Computer Assisted Mechanics and Engineering Sciences, 14: 637–649, 2007.


# 3D measurements and motion analysis supported by passive vision techniques

Piotr Kohut
*AGH-University of Science and Technology, Department of Robotics and Mechatronics*
*Al. Mickiewicza 30, 30-059 Cracow, Poland*

(Received in the final form September 18, 2007)

**The article contains presentation of application of three-dimensional vision methods in realization of vibration measurements and their analysis. For this purpose algorithms were developed of discrete epipolar geometry and structure from motion, with the usage of one camera. Vibration amplitude is determined for selected measurement points on the analyzed object. Each point is represented by flat or three-dimensional marker attached on a construction. The article includes algorithms of the discussed methods and verification of those methods based upon simulation data, as we as preliminary experimental tests carried out on a test bed.**

**Keywords:** 3D vision techniques, epipolar geometry, structure from motion, vibration measurements

## 1. Introduction

Contemporary applications of non-contact methods of signals registration have conformed to a new tendency in construction designing, testing, diagnostic and monitoring. In a great number of optic methods, two categories of systems can be distinguished in the area of vision techniques used for 3-D object geometry measurements: active and passive. The former, in depth measurements, take advantage of additional devices (e.g. lasers, LCD projectors) to generate structured light. The latter execute depth measurements based on image-sequences provided by one or more cameras. The advantage of active methods is very high accuracy, however, the main disadvantage is its very high cost. They are also completely insensitive to texture variation in a scene which restricts its applicability. The passive methods are often more flexible, but computationally more expensive and dependent on the structure of the scene itself.

Stereovision methods belong to the most natural ones, since they make use of two images of the same scene to inference about the three-dimensional properties of the scene. The advantage of a stereoscope method is its reasonable accuracy. The downside is the requirement of using two cameras, which affects the increase in complexity and economic expenses, and failure to apply this method to non-textured surfaces. Another approach to stereovision techniques involves, in general, replacing the pair of cameras by a single moving camera. In this case the single camera records two images of a scene at two different locations and at two different time moments. Reconstruction procedure is identical to stereovision: three-dimensional structure is determined by means of triangulation, based upon the two obtained images. The advantage of this approach is its low cost (single camera) and ergonomic properties. Within the confines of those techniques fits also the methods of structure from motion, which, in contrast to stereovision, applies only a single camera as an assumption.

The paper indicates the use of 3D vision passive measurement system for the analysis of quantities which characterize dynamic properties of constructions. The purpose of the research was developing and creating a vision-based method for reconstructing three-dimensional motion and structure of objects. The project included algorithms allowing for 3D reconstructions with the use of a single camera, matching the characteristics of the assigned objects by means of the images recorded by vision system and algorithms determining relative camera locations in individual motion phases. Motion parameters are specified for planar or three-dimensional markers put on an examined structure.

## 2. 3D ALGORITHMS BASED ON EPIPOLAR GEOMETRY AND STRUCTURE FROM MOTION

For the purpose of measurement of amplitude of vibrations and three-dimensional object structure, an extended application of stereovision techniques was presented. These techniques include discrete methods of epipolar geometry and techniques of „structure from motion". For both methods algorithms and procedures were developed, that enable obtaining amplitudes of vibrations of analyzed scene objects (measurement points) along with their three-dimensional structure based upon data derived from a single fast digital camera. The main difference between those methods stems from the manner of determining motion parameters necessary for drawing the vibration characteristics of analyzed objects. In case of epipolar geometry the motion parameters are determined between two consecutive image frames provided by the camera, whereas in the latter so called „batch method" a whole sequence of images recorded by a camera is necessary. After executing factorization of measurement matrix and superimposing of metric conditions it is possible to obtain three-dimensional object-structure and transformation vector representing object vibrations.

### 2.1. Epipolar geometry

Epipolar geometry is complete generalization of stereo geometry. It describes the geometric relations between two perspective-views of the same 3D scene (Fig. 1). The key feature is that corresponding image points must lie along particular image lines, which implies, that given a point in one image, one can search the corresponding point in the other one along the line, but not in a 2D region. Thus has the epipolar geometry enabled significant reduction of calculation complexity. Search for corresponding elements is reduced to a 1D problem. The epipolar geometry restricts that corresponding points must lie on conjugated epipolar line.

Epipolar geometry is dependant only on the camera parameters (internal and external), and is independent of the 3D structure of a scene. The calibrated reconstruction requires that the camera

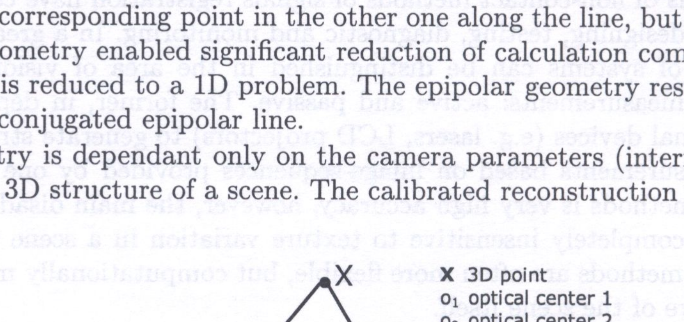



**Fig. 1.** Geometrical framework of epipolar geometry. *Epipolar plane*: any plane containing baseline; *Base line*: a line that, after having been cut by image planes, forms the projection of epipolar lines ($l_1$ and $l_2$) on those planes; *Epipolar line* ($l_1$ , $l_2$): intersection of epipolar plane and image plane; *Epipoles* ($e_1$ , $e_2$): the points where baseline crosses image planes; the projection of the center of the other camera; vanishing point of the direction of camera motion

be internally calibrated. It means that the camera calibration matrix **K** must be known. In this case external camera parameters $(\mathbf{R}, \mathbf{t})$ and 3D structure of the scene can be solved. **R** and **t** can be determined by factoring the essential matrix. Then the 3D reconstruction and depths can be calculated by means of triangulation in the appropriate scale.

