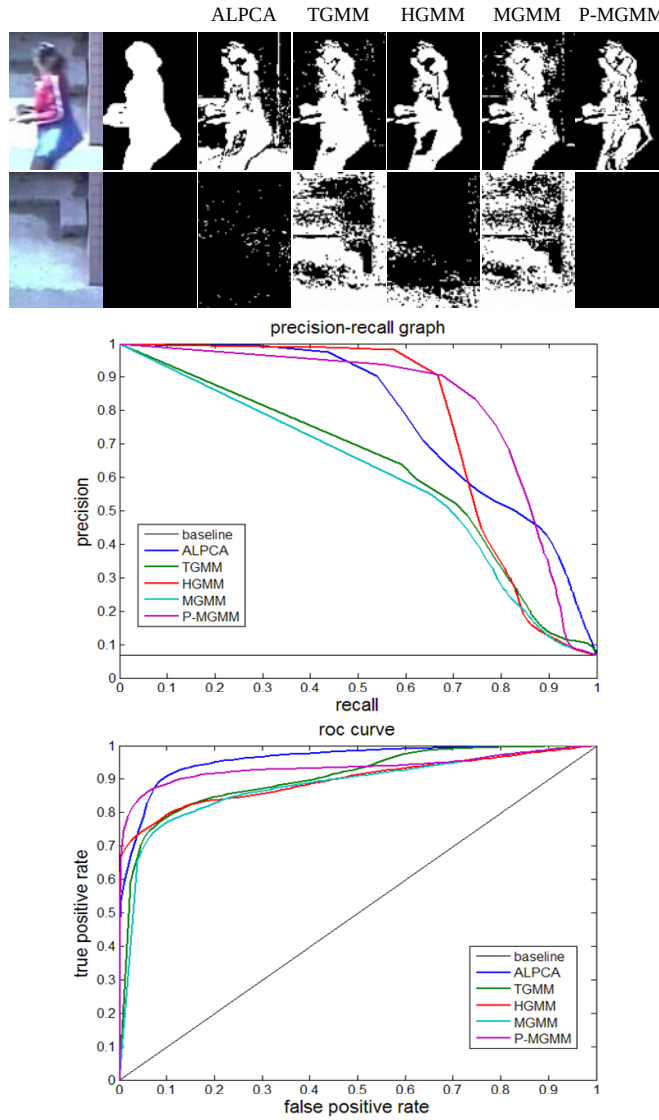


Figure 8. Experimental results on F & GLC containing an extreme fast illumination change accompanied with a pedestrian on high illumination condition ($l = 54$, $M = 118$). The 1st and 2nd column of row 1 and row 2 illustrate the chosen area on different frames and the corresponding ground truths of the 40th video clip. The 1st row shows the detection results of a pedestrian at frame 73. The 2nd row depicts the immunity of the selected methods to fast illumination change at frame 110. The 3rd and 4th row illustrate the precision-recall and ROC curve of the comparative algorithms, respectively. Parameters of P-MGMM for F & GLC are given as follows: $K = 3$, $T = 0.34$, $\sigma_{low} = 2$, $\alpha = 0.005$, and $S = 3$ when the sub-sequence is non-stationary.

