

Aalto University
School of Science
Master's Programme in ICT Innovation

Lichang Zhang

Non-invasive detection algorithm of thermal comfort based on computer vision

Master's Thesis

Stockholm Sweden, June 21, 2018

Supervisor: Haibo Li, Professor of KTH (haiboli@kth.se)

Aalto Supervisor: McGookin David, Associate Professor

Thesis advisor: McGookin David, Associate Professor

Abstract

The waste of building energy consumption is a major challenge in the world. Real-time detection of human thermal comfort is an effective way to deal with this issue. However, due to the difference of personal thermal comfort and changes caused by climatic variations, there is still a long way to reach this target. From another perspective, the current HVAC (heating, ventilating and air-conditioning) systems are reluctant to provide flexible interaction channels to adjust atmosphere which fails to follow continuously increasing requirements from users. All of them indicate the necessity to develop more intelligent detection method for human thermal comfort. In this paper, a non-invasion detection method toward thermal comfort is proposed from two perspectives: macro human postures and skin textures. In posture part, OpenPose is used for detecting the key points' position coordinates of human body in images, which would be functionalized from the term of thermal comfort. In skin textures, deep neural network is used to regress the images of skin to its temperature. Based on Fanger's theory of thermal comfort, the results of both parts are satisfying: subjects' postures can be captured and interpreted into different thermal comfort level: hot, cold and comfort. And the absolute error of prediction from neurons network is less than 0.125 degrees centigrade which is the equipment error of thermometer used in data acquisition. With solutions of this paper, it is promising to non-invasively detect the thermal comfort level of users from postures and skin textures. And the conclusion and future work are discussed in final chapter.

Keywords

Non-invasive detection; deep learning; OpenPose; computer vision; human thermal comfort

Table of Contents

1	Introduction.....	2
1.1	Background	2
1.2	Problem.....	3
1.3	Purpose	4
1.4	Goal.....	4
1.5	Methodology.....	4
1.6	Delimitations	4
1.7	Outline.....	4
2	Literature study.....	6
2.1	Thermal comfort.....	6
2.2	Deep learning.....	6
2.3	OpenPose.....	7
3	Detection based on OpenPose	8
3.1	Introduction:.....	8
3.2	Methods:	9
3.3	Solution:.....	11
3.4	Results:.....	12
3.5	Summary:	13
4	Detection based on skin textures.....	15
4.1	Introduction	15
4.2	Methods.....	15
4.2.1	Algorithm process	15
4.2.2	Data preprocess	16
4.3	Summary.....	22
5	Conclusion	24
5.1	Evaluation	24
5.2	Limitation.....	24
5.2.1	Personality	24
5.2.2	Image quality.....	25
5.2.3	Mini-actions	25
5.2.4	Varies in instances	25
5.2.5	Key area measurement and data preprocess	25
5.2.6	Previous state.....	25
5.3	Future	25
5.3.1	Data gathering	26
5.3.2	Multi-persons.....	26
5.4	Summary.....	26
	References.....	28

1 Introduction

1.1 Background

Thermal comfort[1] is one of key perspectives in improving user experience and reducing energy consumption. The energy consumption of commercial and residential buildings take 21% of the total energy consumption [2]. More essentially, it will increase with 32% annual rate [2]. From [3] and [4], even a slight indoor temperature adjustment, for example, 1 degrees centigrade, has a distinctly impact on energy consumption of the entire building. Not limited in unnecessarily resource costing, 30% of greenhouse gas emissions from commercial and residential buildings [5] point out the urgency for improving resource usage. While from the term of user experience, the definition of thermal comfort is still limited in area such as architecture and quiet unfamiliar to end users. In current HVAC systems [6], the operating is usually restricted with limited options. Most of those options are usually settled based on experience and do not obtain capability to reflect the flexible demand of current users. Although interaction media such as remote controllers, buttons are available in some HVAC system such as air conditioning; they are still unfriendly to users from the term of usage. More importantly, since the particularity of surrounding condition, even with same setting, the performances of HVAC systems are unstable.

To this issue, various kinds of detection methods were developed. They can be divided into three catalogues, namely invasive, semi-invasive and non-invasive. Invasive detection means a part of equipment into the body cavity during inspection. And the nature of this detection strategy makes it hard to be proceeded in the field of thermal comfor. So in semi-invasive detection, the improvements were dramatically reflected in the reducing on size, weight, electromagnetic of equipment.

In part of semi-invasive, iButton (information button) can be regarded as symbol solution which has features such as tiny in size (17.35mm*3.1mm~5.89mm), anti-water, anti-folded, etc. Via using iButton temperature sensor to touch users' skin, the interruption during detection process would be significantly eliminated.



Fig.1 (iButton)

And in [7], an eyeglass frame with infrared sensors installed was used to detect the flow of skin blood (Fig. 2). Although real-time monitoring of the facial skin blood flow can be achieved, contact with users' body is still unavoidable. Not to mention in order to summarize the temperature variances of different part of body, several sensors were used and located on interviewees' body, DuPont lines were used for data transport. All of those would hugely interrupt users.



Fig. 2. (a) Approximate infrared sensing locations on face and (b) 3D view of sensing device with four infrared sensors.

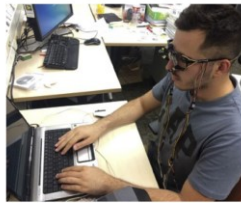


Fig. 3. A male participant wearing the glass frame with infrared sensing system.



Fig. 5. Office space floor plan.

Fig.2 (Invasive detection method)

So with all examples mentioned above, it is safe to say a reliable and high accuracy non-invasive detection method is still waiting for developing and implementing in real life.

1.2 Problem

The reasons for explaining the conflicts mentioned above are related to various perspectives. First, thermal comfort is diverse toward different individual, and even to one specific subject, the feeling toward comfortable would also be changed for surrounding atmosphere, season, mental state, motion state and so on. Users' body temperature instead of thermal comfort should be focused as key feature. Even strong correlation between the two features, body temperature cannot be directly interpreted as thermal comfort and this distortion would definitely affect user experience. Second, according to Fanger's theory of thermal comfort [8], various factors make an impact: the outdoor temperature, the amount of users, humidity, light and so on. Third, to maintain ideal thermal comfort level is a relatively long process. How to continuously obtain data for analysing while eliminating the interruption toward users is a challenge. Fourth, in public scenarios, because of the flexible of users, traditional measurement methods are hard to implement for issues such as privacy, maintenance of equipment and so on. Finally, even remarkable business potentials are mentioned, the direct value brought via improving thermal comfort level is trivial especially compared to cost to upgrade current equipment.

