Modelling land water composition scene for maritime traffic surveillance

Shaoning Pang* and Jing Zhao
Department of Computing,
Unitec Institute of Technology,
Private Bag 92006,
Auckland 1025, New Zealand
Email: ppang@unitec.ac.nz
Email: jzhao@unitec.ac.nz
*Corresponding author

Bruce Hartill
National Institute of Water and Atmospheric Research,
41 Market Place, Viaduct Harbour,
Auckland 1010, New Zealand
Email: bruce.hartill@niwa.co.nz

Abdolhossein Sarrafzadeh
Department of Computing,
Unitec Institute of Technology,
Private Bag 92006,
Auckland 1025, New Zealand
Email: hsarrafzadeh@unitec.ac.nz

Abstract: Background modelling, used in many vision systems, must be robust to environmental change, yet sensitive enough to identify all moving objects of interest. Existing background modelling approaches have been developed to interpret images in terrestrial situations, such as car parks and stretches of road, where objects move in a smooth manner and the background is relatively consistent. In the context of maritime boat ramps surveillance, this paper proposes a cognitive background modelling method for land and water composition scenes (CBM-lw) to interpret the traffic of boats passing across boat ramps. We compute an adaptive learning rate to account for changes on land and water composition scenes, in which a geometrical model is integrated with pixel classification to determine the portion of water changes caused by tidal dynamics and other environmental influences. Experimental comparative tests and quantitative performance evaluations of real-world boat-flow monitoring traffic sequences demonstrate the benefits of the proposed algorithm.

Keywords: background modelling; moving object detection; marine traffic; land and water composition scene; dynamic learning rate.
1 Introduction

Background modelling is a core component for computer vision-based surveillance that uses data from each successive frame to calculate and update a background model that provides a statistical description of the entire background scene (Zhang et al., 2014; Rivera et al., 2013; Yang et al., 2015). Extensive research has been devoted to developing...
a background model that is robust to environmental changes in the background, but sensitive enough to identify all moving objects of interest. In practice, background modelling must be generic enough to handle a variety of changing background scenes to ensure that objects of interest are accurately identified.

Background modelling in the literature has been applied in a variety of situations including: motorways (Unzueta et al., 2012; Mithun et al., 2012), road intersections (Messelodi et al., 2005; Ottlik and Nagel, 2008), car parks (choeychuen, 2012, 2013), swimming pools (Eng et al., 2004; Nuno et al., 2009) and water channels (Bloisi and Iocchi, 2009; Bloisi et al., 2014), etc. In general, we categorise different types of scenes into two groups: land scenes and water scenes, as background dynamics in these contexts differ markedly. On land, the background is usually static, with little or no change in topography. In contrast, water scenes are intrinsically dynamic, as water is a reflective surface that moves continuously, often to varying degrees. The reflection of the sun on water, coupled with the unpredictability of waves caused by the wind, moving vessels (wakes) and tidal flows in the maritime environment creates situations where background modelling is far more challenging.

Many background modelling algorithms have been proposed in the literature, but most of these methods are limited to indoor scenes or very special outdoor scenes, such as highways, inter-sections, etc. In other words, very few previous background modelling studies have considered the marine environment, and the majority of existing work has focused on more static terrestrial situations. Thus, the problem of modelling backgrounds in marine situations and identifying moving objects in these dynamic environments is still far from being completely solved.

This paper considers background modelling for a dynamic maritime environment, at the interface between the land and the sea. Figure 1 shows an example of maritime boat ramp which our research is based on. As seen in this figure, the region-of-interest (ROI), identified as a polygon contains both areas of land and water. The boundary between the water and land changes over time with the rise and fall of the tide. As a consequence, the distribution of water and land varies over time in the ROI, which makes background modelling in this case extremely difficult given the varying area of water and amount of light reflected from this water at differing times of day and sun angle.

In this paper, we propose a cognitive background modelling for land and water composition scenes (CBM-lw), in which the way of human problem solving and mental task processes is simulated in solving difficulties such as noise reduction, land and water scene distinction, and handling illumination changes caused by tide, sunrise and sunset. The CBM-lw approach can be used to dynamically classify areas of an image as either land or water, given ancillary tidal height data so that different strategies can be adapted to model backgrounds on land and on the water, respectively. The impact of sunrise and sunset is also specifically considered by proposed CBM-lw, to allow for changes in outdoor luminance. The CBM-lw approach can also be used to identify corrupted (and therefore unusable) images, to improve the accuracy of background model. In particular, the use of dynamic learning rate and intelligent updating rules for areas of land and water, respectively, significantly increases the robustness of CBM-lw method. Moreover, the low computational cost and low level of memory required for this approach maximises its potential utility. We apply the CBM-lw to real 24-hour boat-flow analysis and counting system and compare it against existing methods for background
modelling; the empirical results show that the proposed CBM-lw achieves better performance.

Figure 1  An example of maritime boat ramp: the ROI includes areas of land and water as seen inside the red polygon, and the boundary between water and land is shown by the yellow line (see online version for colours)

The rest of the paper is organised as follows: Section 2 introduces related works and an overview of our algorithm. Section 3 illustrates the proposed background modelling algorithm, and also explains how our approach differs from existing methods. In Section 4, we describe our experimental setup, compare the proposed algorithm with existing background modelling methods and present experimental results. Finally, we conclude the paper with a discussion and an outlook on further extensions in Section 5.

For the remainder of this paper, we use $I(x, y)$ and $B_t(x, y)$ to represent the luminance intensity and its background estimation at location $(x, y)$ and time $t$, respectively. If irrelevant in the description, the coordinate $(x, y)$ may be dropped.

