e-ISSN 1607-3274 Радіоелектроніка, інформатика, управління. 2020. № 3
p-ISSN 2313-688X Radio Electronics, Computer Science, Control. 2020. № 3

UDC 004.93

REFERENCE POINTS METHOD FOR HUMAN HEAD MOVEMENTS TRACKING

Rakova A. O. – Student of the Software Engineering Department, Kharkiv National University of Radio Electronics, Kharkiv, Ukraine.

Bilous N. V. – PhD, Associate Professor, Professor of the Software Engineering Department, Kharkiv National University of Radio Electronics, Kharkiv, Ukraine.

ABSTRACT

Context. The direction of the human face vector is an indicator of human attention. It has many applications in our daily lives, such as human-computer interaction, teleconferencing, virtual reality and 3D sound rendering. Moreover, determining the position of the head can be used to compare the exercises performed by a person with a certain standard, which brings us to investigation of ways to efficiently track moves. Depth-camera based systems, frequently used for these purposes, have significant drawbacks such as accuracy decreasing on the direct sunlight and necessity of additional equipment. The recognition from the two-dimensional image becomes more widespread and eliminates difficulties related to depth cameras which allows them to be used indoors and outdoors.

Objective. The purpose of this work is creation of the method that will allow us to track human head moves and record only significant vectors of head direction.

Methods. This paper suggests reference points method that decreases set of recorded vectors to minimal amount significant to describe head moves. It also investigates and compares existing methods for determining the vector of the face in terms of use in suggested approach.

Results. Suggested reference points method shows ability to highly decrease set of head direction vectors that describe the move. According to the results of the study, regression-based methods showed significantly better accuracy and independence from light and partial face closure so they were chosen to be used as methods to get head direction vector in reference points approach.

Conclusions. Research confirmed applicability of reference points method for human movements tracking and shown that methods of determining human head vector by two-dimensional image can compete in accuracy with RGBD-based methods. Thus combined with suggested approach these methods expose less restrictions in use than RGBD-based ones.

KEYWORDS: Face orientation vector, head moves, recognition, deep learning.

ABBREVIATIONS

CNN is a convolutional neural network;
FPS is a Frames per Second;
MAE is a Mean absolute error;
PAM is a parameterized appearance models;
RGBD is RGB image with depth information.

NOMENCLATURE

$\overline{a_i}$ is an angle vector on the i-th frame;

$\overline{a_i'}$ is a reference point vector;

(c_x, c_y) is a focal center;

e_i is an absolute error;

f_x and f_y are focal lengths in the x and y directions;

k is a flexibility degree;

n is an amount of observations in experiment;

r_i is an element of the rotation vector;

t_i is an element of camera translation vector;

U, V, W are positions of the object in three-dimensional space;

x, y are coordinates of the points in the image;

x_i is a real value;

y_i is a predicted value.

INTRODUCTION

Human face direction as a part of entire body pose determines positioning in exact time frame. Which means that sequence of face vectors is a move, thus ability to detect human face vector allows us to determine the accuracy of the exercises, which in turn is useful for rehabilitation institutions, fitness centers, entertainment facilities.

In recent years, methods for recognizing the face, its individual parts and looks have evolved in several different directions. In order to recognize human movement in games, the Kinect optical system was developed, which has gained popularity due to its high recognition accuracy. But this system has significant restrictions on use, such as the maximum number of people in the image and the inability to use it in direct sunlight. In parallel, a different direction of recognition was developed from two-dimensional images without the use of depth sensors. This area is also divided into different methods, but in recent years, CNN-based methods have shown the best results in accuracy and speed of image analysis.

Different approaches to determining the vector of face rotation use different metrics. There are two main ways to express the position of the head: the position of the camera relative to the head, the deviation of the head from the position full face.

In a computer vision, position of the object means its relative orientation and camera position. You can change the pose by moving the subject relative to the camera or the camera relative to the subject. If we want to express the position of the head so that the goal is to find the pose of the object when we have a calibrated camera and we

DOI 10.15588/1607-3274-2020-3-11

know the location of n 3D points on the object and the corresponding 2D projections in the image.

A 3D hard object has only two types of camera movement. Translation – moving the camera from its current 3D location *(X, Y, Z)* to a new 3D location *(X', Y', Z')* is called a broadcast. As you can see, the translation has 3 degrees of freedom – you can move in the X, Y or Z direction. The translation is represented by a vector equal to *(X' – X, Y' – Y, Z' – Z)*. You can also rotate the camera around the X, Y, and Z axes. Therefore, rotation also has three degrees of freedom. You can represent it using the Euler angles (turn, step and slope), the 3 x 3 rotation matrix, or the direction of rotation (ie the axis) and angle.

To calculate the 3D position of an object in an image, you need the following information:

– 2D coordinates of multiple points. You need to position multiple points on the image (Fig. 1). In the case of the face, you can choose the corners of the eyes, the tip of the nose, the corners of the mouth, etc.