In a word, solutions which can analysis the thermal comfort non-invasively are demanded urgently for improving the level of thermal comfort and economize the resource.

1.3 Purpose

Focus on this limitation, this paper proposes a non-invasive method to interpret the users' thermal comfort from two perspectives: postures and skin temperature.

1.4 Goal

And via the solution of non-invasive measurement method, the goal is to improve or offer a new possible strategy to update the current HVAC system for improving the experience of thermal comfort and reducing the cost of resource.

1.4.1 Benefits, Ethics and Sustainability

First, the consumption of energy, emission of greenhouse gas would be dramatically reduced, which brings value from both economic and environment.

The user experience would be improved via quickly adjusting the surrounding atmosphere into perfect condition. Users can get rid of remote controllers, buttons even smart phone and ignore the numbers on screen, tell their wants via simplify stay there.

1.5 Methodology

In data collection, 6 semi-interviews were preceded for obtaining information toward thermal characteristics and recording videos about interviewee' habitual behaviors in different environment, for example, hot summer with sunshine, cold winter night, rainy afternoon in later autumn and so on.

And thanks to Xiaogang Cheng, the supervisor of this master project, for providing the data used in skin texture part [9].

In data analysis stage, CNN [10] was used to predict the degrees centigrade of human hand picture. And in the process of training model, control variable [11] was implemented from the perspectives of the depth of CNN network, the parameters for training such as learning rate, drop out value [12], the size of input image.

1.6 Delimitations

The charm of this solution is to incorporate deep learning algorithm into current camera system and brings remarkable improvement to users' experience. And in this structure, users can get rid of various kinds of controller and enjoy what they want without any command. Effort to reduce the size of measurement equipment would be efficiently re-distributed for developing strategies in more complex situations.

1.7 Outline

In this paper, three key definitions will be introduced: thermal comfort [1], deep learning [13] and OpenPose [14]. Thermal comfort is a subject definition to interpret the abstract, personalize thermal concept. And in this paper, to simplify the problem, object parameters were researched to explore users' thermal feeling. And OpenPose is used to detect the key points of human body in real-time. Besides via deep learning, especially neural network, the temperature of human skin can be predicted with common RGB images.

The solutions also present in two parts: first, in the section of detection based on OpenPose, whole process from the data gathering and analysis, results display and to discussions would be illustrated. And the second part: detection based on skin temperature follows the same structures. Finally, the comparisons between both parts and future works will be discussed.

2 Literature study

In this chapter, 3 parts will be illustrated: thermal comfort, deep learning and OpenPose [14]. Thermal comfort is the key definition to whole project, based on the theory of Fanger [8], it is not simply equal to one object value such as temperature and analysis from subject terms. And deep learning provides powerful tool to receive image as input and adjust thousands of parameters for regressing to target value. Finally, OpenPose was used as tool for capturing human key points in image, which built fundamental for further posture analyze.

2.1 Thermal comfort

Thermal comfort is the condition of mind that expresses satisfaction with the thermal environment and is assessed by subjective evaluation [1]. The main factors that influence thermal comfort are those that determine heat gain and loss, namely metabolic rate, clothing insulation, air temperature, mean radiant temperature, air speed and relative humidity. Also, psychological parameters, such as individual expectations, affect thermal comfort.

In naturally ventilated buildings, occupants take numerous actions to keep themselves comfortable when the indoor conditions drift toward discomfort. Operating windows and fans, adjusting blinds/shades, changing clothing, and consuming food and drinks are some of the common adaptive strategies [4]. The thermal sensitivity of an individual is quantified by the descriptionFs, which takes on higher values for individuals with lower tolerance to non-ideal thermal conditions [5].

In this project, based on the theory of Fanger [8], thermal comfort is defined in three different levels: hot, comfort and cold. The standard to judge is based on the posture instead of one specific factor. The knowledge of thermal postures came from interviews and data from Internet, while choosing postures as standard to judge thermal comfort means non-invasive measurement can be realized in common RGB camera continuously.

2.2 Deep learning

Deep learning (also known as deep structured learning or hierarchical learning) is part of a broader family of machine learning methods based on learning data representations, as opposed to take-specific algorithms. Learning can be supervised, semi-supervised or unsupervised [6]. More specifically, deep neural network (DNN) which refers to the neural network with multiple hidden layers between the input and output layers [7] increases the parameters of model.

During training process, three key definitions are essential, namely, back propagation [15], stochastic gradient descent (SGD) [16], and drop-out rate [12].

Back propagation is a method to compute the gradient between expected labels and output of algorithm especially in multi-layers neural network. The gradient used as the direction to update the parameters of model to minimize the difference.

With SGD, it allows algorithm to update parameters with a few examples of input via computing the average gradient. The average gradient indicates the direction of update. This method requires the training data should be unbiased and sufficiently shuffled so that each small set of examples gives a noisy estimate of the average gradient over all examples [16]. The rate of updating parameters decided by learning rate and since the nature of SGD to train with limited part of training data set, in most situations, it would drops into local optimum.

And drop-out rate is a regularization technique to reduce over fitting via drop out units in neural network randomly with pre-defined rate.

Since the space limitation, no more definition will be mentioned in this part, for more knowledge about deep learning, [16] would be perfect introduction.

2.3 OpenPose

OpenPose represents the first real-time multi-person system to jointly detect human body, hand and facial key points on single images [8].

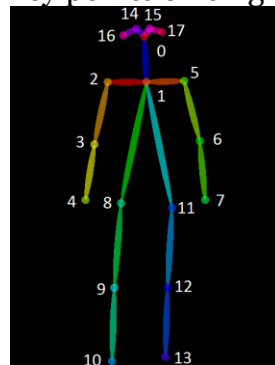


Fig.3 (Human body Key points defined by OpenPose)

Image, video, webcam and IP camera can be used as input, and OpenPose support videos and images as output with labelled key points, also key point saving with format such as json, xml and so on.

With OpenPose, we can reduction the dimension of images from all pixels points to one array with coordinate value and confidence coefficient, which not only reduce the dimension of input, but also release the load of data transportation between OpenPose server and remote camera.

