1.AUTOMATED DROWNING DETECTION AND SECURITY IN SWIMMING POOL

Abstract - Every year, many individuals, including kids under the age of 5 drown in the deeps of the swimming pool, and the lifeguards are not well trained enough to handle these situations. Thus arises the requirement for having a system that will consequently detect the drowning individuals and alarm the life guard at such risk. Swimming pool surveillance systems plays an essential role in safeguarding the premises. In this project differential pressure approach is used for detection of drowning incidents in swimming pools at the earliest possible stage. The children's life is saved during drowning incidents in the swimming pool by lifting the acrylic plate. The proposed approach consists of RF module, Pressure Sensor and Motor Driver. The demo system based on pressure sensor has an advantage of convenience, cost saving and simple algorithm

1.INTRODUCTION

Swimming is a kid's favorite aquatic sport and it's a great stress buster. But in the water, beginners often feel hard to breathe which causes choking actions, loss of balance and results in a drowning accident. Some special circumstances, such as cramps, collide with each other, disease or mental stress and so on may also cause swimmer to drown. Drowning is a leading cause of death and disability for children. Worldwide, drowning produces a higher mortality rate than any other cause of injury in children less than 15 years of age [3]. Younger kids underneath the age of five are at precise threat, suffering the very best drowning mortality rates international. According to the Centers for Disease Control and Prevention, approximately one thousand children die from drowning annually in the world. In this project drowning accidents is avoided automatically by using the acrylic plate. The earliest swimming alarm system appears in the 1976, then there are some patent applications, but due to various reasons, these techniques are not popular[1]. In 2001, the French Vision IQ company produced the world's first set of drowning alarm system Poseidon; this is the first commercial promotion system. In 2003, Singapore Nan Yang, University of Technology design DEWS.

2. LITERATURE SURVEY

2.1 POSEIDON

Video based drowning detection system in the swimming pool Swimming pool drowning monitoring system based on video technology is mostly reported in the literature. There are three kinds drowning monitoring system according to the different position of the camera. One is that the camera is mounted on the underwater swimming pool wall, then monitor underwater swimmer status. A limitation of this equipment is that if too many swimmers, the occlusion problem arises. The other is that the camera is mounted upon the water, and monitors the Swimmer posture change. The reflection and refraction of light in air-water interference will affect the image quality, and drowning man feature this method detected is not easy to distinguish swimmers and divers obviously. The third is a combination of the two, underwater camera and aerial camera matched, monitoring the swimmer posture. This system needs constant observation which is the main disadvantage.

2.2 Wearable devices for early monitoring and alarming for drowning incidents

The wearable drowning monitor device can detect drowning accident and alarm. The device has seven main modules, including microprocessor, power module, SD memory card module, LED warning module, acceleration sensor module, water pressure sensor module, and keys module. When swimming the human arm must constantly waving in the water, if drowning, arm motion of floating is significantly reduced, and if falling into the water, almost motionless. According to the physiological response of human drowning, it can detect drowning accident by recording arm motion real-time through wearable wrist accelerometer device. This accelerometer is packed with embedded functions with edible user programmable options, configurable to two interrupt pins. The pressure sensor is installed to judge whether the human body is in the water. The red LED is used for drowning warning. One blue LED is used to get the work status of the device which wills flash every few seconds in order to save the precious energy. Because LED light emitting angle generally relatively small, 5 red LED lights of upward and around direction is installed to make LED alarm signal caller. Two keys are designed for the demo device One is the switch for power. The other is a self-help button. If drowning danger occurs, the swimmer can push the button and the blue LED will shine for help, and if a swimmer accidentally hit the button, he can push the button to cancel the alarm. If the swimmer lost consciousness because of drowning, the device detects the drowning accident and will ON LED light to inform the lifeguard.

The device is worn on the wrist and move in large amplitude along with the wrist when a human is swimming in the water, and the data acquired from accelerator will dramatically change. If a human is drowning in water, his or her wrist almost motionless, and the data acquired from accelerator will have only small changes due to water movement. The drowning detection method uses threshold. First, data from a water pressure sensor is used to judge whether the human body in the water, if the body in the water, then start drowning judgment process. Then, analog signal obtained from the three axis acceleration sensor is converted to digital signal and three axis acceleration values are gained. Hanning filtering method and the moving average filtering are used to reduce noise error.

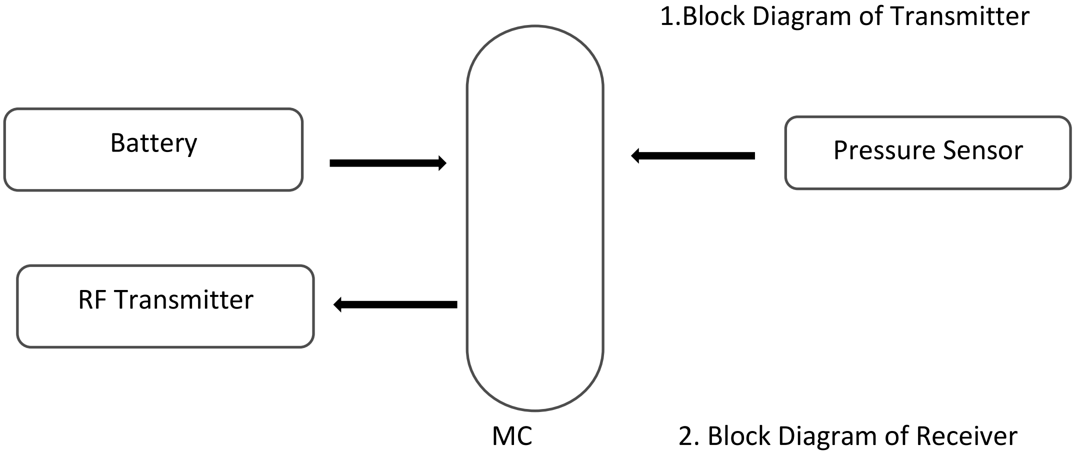
2.3 LDR based automated drowning detection system in the swimming pool

In the proposed method the human identification in the swimming pool depends on the LDR and laser. First, data from a water pressure sensor is used to judge whether the human body in the water, if the body in the water,then start downloading judgment process.The iron metal plate is placed in the floor of the swimming pool. The laser and the LDR source are placed in the side of the wall. Here we are using an ATmega8l microcontroller to control the whole process. Embedded c language is used for the coding. Initially the laser source which spreads over the swimming pool and the LDR which sense the laser light and which produces the resistance value. Depends on the resistance value the process has been taken.When the LDR value will be kept constant then the alarm will be activated.The resistance value will be changed with respect to the human movement. The message will be sent to the administration by using the GSM service. After 30 seconds there is no change which means the plate will lift automatically using the motor and motor driver. The human has safe in this technique.

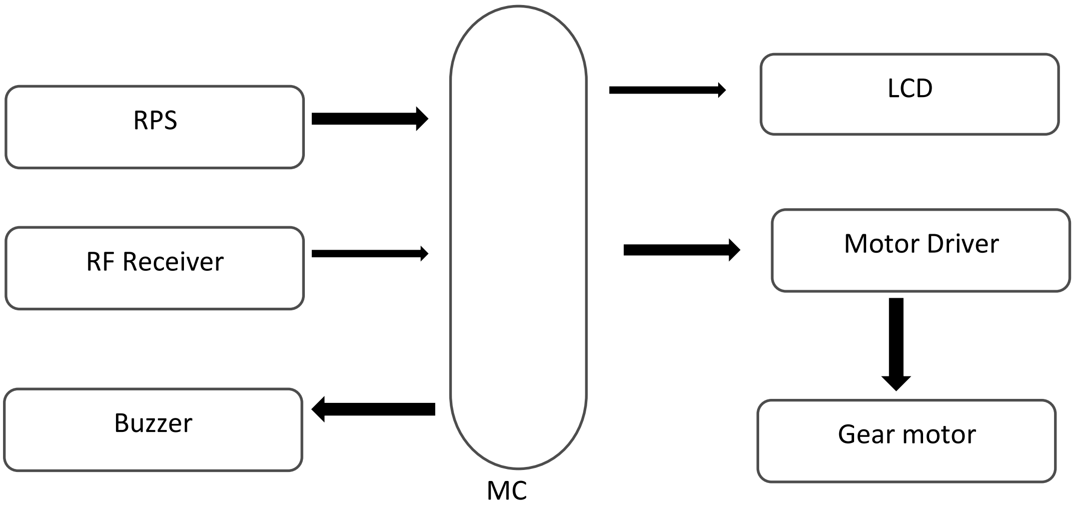
3.PROPOSED SYSTEM

The automated drowning detection system works on the principle of differential pressure. The system contains two fundamental modules: to begin with, the wristband consisting of pressure sensors on the transmitter side. Second, the receiver module at the swimming pool site. The children entering the pool territory should wear the wristband. The Pressure at underwater is different and greater than the pressure at the air - water interface. The pressure at a particular depth is measured and set as the threshold. Once the child gets into the pool, the pressure is continuously measured and monitored by the microcontroller. When the current value surpasses the threshold limit an alerting signal is sent to the receiver. The wireless transmission and reception of signals is done through RF module. On receiving the valid signal microcontroller sets the buzzer ON, turns ON the motor driver which in turn lifts the acrylic plate of the multi-floored swimming pool. The kid is brought to air-water interface, i.e. the top level of swimming pool by the acrylic plate

1.Block Diagram of Transmitter



2. Block Diagram of Receiver



3.1 Microcontroller

AVR microcontroller (Atmega32) is used here. The AVR microcontrollers depend on the propelled RISC design and comprise of 32 x 8-bit universally useful working registers. It is used to monitor the pressure values continuously. As the value exceeds the threshold limit an alerting signal is sent to receiver indicating drowning accident.

3.2 Pressure Sensor

The pressure sensor is used as an input here. The pressure sensor has 4 pins as shown above in figure. This system consists of a diaphragm with the crystal lattice circuit in it. The more the pressure, the more is the bending of the diaphragm and produces a corresponding voltage. This analog voltage is converted into digital and is inputted to the microcontroller.

3.3 RF module

The RF module operates at Radio Frequency. This RF module comprises of an RF Transmitter and an RF Receiver. The transmitter/receiver pair operates at a frequency of 434 MHz. An RF transmitter receives serial data and transmits it wirelessly through RF through its antenna. The transmitted signal is received by an RF receiver operating at the same frequency as that of the transmitter.The RF module is used along with a pair of encoder/decoder. The encoder is used for encoding parallel data for transmission feed while reception is decoded by a decoder. Here HT12E-HT12D is an encoder / decoder pair ICs used.

3.4 Motor Driver (L293D Dual H-Bridge)

IC L293D Dual H-Bridge is the motor driver used here. One IC can interface two DC motors which can be controlled in both clockwise and counter clockwise direction L293D has an output current of 600mA and peak output current of 1.2A per channel.The output supply has a wide range from 4.5V to 36V.

4. CONCLUSIONS

Consistently numerous people, including kids, are suffocated or near suffocating in the deeps of the swimming pools, and the lifeguards are not prepared all around to deal with these issues. In this manner raises the necessities for having a framework that will thus recognize the suffocating people and alert the lifeguards at such hazard. It can be installed in International standardized schools where classes are held for training kids.

2.AngelEye Maximum safety for public pools

Designed for whom has to guarantee every day the safety in public and intensive-use pools, AngelEye LifeGuard detects potential drownings and promptly notifies you. It features the latest artificial intelligence technology and adapts to the needs of the user. It's the ultimate drowning detection system for those who demand the ultimate in safety

Safety first

Active drowning detection.

AngelEye LifeGuard is a drowning detection system that detects every dangerous situation and accident. The AngelEye software works in close integration with the cameras installed in the pool to continuously scan the pool. Thanks to this combination of hardware, software and profound innovations, today AngelEye LifeGuard represents excellence in drowning detection

Position and image of the drowning.

When it comes to swimmers in trouble, every second counts. AngelEye LifeGuard makes itself heard loud and clear in case of danger. The built-in notification system produces alarms within 10 seconds on smartwatches, phones, flashing lights and other configurable devices. In addition, AngelEye's advanced technology can provide real-time location and image of the danger, making rescue operations easier.

Recording of events

The AngelEye LifeGuard system is able to record all the activities in the pools and to classify critical situations from normal ones in order to keep track of what happened. Thanks to its advanced image archiving system, AngelEye LifeGuard meets the legislative requirements for the protection of personal data.

An additional level of security.

The protection of swimmers is ensured to all facilities by a vigilance provided by personnel assigned to control the activities carried out in the pools. These controls have several critical points. The biggest problem is the difficulty in seeing the bottom of the pool. AngelEye LifeGuard is specifically designed to provide support to lifeguards in the supervision of swimmers. It offers an additional level of safety and integrates seamlessly into rescue operations.

ISO 20380:2017.

The standard that regulates drowning detection systems in non-domestic swimming pools larger than 150 m2 is defined by ISO (International Organization of Standardization) 20380:2017.

The standard describes the minimum operational performance and safety requirements as well as simulation methods for drowning detection systems. The aim is to ensure that in the end the customer invests in a high-performance system to ensure the safety of swimmers.

The hazard detection functionalities can be customized according to the specific needs of the system, allowing configurations analyzed and studied for every need and type of structure, whether it is a water parks, a sports facility or others.

AngelEye is able to guarantee the standard defined by ISO 20380:2017 and includes the functionalities for the management and archiving of the scheduled periodic tests.

3. An Automatic Video-Based Drowning Detection System for Swimming Pools using Active Contours

Abstract— Safety in swimming pools is a crucial issue. In this paper, a real time drowning detection method based on HSV color space analysis is presented which uses prior knowledge of the video sequences to set the best values for the color channels. Our method uses a HSV thresholding mechanism along with Contour detection to detect the region of interest in each frame of video sequences. The presented software can detect drowning person in indoor swimming pools and sends an alarm to the lifeguard rescues if the previously detected person is missing for a specific amount of time. The presented algorithm for this system is tested on several video sequences recorded in swimming pools in real conditions and the results are of high accuracy with a high capability of tracking individuals in real time. According to the evaluation results, the number of false alarms generated by the system is minimal and the maximum alarm delay reported by the system is 2.6 sec which can relatively be reliable compared to the acceptable time for rescue and resuscitation.

Video surveillance can be used as a tool for monitoring and security. Observing public and private sites has increasingly become a very sensitive issue. The visual monitoring capabilities can be employed in many different locations to help people live more safely. Videobased surveillance systems are designed and installed in places such as railway stations, airports, and even dangerous environments. Image processing, pattern recognition and machine-vision based methods are efficient ways for real-time intelligent monitoring of the objects or events of interest [1-4]. The existing surveillance systems deliver valued information in monitoring of large areas. Applying intelligence in video surveillance systems allows realtime monitoring of places, people and their activities. The tracking approach can change with varying targets and can change from a single camera to multiple camera configurations [1, 2 and 4]. Tracking methods in video surveillance use different parameters such as objects' motion, position, path of movement and velocity [4, 5], biometrics such as skin color or clothes color [4, 6] and many more. The tracking must be robust and overcome occlusion and noise which are common problems in monitoring [4-6]. One important environment that the need for monitoring systems is crucially sensed is the swimming pool. Each year many people including children are drowned or very close to drowning in the deeps of the swimming pools, and the life guards are not trained well enough to handle these problems [7]. This raises the need for having a system that will automatically detect the drowning person and alarm the lifeguards of such danger. Real-time detection of a drowning person in swimming pools is a challenging task that requires an accurate system. The challenge is due to the presence of water ripples, shadows and splashes and therefore detection needs to have high accuracy.