Epipolar geometry provides solutions of the three basic 3D reconstruction problems:

1. Geometry matching. Given: coordinates of points in the first image $x$; Demanded: locations of the related points on the other image $x'$;

2. Camera geometry (motion). Given: set of corresponded with each other points on images $\{x_i \leftrightarrow x_i'\}$, $i = 1, \ldots, n$; Demanded: camera matrices **P** and **P**′;

3. Scene geometry (structure). Given: set of corresponded with each other points on images $\{x_i \leftrightarrow x_i'\}$ and camera matrices **P** and **P**; Demanded: points' locations in space.

Let us assume that the calibration matrix **K** is known, a camera matrix as $\mathbf{P} = \mathbf{K}[\mathbf{R}|\mathbf{t}]$; and a point in the image $\mathbf{x} = \mathbf{PX}$. Let a pair of camera matrices $\mathbf{P} = [\mathbf{I}|\mathbf{0}]$ and $\mathbf{P}' = [\mathbf{R}|\mathbf{t}]$. Then the essential matrix **E** can be defined [4, 7, 13],

$$\mathbf{E} = [\mathbf{t}]_x \mathbf{R}, \tag{1}$$

where $[\mathbf{t}]_x$ denotes skew symmetric matrix, and the equation for essential matrix,

$$\mathbf{x}_2^T \mathbf{E} \mathbf{x}_1 = \mathbf{0}, \tag{2}$$

where $\mathbf{x}_1$ and $\mathbf{x}_2$ are corresponding points in two image planes.

The essential matrix merges the epipolar constraint and the extrinsic parameter of stereo system. Properties of the matrix **E**:

- **E** is rank 2. Its (right and left) null spaces are the two epipoles,

- Equation (2) is linear and homogeneous in **E** ,

- **E** can be recovered up to scale using 8 points.

The developed algorithm [4, 7, 13] is applied for recovery of unknown camera displacement $(\mathbf{R}, \mathbf{t})$ and 3D scene structure. The eight-point algorithm is employed to estimate motion parameters and then the Direct Linear Algorithm can be used for 3D structure determination.

Given at least 8 pairs of corresponding image points (above 12 in practice): $\{x_1^j, x_2^j\}$, $j = 1, 2, \ldots, n$, $n \geq 8$

1. Estimate essential matrix **E**
2. Decompose the essential matrix **E** (projection onto essential space)
3. Recover the motion parameters $(\mathbf{R}, \mathbf{t})$ from matrix **E** (four solutions)
4. Impose positive depth constraint to obtain right solution
5. Recover 3D structure via triangulation (direct linear transformation)
6. Determine unknown depth up to a single universal scale (the linear least-squares estimate)

## 2.2. Planar homography

In order to correctly determine parameters **R** and **t** by means of eight-point algorithm the points of the examined object must take any given location in space. If those points form degenerated configurations, the algorithm does not bring unambiguous solution. One of such configurations, often encountered in practical applications, is locating all measurement points on one plane (i.e. selected patterns of a given shape on planar measurement marker e.g. ten white circles on a black background of a planar marker put on an examined structure) In such a case the system parameters **R** and **t** are derived from homography matrix **H**.

The planar homography is a non-singular, linear mapping between two corresponding planar points in two views (Fig. 2).

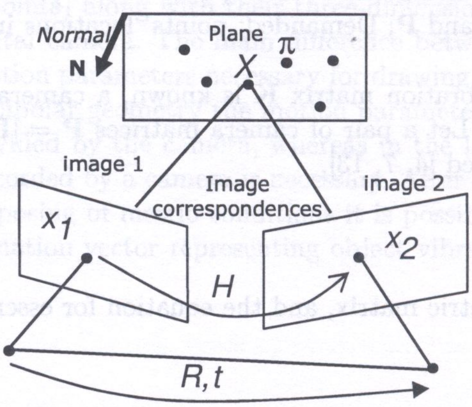



**Fig. 2.** Homography matrix describing relations between projections of points lying on one plane with normal vector **N**, onto image planes of cameras 1 and 2

The **planar homography matrix** is defined as [7]

$$\mathbf{H} = (\mathbf{R} + \frac{\mathbf{T}}{d}\mathbf{N}^T) \tag{3}$$

where $\mathbf{R}, \mathbf{t}$ – motion parameters; $\mathbf{N}$ – unitary normal vector of the plane $\pi$ with respect to the first camera frame; d- distance form the plane $\pi$ to the optical center of the first camera.

The homography matrix describes the relationship between $\mathbf{X_1}$ and $\mathbf{X_2}$ ($\mathbf{X_2} = \mathbf{R}\mathbf{X_1} + \mathbf{t}$).

Homography matrix contains information on motion parameters of the examined object $\{\mathbf{R}, \mathbf{t}\}$ – describing relative location and orientation of the system of cameras, which provided images 1 and 2, as well as on its 3D structure. $\{\mathbf{N}, d\}$ – location of a plane on which the measurement points are laid out, in relation to the coordinate system connected with the first camera. The homography matrix also specifies, within scale accuracy, transformation between corresponding points on images 1 and 2.

$$x_2 \sim \mathbf{H}\, x_1\,. \tag{4}$$

The last equation (4) is called planar homography mapping induced by plane $\pi$.

Four-point algorithm was developed to estimate the homography matrix H and then motion parameters $\mathbf{R}, \mathbf{t}$ between two frames of cameras and structure: $\mathbf{N}$ and $d$.

Given at least 4 pair of corresponding image points (in practice above 10) $\{x_1^j, x_2^j\}$, $j = 1, 2, \ldots, n$, $n \geq 4$

1. Estimate first approximation of the homography matrix $\mathbf{H}$
2. Normalize matrix $\mathbf{H}$
3. Decompose the matrix $\mathbf{H}$ and recover the motion parameters $(\mathbf{R}, \mathbf{t})$ and structure $\mathbf{N}, d$ (four solution)
4. Impose positive depth constraint to obtain right solution
5. Recover 3D structure via triangulation
6. Determine unknown depth up to a single universal scale

## 2.3. Structure from motion (SFM)

In order to obtain 3D structure of an object and vibration measurements in selected measurement points, algorithms were drawn up based on 'structure from motion' method. "Structure from motion" method enables recovery of scene geometry and camera motion tracks based upon the obtained sequence of images. Technique of matrix decomposition is applied to the determination of scene geometry.