2 Related work

Detecting moving objects from image sequences is a fundamental task undertaken in many computer vision applications, such as traffic monitoring and analysis (Buch et al., 2009; Jung et al., 2001), video surveillance (Johansson et al., 2009), human detection and tracking (Buch et al., 2010), gesture recognition (Guo et al., 2008), etc. A very common technique for the detection of moving objects is background subtraction (Gao et al., 2009), where each frame is compared against a background model, and pixels in the current image that deviate significantly from the background model are considered to be moving objects. For accurate foreground object detection, a background model is required to be a good estimation of the scenario after removing all non-stationary elements.
The approaches to background modelling can be divided broadly into two categories, namely non-recursive and recursive methods (Cheung and Kamath, 2004). Non-recursive models use a sliding-window strategy that relies on a buffer of recent images to estimate the background model, which is not influenced by images taken outside of the window period; whereas recursive methods do not maintain a buffer for background estimation. They recursively update a background model based on the current image, thus images from distant past could have influence on the current background model.

2.1 Recursive method

Amongst recursive methods, a variety of Kalman filter-based methods have been developed for background modelling. The major differences between these methods are the state spaces used for tracking. The simplest state tracking method only considers luminance intensity (Wren et al., 1997), whereas Koller et al. (1993) employed the intensity and its spatial derivative for background modelling. Friedman and Russell (1997) proposed first the mixture of Gaussian (MoG) approach, in which multiple Gaussian distributions are taken into account simultaneously by a Kalman filter. Based on this work, Gupte et al. (2002) developed a self-adaptive background subtraction method for segmentation; and Porikli and Tuzel (2005) revised the MoG to deal with dynamic scenarios, in which a set of layers of 3-D multivariate Gaussians are employed to describe each background pixel. A method of using a simple recursive filter (i.e., non-Kalman filter-based method) to estimate the median for background modelling was introduced in McFarlane and Schofield (1995). Based on this work, Remagnino (1997) proposed a new pixel level tracking method, in which the running estimate of the median is increased by one if the input pixel is bigger than the estimation and decreased by one if smaller. An alternative recursive method is online expectation maximisation (EM) proposed by Stauffer and Grimson (1999). Zivkovic (2004) improved the efficiency of the algorithm by developing an adaptive method to update parameters. Later, Lee (2005) revised the same algorithm with short-term statistics to model the non-stationary backgrounds.

2.2 Non-recursive method

In this category, simply use the \( t - 1 \)th frame for building background model at time \( t \), frame differencing is a fast speed background model suitable for stable environment monitoring (Park et al., 2007; Nguyen and Le, 2008). This narrow window approach does not, however, properly account for changes in levels of illumination, noise or periodic movements in the background.

Median filter is also commonly used in simple background model (Cucchiara et al., 2003), in which the background estimation is defined for each pixel as the median of all the images in the buffer. There are many other general methods proposed for background modelling. For example, Elgammal et al. (1999) proposed a non-parametric model for background modelling on the entire history images; Cheung and Kamath (2005) developed a linear predictive filter on the pixels in the buffer to calculate the current background model; and Mittal and Paragios (2004) built background pixel density functions in a higher dimensional space combining intensity with optical flow. However, most of these methods have been originally designed to interpret relatively static scenes.
on land, and relatively few studies have considered background modelling in more dynamic water settings.

2.3 Water background modelling


Another water-based application of background modelling is that of Eng et al. (2004, 2006) who considered outdoor swimming pool surveillance, and modelled the background as regions of dynamic homogeneous processes, and developed a spatial-temporal filtering scheme to enhance the detection of swimmers who were partially hidden by reflections of artificial lighting. Bloisi and Iocchi (2009) and Bloisi et al. (2014) conducted boat traffic monitoring of a Venetian water channel. In this study, a video surveillance system (ARGOS) was developed, that used a Gaussian mixture model-based mean-filter method integrated with an approximated online clustering mechanism. ARGOS was able to build a reliable background model of the water channel and to track the boats navigating the water channel with good accuracy in real-time applications. For specifically dealing with water background, they later proposed an independent multimodal background subtraction (IMBS) approach (Bloisi et al., 2014), which used an online clustering algorithm to capture the multimodal nature of the background as well as a model update mechanism to detect changes in background geometry. The algorithm was successfully demonstrated, and currently provides the basis of a real boat traffic surveillance system in Venice.

2.4 Land and water composition scene processing

Land and water composition scene has been attended in some existing works. Li and Perona (2005) studied coast (a land-water composition scene) and other nature scenes by representing the image of a scene as a collection of local regions. Similarly, Boutella et al. (2004) proposed a framework of multi-label classification, in which a sample may possess multiple properties of multiple classes, and applied it to the problem of semantic scene classification. A natural scene may contain multiple objects, for example, beach may contain sand and water, such that the scene can be described by multiple class labels. Li et al. (2009) dealt with composite scenes by segmenting object components meanwhile
using contextual information (e.g., a list of tags) for classification. Later they (Li et al., 2010) proposed to use objects as attributes of scenes for scene classification, where an image of composite scene is represented as its responses to a large number of object detectors.

2.5 Motivation

The situation we consider in this paper differs from those above, as we are attempting to model the dynamic background at the interface between the land and the sea, at high traffic boat ramps on New Zealand’s north-eastern coast. The background composition of the scene at a boat ramp can change substantially as tidal heights, levels of illumination, and reflectance vary throughout the day and throughout the year.