Related Work

Related Work In swimming pool monitoring intelligent systems, different approaches have been proposed. Most methods perform background processing on input video frames. Some apply background subtraction and image denoising to detect the drowning person . In a Gaussian Mixture Model is used for describing the pixels and the parameters of the model are updated with the EM algorithm. Also, neural networks can be trained toclassify near-drowning and normal swimming patterns . However this requires to have a large dataset of both groups of behavior. The dataset is obtained in by attaching a pressure sensor to a swimmer imitating drowning behavior and normal swimming. Pattern recognition algorithms are also very useful in swimmer detection. In a background model that has prior knowledge about swimming pools is employed. This hierarchical model operates on behavioral traits common in almost all troubled swimmers. Another vision-based system, depend on detection of swimmers' body parts. This approach uses local motion and intensity information from image frames . In the YCbCr color model is selected for detection of the water polo players in water where luminance is separated and the Cb and Cr components are analyzed. Moreover, underwater ultrasonic sensors can detect drowning people up to 70 meters below water in the swimming pool along with a underwater video detection unit that locates and finds the victims .

This research presents a vision-based approach for detecting a drowning person and alarming the life guards of such situations. In this study, the person swimming in the pool is detected and tracked using the HSV color space properties and contour-based methods. As soon as the moving target remains under water for more than a determined period of time, an alarm is sent to the lifeguard rescues. The HSV color space is selected over other color spaces because it is more effective in segmenting the swimmer in various light conditions from the background. The paper is organized as follows. In section II, the proposed method is presented in details. Experimental results, including discussions and reported performance results, are given in section III. Finally, conclusions are summarized in section IV.

II. PROPOSED METHOD

In this paper we have proposed a method for automatic real-time detection of a person drowning in the swimming pools. The overview of the proposed algorithm in this paper is presented in Fig. 1. Our system is based on real time video analysis of the cameras installed around the swimming pool in a way which the entire swimming pool can be covered. Each camera is mounted on pool walls oriented downwards with a sharp angle, so that it can minimize the effect of lightening system which causes occlusions and foreshadowing. In this work, a ODROID-XU as a distributed system is installed in the swimming pool to collect all the video signals collected from cameras and process them using computer vision methods. The used hardware including the distributing system known as ODROID-XU, and our Logitech HD Pro C920 webcam used to record all the video sequences in this paper is illustrated in Fig. 2. The system is used to firstly detect the background of the pool and then decide to send an alarm to rescue team if a previously detected person is missing in video frames for an specific and defined period of time. In the next sections of this paper, we try to explain the concepts we used to detect and track individuals in swimming pools.

A. HSV Color Space Analysis

There are a number of color spaces that are suitable to be used in the area of video tracking and surveillance. They include RGB [15], YCbCr [13], CIE Lab and HSV [16, 17]. Each one can be used in different applications. Since the illumination data is inserted into the three color channels of the RGB color space, normalization of the RGB color space would allow a more robust tracking in this color space. This data can then be transformed into a different color space to separate the brightness effect from color information.

In the HSV color space, there are different layers of information and the luminance data is separated from the color information. The separation of brightness information from the color information makes the HSV color space very suitable for tacking purposes [18-20]. In HSV, the V channel contains the luminance information of the input image, and the H and S channels have the chromaticity information in them. These properties make this color space very effective in segmentation of the target object which is the swimmer. In addition, employment of chrominance in the HSV color space can provide the system with robust tracking. Also, the separation of the brightness information from the chrominance decreases the effect of uneven illumination in an image. Considering light intensity, HSV color model is both scale-invariant and shift-invariant [20]. Due to the vulnerability of color-based tracking algorithms and fluctuation of light conditions, for the proposed system we apply the HSV color model to find the target object which is the swimmer in the input video and also distinguish the background of swimming pool from swimmers. Before starting the detection, a single frame of the input image is given to the system. This frame should be chosen so that it is a suitable sample candidate of all the frames; that is it should contain a person swimming in the swimming pool. This will make the system have a higher accuracy during the detection process. Receiving this single frame, the object of interest which here is the human body in the blue background of the swimming pool, will be manually extracted and marked. With this prior knowledge, the appropriate values for H, S and V channels can be set and tuned. So, once the image is captured through the cameras installed around the swimming pool, its pixel values is converted to the HSV color space. Then, the HSV image obtained from every single frame is converted to binary image by a simple thresholding over HSV values. This threshold is used calculated using the prior knowledge obtained from the initial step in detection phase. As a result, the binary image will be a black and white image in which background will turn black and the foreground (which is the swimmer) would be white.

B. Contour Detection

Contours can be used to find object outlines in images and effectively track targets in videos sequences. In tracking algorithms that are based on contours, the objects are tracked using their outlines as boundary contours. The contours should be updated dynamically in successive frames. In active contours concept, a closed contour is limited to the object's boundary. Hence the contour covers object region and object segmentation is reached. The contours are managed by their energy functions. This function consists of internal, external and shape energy [21, 22]. Active contour representations have been applied in different fields to track non-rigid objects [22, 23, 24]. An active contour representation is defined as in (1). (1) The parameters Rin and Rout are the regions inside and outside the contour C. The function d(x,y,C) returns the smallest Euclidean distance from point (x,y) to the contour C. Segmentation is a technique that segments an image frame into sections to discover the object of interest. In segmentation algorithms, it is very important to have an efficient partitioning method. Once a video frame is segmented, the object of interest is detected for tracking. In many indoor swimming pools, the background only consists of a number of features including the water and the lane drivers. When people are swimming in the pool, the swimmers are the only objects that are distinguishable from the background due to their motion and color. Therefore the first step is to achieve an unsupervised segmentation of the empty pool

After that the input frame is converted to a binary image, the contours in the binary image are found. Out of all the discovered contours, the one with the largest area is selected and tracked in consecutive frames. The resulting contour of the previous frame is taken as initialization in each frame. In Fig. 3 a sample of a given frame from recorded video sequences in swimming pool and the detected contours in this frame is illustrated. An object tracker is important because it can find the motion trajectory of the target object as video frames proceed through time. This is done by identifying the position of the object in every frame of the video. In this paper the tracking procedure is done by applying HSV thresholding algorithm in every single frame and then choosing the contour with largest area available in the result binary image. So, the area that is occupied by the target object is found by the algorithm at every instant and tracked in the subsequent frames.

III. EXPERIMENTAL RESULTS

The proposed system provides an alarm to the lifeguard rescues as soon as the tracking person is detected as being drowning. A visual indicator is used to determine whether the target being tracked in on the surface (green) or below the water (red). A red alarm along with a beep sound is generated when the swimmer is not found by the system for more than a specific number of consecutive frames regarding to the fact that the speed of different boards vary. In this research, we used an ODROID-XU board which contains Exynos5 Octa Cortex™-A15 1.6Ghz quad core and Cortex™-A7 quad core CPUs, and also a 2Gbyte LPDDR3 RAM. For video capturing purpose we used a Logitech HD Pro C920 webcam which is capable of recording in full 1080p at 30 frames per second. This hardware along with the developed algorithm to track swimming objects in pools can process about 6 frames every second. As a result, we can let the alarm go on if the swimming object is not fount after 30 consequent frames. It worth telling that we used the OpenCV library for the implementation of this software. To evaluate the performance of our system, different footages recorded in real swimming pools were used. We used 3 sequence of videos to evaluate the proposed method. Each sequence contains different number of frames and is taken from various views of the swimming to make the evaluation results more reliable. Table I shows the obtained results from 3 video sequences containing their frame counts along with their relative true and false alarms sent by the proposed algorithm in this paper. True alarms (True Positive) represent the situations in which one person is being drowned and the system should raise an alarm to notify the lifeguard in the swimming pool. Also the false alarms (False Positive) represent the conditions in which a drowning alarm has been reported by mistake. All these situations are considered as normal situations that their importance is ranked low compare to false negative ones. In sequence No. 2, we have a drowning condition which takes 20.1 seconds. The presented system was able to detect the drowning person and its position easily, though it reported a short period (1.4 sec.) of a true false situation as a true positive which can be easily overlooked. The sequences No. 1 and No. 3 contain 3 and 2 drowning condition, which the proposed method succeeded to detect them all. Also in these two sequences we had no false Alarms, and this fact can represent high performance and accuracy of our presented work. As could be seen in Table I, the number of false alarms generated by the system is minimal. Table II provides more performance evaluation of our system by depicting the Alarm delays regarding to the length of each drowning conditions occurred in each video sequence. The average detection delay for 3 video sequences is 1.53 seconds which shows high performance and accuracy of the proposed method in this application. Fig. 4 shows the results of applying HSV thresholding on frames' pixels in several frames of 3 video sequences. The result images are excluded from pool's background and are prepared for contour detection.

It shows drowning detection results for 3 video sequences in different conditions including frames in which the object is drowning and also frames in which the object is visible on the surface of the water. In both situations, we achieved the desired results which enable us to use the proposed system for high performance drowning detection in swimming pools.

IV. CONCLUSION

In this paper, we provided a method to robust human tracking and semantic event detection within the context of video surveillance system capable of automatically detecting drowning incidents in a swimming pool. In the current work, an effective background detection that incorporates prior knowledge using HSV color space and contour detection enables swimmers to be reliably detected and tracked despite the significant presence of water ripples. The system has been tested on several instances of simulated water conditions such as water reflection, lightening condition and false alarms. Our algorithm was able to detect all the drowning conditions along with the exact position of the drowning person in the swimming pool and had an average detection delay of 1.53 seconds, which is relatively low compared to the needed rescue time for a lifeguard operation. Our results show that the proposed method can be used as a reliable multimedia video-based surveillance system.

4.Computer Vision Enabled Drowning Detection System

Abstract -Safety is paramount in all swimming pools. The current systems expected to address the problem of ensuring safety at swimming pools have significant problems due to their technical aspects, such as underwater cameras and methodological aspects such as the need for human intervention in the rescue mission. The use of an automated visual-based monitoring system can help to reduce drownings and assure pool safety effectively. This study introduces a revolutionary technology that identifies drowning victims in a minimum amount of time and dispatches an automated drone to save them. Using convolutional neural network (CNN) models, it can detect a drowning person in three stages. Whenever such a situation like this is detected, the inflatable tube-mounted selfdriven drone will go on a rescue mission, sounding an alarm to inform the nearby lifeguards. The system also keeps an eye out for potentially dangerous actions that could result in drowning. This system's ability to save a drowning victim in under a minute has been demonstrated in prototype experiments' performance evaluations. Keywords—Drowning, Lifeguard system, Object detection, Computer vision, Pose estimation, Drone, Convolutional Neural Network (CNN)

INTRODUCTION

Drowning is the third most significant cause of accidental injury globally, according to the World Health Organization (WHO) [1]. Responsible for 7% of all injuryrelated fatalities, an estimated 320 000 people drown each year [2]. An average of 855 persons perished each year in Sri Lanka due to drowning per year, resulting in a drowning rate of 4.4 fatalities per 100,000 people [1]. Drowning may happen in various settings, including bathtubs, natural water bodies, and swimming pools. According to National Vital Statistics System (NVSS), swimming pools account for approximately 16 percent of all drowning deaths, implying a dangerous link between pool swimming and mortality [2], [3]. According to studies, lifeguards may not be adequately trained to deal with a drowning incident [2]. Whether it's due to a lack of training or a failure to spot a drowning victim quickly enough, the result of a life and death scenario may change instantaneously.

Apart from drowning, those who disobey pool laws and regulations cause discomfort to others while causing serious health problems for themselves with dangerous actions such as drinking and running around the pool. shows how lifeguards claim that lack of care is the most probable cause of drowning, among the other reasons that intoxication appears to be present .

Considering the possibility that attempts to save lives in the water using traditional methods fail, it is clear that an intelligent system is needed. The system described in this paper uses computer vision to detect and rescue drowning victims to ensure the safety of pools. The system uses computer vision along with automated electronic equipment to immediately rescue and protect the lives of swimmers in the pool. By integrating the camera above the water surface, it can recognize struggling motions before a fatality occurs. The camera's location captures a complete view of the facility, including swimmers, wanderers, and occupied objects. Swimmers are individually identified using object detection, noise cancellation and individually tracked using deep learning technologies to identify a possible drowning. On detection, the location coordinates of the drowning person are immediately calculated based on the ground coordinates (a grid system linked to x and y blocks) and sent to an autonomous custom drone while sounding an alarm, informing that more security measures should be taken. In addition, the detection of hazardous activities through computer vision concepts and posture detection in the facility's swimming pool ensures the safety and well-being of swimmers. This goal is achieved by using Firebase Cloud Messaging (FCM), the primary notification source to authorized personnel if a dangerous activity is detected. The structure of the article is as follows. Firstly, the literature survey discusses the currently available existing systems and technologies that have similar targets using both software and hardware-based technologies. Secondly, the methodology section explains the steps of how the system tried to solve this problem. Next, the results and discussion section analyze the main experimental results. The final section discusses the advantages and disadvantages of and possible work to improve the system in the future.

II. LITERATURE SURVEY

Vision-based systems and wearable sensor-based systems are two types of existing drowning detection technologies. Vision-based technologies are further subdivided into those that use underwater cameras and those that use above-water cameras. Underwater cameras have the drawback of missing the early struggle above the water. Early on, failure to recognize a drowning scene could result in a longer rescue time, which is a significant issue to consider in a time-critical emergency. The main disadvantage of a wearable-based system is the discomfort of use, which may lead to younger children seeking to alleviate the discomfort by removing the device, which is an unsubstantiated theory [15].

A. Object Detection Using Different Techniques

It is claimed that the usage of Convolutional Neural Network (CNN) architecture in Deep Neural Networks (DNNs) has added a significant shift in learning more complicated, informative characteristics in images as compared to the older techniques . Furthermore, further optimized models such as Fast R-CNN, Faster R-CNN, and YOLO have been constructed since the region-based convolutional neural network (R-CNN) architecture proposal. Fast R-CNN, which improves bounding box (BB) regression and classification Faster R-CNN , which generates area suggestions using an extra sub-network [18]; and YOLO, which detects objects using a fixed-grid regression , are all faster than R-CNN. Bounding box regression is used to recognize generic objects based on basic CNN architectures. Local contrast enhancement and pixellevel segmentation, on the other hand, are used to recognize salient objects . The techniques used in detecting objects under this chapter will be crucial as they establish the groundwork for the methodologies used to identify drowning and hazardous activities.

B. Drowning Detection And Tracking

To avoid drowning events utilizing an alert system, Alshbatat et al. proposed an integrated vision-based monitoring system consisting of a Raspberry Pi, two Pixy cameras, and an Arduino Nano board. They employed two cameras to detect and monitor swimmers by measuring their positions, and the swimmers were obliged to wear passive yellow vests. NEPTUNE , is another unique technology that uses statistical image processing of video sequences to detect drowning victims as soon as possible. The equations utilized in detecting near-drowning victims are based on the variables created by statistical image processing. Another system called VIBE uses background extraction to detect and track drowning victims and updates the motion area by taking the frame difference using the VIBE algorithm, which primarily evaluates the swimmers' positions when making judgments. How-Lung et al. examine some difficulties in spotting drowning victims in a watery environment and offer an automatic detection surveillance system. The key obstacles in the aquatic environment, according to the authors, are water ripples and splashes, as well as background movements of the reflective zones. When it comes to recognizing swimmers, occlusions are also mentioned as a challenging difficulty. Their proposed solution is an algorithm that takes into account all of these issues and detects water crises in complex aquatic environments.

