HUMAN SAFETY DEDUCTION IN COASTAL AND SEA SHORE AREAS
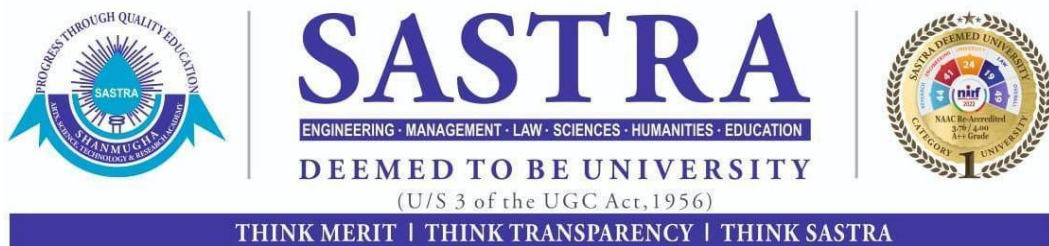
Report submitted to the SASTRA Deemed to be University
as the requirement for the course

CSE300 – MINI PROJECT

Submitted by

HEMACHANDRAN G
(Reg. No: 225003049, CSE
VIGNESH M
(Reg. No: 225003202, CSE
SURYA C
(Reg. No: 225003160, CSE)

MAY 2024

Department of Computer Science and Engineering

SRINIVASA RAMANUJAN CENTER

KUMBAKONAM,TAMIL NADU,INDIA-612001

Department of Computer Science and Engineering

SRINIVASA RAMANUJAN CENTER

KUMBAKONAM, TAMIL NADU, INDIA-612001

Bonafide Certificate

This is to certify that the report titled "HUMAN SAFETY DEDUCTION IN COASTAL AND SEA SHORE AREAS" submitted as a requirement for the course, **CSE300: MINI PROJECT** for B.Tech. is a bonafide record of the work done by **Mr. Hemachandran G(Reg. No. 225003049)**, **Mr. Vignesh M(Reg. No. 225003202)** and**Mr. Surya C(Reg. No. 225003160)** during the academic year 2024, in the Srinivasa Ramanujan Centre, under my supervision.

Signature of Project Supervisor   :

Name with Affiliation   :Dr. Sreedevi. B /Sr. AP/CSE/SRC/SASTRA

Date   :

Mini Project *Viva voce* held on _____

Examiner 1   **Examiner 2**

Acknowledgements

Table of Contents

List of Figures

Abbreviations

PIR   Passive Infrared Sensor

GSM   Global System for Mobile Communication

GPS   Global Positioning System

SMS   Short Message Service

ABSTRACT

In this project, an Arduino microcontroller will be used to construct a camera-based human activity detection system in a beach environment. For motion detection, the system combines a camera module with a passive infrared (PIR) sensor. When the Arduino detects human movement, it sets off an audible alert system and uses a GSM module to send the administrator real-time SMS messages with location data gleaned from a GPS module. Image processing, motion detection, and communication technologies work together to provide quick and effective beach area monitoring that improves security and makes it easier to respond quickly to possible situations. It is appropriate for deployment in beach surveillance applications because of its small size and waterproof construction, which guarantee durability in outside conditions.

CHAPTER 1

SUMMARY OF THE BASE PAPER

Title:"HUMAN SAFETY DEDUCTION IN COASTAL AND SEA SHOREAREAS"

Journal Name :IEEE TRANSACTIONS ON HUMAN-MACHINE SYSTEMS

Publisher: Parham Hadikhani

Year: 2023

SUMMARY :

The project presents a novel approach to enhancing beach surveillance capabilities by incorporating cutting-edge technologies. The system integrates a wide range of hardware components, such as camera modules, Passive Infrared (PIR) sensors, Global System for Mobile (GSM) modules, and Global Positioning System (GPS) modules, with Arduino microcontrollers serving as the primary processing units. The main goal is to create a small, waterproof device that can quickly identify people using beaches, set off alarms, and notify administrators in real time, all of which would improve security.

The majority of beach surveillance systems already in place rely on traditional techniques like CCTV cameras and sporadic security guard patrols. However, the lack of real-time detection and response capabilities in these systems frequently results in a delay in intervening in the event of security incidents. The suggested system aims to provide proactive monitoring and quick response mechanisms by utilising cutting-edge technologies.

The technique is a comprehensive strategy that includes software and hardware development. The project involves putting together the necessary hardware, which includes power supply units, weatherproof enclosures, connectivity cables, mounting hardware, PIR sensors for motion detection, Arduino microcontrollers, camera modules for visual confirmation, GSM modules for communication, GPS modules for location tracking, and power supply units. Every part is essential to the system's overall longevity and effectiveness, which guarantees that it may be used in outdoor settings.

A key component of the project is the software, namely the firmware development for the Arduino microcontrollers. The firmware is in charge of processing data from the numerous sensors, coordinating the integrated system's operations, and triggering the proper actions in response to observed human activity. Furthermore, software development entails

applying image processing algorithms—mostly with Matlab—to recorded video in order to precisely detect human activities. This feature of the system improves its capacity to reliably and accurately discern relevant activity from other motion sources.

The activity detection algorithm, a crucial part of the suggested system, is made to examine recorded photos or video frames and identify instances of human activity. The system uses motion detection techniques to distinguish between motion coming from waves or animals and important human activities. Furthermore, object tracking and detection image processing methods improve the system's ability to precisely identify and categorise human behaviour. The system's monitoring capabilities are based on this strong algorithmic architecture, which also allows the system to function well in a variety of beach environments.

The system includes both local and remote notification features as part of its alert mechanism. A buzzer that detects human activity sounds an audible signal, drawing attention to the area and maybe discouraging intruders. In order to enable remote alerts and notifications for quick responses to security incidents, the system also uses GSM modules to transmit real-time SMS messages to designated administrator numbers. Most importantly, GPS modules supply precise position data that is integrated into the SMS alerts, allowing administrators to precisely identify the site of detected activity and efficiently coordinate reaction actions. Extensive testing and optimisation are carried out during the development process to confirm the functioning and dependability of the system. The system's performance is tested in simulated beach areas under various situations and scenarios. Additionally, debugging and troubleshooting activities are used to quickly address any hardware or software problems that arise during testing. The completed system is installed at the intended beach location after testing and optimisation have been completed successfully. Proper installation and weatherproofing techniques are then put in place to guarantee the system's endurance and robustness in outdoor settings.

The system's total effectiveness is further enhanced by integration with existing beach surveillance infrastructure or security protocols, which facilitates smooth coordination and compatibility with current systems and procedures. To guarantee a seamless deployment and continuous administration, administrators and security staff receive training and instructions on system operation, maintenance, and response protocols.

The system's performance is continuously monitored and maintained using specified processes. Periodic evaluations are carried out to detect and swiftly resolve any malfunctions or maintenance needs. To guarantee the security and safety of beachgoers, administrators continuously monitor incoming SMS notifications and promptly respond to any alerts or events. This proactive approach to upkeep and monitoring contributes to maintaining the system's effectiveness and dependability over time.

CHAPTER 2

MERITS AND DEMERITS OF THE BASE PAPER

2.1 Merits:

1. Enhanced Security: By facilitating the quick identification of human activity and the timely reaction to possible security problems, the incorporation of cutting-edge technology like cameras, PIR sensors, GPS modules, and GSM modules enhances the security of beach areas. By taking the initiative, beachgoers' general safety can be increased and invaders can be discouraged.

2. Real-time Notifications: When the system detects human activity, it may send administrators real-time SMS notifications. This feature gives timely alerts that enable prompt intervention and response. By ensuring that security staff can promptly evaluate and handle security threats, this feature reduces the possibility of an escalation.

3. Accurate Location Tracking: The system's GPS modules allow for precise location tracking of detected activities, which makes it easier to coordinate response actions effectively. By determining the precise location of incidents, administrators can expedite the deployment of personnel and resources to the affected locations.

4. Integration with Current Infrastructure: The suggested method can be easily incorporated to improve overall surveillance capabilities with current security procedures or beach monitoring infrastructure. Instead than necessitating a total redesign of current systems, this interoperability guarantees that the system strengthens and complements current security measures.

2.2 Demerits:

1. Cost: The suggested system's implementation might come with a hefty upfront price tag, which would include buying hardware like cameras, sensors, Arduino microcontrollers, and communication modules. The total cost of deployment may also be increased by continuing maintenance and monitoring costs.

2. Complexity: During the implementation stage, there may be difficulties due to the intricacy of integrating various hardware components and creating software algorithms for activity detection and notification systems. It might take sufficient technical know-how and experience to set up and optimise the system for certain beach conditions.
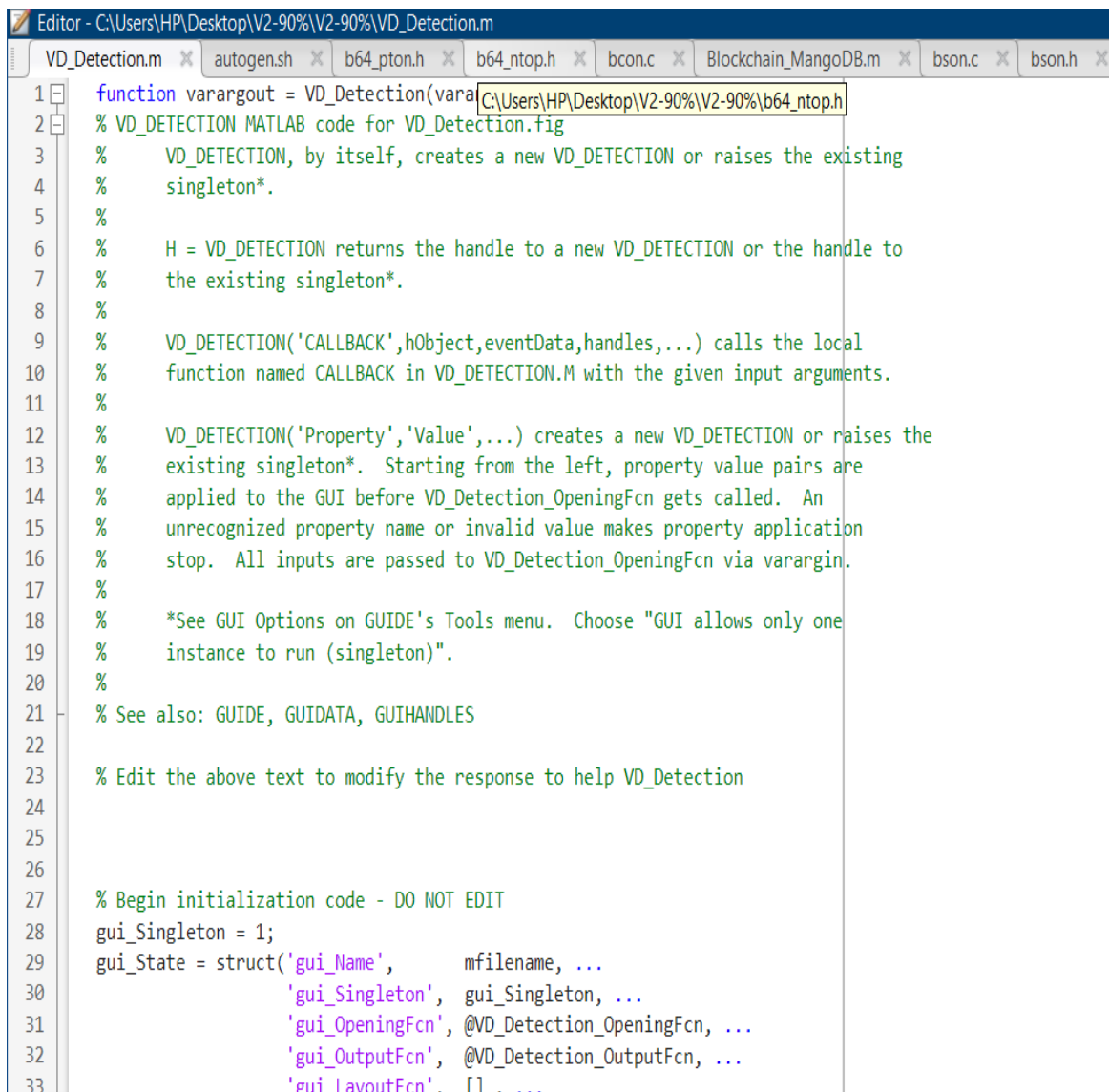
3. False Alarms: When human activity is detected by motion detection algorithms, non-human sources of motion, such animals or environmental conditions, may set out false alarms. This may result in needless disruptions and an increase in the burden for security staff members who are in charge of handling alarm responses.

4. Dependency on Cellular Networks: The system requires consistent cellular network coverage in the beach area in order to transmit SMS notifications, which are dependent on GSM modules. Connectivity problems can occur in remote or underserved areas, jeopardising the system's capacity to notify administrators in real time.

5. Privacy Concerns: The installation of surveillance equipment in public areas gives rise to worries about the gathering and archiving of personal information. It is imperative to establish appropriate mechanisms and measures to guarantee adherence to data protection requirements and minimise any hazards to beachgoers' privacy.