Existing works mostly attempt to develop a generic background model without the used of scene prior knowledge in the modelling. To cope with the extremely challenging land and water composition scene, we conduct cognitive background modelling, following the way of human problem solving and mental task processes for background estimation. The proposed CBM-lw can classify areas of each image as either land or water, given ancillary model data on predicted tidal height, so that different strategies can be adopted to model the background on land and on the water, respectively. The influence of sunrise and sunset is also considered by the CBM-lw approach to account for changes in outdoor luminance. In addition, the proposed CBM-lw approach also makes the use of a dynamic learning rate and can be used to identify and reject corrupted images, to improve the accuracy of background modelling.

3 Proposed cognitive background modelling

The philosophy of our method is to use cognitive modelling to solve dynamically shifting areas of land and water background modelling problems. We address the issue of detecting corrupted images mainly caused by radio interference, which would reduce the quality of any computed background model. As background dynamics in land and water scenes differ markedly, our method separates land and water scenes by constructing a cognitive computing model, in which a geometrical calculation based on tide data gives initial shoreline estimation and pixel classification helps determine the optimal boundary between areas of land and water. Additionally, different learning rates and background updating strategies are exploited for areas of land and water to improve modelling performance. Moreover, our method considers the influence of sunrise and sunset to account for changes in outdoor luminance. Note that occlusion is not an issue of object detection for our boat ramps surveillance because:

1 camera view has been adjusted to avoid occlusions
2 for each boat ramp lane, just one vehicle or boat-vehicle combo is physically able to pass through at one time which minimises the occurrence of occlusions.
**Algorithm 1:** Cognitive background modelling for land and water composition scenes (CBM-lw)

**Input:** Current frame $I_t$, current background $B_t$, ROI coordinates, shoreline area coordinates $S$, and tidal height data $H$.

**Output:** Next background $B_{t+1}$.

1: /* Detecting unusable frame */
2: if detection unusable($I_t$) == 1 then // equation (5)
   
   $B_{t+1} \leftarrow B_t$
3: else
   /* land and water area separation */
5:  $b^* \leftarrow$ tide_interpolation($H$); # Subsection 3.3
6:  pixel_classification($I_t(x, y), S$); # Subsection 3.3
7: /* search $k^*$ in shoreline area */
8:  $k^* \leftarrow \arg \min_{k^*} (\arctan \frac{|k^* - g|}{1 + k^*})$; # equation (9)
9:  $(\alpha_0, \alpha_w) \leftarrow$ compute_learningrate($I_t, B_t, k^*, b^*$); # equation (11)
Modelling land water composition scene for maritime traffic surveillance

10: /* the influence of sunrise and sunset */
11: if \( t \) falls in the sunrise/sunset zone then
12: \((\alpha_l, \alpha_w) \leftarrow \text{sunrisesunset\_adjust}(t)\); \# Subsection 3.5
13: end if
14: /* background updating */
15: \( B_{\text{IB}}(t+1) \leftarrow \alpha_l \text{IB}_t + (1 - \alpha_l)B_t \), and \( B_{\text{IB}}(t+1) \leftarrow \alpha_w \text{IB}_t + (1 - \alpha_w)B_t \); \# equation (12)
16: \( B_{\text{IB}}(t+1) \leftarrow B_{\text{IB}}(t+1) \cup B_{\text{IB}}(t+1) \) \# equation (13)
17: end if

The following subsections describe the proposed CBM-lw in more detail. For an overview of the proposed CBM-lw, Figure 2 gives a block diagram and Algorithm 1 gives pseudocode of the core structure.

3.1 Base background model

Given \( \{I_t\}_{t=1}^T \) as current set of images in observation, and polygon \( R \) as the ROI. We introduce the following signum function of image pixels as,

\[
I(i, j) = \begin{cases} 
I_t(i, j) & \text{if } (i, j) \in R, \\
-1 & \text{otherwise}. 
\end{cases}
\] (1)

By equation (1), we are able to cast any image (or the same size matrix) and related calculations into the ROI block. In practice, we define the ROI according to the prior knowledge that we had from the physical ramp in surveillance. For example of Raglan boat ramp shown in Figure 1, objects of interest (e.g., car or boat) seldom stop on the grass, and boats in sea centre are not relevant to our target. Thus, we are able to define the polygon ROI according to the ramp physical conditions and project requirements. The aim of this operation is to reduce as much as possible the influence of noise to later on all steps image calculation.

According to Gupte et al. (2002), the next background \( B_{\text{IB}} \) is generated using the weighted average of the instantaneous background \( \text{IB}_t \) and the current background \( B_t \):

\[
B_{\text{IB}} = \alpha \text{IB}_t + (1 - \alpha)B_t. 
\] (2)

Here, \( B_t \) is initialised as an image of the background without any moving objects present. \( \alpha \) is the learning rate which determines the updating speed of background. In practice, \( \alpha \) should be big enough so that background modelling algorithms can adapt rapidly to changes in the background, but small enough so that they are not sensitive to momentary changes. Because the background is influenced by changing luminance, different weather conditions, etc., it is essential for the background modelling algorithm to adopt a dynamic learning rate to optimise performance. We have therefore adopted a dynamic learning rate in the proposed method, rather than a static rate in Gupte et al. (2002).

The instantaneous background \( \text{IB}_t \) is generated from the current image but with detected objects removed, and their regions are filled with the current background. Specifically, for each incoming image \( I_t \), we calculate its differences to current background \( B_t \), and threshold the resulting difference image to obtain a binary object mask,
where \( \lambda \) refers to the luminance threshold for object detection. The current instantaneous background \( IB_i \) can then be calculated as,

\[
IB_i(x, y) = \begin{cases} 
B_i(x, y) & \text{if } O_i(x, y) = 1, \\
I_i(x, y) & \text{otherwise.}
\end{cases}
\]

(4)

Note that the above calculations are all in terms of gray-level intensity. In the case of colour image, we simply transform the image to gray-level before any calculation.