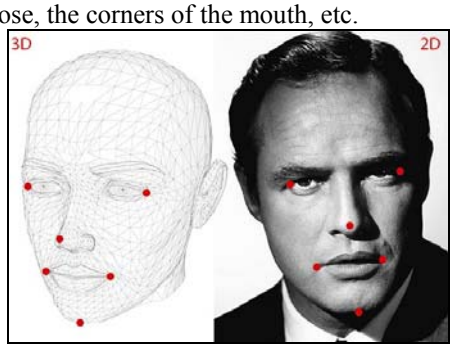

Figure 1 – Matching dots on 2D and 3D

– 3D locations of the same points. You need three-dimensional multi-point placements in an arbitrary frame. Since it is not possible to obtain an accurate three-dimensional model of any head from a single image, a generalized human head model is used.

Indicators that determine the position of the head when using the second way of expressing the position is the angle at which the head deviates from the frontal position in the following three directions: pitch, roll, yaw (Fig. 2). While the first method is suitable for determining the position of any object relative to the camera, the second method can only be used to determine the position of the head. Usually the second method is used in methods that are based on the creation of a CNN model.

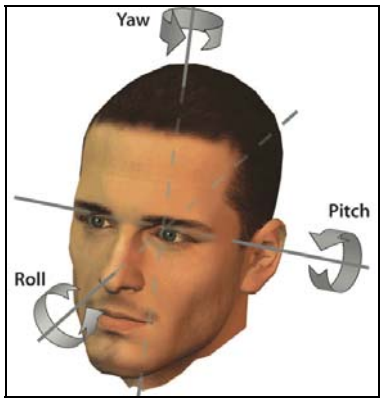

Figure 2 – Head rotation directions

One of the well-known technologies that play a crucial role in determining human postures and head postures is the Kinect camera developed by Microsoft. The Kinect camera has a clear advantage over other 3D cameras, because it gets more accurate depth information and works fast enough. Using Kinect, you can track up to six people at a time, as well as get motion analysis with an extraction function that allows you to determine the locations of the articulation points of the human skeleton. Extremely useful indoors, it cannot be used outdoors because the infrared depth sensor is extremely sensitive to sunlight. Methods based on the use of information obtained with the help of depth sensors allow to achieve accuracy in which the deviation from the real data is no more than two degrees [1].

The limitations on using Kinect, force us to look for ways to recognize a person's head based on a two-dimensional image of an RGB camera. Historically, there have been several basic approaches to face modeling: discriminatory, oriented approaches, and parameterized appearance models, or PAM. Regardless of the approach chosen, each method uses a face detector to obtain a region of the face image that will be further analyzed. There are several approaches to the problem of face recognition, some based on the implementation of the Viola-Jones algorithm [2], the wavelet transform [3], or the principal component method.

Considering the conducted researches and current restrictions **the purpose of this work** is creation of the method that will allow us to track human head moves and record only significant vectors of head direction to represent the entire move with minimal amount of data. **The object** of study is the vector of facial orientation, and **the subject** is the method for determining the reference points of head moves.

1 PROBLEM STATEMENT

Based on the study of existing systems and methods that perform the definition of the face vector on the images, we can describe a problem statement for the head tracking and comparison system.

As an input we have a sequence of frames with human head poses. In order to create a head tracking system, it is necessary to determine human head direction vector $\overline{a_i}, i = \overline{1,n}$ on each frame and distinguish those vectors $\overline{a'_j}, j = \overline{1,m}$ where head changes direction of the move. The output will be a sequence of reference points $\overline{a'_1},\ldots,\overline{a'_m}$, $m < n$ sufficient to describe the head moves.

The main criterion by which existing method should be considered is the accuracy of vector calculation and the stability of operation in different lighting conditions and video quality. The desired angle deviation shall be close to 2° for each rotation direction. The processing speed of one frame should be approximately 40 ms, in order to process high speed moves.

DOI 10.15588/1607-3274-2020-3-11

2 REVIEW OF THE LITERATURE

In recent years, methods that directly take 2D (face landmark) face points using deep learning tools appeared [4]. They have become dominant approaches to the analysis of face rotation due to their flexibility and resistance to extreme posture changes.

Recently, researches that use deep neural networks became leaders in the accuracy of the evaluation of head postures. J. Park and S. Kwon used deep neural networks such as Lenet to evaluate continuous head posture [5]. Massimiliano and Angelo explored the role of adaptive gradient methods for improving CNN performance in the evaluation of head postures [6]. The above works build their core unit using less than five collapsible layers and extract more expressive features from the training datasets. Kumar et al. modified the GoogleNet architecture to jointly predict facial landmarks and head postures. Xu et al. adapted the global and local CNN facial features for coarse-to-fine head posture evaluation. They used the global networks to predict the original head posture and the local networks to update the postures according to the current form. B. Huang et al. created a method of estimating head postures using two-stage groups with averaged top-k regression [7].