CHAPTER 3

SOURCE CODE

```
Editor - C:\Users\HP\Desktop\V2-90%\V2-90%\VD_Detection.m

VD_Detection.m  ✕   autogen.sh  ✕   b64_pton.h  ✕   b64_ntop.h  ✕   bcon.c  ✕   Blockchain_MangoDB.m  ✕   bson.c  ✕   bson.h  ✕

 1 ⊟      function varargout = VD_Detection(vara C:\Users\HP\Desktop\V2-90%\V2-90%\b64_ntop.h
 2 ⊟      % VD_DETECTION MATLAB code for VD_Detection.fig
 3        %      VD_DETECTION, by itself, creates a new VD_DETECTION or raises the existing
 4        %      singleton*.
 5        %
 6        %      H = VD_DETECTION returns the handle to a new VD_DETECTION or the handle to
 7        %      the existing singleton*.
 8        %
 9        %      VD_DETECTION('CALLBACK',hObject,eventData,handles,...) calls the local
10        %      function named CALLBACK in VD_DETECTION.M with the given input arguments.
11        %
12        %      VD_DETECTION('Property','Value',...) creates a new VD_DETECTION or raises the
13        %      existing singleton*.  Starting from the left, property value pairs are
14        %      applied to the GUI before VD_Detection_OpeningFcn gets called.  An
15        %      unrecognized property name or invalid value makes property application
16        %      stop.  All inputs are passed to VD_Detection_OpeningFcn via varargin.
17        %
18        %      *See GUI Options on GUIDE's Tools menu.  Choose "GUI allows only one
19        %      instance to run (singleton)".
20        %
21        % See also: GUIDE, GUIDATA, GUIHANDLES
22
23        % Edit the above text to modify the response to help VD_Detection
24
25
26
27        % Begin initialization code - DO NOT EDIT
28        gui_Singleton = 1;
29        gui_State = struct('gui_Name',       mfilename, ...
30                           'gui_Singleton',  gui_Singleton, ...
31                           'gui_OpeningFcn', @VD_Detection_OpeningFcn, ...
32                           'gui_OutputFcn',  @VD_Detection_OutputFcn, ...
33                           'gui_LayoutFcn',  [] , ...
```

5

```matlab
32                      'gui_OutputFcn',  @VD_Detection_OutputFcn, ...
33                      'gui_LayoutFcn',  [] , ...
34                      'gui_Callback',   []);
35      if nargin && ischar(varargin{1})
36          gui_State.gui_Callback = str2func(varargin{1});
37      end
38
39      if nargout
40          [varargout{1:nargout}] = gui_mainfcn(gui_State, varargin{:});
41      else
42          gui_mainfcn(gui_State, varargin{:});
43      end
44      % End initialization code - DO NOT EDIT
45
46
47      % --- Executes just before VD_Detection is made visible.
48      function VD_Detection_OpeningFcn(hObject, eventdata, handles, varargin)
49      % This function has no output args, see OutputFcn.
50      % hObject    handle to figure
51      % eventdata  reserved - to be defined in a future version of MATLAB
52      % handles    structure with handles and user data (see GUIDATA)
53      % varargin   command line arguments to VD_Detection (see VARARGIN)
54
55      % Choose default command line output for VD_Detection
56      handles.output = hObject;
57      axes(handles.axes1); axis off
58      axes(handles.axes2); axis off
59      set(handles.edit1,'String','**');
60      % Update handles structure
61      guidata(hObject, handles);
62
63      % UIWAIT makes VD_Detection wait for user response (see UIRESUME)
64      % uiwait(handles.figure1);
```

6

VD_Detection.m ✕ | autogen.sh ✕ | b64_pton.h ✕ | b64_ntop.h ✕ | bcon.c ✕ | Blockchain_MangoDB.m ✕ | bson.c ✕ | bson.h ✕ | +

```matlab
67     % --- Outputs from this function are returned to the command line.
68     function varargout = VD_Detection_OutputFcn(hObject, eventdata, handles)
69     % varargout  cell array for returning output args (see VARARGOUT);
70     % hObject    handle to figure
71     % eventdata  reserved - to be defined in a future version of MATLAB
72     % handles    structure with handles and user data (see GUIDATA)
73
74     % Get default command line output from handles structure
75     varargout{1} = handles.output;
76
77
78     % --- Executes on button press in pushbutton1.
79     function pushbutton1_Callback(hObject, eventdata, handles)
80     % hObject    handle to pushbutton1 (see GCBO)
81     % eventdata  reserved - to be defined in a future version of MATLAB
82     % handles    structure with handles and user data (see GUIDATA)
83     count=1;
84     locate_points=65;
85     % w = webcam(1);
86     for c = 1:count
87         url = 'http://192.168.210.195/cam-lo.jpg';
88     img=imread(url);
89         axes(handles.axes1); imshow(img) % display the image
90         title('Input Info','fontsize',12,'fontname','Times New Roman','color','Black')
91         f1=il_rgb2gray(double(img));
92     [ysize,xsize]=size(f1);
93     nptsmax=locate_points;
94     kparam=0.04;
95     pointtype=1;
96     sxl2=4;
97     sxi2=2*sxl2;
98     [posinit,valinit]=STIP(f1,kparam,sxl2,sxi2,pointtype,nptsmax);
99     Test Feat(c 1:locate points) valinit:
```

Command Window

```matlab
        [posinit,valinit]=STIP(f1,kparam,sxl2,sxl2,pointtype,nptsmax);
        Test_Feat(c,1:locate_points)=valinit;
          axes(handles.axes2); imshow(f1,[]), hold on
      axis off;
      showellipticfeatures(posinit,[1 1 0]);
      title('Segmented Info with Points','fontsize',12,'fontname','Times New Roman','color','Black')
      posinit=mean(posinit)
      set(handles.edit1,'String','+++++');
    end


    % --- Executes on button press in pushbutton2.
    function pushbutton2_Callback(hObject, eventdata, handles)
    % hObject    handle to pushbutton2 (see GCBO)
    % eventdata  reserved - to be defined in a future version of MATLAB
    % handles    structure with handles and user data (see GUIDATA)
    count=80;
    locate_points=65;
    % w = webcam(1);
    for c = 1:count
            url = 'http://192.168.210.195/cam-lo.jpg';
    img=imread(url);
    imwrite(img,'23.png')
        axes(handles.axes1);imshow(img) % display the image
          title('Input Info','fontsize',12,'fontname','Times New Roman','color','Black')
      f1=il_rgb2gray(double(img));
      [ysize,xsize]=size(f1);
      nptsmax=locate_points;
      kparam=0.04;
      pointtype=1;
      sxl2=4;
      sxi2=2*sxl2;
      [posinit,valinit]=STIP(f1,kparam,sxl2,sxi2,pointtype,nptsmax);
      Test_Feat(c,1:locate_points)=valinit;
```

VD_Detection.m ✕   autogen.sh ✕   b64_pton.h ✕   b64_ntop.h ✕   bcon.c ✕   Blockchain_MangoDB.m ✕   bson.c ✕   bson.h ✕   **+**

```matlab
130        Test_Feat(c,1:locate_points)=valinit;
131         axes(handles.axes2);imshow(f1,[]), hold on
132       axis off;
133       showellipticfeatures(posinit,[1 1 0]);
134      title('Segmented Info with Points','fontsize',12,'fontname','Times New Roman','color','Black')
135       posinit=mean(posinit)
136
137       if((posinit(1)<250 && posinit(1)>150))
138      tts('VIOLATION DETECTED');
139      % msgbox('VIOLATION DETECTED');
140      set(handles.edit1,'String','VIOLATION DETECTED');
141       end
142
143      end
144
145      % --- Executes on button press in pushbutton3.
146      function pushbutton3_Callback(hObject, eventdata, handles)
147      % hObject    handle to pushbutton3 (see GCBO)
148      % eventdata  reserved - to be defined in a future version of MATLAB
149      % handles    structure with handles and user data (see GUIDATA)
150      axes(handles.axes1); cla(handles.axes1); title(''); axis off
151      axes(handles.axes2); cla(handles.axes2); title(''); axis off
152       set(handles.edit1,'String','--');
153
154      % --- Executes on button press in pushbutton4.
155      function pushbutton4_Callback(hObject, eventdata, handles)
156      % hObject    handle to pushbutton4 (see GCBO)
157      % eventdata  reserved - to be defined in a future version of MATLAB
158      % handles    structure with handles and user data (see GUIDATA)
159      close
160
```

9

```matlab
160
161
162
163    function edit1_Callback(hObject, eventdata, handles)
164    % hObject     handle to edit1 (see GCBO)
165    % eventdata   reserved - to be defined in a future version of MATLAB
166    % handles     structure with handles and user data (see GUIDATA)
167
168    % Hints: get(hObject,'String') returns contents of edit1 as text
169    %        str2double(get(hObject,'String')) returns contents of edit1 as a double
170
171
172    % --- Executes during object creation, after setting all properties.
173    function edit1_CreateFcn(hObject, eventdata, handles)
174    % hObject     handle to edit1 (see GCBO)
175    % eventdata   reserved - to be defined in a future version of MATLAB
176    % handles     empty - handles not created until after all CreateFcns called
177
178    % Hint: edit controls usually have a white background on Windows.
179    %        See ISPC and COMPUTER.
180    if ispc && isequal(get(hObject,'BackgroundColor'), get(0,'defaultUicontrolBackgroundColor'))
181        set(hObject,'BackgroundColor','white');
182    end
183
```

```
1   void setup()
2   {
3   Serial.begin(9600);
4   pinMode(8,OUTPUT);
5   pinMode(7,INPUT);
6
7   }
8
9   void loop()
10  {
11  int PIR=digitalRead(7);Serial.print("PIR:");Serial.println(PIR);delay(1000);
12
13  if(PIR==1)
14  {
15    digitalWrite(8,1);delay(5000);digitalWrite(8,0);
16
17    Serial.println("AT");
18    delay(2000);
19    Serial.println("AT+CMGF=1");
20    delay(2000);
21    Serial.println("AT+CMGS=\"+918056357169\"\r");
22    delay(2000);
23    Serial.println("EMERGENCY ALERT");
24    delay(1500);
25    Serial.println((char)26);
26    delay(2000);
27  }
28  }
29
```

CHAPTER 4

SYSTEM IMPLEMENTATION

4.1 Existing System

To monitor human activity, current beach surveillance systems usually use a mix of security guards, CCTV cameras, and sporadic patrols. However, the ability of these systems to detect and respond in real-time is frequently limited. Conventional CCTV systems could not be nimble enough to quickly detect any incidents and notify the appropriate authorities, which could cause responses to be delayed. Not only that, but communicating immediate notifications to administrators is difficult in the absence of integrated communication tools. When manual surveillance methods are used, there may be gaps in the monitoring, which exposes certain areas to security risks.

Beach monitoring requires security people, but their efficiency is restricted by things like human error and a small coverage area. Patrols can serve as a visual deterrent, but their ability to cover all locations at once is constrained by their manpower requirements. CCTV cameras can record video for subsequent analysis and serve as a passive kind of surveillance, but they might not always deliver real-time alerts that require quick response. Moreover, blind spots could develop from relying solely on fixed camera locations, which would enable occurrences to go unreported.

To identify unwanted entry or infiltration, some current beach surveillance systems include motion detection sensors or alerts. However, environmental factors like wind-blown trash or wildlife movement can cause false alarms from these systems, which would reduce their reliability. Furthermore, these systems' alarms might not reach administrators quickly due to their lack of integrated communication capabilities, which would hinder response attempts.

All things considered, the state of beach surveillance today is defined by a dependence on antiquated techniques that might not be sufficient to handle the ever-changing nature of security risks. The efficacy of security guards, CCTV cameras, and motion sensors in preventing and identifying incidents is restricted by variables including reaction time, coverage area, and communication capacities. Therefore, to improve real-time monitoring,

communication, and response capabilities in beach surveillance applications, creative solutions utilising cutting-edge technologies are required.

4.2 Proposed System

By integrating cutting-edge technology, the suggested system seeks to transform beach area surveillance and provide a holistic solution that would improve security measures and allow for rapid responses. An Arduino microcontroller is at the heart of it all, acting as the main controller for all of the integrated system's functions. This architecture, which is based on Arduino, makes it easy to interface with different hardware elements, guaranteeing quick and effective beach area monitoring.

In order for the system to function properly, a camera module and a passive infrared (PIR) sensor must be included. The camera module is essential to the process of taking pictures or videos of the beach area since it allows for the visual verification of any human activity that is identified. When used in conjunction with advanced image processing algorithms, the camera allows for the accurate identification of pertinent activities within the ever-changing beach scene. Concurrently, the PIR sensor functions as the motion detection element of the system, identifying infrared radiation released by objects in motion, especially people. Through the use of these technologies, the system is able to quickly recognize possible security risks, such unauthorized access or unusual activity, and can take immediate action to address them.

The device sounds a buzzer to initiate a localized alarm mechanism that serves as an on-site warning system to draw attention to the area and discourage trespassers when it detects human movement. Concurrently, the system makes use of a GSM module to send administrators real-time SMS messages, guaranteeing prompt reporting of security events. Most importantly, these alerts contain precise location data gleaned from a GPS module, which lets managers know precisely where the observed activity is. By integrating GPS technology, authorities can better coordinate their reaction and allocate resources to handle security problems in an efficient and effective manner.

The system's small size and waterproof construction further improve its appropriateness for outdoor deployment near beaches. Sturdy, weatherproof enclosures shield the electronic components from elements like wind, rain, and UV rays, guaranteeing their dependability and longevity under harsh circumstances. Furthermore, the power supply of the system—whether it comes from solar or batteries—ensures uninterrupted operation without relying on outside power sources, improving autonomy and resilience.

The system's software development uses advanced algorithms for motion detection and picture processing in addition to firmware for the Arduino microcontroller. These

algorithms can discriminate between relevant human activity and other sources of motion, such waves or animals, by analyzing collected photos or video frames. The system strives to achieve high accuracy and timeliness in identifying security risks while reducing false alarms by optimizing algorithms and fine-tuning settings.

The system must be carefully installed and configured for seamless interaction with any current beach surveillance infrastructure or security measures as part of the deployment and integration process. Administrators and security staff receive the necessary training and instructions to guarantee efficient operation, maintenance, and response protocols. Continuous monitoring and maintenance procedures are designed to evaluate the system's performance on a frequent basis and fix any problems quickly in order to keep it operating at peak efficiency.