3 Detection based on OpenPose

3.1 Introduction:

This part is focus on interpreting thermal comfort via postures. In literature study part, some postures such as operating windows and fans, adjusting blinds/shades, changing clothing, and consuming food and drinks are common for people to adjust temperature level around body. And all those actions are subjective adjustments from users. However the goal is to focus on the actions can be finished in relatively short time slot, have high individual difference and some of them even are unconscious and interpreting their meanings related to thermal comfort.

Postures toward environment temperature also hugely affected by factors such as users' gender, cultural background, emotion, role in current situations.

All those features mentioned above, the challenges in reflecting thermal comfort with postures would be:

- First, define common and meaningful features to reflect users' thermal comfort situation.
- Process and analysis a serial of images within a time slot for capturing users' actions related to thermal comfort.
- Minimal the delay of whole process and keep the accuracy of prediction.

The whole process of detecting postures solution is introduced in Fig.4. First is to have basic information about people postures especially those related to thermal feelings, and the channels of gathering materials were Google image engine and Youtube. With all information gained from first step, six semi-interviews were proceeded to have more knowledge about people thermal postures in indoor environment, and information about their preference and feeling toward this topic.

With all steps mentioned above, five postures which are popular both in Internet and interviewees were summarized and functionalized via the coordinates of human key points. All postures have specific meaning to different thermal comfort level, some of them are single posture, can be captured with single image, while others are actions which combined with a serious of images.

Via successfully captured and interpreted the postures, feedback from our solution can offer thermal comfort level of users under the systems' camera.

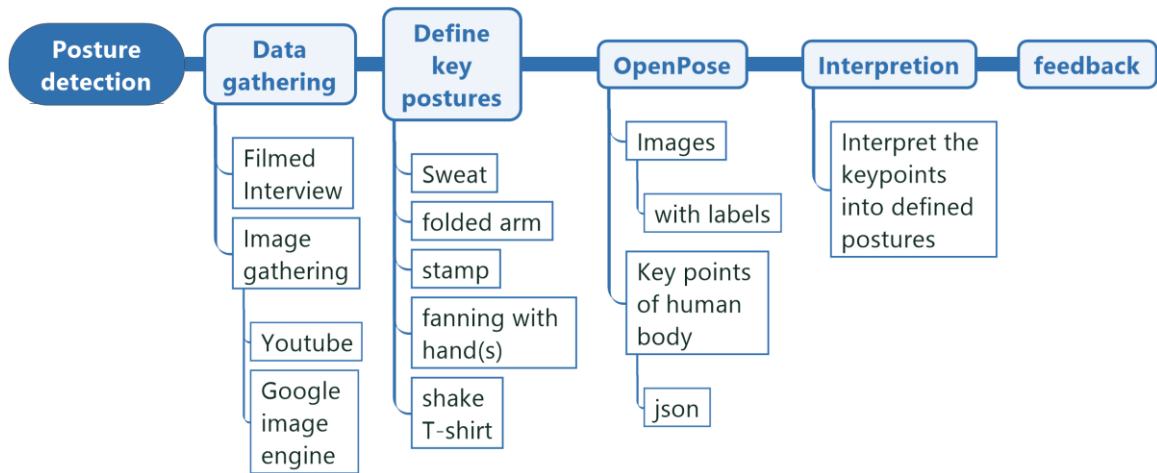


Fig.4 (Process of posture detection)

3.2 Methods:

3.1.1 Data gathering and analysis

Two channels are used to gather data. First one is to gain image or video from YouTube and Google image engine. And with knowledge from first part, filmed interviews are not only gain images material, but also get the views of interviewees.

Before interview, the candidates of thermal postures selected from YouTube and Google image engine are used as options in interview.

And from six interviews (three males and three females, aged from 20 to 30, four Asian and two European) and gathered materials, common and typical postures are being chosen.

In interview, first common postures were took, then, in two postures (standing and sitting), and five scenarios (hot, slight hot, comfort, slight cold and hot), subjects were told to perform the postures they would use. If nothing came to mind, prepared postures were displayed to inspire subjects.

3.1.2 Define key postures

With all data, five key postures are defined and introduced as following. Shaking T-shirt (Fig.5) is to have more cool air flow through skin and clothes. And this frequency of using this action is strongly related to gender, it is males' favourite, especially youth male, while to rule for female. Fig.6 refers to the meaning stamp feet, or in other word, the crossing angle between the first lines: haunch to knee and second line knee to ankle is changed at static value. In this definition, jump and crouch are not qualified. Fig.7 defines as hand folded to the offside elbow. While Fig. 8 defines as left or right hand used to swap the sweat of forehead. Use hand to fan (Fig.9), usually in the same side. Since the hand(s) are dynamic in the process, in our solution, multi-images

are essential for the prediction to prevent situations such as raise hand for answer questions, use hand to hold head and so on.

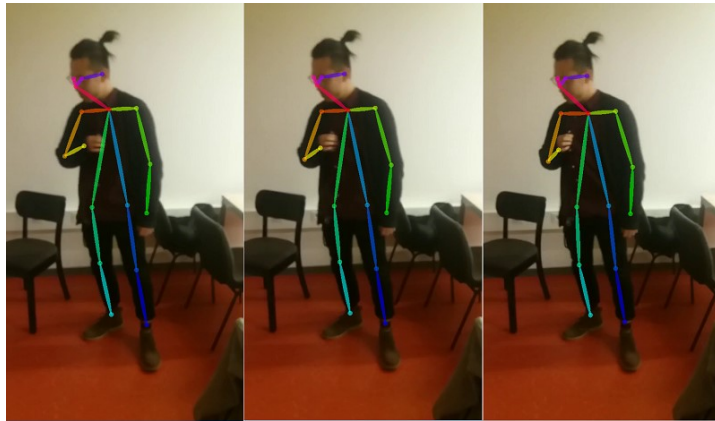


Fig.5 (Shaking T-shirt)

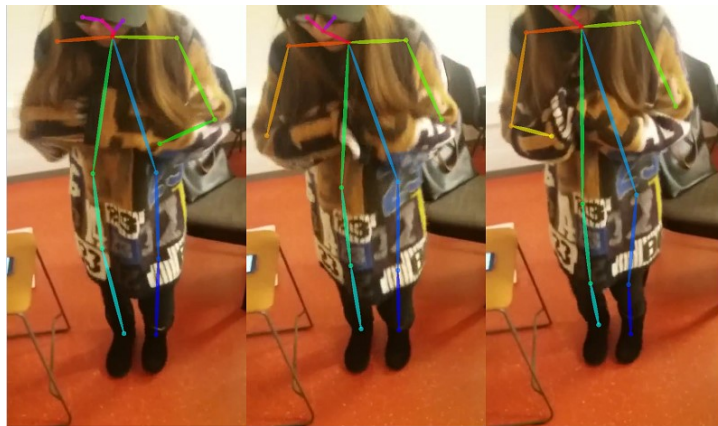


Fig.6 (Stamping feet)