A factorization method belongs to rare methods of "structure from motion".

The assumptions originally made by Kanade [10],[11] are summarized below:

a) the camera model is orthographic;

b) the position of $n$ image points $(u_{fp}, v_{fp})$ have been tracked in $F$ frames $(F \geq 3)$; $n$ image points corresponding to scene points $P$, not all coplanar.

The problem can be the statement: given $(u_{fp}, v_{fp})$, the positions of $n$ image points that have been tracked in $F$ frames $1 \leq f \leq F$, $1 < p \leq n$, to compute the motion of the camera from one frame to another (Fig. 3).

The aim is to track $(u_{fp}, v_{fp})$ in $f$ frames for $p$ points. Feature tracking method was based on the tracker described in the dissertation [3, 6, 13].

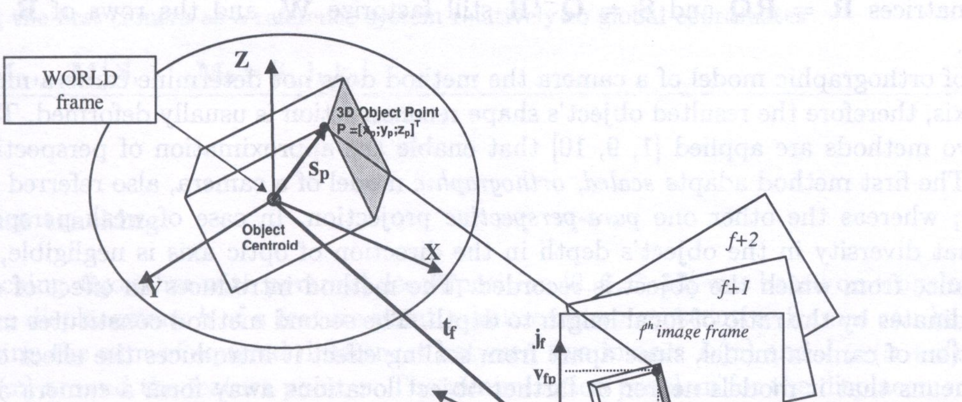



**Fig. 3.** Reference system; $\mathbf{s}_p$ – location of point $P$ feature in global coordinate system, of which the beginning lies in the center of mass of all the points of scene object; $\mathbf{R}$ – matrix of camera orientation for frame $f$ is represented by three versors of axes $\mathbf{i}_f$, $\mathbf{j}_f$, $\mathbf{k}_f$, ($\mathbf{i}_f$ – corresponds to $x$ axis on an image plane and $\mathbf{j}_f$ – $y$ axis); vectors $\mathbf{i}_f$, $\mathbf{j}_f$, are assembled for every frame $F$ as motion matrix $\mathbf{R}$; Coordinate system is centered on the object – its beginning is located in the center of masses of scene points $P_1 \cdots P_n$

After subtract out mean 2D position the measurement equations are:

$$u_{fp} = \mathbf{i}_f^T \mathbf{s}_p\,, \qquad v_{fp} = \mathbf{j}_f^T \mathbf{s}_p\,, \tag{5}$$

where $\mathbf{i}_f$ – rotation, $\mathbf{s}_p$ – position. By connecting them in the form of measurement matrix

$$\tilde{\mathbf{W}} = \mathbf{RS} \tag{6}$$

where: $\mathbf{R} = (\mathbf{i}_1, \ldots, \mathbf{i}_F, \mathbf{j}_1, \ldots, \mathbf{j}_F)^T$ – rotation matrix; $\mathbf{S} = (\mathbf{s}_1, \ldots, \mathbf{s}_P)$ – shape matrix (in a coordinate system attached to the object centroid). The size of measurement matrix is the following: $\tilde{\mathbf{W}} = \mathbf{R}_{2F\times 3}\mathbf{S}_{3\times P}$.

When we employ factorization method, the algorithm is based on SVD:

a) compute

$$\tilde{\mathbf{W}} = \mathbf{U\Lambda V} \tag{7}$$

in which $\mathbf{\Lambda}$ must be rank 3

b) when noise is present, define the adjusted matrices

$$\mathbf{\Lambda}' = \mathbf{\Lambda}(1:3, 1:3), \qquad \mathbf{U}' = \mathbf{U}(:, 1:3), \qquad \mathbf{V}' = \mathbf{V}(:, 1:3), \tag{8}$$

and construct

$$\hat{\mathbf{R}} = \mathbf{U}'\sqrt{\mathbf{\Lambda}'}; \hat{\mathbf{S}} = \sqrt{\mathbf{\Lambda}'}\mathbf{V}'^T. \tag{9}$$

The factorization of $\tilde{\mathbf{W}}$ is straightforward via SVD but is not unique,

$$\mathbf{W} = \mathbf{U\Lambda V}\,, \tag{10}$$

$$\mathbf{W} = \hat{\mathbf{R}}\hat{\mathbf{S}} = \hat{\mathbf{R}}(\mathbf{QQ}^{-1})\hat{\mathbf{S}} = (\hat{\mathbf{R}}\mathbf{Q})(\mathbf{Q}^{-1}\hat{\mathbf{S}}). \tag{11}$$

Fortunately, two constrains can be added:

- the 3D vectors forming the rows of $\mathbf{R}$ must have unit norm
- in $\mathbf{R}$, the $\mathbf{i}_i^T$ must be orthogonal to the corresponding $\mathbf{j}_i^T$.