3 MATERIALS AND METHODS

Face landmark based methods that detect relative to camera head position are quite widespread and are commonly used with face-landmark detectors [8, 9]. The search of the pose is performed by determining the distortion applied to the 3D model. To achieve this, a landmark detector needs to find dozens of dots on the face, such as mouth corners, eye corners, jaw silhouettes, and more. Many algorithms have been developed and implemented in OpenCV. Identifying a person's landmarks begins with identifying the persons in the image and their extents (bounding box). The fastest and easiest way to detect faces in OpenCV is still to use the associated cascade classifiers using the cv :: CascadeClassifier class provided in the core module. The tag detector will work around the detected individuals, starting from the bounding box.

Once we get the landmarks on the face, we can try to determine the direction of the face. 2D face orientations essentially correspond to the shape of the head. Therefore, in view of the three-dimensional model of the human head, we can find approximate corresponding three-dimensional points for a number of faces next way:

$$\begin{pmatrix} x \\ y \\ 1 \end{pmatrix} = s \begin{pmatrix} f_x & 0 & c_x \\ 0 & f_y & c_y \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} r_1 & r_2 & r_3 & t_1 \\ r_4 & r_5 & r_6 & t_2 \\ r_7 & r_8 & r_9 & t_3 \end{pmatrix} \begin{pmatrix} U \\ V \\ W \\ 1 \end{pmatrix}. \quad (1)$$

In the above formula U, V, W are the position of the object in three-dimensional space, and x, y are the coordinates of the points in the image, f_x and f_y focal lengths in the x and y directions, c_x and c_y are focal center, $r_i, i = \overline{1,9}$ are elements of the rotation vector, $t_i, i = \overline{1,3}$ are elements of camera translation vector. OpenCV provides implementation for finding rotation and moving with its cib :: solvePnP functions calib3d module.

Since the general model is commonly used, this can lead to position recognition errors in case of significant differences between the test subjects' head and the model. Another problem in the application of this method depends on the accuracy of determining the anchor points of the individual. If a significant part of the points is absent, it is impossible to calculate the coordinates of the head.

The newest and more advanced methods of determining the posture of the head are methods based on the training of neural networks. Face detectors are trained on images of a person with different discrete poses and combine the outputs of a number of classifiers. These existing methods of estimating head postures are usually taught based on the classification of bin-poses or regression in one pose. Classification-based methods are performed by comparing the image with the labels of discrete poses, while regression-based methods directly output values of continuous posture. However, these methods are unlikely to use classification and regression loss at the same time. Classification-based methods are more robust to changes in non-ideal conditions. Their labels are true, accepting angular intervals (usually greater than 10°), so the corresponding labels are a bottleneck for further improving accuracy. Regression-based methods may more accurately predict head posture, but their effectiveness depends on the initial head position and variations in head posture, and perform poorly in non-ideal conditions.

Nonlinear regression methods use a training set to create nonlinear mapping from images to poses, and CNN is part of these methods. Because CNNs have the ability to reduce size and extract features automatically, they have achieved good results in various areas. The use of CNN has greatly improved the accuracy of head estimation, but excellent performance is only demonstrated in the same type of images and conditions that are found in the training set due to the intense overload of the training set.

FSA-Net [10] is representative of the regression method and uses a smooth stepped regression scheme. Existing methods of function aggregation process the input data as a set of features and thus ignore their spatial relation on the feature map. The method offers a fine-grained mapping of the structure for spatial grouping of features before aggregation.

FSA-Net performs spatial grouping of features before submitting them to the aggregation process. The developers of this method claim that due to this method they achieve accuracy that exceeds the capabilities of systems based on RGBD image processing, that is, images from cameras with depth sensors.

ResNet50 [11] combines a classification and regression approach to create a CNN that determines the

DOI 10.15588/1607-3274-2020-3-11

position of the head without first finding anchor points. The ResNet50 uses three separate losses, one for each corner. Each loss is a combination of two components: a classified posture and a regression component. Any network can be used and complemented by three fully connected layers that predict angles.

The idea behind this approach is that, using the classification, a softmax layer and cross-entropy are used, thus the network learns to clearly predict the adjacent posture. With three cross entropy losses, one for each Euler angle, we have three signals that are transmitted to the network to enhance learning. In order to get clearer forecasts, the expectation of each output angle for the resulting output class is calculated.

Then, regression losses are added to the network, namely the mean square error loss to improve the forecasts. There are three terminal losses, one at each angle, and each is a linear combination of both the corresponding classification and the regression losses.

In order to make progress in the prediction of image intensities, we need to find real datasets that contain accurate posture annotations, multiple images of different people, different lighting conditions, and a significant variety of poses. We identify two very different datasets that meet these requirements.

The BIWI dataset [12] is collected in the laboratory by recording RGB-D videos of different people with different head postures using a Kinect v2 device. It contains approximately 15,000 frames and rotation angles of ± 75° for yaw, ± 60° for pitch and ± 50° for roll. This dataset is commonly used as a benchmark for estimating postures using depth methods, which confirms the accuracy of their labels. Examples of images from the BIWI dataset are shown in Fig. 3.