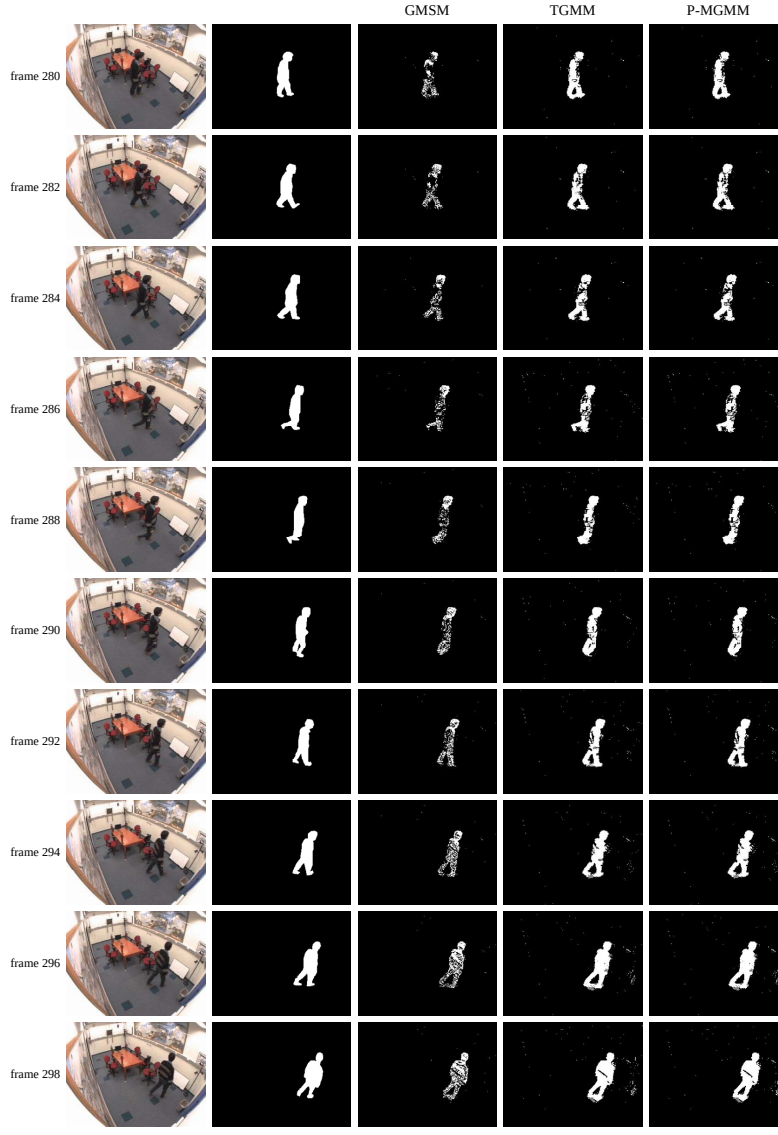


Figure 9. Experimental results on CVRR-S-I containing weak shadows with a pedestrian ($l = 4$, $M = 75$). Rows depict the detection results at a two-frame interval from frame 280 to frame 298. Columns from left to right correspond to original frames, ground truths, detection results of GMSM, TGMM and P-MGMM. Parameters of P-MGMM for CVRR-S-I are same as those for F & GLC.

		ALPCA [16]	LBP [8]	GMM [4]	TGMM [17]	HGMM [18]	GMSM [19]	MGMM [20]	P-MGMM
LS	Precision	0.2111	0.1763	0.1503	0.1504	0.1080	0.1359	0.1549	0.6914
	Recall	0.8884	0.6186	0.8589	0.8589	0.5513	0.6263	0.8927	0.7071
TD	Precision	0.7092	0.9781	0.4558	0.3740	0.4393	0.1643	0.3017	0.9421
	Recall	0.3372	0.2597	0.7261	0.7532	0.7196	0.2435	0.6828	0.5258
WT	Precision	0.5717	0.8663	0.6069	0.5924	0.7394	0.7219	0.6731	0.8198
	Recall	0.8451	0.1707	0.8512	0.8694	0.5480	0.9556	0.8305	0.5406
C	Precision	0.9572	0.8778	0.6789	0.6789	0.7394	0.8182	0.6520	0.8128
	Recall	0.8906	0.7347	0.9835	0.9835	0.0792	0.8769	0.8464	0.8718
B	Precision	0.2964	0.5854	0.4630	0.4747	0.6897	0.6588	0.5521	0.6797
	Recall	0.5925	0.1149	0.5238	0.5579	0.3959	0.5665	0.4315	0.3860
FA	Precision	0.8117	0.8234	0.2478	0.2478	0.2511	0.2745	0.2018	0.6680
	Recall	0.5115	0.5990	0.8033	0.8033	0.6916	0.3969	0.6215	0.5476
MO	Precision	0	0	0	0	0	0	0	0
	Recall	–	–	–	–	–	–	–	–