C. Activity Detection Using Computer Vision

Current work on human motion prediction has been focused on two independent but complementary sub-tasks, according to Anand Gopalkrishnan . 1) Short-term motion prediction, which is quantitatively evaluated by measuring the mean squared error (MSE) over a short period, and 2) long-term motion prediction, qualitatively evaluated by visual inspections of samples over a long period. Shortterm models would be valuable in motion tracking applications because these jobs are applicable in several domains of work. On the other hand, long-term models might be valuable for creating computer graphic tools due to their broad applicability. Additionally, both models could be useful in human gait analysis, kinematics research, and human-computer interaction.

III. METHODOLOGY

The system explained in this paper includes three main functions: detecting drowning victims, sending drones to victims, and detecting dangerous activities. The drowning detection component detects drowning victims through a custom CNN model, which detects drowning in three stages and immediately informs the user through an audio alert. The second component is the rescue drone, activated according to the drowning detection command and sent to the victim's location coordinates. This procedure uses a custom configured x and y coordinate block system to link to ground GPS coordinates. At the same time, potentially dangerous activities, including running around the swimming pool and drinking, will be notified to authorized personnel in the premises through mobile alarms by the hazard detection component. This will prompt authorized personnel (including lifeguards) to make responsible decisions.

A. Drowning Detection And Tracking

1. Creation of the data set: Due to the lack of an existing aquatic human body parts data set, a data set containing 5000 images were constructed. All images in the dataset contain at least one or more swimmers in the water

• Image Collection: The primary source of data collection is the induction of actors and the collection of videos in real-time. The secondary source of data collection is the Internet, using specific keywords, such as "swimmer", "swimming", "drowning", "drowning in a swimming pool",

• Image Labelling: LabelImg – a graphical tool implemented in Python, is used to mark the image. Each image is labelled by creating arbitrary bounding boxes and predefined labels in YOLO format. The predefined tags used in image annotation are "Not_drowning", "Drowning_stage_1", "Drowning_stage_2", and "Drowning_stage_3".

2. Model Creation:

Use Google Colab to create and train models and get weight files every 100 iterations. The created model is then implemented on the NVIDIA Jetson Nano board, which runs on the Quad-core ARM CortexA57 processor. The main reason for using NVIDIA Jetson is to run multiple neural models parallel without complications and with a limited budget. First, swimmers in the pool are detected using an overhead camera and are kept track using the DeepSORT algorithm. YOLO is used to detect objects by locating one or more objects in the image and sorting each object. Yolo works well with a good resolution of entry compared to other models . Most of the problems in the detection and monitoring of swimming players are occlusal, scale changes, changes of appearance. These problems can be overcome using YOLO . The location of the tracked swimmer is also obtained using a predefined coordinate system.

Initially, the detection of swimmers is tracked while using a predefined coordinate system to obtain their position coordinates. At the same time, it will also check whether the swimmer has entered any drowning stage. If such an event is identified, the camera will track video clips in real-time to detect drowning victims. Using CNN, the human detection algorithm will use the image frames to identify the drowning person. Once the location block of the drowning victim is correctly identified, an audible alarm will sound to notify authorized personnel of the event. The images in the data set must include at least one drowning person to identify the chart as a drowning situation. A person's posture and movement can quickly identify a drowning victim. Although it is easy to recognize, one of the most common exercises is the "vertical ladder" exercise, which imitates the movement of a person climbing a ladder in a vertical movement.

Class	Features
Drowning_stage_1	Head is above the water
Drowning_stage_2	Half of the head is underwater, and hand gestures are in the climbing ladder motion
Drowning_stage_3	Head is underwater, and hand gestures are in the climbing ladder
Not_drowning	Regular swimming and floating motions

Table I describes the four categories used to classify swimmers: Drowning_stage_1, Drowning_stage_2, Drowning_stage_3, and Not_drowning. In the event of drowning, a frame pattern of stage 1 to stage 2 and then to stage 3 can be seen. Fig.4 consists of three images, each 743 pixels and 243 pixels in size. According to the climbing ladder motion, people who switched between stages 2 and 3 were identified as drowning victims. Finally, the location block of the drowning victim is passed to the drone.

C.Identification Of Hazardous Activities

In addition, the system can also analyze all visible activities in the pool to ensure the swimmer's safety. Continuous monitoring ensures that dangerous activities are not carried out on the premises. This is accomplished by notifying authorized personnel via mobile alarms when a dangerous activity occurs. The process of identifying hazardous activities at a location is initiated by collecting hazardous and nonhazardous activity data sets. The data set is collected in a pool that contains people hanging out in an environment. The identification process after the alarm notification can be divided into four steps.

1. Masking and noise extraction: Due to the camera arrangement, the noise obtained from the water surface at the site is concentrated in the water waves during the day. As shown in Fig.5, which is of size 1280 x 720 pixels, a fundamental step of masking the image is performed to remove the image's noisy areas. Due to the non-stationary nature of the camera, all frames have static X and y coordinates as masking points.

2. Skeleton sketching: Skeleton sketching is done using OpenPose - a real-time multi-person keypoint detection library , for pose estimation, which is then used to draw skeleton-based human figures to recognize a person's pose in real-time . The poses recognized by pose estimation uses a combination of DNN models to successfully approximate a complex nonlinear mapping function from a random image of a person to match the position, as shown in Fig.6. Each of the images is of size 1280 x 720 pixels due to the static nature of the camera placement.

3. Labelling images using CNN: Using transfer learning, a YOLOv3 model was improved to create a custom model. The sketched dark backgrounded images identified through the skeleton sketching process are integrated into the customized CNN model, identifying the categories (hazardous and non-hazardous activities). The training is carried out in two batches at the Learning Rate of 0.001, and the data set maintains a set of 1000 images for each class labelled for training.

4. Notifying the authorized personnel using mobile alerts: A status alert is sent to one of the authorized personnel devices if a frame is classified as hazardous. This, in turn, will notify the regarded personnel to initiate necessary precautions and measures. The notification signal would be in the form of a Firebase push notification, as it is swift and easily comprehensive. The Android service is compatible with any Android mobile running version from 4.4 (KitKat) and above.

IV. RESULTS AND DISCUSSION

A. Drowning Detection and Tracking Results

The YOLO detection algorithm [19] uses 416 X 416 as its input dimensions. The drowning victims are detected in three stages using a YOLO-based detection technique. Even though the swimmer stayed underwater for an extended period, the DeepSORT algorithm [33] could keep track of them.

The mentioned Fig.7 depicts the model's performance (False Positives and True Positives only) with 500 images.

Accuracy = (TP + TN) / (TP + TN + FP + FN)

TP - True Positives

FP - False Positives

TN - True Negatives

FN - False Negatives

Accuracy Variables	Count
TP	220
TN	208
FP	42
FN	30
Total Accuracy	85.6%

B. Hazardous Activities Because of the noise elimination via picture masking, the posture estimate accuracy was greatly improved. To allow the pose estimation algorithm to make more radical judgments, the default threshold value for the OpenPose body parts heat map was changed from 0.2 to 0.1. Although frame-by-frame identifications were only identified with a probability of 53% due to the threshold adjustment, the total system, which examined a frame in real-time, was able to identify a hazardous activity with much greater ease within 60 seconds, with a mean accuracy of 91.4 percent, after the threshold was changed.

A close examination of the misclassified postures among the testing sets revealed that a posture was more prone to misclassification as it approached the far end of the camera, indicating the need for a secondary camera to improve accuracy and confirm the true positives from the primary camera, as shown in Table II. Although employing a higher quality camera to fix this problem is a good idea, the requisite hardware and the near-real-time CNN techniques used to detect further objects may not be up to standard at present.

CONCLUSION AND FUTURE WORK

This computer vision-enabled automated dronebased lifeguard system consists of three main components, i.e., the drowning detection, the rescuing drone, and the hazardous activity detection. All three components combined will create a system capable of detecting drowning victims, dispatching an inflatable tube using a drone (as depicted in Fig.9) and detecting hazardous activities—eventually becoming an entity that could assist a lifeguard. The system is accessible to its primary user, presumably a pool owner or a lifeguard, in the form of an interface with a sound alarm and an android mobile service that holds the capabilities of receiving Firebase notifications.

Confined with a few of the hardware limitations, such as the use of a single camera and the Jetson Nano at the presence of better-quality hardware, could affect the speed and accuracy of the overall system is becoming a state-of-theart. This limitation could be omitted with the use of multiple cameras that could be placed over the premises in several ground coordinates, increasing the accuracy of the computer vision algorithms. Moreover, due to the inability to fly a drone in extreme weather conditions such as rain, strong winds or lightning, the system is limited to be used under few specifications. As swimming in extreme weather conditions is not preferred either, the system could be further improved to emit a warning signal if a person was to swim in any of the above weather conditions, bypassing the need to fly the drone.

Additionally, all the processing is done on the clientside of the applications on the Jetson Nano board, preventing any security and privacy issues that might arise due to the sensitive information inputted through the cameras. For future developments convenience wise, the system could benefit by having an additional set of cameras to identify and verify a drowning or a hazardous activity on the premises. Accessibility could also be improved by extending the Android service to be an application both in Android and iOS platforms that could hold the details of each premise individually, making a centralized system that watches over the decentralized pool premises. Both drown and hazardous activity detection could be improved by gathering a nighttime dataset that increases the accuracy of the data in low light.

5 . Visual search for drowning swimmers: Investigating the impact of lifeguarding experience

INTRODUCTION

In many applied domains, experienced participants often out-perform novices in relevant visual search tasks. Examples include diverse domains such as driving, sports and radiology (Chapman & Underwood, 1998; Nodine, Mello-Thoms, Kundel, & Weinstein, 2002; Spitz, Put, Wagemans, Williams, & Helsen, 2016). The current work extends this research to lifeguarding, assessing visual search for a drowning swimmer in a swimming pool. Below, we will explore general theories of how experience influences visual search and scene comprehension, before applying these to the lifeguarding context.

Experience influences visual scene processing

The ability to process a visual scene improves with experience via repeated exposure to the same environments and task demands (Stainer, Scott-Brown, & Tatler, 2013; Torralba, Oliva, Castelhano, & Henderson, 2006; Wolfe, Vo, Evans, & Greene, 2012). The following sections will consider some of the experiential effects upon scene processing that are pertinent to lifeguarding.

Contextual knowledge within a scene In real-world scenes

In real -world scenecs, the environment constrains the logical locations where targets appear (Eckstein, 2011). For instance, a “dog” target is likely to be found on the ground, though a cat might equally be found in a tree. The knowledge that cats climb trees, but dogs do not, is gained through experience of these animals. The contextual guidance model (Torralba et al., 2006) describes how such experience combines with bottom-up saliency calculations, creating scene priors to aid prioritization of bottom-up features (i.e., the tree, when searching for a cat).

The scene prior is applied to the visual scene based on a rapid processing of global image features (gist processing). Once the gist of a scene has been processed (typically within 100 ms; Luck et al., 1994; Potter, 1976), the observer can apply contextual knowledge to guide attention to target-relevant areas of the visual scene. Irrelevant, yet salient features of the scene, are de-emphasized if they fall outside of the prior, while salient features within the favoured locations are enhanced. This creates a scene-modulated saliency map which focuses attention to regions of the search scene where the target is expected to be (the cat in the tree, a car on the road, a bird in the sky, etc.). While some scene modulations may appear to require very generalized experience (cats climb trees, birds fly), more finegrained modulations may require increasingly specific scene expertise (e.g., bees like lavender, certain tumours may only be found in certain parts of the body).

While contextual knowledge for target feature guides the searchers' attention to different locations in a search display, observers of real-world scenes often have an incomplete knowledge of all target's features (e.g., location, colour, size). This negatively impacts search performance, particularly when imperfect target templates contain inaccurate information or extraneous features to target previews (Hout & Goldinger, 2015).

Visual processing within a scene

Experience with particular targets lowers the thresholds for their subsequent identification, allowing faster acceptance of these targets in the future, and fewer responses to non-target items (Borowsky & Oron-Gilad, 2013; Randel, Pugh, & Reed, 1996). For example, in category learning, merely learning which features are more diagnostic of category membership increases the speed at which those features are processed (Guest & Lamberts, 2010). Experience in certain domains helps improve visual processing of items in scenes, with shorter fixations and scanning time in people who have a level of experience compared to novices. Konstantopoulos, Chapman, and Crundall (2010) found that driving instructors appeared to have shorter processing times, with shorter fixations distributed across a wider area of the driving display, and broader scanning of the road compared to learner drivers. It appears that, with more experience in driving, overt attention can be moved more quickly, and less processing time is needed.

Situation awareness

Situation awareness refers to the ability to perceive the relevant objects within a scene, comprehend their relationship to one another and predict how the scene will develop (Endsley, 2015). Situation awareness is influenced by domain experience, as viewers may be more aware of the probabilities that certain visual cues may lead to specific outcomes (Endlsey, 1995; Kass, Cole, & Stanny, 2007).

Good situation awareness may improve the searcher's capacity for acquiring information about the events happening around them, particularly in dynamic, real-world search situations such as driving. Situational awareness develops as an individual gains experience of certain domains and any key stimuli that may require attention. These experiences allow the searcher to develop a catalogue of events that are likely to occur in similar situations, allowing viewers to prioritize areas of the scene based on what might happen next (Crundall, 2016). This prioritization could be considered to provide prediction priors. Adopting the terminology of Torralba et al. (2006), such prediction priors could act as a higher-level form of scene selection that is extrapolated following attention to objects within the scene priors.

Dynamic scenes There have been numerous applied visual search studies that aim to assess an individual's search skills and processing speeds (Godwin, Menneer, Cave, Thaibsyah, & Donnelly, 2015; Henderson, Brockmole, Castelhano, & Mack, 2007; Meuter & Lacherez, 2016). Visual stimuli in these applied real-world scenes have often been restricted to static images (although see Kunar & Watson, 2011). However, in real-world search tasks, visual scenes are typically not static, with items in the visual field moving, such as searching through a crowd in surveillance tasks. These types of dynamic searches are more complex and have an additional level of difficulty, with moving targets becoming occluded or undergoing changes in appearance or behaviour over time.

There is evidence to suggest that individuals with certain domain experience will perform better or more effectively in these complex tasks. For instance, Howard, Troscianko, and Gilchrist (2010) found that people with expertise in watching soccer were more likely than non-experts to be looking at task-relevant locations of a videotaped match while monitoring the game for upcoming goals (where “task relevant” locations were actually an emergent property of the eye data analysis). Experts using contextual knowledge to guide search of dynamic scenes to task relevant areas has been also suggested in research of CCTV operators. Howard, Troscianko, Gilchrist, Behera, and Hogg (2013) found that, compared to novices, expert CCTV operators showed greater consistency in eye positions and greater consistency in judgements of suspicious behaviours while monitoring CCTV footage. The authors suggested that the consistency between the trained CCTV operators was a result of their specialized knowledge, and thus knowing what to look for (see also Crundall & EyreJackson, 2017).

Dynamic scenes have also been used to explore visual processing in sport-related domains (see Kredel, Vater, Klostermann, & Hossner, 2017, for a review). For example, Martell and Vickers (2004) explored elite and non-elite gaze strategy in a live defensive zone task for ice hockey. The results showed that the athletes use two different strategies to temporarily regulate their gaze. Furthermore, elite athletes fixated tactical locations more rapidly than non-elites in successful play.