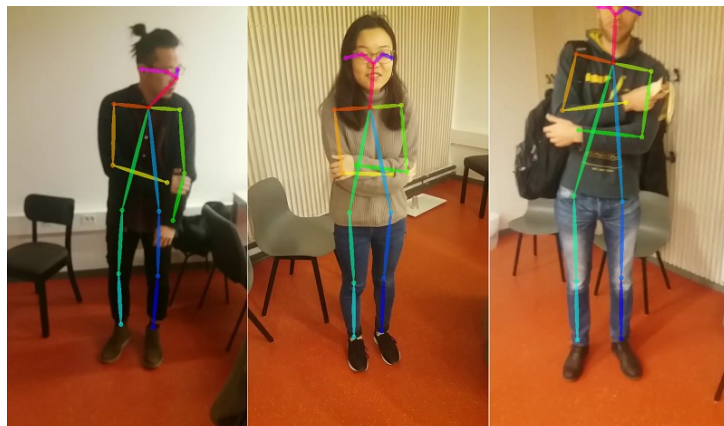


Fig.7 (folded arm(s))



Fig.8 (Fanning with hand(s))



Fig.9 (Swap sweat)

3.3 Solution:

The solution executed with CUDA and OpenPose. Structure of the solution displayed in Fig. 10. Main process is assigned for UI display; adjust the workflow between OpenPose and Analysis module. First sub-process is responsible to enable OpenPose, generate data such json (include information related to key points) and image. Via analysing the files produced in first process, second process can offer feedbacks. During the generating, cleaning the old files and limited the space usage is also considered.

In analyzing, a serial of images would be checked (usually in 60 images, 2 seconds, and 30 frames for 1 second). In recognize actions, it is not a strict real-time response, however users hard to distinguish the delay.

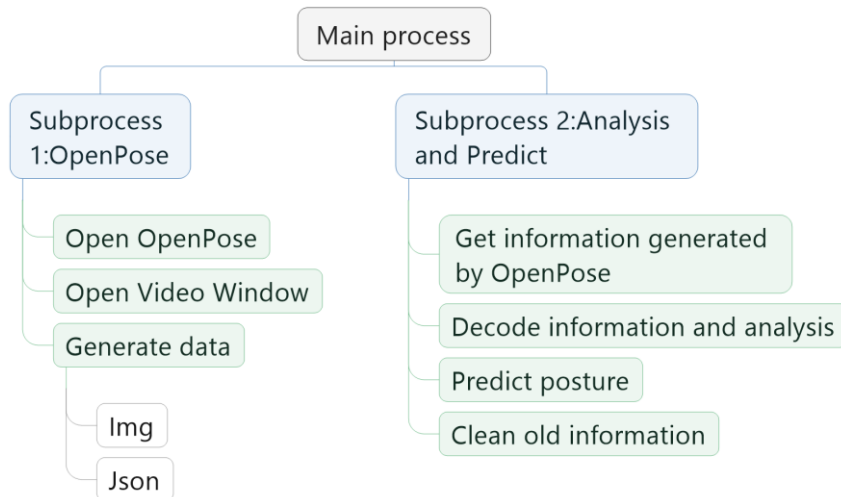


Fig.10 (Process of postures detection)

3.4 Results:

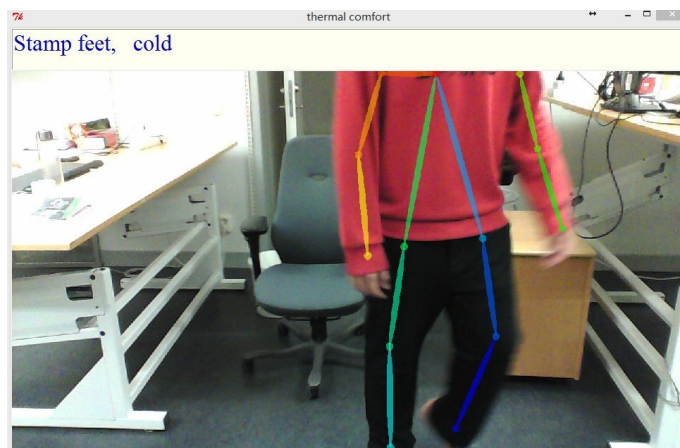


Fig.11 (detect result part 1: stomping foot))

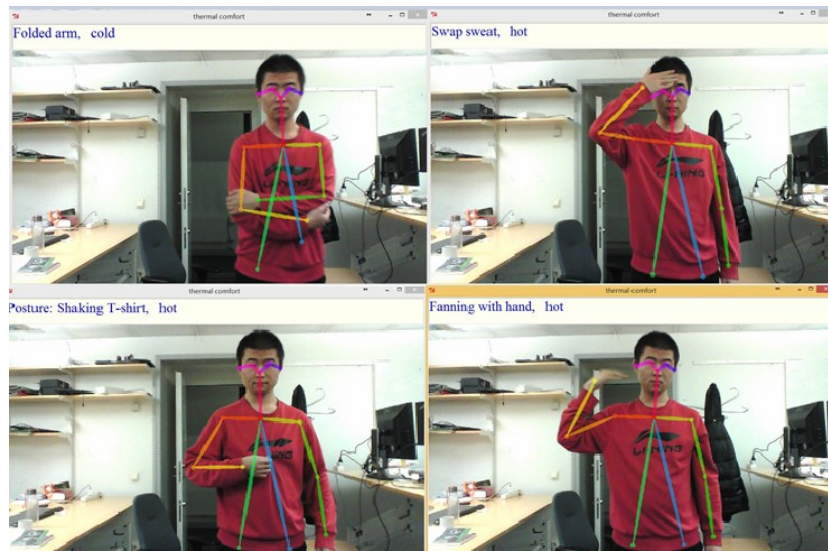


Fig.12 (detection result: part 2)

The feedbacks toward 5 key postures are displayed above. The display window is consist with two parts, blue text as headline to show the name of action being captured and the level of thermal comfort being understood from the captured posture, the action frame also displayed with colourful labels on key postures, which thanks to OpenPose[14] to offer this format of output.