The rows of the matrix $\mathbf{R}$ will not satisfy the constraints mentioned above but if look for a (correction) matrix $\mathbf{Q}$ such that

$$|\mathbf{m}_f|^2 = \mathbf{i}_i^T\mathbf{QQi}_i^T = 1; \qquad |\mathbf{n}_f|^2 = \mathbf{j}_i^T\mathbf{QQj}_i^T = 1; \qquad \mathbf{m}_f\mathbf{n}_f = \mathbf{i}_i^T\mathbf{QQj}_i^T = 0.$$

The new matrices $\mathbf{R} = \hat{\mathbf{R}}\mathbf{Q}$ and $\mathbf{S} = \mathbf{Q}^{-1}\mathbf{R}$ still factorize $\tilde{\mathbf{W}}$, and the rows of $\mathbf{R}$ satisfy the constraints.

In case of orthographic model of a camera the method does not determine camera motion along the optic axis, therefore the resulted object's shape reconstruction is usually deformed. To avert the problem two methods are applied [1, 9, 10] that enable the approximation of perspective camera character. The first method adapts *scaled, orthographic* model of a camera, also referred to as weak perspective; whereas the other one *para-perspective* projection. In case of weak perspective it is assumed that diversity in the object's depth in the direction of optic axis is negligible, compared to the distance from which the object is recorded. The method introduces an effect of scaling the image coordinates by the ratio of focal length to depth. The second method constitutes much better approximation of camera model, since apart from scaling effect it introduces the effect of location. This also means that it models nearer or farther object locations away form a camera as an effect of observation at different angles. The two methods impose other metric limitations on matrix $\mathbf{Q}$. The para-perspective model was selected to estimate amplitude of vibration based on works [9] and conducted research.

Let $x_f$, $y_f$, and $z_f$ designate relative camera locations. The para-perspective model of camera can be distinguished with the following limitations (Table 1) where $z_f$ – depth to the object's center of mass; $x_f$, $y_f$ – components of translation between the origins of the camera and global coordinate system; $f$ – index indicating $f$th frame.

**Table 1.** The metric limitations for para-perspective model

| | |
|---|---|
| Para-perspective | $\lvert m_f\rvert^2/(1+x_f^2) = \lvert n_f\rvert^2/(1+y_f^2) = 1/z_f^2$<br>$m_f \cdot n_f = 0$<br>$\lvert m_1\rvert^2 = 1$ |

The next step of the before-mentioned methods is, based on metric limitations presented in Table 1, determining matrix **Q**. In this research the correction matrix **Q** was determined by means of Newton–Raphson method.

According to papers [1, 9, 10] relations enabling calculating the desired vector $\mathbf{t}_f$, representing the amplitude of scene object vibrations for para-perspective camera model, is enclosed in Table 2.

**Table 2.** Relation describing *tf* vector representing vibration amplitude in analyzed objects

| | | | |
|---|---|---|---|
| Para-perspective model | $t_f = \begin{bmatrix} i_f \\ j_f \\ k_f \end{bmatrix}$ | $\begin{bmatrix} z_f x_f \\ z_f y_f \\ z_f \end{bmatrix}$ | $z_f = \sqrt{\frac{1}{2}\left(\frac{1+x_f^2}{\lvert m_f\rvert^2} + \frac{1+y_f^2}{\lvert n_f\rvert^2}\right)}$ |

Based on the specified $\mathbf{t}_f$ vector it is possible to obtain the camera location for each image frame (at every time moment). This vector describes translations (Fig. 3) between the beginning of coordinate system of the camera and the beginning of global coordinate system (gravity center of an object), and represents amplitude of object vibrations at every moment.

The following algorithm was developed and implemented for para-perspective model in accordance with metric limitations spelled out in Table 1:

1. Calculating: SVD of $\tilde{\mathbf{W}} = \mathbf{UDV}$
2. Specification of structure and orientation: $\hat{\mathbf{R}} = \mathbf{U}'\sqrt{\mathbf{D}'}$, $\hat{\mathbf{S}} = \sqrt{\mathbf{D}'}\mathbf{V}'^T$,
3. Determination of correction matrix **Q** by imposition of metric limitations. Application of Newton–Raphson method.
4. Determination of motion matrix **M** and structure **S** as $\mathbf{M} = \hat{\mathbf{R}}\mathbf{Q}$, $\mathbf{S} = \mathbf{Q}^{-1}\hat{\mathbf{S}}$
5. Compute the translations for each frame $\mathbf{t}_f$ and depth $\mathbf{z}_f$
6. Setting the first camera as a reference system relatively to global coordinates
   $\mathbf{MM}_0$, $\mathbf{M}_0^T\mathbf{S}$, $\mathbf{M}_0 = [\mathbf{i}_1\ \mathbf{j}_1\ \mathbf{k}_1]$

## 2.4. Feature tracking

Feature tracking algorithm with pyramid decomposition [3, 6, 7, 13] was based on translation model in which the displacement $h$ of a feature point $x$ between two consecutive frames can be calculated by minimizing the sum of squared differences between two images $I_i(x)$ and $I_{i+1}(x+d)$ in a small window $W(x)$ around the feature point $x$. The minimization problem for the displacement $d$ can be described as follows,

$$\min_h E(h) = \sum_{\tilde{x}\in W(x)} [I_2(\tilde{x}+h) - I_1(\tilde{x})]^2 . \tag{12}$$

The closed-form solution is given,

$$\mathbf{h} = -\mathbf{G}^{-1}\mathbf{b}, \tag{13}$$

where

$$G = \begin{bmatrix} \sum w(x) I_x^2 & \sum w(x) I_x I_y \\ \sum w(x) I_x I_y & \sum w(x) I_y^2 \end{bmatrix}, \qquad b = \begin{bmatrix} \sum w(x) I_x I_t \\ \sum w(x) I_y I_t \end{bmatrix} \tag{14}$$

and where $I_x$, $I_y$ – are image gradients, and $I_t = I_2 - I_1$ – is a temporal image derivative.