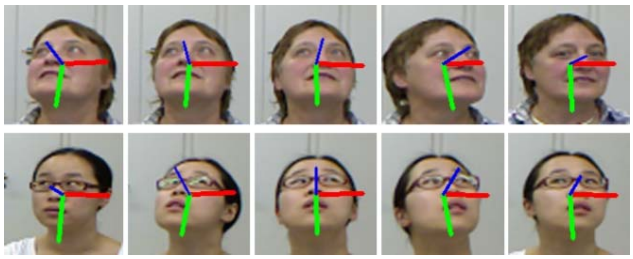

Figure 3 – An example from a BIWI dataset

Another 300W-LP data set [13] is a 300W extension that standardizes multiple alignment databases with 68 landmarks, including AFW, LFPW, HELEN, IBUG, and XM2VTS. This dataset offers synthetically advanced data to prepare landmark models. This is a collection of popular two-dimensional datasets that have been grouped and re-annotated. Synthetics of the set make it possible to create images with different values of deviation in all three directions and at the same time to receive the most accurate marks, which significantly affects the quality of the trained CNN models.

We propose to use head direction estimation methods to track head pose in real time. Considering the possible methods of comparison of head positions, two methods were identified: frame-by-frame comparison, comparison of reference points. The first method is much simpler in terms of implementation, but it has significant drawbacks. First, the accuracy of this method depends on the speed of the face vector search method, if this speed is more than 40 ms (interval between frames when recording FPS 24 video), then some frames will be skipped, which will lead to a false negative scenario. Also, the person performing the exercises can change the pace of movement. Therefore, the fact of performing the exercise should not be associated with time.

We propose the reference point method that simplifies the process of recording the reference exercise. With this method we do not have to record every position of the head, but only those positions in which the movement changes direction. This method tracks the vector by which the face moves in time. Since the vector of the head can be represented as $\bar{a} = (r, p, y)$, where r is the angle of the head along the z axis, p is the angle of the head along the x axis, y is the angle of the head along the y axis, the reference point is a head direction vector in which the way it flows changes its direction. The reference point can be found by the following formula:

$$k < \left|\bar{a}_i\right| - \left|\bar{a}_{i+1}\right| \le \left|\bar{a}_i\right| - \left|\bar{a}_{i-1}\right|, \quad (2)$$

where a_i is the value of the angle on the i frame and k is a flexibility degree. If the inequality holds for at least one of the metrics, the position is stored as a reference point. Flexibility degree is needed to avoid recording reference points caused by variations in vector estimation.

4 EXPERIMENTS

The methods discussed earlier were investigated by the following criterias:

– accuracy of the method;
– speed of work;
– maximal recognition angle.

The accuracy of the method of determining the position of the head is usually calculated in MAE (Mean absolute error). In statistics, the mean absolute error (MAE) is a measure of the difference between two continuous variables. Consider the scatter plot of n points where point i has coordinates (x_i, y_i). Mean absolute error (MAE) is the average vertical distance between each point and an identical line. MAE is also the average horizontal distance between each point and an identical line. MAE is given by the following formula:

$$MAE = \frac{\sum_{i=1}^{n}\left|y_i - x_i\right|}{n} = \frac{\sum_{i=1}^{n}\left|e_i\right|}{n}. \quad (3)$$

Mean absolute error is the mean value of absolute errors $\left|e_i\right| = \left|y_i - x_i\right|$, where y_i is the predicted value and x_i is the real value.

DOI 10.15588/1607-3274-2020-3-11

The experiment was performed using two different datasets: BIWI, AFLW2000.

Speed was tested on frames obtained from a conventional camera with FPS 24. The calculations were performed using an Intel core i7 processor, RAM 16Gb.

The maximum recognition angle was investigated using a BIWI data set containing approximately 15,000 frames, and rotation angles ± 75° for yaw, ± 60° for pitch and ± 50° for roll.

After the desired method was found we performed the experiment on flexibility degree to determine the most appropriate value that will help us to reduce amount of false reference points but not skip relevant points. Considered flexibility degrees were in interval between 1 and 2 degrees with step of 0.25.

5 RESULTS

Experiments mentioned above exposed next results. Comparisons of the main yaw, pitch, roll, and mean error values when using the BIWI data set are presented in figure 4.

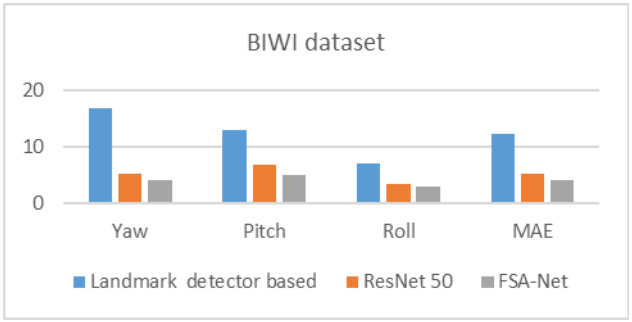

Figure 4 – The results experiment using BIWI dataset

The same experiment was performed with the AFLW2000 data set (Fig. 5). The main difference between these datasets is that BIWI is a laboratory-made data set, so the images have the same illumination and good quality. AFLW2000 [14] is a real-time data set where the brightness, blurring and image quality are very different.

