

## SCALABLE PHOTOGRAMMETRIC MOTION CAPTURE SYSTEM “MOSCA”: DEVELOPMENT AND APPLICATION

V. A. Knyaz<sup>a, \*</sup>,

<sup>a</sup>St. Res. Institute of Aviation Systems (GosNIIAS), 125319, 7, Victorenko str., Moscow, Russia - [knyaz@gosniias.ru](mailto:knyaz@gosniias.ru)

Commission V, WG V/5

**KEY WORDS:** Photogrammetry, accuracy, calibration, motion capture, tracking

### ABSTRACT:

Wide variety of applications (from industrial to entertainment) has a need for reliable and accurate 3D information about motion of an object and its parts. Very often the process of movement is rather fast as in cases of vehicle movement, sport biomechanics, animation of cartoon characters. Motion capture systems based on different physical principles are used for these purposes. The great potential for obtaining high accuracy and high degree of automation has vision-based system due to progress in image processing and analysis. Scalable inexpensive motion capture system is developed as a convenient and flexible tool for solving various tasks requiring 3D motion analysis. It is based on photogrammetric techniques of 3D measurements and provides high speed image acquisition, high accuracy of 3D measurements and highly automated processing of captured data. Depending on the application the system can be easily modified for different working areas from 100 mm to 10 m. The developed motion capture system uses from 2 to 4 technical vision cameras for video sequences of object motion acquisition. All cameras work in synchronization mode at frame rate up to 100 frames per second under the control of personal computer providing the possibility for accurate calculation of 3D coordinates of interest points. The system was used for a set of different applications fields and demonstrated high accuracy and high level of automation.

### 1. INTRODUCTION

Nowadays motion capture as a process of acquiring real 3D movement of an object (or a set of points representing an object) for a further processing is of high demand by many applications. The most known fields of motion capture usage are movie and video game production where accurate registration of 3D motion provides a high impression of reality to virtual creatures.

#### 1.1 Types of motion capture system

A few types of motion capture systems are now in use. Among them there are mechanical, acoustical, magnetic, optical systems.

Mechanical systems use potentiometers and sliders located in the required positions on an actor and provide registration of their spatial positions. They have some advantage such as an interface that is similar to stop-motion systems widely used in the film industry. Other advantages are independence from magnetic fields or reflections and short setting up time. Their main disadvantage is restriction caused by wires which connect sensors to registration system.

In acoustical system a set of acoustic receptors capture sounds from sound transmitters located on the object (actor). The specific sounds from emitters then picked up by receivers and 3D positions of emitters are calculated using registered times between emitting and receiving signal. To determine the 3D

position of each transmitter, a triangulation of the distances between the emitter and each of the receptors is computed.

Acoustical motion capture systems have some problems which make them inconvenient in a number of cases. These problems are: the restrictions to the freedom of movement caused by the cables put on the actor, the limited number of transmitters that can be used and susceptibility to sound reflections or external noise.

Being not comparatively expensive magnetic systems are rather accurate and fast (about 100 fps) for simple movement capture. They use a set of magnets as markers of given points and a set of receivers for measuring the position and orientation of the markers relative to an antenna (Yabukami, 2000). The disadvantages of magnetic systems are also limitations caused by cables and possible interference in the magnetic field caused by various metallic objects and structures.

Typical vision-based motion capture systems usually include a set of cameras capturing video sequences of an actor/object on which special targets are placed. Then video sequences are processed for target detecting, identifying and tracing through the sequence. The level of accuracy of 3D point coordinates calculation is provided by calibration procedure and depends on application needs.

These systems are the most expensive ones in the market due to their cutting-end technological nature, such as the high-

---

\* Corresponding author.

resolution cameras and sophisticated proprietary software. The cost reaches hundreds of thousands USD.

The doubtless advantages of such systems are possibility of capturing at very high speed, no limitation for actor moving in the working space, great potential for automation of the process.

## 1.2 Applications

The initial impulse for creating motion capture systems was done by entertainment industry which had a need for a mean of fast and accurate actor movements transfer into the movie or animation. And now movie and video game production industries are the main users of motion capture systems.