## 3. SIMULATIONS AND EXPERIMENTS

### 3.1. Simulations tests

The developed methods of epipolar geometry and "structure form motion" with various camera models were the subjects of tests carried out by means of series of simulations. Many simulations were carried out on different spatial and planar objects For example, for the method of "structure from motion" and the method estimating major matrix **E**, the simulation object was a cube (Fig. 4a), whereas in the method determining planar homography matrix **H** a flat element was implemented in the form of a square (Fig. 4b). Each of the analyzed objects (spatial and planar) represented measurement point on the examined structure.This article assumes that the vibration amplitude will be determined for every such object. By means of sinusoidal excitation and noise signal applied on its amplitude and frequency, movement (vibrations) of the object was simulated in the directions of all the axes. Simulation data were provided for camera calibrated internally and externally. The error values for epipolar geometry methods were determined by the relation

$$e_{E,H} = \sum (X_{Ti} - X_i)^2;$$

where $X_{Ti}$ – assigned value of amplitude of object vibrations at the moment $i$; $X_i$ – value of amplitude of object vibrations at the moment $i$, derived from reconstruction algorithm.

For "structure from motion" the error was specified as

$$e_{sfm} = \|R * (R^T * R)^{-1} * R^T * W - W\|^2$$

where **R** – camera motion matrix (determining camera orientation in each sequence frame); **W** – measurement matrix, determined after subtraction of the vector designating gravity center of points on the object.

Assumed parameters for sinusoidal excitation signal for each method are assembled in Table 3.

The conducted research proves that all the developed methods, both for sinusoidal excitation signal and for noise signal have reconstructed accurately all the assigned amplitudes of object vibrations along all the axes of global coordinate system, which is illustrated in Figs. 5–7 In case of "structure

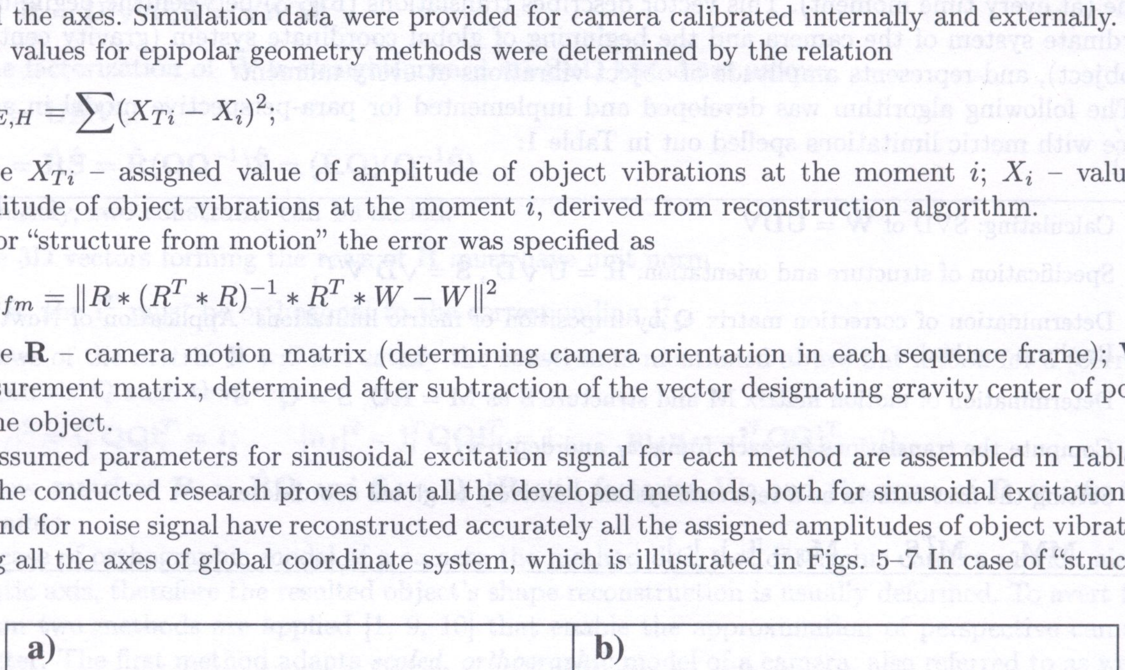



**Fig. 4.** Examined (objects) markers representing measurement points: a) (three-dimensional) cube containing 7 points-features; b) planar element in the shape of a square – 9 features

**Table 3.** Parameters of sinusoidal excitation signals used for testing developed methods

| Method: epipolar geometry – matrix E | | Method: Structure from motion | | Method: epipolar geometry – (Homography H) | |
|---|---|---|---|---|---|
| Ax [mm] – Amplitude in X axis | 8 | Ax [mm] – Amplitude in X axis | 8 | Ax [mm] –Amplitude in X axis | 5 |
| Ay [mm] – Amplitude in Y axis | 4 | Ay [mm] – Amplitude in Y axis | 4 | Ay [mm] – Amplitude in Y axis | 8 |
| Az [mm] – Amplitude in Z axis | 10 | Az [mm] – Amplitude in Z axis | 10 | Az [mm] – Amplitude in Z axis | 10 |
| fx [Hz] – Frequency in X axis | 20 | fx [Hz] – Frequency in X axis | 20 | fx [Hz] – Frequency in X axis | 10 |
| fy [Hz] – Frequency in Y axis | 25 | fy [Hz] – Frequency in Y axis | 25 | fy [Hz] – Frequency in Y axis | 5 |
| fz [Hz] – Frequency in Z axis | 5 | fz [Hz] – Frequency in Z axis | 5 | fz [Hz] – Frequency in Z axis | 10 |

