

The System for Determining the Human Head Position and Orientation for Vehicle Simulators

Alexei Zakharov, Alexei Barinov and Arkady Zhiznyakov

Murom Institute, Vladimir State University, Murom, Russia
aa-zaharov@ya.ru, alexey.barinov.murom@ya.ru, lvovich@newmail.ru

Abstract. The system for determining the human head position and orientation for vehicle simulators is considered in the work. It is necessary to take into account the current position and orientation of the driver's head when simulating a virtual space. The model for determining the human head position and orientation is developed and studied. In the course of the study, the developed position and orientation algorithm was compared with analogs belonging. This model is characterized by the joint use of three-dimensional reconstruction, stereo vision, the spectral theory of graphs, spectral embedding of graphs into a vector subspace, and allowing the head to rotate to 50° with an accuracy of 3° , which exceeds known approaches. A system for determining the human head position and orientation based on stereo images is constructed, which implements the developed algorithms. The developed system can be represented in the form of the following structure: stereo module, initialization module, tracking module, module for calculating the angles of the head orientation, head detection module, information transfer module.

Keywords: Human Head Pose, Vehicle Simulators, Image Recognition.

1 Introduction

To date, an integral part of the high quality training of drivers is the use of simulators, which can be of great help in the training process of driving on the basis of a set of certain exercises.

An integral part of modern simulators is the system of surrounding space visualization. Throughout the development cycle of simulators, much attention has been paid to the development and improvement of this system. But even now, the presence effect created in the simulator is very different from the driver's feelings he experiences in a real vehicle. One of the reasons for this is that in the synthesis of the environment in the transport simulator, the visualization system does not change the displayed picture of the world, depending on the position and orientation of the driver's head in

the cabin (Fig. 1). In turn, this leads to the impossibility of realistic emulation of rear-view mirrors and the entire area around the car to the extent that would enable a person to drive a real vehicle well. To solve this problem, it is necessary to take into account the current position and orientation of the driver's head when simulating a virtual space.

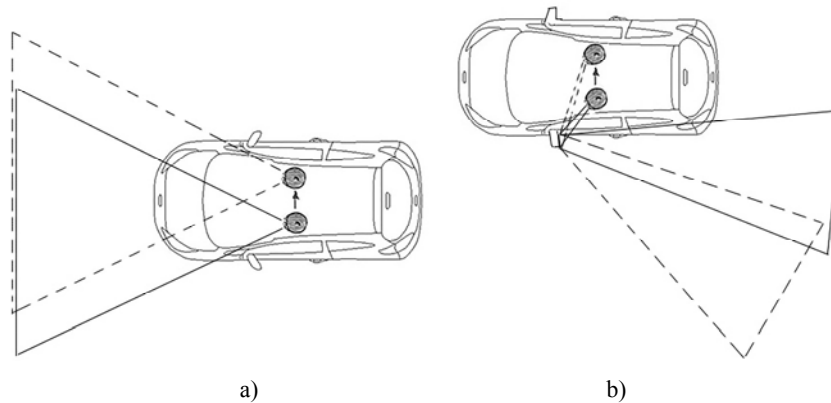


Fig. 1. Change in the overview when the position and orientation of the drivers' head change: a) change in the overview through the vehicle's glass; b) change in the overview in the rearview mirror.

Recently, a large number of methods and systems for determining the position and orientation of the head based on images have been developed. However, the existing systems have low accuracy of determining parameters, do not provide the necessary speed and accuracy. In addition, the use of such systems is limited to specific applications, so they are difficult to use in common vehicle simulators.

The creation of a system for determining the human head position and orientation on the basis of stereo images will increase the effect of the trainee's presence in the synthesized medium by changing the overview. In addition to increasing the realism of the external space display, the trainee will have the opportunity to master such an important skill as controlling the traffic situation through an overview of rear-view mirrors. This will definitely improve the quality of drivers' training process and subsequently reduce the number of accidents

2 The Model for Determining the Human Head Position And Orientation Using Stereo Images

To determine the head position and orientation, a stereo system consisting of two chambers is used. The optical axis of the chambers are parallel to each other. The three-dimensional coordinates of the head point x, y, z are found according to the known distance between the cameras and the camera focal length. Stereo settings are determined during calibration. A preliminarily three-dimensional reconstruction of the

head model based on stereo images using the Sum of Absolute Differences (SAD) algorithm is performed (Scharstein D. et al., 2002). Based on the three-dimensional model, the head current position and orientation are calculated. To track the head orientation, work is done with special points, highlighted in the image using the SURF algorithm. Angles of yaw α_{yaw} , roll α_{roll} and pitch α_{pitch} are calculated with respect to the coordinate system of the stereo settings. The definition of these angles is reduced to the calculation of the angle between the vectors formed by the corresponding key points in the initial and subsequent positions. Such points are: extreme lateral points (A, C); extreme upper and lower points (B, D); the central point (E) (Fig. 2).

