

## New Generation User Interface based on Hand Tracking in 3D Space

**P. Serafinavičius**

JSC "Telesoftas", Butrimonių str. 9, Kaunas, Lithuania, e-mail: paulius.se@gmail.com

**S. Sajauskas<sup>2</sup>**

Department of Electronics Engineering, Kaunas University of Technology  
Studentų str. 50, Kaunas, Lithuania, e-mail: stanislovas.sajauskas@ktu.lt

### Introduction

Traditional human-computer interfaces are not sufficiently effective in order to exploit all of their benefits of modern electronic devices. Keyboards, manipulators, touch screens are the dominant control devices, but they can be realized only in a very narrow channel for the exchange of information between the human brain and electronic devices (or systems). In addition, all these devices work only by the direct human contact with them. Users of the 3D software applications, smartphones and other modern electronic devices require more natural and more effective interactions due to the growing amount of information processed. The current restrictions of human-computer interfaces constitute an obstacle for humans to exploit the advantages of computers in various aspects of life. Human physical capabilities and computer vision technologies offer the level of development of new approaches for interaction between human and machine.

### State of the Art

There were conducted a series of research in order to determine the current state of user interface based on hand tracking and gesture recognition techniques. However, most of the researchers too less mention about the layout of the cameras, the parameters of them, ambient lighting effect on hand detection and tracking quality and accuracy especially in depth (along Z axis) [1]. The experimental research was not carried out on fully realized systems in order to obtain characteristics on how user interface performs in real operating conditions. Most of the research is limited to simulation results. All this leads to the conclusion that no technique of hand positions tracking is available, which could be used to implement a reliable and efficient user interface for modern electronic devices.

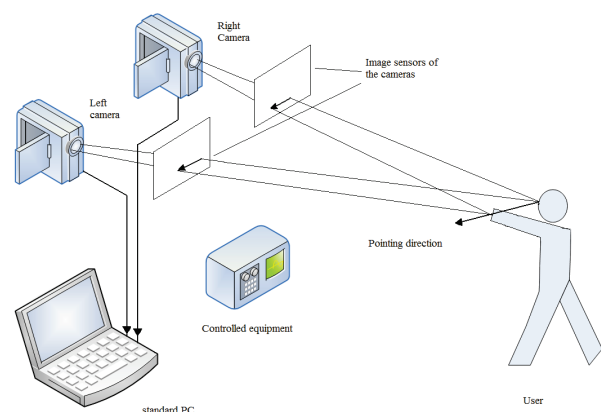
Viola-Jones object detection method, improved by human body skin colour as an additional signal, can help to achieve more reliable user's hand detection. It was necessary to design the algorithm of detection characteristic points of human body and perform

experimental studies of the reliability of hand detection [3, 5].

Since the existing modelling methods of hand's geometric shape is particularly complex and disadvantaged in real-time, the hand tracking in images could be implemented using the technique of optical flow by Lucas-Kanade [2], improved by the identification of human skin colour and the dynamic variation of the parameters for the tracking algorithm. It was necessary to evaluate the robustness of hand tracking, and accuracy of the 3D coordinates of the real environment [5]. The paper presents the further evaluations we made on the implemented hand tracking system.

### Determining the Direction of User Pointing

The basic devices of user hand pointing interaction with the controlled equipments are two web cameras, which are arranged and focused so that user's hand positions fit into the camera field of view (Fig. 1).



**Fig. 1.** The layout of interaction components for the system of determining user's pointing direction

An electronic computerized system performs both the analysis of camera images and calculates the 3D position of user's hand from the projection points of user's hand of

two 2D camera images. The 3D points can be simulated using a pinhole (ideal, non-optical distortions) camera geometric model for both left and right cameras.

On the basis of the stereo camera model (cameras are parallel to their optical axes) and the principle of stereo triangulation, a model of determining user's pointing direction was developed and implemented [4]. At least two points have to be detected of user body (user head and one of his hands) in both of the cameras images in order to estimate his pointing direction.