All things considered, the suggested method is a proactive and responsive strategy to beach area surveillance, utilizing state-of-the-art technologies to improve security protocols and guarantee the protection and welfare of beachgoers.

4.3 Inference

In the realm of safety and surveillance systems, various studies have explored methodologies and technologies aimed at detecting and responding to human activity in diverse environments. These investigations provide valuable insights into the development of effective surveillance solutions, including those applicable to beach area security.

- Amer, Humphries, and Nandeshwar [1] present a methodology for human position detection using point cloud data, focusing on ensuring safety in human-robot interaction systems. By employing advanced sensor technologies, their approach lays the foundation for real-time safety measures, which can be adapted to enhance beach area surveillance by detecting and responding to human presence effectively.

- In a similar vein, Han and Zeng [2] propose a deep learning-based approach for detecting safety helmet wearing on construction sites using multi-scale features. Their methodology leverages convolutional neural networks (CNNs) to analyze image data and identify instances of safety helmet wearing, showcasing the application of image processing algorithms for safety monitoring. This technique can be extrapolated to beach surveillance systems to detect unauthorized activities or individuals not adhering to safety regulations.

- Shi et al. [3] introduce a human fall detection algorithm based on YOLOX-s and Lightweight OpenPose, employing deep learning techniques to accurately detect fall events. Their research highlights the potential of computer vision algorithms in identifying abnormal human movements, which can be adapted to detect suspicious behavior or emergencies on beach premises.

- Du et al. [4] propose an active collision avoidance system for ensuring safety in human-manipulator interaction scenarios. By employing sensor technologies and control algorithms, their approach detects and prevents potential collisions between humans and robotic manipulators, thereby enhancing workplace safety. Such safety mechanisms can be integrated into beach area surveillance systems to mitigate security risks and ensure the safety of beachgoers.

- Zhang et al. [5] present a lightweight neural network-based system for real-time detection of PCB defects, utilizing machine learning algorithms for automated quality control. Their approach demonstrates the application of real-time detection algorithms for identifying anomalies, which can be adapted for detecting suspicious behavior or security threats in beach areas.

4.3 Methodology

➤ **Hardware Setup:**

- Put together the required hardware, such as the GPS, GSM, PIR, camera, and Arduino microcontroller modules.
- Verify that, in accordance with the manufacturer's instructions, the parts are firmly attached to the Arduino microcontroller.

➤ **Software Development:**

- Create the firmware for the Arduino microcontroller, which is used to manage the integrated system's functions.
- Put code into place to configure and communicate with the camera module in order to take pictures or record videos.
- Configure the PIR sensor to recognize motion inside the specified area and initiate appropriate action.
- Create image processing techniques or integrate libraries to recognize human activity in recorded video.
- Set up the GSM module to deliver SMS notifications and connect to the cellular network.
- Get location information using the GPS module and add it to the SMS notifications.

➤ **Activity Detection Algorithm:**

- Create an algorithm to recognize human activity in pictures or video frames that have been taken.
- Use motion detection methods to discern between pertinent human activities and other motion sources, such waves or animals.
- To improve the precision of human activity recognition, apply object detection and tracking image processing methods.
- .

➢ **Alert Mechanism:**

- Create a system that, when it senses human activity, uses the buzzer to sound an alarm.
- Put logic in place to turn on the GSM module and send instant SMS alerts to pre-configured administrator numbers.
- For an efficient reaction, incorporate pertinent data into the SMS notifications, such as the timestamp and GPS coordinates of the observed activity.

➢ **Testing and Optimization :**

- To confirm the integrated system's dependability and functionality, thoroughly test it in a beach simulation setting.
- Adjust settings and algorithms to maximize human activity detection's responsiveness and accuracy.
- Take care of any possible software or hardware problems that come up during the testing and debugging stages.

➢ **Deployment and Integration:**

- Install the completed system at the desired beach location, making sure that all necessary weatherproofing measures and installation are in place.
- To improve total surveillance capabilities, integrate the system with the security protocols or beach surveillance equipment that is already in place.
- Administrators and security staff should receive education and training on system upkeep, response protocols, and operation.

➢ **Monitoring and Maintenance:**

- Create a monitoring procedure to evaluate the system's performance on a regular basis and quickly handle any maintenance or operational problems.
- In order to protect beachgoers' safety and security, keep an eye on incoming SMS notifications and promptly respond to any alerts or occurrences.

4.4 Module Description

1. Arduino Microcontroller: - Overview: Acting as the system's brain, the Arduino microcontroller manages and synchronizes the different parts. It gathers data from the GPS, GSM, PIR, and camera modules and uses it to trigger actions when it detects human activity.

2. Camera Module: - Overview: This module records pictures or videos of the beachfront. It is essential to the system's image processing component because it provides visual confirmation of any identified human movement.

3. Passive Infrared (PIR) Sensor: - Overview: PIR sensors pick up infrared radiation that is released by moving objects, including people. It performs the function of a motion detection component, alerting the system upon detection of human activity.

4. Buzzer: - Description: When the PIR sensor picks up human activity, the buzzer on-site sounds an alert. It functions as a warning system to draw attention to the area and discourage any trespassers.

5. GSM Module: - Overview: When human activity is recognized, the GSM module notifies managers via SMS in real time, facilitating communication. It makes remote notifications and alerts easier to use so that security incidents may be handled quickly.

6. GPS Module: - Overview: The SMS notifications contain precise position data that is provided by the GPS module. This guarantees that administrators obtain accurate information regarding the location of the observed activity, facilitating effective coordination of responses.

7. Power Supply: - Description: In outdoor settings, the components must have an appropriate power source to ensure continuous functioning. Examples of these are batteries and solar power systems. It is necessary for the independence and dependability of the system.

8. Weatherproof Enclosure: - Overview: This enclosure ensures the system's longevity in outside environments by shielding its components from the elements. It protects the electronics from elements including wind, rain, and UV rays.

9. Connectivity Cables and Wiring: Well arranged and insulated cables link the different parts of the system, guaranteeing smooth power distribution and communication.

10. Mounting Hardware: This part makes it easier to place the system in a permanent spot on the beach. For efficient surveillance, it makes sure the camera and sensors are stable and positioned correctly.
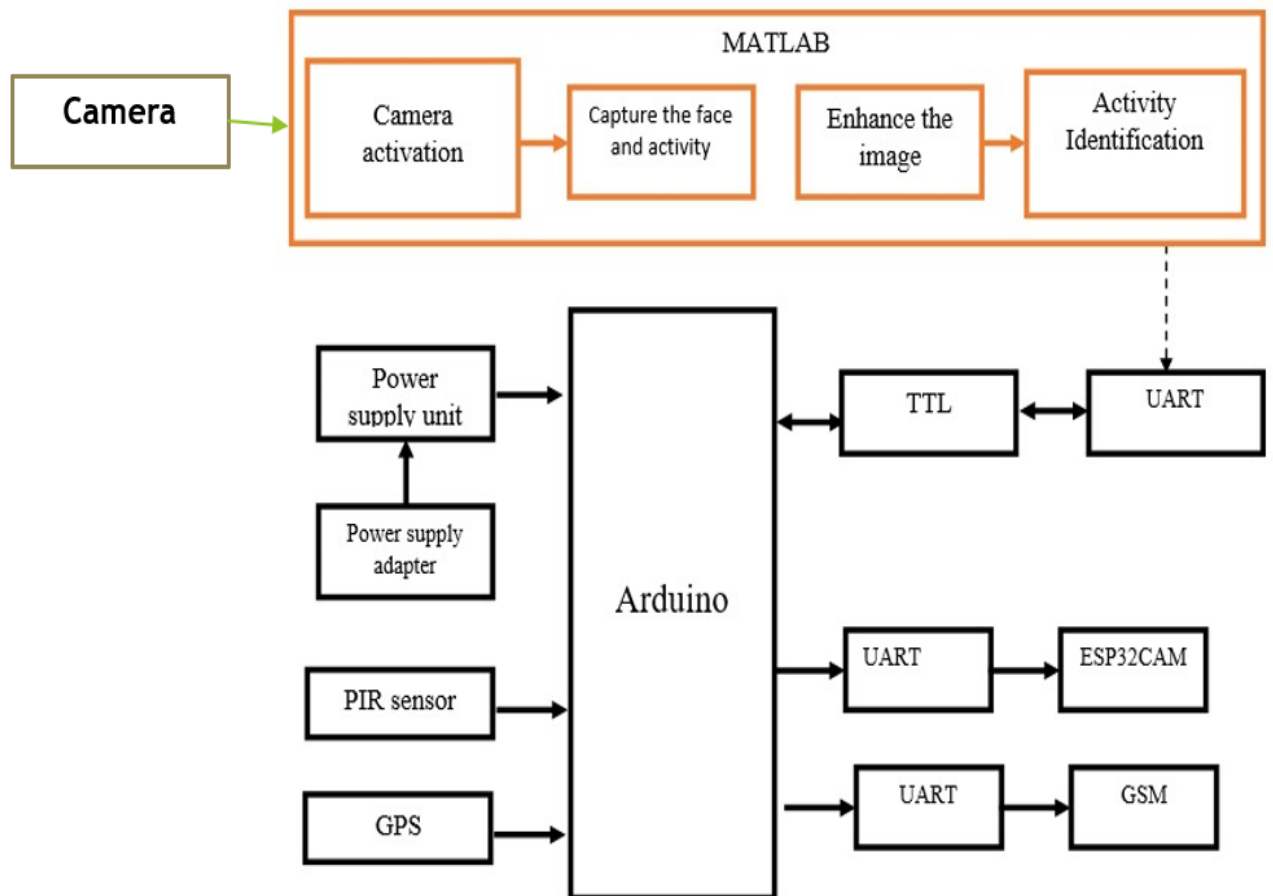
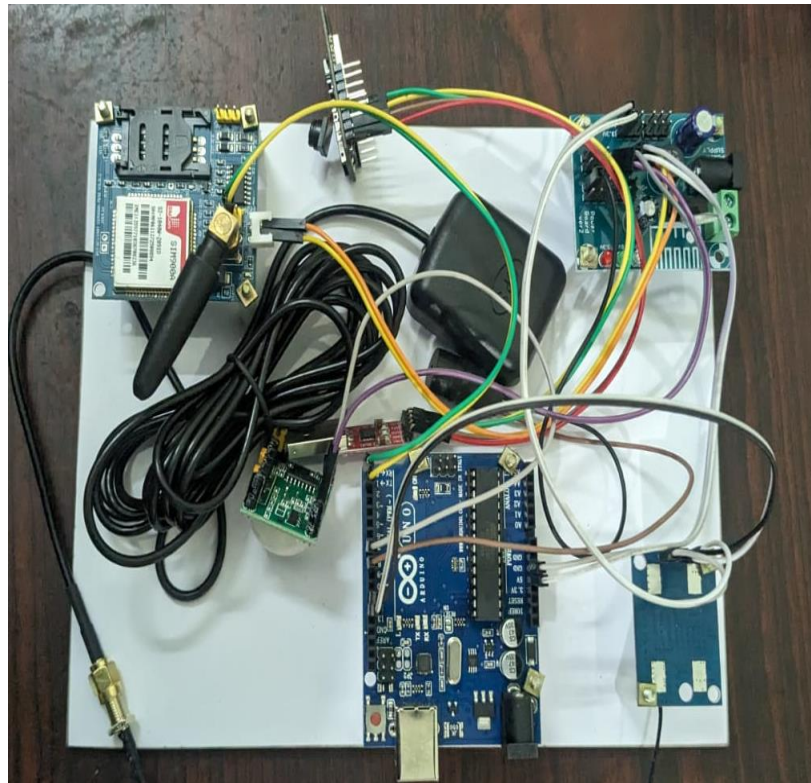## 4.5 Block Diagram

Fig 4.1 Block Diagram

Snapshots

Fig 4.2 Hardware setup

Fig 4.3 Software setup
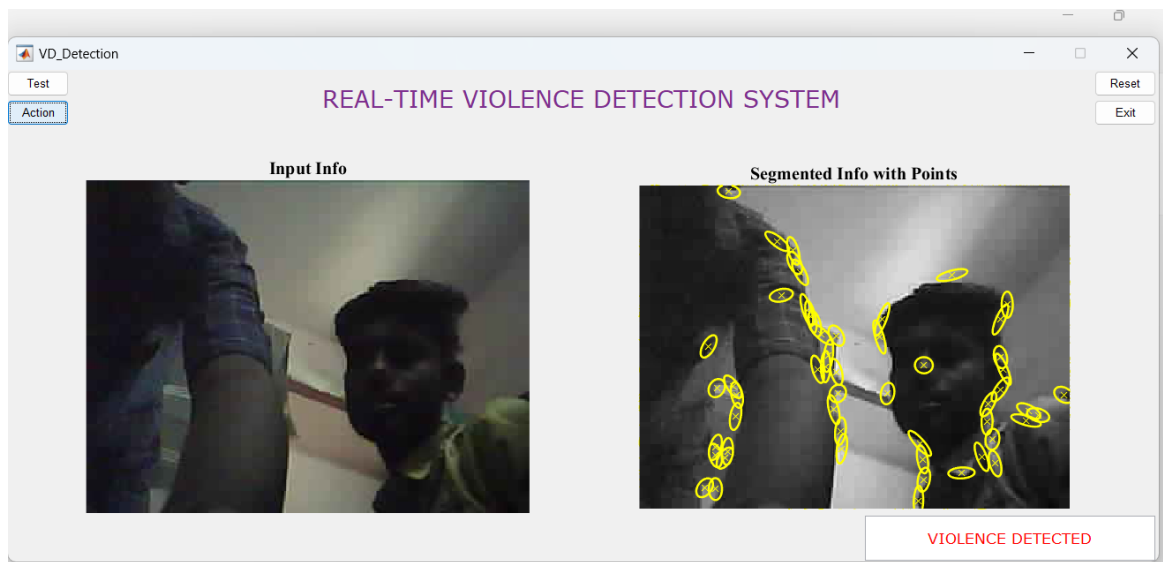
Output Screenshots