Both parts are real-time response based on input video stream. With the response, it is safe to say the solution can detect the 5 defined postures, and offered feedback with key frame of making judgment.

3.5 Summary:

To pre-defined 5 postures, the solution can keep high accuracy toward postures and offer meaningful feedback to users' thermal comfort.

The analysing starts from RGB images token via common cameras and the whole process can be processed non-invasively. It means our solution can be easily corporate into systems such as security monitoring systems. And to scenarios which have rapid flow of users, it can realize high sensitive response to current users. Besides the performances of solution are quiet stable in 3 types of image quality: 640 pixels * 480 pixels, 1080 pixels and 720 pixels and 1920 pixels and 1080 pixels.

However since the limitation in data for analysis, only 5 postures are being normalized and can be recognized. And the 5 postures cannot reflect users' personality. Even within 5 postures, the probability for user to act would still be diverging in factors such as gender and so on.

In 5 key postures, the detection can be divided into two types: detection based on single frame, which refers to folded arm and swap sweat and detection based on a serious of continuous frames, which are as mentioned above 60 frames for detecting one action. The number is based on the trade-off between delay and accuracy of detection. And the delay is consisting of gathering

images and calculating cost. The 2 seconds delay needs more attention to reduce.

Finally, no feedback is available for any other postures. Even personalized postures are offered for specific users', it still hard to guarantee the accuracy for small movements. And since the method of functionalize is based on the relative position relations among human key points. It can guarantee high accuracy of recognizing defined postures, while pain to some postures with similar position relations but have totally different meaning.

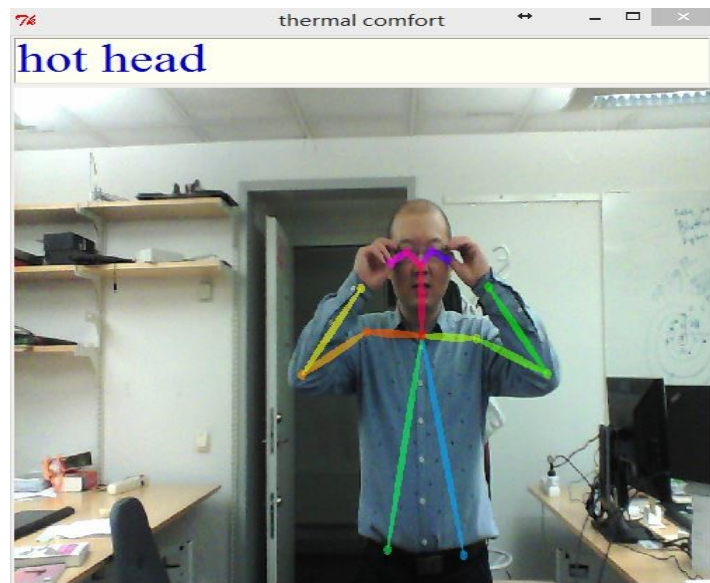


Fig.13 (False detection)

In graph, the solution falsely recognized the posture of adjusting glasses to swap sweat. So how to avoid the affection from false recognition is a key issue to current solution.

4 Detection based on skin textures

4.1 Introduction

Toward the limitation in previous solution, in this part, deep neural network is used to predict the temperature of skin. So from perspective of skin temperature, we assume human should have an ideal temperature which they can achieve highest thermal comfort level. And via predict the skin temperature and make sure it varies the ideal value; we can guarantee the users' thermal comfort level.

The dataset is also used in [9], and generated from two terms: strong simulation and weak simulation. In strong simulation, first to put the two hands of subjects into hot water (45 degrees centigrade), then record video of the back of hand and log the temperature value. While in weak simulation, a warmer is used to improve the skin temperature slowly.

16 instances are available for both. And all of them are Asian youth female. The duration of video in strong simulation is longer than 48 min, and longer than 40 min in the weak simulation. The original video size is 1920*1080, 30 frames per second. The equipment error of temperature value is 0.125 degrees centigrade.

The general process of skin detection is mentioned in graph. First part is data preprocess, and the goal is to transfer original data into format which can be used as input for deep learning model. Second stage is choosing a proper model which the absolute error of prediction can be reduced to 0.1 degrees centigrade. While during the training, lots of methods were introduced: normalization, data shuffle, decaying learning rate and so on. Finally, with the trained model, all images from 16 instances were predicted.

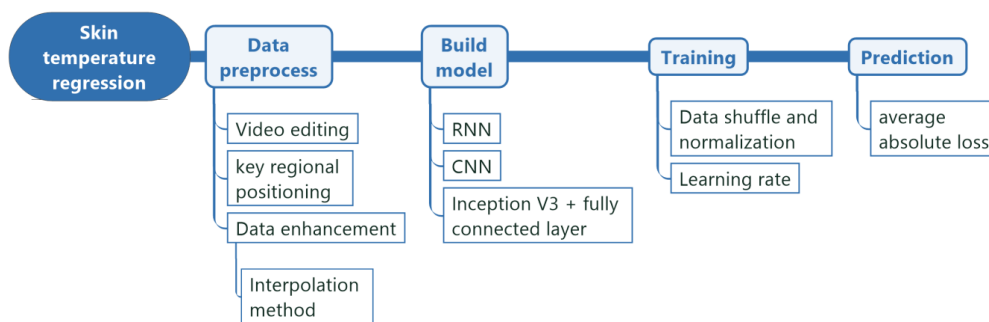


Fig.14 (Process of detection skin temperature)

4.2 Methods

4.2.1 Algorithm process

First, to every model, training toward single instances, record the rate of convergence in 10 rounds. Round means the times to train model with training dataset. And the goal of regression is the absolute error between prediction value and original value less than 0.1 degrees centigrade (the equipment error of temperature sensor). If the model fails to meet the requirement, increasing the size of images, or adding the levels and parameters of models were the methods for improving.

Divide all images as training set and validation set. Choose 100 images from validation datasets as test and record the absolute error of test images during training.

Saving the model after 10000 steps or the absolute error reach the expectation: less than equipment error, namely 0.125 degrees centigrade.