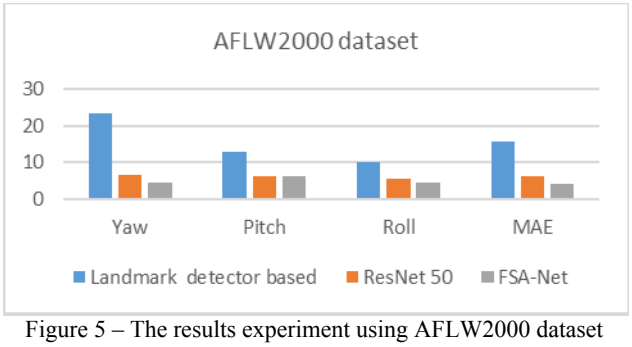

Figure 5 – The results experiment using AFLW2000 dataset

Speed test was performed under the same circumstances using same pictures for each method. The rate of calculation of postures for one frame and one face in the frame is shown in Table 1.

Table 1 – Speed test

№	Method	Average execution speed
1	Landmark-based detector	75 ms
2	ResNet 50	105 ms
3	FSA-Net	133 ms

The results of the accuracy of angle recognition depending on its value are shown in Fig. 6.

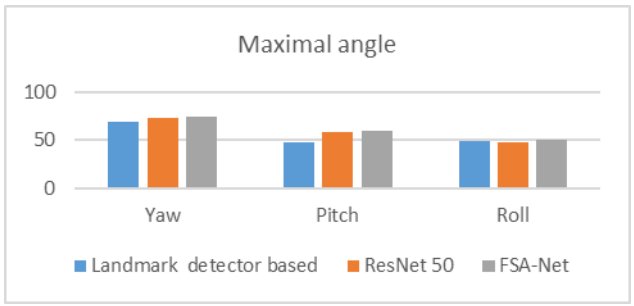

Figure 6 – Maximal angles of recognition

This coefficient was tested by next indicators: percent of correctly detected points, extra points.

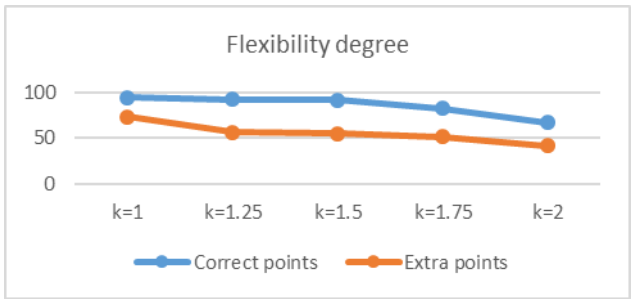

Figure 7 – Flexibility degree test results

6 DISCUSSIONS

The data obtained during the experiment shows that methods that are not based on the search for key points show quite good results, even on the AFLW2000 dataset. The average error for the landmark-based method is 15.8, which means an error of 15.8°. While ResNet 50 and FSA-Net allow for accuracy of 6.4° and 4.2° respectively. This leads us to believe that the potential of CNN-based methods is higher than key point methods.

Based on the results we have obtained during the speed test, the most accurate FSA-Net method shows the worst result of 133 ms, which is 3 times higher than the desired processing speed of 40 ms. If we use this method, we will be able to compare every fourth frame in the video. Such a low speed is due to the large number of image processing steps and is highly dependent on the speed of the face detector, which no method can do without.

All methods almost equally well recognize all available angles of the dataset. However, the landmark-based method showed the least accurate recognition of the difference between the boundary angles, which is due to the hidden part of the reference points when turning the head. For landmark-based methods, the lack of a large part of the points due to overlapping of the face with other

DOI 10.15588/1607-3274-2020-3-11

parts of the body or the headpiece makes it impossible to determine the angle.

According to the research, algorithms that are not based on facial landmark have shown much better accuracy results, especially on a real-time and poor lighting conditions. The 2° difference for the classification regression models is a very important step in improving the accuracy of face vector recognition. FSA-Net, which delivers 98.8% accuracy (based on an error of 4.2° from 360°), is the best candidate today for use in human head monitoring systems.

A method proposed to determine reference points shown that with flexibility degree close to 1°, 95% of all reference points are detected but too many extra points are recorded because of variations in head vector estimation. With flexibility degree equal to 1.25 or 1.5 accuracy stays the same while amount of extra points decreases. These values must be considered as most appropriate.

CONCLUSIONS

The practical value of the study is in creation of a reference points method. The proposed approach of search allows us to eliminate intermediate angles and store about 95% of all points where head move direction changes. This gives us ability to express head move with minimal amount of significant vectors and drastically reduces space needed do store the move representation. Further these reference points can be used to track human head moves by being compared to vectors determined in runtime. However further optimization is required by reference points search algorithm to decrease amount of extra points.