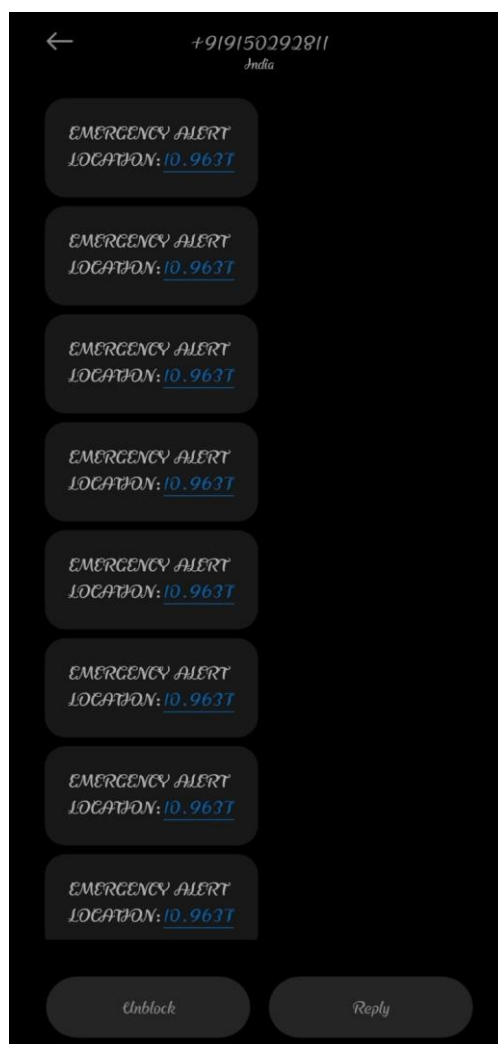Fig 4.4 Detection of motion

Fig 4.5 Response from GPS and GSM module

CHAPTER 5

CONCLUSION AND FUTURE PLANS

5.1 Conclusion:

To sum up, the suggested camera-based human activity detection system for beach surveillance is a big step forward in improving coastal areas' security and safety protocols. Modern technology like an Arduino microcontroller, a camera module, a PIR sensor, a GSM module, and a GPS module are all integrated into the system to provide a complete solution for quickly and effectively monitoring human activity on beaches.

The project adds real-time detection and reaction capabilities to overcome the shortcomings of the current beach monitoring systems. By combining image processing, motion detection, and communication technologies, administrators may instantly be notified when they see human activity. This allows them to take prompt action to protect beachgoers.

In addition, the system's weatherproof and small design guarantees its resilience in outdoor settings, which makes it appropriate for use in beach monitoring applications. Through the utilization of sophisticated sensors and an Arduino microcontroller, the system offers an affordable and expandable approach to beach security.

5.2 Upcoming Initiatives:

Going forward, there are other ways to enhance and grow the suggested beach surveillance system:

1. Enhanced Image Processing Algorithms: Ongoing research and development in image processing algorithms can reduce false positives and boost overall system performance by increasing the precision and dependability of human activity detection.

2. Artificial Intelligence Integration: By using AI methods like machine learning and deep learning, the system will be able to adjust to various beach conditions and become more adept at differentiating between false alerts and human activity.

3. Multi-Sensor Fusion: Adding other sensors, including environmental or acoustic ones, can provide further information for more thorough beach monitoring, which includes identifying odd noises or anomalies in the surroundings.

4. Remote Monitoring and Control: By putting in place remote monitoring and control features via a mobile application or web-based interface, administrators may keep an eye on activities at the beach in real-time and take preventative action when there are security issues.

5. Cooperation with Local Authorities: Working together with law enforcement, beach management, and local authorities can make it easier to integrate the system into current emergency response plans and security protocols, guaranteeing smooth coordination in the event of a security incident.

To sum up, the suggested beach surveillance system forms the basis for improving security protocols and guaranteeing beachgoers' safety. Communities may enjoy safer and more intelligent beach settings if we embrace innovation and continual development.

REFERENCES

1) N. Amer, J. Humphries and N. Nandeshwar, "Human Position Detection Using Point Cloud Data for Human-Robot Safety Systems," 2022 7th International Conference on Mechanical Engineering and Robotics Research (ICMERR), Krakow, Poland, 2022, pp. 219-223, doi: 10.1109/ICMERR56497.2022.10097801.

2) D. Xinvu, M. Yang, X. Zhihui and G. Haixia, "Human Performance Detection Using Historical Operation Data of Nuclear Power Plants," 2022 4th International Conference on System Reliability and Safety Engineering (SRSE), Guangzhou, China, 2022, pp. 167-173, doi: 10.1109/SRSE56746.2022.10067699.

3) E. Tavanti et al., "RFID-based Monitoring of Human Operators for Safety in Outdoor Working Sites," 2023 IEEE 13th International Conference on RFID Technology and Applications (RFID-TA), Aveiro, Portugal, 2023, pp. 205-208, doi: 10.1109/RFID-TA58140.2023.10290563.

4) G. Shen, B. Zhao, X. Chen, L. Liu, Y. Wei and T. Yin, "Human Fall Detection Based on Re-Parameterization and Feature Enhancement," in IEEE Access, vol. 11, pp. 133591-133606, 2023, doi: 10.1109/ACCESS.2023.3335833.

5) S. B. Selvaraj, R. Canale, T. Piriyatharawet, R. Xiao, P. Vyas and C. S. Horng, "Towards Safe and Efficient Human-Robot Collaboration: Motion Planning Design in Handling Dynamic Obstacles," IECON 2023- 49th Annual Conference of the IEEE Industrial Electronics Society, Singapore, Singapore, 2023, pp. 1-5, doi: 10.1109/IECON51785.2023.10311798.

A Novel Skeleton-Based Human Activity Discovery Using Particle Swarm Optimization With Gaussian Mutation

Parham Hadikhani [ORCID], Daphne Teck Ching Lai [ORCID], and Wee-Hong Ong [ORCID]

Abstract—Human activity discovery aims to cluster human activities without any prior knowledge of what defines each activity. However, most existing methods for human activity recognition are supervised, relying on labeled inputs for training. In reality, it is challenging to label human activity data due to its large volume and the diversity of human activities. To address this issue, this article proposes an unsupervised framework for human activity discovery in 3-D skeleton sequences. The framework includes a data preprocessing step that selects important frames based on kinetic energy and extracts relevant features, such as joint displacement, statistical displacement, angles, and orientation. To reduce the dimensionality of the extracted features, the framework uses principle component analysis. Unlike many other methods for human activity discovery, the proposed framework is fully unsupervised and does not rely on presegmented videos. To segment the time series of activities, the framework uses a sliding time window with some overlapping. The hybrid particle swarm optimization (PSO) with Gaussian mutation and K-means algorithm is then proposed to discover the activities. PSO is chosen for its powerful global search capability and simple implementation. To further improve the convergence rate of PSO, K-means is applied to the outcome centroids from each iteration of PSO. The experimental results on five datasets demonstrate that the proposed framework has superior performance in discovering activities compared to other state-of-the-art methods. The framework achieves an average increase in accuracy of at least 4%.

Index Terms—Clustering, dimension reduction, feature extraction, human activity discovery, particle swarm optimization (PSO), skeleton sequence, unsupervised learning.

I. INTRODUCTION

HUMAN activity recognition (HAR) has attracted much attention due to its applications in fields such as human–computer interaction, intelligent transportation systems [1], [2],

Fig. 1. Overview of HAR system: (a) Performed activities are (b) captured by a Kinect sensor. (c) After that, pose of humans are estimated by extracting joints. (d) To make raw data more usable, their salient and defining features are identified. (e) Based on the similarities and differences, activities are discovered. (f) Afterwards, the system begins to learn from the discovered activities and (g) finally human activities are recognized.

and monitoring applications [3]. Activity recognition aims to identify actions and activities that humans perform in different environments automatically. The input to a vision-based HAR system is a sequence of frames of a person performing different movements. The output is a set of labels representing the actions taken or activities in those movements. Many existing works use visual data as input. But such data have considerable complexity detrimental to the performance of HAR systems. These complexities include cluttered background, changes in brightness, and points of view. 3-D skeleton data partially overcomes these complexities and protects people's privacy when red, green, and blue (RGB) data is not captured. Each frame represented by 3-D coordinates of the main body joints is appropriate for representing human actions [4] and can be easily obtained in real-time with low-cost depth sensors [5].

As shown in Fig. 1, there are at least seven steps in vision-based HAR systems. Vision sensors capture activities performed by a person. The skeletal information comprising joints coordinate are extracted from captured videos, containing image sequences called frames. Meaningful features are then extracted

for more accurate activity discovery. The system without using manual annotations and having any guidance for activities discovers them by clustering the most similar activities from a set of different activities. In other words, the system tries to differentiate observed activities based on the likeness of extracted features. The discovered activity clusters are used in the learning process to model each cluster of activity and recognize future activities. Significant progress has been made in the supervised learning of activity models [6], illustrated in Fig. 1(f) and (g). The learning and recognition steps rely on human-labeled training data to categorize activities if activity discovery [block (e)] was not performed. Human activity discovery is a part of the HAR process where activities are categorized based on their similarities without any knowledge of activity labels or any information that characterizes an activity, making this step particularly challenging. In other words, activity discovery is like a child's learning. There is no prior information to define a specific sequence of movements to mean a particular activity such as crawling or waving and so forth to the child learner. Using the ability to differentiate, they learn from unlabeled data and form a model that can postlabel new data based on that training. In human activity discovery, there is no known information or knowledge about a particular movement, including its start to end, e.g., when someone is picking up something. This means the input is a series of movements without knowing the start and endpoints to indicate each activity. Some existing work has segmented the input data by activity [4]. Thus, the start and endpoints of the activities are already known, although the method of grouping activities may be unsupervised.

In this article, we focus on the less developed activity discovery comprising the block (d) to (e) of Fig. 1 by developing an effective methodology to extract good features and cluster activities without any label. Keyframe selection [7] and principle component analysis (PCA) are used to remove redundant frames and features to reduce time complexity and increase accuracy. A preprocessing and feature extraction methodology are proposed to prepare information and extract features from the most informative joints and bones, including joint displacement, joint orientation, and statistical time domain. As our first study, a hybrid PSO with Gaussian mutation and K-means (HPGMK) clustering that requires a known prior of the cluster number is proposed to find activities. Sometimes particles converge to a specific point between the best global and personal positions and get trapped in local optima. This difficulty arises when the swarm's variety reduces and the swarm cannot escape from a local optimum [8]. To address this, a hybrid PSO with Gaussian mutation is proposed to promote diversity to avoid early convergence. Then, K-means is applied to the centroids obtained by PSO to refine their location and get the best possible solution. Our methodology performs activity discovery using unsegmented input data and the proposed techniques used are unsupervised with no prior knowledge of the labels of the different activities. The main contributions of this article are as follows.

1) A methodology consists of keyframe selection, feature extraction, and PCA to represent human activities. The features based on displacement, statistical, and orientation

are extracted simultaneously to represent all the movement aspects of human activities. This makes the discovery part perform better because it has comprehensive information about the activities.

2) A hybrid clustering algorithm called HPGMK to discover and group unlabelled human activities observations into individual activity classes. PSO is customized by applying the Gaussian mutation, based on the advantages of two methods [9] and [10], on the global best's centroids to increase the diversity of selected clusters in the global best. Also, due to the increase in the variety of solutions, the proposed algorithm reaches the desired solution in a smaller number of iterations. That is why the clustering time is reduced.

3) Integrating K-means to refine the obtained cluster centers from the PSO to improve the exploitation of the algorithm. In PSO, when approaching the final solution, the speed of the particles decreases greatly and it becomes difficult to reach the best optimum solution. For this reason, the problem of PSO is solved by using the advantage of K-means in local search and applying it to the obtained solution from PSO.

The rest of this article is organized as follows. In this article, the background and related methods are discussed in Section II. The methodology is described in Section III. We present the evaluation of the proposed approach, comparing with state-of-the-art (SOTA) techniques in Section IV. Finally, Section V concludes this article.

II. RELATED WORKS

Feature extraction from 3-D skeletal human activities data: Skeletal data includes the number of joints, and each joint contains 3-D coordinates. Since the motion of the joints has essential information for any activity, feature extraction is vital. There are various methods for representation of the motion of skeletal joints such as calculating the difference between the joints in the same frame and the same joints in different frames [11], using histogram oriented of joints [12], dynamic time warping algorithm [13], covariance of 3-D joints [14], generating joint rotation matrix concerning the person's torso [15], and extracting the angles and orientations of the most informative body joints [16]. However, most of these methods extract one aspect of the skeletal data features, leading to other important aspects of activities being overlooked. As a result, there is a decrease in accuracy in the final result because of the insufficient discriminating ability of the extracted features. Moreover, due to the complexity of feature calculations, some of these methods cause computational latency. The difference between our article with previous works for extracting features is that We have combined the feature extraction techniques from [11], [16], and [17] to extract three skeletal features from informative joints and keyframes. Previously, these features have been applied separately and most of them have used all of the joints and frames, which increases additional information. This increases the time complexity and reduces identifying the activities' performance. Some methods like [7] and [18] have tried to select some frames that are more

27

distinguishing compared to other frames and remove redundant information with the assistance of kinetic energy. However, these methods are applied to each activity sample separately. In other words, in [7] and [18], keyframes were selected in a supervised manner. In contrast, we apply the above methods in our proposed method to select keyframes without knowing the beginning and end of the activities.