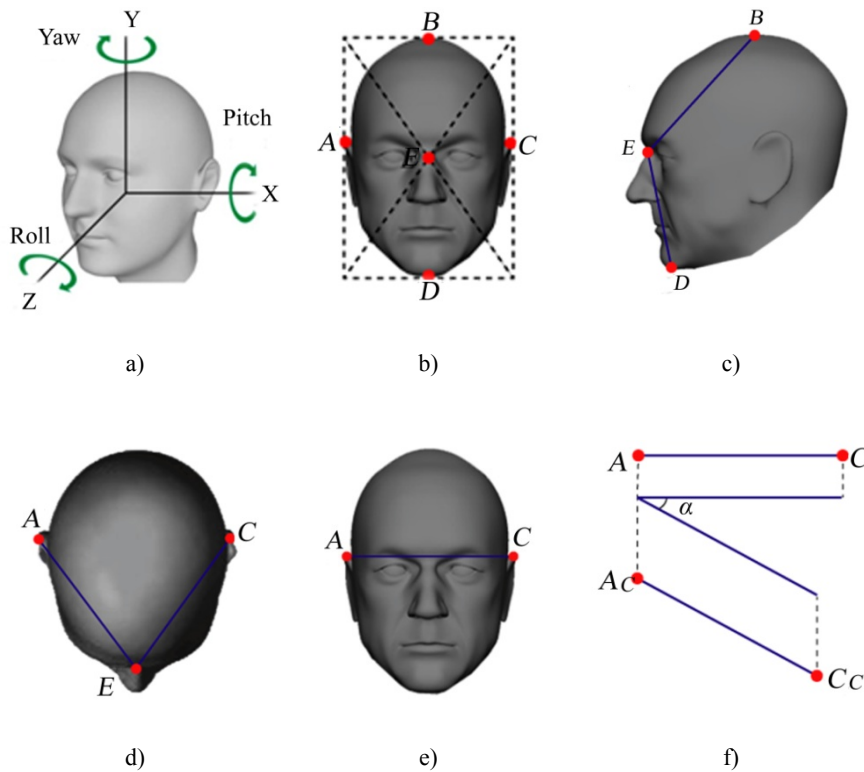


Fig. 2. Determination of the head orientation with the help of singular points. a) angles of rotation of the head; b) key points; c) pitch calculation; d) yaw calculation; e), f) roll calculation.

Thus, for example, to calculate the head tilt, the following formula is used

$$\alpha_{roll} = \arccos \frac{\overline{AC} \cdot \overline{A_C C_C}}{|\overline{AC}| |\overline{A_C C_C}|}, \quad (1)$$

where \overline{AC} is a vector, formed by points in the initial position, $\overline{A_C C_C}$ is a vector formed by points in the current position.

The model for determining the human head position and orientation is a model of actions and consists of the following steps: initialization of the head image using the clustering of features based on the spectral theory of graphs and constraints; the head tracking on images by finding matches using graphs; the head detection in the image based on the embedding of the graph of singularities in the vector space; calculating the angles of nod, rotation and tilt.

2.1 Initialization of the Head Image using the Clustering of Features Based on the Spectral Theory of Graphs and Constraints

For the purpose of controlling the clusterization process limiting conditions are used (Barinov A.E. et al., 2016). For this the prior information about the proportions of the head of the man is used. It is proposed to use the network of the head proportions. This idea was borrowed from art for depicting a human head. It was found out that regardless of the sex and the race the proportions of the faces of different people are equal.

The peculiarities of the photos are used to create the network of proportions. The network is applied on the image of the face containing 5 cells horizontally and 7 vertically. For the purpose of using the network of proportions in the course of clusterizations it is necessary to calculate an averaged template. To achieve this purpose 500 images of people were chosen, special points were calculated, a network of proportions was applied manually (Fig. 3, a).