An introduction to the lifeguarding context

The various sub-processes considered above have been studied in a variety of applied settings including driving, airport security, and radiology (Biggs & Mitroff, 2014; Crundall, 2016; Nodine et al., 2002). One under-researched area of application however is that of lifeguarding. Lifeguards have an important, but extremely difficult job of supervising swimmers in a pool or beach setting. This includes searching for any swimmers that may be experiencing distress or drowning in the water. Explicit practical training in visual search of a pool is not currently part of lifeguard training in the UK, though search techniques are discussed with trainees (e.g., how to monitor a particular “zone”). Beyond problems with limited training, the swimming environment makes scanning difficult due to factors such as heat, long periods on duty and a large overlap in drowning and swimming characteristics (Griffiths & Griffiths, 2013; Lanagan-Leitzel, Skow, & Moore, 2015). While drowning in lifeguarded pools within the UK is incredibly rare, there are instances where supervision fails, resulting in injury or death. To prevent these fatal incidents, UK lifeguards are trained to recognize certain behaviours that are associated with drowning and distress.

A common form of drowning behaviour is termed active drowning, where targets typically display the Instinctive Drowning Response (Pia, 1974). These swimmers will usually be vertical in the water, with their arms flailing and splashing the water. The head will typically submerge and re-emerge and will be usually tossed back as the swimmer gasps for air. They will have no forward propulsion through the water and are unlikely to respond to shouted instructions. This struggle will last for as long as the person's energy permits, however research suggests a 60 s struggle is typical before energy is fully depleted. In nonswimmers and children this struggle may only last for 20 s (Pia, 1974). Swimmers in a crisis stage of drowning will not be able to call for help, as breathing takes precedence (Doyle & Webber, 2007).

In contrast to active drownings, passive drownings refer to those swimmers who have lost consciousness in the water, usually from some form of medical emergency. The transition from normal swimming to unconsciousness can happen quickly and the victim will either slip slowly under the water or remain face down and motionless on the surface (Fenner, Leahy, Buhk, & Dawes, 1999). Once at the bottom of a pool, the swimmer may be left unattended for a prolonged period with greater risk of permanent brain damage. In a study reported by Brener and Oostman (2002) lifeguards' responses to detect a submerged manikin were recorded. The manikin was introduced in a live pool setting, unbeknownst to the lifeguards. On average it took successful lifeguards 1 min and 14 s to notice the manikin, with only 9% of lifeguards detecting the manikin within 10 s (which is the target time as taught by the “10:20 system” which gives lifeguards 10 s to a spot and 20 s to respond; Ellis and Associations, 2007). The response to the submerged manikin was highly variable, with 14% failing to spot the manikin before 3 min had elapsed.

Although lifeguards are taught to recognize characteristics of drowning and distress, these behaviours are not always indications that a swimmer is in trouble. For example, splashing or submersion on their own are also common in swimming play behaviours, or even in swimmers with a weak technique. The lifeguard needs to be flexible in appraising these behaviours. Lifeguards also need to be aware of behaviours that could lead to drowning and distress, such as dangerous behaviour, poor swimmers entering deep areas of water, or otherwise vulnerable swimmers. These complexities have led to lifeguards differing in opinion in regard to which behaviours and events are critical. Lanagan-Leitzel (2012) found that when watching video footage of people swimming at a leisure facility, lifeguards disagreed on the events that should be rated as critical and in need of monitoring. Trainers reported nearly double the number of critical events compared to the lifeguards, however there was limited consistency in the different events reported. The events that were reported by the majority of lifeguards and trainers were also reported by a large number of non-lifeguard participants, suggesting that salience of many critical events was more important for detection than expertise.

Lifeguard expertise and visual search

Of the limited literature on visual search in lifeguards, experience has been shown to have a positive impact upon their search skills, leading to a greater frequency of critical events being identified. For instance, Lanagan-Lietzel and Moore (2010) found that lifeguards and participants who had received short instruction in drowning detection performed better in the detection of critical events than naive participants with no knowledge of drowning behaviours. They found lifeguards only identified 54% of critical events, while those participants who were given brief training identified 45%. Eye movement data suggested that lifeguards had superior search in terms of their fixations, with shorter fixations, fewer fixations of the water and more fixations to the critical events.

In a similar study using schematic animations of swimmers, Page, Bates, Long, Dawes, and Tipton (2011) found that lifeguards were significantly more accurate in detecting swimmers who disappeared under the water. Even though the provision of contextual knowledge improved performance (e.g., the location of a rip current in the simulated scene; in effect, a domain-specific scene prior) performance remained poor in absolute terms for both experienced lifeguards (31.6%) and novices (16.7%). Page et al. (2011) found only 12 out of 69 lifeguards looked in the correct area of the beach scene in unbiased conditions (when no contextual information was given), with only 7 of those 12 detecting the drowning victim, possibly arguing for a Look But Fail To See Error (LBFTS; Crundall, Crundall, Clarke, & Shahar, 2012). This error has been well researched in the applied domain of driving (Clabaux et al., 2012; Herslund & Jørgensen, 2004; Underwood, Humphrey, & van Loon, 2011). Commonly, drivers who report a LBFTS error look directly at the other road user but see them when it is too late or not at all. It has been noted that when an individual experiences a LBFTS error their cognitive resources are often engaged in another task, such as approaching a junction (Casner & Schooler, 2015). It is also possible that a person's expectations on what they will see will bias their processing. For example, when a driver approaches a junction, they believe that looking down the road will either reveal a car or an empty road. However, in the absence of a car, the driver is may be tempted to make a quick decision that the road is empty before they realize that they are actually looking at an unexpected oncoming motorcycle (Crundall et al., 2012).

The findings of Page et al. (2011) either suggest that a lot more could be done to improve the detection of drowning victims in aquatic environments, or alternatively, that we need more realistic tests that better capture lifeguards' skills. While Lanagan-Lietzel and Moore (2010) used naturalistic video, events were uncontrolled and possibly confounded many factors. Unfortunately, the highly controlled yet highly artificial stimuli used by Page et al. (2011) may have swung too far in the other direction, with the simplistic animated footage making generalizations back to pool or beach environments difficult (Page et al., 2011).

In a recent study, Laxton and Crundall (2018) aimed to bridge the gap between the previous two studies by using dynamic, naturalistic stimuli with experimental control. This study employed video clips of regimented swimming across the width of the pool (Figure 1). While distractor swimmers engaged in regimented lap swimming, some naturalistic behaviours were also captured, including pauses to alter goggles, chatting with others in the pool and underwater swimming. The results showed that lifeguards were superior in both accuracy and speed of responses to a mock drowning incident. In addition, fewer active drownings were missed compared to passive drownings, but response times to active drownings were slower. While the results were promising in terms of identifying superior lifeguard performance the research was not without limitations, as any false alarm responses (made before drowning onset) ended the trial prematurely, potentially systematically reducing performance of participants with low thresholds for reporting events. Non-lifeguard participants were overrepresented in their premature responses (17% vs. 7% for nonlifeguards and lifeguards, respectively), raising the possibility that, if given the opportunity to see the full trial, the non-lifeguards may have performed similarly to the lifeguard participants in detecting actual drowning targets.

Despite shortcomings in all the above studies, they consistently demonstrated a superiority in some aspects of lifeguards' performance. The current studies aim to build upon these findings, replicating the superiority effect and increasing our understanding of it, using a more exacting design. The first experiment explores differences in eye-movements and behavioural responses between non-lifeguards and lifeguards using a modified version of the drowning detection task used by Laxton and Crundall (2018). By removing the possibility of a false alarm terminating the clip early, it is possible that non-lifeguards' performance on this task will improve above performances noted in Laxton and Crundall (2018). Even though this will create a fairer comparison between groups, we still predict that the lifeguard superiority effect will be present. We further predict that eyemovement data will provide additional insight into the mechanisms underlying this superiority (e.g., faster first fixations to the target and greater probability of fixating the target).

The second experiment manipulated the instructions, with half of the participants given explicit guidance on behavioural characteristics of active and passive drownings. This manipulation was intended to assess whether non-lifeguards' poor performance was affected by a lack of expectations regarding drowning characteristics. A further improvement to the design of the second experiment was to include a wider range of participant expertise, by assessing four distinct groups whom we predicted to show increasing levels of superiority benefit: non-lifeguards, lifesavers, lifeguards and lifeguard trainers (described further in experiment 2). A final innovation for was to collect localized responses via a touch screen, adding a dimension of spatial accuracy that was missing from Experiment 1.

EXPERIMENT 1

The first experiment manipulated the number of swimmers in the search array (3, 6, or 9 swimmers) and the type of drowning (active, passive and a no-drowning control condition) across 45 staged video clips. Lifeguard and non-lifeguard performance was measured in terms of response times and accuracy to drowning targets, and their eye movements were recorded in an effort to detect visual search differences across the two groups. We predicted that lifeguards would perform better than non-lifeguards, and that this would be reflected in their eye movements. This superiority should be exacerbated in the hardest conditions (i.e., the largest set size, which will be more akin to the typical number of swimmers they might have experience of observing, and when the target is a passive drowning victim), though it was also a possibility that the hardest conditions may be so difficult as to cause a floor effect across all participants, nullifying the group differences that occur in the easier conditions.

METHOD

Participants

Forty-two participants were recruited to take part in a visual search study (with a mean age of 24.01, SD = 6.07, 22 female). Twenty-one of these participants (mean age 21.14, SD = 4.27, 23–47 age range, 11 females) had completed the UK National Pool Lifeguard qualification prior to testing and had a varying amount of experience in poolside lifeguard duties (2.46 years of lifeguarding experience on average, SD = 3.25). There were 5 lifeguards who worked on a full time basis (30–40 hours a week), a further 8 worked on a weekly part-time basis (between 5 and 20 hours a week), and 4 lifeguards worked less than 10 hours over a month. There were an additional 4 lifeguards that only worked during school holidays (both full and part time hours).

The remaining 21 participants (mean age 27.97, SD = 5.87, 16–31 age range, 11 females) had no lifeguarding experience. Lifeguards were recruited from a local leisure center and non-lifeguard participants were an opportunistic sample.

Design

A 2 × 2 × 3 mixed design was employed, comparing experience (lifeguards to non-lifeguard participants), drowning type (15 active drowning trials and 15 passive drowning trials) and set size of the search array (with 3, 6, or 9 swimmers). In addition to the active and passive drowning targets, 15 non-drowning trials were also included. Of the 15 trials for each of the drowning and control stimuli sets, five trials contained three swimmers, five trials contained six swimmers and five trials contained nine swimmers. During presentation to participants, all trials were randomized within a single block. All participants viewed all trials. Accuracy and response times to detect the drowning target were recorded. A response was considered accurate if the participant pressed the button during following the onset of the target until the end of the clip (i.e., when the swimmer began to drown). Response times were measured from target onset. To overcome the problem with premature responses being recorded as incorrect in a previous experiment (Laxton & Crundall, 2018) participants in this experiment could make multiple responses. However, if participants made a premature response that was not followed by a correct response in the target time window this was coded as an incorrect false alarm. Alternatively, if no response was made during a clip this was also coded as incorrect (a “miss”). Participants' eye movements in each trial were also recorded.

Apparatus and stimuli

The stimuli were the same as those used by Laxton & Crundall, 2018 (Figure 1), and were recorded on a Samsung Galaxy EK-GC 100, 23 mm handheld digital camera, with a field of view of approximately 70. The videos were presented on a Dell computer screen connected to an SMI RED500 eye tracker sampling at 500 Hz. Participants were tracked from an ideal distance of 60 cm from a display screen measuring 49 cm × 29.5 cm (44 × 28 of visual angle), with a resolution of 1600 × 900. In total there were 45 video-clips, and these were presented in colour with either 3, 6 or 9 swimmers traversing the width of the pool. The choice of filming swimmers crossing the width, rather than swimming the length of the pool, was made due to restrictions in the visible angle available to the camera. Nonetheless, the camera position reflects an operational standing lifeguard position employed during unstructured swimming sessions. In twothirds of the videos a staged active or passive drowning would occur, and the other one-third were catch trials with no drowning event.

The clips lasted 29 s on average (SD = 1.5 s) and were presented without sound. The drowning incidents lasted an average of 11.9 s (SD = 2.9 s) with clips ending immediately following the drowning. This should have allowed all lifeguards sufficient time to spot the drowning victim if following the 10:20 method (Ellis and Associates, 2007). Both types of drownings happened quasi-randomly within the second half of an average length video clip.

Volunteer actors were recruited to display both active and passive drowning behaviours and the silent presentation of the final clips allowed the cameraperson to use verbal cues and a whistle during filming to direct the action. Background swimmers (distractor swimmers) were instructed to engage in lap swimming, however, were free to choose their own style of swimming, which varied in pace and whether it was done above or below the surface. Swimmers were also permitted to take pauses for natural behaviours, such as taking a rest, talking with others, or altering goggles/swim hats, although these behaviours typically occurred at the sides of the pool. Importantly the freedom of choice about swimming behaviour for distractor swimmers meant a range of behaviours was in play at the time at which a drowning occurred (in the drowning clips). Table 1 shows the range of behaviours that occurred at the point of drowning onset. In all set sizes, the most frequent behaviour was continuous swimming; however, there were enough other behaviours present to ensure that clips contained novelty of action.

The trials were presented in a single randomized block. Each clip was preceded by a gaze-sensitive fixation cross that would only allow the trial to start if the participant was fixating the cross for a minimum of 500 ms. This ensured that calibration was retained through-out the study. If problems occurred, it was possible to recalibrate participants.

Procedure

Testing sessions were arranged at various pools and leisure centers, to better recruit lifeguards, with a quiet office or side-room acting as the laboratory. Non-lifeguard participants were tested under similar conditions on University premises. Participants were given written instructions and asked to fill in a consent form and demographic questionnaire. Prior to the study, participants were made aware that they would be searching for any potentially drowning victims from a lifeguard's perspective, and that the study would contain active, passive and non-drowning trials. Definitions of the drowning types were also provided (no such descriptions were provided in the Laxton & Crundall, 2018, study). Participants were told they could make multiple responses, if they thought that a later potential drowning event superseded a potential event that they had already responded to in the same clip. They were however encouraged to only respond once to any drowning incidents they observed and were told that a maximum of one simulated drowning event would occur per clip to reduce the number of premature responses. If a drowning was identified, participants were told to press the zero key on the number pad of a standard keyboard. Once all instructions had been given, participants had the opportunity to complete a practice trial, which was followed by a final opportunity to ask any remaining questions before the study began. Participants' eye movements were calibrated at a distance of 60 cm without head restraint using an 8-point calibration, followed by a similar 8-point validation test. If the validation procedure recorded fixations that deviated more than 0.5 from a validation target, a recalibration was undertaken. Upon finishing the test, the participants were fully debriefed and thanked for their time and participation. The total testing time took approximately 30 mins.

Data analysis First the data were scanned for influential outliers above or below 3SD from the means of the behavioural measures. None were removed from the final data sets. A number of analyses were undertaken to explore the data. The measure of d' (a measure of sensitivity to the signal; zHits – zFalse Alarms) and c (the criterion bias to say “yes” regardless of the information; (zHits + zFalse Alarms)/2) were calculated for each experience group and then compared. Mixed Analyses of Variance (ANOVA) compared set size (3, 6, and 9) across group (lifeguards and non-lifeguards) and drowning type (active or passive). As participants' lifeguarding experience was the focus of this research, only significant interactions including this factor are explored. If set size produced a significant main effect or was involved in a significant interaction with experience, then planned comparisons were employed, comparing set sizes 3 and 6, and set sizes 6 and 9 (including the experience factor in order to identify the locus of the interaction). Where significant interactions required further exploration, t-tests were used. Bonferroni corrections compensated for these multiple comparisons. All eye-movement data was processed by and prepared for analysis using the programme BeGaze. The minimum duration for a fixation to be measured was 80 ms and fixations were calculated from saccadic velocity, with a peak velocity of 400/s. The measures explored within these eye-movement data were the number of targets fixated, time taken to first fixate the targets (measured from drowning onset), percentage dwell time (number on samples on target post-drowning onset/total possible time that one could have looked at the target) and number of fixations on targets. The time between first fixation and first correct response time was also analyzed to provide a measure of processing time.