### Adaptation of the Tracking Method

The detection algorithm based on adapted Viola-Jones method can find the user's head and hand only in certain positions, which were trained on a cascade of boosted classifiers [4]. The user will not maintain of these positions for long period due to fatigue, and the detected areas of the cameras' images will be lost. This leads to the necessity of the continued tracking of the hand position detected once. The method of tracking the features in the image sequence by Lucas-Kanade can be adapted to solve the problem, but it requires additional constraints in order to achieve higher quality of the tracking. Therefore we developed the concentration algorithm for the characteristic points of human body, specified by the good-to-track features in the detected regions. The procedure is as follows [3]:

1. Evaluation of the distance between recently added and the adjacent characteristic points.
2. Calculation of the average characteristic point coordinates which denotes the centre of user's hand or head.
3. Evaluation of distances between the centre point and each of the existing characteristic point.
4. Rejection of the characteristic point and addition a new characteristic point closer to the centre point.
5. Relocation of each characteristic point to a new position based on the maximum likelihood of skin colour.

The result of the algorithm is illustrated in Fig. 2 and Fig. 3. The big white circle denotes the central point, the small ones – characteristic points of hand.

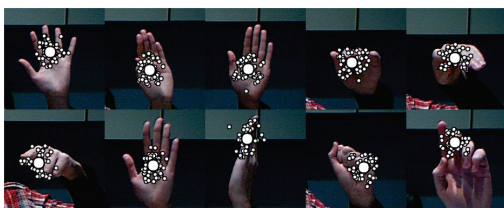


Fig. 2. User's hand views while tracking it at different positions using the algorithm of concentration the characteristic points

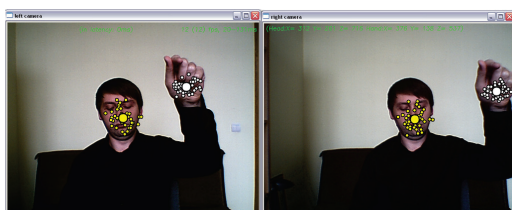


Fig. 3. The concentration of characteristic points while tracking user hand and head for left and right cameras images

### Experimental Evaluation of the System

The evaluation was performed on the hand tracking system (Fig. 4) [5]. It was implemented using two web cameras and a standard notebook PC.

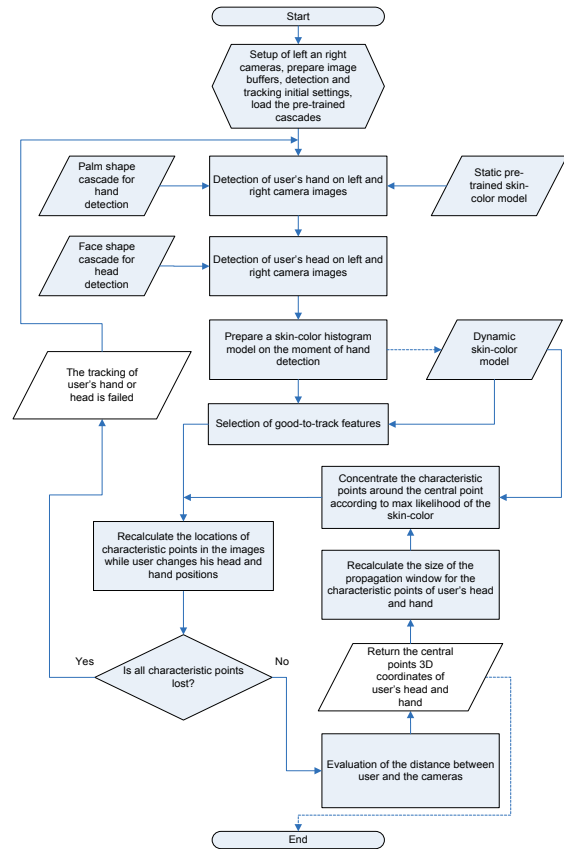


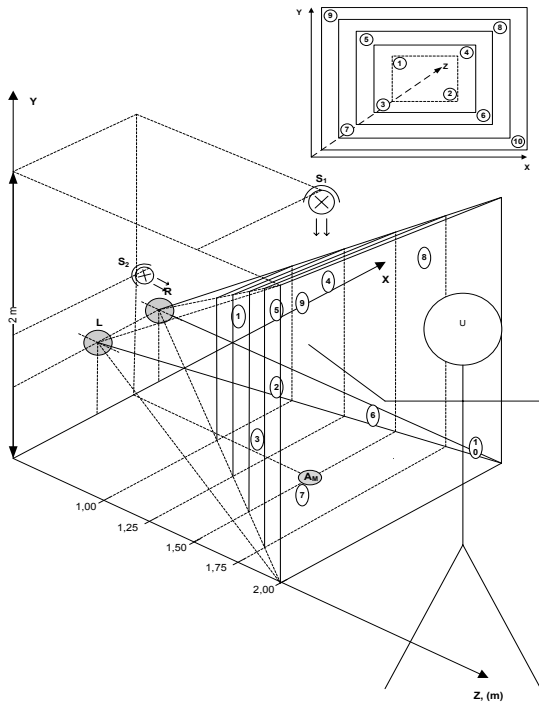
Fig. 4. The general structure-algorithm of the electronic hand position tracking system