But the field of application for motion capture systems grows very dynamically. Among major areas of application there are medicine, sport, various branches of industry, scientific researches (Moeslund, 2006).

Motion capture systems in medical applications are used for accurate analysis of human motion which cannot be registered by other means. 3D study of human motion allows to find abnormalities and propose the way of rehabilitation.

In sport of high results motion capture systems are of great demand because they provide valuable information about high speed motion of a sportsman during competition. This information is the basis for improving the sport technic and achieving better results. Golf is one of the major users of motion capture systems for analysis and correction of sport technics.

Motion capture is often the single tool for scientific research of specific tasks where information about object 3D movement could not be obtained by other means. The typical examples are very fast dynamical process analysis, analysis of vehicle vibration, object 3D trajectory estimation and analysis and similar projects.

## 2. SYSTEM OUTLINE

For a vision-based photogrammetric motion capture system a reliability and a convenience for a user are the key features defining the quality of the system. In this aspect a detection and tracking of given object points required for an application tasks play the essential role. Also if the area of application and scale of the captured process can be variable it is required that the motion capture system can work with different working area size and can be easily reconfigured and recalibrated. These are the main requirement which were in consideration in developing the system.

### 2.1 Hardware

The developed scalable 3D motion capture system “Mosca” is based on photogrammetric techniques for 3D measurements and provides high speed image acquisition, high accuracy of 3D measurements and high level of automation of captured data. Depending on the application the system can be easily modified for different working areas from 100 mm to 10 m. The developed motion capture system uses from 2 to 4 technical vision cameras for video sequences of object motion acquisition. All cameras work in synchronic mode at frame rate up to 100 frames per second under the control of personal

computer (PC) providing the possibility for accurate calculation of 3D coordinates of interest points. The system could be extended to more cameras by including an additional PC station in the system.

The original camera calibration and external orientation procedure is used to reach high accuracy of 3D measurements. The calibration procedure is highly automated due to applying original coded targets (Knyaz, 1998) for identifying and measuring image coordinates of reference points. The system calibration provides accuracy of 0.01% of working space (WS) of the motion capture system.

Technical characteristics of the motion capture system are presented in table 1.

Camera resolution	656 x 491 pixels
Acquisition speed	up to 100 fps
Number of tracking points	up to 200
Working space (WS)	Scalable: from 0.1x0.1x0.1m to 10x10x10 m
3D point coordinate accuracy	0.01% of WS

Table 1. Technical characteristics of the Mosca motion capture system

The possibility for varying the scale of imaging is provided by fast and highly automated procedures for calibration and exterior orientation.

### 2.2 Calibration

Calibration is performed using original technique and original software. Classical central projection model is used for camera imaging process. With given centre of projection  $\mathbf{X}_o = (X_0, Y_0, Z_0)$  for object point  $a$  with spatial coordinates  $\mathbf{X} = (X, Y, Z)$  its image coordinates  $\mathbf{x} = (x, y)$  can be found from the collinearity equation:

$$\mathbf{X} = \mathbf{X}_o - \mu \mathbf{A}^T (\mathbf{x} - \mathbf{x}_p)$$

Where  $\mathbf{A}$  is transition matrix,  $\mathbf{x}_p$  – image coordinates of principal point,  $\mu$  – scale factor.

The additional parameters describing CCD camera model in collinearity conditions are taken in form:

$$\begin{aligned} \Delta x &= a\bar{y} + \bar{x}r^2 K_1 + \bar{x}r^4 K_2 + \bar{x}r^6 K_3 + (r^2 + 2\bar{x}^2)P_1 + 2\bar{x}\bar{y}P_2 \\ \Delta y &= a\bar{x} + \bar{y}r^2 K_1 + \bar{y}r^4 K_2 + \bar{y}r^6 K_3 + 2\bar{x}\bar{y}P_1 + (r^2 + 2\bar{y}^2)P_2 \\ \bar{x} &= m_x (x - x_p); \bar{y} = -m_y (y - y_p); r = \sqrt{\bar{x}^2 + \bar{y}^2} \end{aligned}$$

where  $x_p, y_p$  - the coordinates of principal point,  
 $m_x, m_y$  - scales in  $x$  and  $y$  directions,  
 $a$  - affinity factor,  
 $K_1, K_2, K_3$  - the coefficients of radial symmetric distortion  
 $P_1, P_2$  - the coefficients of decentring distortion

The common procedure for determining unknown parameters of camera model is bundle adjustment procedure using observations of test field reference points with known spatial coordinates (Knyaz, 2002).

