**1. Introduction**

1

Nowadays, the global positioning service (gps) has become a trend which let people could use it for locating their locations, and the extended applications of gps includes ubereat , foodpanda significantly affect people’s life.

Although gps has high precise outdoors, it’s performance could be influenced due to the signal obstacles by buildings and walls so that it’s inappropriate for indoor positioning.

2

Before introducing the techniques of indoor positioning, we first introduce the applications based on it.

First, department store could send the coupons to customers by knowing their positions.

Second,**indoor positioning can also help to track the movement of patients in hospitals.**

Third, high precision positioning could help personnel control in factory, especially the dangerous zone.

Last, it could be used in exhibition such as gulide tour in museum.

3

Currently\*\*, the common approachs of indoor positioning is mainly divided into two categories.The first is positioning based on wireless signals, and the second is based on images.\*\*

We first introduce the Wireless based solution, there are several signal based methods shown below. Wi-Fi, Zigbee, Bluetooth, Ultra-wide band and infared.

Howeber, these methods have the following problems. Wi-Fi Zigbee and Bluetooth could not have high precise positioning, and although uwb and Ir have high precision, they both need very high cost of environment construction and additional equpments on human, and environment noise have dramtical impact on all of them.

4

The second positioning method is based on computer vision.

The pedestrian position estimation and positioning method captures the foot point position of the person in the image through surveillance cameras, and then use 2D direct linear transform to the coordinates of the human feet position in the image into plane coordinates. This method is usually based on deep learning. The recognition model, when the environment changes, as long as the person appears in the image, the human body can still be positioned, so it is suitable for complex environmental backgrounds. Pedestrian estimation and positioning usually has high accuracy, but it cannot identify the identity of the person. When the human body is blocked by the occluder, it causes to reduce the accuracy, so an additional camera might require to avoid the occlusion of the occluder.

5

Next, this table shows the comparison between wireless signal based and computer vision based solution.

Wireless signal positioning is mainly based on mathematical models and deep learning models are used for images. Therefore, pure mathematical model is faster than images in terms of speed. Regard accuracy, wireless signal positioning is usually used for rough positioning, and the accuracy falls within 2 to 3 meters. Vision based positioning is used for high-precision positioning, and the error falls within 0.5 to 1 meter. Next is the requirement for additional equipment, wireless signal positioning generally needs user to carry additional devices for data collection, such as UWB and infrared, but images are monitored by cameras without the need for additional equipment, so wireless signal based solution can be used to identify through device information.

6

In recent years, due to people's attention to security, the global monitor will exceed one billion in 2021, and wifi is everywhere as the Internet of Things devices. Therefore, through these existing devices, we can collect these data without adding additional devices, and use existing resources to locate person.

Our motivation is to combine the advantages of wireless based and vision based for positioning. Although Wi-Fi can identify the user's device, it has low accuracy, and pedestrian location estimation has high accuracy, but it cannot get the information of the user's device. Therefore, this study will combine wifi and image to establish a two-stage positioning system. Firstly, wi-fi is used for user identification and rough location, and then the pedestrian position estimation method of image is used to achieve accurate location, finally these two are matched to get the precise position.

7

Our proposed method use smart phone to recieve the surrunding wifi signals strength and send it to our system. At the same time, camera captures the image and reconize the feet position of people in photo,and transform the image coordinates to plane coordinates and match the wifi positioning result and image positioning result to get the final posion

# 2. Related work

In the related work, we’ll introduce the techniques about the pedestrian location estimation.

The traditional method to detect human is background substraction, it is based on a Gaussian Mixture Model for background [modelling. It](http://modelling.It) will generate a realtime changging model from background of images, and Human part is computed by subracting the current frame from the background model. The shortcoming of the method is evident in the case of shadows or sudden change in illumination.