3.2 Detection of unusable images

The image data used in this study is provided by web cameras overlooking boat ramps. These web camera systems wirelessly transmit video data to a nearby PC for storage. Unfortunately, images are sometimes corrupted by radio interference, or because some other part of the system has failed, and it is necessary to detect when this has occurred. Corrupted images are of no practical use, because any detection of objects from these images would be unreliable. Further, the use of these corrupted images would reduce the quality of any background model produced from an affected time series of images.

In proposed method, every input frame is inspected and classified as usable or unusable, and only usable frames are employed to generate the next background. Consider boat ramps surveillance, a major portion of each image is composed of the background, unless a big amount of noise is introduced in image. In other words, the binary object mask for a usable image would consist of a large number of pixels having the value 0 (i.e., background pixel), and a small number of pixels having the value 1 (i.e., object pixel).

We calculate a histogram of the binary object mask \( O_i \) in equation (3) as \( h(O_i) \). Let \( h(O_i)^1 \) and \( h(O_i)^0 \) denote the number of 1 and 0 in \( h(O_i) \), respectively. In determining if the current image is a unusable image, we simply calculate the ratio of pixel number as,

\[
U(O_i, \delta) = \begin{cases} 
1 & \text{if } h(O_i)^1 / h(O_i)^0 > \delta, \\
0 & \text{otherwise}
\end{cases}
\]

(5)

where \( \delta \) denotes the threshold which can be easily determined by cross validation as the histogram of the binary object mask is expected to have much higher number of value 0 than that of value 1.

3.3 Land and water area separation through cognitive modelling

The goal of separation is to segment the ROI into areas of land and water. In the image coordinate system shown in Figure 3, the distinction between areas of land and water can be interpolated as a geometric problem, which is to find/fix a straight line as,

\[
n = km + b
\]

(6)
where $k$ and $b$ refer to the slope and intercept of straight line, respectively. Thus to determine the shoreline, the task is to search for optimal values of $k$ and $b$.

Physically, we look the sea as a large container, with the amount of water determined the position of tidal boundary. In this sense, for a specific ramp, we are able to determine the optimal $b^*$ given tidal height data $H$ provided by an ancillary model using interpolation methods such as linear interpolation, polynomial interpolation, or spline interpolation, etc. As a result, we have the revised shoreline function as,

$$n = km + b^*$$  \hspace{1cm} (7)

**Figure 3** An illustration of searching optimal boundary between water and land (see online version for colours)

However, the slope $k$ varies over time, as the direction of the boundary between the land and the water is not only determined by the shape of container, but also by the prevailing weather conditions such as the wind direction. Here, the proposed solution is to classify all ROI pixels into land and water area, then we seek the optimal slope $k^*$ by a searching process described below.

Let $D_t$ denote a land-water distribution matrix of current image $I_t$, we can find the land water border line by accurately classifying every pixel as covering either land or water. $D_t$ can be obtained by a binary pixel classification, which can be formulated as a convex optimisation problem, i.e., the task of finding a minimiser of a convex function $f$ that depends on a variable vector $w$. Formally, we formulate this as an optimisation problem, where the objective function is of the form

$$f(w) = \frac{1}{2} w^T \omega + C \sum_{i=1}^l \max \{1 - y_i w^T x_i, 0\},$$  \hspace{1cm} (8)

Here, the vectors $x_i \in \mathbb{R}^d$ are the training data examples, for $1 \leq i \leq r$, and $y_i \in [-1, 1]$ are their corresponding labels, which we want to predict. Consequently with the SVM trained, every pixel in $S$ is classified as either land or water. It is not difficult to model a line $n = gm + l$ that gives a pixel level shoreline approximation regardless of tide change.

Consider shoreline approximation in Figure 3, by equation (7) we have point $E$ that the actual land/water boundary should have gone through, and its distance to $O$ is $b^*$. Without loss of generality, we can define for every ramp in surveillance a maximum
margin for all possible shorelines. In the example of Figure 3, rectangle $ABCD$ is the margin area which we denote hereafter as $S$. To find the optimal slope $k^*$, we rotate line [equation (7)] around $E$ by trying every possible slope $k'$ that directs the line going through margin $S$. For each test line, we calculate its angle to the land-water border line comes from equation (8) for land and water pixel classification. Thus, we have optimised slope $k^*$ calculated as,

$$k^* \leftarrow \arg \min_{k \in S} \left( \frac{k' - g}{1 + k'g} \right)$$

(9)

3.4 Background learning rate calculation

Under the condition of land and water area distinction, we are able to model land and water backgrounds separately by applying equation (2) to land and water image block as,

$$B_{l,t+1} = \alpha_l B_{l,t} + (1 - \alpha_l) B_t,$$

$$B_{w,t+1} = \alpha_w B_{w,t} + (1 - \alpha_w) B_t,$$

(10)

where $\alpha_l$ and $\alpha_w$ represent the learning rate for land and water scenes respectively.

The proposed method adopts a dynamic learning rate for land and water area respectively rather than a static one in Gupte et al. (2002).

We maintain a 24-hour learning rate buffer, in which pair of land and water learning rates ($\alpha_l$, $\alpha_w$) is stored at every minute. The process for computing learning rates is described below.