Image interior orientation and image exterior orientation ( $X_i, Y_i, Z_i$  – location and  $\alpha_i, \omega_i, \kappa_i$  and angle position in given coordinate system) are determined as a result of calibration. The residuals of co-linearity conditions for the reference points after least mean square estimation  $\sigma_x, \sigma_y$  are concerned as precision criterion for calibration.



Figure 1. Exterior orientation

The results of the cameras interior orientation parameters estimation by described technique are presented in Table 2.

	Camera 1	Camera 2	Camera 2
$m_x$	0.00943414	0.00935769	0.00929585
$m_y$	0.00944242	0.00935618	0.00928185
$b_x$	291.4501	298.077039	289.09868
$b_y$	255.45588	255.501334	253.84673
$a$	-0.00668	-0.00617081	-0.005961
$K_1$	0.0008712	0.001612895	0.0007685
$K_2$	-0.0000182	0.000091961	0.00001594
$K_3$	0.00000032	-0.000007424	-0.00000041
$P_1$	0.0002827	-0.000030034	-0.000110274
$P_2$	0.00004459	0.00014336	0.000081377

$\sigma_x$	0.30 mm
$\sigma_y$	0.31 mm

Table 2. Sample of calibration results for system configuration for 3x3x3 m working area

The external orientation of the motion capture system is performed after choosing a working space and camera configuration for motion capture. For external orientation the

same test field is used. It defines the global coordinate system in which 3D coordinates are calculated.

### 3. ALGORITHMS FOR AUTOMATION

The problem of given object point detection and tracing is of great importance for the motion capture system. For most part of applications it is needed to measure 3D coordinates of given points so these points have to be marked on the objects by special targets which should be detected and identified in the image. Coded targets could not be applied in this case because of their rather large size required for reliable identification. So for automation of the target detection and identification some techniques were developed.

#### 3.1 Algorithms for target detection

**3.1.1 Algorithm assumption:** Algorithm works in assumption that the target is a connected region in the image which meets to three conditions:

1. There is a single maximum of intensity for any section through the center of the probable region.
2. The value of this maximum is greater than given threshold B.
3. Dimensions of the region belong to given range between  $D_{min}$  and  $D_{max}$ .

**3.1.2 Algorithm description.** Algorithm is based on image  $I(x,y)$  binarization by the sequence of thresholds  $h_{max}, h_{max}-s, h_{max}-2s, \dots, h_{max}-ns$ ; finding all connected regions in every binary image and selecting only such regions which meet to conditions 1-3.

Algorithm's parameters:

$h_{max}$  – maximum value of intensity for binary image. The recommended value of  $h_{max} = I_{max} - B/2$ .

$h_{min}$  – minimum value of intensity for current binary image.

$s$  – a step in intensity increasing.

$B$  – minimal intensity of target to search.

$D_{min}$  and  $D_{max}$  – minimum and maximum values of a target dimension.