4.2.2 Data preprocess

4.2.2.1 Video preprocess

4.2.2.1.1 Normalize duration

The durations of raw videos are not exactly the same, and in the beginning of videos, the fluctuate brought by the movement of hands make the difference of images in this part are larger than images in other time period both in visualization and absolute value between images' pixels.

In the strong simulation, the duration of videos is shorten to 48min, the amount of data is: $1382400 = 48 * 60 * 30 * 16$ (48 minutes, 60 seconds per minute, 30 frames per second and 16 in all) .In the weak simulation, the duration of videos is 40 min, the amount of data is: 1141863 (the duration of instance 10 is 34min 21sec).

4.2.2.1.2 Key area

The original size of video is $1280 * 720$, and both hands were taken. The temperature sensor was addressed on the right hand of instance (left side) of images. And we assume both hands have same skin temperature.



Fig.15 (frame of original video, subject No.11)

And in the right half of original size, namely the left hand of instance, is the key element for further capturing.



Fig.16 (left hand of subject, subject No.11)

4.2.2.1.3 *Image size for processing*

First option is to take the whole right side of original size, namely 720*650, which keeps most information, and also requires more calculation resource.



Fig.17 (left hand of subject, size 720pixels * 650 pixels)

Second option is to take the back of left hand, and the size was 320*320. This size was considered for CNN network. The performance of CNN is better when the step of pooling in both x coordinate and y coordinate is 2. 320*320 not only guarantees the key information in original video, and supports up to 6 pooling layers.



Fig.18 (left hand of subject, size 320pixels * 320pixels)

And size of 160*160 pixels is the third options since it is hard to estimate the complexity of this regression assignment. And the equipment error in logging temperature was 0.1 degrees centigrade. If the neural network can reach the level that the absolute error between label and prediction is less than the equipment error, namely 0.1 degrees centigrade, with small images, it means

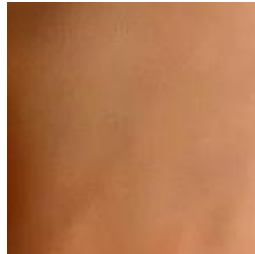


Fig.19 (left hand of subject, size 160pixels * 160 pixels)

4.2.2.2 Data enhancement

4.2.2.2.1 Interpolation method

In original dataset, the logging information of temperature was being recorded per minute once, while the available frames in 1 minute would be 1800. And it is clear that the changing of skin temperature was a continuous process, with nearly single trend: in strong simulation the temperature went down when hands being moved out from hot waters (45 degrees centigrade), while in weak simulation, the temperature was slowly growing because of the warmers under hands.

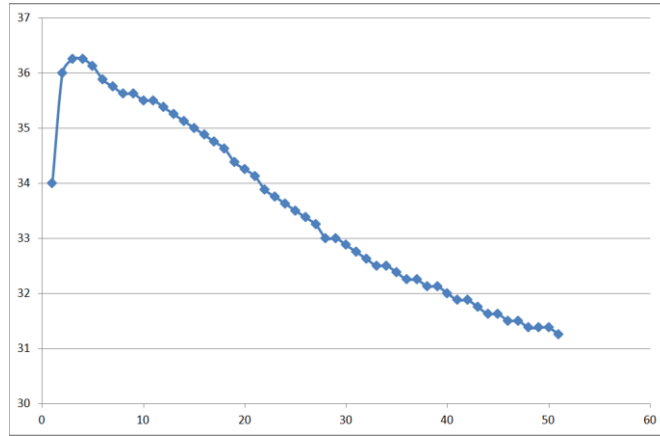


Fig. 20 (temperature label in strong simulation)

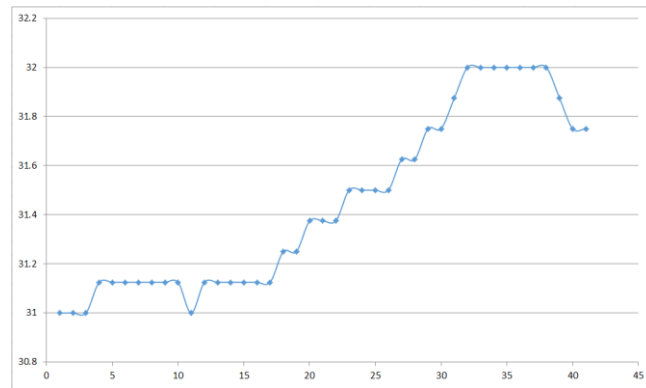


Fig. 21 (temperature label in weak simulation)

In order to reasonably label all frames, interpolation method was used to insert more temperature values via following formula.

$$gap = \frac{T_{i+1} - T_i}{12} \quad i = 0 \dots n - 1 \quad \text{Equation (1)}$$

$$T_{new} = T_i + gap * j \quad j = 1 \dots 11 \quad \text{Equation (2)}$$

In other words, after interpolating, one temperature value for 5 seconds, namely 150 frames use same value and the difference between neighbored values is less than 0.125 degrees centigrade.

4.2.2.3 Model design

4.2.2.3.1 Convolutional neural network (CNN)

CNN is a class of deep, feed-forward artificial neural networks. It consists of an input, an output layer and multiple hidden layers. The hidden layers of a CNN typically consist of convolutional layers, pooling layers, fully connected layers and normalization layers.

Since its superiority in processing image information, especially images' details and textures, we assume CNN can perform its ability in analyzing the change of images brought by the vasodilatation and vasoconstriction of instances' left hand.

● Parameters

Layer		Layer Type	Hyper-parameters
Input		Input	Image size: 320 * 320
NC1		Normalization and convolution	Filter size:5*5 Filter number:32 Stride:[1,1,1,1] Activation Function: Tensorflow.nn.relu
P1		Max pooling	Pooling region size: [1,2,2,1] Stride:[1,2,2,1] Pooling method:max_pooling (tf.nn.max_pooling)
NC2		Normalization and convolution	Filter size: 3*3 Filter number:64 Stride:[1,1,1,1] Activation Function: Tensorflow.nn.relu
P2		Max pooling	Pooling region size: [1,2,2,1] Stride:[1,2,2,1] Pooling method:max_pooling (tf.nn.max_pooling)
NC3		Normalization and convolution	Filter size: 3*3 Filter number:64 Stride:[1,1,1,1] Activation Function: Tensorflow.nn.relu
P3	Max pooling	Pooling region size: [1,2,2,1] Stride:[1,2,2,1] Pooling method:max_pooling (tf.nn.max_pooling)	80*80 →40*40
Flatten layer	Flatten layer		API: tensorflow.reshape [batch_size,40,40,feature_point] (current feature points: 40*40*256 for single image)

Fully-connected-layer1	Fully-connected	Neurons number: 40* 40 *256*64 Dropout rate: 0.5 Activation function: Tensorflow.nn.relu	To [batch_size, all points after max-pooling] 40* 40 *256 → 64
Fully-connected-layer2	Fully-connected	Neurons number: 64	64 → 1
Output	Output	Output size: 1	Temperature float

4.2.2.3.2 Inception V3 model

Here we used pre-trained InceptionV3 model, its depth is 159 with 23852784 parameters. Inception V3, with its factorization features, increases the depth of network while speeds up the calculation. Since the size of image is relatively small, and the similarity among all images is quiet strong, the over-fitting issue during training process should be carefully considered. Compared with previous CNN model, we used 1 million images for training and left for validation.