According to the results of the study, regression-based methods showed significantly better accuracy and independence of accuracy from light and partial face closure. All methods, and in particular the method based on facial landmark showed deterioration in the recognition of boundary angles due to the slight difference in the characteristics of the face. This is a window for the possible improvement of the method of operation in the conditions of tracking the person turned to the camera at an angle greater than 70°.

Significant optimization requires the speed of the FSA-Net method because at it works 2 times more slowly than landmark-based. Because the process of determining the posture of the head is multi-step, optimization can be performed both in the step of finding faces in the image, using faster detectors, and in the step of finding poses.

Research shows that methods of finding a person's head posture can compete with accuracy based on the use of depth sensors. However, methods of working with a simple RGB image have far fewer restrictions on the place of use.

ACKNOWLEDGEMENTS

The work is carried out in the framework of research directions of Software Engineering Department and with the support of researchers from the Scientific Research Laboratory "Information Technologies in Learning and Computer Vision Systems" of Kharkiv National University of Radio Electronics.

REFERENCES

1. Borghi G., Fabbri M., Vezzani R. et al. ace-from-Depth for Head Pose Estimation on Depth Images, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2018, Vol. 42, pp. 596–609. DOI: 10.1109/TPAMI.2018.2885472
2. Viola P., Jones M. Rapid object detection using a boosted cascade of simple features, *Computer Society Conference on Computer Vision and Pattern Recognition, 8–14 December 2001: proceedings*. Kauai, IEEE, 2001. DOI: 10.1109/CVPR.2001.990517
3. Shcherbakova G., Krylov V. N., Bilous N. V. Methods of automated classification based on wavelet-transform for automated medical diagnostics, *Information Technologies in Innovation Business Conference (ITIB), 7–9 October 2015: proceedings*. Kharkiv, IEEE, 2015, pp. 7–10. DOI: 10.1109/ITIB.2015.7355048
4. Wallhoff F., AblaBmeier M., Rigoll G. Multimodal Face Detection, Head Orientation and Eye Gaze Tracking, *International Conference on Multisensor Fusion and Integration for Intelligent Systems, 3–6 September 2006: proceedings*. Heidelberg, IEEE, 2006, pp. 13–18. DOI: 10.1109/MFI.2006.265612
5. Ahn B., Park J., Kweon I. S. Real-time head orientation from a monocular camera using deep neural network, *Asian Conference on Computer Vision: 12th Asian Conference on Computer Vision, 1–5 November 2014: proceedings*. Singapore, ACCV, 2014, pp. 82–96. DOI: 10.1007/978-3-319-16811-1_6
6. Patacchiola M., Cangelosi A. Head pose estimation in the wild using convolutional neural networks and adaptive gradient methods, *Pattern Recognition,* 2017, Vol. 71, pp. 132–143. DOI: 10.1016/j.patcog.2017.06.009
7. Huang B., Chen R., Xu W. et al Improving head pose estimation using two-stage ensembles with top-k regression, *Image and Vision Computing*, 2020, Vol. 93. DOI: 10.1016/j.imavis.2019.11.005
8. Kumar A., Alavi A., Chellappa R. KEPLER: Keypoint and pose estimation of unconstrained faces by learning efficient H-CNN regressors, *Image and Vision Computing*, 2018, pp. 258–265. DOI: 10.1016/j.imavis.2018.09.009
9. Hien L. T., Toan D. N., Lang T. V. Detection of Human Head Direction Based on Facial Normal Algorithm, *International Journal of Electronics Communication and Computer Engineering*, 2015, Vol. 6, pp. 110–114
10. Tsun-Yi Y., Yi-Ting C., Yen-Yu L. et al. FSA-Net: Learning Fine-Grained Structure Aggregation for Head Pose Estimation from a Single Image, *Conference on Computer Vision and Pattern Recognition (CVPR), 15–20 June 2019: proceedings*. Long Beach, IEEE, 2019, pp. 1087–1096. DOI: 10.1109/CVPR.2019.00118
11. Ruiz N., Chong E., Rehg J. M. Fine-grained head pose estimation without keypoints, *Conference on Computer Vision and Pattern Recognition Workshop, 18–22 June 2018: proceedings.* Salt Lake City, IEEE, 2018, pp. 1821–1829. DOI: 10.1109/CVPRW.2018.00281
12. Fanelli G., Dantone M., Gall J. et al. Random forests for real time 3D face analysis, *International Journal of Computer Vision,* 2013, Vol. 101, pp. 437–458. DOI: 10.1007/s11263-012-0549-0
13. Zhu X., Lei Z., Liu X. et al. Face alignment across large poses: A 3D solution, *Conference on Computer Vision and*

DOI 10.15588/1607-3274-2020-3-11

Pattern Recognition, 27–30 June 2016: proceedings. Las Vegas, 2016. – P. 146–155. DOI: 10.1109/CVPR.2016.23

14. Koestinger M., Wohlhart P., Roth P. M. et al. Annotated facial landmarks in the wild: A large-scale, real-world database for facial landmark localization, *International Conference on Computer Vision Workshops, 6–13 November 2011: proceedings*. Barcelona, IEEE, 2011, pp. 2144–2151. DOI: 10.1109/ICCVW.2011.6130513

Received 15.05.2020.
Accepted 22.09.2020.