To solve the occlusion and environment problem, bi-box regression was invented to perform both Pedestrian Detection and Occlusion Estimation. It generates two bounding boxes: one for the Full Human body and the other for the Visible Part of the Body.The model structure is divided into two branches for different processing. One of their outputs is the classification and the other is the regression of the bounding box.

Although bi-box regression can solve the occlusion problem, it’s performance is not well in real-time of many people. When the number of human bodies increases, it is necessary to continuously detect people and then detect the complete human body, which leads to slower speed.

However, the OpenPose pose estimation model solves this problem. The method it adopts is using Bottom-up method to get keypoints and connect body points. that is, the key points of the human body are detected first, and then the key points are connected.

OpenPose can not only identify the streaming images in real time, but also distinguish each human body in the complex background of many people, which solves the problem of occlusion and overlap, that’s why we use this method as the image processing method

# 3. Proposed Method

We have three main objectives.

First is Using machine learning models to achieve highly accurate Wi-Fi positioning.

This graph shows the architecture of our offline system.First, user use smartphone to receive wi-fi rssi for 1 minutes, and send the request to backend api.Then backend api will save wi-fi signal strength and capture the image from monitor at the same time.Then extract human posture and save all the data in database

In the online flow, user still use smartphone to recieve wi-fi rssi, and post it to the backend. Then system will first predict the wi-fi position by using wi-fi rssi, and capture the image, and extract several human postures then used machine learning model to predict many positions. After caclulating these two position, the system will do many-to-one match and return only one position to smart phone.

The following picture introduces the flow of our feet position estimation flow. First,we use openpose model to perform posture estimation, and then the detected key points are calculated to obtain the pedestrian feetposition position. In order to make the foot points more accurate, we propose two methods for pedestrian foot point extraction, the white point is the first one, and the blue point is the improved one.

And this the algorithm flow chart.

In order to convert the coordinates of the foot point in the image to the coordinates on the plane, we need to convert through 2D direct linear transformation. Direct linear transformation can map the correspondence between two similar distributions.

The formula is shown on the lower left. The coordinates of the image uv can be transformed to the plane coordinate xy through the H matrix. In order to obtain the H matrix, we bring the four sets of plane coordinates xy and the coordinate uv of the image into the equation, we can get the formula on the right, solve the h matrix, and finally we can Transform the uv coordinate to the xy coordinate by the h matrix.

These red points are the reference points to solve the h matrix.Then we can use the fomula to transform from uv points to xy coordinates.

From the previous fomula we could use the 2D direct linear transformation to transform the points on the floor to plane coordinates. However, if the feet position extrated from our feet position algorithm is not on the floor, it causes big error. As you can see in the picture, after performming the direct linear transformation, the plane coordinates is far from the actual position. Therefore, we improve the previous feet position estimation algorithm to get the more accurate feet position.

These five steps are as same as the previous ones. If we find ankles we return middle point of ankles. Then if we find knees or bows, we take the x-coordinate of the neck as x and take the y coordinate of neck and extend twice distance from bow to neck.The white point is the previous algorithm and the blue point is the imporved one. You can see the blue point is more close to the floor.

Then if we find neck, we also take x-coordinate of the neck as x, and extend third distance from neck to nose as y-coordinate.

We perform the 2D DLT to different feet positions, you can see after we improve the feet position algorithm, the blue point is much closer to actual position.

Although our improved method can shorten the error, it still have some distance error.

The result will be shown in the experiment part.

Next, we introduce the match approach of wi-fi and pedestrian. The orange points represent the pedstrian from the image, so there are many orange points.The blue point is wifi position calculated from the recieve signal from smart [phone.](http://phone.So) So it’s only one point. The method we use to match them is simply using the e**uclidean distance. We find the min distance between wi-fi position and pedestrian position. In the example, we take p1 as the user’s postion.**

We take our lab as the experiment area. The background is complicated and people here could be blocked by computers. And the space size is 6.8 meters cross 5.8 meters. And we collect 75 sets of coordinates for data collection.