4.2.2.4 Results

4.2.2.4.1 Strong simulation

- Inception V3

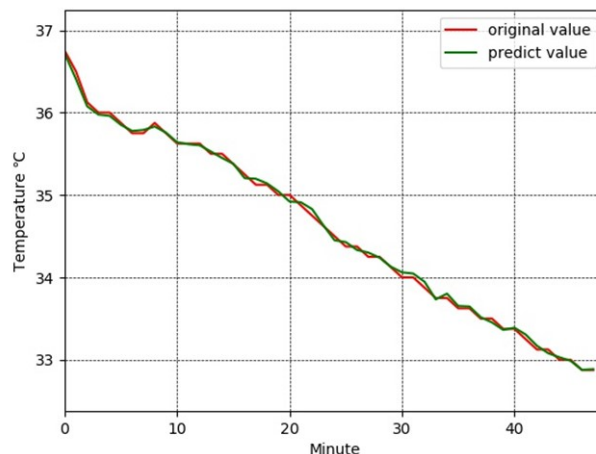


Fig. 22 (predict result, instance 11, strong simulation, Inception V3)
The rate between training set and test set is nearly 6:1; while in 2 round, the absolute loss of predict already reached 0.05 degrees centigrade, which far more less than the equipment loss of temperature sensor.

4.2.2.4.2 Weak simulation

- CNN

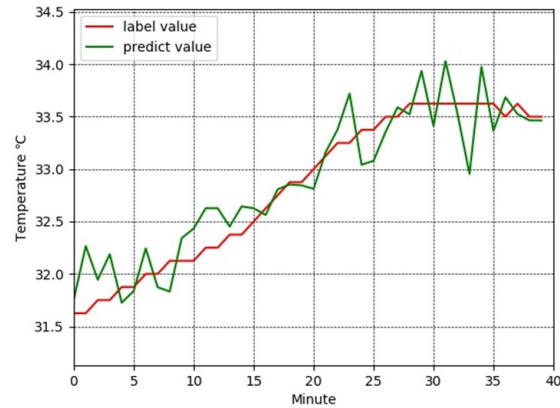
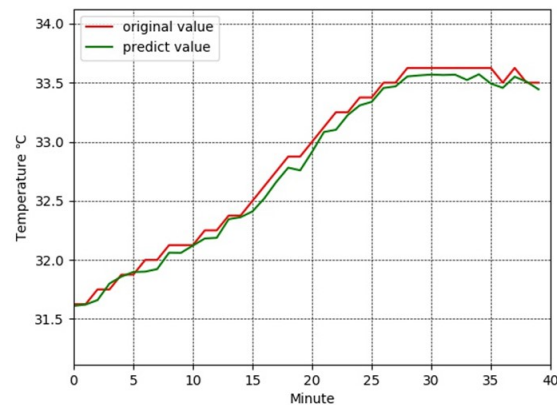


Fig.23 (predict result, instance 11, weak simulation, CNN model)

In weak simulation part, after 15 rounds and $1e-7$ learning rate, the absolute loss was reduced to 0.3 degrees centigrade. And the performance of prediction also related to the scope of temperature, in 32.0 to 33.0 degrees centigrade, the prediction was reliable, while in other range, the prediction was hard to be satisfied.

- Inception V3



- Fig. 24 (predict result, instance 11, weak simulation, Inception V3)

4.3 Summary

While the CNN reached the 0.3 absolute losses in weak simulation toward all instances after 15 rounds training. The images size for training is 320 pixels*320 pixels. While in strong simulation, the model fall into local solution

after 10000 training step: to all 100 images in testing set, only one value is predicted.

InceptionV3 reaches the goal in weak simulation in 2 rounds, and 3 rounds in strong simulation with nearly 85% of data as training set. So with all results mentioned above, it is promising to use RGB images to predict human skin temperature.

Nevertheless limitations also need to be mentioned here. First, all original data from 16 youth, female Asian, the loss of prediction among different gender, race is not guaranteed. And the skin temperature only can be used as one key feature to predict the thermal comfort level, more parameters are essential for improving the user experience. Finally, the changing of temperature was caused by simulators such as hot water and warmer. To common situations without obvious simulators, the changes of human skin image should be further researched.

5 Conclusion

5.1 Evaluation

In this paper, toward the goal of non-invasive thermal comfort detection method, solutions are introduced from two perspectives: postures and skin texture. Compared with current semi-invasive methods, the improvements are conspicuous.

First, all source of data is common RGB images, and all predictions and interpretations are generated from only images, which insure the solutions introduced in this paper are truly non-invasive.

Second, compared with previous solutions used temperature as the key even only one element, in postures part, people's actions and postures are being interpreted as feature to present users' thermal comfort via data mining and interviews.

Third, using deep learning to predict the temperature value of human skin is reliable and promising.

Fourth, non-invasive method using RGB images as input can show its superiority in spreading. Camera is one of most popular sensors which are equipped on computer, smart phone and varies computer vision system. It means improving thermal comfort can be easily embodying into nearly all cameras all over the world.

Finally, non-invasive method makes it attractive to scenarios such as shopping hall, public gym, libraries and stations. While with previous solutions, if thermal comfort wanted to be improved, sensors have to be spread out to every user, which causes potential problems such as the management of sensors, wear and tear of equipment, dangerous of spread skin disease. And now, with minimal consideration, the expectation to improve thermal comfort and reduce resource cost of HVAC would have more possibility to be satisfied.