The detection rate of the system is the number of tries of successfully detected user head and hand in relation to total number of tries. 32 different persons were involved into the experiment. Each person had to try detecting his head and hand 10 times by using the system. The results are provided in Table 1 when lighting intensity was  $I_a = 300-600$  lx and the type of lighting was a natural daylight.

Table 1. The results of detection rate of the system

	$P_D, \%$	$P(R), \%$	$P(G), \%$
$\bar{P}_D$	91±2	93±2	98±1

The tracking robustness describes the resistance of tracking algorithm to various kind of disturbance, e.g. various objects in the scene, different lighting conditions while tracking user's hand which can lead to tracking failures. The tracking robustness is expressed as a rate of successful tasks in relation to the total tasks user can perform by the system [3]. The evaluation of the tracking robustness of the system was made in different lighting conditions as well. 32 persons were involved into the experiment. Each person had to point his hand to the virtual targets in 3D space, which denote user's tasks to be performed using the system as user interface. The scheme of the experiment is provided in Fig. 5.



**Fig. 5.** Schematics of the experimental evaluation of tracking robustness of the system. S1 – top-oriented tungsten lamp; S2 – front-oriented natural daylight or halogen lamp; AM – measurement point of lighting intensity; L, R – left and right cameras; 1–10 – tasks (virtual points in 3D); U – user

The results of the evaluation the tracking robustness are provided in Table 2.

**Table 2.** The tracking robustness of the system

Natural daylight ( $I_a=300...600$ lx)	35W halogen lamp ( $I_a=130$ lx)	100W tungsten lamp ( $I_a=230$ lx)	60W tungsten lamp ( $I_a=160$ lx)
$P_s = 93 \pm 2 \%$	$P_s = 92 \pm 2 \%$	$P_s = 89 \pm 2 \%$	$P_s = 88 \pm 2 \%$

All the evaluation has been performed in complex background of the scene – pale yellow wall and a variety of colours, shapes and sizes of objects which could interfere the tracking of hand while performing a task. A scene for the investigation was prepared so, that half of all the tasks must be carried out by tracking the hand over the homogeneous background, the other half – in a variety of “barriers”, which consists of visible objects, the colour contrast. The results are presented in Table 3.

**Table 3.** The average tracking robustness rates for each task in natural daylight ( $I_a = 300-600$  lx)

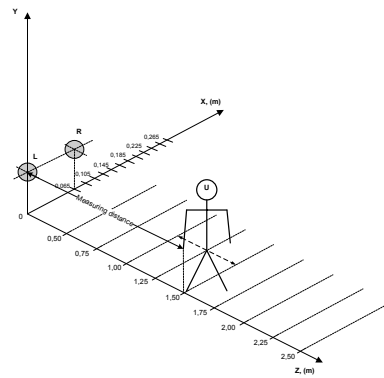
	T <sub>1</sub>	T <sub>3</sub>	T <sub>5</sub>	T <sub>7</sub>	T <sub>9</sub>	T <sub>2</sub>	T <sub>4</sub>	T <sub>6</sub>	T <sub>8</sub>	T <sub>10</sub>
$\overline{P_s}$	90,6	87,5	87,5	81,3	84,4	100,0	100,0	100,0	100,0	96,9
%	$\pm 5,2$	$\pm 5,9$	$\pm 5,9$	$\pm 7,0$	$\pm 6,5$	$\pm 0,0$	$\pm 0,0$	$\pm 0,0$	$\pm 0,0$	$\pm 3,1$
	Complex background, $\overline{P_s} = 86 \pm 6$					Homogeneous background, $\overline{P_s} =$				
	%					99 $\pm$ 1 %				