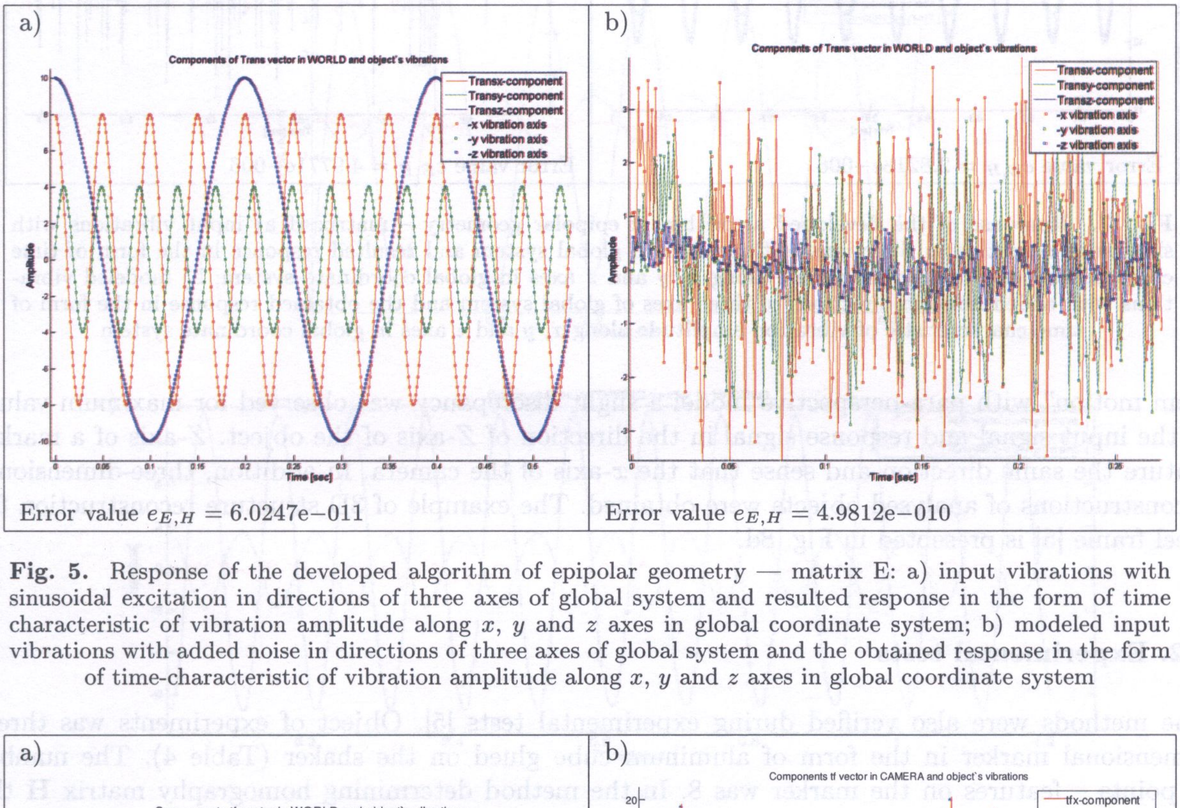



**Fig. 5.** Response of the developed algorithm of epipolar geometry — matrix E: a) input vibrations with sinusoidal excitation in directions of three axes of global system and resulted response in the form of time characteristic of vibration amplitude along $x$, $y$ and $z$ axes in global coordinate system; b) modeled input vibrations with added noise in directions of three axes of global system and the obtained response in the form of time-characteristic of vibration amplitude along $x$, $y$ and $z$ axes in global coordinate system

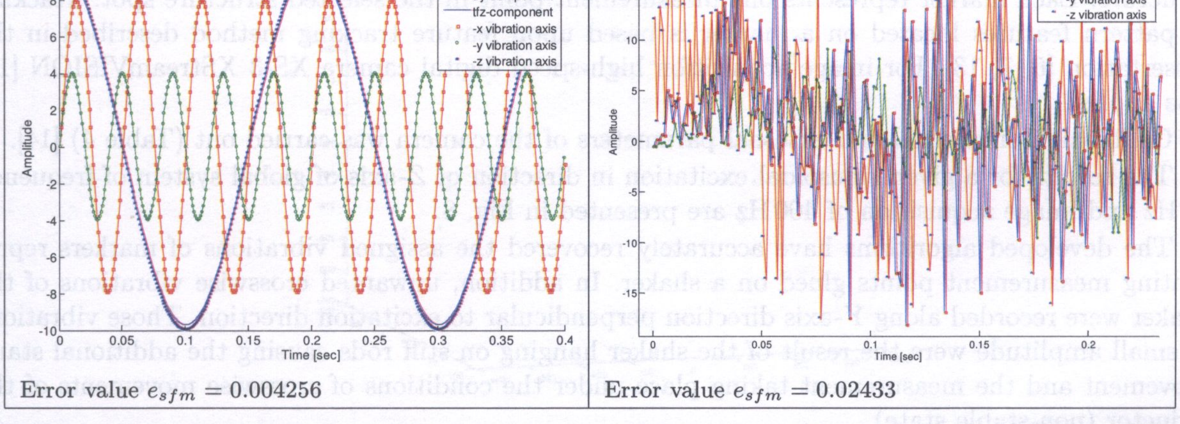
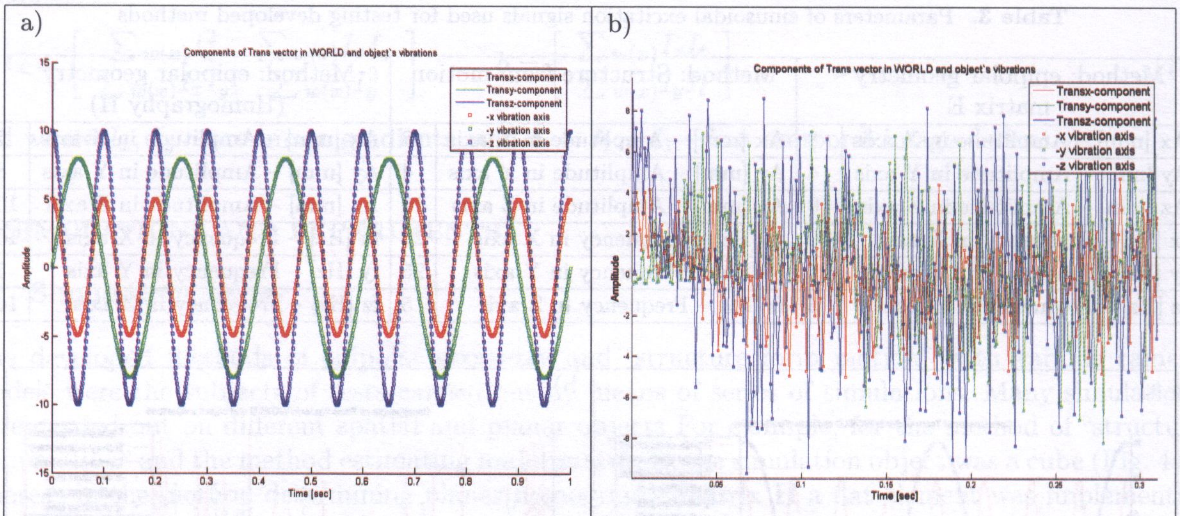