PSO clustering and dimension reduction: One of the important methods to do discovery is to use clustering. Clustering is a method that categorizes data points based on similarities and dissimilarities [19], [20]. One of the common methods of clustering is K-means. But it has problems such as poor convergence rate and local optimum. One of the ways to overcome these problems is to use evolutionary algorithms like PSO. PSO is a population-based algorithm where each individual represents a potential solution, which makes a better problem space search. PSO greatly reduces the possibility of getting stuck in local optimum due to using local search and global search simultaneously [21], [22]. To create a novel clustering algorithm, Malarvizhi et al. [23] combined PSO and feature linkage-based weight reduction. The method determined the weight of each feature using the Mahalanobis distance to choose the feature and do the clustering automatically. Sharma et al. [24] developed a hybrid PSO clustering for network-based sustainable computing. They applied the mutation operator to ensure diversity among the solutions to keep the algorithm's balance between exploration and exploitation. Rengasamy et al. [25] introduced a new memory dimension termed family memory and added to the two already existing ones of cognitive memory and social memory. This memory was used to collect the information from the particles that favor a certain cluster. Additionally, they utilized the K-means to initialize centroids for PSO clustering to enhance the traditional PSO. Cai et al. [26] proposed a new clustering method based on combining density peaks clustering with PSO. They employed a technique to compute density peaks to avoid falling into a local. They also presented a new fitness criteria function to optimally explore K cluster centers to obtain the optimal global solutions. However, these methods do not address all the weak points of PSO. Some of them either deal with the issue of reducing the speed of PSO when approaching the optimal solution or the issue of reducing the diversity among the solutions during the search. Different from the abovementioned methods, we overcome the weakness noted for PSO simultaneously by using Gaussian mutation and K-means. By employing Gaussian mutation, the variety of the solutions the PSO's capacity to exploit around potential solutions gets effectively enhanced. We also utilize K-means to solve the PSO problem: Its convergence speed slows down when it approaches the global optimum. We also use K-means to solve the problem of reducing the convergence speed of PSO when approaching the global optimum by using the fast speed of K-means in local search. Zhang et al. [27] introduced a feature selection method based on PSO that combined fuzzy clustering and feature importance (PSOFS-FC). They presented a new objective function based on F-measure and filling risk for PSO with fuzzy clustering to assess the impact of missing data in class imbalance. To overcome the dimensionality curse, Song et al. [28]

presented a three-phase feature selection technique based on correlation-guided clustering and PSO. First, they combined a filter approach and a feature clustering-based method to reduce the search space. Then, an enhanced integer PSO was used to select the best feature subset. Unlike PSOFS-FC and other mentioned methods, which combine clustering and feature selection, we reduce extracted features' dimensions before clustering. Although good results have been obtained in these methods, due to the simultaneity of feature selection and clustering, clustering becomes problematic in high-dimensional data and the clustering time increases greatly. In human activity discovery (HAD), spending time to discover activities is important because of its applications, such as use in security areas to identify suspicious behavior or in hospitals to check patients' status. For this reason, our proposed method reduces the feature dimensions before clustering by using PCA to make the data more clusterable. PCA speeds up the clustering algorithm by removing correlated features that do not influence decision-making. As a result, the algorithm's clustering time decreases dramatically with fewer features. Thus, not only the speed of clustering increases but also the accuracy of clustering increases because the clustering process is performed on highly important features. Regarding detecting outlier and noisy data, Hubert et al. [29] proposed a combination method to make PCA robust to outliers. They combined projection pursuit [30] with h robust covariance estimation in lower dimensions [31]. Moreover, they applied a diagnostic plot to detect the outliers. Candes et al. [32] proposed a technique to improve the performance of PCA. They used a low-rank and sparse component for PCA to avoid outliers and achieved good performance in the application of Alzheimer's Disease Recognition [33]. Rahmani et al. [34] presented a provable algorithm to identify the outliers based on PCA. For this reason, they employed a convex optimization problem to evaluate the data points based on the innovation search method. Despite the very good performance of the presented methods to improve PCA, they have more execution time than the original PCA. On the other hand, the focus of our work is on the improvement of feature extraction and discovery for human activities. As mentioned before, time is very important in HAD. That is why we use PCA to prevent the increase in the computation of time for the presented framework. It is worth mentioning that the obtained results show that PCA has reduced the dimensions of the features and improved HAD performance significantly.

Recognition and discovery of 3-D skeletal human activity: Many studies in HAR used supervised approaches [35], [36], [37]. Yadav et al. [38] combined long-short term memory networks and convolutional neural networks for recognizing human activity and fall detection. They used some handcrafted features, including geometrical and kinematic features to guide their proposed model. Zhang et al. [39] proposed an end-to-end semantics-guided neural networks framework. They provided two semantic forms based on joints and frames and used graph convolutional network (GCN) and convolutional neural network (CNN) layers to find the dependence of joints and frames, respectively. Si et al. [40] proposed a novel model based on a recurrent network. They applied a graph convolutional layer into the LSTM network to improve the performance

for more accurate activity discovery. The system without using manual annotations and having any guidance for activities discovers them by clustering the most similar activities from a set of different activities. In other words, the system tries to differentiate observed activities based on the likeness of extracted features. The discovered activity clusters are used in the learning process to model each cluster of activity and recognize future activities. Significant progress has been made in the supervised learning of activity models [6], illustrated in Fig. 1(f) and (g). The learning and recognition steps rely on human-labeled training data to categorize activities if activity discovery [block (e)] was not performed. Human activity discovery is a part of the HAR process where activities are categorized based on their similarities without any knowledge of activity labels or any information that characterizes an activity, making this step particularly challenging. In other words, activity discovery is like a child's learning. There is no prior information to define a specific sequence of movements to mean a particular activity such as crawling or waving and so forth to the child learner. Using the ability to differentiate, they learn from unlabeled data and form a model that can postlabel new data based on that training. In human activity discovery, there is no known information or knowledge about a particular movement, including its start to end, e.g., when someone is picking up something. This means the input is a series of movements without knowing the start and endpoints to indicate each activity. Some existing work has segmented the input data by activity [4]. Thus, the start and endpoints of the activities are already known, although the method of grouping activities may be unsupervised.

In this article, we focus on the less developed activity discovery comprising the block (d) to (e) of Fig. 1 by developing an effective methodology to extract good features and cluster activities without any label. Keyframe selection [7] and principle component analysis (PCA) are used to remove redundant frames and features to reduce time complexity and increase accuracy. A preprocessing and feature extraction methodology are proposed to prepare information and extract features from the most informative joints and bones, including joint displacement, joint orientation, and statistical time domain. As our first study, a hybrid PSO with Gaussian mutation and K-means (HPGMK) clustering that requires a known prior of the cluster number is proposed to find activities. Sometimes particles converge to a specific point between the best global and personal positions and get trapped in local optima. This difficulty arises when the swarm's variety reduces and the swarm cannot escape from a local optimum [8]. To address this, a hybrid PSO with Gaussian mutation is proposed to promote diversity to avoid early convergence. Then, K-means is applied to the centroids obtained by PSO to refine their location and get the best possible solution. Our methodology performs activity discovery using unsegmented input data and the proposed techniques used are unsupervised with no prior knowledge of the labels of the different activities. The main contributions of this article are as follows.

1) A methodology consists of keyframe selection, feature extraction, and PCA to represent human activities. The features based on displacement, statistical, and orientation

are extracted simultaneously to represent all the movement aspects of human activities. This makes the discovery part perform better because it has comprehensive information about the activities.

2) A hybrid clustering algorithm called HPGMK to discover and group unlabelled human activities observations into individual activity classes. PSO is customized by applying the Gaussian mutation, based on the advantages of two methods [9] and [10], on the global best's centroids to increase the diversity of selected clusters in the global best. Also, due to the increase in the variety of solutions, the proposed algorithm reaches the desired solution in a smaller number of iterations. That is why the clustering time is reduced.

3) Integrating K-means to refine the obtained cluster centers from the PSO to improve the exploitation of the algorithm. In PSO, when approaching the final solution, the speed of the particles decreases greatly and it becomes difficult to reach the best optimum solution. For this reason, the problem of PSO is solved by using the advantage of K-means in local search and applying it to the obtained solution from PSO.

The rest of this article is organized as follows. In this article, the background and related methods are discussed in Section II. The methodology is described in Section III. We present the evaluation of the proposed approach, comparing with state-of-the-art (SOTA) techniques in Section IV. Finally, Section V concludes this article.

II. RELATED WORKS

Feature extraction from 3-D skeletal human activities data: Skeletal data includes the number of joints, and each joint contains 3-D coordinates. Since the motion of the joints has essential information for any activity, feature extraction is vital. There are various methods for representation of the motion of skeletal joints such as calculating the difference between the joints in the same frame and the same joints in different frames [11], using histogram oriented of joints [12], dynamic time warping algorithm [13], covariance of 3-D joints [14], generating joint rotation matrix concerning the person's torso [15], and extracting the angles and orientations of the most informative body joints [16]. However, most of these methods extract one aspect of the skeletal data features, leading to other important aspects of activities being overlooked. As a result, there is a decrease in accuracy in the final result because of the insufficient discriminating ability of the extracted features. Moreover, due to the complexity of feature calculations, some of these methods cause computational latency. The difference between our article with previous works for extracting features is that We have combined the feature extraction techniques from [11], [16], and [17] to extract three skeletal features from informative joints and keyframes. Previously, these features have been applied separately and most of them have used all of the joints and frames, which increases additional information. This increases the time complexity and reduces identifying the activities' performance. Some methods like [7] and [18] have tried to select some frames that are more

29

of traditional LSTM. They also introduced an attention gate inside the LSTM to capture discriminative features. Xia et al. [41] provided a graph convention network based on spatial and temporal. They applied an attention layer to the model to generate discriminative features and modified feature maps. Then, a softmax classifier was used to categorize the activities. Cai et al. [42] introduced a scheme to capture visual information surrounding each skeleton joint and achieve local motion cues. They extracted features from skeleton and RGB data using two graph convolutional networks. Then, both types of features were concatenated and activities were classified by calculating a score based on linear blending. The problem with these approaches is that they require activity labels in the training data. Humans annotated the labels during data preparation. It makes these methods impractical with real-life data that are mainly unlabeled. Our work does not use labels for training in our algorithm. The algorithm discovers activities by looking for similar features between them. In addition, the methods mentioned above use deep learning techniques. In contrast, as a first study, we do not use them in our method and the focus is on developing a comprehensive model for HAD as a baseline.

Several approaches try to address the HAR in an unsupervised way. Wang et al. [43] presented a deep clustering method based on a dual-stack autoencoder to map raw data to spatio-temporal features. After extracting features, the radial basis function neural network was used to classify the activities. Su et al. [6] provided an unsupervised model by employing a bidirectional recurrent neural network and used K-NN to classify the activities. Liu et al. [44] designed a spatial-temporal asynchronous normalization method to reduce redundant information related to time and normalize the spatial features. Next, they used a gated recurrent unit autoencoder to feature vectors. First, all of these methods received the activities already segmented which has enabled them to be aware of the differences between the activities before performing the recognition. Second, in most of these methods, only feature extraction was performed without supervision. The supervised classification method was used for the rest of the operations to learn activity models using activity labels.

On the other hand, human activity discovery can automatically categorize human activity in a fully unsupervised way and the challenge is learning from unlabeled data. The majority of existing methods were developed for sensor-based [45], [46] and RGB video data [47]. The challenges of the sensor-based approach are difficult to implement in the environment and take a long time to install [48]. Furthermore, it is impractical for people to wear sensors everywhere. With RGB videos, the problems faced are millions of pixel values, illumination variations, viewpoint changes, and cluttered backgrounds [5]. In this work, we concentrate on 3-D skeleton-based data as it does not have the problems of the other two data types. One of the first works in HAD was performed by Ong et al. [48]. They proposed an autonomous learning technique based on the mixture of the Gaussian hidden Markov model. They introduced an incremental clustering approach based on K-means to discover the activities to deal with the undefined number of clusters. An issue with their approach is that they have used K-means to discover the activities that get stuck into the local optimum

easily and they have not examined all aspects of the skeleton data features. Moreover, they extracted all the features from all joints, resulting in more redundant data and increased discovery errors. Recently several approaches have been proposed by [4] to solve HAR without labels. In their proposed methods, several clustering methods, including spectral clustering (SC), elastic net subspace clustering (ENSC), and sparse subspace clustering (SSC) were used, which used covariance descriptors. They used an affinity matrix to find similarities and then applied spectral clustering. In addition, a time stamp pruning approach was used to remove redundant data to normalize temporal features. Although they have achieved impressive results, the data used were already segmented by activity before performing discovery. It means that the activities are already categorized. Because each sample contains an activity that performs completely. In other words, the beginning and end of each activity are clear.

In a nutshell, many methods have been proposed to recognize human activities in a supervised and unsupervised manner and have obtained very acceptable results. But the problem with these methods is that they ignore the discovery step. These methods are useless because labeling activities do not occur in real time. They also need a lot of computation time for training. In reality, training data are not available. If we have a dynamic big and growing video related to human activity, we are not sure of the labels to predefine the rules. This can be a real challenge. On the other hand, due to the variety of human activities, these methods need to be retrained for new activities, making them not scalable. In the case of HAD, in addition to the fact that there are very limited methods, these methods have problems including using shallow methods which are not very accurate or not fully performing the discovery process in an unsupervised manner. In this article, we propose an approach to discovering activities from untrimmed videos without knowing the label of activities. It makes this method suitable for use in real-world scenarios. In addition, we use a feature extraction approach to examine most aspects of skeleton data along with a keyframe selector to reduce redundant information and discovery accuracy.