Finally an averaged position of special points was obtained with respect to the network of proportions (Fig. 3, b).

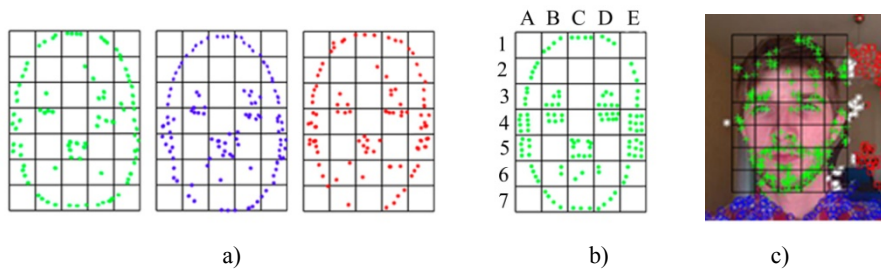


Fig. 3. The network of proportions for identifying points belonging to the head of the man a) The examples of the identification of different faces; b) The averaged template of the network of proportions; c) The results of the clusterization.

2.2 The Head Tracking on Images by Finding Matches Using Graphs

The use of computer vision is limited by sound and optical effects, the textured background, mutual overlapping of objects. To increase the reliability it is proposed to use structured information in the form of graphs. The advantage of structural methods is that they give an opportunity to analyze a big set of elements on the basis of a small quantity of simple components and rules of forming the graphic model. Also structural methods allow to describe the characteristics of the object excluding its reference to another class that increases the reliability of the recognition. Scott and Longuet-Higgins used graphs to find the corresponding. Our method of finding correspondences based on the combined use of graphs and descriptors (Zakharov A. et al., 2015), (Fig. 4).



Fig. 4. The head tracking on images by finding matches using graphs.

2.3 Detection of Human Head on Images Using Graph Embedding In Vector Space

Upon detection of the head on the image, it is proposed to compare spectral characteristics of graphs at the current frame and at the picture obtained at the stage of initialization. It is proposed to use the thermodynamic equation. This approach is widely used when studying temperature effects on gases, with the gradual tracking of changes in the current state. The practical application of thermodynamic equations can be a solution of matching images problem.

Embedding the graph in a vector subspace, we have to deal with differential geometry where to describe relationships between points, we have to use composite curves. The construction of the graph is shown in figure 5.

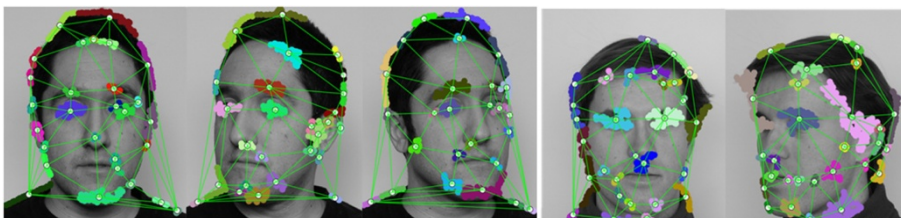


Fig. 5. Construction of Graphs on the Basis of Images.

The results of embedding graphs are shown in the figure 6.

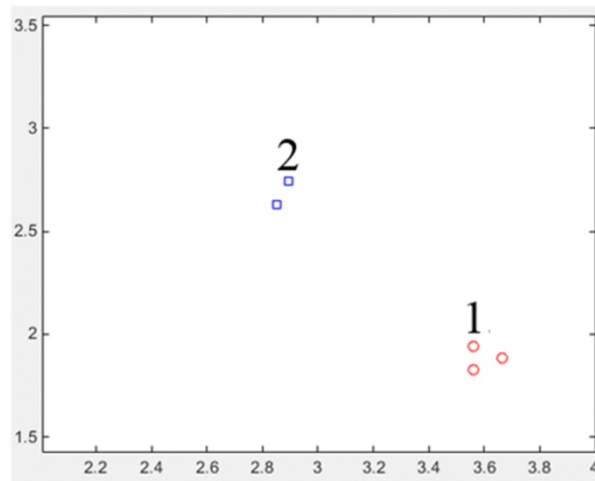


Fig. 6. Results of embedding graphs

3 The Structure of the System for Determining the Head Position and Orientation Using Stereo Images

The developed system is implemented in the Visual Studio Community 2015 programming environment using the C # programming language. When developing the system, both standard functions of the environment and third-party libraries were used: OpenCV library (for working with video images), ALGLIB library (for using advanced mathematical functions), WebCamLib interface (for working with a video camera), Camera calibration functions Camera Calibration Tools.