**Fig. 6.** Response of the algorithm with para-perspective model; a) input vibrations with sinusoidal excitation in directions of three axes of global system and the output response in the form of three components of translation vector $tf$ representing the amplitude of object vibrations; b) model vibrations with superimposed noise in directions of three axes of global system and resulted response in the form of three vector $tf$ components representing object vibrations




Error value $e_{E,H} = 2.8218\text{e}{-006}$ Error value $e_{E,H} = 4.9771\text{e}{-006}$

**Fig. 7.** Response of the developed algorithm of epipolar geometry – matrix H a) input vibrations with sinusoidal excitation in directions of three axes of global system and resulted response in the form of time characteristic of vibration amplitude along $x$, $y$ and $z$ axes in global coordinate system; b) modeled vibrations with added noise in directions of three axes of global system and the obtained response in the form of time-characteristic of vibration amplitude along $x$, $y$ and $z$ axes in global coordinate system

from motion" with para-perspective model a slight discrepancy was observed for maximum values of the input signal and response signal in the direction of $Z$-axis of the object. $Z$-axis of a marker feature the same direction and sense that the $x$-axis of the camera. In addition, three-dimensional reconstructions of analysed objects were obtained. The example of 3D structure reconstruction for steel frame [5] is presented in Fig. 8d.

## 3.2. Experimental tests

The methods were also verified during experimental tests [5]. Object of experiments was three-dimensional marker in the form of aluminum cube glued on the shaker (Table 4). The number of points – features on the marker was 8. In the method determining homography matrix **H** the pattern points were located only on one plane. The marker can be put on any place of the analyzed structure. Each marker represents one measurement point in the selected structure spot. Tracking of pattern features located on a marker is based upon feature tracking method described in the dissertation [6, 7, 13]. For image acquisition high-speed digital camera XS-3 XStreamVISION [15] was applied.

Calibration of internal and external parameters of the camera was carried out (Table 4) [14].

The results for a given sinusoidal excitation in direction of $Z$-axis of global system of frequency 10 Hz and image acquisition of 400 Hz are presented in Fig. 8.

The developed algorithms have accurately recovered the assigned vibrations of markers representing measurement points glued on a shaker. In addition, unwanted crosswise vibrations of the shaker were recorded along $Y$-axis direction perpendicular to excitation direction. Those vibrations of small amplitude were the result of the shaker hanging on stiff rods causing the additional stand movement and the measurement taking place under the conditions of crosswise movements of the inductor (non-stable state)

Comparing the response of the methods proves that smooth vibration characteristics were yielded by methods based on homography matrix H and on "structure from motion". As for values of vibration amplitudes conforming results were obtained by means of both methods of epipolar geometry (matrices E and H). In this respect the method of "structure from motion " requires proper scaling.

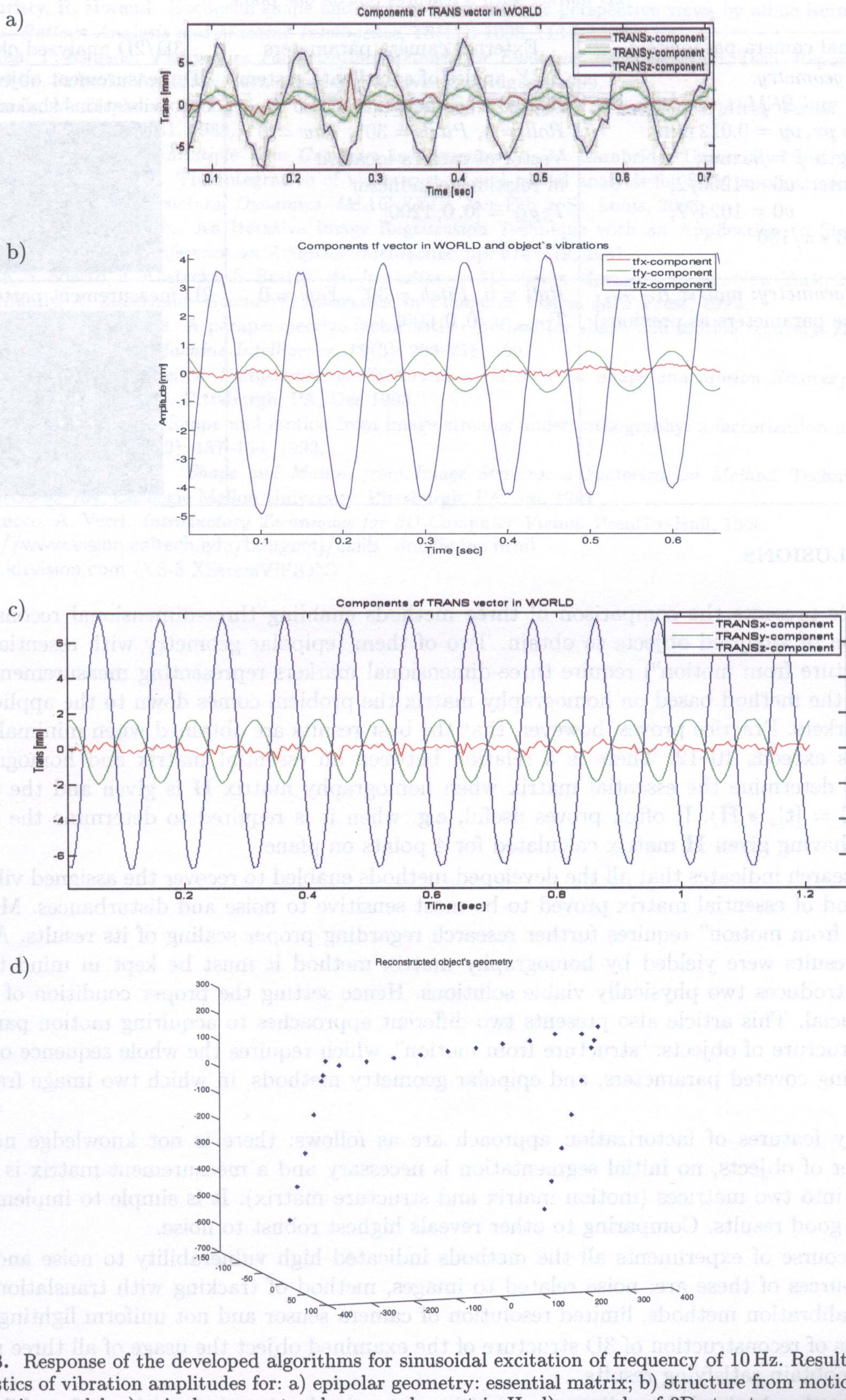
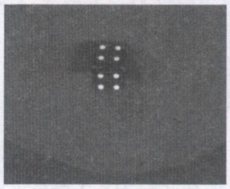
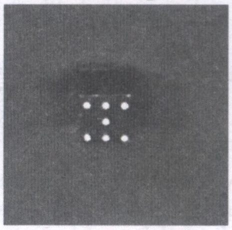