5.2 Limitation

5.2.1 Personality

As mentioned in the previous part, how to personalize the postures of users especially in the scenarios of family and official are essential for improving performance. It requires a self-learning process: gathering videos of specific users in different temperature sections summarize the personal postures and definite as functions.

Besides, even to one specific user, the ideal temperature would be changed in different status, for example, health or in sick. And would the personal status

change his or her thermal comfort, if so, and the changes would happen in which perspective? All those problems still need more attention and efforts.

5.2.2 Image quality

Even 3 size of images being tested, and postures solution performs well to all of them. The challenges here is the available images in current system cannot be directly used to provide thermal comfort information. For example, the security cameras of shopping halls are usually located at extreme, where temperature is relatively low. Besides, the angles of images are various; some are recording from top to bottom, while others only take part of human body. In all situations mentioned before, the performance of our solution would be limited.

5.2.3 Mini-actions

The key for judge postures and actions is to have clear movement, so toward mini-actions such as curl lips, shake, it is hard to functionalize and corporate into current structure.

5.2.4 Varies in instances

The data for training was all from youth Asian female, so gathering more data with more variants in gender, race, cultural and education background would be essential for improving robustness of solution.

5.2.5 Key area measurement and data preprocess

Current the back of hand is defined as key area, and in the videos, the hand was clear took and stayed in nearly same position. In real world, it is hard to track user' hand and capture the key are since the varying in angle, being shadowed or not, strength and direction of light.

Compared with hand, the heads of people are rare being shadowed and can be easily tracked with facial recognition technologies. So capture human faces instead of hands via facial recognition solutions would simplify the complexity of realization.

5.2.6 Previous state

The change of temperature in data was generated by the hot water (45 degrees centigrade) and warmer. To confirm the change brought by simulation would also happen in daily life is important.

5.3 Future

To achieve the goal of improving user' thermal comforts, lots of efforts are still essential.

5.3.1 Data gathering

Lack of data is a key challenge in current solutions. In postures part, the materials gathered from Internet are mainly from outdoor, while the 6 interviews cannot support serials of postures for widely target users. While the data in detection skin temperature part was all from youth, female Asian. More data toward different gender, race and age is essential to research the ubiquity of solutions.

5.3.2 Multi-persons

It is not a challenge to distinguish multi-persons in single image and judge their thermal comfort, the issue here is to find a strategy to decide the output of HVAC system, or more generally, to one specific temperature, different people show various thermal comfort statuses, How to find a strategy for HVAC to decide the output is the issue for realization.

5.3.3 Feedback frequency of HVAC

The posture solution offer near real-time feedback and the duration to predict skin temperature from capturing, data process and model predict is minimal. Nevertheless the interval of adjustment toward HVAC should be relative long for avoiding meaningless the wear and tear of equipment and reducing cost of recourse.

Although lots of works need to be proceeded, it is still promising to use non-invasive thermal comfort detection to improve users' thermal comfort level, and reduce the resource cost of HVAC system.

5.4 Summary

In this article, non-invasive thermal comfort detection method is proposed for improving the user experience via combining existing HVAC system, and reducing the consumption of energy. Although lots of limitations prevent the solution to reach real life, it still novel in solving this challenge which is to combine the strength of computer visual and deep learning algorithm instead of rely on smart equipment.

Thermal comfort is the term which is hugely affect people daily experience, hope this project can introduce this definition to more people and contribute to the goal for dynamic resource distribution in HVAC system and satisfy personal preference especially in thermal comfort.

References

- [1] “ANSI/ASHRAE Standard 55-2013, Thermal Environmental Conditions for Human Occupancy.”
- [2] U. S. Energy Information Administration., “International Energy Outlook 2017. IEO2017 Report,” Sep. 2017.
- [3] A. Ghahramani, K. Zhang, K. Dutta, Z. Yang, B. B. Gerber, “Energy Savings from Temperature Setpoints and Deadband: Quantifying the Influence of Building and System Properties on Savings. *Applied Energy*,” vol. 165: 930-942, 2016.
- [4] A. Ghahramani, K. Dutta, Z. Yang, G. Ozcelik, B. Becerik-Gerber, “Quantifying the Influence of Temperature Setpoints. Building and System Features on Energy Consumption,” vol. 1000–1011, 2015.
- [5] “C. Initiative, Buildings and Climate Change,” 2009.
- [6] “Definition of HVAC.” [Online]. Available: <https://www.merriam-webster.com/dictionary/HVAC>. [Accessed: 10-Feb-2018].
- [7] A. Ghahramani, G. Castro, B. Becerik-Gerber, and X. Yu, “Infrared thermography of human face for monitoring thermoregulation performance and estimating personal thermal comfort,” *Building and Environment*, vol. 109, pp. 1–11, Nov. 2016.
- [8] Fanger, P.O, “Thermal comfort. Robert E. Krieger Publishing Company, Malabar, FL.,” 1982.
- [9] X. Cheng, B. Yang, T. Olofsson, G. Liu, and H. Li, “A pilot study of online non-invasive measuring technology based on video magnification to determine skin temperature,” *Building and Environment*, vol. 121, pp. 1–10, Aug. 2017.
- [10] DeepLearning o.1. LISA Lab, “Convolutional Neural Networks (LeNet) - DeepLearning o.1 documentation,” Aug. 2013.
- [11] “Variables in Your Science Fair Project.” [Online]. Available: <https://www.sciencebuddies.org/science-fair-projects/science-fair/variables>. [Accessed: 24-Apr-2018].
- [12] Hinton, Geoffrey E.; Srivastava, Nitish; Krizhevsky, Alex; Sutskever, Ilya; Salakhutdinov, Ruslan R., “Improving neural networks by preventing co-adaptation of feature detectors,” 2012. .
- [13] Y. LeCun, Y. Bengio, and G. Hinton, “Deep learning,” *Nature*, vol. 521, no. 7553, pp. 436–444, May 2015.
- [14] *openpose: OpenPose: Real-time multi-person keypoint detection library for body, face, and hands estimation*. CMU-Perceptual-Computing-Lab, 2017.
- [15] Y. LeCun *et al.*, “Handwritten Digit Recognition with a Back-Propagation Network,” p. 9.
- [16] “Large-Scale Machine Learning with Stochastic Gradient Descent | SpringerLink.” [Online]. Available: https://link.springer.com/chapter/10.1007%2F978-3-7908-2604-3_16. [Accessed: 25-Apr-2018].