**3.1.3 Algorithm steps:** Algorithm includes the following steps:

1. Building the intensity histogram  $HIST[0.. I_{max}]$  for initial image  $I(x,y)$ .

2. Choosing the initial intensity level:

$$h = h_{max}$$

3. Choosing the new intensity level of binary image:

$$\text{until } HIST[h]=0 \text{ } h:=h-1$$

4. Image  $I(x,y)$  binarization at threshold  $h$ :

$$\text{if } I(x,y) > h \text{ then } B(x,y):=1 \\ \text{else } B(x,y):=0$$

5. Searching for all connected regions in image  $B(x,y)$  and creating the array of descriptors for every detected region R:

- coordinates of upper left  $(x_1, y_1)$  and lower right  $(x_2, y_2)$  corners of the minimal rectangle containing detected region R
- maximal  $B_{max}$  and minimal  $B_{min}$  values of intensity in detected region R;
- the number N of regions detected at previous binary image, belonging to region R;
- unique number  $M(x, y)$  of region R which allow to check if any pixel  $(x, y)$  belongs to the region R

6. Output image  $R(x, y)$  includes only that regions for which the condition 1-3 are carried out:

if  $(x_2 - x_1 > A_{min})$  and  $(y_2 - y_1 > A_{min})$   
and  
 $(x_2 - x_1 < A_{max})$  and  $(y_2 - y_1 < A_{max})$   
and  
 $(B_{max} - B_{min} > B)$  and  $(N < 2)$   
than  $R(x, y) := 1$   
else  $R(x, y) := 0$

7.  $h := h - s$

8. If  $h > h_{min}$  then go to 3.

9. Finding all connected regions in the image  $R(x, y)$ .

10. For every region R in the binary image  $R(x, y)$  coordinates of its center of mass are calculated. The resulting sub-pixel coordinates  $(x^*, y^*)$  are the coordinates of the target:

$$x^* = \frac{\sum_{x, y \in R} xI(x, y)}{\sum_{x, y \in R} I(x, y)}$$

$$y^* = \frac{\sum_{x, y \in R} yI(x, y)}{\sum_{x, y \in R} I(x, y)}$$

### 3.2 Algorithm for searching corresponding points in the captured images

After detection all  $n_t$  targets by described above algorithm their coordinates  $(x_i^j, y_i^j)$  for every target  $t_i, i=1, \dots, n_t$  are known for every image  $I_j, j=1, \dots, n_t$ . Also parameter of exterior orientation for every image  $I_j, j=1, \dots, n_t$  are known due to preliminary exterior orientation procedure.

For determining 3D coordinates of the target it is required to find the correspondence between detected targets in different images. Because all target have the same shape it is impossible to apply correlation or descriptor-based methods. If the number of cameras more than two epipolar geometry could be used for identification of similar targets in different images.

For point  $p_i^1$  in the frame from the first camera its image in the frame from second camera should lay on the epipolar line  $r_1^1$  which is an intersection of the plane defined by three points (center of projection of the first camera, center of projection of the second camera and the image of the point  $p_i^1$  in the frame from the first camera) and projection plane of the second camera (figure 2).

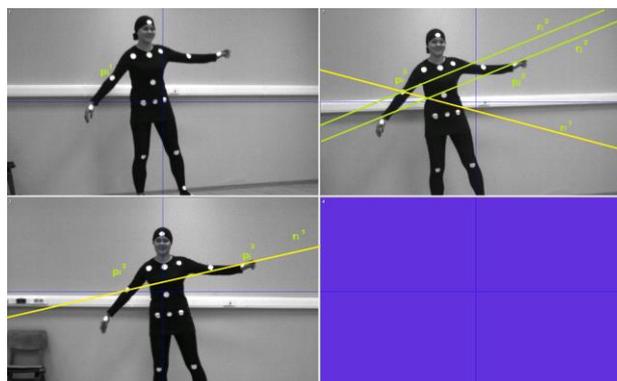


Figure 2. Epipolar points searching

So the targets  $p_i^2$  and  $p_j^2$  detected in the frame from the second camera and lying on the epipolar line  $r_1^1$  are potential images of point  $p_i$ . In a similar way point  $p_i^1$  will be imaged in the frame from the third (fourth, etc) camera as epipolar line and the points regarded as possible images of  $p_i^1$  will be presented in the frame from the third camera as a set of epipolar lines  $r_1^2, r_j^2$  from the second camera. So the corresponding point  $p_i^3$  to point  $p_i^1$  in the third image will be the point of target location in which epipolars are intersected. Figure 2 illustrates the algorithm for points correspondence searching.