III. PROPOSED HUMAN ACTIVITY DISCOVERY

The goal of this article is to pave the path to make robots and machines learn like toddlers, who learn automatically, gain knowledge every day, and their intelligence will gradually increase until they resemble humans. The crucial aspect is the increase in learning without human intervention and planning. The proposed approach can be used in developing other business technologies by identifying and analyzing human activities. For example, in security areas, new activities can be discovered and the new activity can be determined whether is suspicious or not. Another application is in hospitals to care for patients. New patient conditions are discovered through their activities and can be used to help better monitor the patient. To reach this goal, we have presented an unsupervised framework for 3-D skeleton-based human activity to discover the performed activities from untrimmed streaming and unlabeled data in which each sequence contains multiple activity samples without prior knowledge about the performed activities. Our proposed

framework consists of two main stages (the diagram of the proposed framework is given in Fig. 1 of supplementary). In the first stage, we propose an approach that can extract high-quality features in an unsupervised manner. Three crucial factors should be taken into account in this matter. First, not all the captured frames are important. Due to the similarity of frames and noises, HAD performance reduces sharply. Therefore, we need frames that show the salient features of the activities. Also, not all the joints have a significant role in discovering activities, such as the torso, which is constant in most activities. Extracting features from these joints increases time complexity. Second, the extracted features should accurately reflect all aspects of human activity by considering all factors such as spatial and temporal. Third, the feature dimension needs to be minimal to ease the clustering process (a larger dimension confuses the clustering process). To meet the requirement in the first stage, we present a preprocessing method [shown in Fig. 1 (Stage 1) of supplementary]. We employ an innovative approach based on kinetic energy to select representative frames of the video sequences as keyframes. In other words, we seek to select the frames that show the most prominent characteristics of the activities as the keyframes. However, the selection of keyframes without losing the required information is a challenging task. When only the local maximum kinetic energies are considered keyframes, the sequence of activities may break up and no longer represent activities. For example, in the walking activity, considering keyframes with high local energy, only positions where both legs move away from each other are considered keyframes. In contrast, positions, where both legs are placed together, show a part of the walking process that is lost, if only the maximum kinetic energies are considered. For this reason, we perform keyframe selection based on local maximum and minimum kinetic energy that can find a sequence of representative frames while reducing complexity. To minimize computational time and overlapping among activities, we select joints (informative) that have a vital role in displaying activities and avoid similarity between different activities based on experimental tests. To increase the discovery performance, it is necessary to extract features to represent all aspects of each activity. for this purpose, we design a method to represent the activities based on spatial and temporal displacement, statistical, and orientation features. The displacement-based representations provide the view-invariant spatio-temporal human representations independent of the position and orientation of the person with respect to the camera. To obtain features that are invariant to human scale changes, orientation-based representations are extracted to find relative information between human joints. Statistical features describe how activity evolves over time, especially when distinguishing the actions of the arms from the legs. Therefore, statistical time-domain characteristics represent changes in a set of postures for a time-domain activity. We employ PCA to address the third factor to reduce dimensionality and make the extracted feature more clusterable while high-importance features are kept. We adopt a sliding window over the stream of skeleton sequences to perform activity discovery. To increase the number of samples and avoid pruning important events

like a transition between activities we employ overlapping sliding windows.

In the second stage, we propose a novel clustering algorithm based on hybrid PSO called HPGMK to discover human activities [shown in Fig. 1 (Stage 2) of supplementary]. The key benefit of PSO is that there are fewer parameters to set. Contrary to genetic algorithms, PSO does not use complex evolutionary operators like the crossover, making it less complicated. However, the issue is that its convergence speed slows down when it approaches the global optimum. For this reason, we combine PSO with K-means to use the fast speed of K-means to reach the local optimum to improve PSO's performance. The Gaussian mutation is employed on the global best particle to search for areas around it to generate diverse solutions and strike a balance between exploitation and exploration. In the following subsections, the details of each part of the methodology are described.

A. Keyframe Selection

Keyframe selection is a process of selecting frames reflecting the main activities in the video. Some methods like [7] and [18] have tried to select some frames that are more distinguishing compared to other frames. However, these methods were applied to each activity sample separately. In other words, in methods [7] and [18], keyframes were selected in a supervised manner. In our proposed method, keyframes are selected without knowing the beginning and end of the activities. To find the keyframes, the kinetic energy $E(f_i)$ of each frame f_i is calculated [7] using (1), based on the displacement of joints over time. In this way, the movement of a joint j between frame i and i-1 is calculated for all joints (J). The sum of the movements for all joints is the energy of the current frame. Frames with local maxima and minima amount of kinetic energy compared to neighboring frames are considered keyframes (see Fig. 2 of supplementary) Because these are the energy's extreme points, which are meant to resemble crucial posture data.

$$E(f_i) = \sum_{j=1}^{J} E(f_i^j) = 1/2 \sum_{j=1}^{J} (f_i^j - f_{i-1}^j)^2. \qquad (1)$$

B. Feature Extraction

To represent the activities, a set of statistical displacements, angles, and orientation features for encoding key aspects of activities are extracted. These important features are extracted from selected (informative) joints in the data to describe the shape and movement of human. Selected joints have been obtained based on experimental tests that included left and right hand, foot, hip, shoulder, elbow, and knee. These joints have more movement and contribution than other joints such as torso in activities. We use information related to the position and movement of joints, the orientation and angle between a pair of bones, and activity variation over time. The normalization procedure [11] is performed on all features.

1) Displacement Features: Joint displacement-based features encode information on the position and motion of body
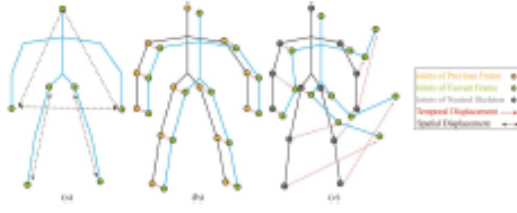
31

Fig. 2. (a) Spatial displacement of pairwise joints in the same frame. Temporal displacement of current frame from (b) previous frame and (c) the neutral frame.

Fig. 3. (a) Illustration of the rotation between two bones A and B. α, β, and γ are the orientation of angles. (b) The angles of the selected body bones. The angles of elbow-wrist and shoulder-elbow at both sides and angles between the bones of hip-knee and knee-ankle at both sides are used to calculate angle features.

joints, particularly displacement between joints of a pose and position differences of skeleton joints across time [11] as follows:

1) Spatial joint displacement is computed using pairwise Euclidean distances between joints P_i and P_j ($i \neq j$) in 3-D space in the same frame (2). The joint pairs used are both hands, hands and head, and hip and feet at both sides, giving five features per frame, [see Fig. 2(a)].

$$\text{Pairwise distances} = \sqrt{\sum_{x,y,z} (P_i - P_j)^2}. \quad (2)$$

2) Temporal joint displacement is calculated based on two modes. T_{cp} is the difference between each selected joint P_i in current frame P_i^c and previous frame P_i^{c-1} [see Fig. 2(b)] to determine the small changes in joint movement over time (3). T_{cn} is the difference between each selected joint of current frame and the frame of neutral pose (we randomly select a standing position as a neutral position) P_i^n, illustrated in Fig. 2(c), to find general changes in joint movements as given in (4).

$$T_{cp} = P_i^c - P_i^{c-1} \quad (3)$$

$$T_{cn} = P_i^c - P_i^n. \quad (4)$$

2) *Statistical Features:* The mean and standard deviation of time-domain features express how activity changes over time, particularly in distinguishing between activities related to the arms and legs. Thus, statistical time-domain features encode variations across a collection of poses of an activity in time-domain. These features are calculated by the difference of selected joint P_i^c in current frame from mean $P_{(i,mean)}$ and standard deviation $P_{(i,std)}$ of the selected joint coordinates within an activity sequence as given by (5) and (6) [11]. They are as follows:

1) Joint coordinate-mean difference

$$P_{i_{(mean)}}^c = P_i^c - P_{(i,mean)}, P_{(i,mean)} = \frac{1}{N}\sum_{c=1}^{N} P_i^c; \quad (5)$$

N is the number of frames.

2) Joint coordinate-standard deviation difference

$$P_{(i,std)}^c = P_i^c - \sqrt{\frac{\sum_{i=c}^{N}(P_i^c - P_{(i,mean)})^2}{N}}. \quad (6)$$

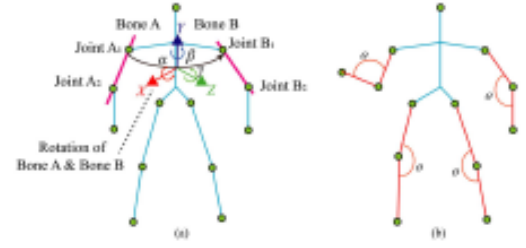3) *Orientation Features:* The 3-D coordinate system $\{x, y, z \in R^3\}$ represents points as joints. x, y, and z denote the 3-D coordinates of joints. Joints and bones can be described by the orthonormal vectors [17] as follows:

$$P_i^f = x_i^f e_1 + y_i^f e_2 + z_i^f e_3 \quad (7)$$

$$B_{ij}^f = (x_i^f - x_j^f)e_1 + (y_i^f - y_j^f)e_2 + (z_i^f - z_j^f)e_3 \quad (8)$$

where P_i^f is the ith skeleton joint in the fth frame and e_1, e_2, e_3 are orthonormal vectors. B_{ij}^f is the bone between two adjacent joints P_i^f and P_j^f. Moreover, magnitude and direction of two bones a and b are represented by geometric product where this product is the sum of internal $(a.b)$ and external $(a \wedge b)$ product. where the inner product is used to compute the length and angle between two bones a and b. The outer product of two bones can be regarded as an oriented plane containing a and b. The orientation and angles between bones features are obtained in the process described as follows.

1) The rotation matrix is a transformation matrix that describes the rotation from a bone to another. Three angles are required to define the rotation matrix between two bones. The rotation angles are considered as orientation features. The elements of these features are the rotation of bones relative to the x, y, and z axes [see Fig. 3(a)].

2) The angle features consist of the angles between the bones of elbow-wrist and shoulder-elbow at both sides and the angles between the bones of hip-knee and knee-ankle at both sides. These angles are highlighted in Fig. 3(b) and calculated below as follows:

$$\theta = 180 \times \frac{\arctan^2(\frac{\|bone_i \wedge bone_j\|}{\|bone_i . bone_j\|})}{\pi} + 180 \quad (9)$$

where $bone_i$ and $bone_j$ are determined by (8).

C. *Feature Selection and Sampling*

For fast clustering and complexity reduction, key features are extracted by PCA. Then, sliding windows are used to segment frames into time windows. Each window comprises of 15 frames. The overlap of sliding windows increases performance, because it increases the number of samples and
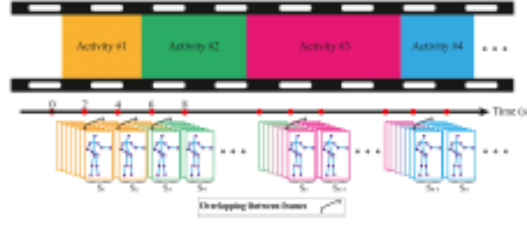
32

Fig. 4. Illustration of sampling based on overlapping sliding windows. Each sample ($S_1, S_2, \ldots, S_n$), except the first sample, starts with the last frame of the previous sample.
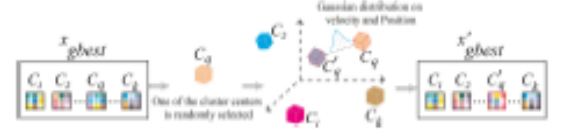

Fig. 5. Visualization of the Gaussian mutation operator. In each iteration of hybrid PSO, one centroid (C_q) is chosen from x_{gbest} randomly. Then, Gaussian distribution is applied on position and velocity of the selected centroid based on (16) and (17) to create a new offspring C'_q. The new global best (x'_{gbest}) is then compared to x_{gbest}. If x'_{gbest} has better fitness value than x_{gbest}, x'_{gbest} is replaced with new global best.

avoid pruning important events like transition between activities [49]. The first 15 frames do not overlap while in other samples, their first frame starts from the last frame of the previous sample (see Fig. 4).

D. Proposed Clustering

PSO is a population-based optimization algorithm [50]. A population is made up of a number of particles and each particle represents a solution and moves according to its speed. The changes in velocity and position of the particles are calculated based on the following formula (the illustration of how to decode the PSO particles to obtain a set of clusters and evolving particles is shown in Fig. 3 of supplementary material):

$$v_i(t+1) = w \times v_i(t) + c_1 \times rand_1 \times (pbest_i(t) - x_i(t))$$
$$+ c_2 \times rand_2 \times (gbest(t) - x_i(t)) \tag{10}$$

$$x_i(t+1) = x_i(t) + v_i(t+1) \tag{11}$$

$$w = w_{max} - \frac{t}{t_{max}} \times (w_{max} - w_{min}) \tag{12}$$

$$c_1(t+1) = c_{1_{max}} - (c_{1_{max}} - c_{1_{min}}) \times \frac{t}{t_{max}} \tag{13}$$

$$c_2(t+1) = c_{2_{min}} + (c_{2_{max}} - c_{2_{min}}) \times \frac{t}{t_{max}} \tag{14}$$

In (11) and (10) $x_i(t)$ and $v_i(t)$ are the position and velocity of the particle i at time t, respectively. $pbest_i$ is the best position found by the particle i. $gbest$ is the best position found in the population. w is the inertial weight defined by (12) and starts to decrease from w_{max}. c_1 and c_2 are acceleration coefficients expressed by (13) and (14). The $c_{1_{max}}, c_{2_{max}}$ and $c_{1_{min}}, c_{2_{min}}$ are initial and final values, respectively, t is the number of iterations, and t_{max} is the maximum number of iterations [26]. $rand_1$ and $rand_2$ are random variables between 0 and 1. Each solution is evaluated by (15) which should be minimized to achieve proper clustering.

$$SSE = \sum_{k=1}^{K} \sum_{\forall x \in c_k} \|x_i - \mu_k\|^2 \tag{15}$$

x_i is a data point belonging to the cluster C_k and μ_k is the mean of the cluster C_k. k is the number of clusters specified. To avoid in local optimum, a Gaussian mutation operator based on [9] and [10] is applied to the global particle as follows:

$$v'_{gbest}(d) = v_{gbest}(d) \times G(0, h)$$
$$\times (x_{max}(d) - x_{min}(d)) \tag{16}$$

$$x'_{gbest}(d) = x_{gbest}(d) + G(0, h) \times v'_{gbest}(d) \tag{17}$$

where x_{gbest} and v_{gbest} represent the position and velocity of global best particle. x_{max} and x_{min} are the maximum and minimum value in dth dimension. Gaussian $(0,h)$ is Gaussian distribution with the mean 0 and the variance h. The value of h starts with a high value ($h(0) = 1$) to increase the exploration ability of the algorithm to find an interesting region at the beginning of the search. Then, h decreases linearly during each iteration according to (18), to increase the power of exploitation at the end of the search to reach the optimum solution.

$$h(t+1) = h(t) - (1/t_{max}) \tag{18}$$

where t_{max} is the maximum number of iterations. Fig. 5 is an illustration of the Gaussian mutation.