According to the investigation scheme (Fig. 6) the points T<sub>2</sub>, T<sub>4</sub>, T<sub>6</sub>, T<sub>8</sub>, T<sub>10</sub> were reached by tracking the user's hand over the homogeneous, and the T<sub>1</sub>, T<sub>3</sub>, T<sub>5</sub>, T<sub>7</sub>, T<sub>9</sub> – complex background. The evaluation of the distance accuracy determined by the system was performed by the user. He had to move slowly along the line away from the

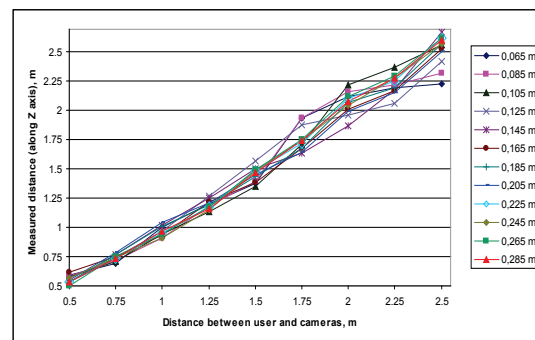
cameras by the graded line. The other user had to register the measured distance by the system. The scheme of the experiment is illustrated in Fig. 6.

12 tests were performed with the same user by changing the distance between the cameras for each trial. The results are provided in Fig. 7 and uncertainty of the measurements – in Fig. 8.

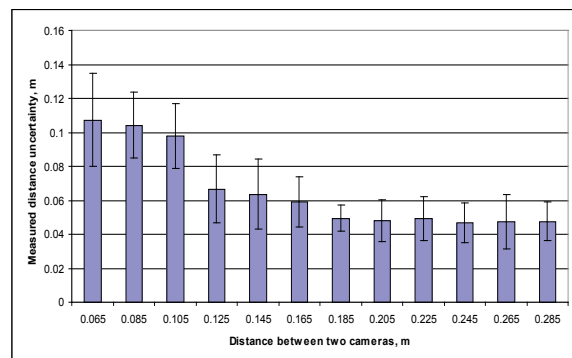
The results show that reasonable distance between the two cameras is 0.185 m. The uncertainty of measurement is up to 0.06 m. Note that these results are camera dependent. In the case of higher quality camera (higher resolution, wider field of view and lower optical distortion) is expected wider measuring range, while the measurement uncertainty remains the same.



**Fig. 6.** Scheme of measuring distance between the user and cameras. L, R – left and right cameras, U – user



**Fig. 7.** Relation of measured user's hand distance (by Z coordinate) and real distance between the cameras and user's hand. 12 tests were conducted by changing the distance between left and right camera for each test



**Fig. 8.** Influence of the distance between the cameras to 3D depth accuracy. Uncertainties for each distance are provided with a standard deviation of them

## Conclusion

Evaluated detection rate of the system in the presence of 32 different users is sufficiently high, and is equal to  $(91 \pm 2)\%$ .

The robustness rate of user's hand tracking in 3D space is equal to  $(93 \pm 2)\%$ . Dependence of the robustness rate on lighting conditions are low, however, to ensure high reliability of the tracking, the illumination of  $\geq 130$  lx and halogen lamps or natural daylight, oriented in front of the user, are recommended. The best tracking results were obtained for homogeneous background (tracking robustness rate –  $(99 \pm 1)\%$ ); the reliability of the tracking rate is  $(86 \pm 6)\%$  for a complex background case. Small variations of the robustness rates, while changing ambient lighting type, intensity and background conditions, are acceptable for alternative computer vision-based user interfaces.

The uncertainty of measuring distance in a 3D space by the system increases by 1.5 times where distance between the cameras is less than optimal (0.185 m).

## References

1. **Lipnickas A., Knyš A.** A Stereovision System for 3-D Perception // *Electronics and Electrical Engineering*. – Kaunas: Technologija, 2009. – No. 3(91). – P. 99–102.
2. **Choi C., Baek S. M., Lee S.** Real-time 3D object pose estimation and tracking for natural landmark based visual servo // *Intelligent Robots and Systems, 2008. IEEE/RSJ International Conference*. – P. 3983–3989.
3. **Serafinavičius P.** Estimating Characteristic Points of Human Body for Automatic Hand Pointing Gesture Recognition // *Electronics and Electrical Engineering*. – Kaunas: Technologija, 2007. – No. 8(80). – P. 83–86.
4. **Serafinavičius P., Daunys G.** Detection of Hand Position using 3-D Computer Vision // *Electronics and Electrical Engineering*. – Kaunas: Technologija, 2006. – No. 7(71). – P. 63–66.
5. **Serafinavičius P., Sajauskas S., Daunys G.** Evaluation of Hand Pointing System Based on 3-D Computer Vision // *Electronics and Electrical Engineering*. – Kaunas: Technologija, 2008. – No. 8(88). – P. 95–98.