Table 2. Quantitative comparison on Wallflower

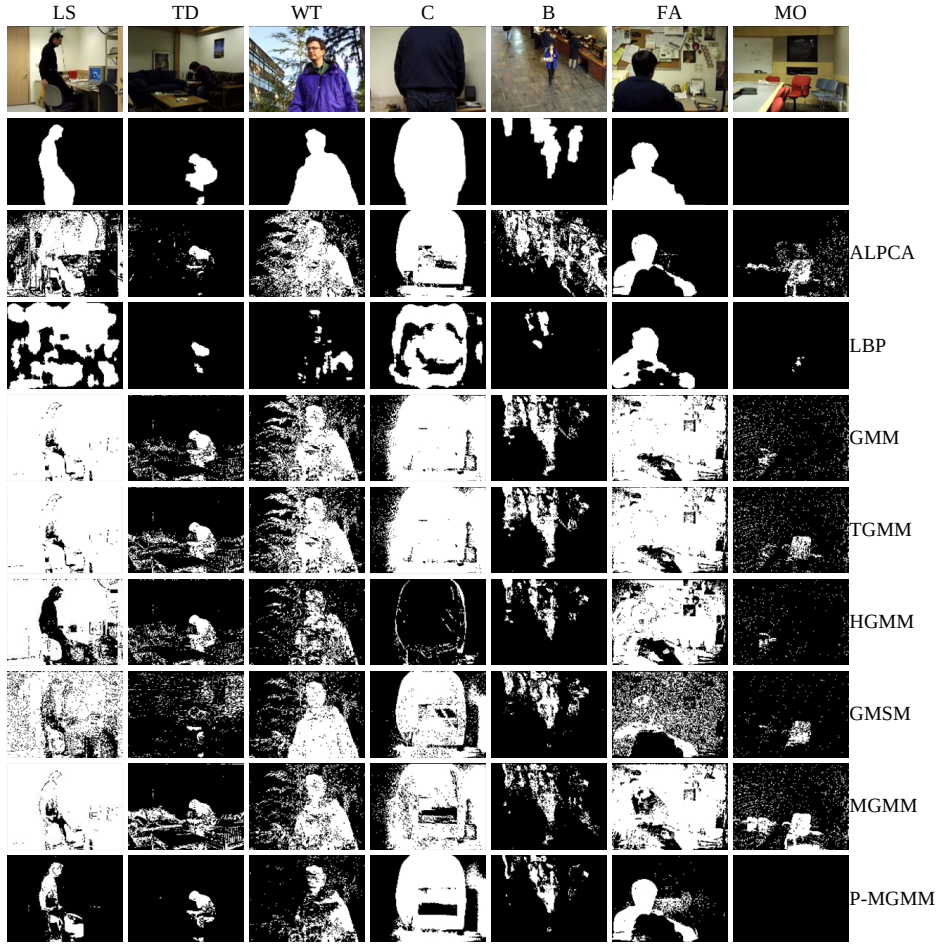


Figure 10. Experimental results on Wallflower with $M = 100$. The 1st and 2nd row illustrate the original frame and corresponding ground truth. Detection results including ALPCA, LBP, GMM, TGMM, HGMM, GMSM, MGMM and P-MGMM are illustrated from the 3rd to the 10th row.

Of course, the problems of *camouflage* and *foreground aperture* are left to be solved using erosion and dilation operations.

6 CONCLUSION

A piecewise memorizing framework, which is capable of solving most typical problems on background subtraction, especially forgetting of long-term memory back-

ground, is applied to GMM in this paper. Inspired by the metaphors of psychological selective attention theory, a prior perception-concerned recognition for stationary intensity test is presented, followed by a hierarchical memorizing mechanism containing real-time state record, spatial background memory and global difference memory. Real-time state record duplicates establishment and maintenance of prevailing exponential forgetting models such as GMM and LBP for short-term background memory. Enlightened from spatial information for fast adaptation to background variations (e.g. ALPCA and HGMM) and hierarchical memory strategy for enlarging memory capacity (e.g. MGMM), spatial background memory is developed. In order to ensure the robustness of the framework, global difference memory is designed and can be partially viewed as the first order difference of spatial background memory. Experimental results with various benchmark sequences have quantitatively and qualitatively demonstrated the effectiveness of our P-MGMM compared with many other statistical and structural background models on foreground detection with various disturbances of complex environments. Next, we will concentrate on applying the proposed framework to other background models in the literature.