УДК 004.93

МЕТОД ОПОРНИХ ТОЧОК ДЛЯ ВІДСТЕЖЕННЯ РУХІВ ГОЛОВИ ЛЮДИНИ

Ракова А. О. – студент кафедри програмної інженерії, Харківський національний університет радіоелектроніки, Харків, Україна.

Білоус Н. В. – канд. техн. наук, доцент, професор кафедри програмної інженерії, Харківський національний університет радіоелектроніки, Харків, Україна.

АНОТАЦІЯ

Актуальність. Напрямок вектору обличчя людини є показником уваги людини. У нашому повсякденному житті він має багато застосувань, такі як взаємодія людина-комп'ютер, телеконференції, віртуальна реальність та 3D-передача звуку. Більше того, визначення положення голови можна використати для порівняння вправ, які виконує людина, з певним стандартом, що приводить нас до дослідження способів ефективного відстеження рухів. Системи на основі глибинних камер, які часто використовуються для цих цілей, мають суттєві недоліки, такі як зниження точності від прямого сонячного світла та необхідність додаткового обладнання. Розпізнавання від двовимірного зображення набуває все більшого поширення та усуває труднощі, пов'язані з глибинними камерами, що дозволяє використовувати їх у приміщенні та на відкритому повітрі.

Мета. Метою даної роботи є створення методу, який дозволить нам відстежувати рухи голови людини і записувати лише значні вектори напрямку голови.

Методи. У цій роботі пропонується метод опорних точок, який зменшує набір записаних векторів до мінімальної кількості, значущої для опису рухів голови. Він також досліджує та порівнює існуючі методи визначення вектору обличчя з точки зору використання у запропонованому підході.

Результати. Запропонований метод опорних точок показує здатність сильно зменшувати набір векторів напрямку голови, які описують рух. Відповідно до результатів дослідження, методи, засновані на регресії, показали значно кращу точність та незалежність від світла та часткового закриття обличчя, тому їх було обрано для використання в якості методів отримання вектору напрямку голови в підході опорних точок.

Висновки. Дослідження підтвердили застосовність методу опорних точок для відстеження рухів людини і показали, що методи визначення вектору голови людини за двовимірним зображенням можуть конкурувати в точності з методами на основі RGBD. Таким чином, у поєднанні із запропонованим підходом ці методи мають менше обмежень у використанні, ніж такі, що базуються на RGBD.

КЛЮЧОВІ СЛОВА: вектор направленості обличчя, рухи голови, розпізнавання, глибоке навчання.

УДК 004.93

МЕТОД ОПОРНЫХ ОТЧЕК ДЛЯ ОТСЛЕЖИВАНИЯ ДВИЖЕНИЙ ГОЛОВЫ ЧЕЛОВЕКА

Ракова А. О. – студент кафедры программной инженерии, Харьковский национальный университет радиоэлектроники, Харьков, Украина.

Белоус Н. В. – канд. техн. наук, доцент, профессор кафедры программной инженерии, Харьковский национальный университет радиоэлектроники, Харьков, Украина.

АННОТАЦИЯ

Актуальность. Направление вектора человеческого лица является индикатором человеческого внимания. Оно имеет множество вариантов применения в нашей повседневной жизни, таких как взаимодействие человека с компьютером, телеконференций, виртуальной реальности и 3D-рендеринга звука. Более того, определение положения головы можно использовать для сравнения упражнений, выполняемых человеком с определенным стандартом, что приводит нас к исследованию способов эффективного отслеживания движений. Системы на основе глубинных камер, часто используемые для этих целей, имеют существенные недостатки, такие как снижение точности на прямом солнечном свете и необходимость дополнительного оборудования. Распознавание по двумерному изображению становится все более распространенным и устраняет трудности, связанные с глубинными камерами, что позволяет использовать их в помещении и на улице.

Цель. Целью данной работы является создание метода, который позволит нам отслеживать движения головы человека и фиксировать только значимые векторы направления головы.

Методы. Эта статья предлагает метод опорных точек, который уменьшает набор записанных векторов до минимального значения, значимого для описания движений головы. Он также исследует и сравнивает существующие методы определения вектора лица с точки зрения использования в предлагаемом подходе.

Результаты. Предложенный метод опорных точек показывает способность значительно уменьшить набор векторов направления головы, которые описывают движение. Согласно результатам исследования, методы на основе регрессии

DOI 10.15588/1607-3274-2020-3-11

показали значительно лучшую точность и независимость от легкого и частичного закрытия лица, поэтому они были выбраны для использования в качестве методов для получения вектора направления головы в методе опорных точек.