Received 2009 09 15

### **P. Serafinavičius, S. Sajauskas. New Generation User Interface based on Hand Tracking in 3D Space // *Electronics and Electrical Engineering*. – Kaunas: Technologija, 2010. – No. 3(99). – P. 69–72.**

The hand tracking system in 3D space dedicated for design of new generation user interface is presented. The system is based on 3D computer vision and implemented using open source computer vision library (OpenCV). The further experimental evaluation of the system was performed closer as possible to meet the real environment conditions. The system is able to detect and track user's head and hand and return 3D coordinates in real time. The detection of head and hand is based on adapted Viola-Jones detector. The detection rate is equal to  $(91 \pm 2)\%$ . The tracking is based on Lucas-Kanade optical flow calculation technique which was customized and improved for hand tracking. High tracking robustness ( $(93 \pm 2)\%$ ) was achieved. Small variations of the robustness rate, while changing ambient lighting type, intensity and background conditions, are acceptable for alternative computer vision-based user interface. Ill. 8, bibl. 5 (in English; abstracts in English, Russian and Lithuanian).

### **П. Серафинавичюс, С. Саяускас. Пользовательский интерфейс нового поколения, основанный на слежении руки в 3D-пространстве // *Электроника и электротехника*. – Каунас: Технология, 2010. – № 3(99). – С. 69–72.**

Представлена система слежения в 3D-пространстве, предназначенная для разработки нового поколения пользовательского интерфейса. Система основана на 3D компьютерном зрении и реализована используя библиотеку открытого кода компьютерного зрения (OpenCV). Экспериментальная оценка системы выполнена для реальной окружающей среды. Система способна обнаруживать и отслеживать голову и руку пользователя в реальном масштабе времени, а также вернуть их 3D координаты. Разработанный детектор обнаружения головы и руки выполнен на основе метода Виола-Джонс. Уровень обнаружения равен  $(91 \pm 2)\%$ . Отслеживание проводится по методике Лукас-Канаде для расчета оптического потока, примененного для слежения рука. Достигнут высокий уровень надежности отслеживания ( $(93 \pm 2)\%$ ). Небольшие изменения надежности вследствие типа и изменения интенсивности освещения окружающей среды, фоновых условий являются приемлемыми для альтернативного интерфейса пользователя, основанного на компьютерном зрении. Ил. 8, библи. 5 (на английском языке; рефераты на английском, русском и литовском яз.

### **P. Serafinavičius, S. Sajauskas. Rankos sekimo 3D erdvėje pagrįsta naujos kartos naudotojo sąsaja // *Elektronika ir elektrotechnika*. – Kaunas: Technologija, 2010. – Nr. 3(99). – P. 69–72.**

Pateikiama rankos sekimo 3D erdvėje sistema, skirta naujos kartos naudotojo sąsajai sudaryti. Ši sistema buvo sukurta stereoregos pagrindu ir realizuota naudojant atvirojo kodo kompiuterinės regos biblioteką (OpenCV). Pristatomi tolimesni sistemos eksperimentiniai tyrimai. Sistema gali aptikti ir sekti žmogaus galvą ir ranką realiu laiku bei grąžinti jų 3D koordinates. Galvos ir rankos detektorius buvo sukurtas adaptavus Violos ir Joneso detektavimo metodą. Gautas detektavimo patikimumas yra  $(91 \pm 2)\%$ . Rankos ir galvos sekimas yra pagrįstas Lucaso ir Kanade'o optinio srauto skaičiavimo metodu. Pasiektas didelis sekimo patikimumas ( $(93 \pm 2)\%$ ). Sekimo patikimumo rodiklio svyravimai priklausomai nuo aplinkos apšvietos tipo, intensyvumo ir foninio vaizdo sąlygų yra priimtini alternatyvioje, kompiuterine rega pagrįstoje naudotojo sąsajoje. Il. 8, bibl. 5 (anglų kalba; santraukos anglų, rusų ir lietuvių k.).

DOI: 10.5755/j02.eie.9908