Acknowledgements

This work was partially supported by the Specialized Personnel Start-up Grant (No. 41112419) and Fundamental Research Funds for the Central Universities (No. 2572017CY08) from Northeast Forestry University.

REFERENCES

[1] Zhou, X.—Yang, C.—Yu, W.: Moving Object Detection by Detecting Contiguous Outliers in the Low-Rank Representation. IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 35, 2013, No. 3, pp. 597–610.

[2] Srinivasan, R.—Rudolph, C.—Roy-Chowdhury, A. K.: Computerized Face Recognition in Renaissance Portrait Art: A Quantitative Measure for Identifying Uncertain Subjects in Ancient Portraits. IEEE Signal Processing Magazine, Vol. 32, 2015, No. 4, pp. 85–94, doi: 10.1109/MSP.2015.2410783.

[3] Chen, L. L.—Wang, W.—Panin, G.—Knoll, A.: Hierarchical Grid-Based Multi-People Tracking-by-Detection with Global Optimization. IEEE Transactions on Image Processing, Vol. 24, 2015, No. 11, pp. 4197–4212, doi: 10.1109/TIP.2015.2451013.

[4] Stauffer, C.—Grimson, W. E. L.: Learning Patterns of Activity Using Real-Time Tracking. IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 22, 2000, No. 8, pp. 747–757, doi: 10.1109/34.868677.

[5] Elgammal, A.—Harwood, D.—Davis, L.: Non-Parametric Model for Background Subtraction. In: Vernon, D. (Ed.): Computer Vision – ECCV 2000. Springer, Berlin, Heidelberg, Lecture Notes in Computer Science, Vol. 1843, 2000, pp. 751–767.

[6] Sheikh, Y.—Shah, M.: Bayesian Modeling of Dynamic Scenes for Object Detection. IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 27, 2005, No. 11, pp. 1778–1792, doi: 10.1109/TPAMI.2005.213.

[7] Monnet, A.—Mittal, A.—Paragios, N.—Ramesh, V.: Background Modeling and Subtraction of Dynamic Scenes. Proceedings of the 9th IEEE International Conference on Computer Vision (ICCV '03), Nice, October 2003, pp. 1305–1312, doi: 10.1109/ICCV.2003.1238641.

[8] Heikkilä, M.—Pietikäinen, M.: A Texture-Based Method for Modeling the Background and Detecting Moving Objects. IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 28, 2006, No. 4, pp. 657–662, doi: 10.1109/TPAMI.2006.68.

[9] Maddalena, L.—Petrosino, A.: A Self-Organizing Approach to Background Subtraction for Visual Surveillance Applications. IEEE Transactions on Image Processing, Vol. 17, 2008, No. 7, pp. 1168–1177, doi: 10.1109/TIP.2008.924285.

[10] Kim, K.—Chalidabhongse, T. H.—Harwood, D.—Davis, L.: Real-Time Foreground-Background Segmentation Using Codebook Model. Real-Time Imaging, Vol. 11, 2005, No. 3, pp. 172–185.

[11] Seki, M.—Wada, T.—Fujiwara, H.—Sumi, K.: Background Subtraction Based on Cooccurrence of Image Variations. Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR '03), Madison, June 2003, pp. 65–72, doi: 10.1109/CVPR.2003.1211453.

[12] Toyama, K.—Krumm, J.—Brumitt, B.—Meyers, B.: Wallflower: Principles and Practice of Background Maintenance. Proceedings of the 7th IEEE International Conference on Computer Vision (ICCV '99), Corfu, September 1999, pp. 255–261, doi: 10.1109/ICCV.1999.791228.