Выводы. Исследования подтвердили применимость метода опорных точек для отслеживания движений человека и показали, что методы определения вектора головы человека по двумерному изображению могут конкурировать в точности с методами на основе RGBD. Таким образом, в сочетании с предлагаемым подходом, эти методы создают меньше ограничений в использовании, чем основанные на RGBD.

КЛЮЧЕВЫЕ СЛОВА: вектор направленности лица, движение головы, распознавание, глубокое обучение.

ЛІТЕРАТУРА / ЛИТЕРАТУРА

1. Face-from-Depth for Head Pose Estimation on Depth Images / [G. Borghi, M. Fabbri, R. Vezzani et al.] // IEEE Transactions on Pattern Analysis and Machine Intelligence. – 2018. – Vol. 42. – P. 596–609. DOI: 10.1109/TPAMI.2018.2885472
2. Viola P. Rapid object detection using a boosted cascade of simple features / P. Viola, M. Jones // Computer Society Conference on Computer Vision and Pattern Recognition, 8–14 December 2001: proceedings. – Kauai : IEEE, 2001. DOI: 10.1109/CVPR.2001.990517
3. Shcherbakova G. Methods of automated classification based on wavelet-transform for automated medical diagnostics / G. Y. Shcherbakova, V. N. Krylov, N. V. Bilous // Information Technologies in Innovation Business Conference (ITIB), 7–9 October 2015: proceedings. – Kharkiv: IEEE, 2015. – P. 7–10. DOI: 10.1109/ITIB.2015.7355048
4. Wallhoff F. Multimodal Face Detection, Head Orientation and Eye Gaze Tracking / F. Wallhoff, M. AblaBmeier, G. Rigoll // International Conference on Multisensor Fusion and Integration for Intelligent Systems, 3–6 September 2006: proceedings. – Heidelberg: IEEE, 2006. – P. 13–18. DOI: 10.1109/MFI.2006.265612
5. Ahn B. Real-time head orientation from a monocular camera using deep neural network / B. Ahn, J. Park, I. S. Kweon // Asian Conference on Computer Vision: 12th Asian Conference on Computer Vision, 1–5 November 2014: proceedings. – Singapore: ACCV, 2014. – P. 82–96. DOI: 10.1007/978-3-319-16811-1_6
6. Patacchiola M. Head pose estimation in the wild using convolutional neural networks and adaptive gradient methods / M. Patacchiola, A. Cangelosi // Pattern Recognition. – 2017. – Vol. 71. – P. 132–143. DOI: 10.1016/j.patcog.2017.06.009
7. Improving head pose estimation using two-stage ensembles with top-k regression / [B. Huang, R. Chen, W. Xu et al.] // Image and Vision Computing. – 2020. – Vol. 93. DOI: 10.1016/j.imavis.2019.11.005
8. Kumar A. KEPLER: Keypoint and pose estimation of unconstrained faces by learning efficient H-CNN regressors / A. Kumar, A. Alavi, R. Chellappa // Image and Vision Computing. – 2018. – P. 258–265. DOI: 10.1016/j.imavis.2018.09.009
9. Hien L.T. Detection of Human Head Direction Based on Facial Normal Algorithm / L. T. Hien, D. N. Toan, T. V. Lang // International Journal of Electronics Communication and Computer Engineering. – 2015. – Vol. 6. – P. 110–114
10. FSA-Net: Learning Fine-Grained Structure Aggregation for Head Pose Estimation from a Single Image / [Y. Tsun-Yi, C. Yi-Ting, L. Yen-Yu et al.] // Conference on Computer Vision and Pattern Recognition (CVPR), 15–20 June 2019: proceedings. – Long Beach: IEEE, 2019. – P. 1087–1096. DOI: 10.1109/CVPR.2019.00118
11. Ruiz N. Fine-grained head pose estimation without keypoints / N. Ruiz, E. Chong, J. M. Rehg // Conference on Computer Vision and Pattern Recognition Workshop, 18–22 June 2018: proceedings. – Salt Lake City : IEEE, 2018. – P. 1821–1829. DOI: 10.1109/CVPRW.2018.00281
12. Random forests for real time 3D face analysis / [G. Fanelli, M. Dantone, J. Gall et al.] // International Journal of Computer Vision. – 2013. – Vol. 101, P. 437–458. DOI: 10.1007/s11263-012-0549-0
13. Face alignment across large poses: A 3D solution / [X. Zhu, Z. Lei, X. Liu et al.] // Conference on Computer Vision and Pattern Recognition, 27–30 June 2016: proceedings. – Las Vegas, 2016. – P. 146–155. DOI: 10.1109/CVPR.2016.23
14. Annotated facial landmarks in the wild: A large-scale, real-world database for facial landmark localization / [M. Koestinger, P. Wohlhart, P. M. Roth et al.] // International Conference on Computer Vision Workshops, 6–13 November 2011: proceedings. – Barcelona : IEEE, 2011. – P. 2144–2151. DOI: 10.1109/ICCVW.2011.6130513

DOI 10.15588/1607-3274-2020-3-11