RESULTS

Signal detection analysis Accuracy for detecting a drowning target (i.e., making a response within the drowning window) was subjected to signal detection analysis. Neither d' ($t(40) = 1.01$, $p = .320$) (2.8 vs. 2.5 non-lifeguards and lifeguards, respectively) or c ($t(40) = -1.27$, $p = .208$) (−2.3 vs. -1.9 non-lifeguards and lifeguards, respectively) were found to differ significantly between the two groups. This suggests that there was no difference between the participants likelihood to detect the target and their likelihood to say "yes" to the signal. All subsequent analysis focuses on trials on which there was a target. 4.2 | Behavioural responses The percentage of trials with a drowning target that were correctly responded to were then analyzed. Trials with a drowning target were considered incorrectly responded to if no response was made following the onset of drowning activity. Premature responses made to drowning-present trials were analyzed. Premature responses occurred on 9.05% of all trials. Nonlifeguard participants were responsible for 3.02% of premature responses and lifeguards were responsible for 6.03%. There was no statistical difference in the premature responses made by lifeguards and non-lifeguards ($t(40) = -1.8$, $p = .07$). Correct responses were converted into percentages of the total drowning trials in each condition (Table 2) and subjected to a group x drowning type × set size (2 × 2 × 3) mixed ANOVA. Unlike Laxton and Crundall (2018), a main effect was not forthcoming for participant group on accuracy rates ($F(1,40) = 1.3$, $MSe = 387.5$, $p = .259$, $\eta 2\ p = .03$). Though the lifeguards identified 89.5% of targets compared to the non-lifeguards 84.6%, this difference was not significant. The difference between accuracy for active trials and passive trials, and the main effect of set size also failed to reach significance. One important interaction was noted between set size and experience ($F(2,80) = 4.6$, $MSe = 231.8$, $p = .012$, $\eta 2\ p = .10$). Repeated contrasts revealed this interaction to lie between set size 6 and set size 9 ($F(1,40) = 8.1$, $MSe = 461.9$, $p = .007$, $\eta 2\ p = .17$) . As can be seen from Figure 2, lifeguards only show superiority at the two lower set sizes. Response times were then subjected to a similar 2 × 2 × 3 ANOVA (group x drowning type x set size. One participant, who did not respond to any drownings in the set size 6 condition, was removed from the analysis. Main effects were found for all three factors. First a group effect was noted ($F(1,39) = 4.2$, $MSe = 2603666$, $p = .026$, $\eta 2\ p = .10$), with lifeguards identifying drowning targets faster than non-lifeguard participants (4,215 ms vs. 4,935 ms). The main effect of drowning type ($F(1,39) = 2.80$, $MSe = 3{,}198{,}316$, $p < .001$, $\eta 2\ p = .35$) revealed passive drownings were identified over a second faster than active drownings (4,051 ms vs. 5,092 ms). The main effect of set size ($F(2,70) = 8.7$, $MSe = 1{,}449{,}725$, $p < .001$, $\eta 2\ p = .18$) reflects an ostensible increase in RTs with an increase in distractors (4,125 ms, 4,723 ms and 4,865 ms for set sizes 3, 6, and 9, respectively). Planned repeated contrasts demonstrate that set size 3 evoked faster RTs than set size 6 ($F(1,39) = 12.2$, $MSe = 2{,}274{,}287$, $p = .009$, $\eta 2\ p = .25$) however there was no difference in RTs between set size 6 and 9 ($F(1,39) = .47$, $MSe = .499$, $p = .270$, $\eta 2\ p = .01$). No interactions were found with experience.

Eye-movement measures

How many targets were fixated?

The number of drowning swimmers that received a fixation after drowning onset were analyzed. A fixation on a drowning target was only considered relevant if it occurred within the drowning window. Two potential participants did not reach a satisfactory calibration with the eye-tracker and these results were omitted from the results. There was a good tracking ratio average for all trials (average 90.23%), however it should be noted that two participants' averages fell between 70–80%. The number of targets that received a fixation were converted into percentages of total targets (see Table 3) and subjected to a group x drowning type x set size (2 × 2 × 3) mixed ANOVA. The main effect of group was not significant. However, main effects were found for both drowning type and set size. The main effect of drowning type ($F(1,40) = 4.6$, $MSe = 34.6$, $p = .038$, $\eta 2\ p = .10$) identified that passive drownings were more likely to be fixated than active drownings (95.4% vs. 93.8%). The main effect of set size was also significant ($F(2,80) = 4.6$, $MSe = 77.9$, $p = .013$, $\eta 2\ p = .10$). Planned repeated comparisons between set size 3 vs. 6 and set size 6 vs. 9 showed no significant differences in fixation percentages. As such the additional t-test (Bonferroni adjusted) between set size 3 and 9 was run which showed that fewer targets were fixated at set size 3 than set size 9 (92.4% vs. 97.4%) ($t(41) = -2.6$, $p = .012$). A three-way interaction between group x drowning type x set size was found to be significant ($F(2,80) = 3.3$, $MSe = 91.9$, $p = .043$, $\eta 2\ p = .08$). Figure 3 shows that this appears to be driven by the number of targets fixated by lifeguard participants, which seem to be differentially affected by the increase in set size across drowning target type. Lifeguards are close to ceiling in terms of the number of targets fixated in set size 6 for passive drowning trials, though this number decreases slightly in set size 9. However, with active drownings there is an increase in the number of fixated targets at set size 9 compared to set size 6. Non-lifeguard participants' likelihood of fixating the targets is the same, regardless of drowning type, and follows the pattern of results produced by lifeguards when fixating active targets. To unpack this interaction two drowning type x set size mixed ANOVAs were carried out for each group. In the non-lifeguard conditions the main effects of set size and drowning type were not significant, however the interaction effect between drowning type and set size approached significance ($F(2,40) = 3.08$, $MSe = 203.18$, $p = .057$, $\eta 2\ p = .13$). The second drowning type x set size ANOVA for the lifeguard group revealed a main effect of drowning type ($F(1,20) = 4.71$, $MSe = 43.16$, $p = .042$, $\eta 2\ p = .07$), with passive targets more likely to be fixated than active targets (96.2% vs. 93.7%, respectively). The interaction effect between drowning type and set size did not reach significance ($F(2,40) = 2.65$, $MSe = 108.89$, $p = .083$, $\eta 2\ p = .12$).

Time taken to first fixate the targets

The time (ms) to make the first fixation on the target (calculated from drowning onset) was subjected to a similar 2 x 2 x 3 ANOVA. A main effect for drowning type was found ($F(1,40) = 16.0$, $MSe = 1{,}073{,}289$, $p < .001$, $\eta 2\ p = .26$), with passive drowning trials receiving an initial fixation an average of 500 ms before active drowning trial (1,615 ms vs. 2,136 ms). The other two main effects failed to reach significance. There were no interactions with experience.

Dwell times on targets

Dwell times (the amount of time the participants' eyes were on the target as a percentage of the time it was available for inspection; Table 3) were also analyzed with a 2 x 2 x 3 ANOVA. There was a main effect of both drowning type ($F(1,40) = 7.3$, $MSe = 1.9$, $p = .010$, $\eta 2\ p = .15$) and set size ($F(2,80) = 6.7$, $MSe = 15.3$, $p = .002$, $\eta 2\ p = .14$). Passive drownings received a shorter dwell compared to active (35.5% vs. 38.4%; possibly reflecting the more captivating antics of active targets) and dwell decreased as set size increased (38.9% vs. 37.8% vs. 34.1%, respectively; presumably caused by the increase in other stimuli trying to capture attention)

Fixations to targets

The mean number of fixations on the targets was subjected to a 2 × 2 × 3 mixed ANOVA (group x drowning types x set size). No difference was found between experience groups (non-lifeguards 9.8 and lifeguards 10.0), however, main effects were found for drowning type and set size. First, drowning type ($F(1,40) = 17.6$, $MSe = 6.7$, $p < .001$, $\eta^2_p = .31$) revealed that active drowning targets received more fixations than passive (10.5 vs. 9.2). The main effect of set size ($F(2,80) = 13.6$, $MSe = 5.8$, $p < .001$, $\eta^2_p = .25$) noted a linear increase in the number of fixations as set size increased (8.9 vs. 10.0 vs. 10.8). Planned repeated contrasts revealed that set size 3 was different from set size 6 ($F(1,40) = 9.2$, $MSe = 11.2$, $p = .004$, $\eta^2_p = .19$), and set size 6 was different from set size 9 ($F(1,40) = 5.6$, $MSe = 10.2$, $p = .023$, $\eta^2_p = .12$).

One interaction was subsumed by a 3-way interaction between group x drowning type x set size ($F(2,80) = 3.9$, $MSe = 4.6$, $p = .025$, $\eta^2_p = .09$). From Figure 4, this appears to be driven by the difference in the number of fixations on active and passive targets made by nonlifeguard participants at set sizes 3 and 9. Lifeguard participants also appear to differ in the number of fixations given to active and passive targets at set size 9. To unpack this interaction two drowning type x set size ANOVAs were conducted for each experience group. Some key differences were noted between the analysis for lifeguards and non-lifeguards. First, the main effect of drowning type remained for both groups, with active drownings receiving more fixations. This effect was stronger in the lifeguard group ($\eta^2_p = .04$ vs. .24, for non-lifeguards and lifeguards, respectively). The effect of set size also remained for both groups, however planned contrasts revealed that non-lifeguards fixated more targets at set size 9 compared to 6, while the lifeguards fixated more targets at set size 6 compared to 3. Finally the interaction between drowning type and set size was present for both groups, with active targets being fixated more than passive at set size 9. However, the increase from set size 6 to 9 for the non-lifeguards' fixations to active targets caused a cross over effect with the passive targets.

6. Drowning Detection System using LRCN Approach

Abstract

This project provides the insights of a real-time video surveillance system capable of automatically detecting drowning incidents in a swimming pool. Drowning is the 3rd reason for the highest unintentional deaths, and that’s why it is necessary to create trustable security mechanisms. Currently, most of the swimming pool\'s security mechanisms include CCTV surveillance and lifeguards to help in drowning situations. But this method is not

enough for huge swimming pools like in amusement parks. Nowadays, some of the security systems are using AI for drowning detection using cameras situated underwater at a fixed location and also by using floating boards having a camera mounted on the bottom side so that underwater view can be captured. But the main problems in these systems arise when the pool is crowded and vision of cameras is blocked by people. In this project, rather than using underwater cameras, we are using cameras situated on top of the swimming pool to get an upper view of the swimming pool so that entire swimming pool will be under surveillance all time.

Introduction

I. BACKGROUND AND MOTIVATION

Drowning is the 3rd reason for the highest unintentional deaths, and that’s why it is necessary to create trustable security mechanisms. Currently, most of the swimming pool's security mechanisms include CCTV surveillance and lifeguards to help in drowning situations. But this method is not enough for huge swimming pools like in amusement parks.

Some of security systems are using AI for drowning detection using cameras situated underwater at a fixed location and also by using floating boards having a camera mounted on the bottom side so that underwater view can be captured. But the main problems in these system arises when the pool is crowded and vision of cameras are blocked by people. In this project, rather than using underwater cameras, we are supposed to use cameras situated on top of the swimming pool to get an upper view of the swimming pool.

II. INTRODUCTION

Drowning is the 3rd reason for the highest unintentional deaths, and that‘s why it is necessary to create trustable security mechanisms. This project aims to create a system that will be able to automatically detect drowning incidents in the swimming pool using human action detection. The drowning detection model will be used to process and classify video that will be given to the system which will be recorded using live surveillance cameras. The system will break this video in image frames and apply model over it and if the early actions of drowning like hand waving, water splashing or diving is detected then the system will set the alarm so that the lifeguards can initiate their rescue operations. The classifier model is trained using a Long-term Recurrent

Convolutional Network which is a combination of convolutional neural network and recurrent neural network which is suitable for large-scale visual understanding tasks such as activity recognition and image captioning.

III. LITERATURE REVIEW

This section aims to identify and discuss the lacunae and similarities respectively, in some of the previous works related to drowning detection systems.

Lei Fei, Wang Xueli, Chen Dongsheng, proposed a background subtraction method for drowning detection and swimmer identification using visual surveillance in their research paper. This method fails to reflect real background accurately thus restricting model accurate shape detection of moving objects. It also fails to reflect sudden background changes.

Ajil Roy, Dr. K. Srinivasan, proposed drowning detection using RFID-based swimming goggles, however, this model also fails to overcome the limitation of accuracy since the water sensor is not placed very close to the mouth and nose. But this model successfully overcomes limitations of video surveillance-based drowning detection systems like the need for high power computing devices.

Chi Zhang, Xiaoguang Li, Fei Lei, proposed ?A Novel Camera-Based Drowning Detection Algorithm? using input video sequences obtained from underwater cameras. In this case, to detect drowning swimmers an implementable real-time detection system with high accuracy will be needed.

IV. PROPOSED METHODOLOGY

In particular, a specific type of neural networks called Convolutional Neural Networks (CNNs) is best suited for the task of image recognition. So implementation of Long Term Recurrent Convolution Network (LRCN) approach suitable for Video Classification & Action Recognition.

The Long Term Recurrent Convolution Network methodology is a combination of Convolutional Neural Network (CNN) & Recurrent Neural Network (RNN). LRCN is end-to-end trainable and appropriate for vast visual understanding tasks such as video description, activity recognition and image captioning. The main idea is to learn visual features from video frames with the help of CNN & then use LSTM layers to transform a sequence of image embeddings into a class label, sentences, probabilities, etc.

In this segment, we empirically propose LRCN approach for the implementation of drowning detection as CNN extracts the features from the input provided to the model and then the LSTM layers predict the action of the human whether one is drowning, swimming or diving.

A. Convolutional Neural Networks

Convolutional Neural Network (ConvNet/CNN) is a Deep Learning algorithm that can take in an input image and assign learnable weights to various features in the image. As compared to other classification algorithms, Convnet requires less preprocessing. Filters are hand-engineered, with enough training in primitive methods and ConvNets can learn these filters/characteristics.

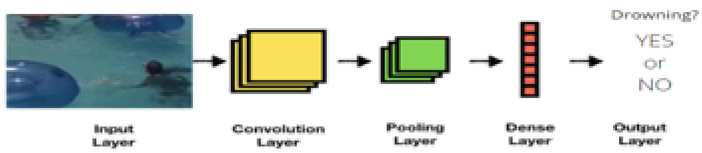


Fig 2: CNN Network

B.LONG SHORT-TERM MEMORY NETWORKS

LSTM Networks is a supplement of recurrent neural networks (RNNs) mainly introduced to handle situations where RNNs fail. RNN is a network that works on the present input by taking into account the previous output (feedback) and storing it in its memory for a short period of time (short-term memory). Long Short-Term Memory (LSTM) overcomes the vanishing gradient problem as the training model is left unaltered. Long time lags in certain problems are crossed using LSTMs which also handle noise, distributed representations, and continuous values. LSTMs provide us with an immense range of parameters such as learning rates, and input and output biases. Hence, no need for fine adjustments. The update complexity weight is reduced by O(1) in LSTM’s.