[13] Liu, C.—Yuen, P.—Qiu, G.: Object Motion Detection Using Information Theoretic Spatio-Temporal Saliency. Pattern Recognition, Vol. 42, 2009, No. 42, pp. 2897–2906.

[14] Zhao, X.—Satoh, Y.—Takauji, H.—Kaneko, S.—Iwata, K.—Ozaki, R.: Object Detection Based on a Robust and Accurate Statistical Multi-Point-Pair Model. Pattern Recognition, Vol. 44, 2011, No. 6, pp. 1296–1311.

[15] Chen, Y.—Chen, C.—Huang, C.—Hung, Y.: Efficient Hierarchical Method for Background Subtraction. Pattern Recognition, Vol. 40, 2007, No. 10, pp. 2706–2715.

[16] Dong, Y.—DeSouza, G. N.: Adaptive Learning of Multi-Subspace for Foreground Detection under Illumination Changes. Computer Vision and Image Understanding, Vol. 115, 2011, No. 1, pp. 31–49.

[17] Tian, Y.—Senior, A.—Lu, M.: Robust and Efficient Foreground Analysis in Complex Surveillance Videos. Machine Vision and Applications, Vol. 23, 2012, No. 5, pp. 967–983.

[18] Sun, Y.—Yuan, B.: Hierarchical GMM to Handle Sharp Changes in Moving Object Detection. Electronic Letters, Vol. 40, 2004, No. 13, pp. 801–802.

[19] Martel-Brisson, N.—Zaccarin, A.: Learning and Removing Cast Shadows Through a Multidistribution Approach. IEEE Transactions on Pattern Anal-

ysis and Machine Intelligence, Vol. 29, 2007, No. 7, pp. 1133–1146, doi: 10.1109/TPAMI.2007.1039.

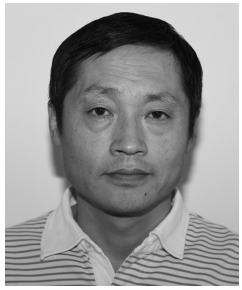
[20] Qi, Y.—Wang, Y.: Memory-Based Gaussian Mixture Modeling for Moving Object Detection in Indoor Scene with Sudden Partial Changes. Proceedings of the 10th International Conference on Signal Processing (ICSP '10), Beijing, October 2010, pp. 752–755, doi: 10.1109/ICOSP.2010.5655913.

[21] Friedman, N.—Russell, S.: Image Segmentation in Video Sequences: A Probabilistic Approach. Proceedings of the 13th Conference on Uncertainty in Artificial Intelligence (UAI '97), Paperback, August 1997, pp. 175–181.

[22] Zhao, W.—Zhao, X. D.—Liu, W. M.—Tang X. L.: Long-Term Background Memory Based on Gaussian Mixture Model. Proceedings of Visual Communications and Image Processing (VCIP 2013), 2013, Kuching, November 2013, pp. 1–5.

[23] Galotti, K. M.: Cognitive Psychology in and out of the Laboratory. 4th ed. Thomson Wadsworth Press, Belmont, CA, 2008.

[24] Priestley, M.: Non-Linear and Non-Stationary Time Series Analysis. Academic Press, London, UK, 1988.



Songbo Liu works in the School of Computer Science, Harbin Institute of Technology, Harbin, China. He is now Assistant Professor in Harbin Institute of Technology. His research interests focus on pattern recognition, embedded systems, and intelligent control.

Xudong Zhao works in the College of Information and Computer Engineering, Northeast Forestry University, Harbin, China. He received his Bachelor, Master and Ph.D. degrees from Harbin Institute of Techology in 2003, 2007 and 2013, respectively. After that, he pursued his postdoctoral fellowship in Chinese University of Hong Kong, and now he is Assistant Professor in Northeast Forestry University. His research interests focus on pattern recognition, machine learning, and biostatistics.

Xianglong Tang works in the School of Computer Science, Harbin Institute of Technology, Harbin, China. He is now Professor in Harbin Institute of Technology. His research interests focus on pattern recognition, machine learning, and image processing.