In general, the core of HPGMK is based on PSO which is a population-based algorithm. The position of each particle in the population represents a solution. In other words, each particle contains the position of the cluster centers. Each particle updates its position using its velocity to reach the optimum solution [51]. Furthermore, we have one objective function which is SSE (sum square error). We evaluate each individual based on SSE (15). In this process, an individual is chosen as the global best in each iteration with the lowest SSE value among the rest of the individuals. To increase the diversity of solutions, a Gaussian mutation is applied to the position and velocity of the global best particle. The velocity of the particles reduces quickly as PSO approaches the global optimum, and in most circumstances, the ideal solution is not achieved. For this reason, K-means is applied to the obtained centroids from PSO to refine them. After the completion of the PSO process, the global best solution is selected based on the SSE value. Then, the selected solution is modified by averaging the position of data points in each cluster to select the best position for the cluster centers. This continues until the position of the clusters does not change. The routine of the proposed clustering algorithm is given in supplementary.

IV. EXPERIMENTS

A. Datasets

Five datasets were used to evaluate the effectiveness of proposed method: Cornell Activity Dataset (CAD-60) [15], UTKinect-Action3D (UTK) [12], Florence3D (F3D) [52], Kinect Activity Recognition Dataset (KARD) [37], and MSR DailyActivity3D (MSR) [53]. These datasets have different dimensions, features, and activities. More details about the datasets are given in the supplementary.

B. Method

The performance of our proposed method (HPGMK) was with three SOTA methods for HAD incuding ENSC, SSC, and SC [4] and three recent deep clustering methods deep clustering network (DCN) [54], structural deep clustering network (SDCN) [55], and incomplete multiview clustering via contrastive prediction (Completer) [56]. In addition, we compared our method with conventional and well-known clustering methods including K-means clustering (KM) and PSO. All parameters of each compared method, such as dimensions and numbers of layers, have been adjusted as described in their papers. KM and PSO have been chosen for comparison as our proposed HPGMK is based on them. ENSC was found to be most similar to our work as an unsupervised algorithm requiring known cluster number while SSC and SC were the original algorithms that ENSC was based on. The three deep clustering methods were chosen to compare our method with the latest methods that use deep learning tools for clustering. To compare the performance of the methods, the accuracy metric (calculated based on [57]) was used. Moreover, F-score was used to show the performance of each method in categorizing each activity and the confusion between them was shown in the confusion matrix. The convergence test and clustering time of HPGMK were measured to evaluate the benefits of each component used in the HPGMK on its performance. Information about parameters setting and the computation complexity of HPGMK are given in the supplementary.

C. Results and Discussion

Fig. 6 shows the accuracy of HPGMK with the SOTA techniques for all subjects of each dataset based on the maximum, minimum and average accuracies. The average overall accuracy of the HPGMK was 77.53% for CAD-60, 56.54% for UTK, 66.84% for F3D 46.02% for KARD, and 40.12% for MSR. As seen in Fig. 6, HPGMK has the best performance in terms of maximum and average accuracy in all datasets. This shows the effectiveness of the HPGMK for human activity discovery. By utilizing the Gaussian mutation and KM along with PSO, our approach brings performance improvement compared to the other methods. ENSC and SSC, which are subspace clustering algorithms, do not use an efficient search strategy [58]. In these methods, there is no strategy for maintaining the balance between exploitation and exploration in their search. Moreover, parameters are required to be set and finding the right values for them is tricky and complex such as size of subspace [59].
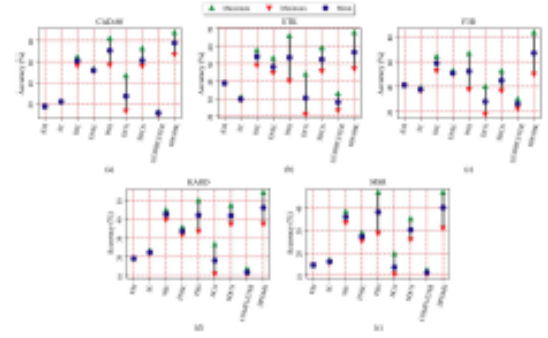

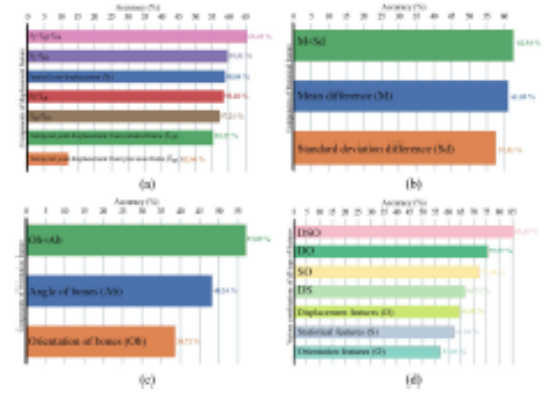Fig. 6. Average accuracy for all subjects in (a) CAD-60, (b) UTK, (c) F3D, (d) KARD, and (e) MSR.

Fig. 7. Effect of each component of each type of feature approach including (a) displacement (D), (b) statistical (S), (c) orientation (O) features, and (d) their various combinations together. Capital letters stand for different methods and putting these letters together means combining relevant methods.

In contrast, HPGMK are not dependent to parameters like SSC and ENSC and has several strategies for searching. First, it used the PSO to search in large space area by using several particles as potential solutions. To promote diversity, Gaussian mutation is used. KM is also used to search in a small area of the global best solution to refine the obtained centroids from PSO. These search strategies enable HPGMK to have a relatively good performance compared to the SSC and ENSC. Moreover, HPGMK has performed better than the methods that use deep learning. In HPGMK, both spatial features from each 3-D skeleton frame and temporal features from sequences along with Orientation and statistical information are extracted. However, in the deep clustering this information is ignored. On the other hand, unlike deep clustering methods that have used shallow clustering, HPGMK has different search strategies for exploration and exploitation to determine better clusters. Fig. 7(a)–(c) show the effect of each component of each type of the proposed hybrid feature extraction method based on the discovery accuracy of the activities performed by subject one in the CAD-60 dataset. Percentages represent the discovery accuracy using the different
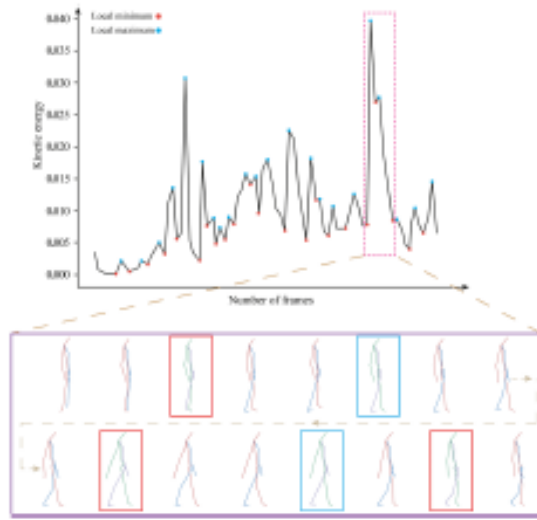
Fig. 8. Illustration of the effect of selecting the keyframe based on kinetic on "walking" in MSR where distinct frames compared to adjacent frames are selected from similar frames. A window of frames is shown with a few selected frames based on the local maximum (blue) and local minimum (red) energy.

combination of features and each piece of graphs shows the ratio of the impact of one component in discovery improvement to the rest of the other components in each type of feature. This ratio is obtained based on dividing *discovery accuracy obtained by one of the components from a feature type* by *summation of discovery accuracy obtained by all components of that feature type*. Overall, in three feature extraction methods comprising displacement, statistical, and orientation features, when all their components are combined, the discovery accuracy significantly increased and obtained 65.45%, 62.54%, and 57.09%, respectively. By contrast, if each component of the feature extraction methods is used alone without considering other components in the feature extraction, the accuracy of discovery decreases. Fig. 7(d) shows the effect of different combinations of each of the feature extraction methods. The size of each circle indicates the effectiveness of the features. Based on the obtained results, it is shown that the highest detection accuracy of 85.45% is obtained by combining all the three methods. It indicates that in order to better differentiate between activities, it is necessary to extract features from different aspects of activities. Fig. 8 shows the selected keyframes from walking activity in MSR. As is shown, there are a lot of frames with high similarity that by extracting their features not only do not help to improve discovery, but also increase the computational complexity and increase the overlap between other activities because these gestures occur in other activities. However, using local maximum and minimum kinetic energy can find representative frames and reduce complexity. Looking at a window of frames, we can see the selected frames based on the maximum and minimum local energy value. The selected frames show the most differentiation to display the activity sequence. It is worth mentioning that selecting keyframes maintains the order of the activity.

It is worth to mention that the results related to confusion matrix, clustering time of different components of HPGMK, average F-score, and the Kruskal-Wallis test (p-value) between HPGMK and KM, HPGMK and PSO, HPGMK and SSC, HPGMK and ENSC, and HPGMK and SDCN are reported in the supplementary.

V. CONCLUSION

Most of the proposed HAR frameworks are supervised or semisupervised, making them unusable in real-world situations due to a lack of access to the ground truth. In this article, a HPGMK approach was proposed to solve human activity discovery on skeleton-based data with no prior knowledge of the label of the activities in the data. Five different datasets were used to assess the performance of the method. The results obtained have shown that HPGMK achieved an average overall accuracy of 77.53%, 56.54%, 66.84%, 46.02%, and 40.12% in datasets CAD-60, UTK, F3D, KARD, and MSR, respectively, and validate the superiority of HPGMK over other methods compared. In activities with high intraclass variation, corrupted data and the same activity performed in sitting and standing positions, HPGMK has performed better in activity discovery than other SOTA methods.

We have examined the impact of each feature used. It was found that the simultaneous combination of features together further improves the results. The impact of the different components in the proposed algorithm has shown that Gaussian mutation has evolved particles to improve search algorithm and K-means has increased discovery efficiency and improved the convergence rate.

This work paves the way toward implementing fully unsupervised human activity discovery in practical applications using skeleton-based data. There are various factors in HPGMK that need to be addressed to develop an effective HAD algorithm. One factor is the number of clusters that were pre-configured in the proposed algorithm. The HPGMK needs to be further extended to automatically address human activity discovery by estimating the number of activities by itself. Another factor is detecting outliers. Outliers shift the cluster centers towards themselves, thus affecting optimal cluster formation. Using outlier detection methods in HPGMK to reject outliers will be beneficial. The manual procedure used to build the PSO structure in the suggested technique was based on the information and expertise that was gained. It takes a long time to manually introduce changes because of trial and error, which makes it challenging to thoroughly explore all potential algorithm setups. A potential future study to address these concerns is automating the suggested method's setup to make it more effective in handling various circumstances and datasets. Moreover, HPGMK's potential for wider application and improvement through PSO variants like learning-aided [60], region-encoding [61], and triple archive PSO [62] can be investigated.

REFERENCES

[1] P. Hadikhani, M. Eslaminejad, M. Yari, and E. Ashoor Mahani, "An energy-aware and load balanced distributed geographic routing algorithm for wireless sensor networks with dynamic hole," *Wireless Netw.*, vol. 26, no. 1, pp. 507–519, 2020.

[2] M. Yari, P. Hadikhani, and Z. Asgharzadeh, "Energy-efficient topology to enhance the wireless sensor network lifetime using connectivity control," *J. Telecommun. Digit. Economy*, vol. 8, no. 3, pp. 68–84, 2020.

[3] H. V. Chandrashekhar et al., "Human activity representation, analysis, and recognition," 2006.

[4] G. Paoletti, J. Cavazza, C. Beyan, and A. Del Bue, "Subspace clustering for action recognition with covariance representations and temporal pruning," in *Proc. 25th Int. Conf. Pattern Recognit.*, 2021, pp. 6035–6042.

[5] J. Han, L. Shao, D. Xu, and J. Shotton, "Enhanced computer vision with microsoft kinect sensor: A review," *IEEE Trans. Cybern.*, vol. 43, no. 5, pp. 1318–1334, Oct. 2013.

[6] K. Su, X. Liu, and E. Shlizerman, "Predict & cluster: Unsupervised skeleton based action recognition," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2020, pp. 9631–9640.

[7] J. Shan and S. Akella, "3D human action segmentation and recognition using pose kinetic energy," in *Proc. IEEE Int. Workshop Adv. Robot. Social Impacts*, 2014, pp. 69–75.