This algorithm allows to establish the correspondence between targets images from more than two cameras and then provide the possibility to resolve collision when tracing target along the recorded video sequence.

### 3.3 Software

The original software for synchronic video sequences capture and their automated processing was developed. The software supports a set of procedures for motion capture and processing such as:

- automated system orientation
- video sequences capture in synchronic mode
- automated target detection and correspondence problem solution
- automated target tracing
- 3D trajectory calculation and visualisation

## 4. APPLICATIONS

The developed photogrammetric motion capture system is applicable for wide variety of applications where accurate, fast and reliable data on object (object points) is needed. Among application in which Mosca was used are biometry and biomechanics, robot dynamical model identification, unmanned aerial vehicle (UAV) self-orientation accuracy evaluation, virtual objects control.

### 4.1 Human motion capture

Human biomechanics is one of the important applications which need a mean for fast and accurate human motion in different modes. In this case usually it is required to obtain accurate 3D trajectories of given points of a human body. Depending on the task to solve the Mosca could capture full human body motion or movements of body parts (e.g., facial expressions) with higher accuracy.



Figure 3. An actor with targets placed according BVH model

For human motion capture the Mosca is configured for working space of about 2.5x2.5x2.5 m so that the required movement of an actor could be captured. Figure 3 presents an actor with targets placed according BVH model during the acquisition process.

In figure 4 software interface is shown with detected and identified targets and biped 3D model generated using captured data.

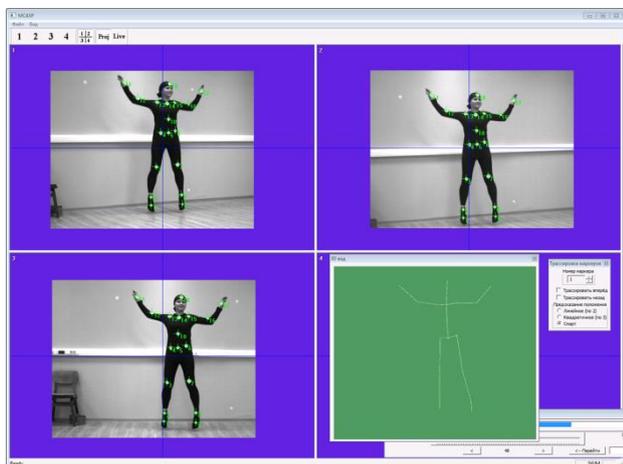


Figure 4. Software interface with detected and identified targets and biped 3D model

#### 4.2 UAV accuracy estimation

The Mosca system was used for estimation the accuracy of self-orientation of unmanned aerial vehicle (UAV). UAV Parrot AR.Drone2.0 was used for experiments.

Its basic technical characteristics are given in Table 3.

Parameter	Value
Velocity	0.5 m/s
Weight	366 g
Dimensions	515x515 mm

Frontal camera	
- Frame rate	30 fps
- Resolution:	1280x720 pix
Vertical camera	
- Frame rate	60 fps
- Resolution:	320x240 pix

Table 3. UAV Parrot AR.Drone2.0 technical characteristics

Figure 5 presents Parrot AR.Drone2.0 bottom view with indication of targets' placement (1, 2, 3, 4, 5) and the position of the frontal camera (*a*).