Consider in our case that, the objective is traffic analysis; more specifically, to count the number of boats/cars passing through the ramps. Thus, the number of objects is the ground truth of our background modelling. When determining the learning rate at time $t$, we search the optimal rates in the criterion of minimising the error of objects counting as,

$$\left( \alpha_l^*, \alpha_w^* \right) \leftarrow \arg \min_{\alpha_l, \alpha_w \in [0, 1]} \left\{ \left\| \psi_l - \eta_l \right\| + \left\| \psi_w - \eta_w \right\| \right\},$$

(11)

where $\eta_l$ and $\eta_w$ are the predicted number and $\psi_l$ and $\psi_w$ are the actual number of objects in land and water area respectively. Here $\psi_l$ and $\psi_w$ can be obtained by manually viewing each frame. In practice, this is a time-consuming process. For simplicity, we implement equation (11) by counting the total number of objects using currently computed background regardless of water or land scene.

3.5 Accounting for the influence of sunrise and sunset

In complex outdoor scenes, the level of luminance is easily influenced by several factors, such as time of day, cloud cover, time of year, available street lighting, etc. The rate of luminance change is obviously higher during sunrise and sunset than that of other time. It is therefore necessary to specifically consider the influence of sunrise and sunset as part of our background modelling.

Sunrise and sunset occur during two brief periods every 24 hours. Publicly available sunrise/sunset data provided by Land Information New Zealand (http://www.linz.govt.nz) gives an accurate estimate of when sunrise and sunset occur, which changes throughout
the year. In practice, the influence of sunrise and sunset on the learning rate follows a specific pattern. We assume the pattern gradually changes throughout the year, which gives a fixed learning rate pattern for sunrise and sunset on different days.

According to the learning rate pattern, we refresh the learning rate for every minute during a 24-hour loop (i.e., learning rate buffer), given the shifting timing of sunrise and sunset during the year. For every incoming image, we search in the buffer suitable rates by time $t$, if $t$ falls during the expected period of sunrise or sunset, the pattern learning rates are assigned to $\alpha_w$ and $\alpha_l$.

### 3.6 Background updating

Our strategy for updating background consists of two steps: firstly update land and water background respectively as,

$$B_{l,t+1} = \alpha^*_w IB_t + (1 - \alpha^*_w) B_t,$$

$$B_{w,t+1} = \alpha^*_l IB_t + (1 - \alpha^*_l) B_t.$$  \hspace{1cm} (12)

Then merge obtained land and water backgrounds into the next background,

$$B_{t+1} = B_{w,t+1} \cup B_{l,t+1}.  \hspace{1cm} (13)$$

Note that $B_{t+1}$ here is an ROI image. To have the entire background image, we can simply merge $B_{t+1}$ with the ROI residual image which can be obtained by applying a reversed function (1) on $I_t$.

### 4 Experimental results

The image data we used for our experiments was collected by New Zealand’s National Institute of Water and Atmospheric Research (NIWA). NIWA has established a network of web cameras overlooking key boat ramps, on behalf of the Ministry for Primary Industries, to monitor trends in recreational fishing effort over time. In this monitoring system, one image is captured per minute for each web camera, providing 1,440 images of a monitored ramp on each day. These images are viewed in series by a technician who manually interprets these images and records a count of returning boats for that day. Some form of random stratified sub-sampling is applied in practice, and images on 60 days per year are interpreted to provide a cost effective estimate of the whole year.
Table 1 describes our experimental data, which are the 2010–2012 image series captured at Waitangi, Takapuna and Raglan boat ramp. We compared first with the SABS method (Gupte et al., 2002), which is initially designed for detecting vehicles in a terrestrial situation, such as on a highway, and which is proposed CBM-lw based on. In all experiments, we set the learning rate $\alpha$ in equation (2) to 0.1 for SABS according to the authors’ recommendations (Gupte et al., 2002). Here, three measures were used to assess the relative utility of the CBM-lw and SABS background modelling approaches: their ability to detect data provided by corrupted images; their robustness to changes in levels of luminance at sunrise and sunset, and robustness to changes in tidal height. To further demonstrate the background modelling contribution to moving object detection, we evaluate the object detection performance in land and water composition scenes, the most difficult scenarios for background modelling, using the background model from our proposed method with comparison to that of four state-of-the-art methods. Quantitatively, we use object detection accuracy to evaluate the performance of background modelling, which is defined as the ratio of the number of objects correctly detected by the algorithm against the number of objects from manual counting.

**Figure 4** Object detection accuracy with respect to the luminance threshold $\lambda$ for object detection, (a) Waitangi (b) Takapuna (c) Raglan (see online version for colours)

Note: Each curve represents different values of $\delta$ used to detect unusable image.
Figure 5  DET curves for object detection performance evaluation, (a) Waitangi (b) Takapuna (c) Raglan (see online version for colours)

Note: Each curve represents different values of the luminance threshold $\lambda$ and $\delta$ used to detect unusable image.

### 4.1 Parameter adjustment

In the proposed CBM-lw method, two parameters are required. The first one is the luminance threshold $\lambda$ for object detection. The second one is the threshold $\delta$, which is used to detect unusable image. To evaluate effects of different $\lambda$ and $\delta$, we set the range of $\lambda$ from 0.1 to 0.6 and the range of $\delta$ from 0.1 to 0.5. Figure 4 shows object detection accuracy of each boat ramp with respect to different values of $\lambda$, respectively. Each curve represents different values of $\delta$. As shown in Figure 4, if the value of $\delta$ is fixed, the best object detection accuracy is achieved when $\lambda = 0.2$ for all three boat ramps, we then select $\lambda = 0.2$. If $\lambda < 0.2$, lower values of $\lambda$ achieve lower detection accuracy. In contrast, if $\lambda > 0.2$, higher values of $\lambda$ achieve lower detection accuracy. These results are reasonable, because lower value of $\lambda$ will cause more false positives, i.e., more noise will be detected as moving objects; whereas higher value of $\lambda$ will cause more false negatives, i.e., more moving objects cannot be identified from images. Similarly, if the value of $\lambda$ is fixed, the best object detection accuracy is achieved when $\delta = 0.3$ for all three boat ramps as shown in Figure 4. Thus, we select $\delta = 0.3$ for the threshold of unusable image detection. Further, the optimal combination of $\lambda = 0.2$ and $\delta = 0.3$ is confirmed in
Figure 5, where system detection error trade-off curves are studied with the setup of different $\lambda$ and $\delta$ values.