To save the program settings, the recording technology in the ".ini" files was used.

The developed system can be represented in the form of the following structure (Fig. 7).

Stereo module. It is responsible for the coordinated work of video cameras that are a part of the stereo settings. The module includes a calibration algorithm.

Initialization module. It is necessary to perform the initial search process of the user's head on the image and calculate the descriptor for later detection. It contains an algorithm for spectral clustering of image features with constraints and an algorithm for the structural descriptor calculation.

Tracking module. This module tracks the user's head descriptor between the frames of the video sequence. It contains an algorithm for finding correspondences on images using thermal cores on graphs and calculating the model for determining the human head position and orientation.

Module for calculating the angles of the head orientation. Using the developed model, the current values of orientation angles and three-dimensional coordinates of the head in space are calculated.

Head Detection Module. It searches for an object if it is lost from the stereo view overview area. It contains an algorithm of spectral clustering with constraints and an algorithm for detecting the head based on embedding graphs in a vector space.

Information transfer module. It is an interface between the system for determining the human head position and orientation and the transport simulator.

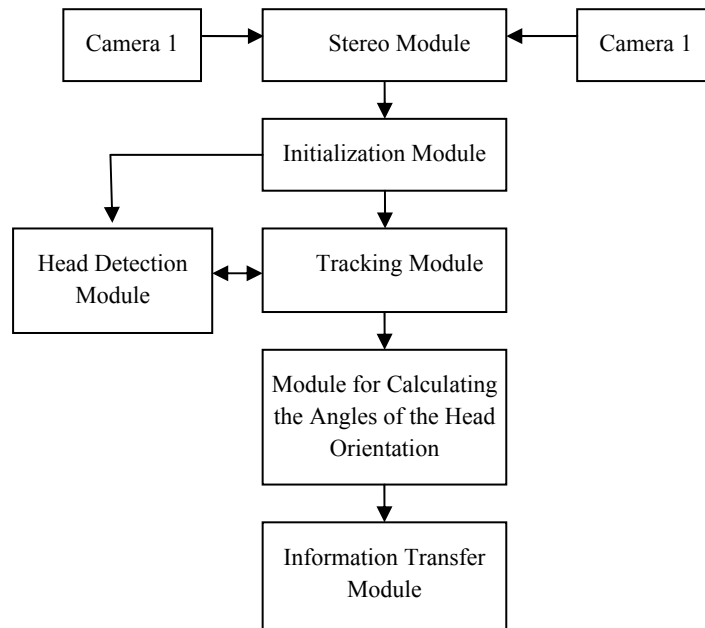


Fig. 7. The structure of the developed system of determining the human body position and orientation with the help of stereo images using graphs

The hardware of the system includes two webcams Logitech C300, with a matrix of 1.3 megapixels. They are mounted on a special bracket, which in turn is mounted on the monitor (Fig. 8). Each camera is connected to the computer using a USB 2.0 cable.

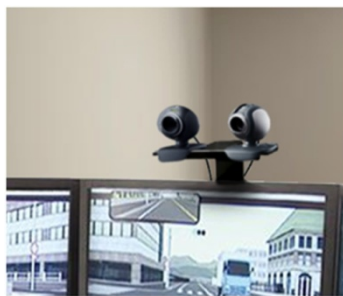


Fig. 8. Fixing the stereo system on the simulator monitor

4 The Investigation of The Model for Determining the Human Head Position And Orientation

In the course of the study, the developed position and orientation algorithm was compared with analogs belonging to well-known groups of methods: appearance template methods (Sharma S., 2013), detector arrays (Jones M. et al., 2003), nonlinear regression (Drouard V. et al., 2015), manifold embedding methods (Sundararajan K. et al., 2015), flexible models (Chen Y. et al., 2014) [64], geometric methods (Hatem H. et al., 2015), hybrid methods [Liao W.K. et al., 2010; Burger P. et al., 2013; Cabrera C.R. et al., 2015].