[8] O. M. Nezami, A. Bahrampour, and P. Jamshidlou, "Dynamic diversity enhancement in particle swarm optimization (DDEPSO) algorithm for preventing from premature convergence," *Procedia Comput. Sci.*, vol. 24, pp. 54–65, 2013.

[9] B. Jana, S. Mitra, and S. Acharyya, "Repository and mutation based particle swarm optimization (RMPSO): A new PSO variant applied to reconstruction of gene regulatory network," *Appl. Soft Comput.*, vol. 74, pp. 330–355, 2019.

[10] C. Li, S. Yang, and I. Korejo, "An adaptive mutation operator for particle swarm optimization," 2008.

[11] D. A. Adama, A. Lotfi, C. Langensiepen, K. Lee, and P. Trindade, "Human activity learning for assistive robotics using a classifier ensemble," *Soft Comput.*, vol. 22, no. 21, pp. 7027–7039, 2018.

[12] L. Xia, C.-C. Chen, and J. K. Aggarwal, "View invariant human action recognition using histograms of 3D joints," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. Workshops*, 2012, pp. 20–27.

[13] M. Tabejamaat and H. Mohammadzade, "Embedded feature representation in dynamic time warping space for 3D action recognition using kinect depth sensor," *J. Mach. Vis. Image Process.*, vol. 9, no. 3, pp. 19–34, 2022.

[14] M. E. Hussein, M. Torki, M. A. Gowayyed, and M. El-Saban, "Human action recognition using a temporal hierarchy of covariance descriptors on 3D joint locations," in *Proc. 23rd Int. Joint Conf. Artif. Intell.*, 2013, pp. 2466–2472.

[15] J. Sung, C. Ponce, B. Selman, and A. Saxena, "Unstructured human activity detection from RGBD images," in *Proc. IEEE Int. Conf. Robot. Automat.*, 2012, pp. 842–849.

[16] W. Cao, Y. Lu, and Z. He, "Geometric algebra representation and ensemble action classification method for 3D skeleton orientation data," *IEEE Access*, vol. 7, pp. 132049–132056, 2019.

[17] X. Liu, Y. Li, and R. Xia, "Rotation-based spatial–temporal feature learning from skeleton sequences for action recognition," *Signal Image Video Process.*, vol. 14, no. 6, pp. 1227–1234, 2020.

[18] M. M. Arzani, M. Fathy, A. A. Azirani, and E. Adeli, "Switching structured prediction for simple and complex human activity recognition," *IEEE Trans. Cybern.*, vol. 51, no. 12, pp. 5859–5870, Dec. 2020.

[19] P. Hadikhani, D. T. C. Lai, W.-H. Ong, and M. H. Nadimi-Shahraki, "Improved data clustering using multi-trial vector-based differential evolution with Gaussian crossover," in *Proc. Genet. Evol. Computation Conf. Companion*, 2022, pp. 487–490.

[20] P. Hadikhani, D. T. C. Lai, W.-H. Ong, and M. H. Nadimi-Shahraki, "Automatic deep sparse multi-trial vector-based differential evolution clustering with manifold learning and incremental technique," 2023.

[21] Z.-G. Chen, Z.-H. Zhan, S. Kwong, and J. Zhang, "Evolutionary computation for intelligent transportation in smart cities: A survey," *IEEE Comput. Intell. Mag.*, vol. 17, no. 2, pp. 83–102, May 2022.

[22] Z.-H. Zhan, L. Shi, K. C. Tan, and J. Zhang, "A survey on evolutionary computation for complex continuous optimization," *Artif. Intell. Rev.*, vol. 55, pp. 59–110, 2022.

[23] K. Malarvizhi and K. Amshakala, "WITHDRAWN: Data clustering using hybrid of feature linkage weight based feature reduction and particle Swarm optimization," 2021.

[24] M. Sharma and J. K. Chhabra, "Sustainable automatic data clustering using hybrid PSO algorithm with mutation," *Sustain. Comput.: Inform. Syst.*, vol. 23, pp. 144–157, 2019.

[25] S. Rengasamy and P. Murugesan, "PSO based data clustering with a different perception," *Swarm Evol. Computation*, vol. 64, 2021, Art. no. 100895.

[26] J. Cai, H. Wei, H. Yang, and X. Zhao, "A novel clustering algorithm based on DPC and PSO," *IEEE Access*, vol. 8, pp. 88200–88214, 2020.

[27] Y. Zhang, Y.-H. Wang, D.-W. Gong, and X.-Y. Sun, "Clustering-guided particle swarm feature selection algorithm for high-dimensional imbalanced data with missing values," *IEEE Trans. Evol. Comput.*, vol. 26, no. 4, pp. 616–630, Aug. 2022.

[28] X.-F. Song, Y. Zhang, D.-W. Gong, and X.-Z. Gao, "A fast hybrid feature selection based on correlation-guided clustering and particle swarm optimization for high-dimensional data," *IEEE Trans. Cybern.*, vol. 52, no. 9, pp. 9573–9586, Sep. 2022.

[29] M. Hubert and S. Engelen, "Robust PCA and classification in biosciences," *Bioinformatics*, vol. 20, no. 11, pp. 1728–1736, 2004.

[30] M. Hubert, P. J. Rousseeuw, and S. Verboven, "A fast method for robust principal components with applications to chemometrics," *Chemometrics Intell. Lab. Syst.*, vol. 60, no. 1-2, pp. 101–111, 2002.

[31] M. Hubert, P. J. Rousseeuw, and K. Vanden Branden, "ROBPCA: A new approach to robust principal component analysis," *Technometrics*, vol. 47, no. 1, pp. 64–79, 2005.

[32] E. J. Candès, X. Li, Y. Ma, and J. Wright, "Robust principal component analysis?," *J. ACM (JACM)*, vol. 58, no. 3, pp. 1–37, 2011.

[33] M. Alessandrini, G. Biagetti, P. Crippa, L. Falaschetti, S. Luzzi, and C. Turchetti, "EEG-based Alzheimer's disease recognition using robust-PCA and LSTM recurrent neural network," *Sensors*, vol. 22, no. 10, 2022, Art. no. 3696.

[34] M. Rahmani and P. Li, "Outlier detection and robust PCA using a convex measure of innovation," *Adv. Neural Inf. Process. Syst.*, vol. 32, 2019.

[35] L. Shi, Y. Zhang, J. Cheng, and H. Lu, "Two-stream adaptive graph convolutional networks for skeleton-based action recognition," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2019, pp. 12026–12035.

[36] C. Li, C. Xie, B. Zhang, J. Han, X. Zhen, and J. Chen, "Memory attention networks for skeleton-based action recognition," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 33, no. 9, pp. 4800–4814, Sep. 2022.

[37] S. Gaglio, G. L. Re, and M. Morana, "Human activity recognition process using 3-D posture data," *IEEE Trans. Human-Mach. Syst.*, vol. 45, no. 5, pp. 586–597, Oct. 2014.

[38] S. K. Yadav, K. Tiwari, H. M. Pandey, and S. A. Akbar, "Skeleton-based human activity recognition using ConvLSTM and guided feature learning," *Soft Comput.*, vol. 26, no. 2, pp. 877–890, 2022.

[39] P. Zhang, C. Lan, W. Zeng, J. Xing, J. Xue, and N. Zheng, "Semantics-guided neural networks for efficient skeleton-based human action recognition," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2020, pp. 1112–1121.

[40] C. Si, W. Chen, W. Wang, L. Wang, and T. Tan, "An attention enhanced graph convolutional LSTM network for skeleton-based action recognition," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2019, pp. 1227–1236.

[41] H. Xia and X. Gao, "Multi-scale mixed dense graph convolution network for skeleton-based action recognition," *IEEE Access*, vol. 9, pp. 36475–36484, 2021.

[42] J. Cai, N. Jiang, X. Han, K. Jia, and J. Lu, "JOLO-GCN: Mining joint-centered light-weight information for skeleton-based action recognition," in *Proc. IEEE/CVF Winter Conf. Appl. Comput. Vis.*, 2021, pp. 2735–2744.

[43] T. Wang et al., "A deep clustering via automatic feature embedded learning for human activity recognition," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 32, no. 1, pp. 210–223, Jan. 2021.

[44] M. Liu, Y. Bao, Y. Liang, and F. Meng, "Spatial-temporal asynchronous normalization for unsupervised 3D action representation learning," *IEEE Signal Process. Lett.*, vol. 29, pp. 632–636, 2022.

[45] P. Gupta, R. McClatchey, and P. Caleb-Solly, "Tracking changes in user activity from unlabelled smart home sensor data using unsupervised learning methods," *Neural Comput. Appl.*, vol. 32, no. 16, pp. 12351–12362, 2020.

[46] W. Qi, H. Su, and A. Aliverti, "A smartphone-based adaptive recognition and real-time monitoring system for human activities," *IEEE Trans. Human-Mach. Syst.*, vol. 50, no. 5, pp. 414–423, Oct. 2020.

[47] X. Yan, Y. Ye, X. Qiu, and H. Yu, "Synergetic information bottleneck for joint multi-view and ensemble clustering," *Inf. Fusion*, vol. 56, pp. 15–27, 2020.

[48] W.-H. Ong, L. Palafox, and T. Koseki, "Autonomous learning and recognition of human action based on an incremental approach of clustering," *IEEJ Trans. Electron., Inf. Syst.*, vol. 135, no. 9, pp. 1136–1141, 2015.

[49] L. L. Presti and M. La Cascia, "3D skeleton-based human action classification: A survey," *Pattern Recognit.*, vol. 53, pp. 130–147, 2016.

[50] P. Hadikhani and P. Hadikhani, "An adaptive hybrid algorithm for social networks to choose groups with independent members," *Evol. Intell.*, vol. 13, no. 4, pp. 695–703, 2020.

36

[51] J. Kennedy and R. Eberhart, "Particle swarm optimization," in *Proc. - Int. Conf. Neural Netw.*, 1995, vol. 4, pp. 1942–1948.

[52] L. Seidenari, V. Varano, S. Berretti, A. Bimbo, and P. Pala, "Recognizing actions from depth cameras as weakly aligned multi-part bag-of-poses," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops*, 2013, pp. 479–485.

[53] J. Wang, Z. Liu, Y. Wu, and J. Yuan, "Mining actionlet ensemble for action recognition with depth cameras," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2012, pp. 1290–1297.

[54] B. Yang, X. Fu, N. D. Sidiropoulos, and M. Hong, "Towards k-means-friendly spaces: Simultaneous deep learning and clustering," in *Proc. Int. Conf. Mach. Learn.*, 2017, pp. 3861–3870.

[55] W. Van Gansbeke, S. Vandenhende, S. Georgoulis, M. Proesmans, and L. Van Gool, "Scan: Learning to classify images without labels," in *Proc. Eur. Conf. Comput. Vis.*, 2020, pp. 268–285.

[56] Y. Lin, Y. Gou, Z. Liu, B. Li, J. Lv, and X. Peng, "Completer: Incomplete multi-view clustering via contrastive prediction," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2021, pp. 11 174–11 183.

[57] B. Peng, J. Lei, H. Fu, L. Shao, and Q. Huang, "A recursive constrained framework for unsupervised video action clustering," *IEEE Trans. Ind. Informat.*, vol. 16, no. 1, pp. 555–565, Jan. 2020.

[58] Y. Lu, S. Wang, S. Li, and C. Zhou, "Particle swarm optimizer for variable weighting in clustering high-dimensional data," *Mach. Learn.*, vol. 82, no. 1, pp. 43–70, 2011.

[59] P. Agarwal, S. Mehta, and A. Abraham, "A meta-heuristic density-based subspace clustering algorithm for high-dimensional data," *Soft Comput.*, vol. 25, pp. 10237–10256, 2021.

[60] Z.-H. Zhan, J.-Y. Li, S. Kwong, and J. Zhang, "Learning-aided evolution for optimization," *IEEE Trans. Evol. Comput.*, early access, Dec. 29, 2022, doi: 10.1109/TEVC.2022.3232776.

[61] J.-R. Jian, Z.-G. Chen, Z.-H. Zhan, and J. Zhang, "Region encoding helps evolutionary computation evolve faster: A new solution encoding scheme in particle swarm for large-scale optimization," *IEEE Trans. Evol. Comput.*, vol. 25, no. 4, pp. 779–793, Aug. 2021.

[62] X. Xia et al., "Triple archives particle swarm optimization," *IEEE Trans. Cybern.*, vol. 50, no. 12, pp. 4862–4875, Dec. 2020.

Parham Hadikhani received the bachelor's degree in computer engineering in 2016, the master's degree in information technology engineering in 2019, and the Ph.D. degree in computer science, specializing in computer vision.

His research interests include but not limited to pattern recognition, deep clustering, symbolic regression, transfer learning, and evolutionary algorithms.

Daphne Teck Ching Lai received the B.Sc. degree in computer science from Strathclyde University, Glasgow, U.K., in 2004, the M.Sc. degree from the University of Kent, Canterbury, U.K., in 2006, and the Ph.D. degree from University of Nottingham, Nottingham, U.K., in 2014.

She is currently a Senior Assistant Professor with the School of Digital Science, Universiti Brunei Darussalam, Bandar Seri Begawan, Brunei. In 2018, she was with Hosei University, Tokyo, Japan, under the Hosei International Fund Foreign Scholars Fellowship. Her research interests lie in data mining, artificial intelligence, and metaheuristics.

Wee-Hong Ong is currently an Assistant Professor in computer science with the School of Digital Science, Universiti Brunei Darussalam, Bandar Seri Begawan, Brunei. In particular, he is exploring the application of artificial intelligent techniques and info-communication technologies in developing intelligent systems. His current projects include unsupervised human activities recognition, file-based IoT system, mobile robot navigation, human emotion perception for human.robot interaction and self-supervised object recognition. His research focuses on the development of intelligent systems for personal robots and ambient intelligence.