### 4.2 Detecting unusable images

Figure 6 compares the ability of the existing SABS and proposed CBM-lw approaches to detect images that were corrupted due to system failure. The first column of Figure 6 corresponds to the current frame of an image sequence (i.e., T1, T4, T5 and T6) in observation. According to equation (5), the noise ratios for frame T4-T6 are 1.16, 1.31 and 1.22 respectively, which are all far greater than the acceptable noise threshold 0.3. These images are unusable as they do not provide a sufficiently clear image of the boat ramp. The background modelled using the SABS approach is shown in the second column, whereas the third column gives the background from proposed CBM-lw. The background model provided by the SABS quickly degrades as it has assimilated the noise from the unusable images. In contrast, proposed CBM-lw approach has discounted data from the corrupted images, providing a background model which is unpolluted by noisy images.

**Figure 6** Comparisons of SABS and proposed CBM-lw modelled backgrounds derived from a series of images including some which were corrupted by interference, (left to right) current image, background from SABS and background from proposed CBM-lw
4.3 Robustness to changes in luminance at sunrise and sunset

In this comparative study, we demonstrate algorithm robustness to sunrise and sunset luminance changes. As we know, the periods of sunrise or sunset last for approximately 30 minutes each day, for which a corresponding 30 frame images are collected by each web camera system. Starting from the first frame, we select frames of minutes with an interval of six and observe algorithm performance in the whole procedure of luminance changes. Figures 7 and 8 give the comparison matrix, in which the first column shows input frames, and the second and third column show the backgrounds modelled using the SABS and proposed CBM-lw algorithms, respectively.

**Figure 7** Comparisons of SABS and proposed CBM-lw modelled backgrounds for a sunrise sequence of images, (left to right) current image, background from SABS and background from proposed CBM-lw

<table>
<thead>
<tr>
<th>T1</th>
<th>T7</th>
<th>T13</th>
<th>T19</th>
<th>T25</th>
<th>T31</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image1" alt="T1 Current Image" /></td>
<td><img src="image2" alt="T1 Background (SABS)" /></td>
<td><img src="image3" alt="T1 Background (CBM-lw)" /></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><img src="image4" alt="T7 Current Image" /></td>
<td><img src="image5" alt="T7 Background (SABS)" /></td>
<td><img src="image6" alt="T7 Background (CBM-lw)" /></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><img src="image7" alt="T13 Current Image" /></td>
<td><img src="image8" alt="T13 Background (SABS)" /></td>
<td><img src="image9" alt="T13 Background (CBM-lw)" /></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><img src="image10" alt="T19 Current Image" /></td>
<td><img src="image11" alt="T19 Background (SABS)" /></td>
<td><img src="image12" alt="T19 Background (CBM-lw)" /></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><img src="image13" alt="T25 Current Image" /></td>
<td><img src="image14" alt="T25 Background (SABS)" /></td>
<td><img src="image15" alt="T25 Background (CBM-lw)" /></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><img src="image16" alt="T31 Current Image" /></td>
<td><img src="image17" alt="T31 Background (SABS)" /></td>
<td><img src="image18" alt="T31 Background (CBM-lw)" /></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
As would be expected, levels of illumination increased rapidly in the sunrise sequence, being darkest in T1 to brightest in T31; whereas the reverse occurred at sunset. The background models generated by both the CBM-Iw and SABS approaches adapt to changes in levels of illumination at either end of the day, but the proposed CBM-Iw approach performs much more rapidly and accurately than the SABS approach. The background model luminance generated by the CBM-Iw approach closely matches that of the actual image at the time, but the background images generated by the SABS approach do not track the actual change in level of luminance seen in the first column. By T31 the background model generated by the proposed CBM-Iw approach is very different from that generated by the less accurate SABS method.
4.4 Robustness to changes in tidal height

Background modelling in coastal situations also needs to consider changes in tidal height. In this section we compare the performance of the proposed CBM-lw and existing SABS approaches at differing tidal states at three sites: Waitangi, Takapuna and Raglan. The results of these comparisons are shown in Figures 9–11, in which SABS background and the background from proposed CBM-lw are given in the middle and right column, and the ellipses in red colour highlight the shoreline area of each predicted background.

Figure 9  Comparisons of SABS and proposed CBM-lw modelled backgrounds at different tidal states at Waitangi, (left to right) current image, background from SABS and background from proposed CBM-lw (see online version for colours)

Figure 10  Comparisons of SABS and proposed CBM-lw modelled backgrounds at different tidal states at Takapuna, (left to right) current image, background from SABS and background from proposed CBM-lw (see online version for colours)
Since the moving objects presented in previous images have not yet been completely forgotten, the ghost cars and boats are evident in the land water boundary area of the SABS backgrounds and the shoreline is blurred. This suggests that the SABS approach has difficulty with modelling backgrounds in areas where the movement of water is highly variable, whereas the dual area CBM-lw approach can readily account for this variability and generates more reliable background image for each frame. Results of moving objects detection provided by the CBM-lw approach should therefore be more accurate.