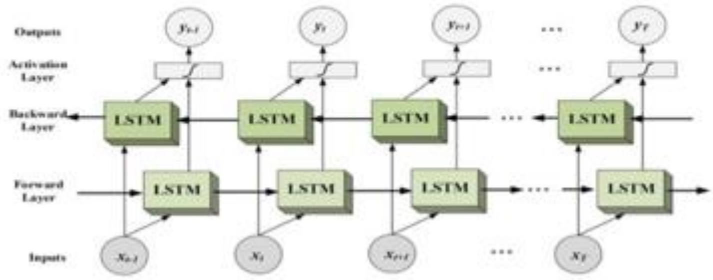


Fig 3: LSTM Architecture

Thus, we suggest two approaches for implementing the drowning detection-

1. Using ConvLSTM2D layers in the model
2. Using LRCN approach
3. As shown in the figures below, both of this approaches provide convincing results and proper detection if someone’s drowning in the swimming pool.

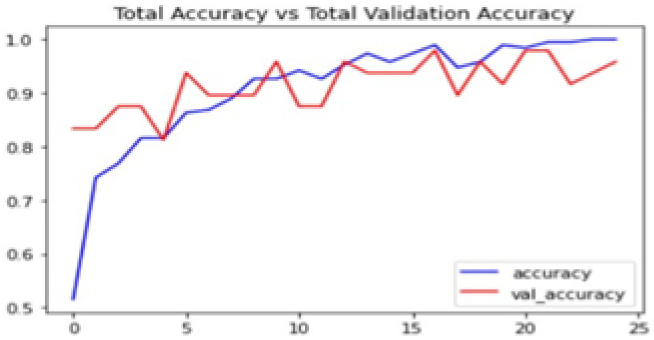


Fig.4 Accuracy Graph for ConvLstm2D

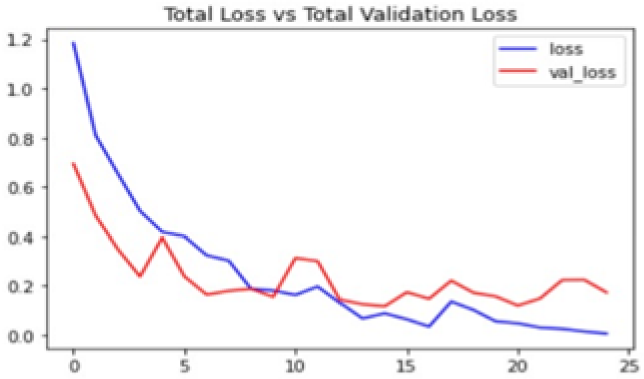


Fig.5 Loss Graph for ConvLstm2D

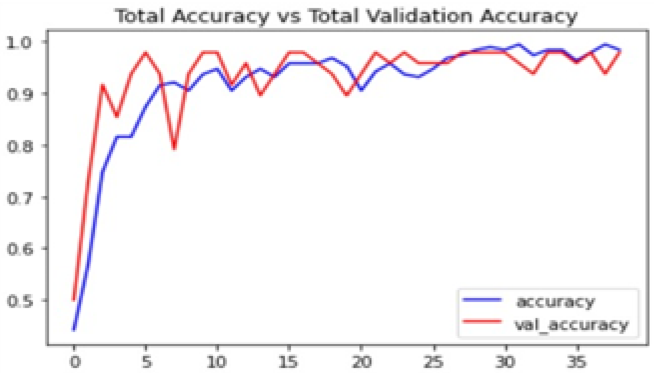


Fig 6 Accuracy Graph for LRCN

4.

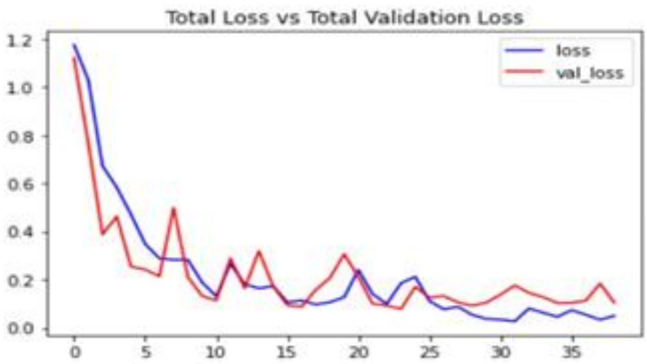


Fig 7 Loss Graph for LRCN

The project has three phases:

1. *Preprocessing of Data:* The Video Dataset used for training the LRCN model is a blend of the UCF50 dataset & videos that are downloaded from Youtube. The video dataset of swimming and drowning is not available on any websites so we had to download related videos for the same, one by one from youtube. But these videos had extra activities like in swimming videos there was a person approaching the swimming competition, taking their stance and handshakes after the race. In drowning-related videos, there was a rescue operation as well in video. So we cropped such extra activities and kept only the required part of the video which were actual swimming and drowning actions. Finally, the videos of diving, breaststroke swimming from the UCF50 video dataset were combined with the data that was created by us for further training of the models. For training, the model image frames were extracted from each video and used for model training, and the same for testing.
1. *Creating Deep Learning Model:* In this project, we have used two models which are the ConvLstm2D model and the LRCN model. In this phase, we used training set videos to train the model by passing them through multiple layers of the ConvLstm2D network and also implemented the LRCN network to compare the better results provided by the respective models. The architecture of both models are as follows:

A. ConvLstm2D Model

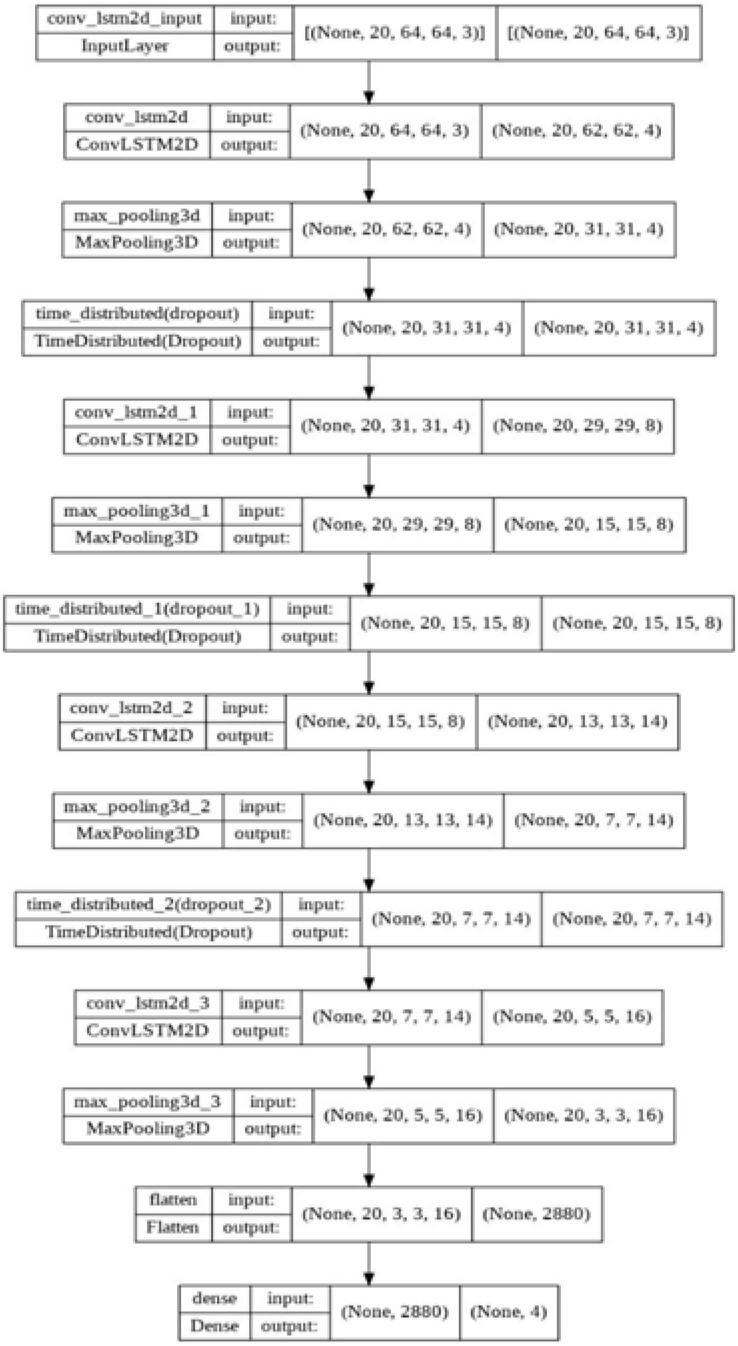


Fig 8 ConvLstm2D Model Architecture

B. LRCN Model

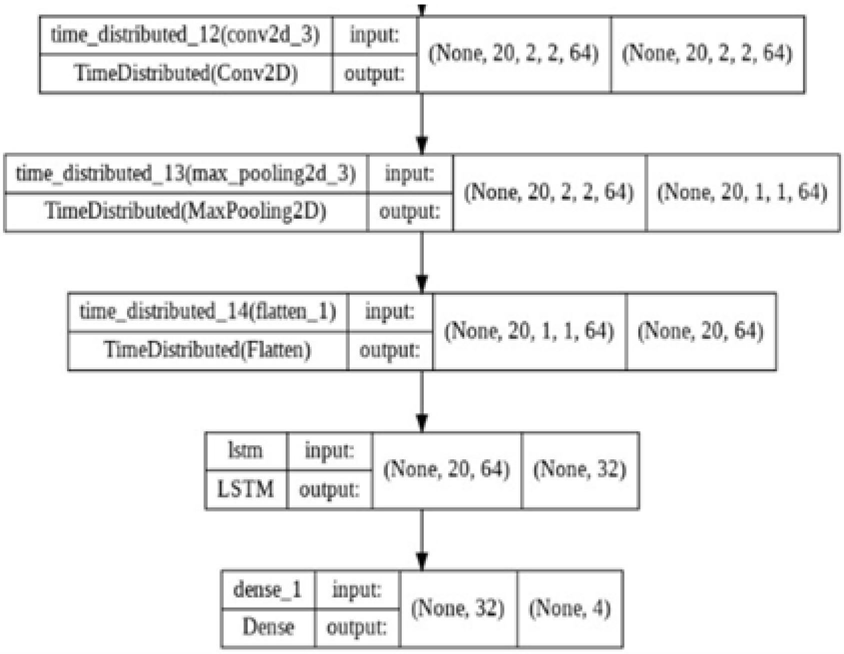


Fig 9 LRCN Model Architecture

V. SOFTWARE DESIGN

A. Incremental Model

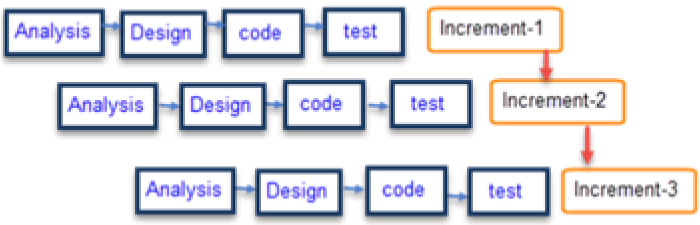


Fig 10: Incremental Model

The incremental model is a process of software development where requirements are broken down into multiple standalone modules of the software development cycle. Incremental development is done in various stages such as analysis design, implementation, testing/verification, and maintenance.

Each iteration passes through the analysis, design, coding, and testing phases, and each subsequent release of the system adds function to the previous release until all designed functionality has been implemented and accepted by the user.

The first increment is often a core product where the basic functionalities are addressed, and supplementary features are added in the next increments. After completion of the first iteration, the product is handed over to the end-users for their feedback. Once the core product is analyzed by the client, further improvements are done in the next increment.

VI. RESULT AND DISCUSSION

The outputs achieved will predict the class names for a batch of frames of the videos given as input. The predicted class name having the highest probabilities will be detected as the action being performed in the swimming pool.

1. The predicted class name having maximum probability can be displayed as the confidence variable.
2. After the successful completion of the project, one can:
3. Observe the video surveillance and rely on the drowning detection system.
4. An alarm will be raised if someone is detected as drowning.
5. Drowning preventive measures can be performed due to early alerts raised by the system.
6. The project should examine the actions performed by swimmers to detect drowning more accurately.
7. **VII. FUTURE SCOPE**
8. Availability of better dataset, modern methodologies, and technologies with high computational power accompanied by high-quality surveillance cameras, will help to improve the accuracy of drowning detection & even can be used in adverse conditions.
9. After the implementation of all these essentials, this system also can be used on sea beaches for drowning detection

10. Conclusion

Once we have the working drowning detection model we can feed live video footage of the swimming pool to it so that it can keep detecting continuously for any drowning activities. If drowning is detected it will be highlighted on the system screen as well as alarms will be raised to alert security guards so that they can initiate rescue.

7. Drowning detection systems for public swimming pools

General

The term 'drowning detection system' (DDS) is used to describe various electronic systems that are designed to assist with the surveillance of swimmers within the water of a swimming pool. Approximately 30 systems have been installed in public swimming pools in the UK to date.

There are 3 general types:

CCTV cameras	that give a lifeguard additional underwater views
CCTV cameras with computer monitoring and automatic alarms	that detect 'static' solid objects of appropriate size
Wristband tags with computer monitoring and automatic alarms ,	that detect wristbands that are outside set depth movement and time parameters.

The concept is somewhat controversial, and there is debate within the industry on the extent that such systems should complement traditional lifeguarding arrangements. The advantages of an additional safety system for a pool are set against other factors such as: • Concerns over inconsistent levels of reliability of systems and situations where glare, swimming aids or high occupancy / activity rates can cause false alarms • Impact of the additional cost on financial viability • Risk that such systems can create a false sense of security for lifeguards • Risk that numerous false alarms can cause lifeguards to ignore a genuine emergency situation • Limited level of in-use knowledge and experience in UK pools. These systems all require various items of equipment to be built integrally with the pool structure and it is recommended that for all new and refurbishment projects, their inclusion is considered at an early stage in the briefing and design processes. However, some systems can be retro-fitted in existing pools. The issues outlined should be taken fully into account during the development of the operator's 'Pool Safety Operating Procedure' 1 whilst considering the priority of maintaining high levels of lifeguard vigilance and effectiveness.

Available legislation and advice

There is no legislation for drowning detection systems, with only limited published guidance available. There is a one line reference in the HSG 179 'Managing Health and Safety in Swimming

Pools', the standard UK reference for swimming pool safety and management, which states that 'cameras and computer aided surveillance systems may be used to assist supervision'. The European standard for swimming pool safety BSEN 15288-1 and 2 : 2008 mentions 'special installation / detection technologies' as requiring consideration at the design stage and a factor to be considered in a risk assessment on supervision requirements. The ISRM 2 website has limited information on trials, research and articles about such systems but with some examples accompanied by a disclaimer by the organisation.

Drowning detection systems

Key issues

Swimming pools may have potential blind spots due to their size and shape and these should be carefully assessed, analysed and appropriately mitigated during the design process. The pool operator has a responsibility to undertake a full risk assessment of all relevant factors, including potential blind spots, before operating the pool. The HSG 179 guidance stresses the importance of lifeguards having good visibility beneath the water and states that a minimum number of lifeguards should be on duty for programmed and un-programmed swimming sessions. For example, in the case of a 50 m pool, recommendations are for a minimum of 4 lifeguards in normal conditions and 6 lifeguards in busy conditions. The document also states that glare and specular reflection3 should be avoided and points to particular visibility problems for pools that are more than 16 m wide. Even when glare is not an issue, it is likely to be difficult for a lifeguard on one side of the pool to see objects beneath the water on the far side. There can also be visibility issues on the near side of the pool unless the lifeguard is positioned close to the pool edge. This is mentioned in a number of Health and Safety Executive (HSE) reports of prosecutions after fatalities in swimming pools and where fines have been imposed under the Health and Safety at Work Act (HASAWA) 1974. Important detailed design considerations are: • The specialist suppliers of DDS systems should be consulted on the suitability of their equipment for a particular swimming pool project • Underwater cameras that project from the pool sidewalls will not be compatible with moving floors and bulkheads. In this situation, overhead cameras can be added into the system which can be effective in pools up to 2.6 m in depth • Cameras can be fitted with different lenses to increase coverage and avoid 'blind spots' • Where underwater cameras are built into the pool tank wall, the depth and size of the camera and housing can be co-ordinated to achieve an aesthetically pleasing appearance to the pool light fittings.