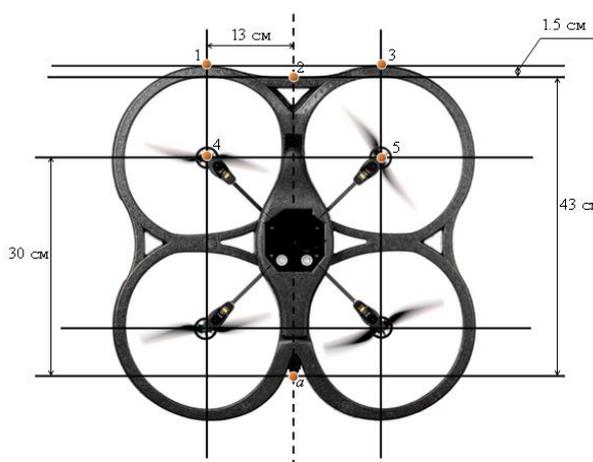


Figure 5. Parrot AR.Drone2.0 bottom view

The problem to be solved was to determine with what accuracy the UAV could determine its own position and orientation basing on processing video information from the frontal camera. Preliminary calibration of UAV frontal camera was carried out using a set of images of the test field captured by UAV frontal camera. Figure 6 shows the test field image acquired by the frontal camera. The accuracy of the UAV camera calibration was at the level of 0.5 mm.



Figure 6. The test field image acquired by frontal camera.

Then the flight trajectory of the set of targets placed on the UAV was captured by the Mosca along with acquiring video sequences of this flight by UAV frontal camera. Both data sets were processed resulting in two 3D trajectories: captured by the

motion capture system and self-estimated by UAV. Both trajectories were registered in common coordinate system defined by the test field. Coded targets were used to provide high accuracy and automation of the process. The synchronization between these two trajectories was performed using special light marker in the acquired video sequences.

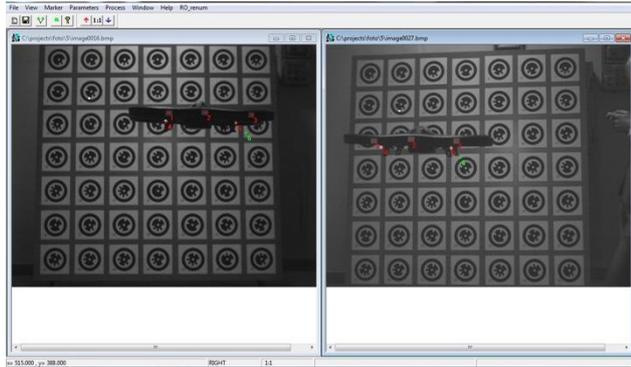


Figure 7. The images of UAV during the flight acquired by the motion capture system.

The position of the UAV frontal camera was estimated by using coordinates of the targets in coordinate system connected with the UAV and coordinates of the targets in the motion capture coordinate system. The mean errors in UAV position and UAV angular orientation is given in table 3.

$\alpha$	$\omega$	$\kappa$
0.0232	-0.08392	0.03281
X, mm	Y, mm	Z, mm
-5.2618	8.0955	27.2022596

Table 4. The mean errors in UAV position and UAV angular orientation

The results of verifying the UAV self-orientation by video from frontal camera demonstrated a high potential for vision-based techniques for UAV navigation.

### 4.3 Skied-steered robot dynamic model identification

In system identification problem it is important to have accurate data about system output on given input. In case of skied-wheel robot it is needed to register output velocity and angular orientation of the vehicle and this data has to be synchronized with input commands. For obtaining the required information accelerometers could be used but the accuracy of state vector components is not enough and synchronized problem requires some additional hardware.

The developed Mosca system was used for synchronized output registering during model identification of Hercules skied-wheel robot. The accuracy and sample rate of the motion capture system are adequate to the task of dynamic model identification. Special program block for synchronization of the robot input commands and capturing frames was developed and implemented.

For registering the robot moving during the experiment a set of circular targets located on the robot upper deck. The central target (#8) defines the centre of the robot coordinate system.

Targets #6 and #10 defines X axis of the robot coordinate system, targets #3 and #13 defines Y axis. These points were used for calculation of output parameters  $v_x$  and  $\theta$ .