4.5 Performance evaluation on moving object detection

To evaluate contribution of background modelling to moving object detection, we use the alternative backgrounds generated by algorithms when detecting moving objects appearing at all three boat ramps. We compare our method to four state-of-the-art methods. The first one is MOG (KaewTraKulPong and Bowden, 2002). The second one is GMG (Godbehere et al., 2012). Both methods are Gaussian mixture-based background models. The third one is SABS (Gupte et al., 2002) which is proposed CBM-lw based on. We also compare with IMBS (Bloisi et al., 2014) which is developed on purpose for water background modeling. MOG and GMG are implemented in OpenCV, the software of IMBS (available at http://www.dis.uniroma1.it/bloisi/software/imbs) is provided by the authors.

1 Qualitative results: Figures 12–14 illustrate the experimental results of the Waitangi, Takapuna and Reglan image series, where column (b) and (c) shows the frame image and the ground truth with respect to the frame, respectively. The selected frame image includes the case of no moving objects, single and multiple objects, and scene with different tidal states. The ground truth is manually labelled in advance for comparisons.
As seen in Figure 12(a), the water area (top left of each frame) of the Waitangi boat ramp changes over time with the rise and fall of the tide. In addition, the scene contains a big tall waving tree which increases the uncertainty of the background. The results of SABS and MOG are given in Figures 12(c) and 12(d), respectively. As shown in the 1,037th and 2,403th frames, SABS and MOG can still extract the foreground object on the ground. However, they are hard to overcome dynamic illumination changes in the area of water; objects in such area are mis-detected. Moreover, many minor false alarms can be observed in the area of waving tree as shown in both the 574th and 2,867th frames. Figure 12(f) presents the results of GMG, which contain fewer false alarms. However, this approach is still difficult to resolve the issue of dynamic water background, and it is easy to misidentify pixels with large variations of intensity values as shown in the 537th and 2,403th frames. The results of IMBS are demonstrated in Figure 12(e). IMBS performs relatively better foreground extraction than the aforementioned methods, since it is specifically designed for dealing with water background. However, in the 2,403th frame, many minor false alarms can be observed in the border area of land and water. Our results are displayed in Figure 12(g). As compared to the state-of-the-art methods, our method can obtain much clear foregrounds and fewer false alarms in the land as well as water area during illumination changes.

The Raglan and Takapuna image sequences also have similar results as the Waitangi sequence as shown in Figures 13 and 14, respectively. Again, compared with the state-of-the-art approaches, our method can extract the shapes of moving objects with much fewer false alarms in various scenarios as shown in Figures 13(g) and 14(g).

Figure 12 Foreground extraction results of the Waitangi sequence, (a) original frame (b) ground truth (c) SABS (Gupte et al., 2002) (d) MOG (KaewTraKulPong and Bowden, 2002) (e) IMBS (Bloisi et al., 2014) (f) GMG (Godbehere et al., 2012) (g) proposed method
Figure 13  Foreground extraction results of the Raglan sequence, (a) original frame (b) ground truth (c) SABS (Gupte et al., 2002) (d) MOG (KaewTraKulPong and Bowden, 2002) (e) IMBS (Bloisi et al., 2014) (f) GMG (Godbehere et al., 2012) (g) proposed method (see online version for colours)

Figure 14  Foreground extraction results of the Takapuna sequence, (a) original frame, (b) ground truth (c) SABS (Gupte et al., 2002) (d) MOG (KaewTraKulPong and Bowden, 2002) (e) IMBS (Bloisi et al., 2014) (f) GMG (Godbehere et al., 2012) (g) proposed method
Quantitative results: the overall relative performance of these background modelling approaches is also evaluated quantitatively on moving object detection. We use the alternative backgrounds generated by these five algorithms when detecting moving objects appearing at all three boat ramps, at different times of day, under differing weather conditions, and at different tidal states. For performance measurement, we calculate object detection accuracy as the ratio of the number of objects correctly detected by the algorithm against the number of objects from manual counting.

Table 2  Comparison of object detection accuracy on Waitangi boat ramp

<table>
<thead>
<tr>
<th></th>
<th>Detection accuracy (%)</th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>MOG (KaewTraKulPong and Bowden, 2002)</td>
<td>IMBS (Bloisi et al., 2014)</td>
<td>GMG (Godbehere et al., 2012)</td>
<td>SABS (Gupte et al., 2002)</td>
<td>Proposed</td>
</tr>
<tr>
<td>Time</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sunrise</td>
<td>62.81</td>
<td>74.18</td>
<td>67.32</td>
<td>65.48</td>
<td>91.62</td>
</tr>
<tr>
<td>Sunset</td>
<td>63.32</td>
<td>72.72</td>
<td>65.69</td>
<td>64.29</td>
<td>91.67</td>
</tr>
<tr>
<td>Daytime</td>
<td>71.15</td>
<td>83.14</td>
<td>76.32</td>
<td>75.01</td>
<td>91.78</td>
</tr>
<tr>
<td>Night</td>
<td>67.34</td>
<td>78.56</td>
<td>71.47</td>
<td>68.01</td>
<td>91.69</td>
</tr>
<tr>
<td>Weather</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rainy</td>
<td>63.56</td>
<td>75.79</td>
<td>68.55</td>
<td>66.17</td>
<td>91.58</td>
</tr>
<tr>
<td>Foggy</td>
<td>60.35</td>
<td>73.65</td>
<td>65.75</td>
<td>63.19</td>
<td>91.37</td>
</tr>
<tr>
<td>Windy</td>
<td>64.12</td>
<td>76.65</td>
<td>69.56</td>
<td>65.01</td>
<td>91.63</td>
</tr>
<tr>
<td>Sunny</td>
<td>70.92</td>
<td>83.78</td>
<td>76.92</td>
<td>74.82</td>
<td>92.01</td>
</tr>
<tr>
<td>Tide</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Low</td>
<td>62.33</td>
<td>72.45</td>
<td>67.56</td>
<td>66.48</td>
<td>91.39</td>
</tr>
<tr>
<td>Mid</td>
<td>66.56</td>
<td>77.67</td>
<td>72.76</td>
<td>71.29</td>
<td>91.61</td>
</tr>
<tr>
<td>High</td>
<td>71.87</td>
<td>82.99</td>
<td>76.65</td>
<td>75.21</td>
<td>91.92</td>
</tr>
</tbody>
</table>