	CCTV Viewing	CCTV + Monitoring + Alarms	Wristband tags + Monitoring + Alarms *
CCTV Viewing	A passive aid to the lifeguard that gives an underwater view via CCTV cameras. However it does not incorporate a detection system. Glare can be a potential problem with viewing monitors in bright environments. However, it is possible to incorporate shrouds around monitors.	A computer aided detection system, that is not primarily a CCTV system and does not require a lifeguard dedicated to watch the TV monitor. The system monitors the swimming pool water and when it detects a person in potential difficulty (i.e. a static solid object of appropriate size and shape), it alerts the lifeguard via an LED monitor. The LED monitor visually flashes and produces an audible alarm. The lifeguard will be given a location on the LED screen and can also see the 'casualty' on the supervision workstation monitor.	Individual wristbands, about the size of a small wrist watch, that are worn by users of the pool. If a bather (with wristband) approaches the preset parameters of the system, an alert via radio and / or ultrasonic is generated. Initially a small audio sound and flashing LED light will remind the bather to return to a safer location. If the bather does not respond appropriately, the system activates alarms on the receivers worn by the lifeguards. * System mainly targeted at the Hotel sector for pools

		The computer continually monitors the pool assessing any potential problem.	without constant lifeguard supervision
Real time underwater images on a viewing monitor	(Viewing monitor adjacent to the lifeguard chair)	(Viewing monitor adjacent to the lifeguard chair)	
Underwater CCTV cameras			
Overhead CCTV cameras	(Required if a moving floor is installed)	(Required if a moving floor is installed)	
Computer monitoring for static solid objects			
Computer monitoring for set parameters (depth / movement / time of wristband tag)			
Automatic alarm			
Audio and video recording facility			
Installation cost (for a typical 25m pool)	> £35k	> £100k	> £15k
Operating costs (for a typical 25m pool)	> £2k	> £2k	> £2k

8. A novel drowning detection method for safety of swimmers

I. INTRODUCTION

Safety in water has been a concern for many centuries for the survival of human lives. The latest technology advancements have enabled to come up with effective drowning detection methods (DDM). A recent report from World Health Organizations (WHO) gives us some insight into the drowning incidents globally. The number of reported drowning deaths globally is 37200. The highest numbers of deaths are in low and middle-income countries. The survey also points that children have the largest death ratio compared to adults. Majority of the drowning deaths are reported from open water bodies likes lakes and sea, and not in pools. In the report WHO has recommended various drowning prevention techniques like constructing fences across the lakes, to prevent accidental fall to teaching school age children swimming as a part of their curriculum in schools. According to Jeff Ellis and Associates, an International Aquatic Safety and Risk Management Consulting firm, drowning is divided into five stages .

Stage 1 - Shocking surprise: This initial stage is characterized by the shock of   and drowning difficulty in breathing. The person starts to show a higher level of distress and attempts to reach the surface of the water, but in the vertical direction.

Stage 2 - Involuntary Breath Holding: In this stage attempt to come to surface of water stops. He starts involuntary breath holding. Water has entered the mouth, causes the epiglottis to close. The victims gradually become unconsciousness, as breath is stopped.

Stage 3- Unconsciousness: The victim becomes unconscious, and the body starts to sink to the bottom of the water. Unless breathing is re-established, the victim remains unconscious.

Stage 4- Hypoxic Convulsions: The oxygen level in the brain reduces drastically. The victim’s skin turns blue, especially in the lips and fingernail beds.

Stage 5- Clinical Death: Death is the final stage of drowning

Any prediction of drowning during the early stages always reduces risks during the rescue operations.

Drowning is classified into two, active drowning and passive drowning. In the active drowning, the victim express distress that is noticeable to others. In passive drowning, there is no distress exhibited by the victim. The passive drowning happens due to medical reasons like stroke; heart attack etc. or it could be that the person has become unconscious. Passive drowning victims generally have their face down underwater and in some cases will be floating below the surface. All these characteristics make the detection of drowning difficult for even professional lifeguards. With the advancement in technology, various drowning detection methods are available. Some detectors are wearable like Kingii . Another proposed wearable solution is a wearable life jacket which communicates with the life guard and also gets inflated during danger. However many of these wearable systems are designed with a specific use case scenarios. A solution developed for pool might not be effective in the ocean. Another type of detectors is non-wearable. E.g. Video surveillance based systems that focus on taking images of the swimmer and water. The video surveillance comes with few design limitations like unable to work in darkness, prediction depends on the quality of the image, and need of high power computing devices. Various enhancements are provided to the video surveillance methods.

Many DDM designed today have limited scope of usage. DDM developed for pools requires lots of customisation work to make it work for the ocean. E.g. A video surveillance method used in the pool may not be useful to detect drowning in oceans with high waves. The paper address few of the points mentioned in the WHO report, like ease of use to children, economically viable solutions. The children playful nature makes them curious to water and less cautious about the dangers. A wearable DDM should not be impacting his fun while swimming. This will be the most important deciding factor that enables children to use the DDM. To address the above needs, we are proposing a DDM that can be attached to the swimming goggle. During drowning detection, the alarms are transmitted through water and picked up by the receivers/hydrophone placed at different locations in the water body. This is further processed and transmitted to the lifeguard for the rescue operation. The organisation of paper includes the design of swimming goggles. A system for drowning detection is reported in section II. Section III discusses the block diagram and flowchart representation. The simulation and experimentation is reported in section IV. Section V discusses results. The conclusion and future scope is presented in section VI.

II. SWIMMING GOGGLES BASED DROWNING DETECTION METHOD

Drowning detectors detect the drowning by analysing the various readings exhibited during drowning distress, by the victim. This could be like monitoring the waves generated due to panic to monitoring the irregular pressure variations from the gadget, used by the victim. In the method described in this paper, we try a novel method which could predict the drowning. It is obvious that the distress starts as soon as the person reaches the mental barrier point of his breath holding capacity. Maximum breath holding time varies from person to person. An average person can hold breath for 1 to 2 minutes while a 4 year old kid could hold very less duration.

If we can develop a system that can trigger alarm in case both mouth and nose are under the water, beyond the breath holding time, we can predict the drowning.

The swimmers, especially children get easily disturbed by placing any sensors very closer to the mouth and nose. They may also try to remove it because of the disturbance. The regular touching on the sensors can also break the unit. Having tried out many alternatives, we have found out swimming goggles are the best place to fit the DDM. This makes the solution acceptable to all age group and does not disturb the swimming pleasure. On the negative side, since water sensor is not placed very close to the nose and mouth, the accuracy of the detection has to depend on redundant water detecting sensors. the parts of the drowning detection system attached to the goggles. Two drowning detection sensors are placed on the side elastic of the goggles. A simple resistive based circuit interfaced with the microcontroller is used. The two Input Output lines of the microcontroller form the resistive circuit. When in air, due the high resistance of air between these wires, the circuit remains open. When immersed in water, the circuit gets completed with the water. The two detectors represent that both sides of the nose and mouth are closed. The Alarm transmission module is used to send the alarms when the drowning is detected. Alarm transmission modules are triggered by drowning detection unit. The Alarms are transmitted using the underwater communication .

III. SYSTEM DESCRIPTION AND FLOWCHART

The entire system consists of drowning detection enabled goggles and an alarm receiver. The alarms from goggles are transmitted under water through existing under water communication technology like acoustic waves, infrared etc. The selection of communication technology is crucial from the range perspective. The bottom portion of the alarm receiver is immersed underwater and top portion faces outside water. The bottom portion receivers picks the alarms transmitted from the victims goggles. These alarms are further processed and transmitted to life guard or concerned authority wired or wireless, depending on the need.

a) Drowning detection sensor: Sensor consists of a simple circuit to detect the presence of water. Output goes HIGH when water is detected and LOW in the presence of no water.

b) Transmitter: Signals or information obtained from the detection sensor is transmitted for further process by Transmitter.

c) Receiver: Ultrasonic waves of 40 kHz are received and convert the signal into useable form. The drowning detection sensor and the Ultrasonic transmitters are both connected to the Arduino board. The board is programmed with a timer that simulates the user configurable breath holding time. When the drowning detection sensor comes in contact with the water, the timer is started. Another Arduino board is connected with the ultrasonic receiver. As the timer gets expired the ultrasonic transmitter transmits the signal to receiver. The ultrasonic communication was used for the experiment mainly because the miniature circuit will enable easy interfacing. This is also less prone to interference and noise. The ultrasonic module used was HC-SR04[13]. The Yellow LED in the Receiver board gets ON upon receiving of the signals. A audio siren is also programmed in the second Arduino board to simulate drowning. The experimentation setup is shown in Figure 7. Fig.7. Prototype unit of proposed DDM

V. RESULTS AND DISCUSSION

Waveform in yellow color is the output waveform received from the buzzer. Fig. 8. Output of the prototype seen in Oscilloscope The model was simulated using Proteus design suite to check its working before the actual physical prototype is constructed. The waveform in yellow colour is the output waveform received from the buzzer. The buzzer is triggered by receiving ultrasonic waves. The system is tested successfully, and the alert siren has been blown to alert the lifeguards. The drowning detection system will be ready to be used underwater when replaced it with the underwater module. According to statistics, the average time a child of age between 5-10 years can hold their breath for 10 sec underwater. We estimated the probability of danger based on the time for which the child is in water and plotted a graph between time and probability of danger. The probability of danger denotes threshold time. The same has been done for an adult, and the graph is plotted. We can infer two scenarios from this which are: Scenario 1: Person comes out of water before the threshold time. In this case, the threshold time, i.e. time to withstand is 10 seconds. If the person comes out on or before 10 seconds, the danger associated with him is zero, entire setup gets reset once he comes out of the water. From the graph, the time a person stays inside water is shown by an increasing slope, as soon as the person comes out the graph drops abruptly and comes to 0 probability of danger and remains the same until the person is above water. The whole system is reset automatically when the person comes out of the water. Person doesn't come out of water before threshold time In this case, if a person stays inside water for more than his threshold time, it indicates danger, this is shown in the graph using red color. An increasing slope shows the time a person stays inside water. If the person stays in water even after the threshold time, the value of probability of danger reaches 1 and remains constant thus indicating that the person is drowning. The whole system is reset automatically when the person is saved and brought out of water. Both the scenarios are depicted

VI. CONCLUSION

Life safety in water has been a concern for many centuries. Latest technology advancements has enabled us to come up with effective drowning detection systems. However many of those solutions are costly and limited to few. Survey reports show us that highest numbers of deaths are reported in low and middle income countries. The survey report also mentions the children have the largest death ratio compared to adults. Also the deaths reported in these incidents are more from open water bodies than closed water bodies like swimming pools. The solution described above will be able to address these issues. The swimming goggles with drowning detection unit can be economically viable solution. The range of the alarms transmission can be improved by using underwater acoustics. Any age groups will be comfortable wearing the goggles, without hampering the recreational joy while swimming. The goggles can be useful even in sea. The alarm receivers can be placed at different locations in the water bodies which is having high chance of drowning. Another major advantage of this approach unlike other approach is the ease of use in all atmospheric conditions, like rain or wind to day or night. This solution is also a reliable solution where the life guards have difficulty to monitor the swimmers like a highly crowded sea. This is one of the biggest challenges the lifeguards face. Many of the training to life guards includes how to monitor drowning in a large crowd like in beeches. The future research plans include improving the underwater communication range by using various other technologies. This will enable to use this system in seas also. The alarm receivers can be easily connected to the buoy. With the help of establishing a standard communication protocol, we will be able to communicate more information to the lifeguards, as the name of the victim etc. This will help the lifeguard to search for his previous medical records as does the patient had any heart or lungs diseases etc. This information will provide an additional advantage while doing the rescue operation and while doing first aid. We also have plans to integrate a

Global Positioning System (GPS) and pressure sensor. As pressure increases with the depth, the pressure reading will let the lifeguard know the depth at which victim is located. The GPS reading will be saved whenever the signal is available. If the system can tell the last previous GPS reading that was stored, it will enable the lifeguard to know what the approximate location of the victim. This feature will be very timesaving for lifeguards reducing their search time to near locations especially in case of lakes and oceans. However, this feature should be valid only if the GPS connectivity was alive with a minimum of 10 minutes before the drowning, as a very old GPS value will give a wrong location itself.

9. DROWNING DETECTION SYSTEM USING CNN

I. INTRODUCTION

The term drowning detection system is used to describe various electronic system. Safety in a water has a main concern in many centuries for survival of human life. The latest technology advancement have enables to come up with effective drowning detection methods. This project uses CNN architecture to classify different object with their dimension (In general height and width of the object),so we detect human from the video frame ,then we calculate height and width for that object. If the swimmer gets difficulty then the system throw alert for security. By using this logic we are getting up to 85% accuracy with 480p video quality and minimum spaces and with higher video quality and processing power we are getting up to>90% accuracy.

II. LITERATURE REVIEW.

On existing system there are two similar system created but these systems useful when the swimmer goes die or goes down on the surface. The existing system is more costly as well as they need large appliances, but in our system we used small appliances as well as it is more friendly for environment. The system is useful before swimmer gets difficulty in water. Existing system is only used in reached country on only on swimming pool the space is allocated for these is too small area, but in these newer version we used any ware in open surface for example river, sea

CONCEPT OF THE SYSTEM

First we get input that is our photos for example swimmer swim in swimming pool then camera take the picture on process of video processing. This video processing will capture the image. This images send to the Convolution Nural Network (CNN), then Convolution Nural Network (CNN) check if fattle is occured then alarm, if fattle is not occured then continue for the input process.

RESULT

We are getting upto 85% accuracy with 480p video quality and minimum specs and with a higher video quality and possessing power we are getting upto >90 accuracy.

CONCLUSION

The system has been tested on sseveral instances of simulated water conditions such as wwater reflection, lightening condition and false alarms. Our algorithm was able tto detect all the drowning ccondition along with the exact position of the drowning person to the swimming pool. Our results show that the proposed method can be used as a reliable multimedia video-based surveillance system.

ACKNOWLEDGMENT

Report is on the topic: “Downing detection System ”All the Relevant and essential details are included in the paper . At the beginning we have given the summarized details of the project which we are building and we have also proceed details about how the project is going to be implement and which technologies we are going to use to develop this project. We are thankful to Prof.Mr.Amar Palvankar who guides us and helps in preparing the paper .We thank him for providing us the confidence and most importantly giving us the track regarding the project topic whenever we needed it.

10.A CAMERA-BASED SYSTEM FOR EARLY DETECTION OF DROWNING INCIDENTS

ABSTRACT

We present in this paper a camera-based system for detecting drowning incidents in a swimming pool at the earliest possible stage. The system consists of two main parts: a vision component which can reliably detect and track swimmers in spite of large scene variations of monitored pool areas, and an event-inference module which parses observation sequences of swimmer features for possible drowning behavioral signs. The vision component employs a model-based approach to represent and differentiate background pool areas and foreground swimmers. The event-inference module is constructed based on a finite state machine, which integrates several reasoning rules formulated from universal motion characteristics of drowning swimmers. Possible drowning incidents are quickly detected using a sequential change detection algorithm. The proposed system has been applied to a number of video clips of simulated drowning, and promising results have been obtained.