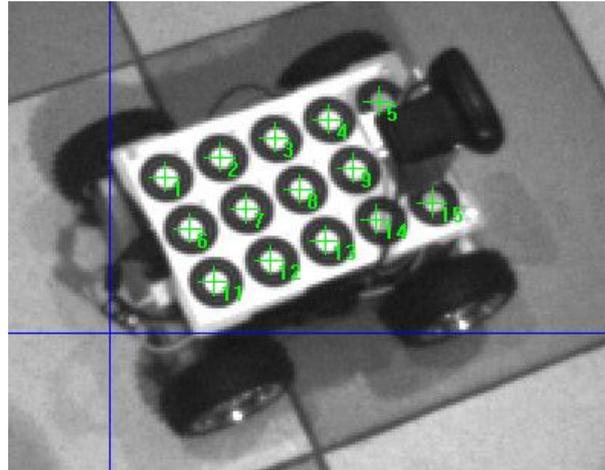


Figure 8. Hercules skied-steered robot with targets

To estimate the full dynamics model of Hercules robot a white noise command sequence was generated. Resulting motion was recorded using motion capture system. The input signal used for the experiment is given by:

$$U(t) = \begin{bmatrix} a_m \cdot \text{rand}(t) \\ d_m \cdot \text{rand}(t) \end{bmatrix}$$

Here  $a_m$  – is maximum average PWM value,  $d_m$  – is maximum command difference,  $\text{rand}(t)$  – random number uniformly distributed over an interval  $[-1, 1]$ . The resulting command sequence, longitudinal speed  $v_x$  and rotational speed  $\theta$  captured using the motion capture system for an experiment with  $a_m = 30$  PWM and  $d_m = 60$  are shown on figure 9.

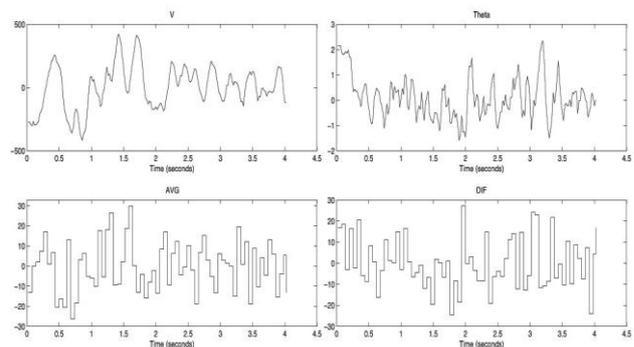


Figure 9. Output  $v_x$ ,  $\theta$  and input commands used for white noise signal experiment

## 5. CONCLUSION

The photogrammetric scalable motion capture system Mosca is developed. The Mosca system uses from 2 to 4 technical vision cameras for video sequences of object motion acquisition. It can be easily modified for different working areas from 100 mm to 10 m. The original algorithms for object point detecting, identifying and tracking is developed which provides high level of automation for video sequence processing.

Some results of using the Mosca photogrammetric system in various fields of applications such as biomechanics, dynamical model identification, self-orientation accuracy estimation are presented and discussed. The application results show high accuracy and high reliability of the developed photogrammetric system for 3D motion capture.

#### ACKNOWLEDGEMENTS

The work was performed with the support by Grant of RF President for Leading Science schools of Russia (HIII-3477.2014.8) and by Grant 15-08-99580a of Russian Foundation for Basic Research (RFBR).

#### REFERENCES

Knyaz V.A., Sibiriyakov A.V., 1998. The Development of New Coded Targets for Automated Point Identification and Non-contact 3D Surface Measurements. In: *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, Hakodate, Japan, Vol. XXXII, part 5, pp. 80-85.

Knyaz V.A., 2002. Method for on-line calibration for automobile obstacle detection system. In: *International Archives of Photogrammetry and Remote Sensing*, Proceedings of ISPRS Commission V Symposium "CLOSE-RANGE IMAGING, LONG-RANGE VISION", Vol. XXXIV, part 5, Commission V, September 2-6, Corfu, Greece, pp. 48-53

Moeslund, T. B., Hilton, A., and Kruger, V., 2006. A survey of advances in vision-based human motion capture and analysis. In: *Computer Vision and Image Understanding* 104, pp. 90–126.

Yabukami S., Kikuchi H., Yamaguchi M., Arai K. I., Takahashi K., A. Itagaki, Wako N., 2000. Motion capture system of magnetic markers using three-axial magnetic field sensor. In: *IEEE Transactions on Magnetics*, Vol: 36, Issue: 5, pp: 3646-3648. 2000.