The comparison was made according to the following parameters: the range of defined head positions; the maximum value of the error; the presence of automatic initialization, which does not require human intervention at the beginning of the system and in case of loss of the monitored object during operation; the processing of situations of mutual overlapping of objects in the scene; the ability to produce the correct result when a person is wearing accessories in the form of glasses or beards; the type of position and orientation determination (discrete or continuous).

It can be seen from the table that majority of the algorithms under consideration have a similar range of detectable angles of rotation, nod and tilt. At the same time, the maximum error belongs to the methods of arrays of detectors and is 9° . The maximum error of the developed algorithm does not exceed 3° . All hybrid methods have automatic initialization and search of the monitored object in case of its loss from the scope of the camera. Also, the methods of nested varieties and the developed algorithm allow further determination of position and orientation in the presence of situations with mutual overlap. Most of the methods considered before work need to be trained, which can cause some difficulties. Also, all hybrid methods and the developed algorithm make it possible to conduct a continuous tracking of the position and orientation, which allows us to obtain the rotation, nod and tilt angles, as well as three-dimensional head coordinates at any time. The latter is an important characteristic when used in transport simulators.

When compared with performance algorithms, it is established that processing is performed at an average rate of 30 frames per second. This allows us to use the developed algorithm in real applications.

5 Conclusion

The model for determining the human head position and orientation is developed and studied. This model is characterized by the joint use of three-dimensional reconstruction, stereo vision, the spectral theory of graphs, thermal nuclei on heat kernels and spectral embedding of graphs into a vector subspace, and allowing the head to rotate to 50° with an accuracy of 3° , which exceeds known approaches.

A system for determining the human head position and orientation based on stereo images is constructed, which implements the developed algorithms. The system has a modular structure and is implemented in Visual Studio Community 2015 using

OpenCV and ALGLIB libraries, WebCamLib interface and Camera Calibration Tools functions. The system allows us to increase the effect of the trainee's presence in the synthesized virtual environment of the transport simulator due to the change in the field of view.

Acknowledgements. This work was supported by the RFBR grant 16-37-00235, project number 2.1950.2017/ПЧ in the framework of the basic tasks of the state of the Russian Ministry of Education.

References

1. Barinov A.E., Zakharov A.A.: Clustering using a random walk on graph for head pose estimation. 2015 international conference on mechanical engineering, automation and control systems (MEACS) (2016).
2. Burger P., Rothbucher M.: Self initializing head pose estimation with a 2D monocular USB camera. Technischer bericht, technische universität münchen (2013).
3. Cabrera C.R., García-Montero M., López-Sastre R., Tuytelaars T.: Fast head pose estimation for human-computer interaction. Iberian conference on pattern recognition and image analysis (2015).
4. Chen Y., Fu M., Yang Y., Song W.: A method of head pose estimation based on active shape model and stereo vision. Control conference (2014).
5. Drouard V., Ba S., Evangelidis G., Deleforge A., Horaud R.: Head pose estimation via probabilistic high-dimensional regression. In Proc. IEEE International Conference on Image Processing, (2015).
6. Hatem H., Beiji Z., Majeed R., Waleed J., Lutf M.: Head pose estimation based on detecting facial features. International journal of multimedia and ubiquitous engineering, Vol. 10, №. 3, pp. 311-322, (2015).
7. Jones M., Viola P.: Fast multi-view face detection. Mitsubishi electric research laboratories (2003).
8. Liao W.K., Fidaleo D., Medioni G.: Robust, real-time 3D face tracking from a monocular view. EURASIP Journal on image and video processing (2010).
9. Scharstein D., Szeliski R.: A taxonomy and evaluation of dense two-frame stereo correspondence algorithms. International journal of computer vision, №. 47, pp. 7-42 (2002).
10. Sharma S.: Template matching approach for face recognition system. International Journal of Signal Processing Systems, Vol. 1, №. 2, pp. 284-289 (2013).
11. Sundararajan K., Woodard D. L.: Head pose estimation in the wild using approximate view manifolds. In IEEE Conference on Computer Vision and Pattern Recognition Workshops, pp. 50-58, (2015).
12. Zakharov A., Tuzhilkin A., Zhiznyakov A.: Finding correspondences in images using descriptors and graphs. Procedia Engineering, № 129, pp. 391-396 (2015).