1. INTRODUCTION

This paper presents a novel camera-based system designed to detect potential drowning incidents in a swimming pool. There are numerous swimming pools worldwide located in public places and private houses. Drowning is one of the leading causes of death from unintentional injury. According to the Centers for Disease Control and Prevention, 4,406 people drowned in the United States in 1998. Many drowning incidents happened in public swimming pools staffed with professional lifeguards. Even more nearly drowning victims are left with irreversible injuries, mostly to the brains, due to lack of timely rescue. Therefore, there is a clear need for automated drowning detection systems to provide useful assistance to lifeguards on duty or to enhance the safety of unattended pools.

Existing camera-based drowning detection systems are surprisingly rare except few reported in patents [1, 2]. In general, these systems make use of underwater cameras to detect motionless bodies at the bottom of a swimming pool; none of them involves analyzing early drowning behavioral signs, such as struggles on the water surface, which are important to both timely rescue and reliable drowning detection. Furthermore, the use of underwater cameras not only incurs high installation and maintenance cost but also faces the problem of cameras being easily occluded by nearby swimmers. Our objective is to build a camera-based system that is capable of detecting potential drowning incidents at the earliest possible stage using only off-the-shelf overhead cameras.

1.a vision component to reliably detect and track swimmers despite large scene variations of monitored pool areas, as well as to extract their motion and shape features into observation sequences; and 2) an event-inference module to parse the observation sequences of swimmer features to determine whether a swimmer has lapsed into a drowning situation.

The remainder of the paper is organized as follows. Section 2 is devoted to the vision component of the system and Section 3 describes the extraction of robust motion and shape features of swimmers. The event-inference module is presented in Section 4, and in Section 5, we summarize our work and report experimental results of the proposed system.

2. VISION COMPONENT

The main function of the vision component is to reliably detect and track multiple swimmers in order to extract their motion and shape features into observation sequences. Due to swimmer movements, wind blow, and sunlight reflection, the non-stationary nature of swimming pools precludes direct application of existing object/motion segmentation methods [5-7]. For example, when swimmers are moving about in a pool, a background pixel can be image of black stripes on the pool floor at one frame, water at another frame, and ripple at a subsequent frame. In each case the background pixel has a different color. Given such a demanding environment, our vision component is developed by exploiting the scene homogeneity of a swimming pool to handle large shift of background pixels and by constructing swimmer appearance models to obtain good segmentation results.

Swimmer Detection

A high-mounted overhead camera with an oblique view is used in the system to monitor swimmers. Figure 2 shows the block diagram of the proposed swimmer detection and segmentation method. When a swimmer just enters the monitored pool area, his/her presence can be detected by identifying a sizable image region which does not conform to a background model of the monitored pool area. The pixels of a newly detected swimmer are accumulated to construct his/her color appearance model, which will then be used to segment the swimmer in subsequent frames. To suppress erroneously classified pixels, a heuristic constraint is imposed on the intensity change of classified pixels so that foreground regions in segmentation maps should only contain 2D motion blobs corresponding to swimmers. The details are presented in the following subsections.

Background Modeling

Unlike many other natural scenes, the background scenes of most swimming pools are rather homogeneous, consisting of only a few distinct scene classes. Here, a scene class refers to a group of pixels that share similar color; typical examples include pixels of the water or those of the black strips on the floor. As can be seen from Figure 3, pixels of two scene classes, the water and the black floor strips, tend to form tight clusters in HSV (hue, saturation, and value) color space. In the proposed method, the background scene of the entire monitored pool area is modeled with a multivariate Gaussian mixture model (GMM), of which each Gaussian component accounts for the color distribution of one scene class. The standard Expectation-Maximization (EM) algorithm [11] is used to estimate the parameters of the Gaussian mixture model. To provide the EM algorithm with the number of Gaussian components and the initial estimate of model parameters, we use a mean-shift clustering algorithm [4] to identify the dominant scene classes of the monitored pool area. The number of Gaussian components is set to the number of the resultant scene classes, while the initial model parameters (i.e., the mean vector, covariance matrix, and mixing proportion) of each Gaussian component are computed from pixels in the corresponding scene class. Figure 3 shows that there is a reasonable fit between each Gaussian component (where the ellipsoid shows the surface of equal probability density of the Gaussian component) and the feature points of the corresponding scene class. Specifically, let B denote the background model consisting of a mixture of Gaussian components $\{B_i\}_{i=1}^{k}$, each characterized by its mean vector μ_{B_i} and covariance matrix Σ_{B_i}, in some proportions $\alpha_1,\ldots,\alpha_k$, where $\sum_{i=1}^{k} \alpha_i = 1$ and $\alpha_i > 0$. The probability that a pixel j with HSV color value X_j belongs to the background model B can be computed as

$$p(X_j|B) = \sum_{i=1}^{k} \alpha_i \frac{\exp\left(-\frac{1}{2}(X_j - \mu_{B_i})^T \Sigma_{B_i}^{-1} (X_j - \mu_{B_i})\right)}{(2\pi)^{3/2} |\Sigma_{B_i}|^{1/2}} \quad (1)$$

If $p(X_j|B)$ is less than a preset threshold, the pixel is considered not belonging to the background scene. All such pixels sharing some spatial proximity are grouped together using a standard connected component algorithm. The presence of a new swimmer is assumed if a sizable connected region of these non-background pixels has been detected.

Swimmer Modeling

Once a sufficient number of pixels corresponding to a swimmer have been accumulated, a color appearance model for the swimmer can be constructed. Similar to the background modeling, a Gaussian mixture model is used to model the color distribution of the swimmer pixels. That is, the color value of each swimmer pixel is viewed as a realization from a Gaussian mixture model S consisting of Gaussian components $\{S_i\}_{i=1}^{k}$, each characterized by its mean vector μ_{S_i} and covariance matrix Σ_{S_i}, in some proportions $\gamma_1,\ldots,\gamma_k$, where $\sum_{i=1}^{k} \gamma_i = 1$ and $\gamma_i > 0$. It is found that $3 \sim 5$ Gaussian components are normally enough to well model the color distribution of a swimmer. Figure 3 shows an example of a swimmer model obtained from a test video. The probability that a pixel j with HSV color value X_j belongs to the swimmer model S can be similarly computed as

$$p(X_j|S) = \sum_{i=1}^{k} \gamma_i \frac{\exp\left(-\frac{1}{2}(X_j - \mu_{S_i})^T \Sigma_{S_i}^{-1} (X_j - \mu_{S_i})\right)}{(2\pi)^{3/2} |\Sigma_{S_i}|^{1/2}} \quad (2)$$

Swimmer Segmentation

The segmentation of swimmers is closely coupled with the tracking process, which will be presented in Section 2.2. Since the location and size of a swimmer can only change smoothly over subsequent frames when the processing frame rate is high, based on the segmentation and tracking results in the previous frame the vision component can define in the current frame a local search window where each tracked swimmer is likely to appear. The segmentation of the swimmer in the current frame can be done by maximum likelihood classification of pixels within the local search window. Specifically, given the background model B and the swimmer model S, a pixel with color value X is considered to be a swimmer pixel if

$$p(X|S) > p(X|B), \quad (3)$$

where $p(X|S)$ and $p(X|B)$ are computed by (1) and (2), respectively.

However, the pixels satisfying (3) may correspond to some motion clutters, mainly resulting from water reflection and splashes, which are not explicitly accounted for by either the background model or the swimmer model. As the presence of a swimmer normally decreases the scene intensity, while water reflection and splashes always make the scene brighter, an intensity constraint is imposed to refine the results of swimmer segmentation. Specifically, a pixel j satisfying (3) is considered a valid swimmer pixel only if: (IW −Ij)/IW > τ where Ij is the intensity of pixel j and IW is the average water intensity. As the pool mainly contains the water, the average water intensity IW can be obtained from the mean intensity of the scene class having the largest sample size. In our implementation τ is set to 0.1, i.e., a swimmer pixel is normally at least 10% darker than the average water intensity.

Tracking of Swimmers

The vision component uses a first order Kalman filter to track the 2D motion blob of each swimmer and predict his/her position in the subsequent frame. The matching distance between a swimmer in the previous frame and a motion blob in the current frame is defined as the sum of: 1) the Euclidean distance between the swimmer's centroid location in the previous frame and the blob's centroid location measured in the current frame, and 2) the Euclidean distance between the swimmer's predicted centroid location and the blob's centroid location measured in the current frame. A correspondence between a swimmer and a blob is established if the matching distance between them is minimum compared to all other swimmers and it is less than the maximum possible distance traveled by a swimmer within the interframe interval.

3. FEATURE EXTRACTION

For each tracked swimmer, the system extracts multiple features, including global motion feature (moving speed), local motion feature (size variation), and shape feature (elongation measure), from every M consecutive swimmer observations. Here, swimmer observation refers to the swimmer's 2D motion blob, and M is set to 8 which is the target number of frames processed every second.

Global motion feature: moving speed As the camera in our system is mounted high above the pool, a swimmer's moving speed can be roughly approximated by the displacement of his/her centroid over subsequent frames.

Local motion feature: size variation Struggles of a drowning swimmer (characterized by fast limb movements) normally cause the size of the swimmer's 2D motion blob to change rapidly within a short duration. Hence, a normalized standard deviation of sizes for every M consecutive swimmer observations can be used to characterize the local motion of the swimmer, defined as

SV = 1 M M i=1 (si − ms)2 ms (4) where ms = 1 M M i=1 si and si is the size of the i-th observation.

Shape feature: elongation measure

In general, swimmers have two common body postures, being nearly vertical in water or being almost parallel to the water surface. As observed by an overhead camera, these two body postures can be characterized by an elongation measure, which is the ratio between the major and minor axes of a best-fit ellipse [8] (referring to Figure 4) around the swimmer. When the elongation measure is close to 1, it indicates the motion blob having a circular shape and the swimmer having nearly vertical body posture. On the other hand, when elongation measure is much larger than 1, it shows that the swimmer maintains his/her body parallel to the water surface.

4. EVENT-INFERENCE MODULE

The event-inference module is used to detect in observation sequences of swimmer features unusual patterns which indicate that a swimmer has lapsed into a drowning situation.

4.1. Rules for Evaluating Swimmer Conditions

The event-inference module of the proposed system is constructed based on the universal motion characteristics of drowning swimmers [9], which have been used by lifeguards to identify drowning swimmers. These motion characteristics can be summarized into the following rules for detecting an active drowning swimmer, who

Rule I: moves slowly or remains within a small neighbrhood,

Rule II: has nearly vertical body posture in water, and

Rule III: exhibits irregular and possibly fast limb movements.

There are three possible swimmer conditions considered in the proposed system, namely 1) normal, 2) treading water, and 3) drowning, which are categorized according to swimmers' motion characteristics and body postures. However, these three possible swimmer conditions may have some similar characteristics. For example, both a drowning swimmer and a swimmer treading water exhibit slow moving speed and nearly vertical body posture; and an energetic swimmer can have very fast limb movements while swimming freely in the pool.

4.2. Integrating Rules with a Finite-state Machine

To differentiate three swimmer conditions, a finite state machine (see Figure 5) is used to systematically integrate the formulated rules so that these rules are properly applied to swimmers in different conditions (or states) using suitable swimmer features. The initial state of a newly appearing swimmer is assumed to be 'Normal'. In this state, Rules I and II are used to differentiate normal swimmers from drowning swimmers or swimmers treading water. Only when the swimmer is detected to be moving unusually slow and having nearly vertical body posture, the event-inference module suspects an onset of a possible drowning and changes the swimmer state to 'Possible drowning/treading'. In this new state, besides Rules I and II, the swimmer condition will be further evaluated using Rule III. Only when all three rules are satisfied, the swimmer is considered to be in the state of 'Drowning'. Alarms will be triggered after the swimmer remains in the 'Drowning' state for a period longer than a preset duration. In each state the evaluation of a swimmer's condition can be formulated as a sequential hypothesis testing on an observation sequence of suitable swimmer features O = {o1o2 ...oT }. At an instance t, the observation ot is assumed to be generated according to two possible underlying feature distributions, Pw1 and Pw2 . Referring to Figure 6, when the swimmer is in 'Normal' state, Pw1|1 and Pw2|1 correspond to the distributions of moving speed and elongation measure for normal swimmers and drowning swimmers or swimmers treading water, respectively. While the swimmer is in 'Possible Drowning/Treading' state, Pw1|2 and Pw2|2 become the distributions of size variation for drowning swimmers and swimmers treading water, respectively. By using suitable rules at different states, if we can decide from which underlying distribution the observation sequence of swimmer features has been generated, we can determine the swimmer condition accordingly. Furthermore, to assist lifeguards to perform timely rescue, it is highly desirable to detect as soon as possible the moment when a swimmer lapses into drowning situation. To achieve this, the system has to quickly detect the change of the underlying distribution corresponding to either a change of the swimmer condition (Figure 6: State transition type 1) or a confirmation that the swimmer has lapsed into a drowning situation .

4.3. Sequential Detection of State Transition

The detection of abrupt change in swimmer condition can be formulated as sequential hypothesis testing between a hypothesis H1(without abrupt change) and a composite hypothesis H2 (with abrupt change at time tc), where H1 : ot ∼ Pw1 for 1 ≤ t ≤ T H2 : ot ∼ Pw1 for 1 ≤ t ≤ tc − 1 ot ∼ Pw2 for tc ≤ t ≤ T To detect the abrupt change time tc as soon as possible while minimizing false detection, we adopt a modified cumulative sum (CUSUM) algorithm [3], which examines the log-likelihood ratio of an observation sequence of swimmer features O = {o1o2 ...oT } and detects the existence of the abrupt change at tc = min{t : Λt > THD}, (5) where Λt = max {Λt−1 + log(Pw2 (ot)/Pw1 (ot), 0}, Λ0 = 0, and THD is a threshold which is set based on the priors of the underlying hypotheses and associated decision costs. Figure 7 shows one example of detecting 'State transition type 1'.

5. DETECTION RESULTS AND CONCLUSION

In developing the proposed system we faced sheer difficulty in collecting video data of real drowning incidents. In our experiments, we invited three volunteers to simulate the motion of drowning swimmers recorded in a documentary video 'On Drowning' [10] (which has been used by the American Red Cross to train their lifeguards) under the guidance of a professional lifeguard. In total, we captured more than 30 minutes video data of simulated drowning, which were processed at about 8 frames/second. We manually archived the data into 15 clips with the length ranging from 700 frames to 1200 frames. In each clip, there is only one drowning incident. Besides drowning simulation, we also captured about 45 minutes video of the volunteers swimming in different swimming styles and playing various water antics (including treading and splashing water). The video data were all captured from a standard teaching pool and partitioned into training and test sets. The training set was used to model the distributions of swimmer features under different conditions (i.e., Pw1 and Pw2 in different states, as shown in Figure 6), while the test set was used to test the system's ability in detecting possible drowning incidents while minimizing false alarms. Table 1 shows the overall performance of the proposed system. The table shows that the system is able to detect all the simulated drowning incidents and does not make any false detection for normal swimmers in the independent test set. On the other hand, the system makes two false detections for swimmers treading water. These false detections are caused by swimmers kicking water violently. Nevertheless, normal swimmers usually do not exhibit this type of motion characteristic, which resembles that of drowning swimmers. In sum, the proposed system makes a novel attempt to evaluate swimmers' conditions by analyzing their motion and shape features, and provides solutions to several key problems in detecting drowning incidents. While challenging in many aspects, a successful system will bring inestimable value in saving human lives.

Event Category	No of events	false	missed
Simulated drowning	8	0	0
Normal swimming	44	0	0
Treading water	17	2	0

Table 1. Drowning detection performance of the proposed system.