Table 3  Comparison of object detection accuracy on Takapuna boat ramp

<table>
<thead>
<tr>
<th></th>
<th>Detection accuracy (%)</th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>MOG (KaewTraKulPong and Bowden, 2002)</td>
<td>IMBS (Bloisi et al., 2014)</td>
<td>GMG (Godbehere et al., 2012)</td>
<td>SABS (Gupte et al., 2002)</td>
<td>Proposed</td>
</tr>
<tr>
<td>Time</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sunrise</td>
<td>60.38</td>
<td>72.45</td>
<td>65.12</td>
<td>62.39</td>
<td>90.37</td>
</tr>
<tr>
<td>Sunset</td>
<td>62.12</td>
<td>71.39</td>
<td>66.34</td>
<td>67.46</td>
<td>90.52</td>
</tr>
<tr>
<td>Daytime</td>
<td>69.46</td>
<td>81.72</td>
<td>75.32</td>
<td>72.12</td>
<td>90.73</td>
</tr>
<tr>
<td>Night</td>
<td>65.43</td>
<td>76.76</td>
<td>70.24</td>
<td>65.34</td>
<td>90.69</td>
</tr>
<tr>
<td>Weather</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rainy</td>
<td>62.67</td>
<td>74.12</td>
<td>67.50</td>
<td>63.34</td>
<td>90.57</td>
</tr>
<tr>
<td>Foggy</td>
<td>60.15</td>
<td>71.34</td>
<td>64.32</td>
<td>60.32</td>
<td>90.21</td>
</tr>
<tr>
<td>Windy</td>
<td>65.34</td>
<td>77.78</td>
<td>69.67</td>
<td>62.34</td>
<td>90.45</td>
</tr>
<tr>
<td>Sunny</td>
<td>68.78</td>
<td>82.08</td>
<td>73.99</td>
<td>71.91</td>
<td>90.96</td>
</tr>
<tr>
<td>Tide</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Low</td>
<td>60.25</td>
<td>73.34</td>
<td>65.45</td>
<td>63.62</td>
<td>90.22</td>
</tr>
<tr>
<td>Mid</td>
<td>65.23</td>
<td>78.73</td>
<td>69.95</td>
<td>68.42</td>
<td>90.56</td>
</tr>
<tr>
<td>High</td>
<td>69.51</td>
<td>82.51</td>
<td>74.35</td>
<td>72.32</td>
<td>90.91</td>
</tr>
</tbody>
</table>
Table 4  Comparison of object detection accuracy on Raglan boat ramp

<table>
<thead>
<tr>
<th>Raglan</th>
<th>Detection accuracy (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>MOG (KaewTraKulPong and Bowden, 2002)</td>
</tr>
<tr>
<td>Time</td>
<td></td>
</tr>
<tr>
<td>Sunrise</td>
<td>66.06</td>
</tr>
<tr>
<td>Sunset</td>
<td>67.89</td>
</tr>
<tr>
<td>Daytime</td>
<td>73.15</td>
</tr>
<tr>
<td>Night</td>
<td>70.48</td>
</tr>
<tr>
<td>Weather</td>
<td></td>
</tr>
<tr>
<td>Rainy</td>
<td>68.79</td>
</tr>
<tr>
<td>Foggy</td>
<td>66.21</td>
</tr>
<tr>
<td>Windy</td>
<td>69.73</td>
</tr>
<tr>
<td>Sunny</td>
<td>73.73</td>
</tr>
<tr>
<td>Tide</td>
<td></td>
</tr>
<tr>
<td>Low</td>
<td>64.67</td>
</tr>
<tr>
<td>Mid</td>
<td>69.12</td>
</tr>
<tr>
<td>High</td>
<td>74.01</td>
</tr>
</tbody>
</table>

As the results, the object detection accuracy of all the approaches on the Waitangi, Takapuna and Raglan image series is presented in Tables 2, 3 and 4, respectively. Among the five methods in comparison, the proposed CBM-lw is seen giving consistently the best performance for all conditions, which indicates that our approach is able to overcome the extremely dynamic land and water composition scene problems, and capable of adapting to surveillance sequences captured from variant environments of different boat ramps.

5  Conclusions

In this paper, we propose a new cognitive background modelling algorithm (CBM-lw) intended to be used at maritime boat ramps where areas of both land and water are in frame. Background modelling in this context is especially challenging. The influence of sunrise and sunset is taken into account by CBM-lw approach which accounts for changes in outdoor luminance throughout the day. To improve the accuracy of background model, the CBM-lw algorithm can also be used to detect and reject corrupted images which may result in misleading backgrounds. In particular, our invention of cognitive water and land scene classification and use of dynamic learning rate and intelligent updating rules significantly increases the robustness of backgrounds generated by this approach. Some experimental tests and evaluations of its performance have been presented on a real 24-hour boat-flow analysis and counting system, where the proposed CBM-lw approach is compared with an existing method. These tests demonstrate that a much more stable background model is obtained by the CBM-lw algorithm in this context. The optimal learning rate $\alpha$ is mostly determined by the luminance threshold $\lambda$ which is used in equation (3) for objects detection. Future work will therefore focus on how to estimate the optimal value for this luminance threshold.
References