**Fig. 8.** **Response of the developed algorithms for sinusoidal excitation of frequency of 10 Hz. Resulted characteristics of vibration amplitudes for: a) epipolar geometry: essential matrix; b) structure from motion: para-perspective model; c) epipolar geometry: homography matrix H; d) example of 3D structure reconstruction for steel frame [5] as a result of SFM with para-perspective model**

**Table 4.** Internal and external and examination objects: 3D cube with markers put onto two surfaces and 2D marker with 9 features put on one plane

| Internal camera parameters | External camera parameters | 3D/2D analyzed objects |
|---|---|---|
| *Epipolar geometry: matrix E and SFM*<br>Pixel size $px, py = 0.012$ mm<br>Focal length $f = 55$ mm;<br>Image center: $u0 = 1260/2$;<br>$v0 = 1024/2$;<br>$skew = 90 * \pi/180$ | *RPY* angles of coordinate system related to camera:<br>$Roll = 0$, $Pitch = 30°$, $Yaw = 0$<br>Vector of camera location in relation to inductor<br>$T_{SFM} = [0, 0, 1200]$ | 3D measurement object glued onto vibrations shaker<br> |
| *Epipolar geometry: matrix H*<br>(The same parameters as previous) | $Roll = 0$, $Pitch = 30°$, $Yaw = 0$<br>$T_{Hom} = [0, 0, 600]$ | 2D measurement pattern<br> |

## 4. CONCLUSIONS

This article presents the comparison of three methods enabling three-dimensional reconstruction and motion of analyzed objects to obtain. Two of them (epipolar geometry with essential matrix and "structure from motion") require three-dimensional markers representing measurement points. In case of the method based on homography matrix the problem comes down to the application of planar markers. Practice proves, however, that the best results are obtained when minimal number of markers exceeds 10–12. There is a relation between an essential matrix and homography. It enables to determine the essential matrix when homography matrix **H** is given and the other ay around ($\mathbf{E} = [\mathbf{t}]_x * \mathbf{H}$). It often proves useful, e.g. when it is required to determine the essential matrix **E** having given **H** matrix calculated for 4 points on plane.

The research indicates that all the developed methods enabled to recover the assigned vibrations. The method of essential matrix proved to be most sensitive to noise and disturbances. Method of "structure from motion" requires further research regarding proper scaling of its results. Although the best results were yielded by homography matrix method it must be kept in mind that this method introduces two physically viable solutions. Hence setting the proper condition of solution is very crucial. This article also presents two different approaches to acquiring motion parameters and 3D structure of objects: "structure from motion", which requires the whole sequence of images for providing coveted parameters, and epipolar geometry methods, in which two image frames are analyzed.

The key features of factorization approach are as follows: there is not knowledge needed of the number of objects, no initial segmentation is necessary and a measurement matrix is globally factorized into two matrices (motion matrix and structure matrix). It is simple to implement and gives very good results. Comparing to other reveals highest robust to noise.

In the course of experiments all the methods indicated high vulnerability to noise and disturbances. Sources of these are: noise related to images, method of tracking with translation model, errors of calibration methods, limited resolution of camera sensor and not uniform lighting.

In terms of reconstruction of 3D structure of the examined object the usage of all three methods enabled to obtain satisfying results.

Based on the obtained three-dimensional vision data, the developed algorithms of modal analysis will be extended, including energetic method for detection and localization of damages appearing in the structure [5].

## REFERENCES


[1] S. Christy, R. Horaud. Euclidean shape and motion from multiple perspective views by affine iteration. *IEEE Trans. Pattern Analysis and Machine Intelligence*, **18**(11): 1098–1104, 1996.

[2] M. Han, T. Kanade. *Perspective Factorization Methods for Euclidean Reconstruction*. Tech. Report CMU-RI-TR-99-22, Robotics Institute, Carnegie Mellon University, August, 1999.

[3] C. Harris, M. Stephens. A combined corner and edge detector. *Proc. of The Fourth Alvey Vision Conference*, Manchester, pp. 147–151. 1988.

[4] Hartley, A. Zisserman. *Multiple View Geometry in Computer Vision*. Cambridge University Press, 2004.

[5] P. Kohut, P. Kurowski. The integration of vision system and modal analysis for SHM application. *A Conference&Exposition on Structural Dynamics, IMAC XXIV*, Jan–Feb 2, St. Louis, 2006.

[6] B.D. Lucas, T. Kanade. An Iterative Image Registration Technique with an Application to Stereo Vision. *International Joint Conference on Artificial Intelligence*, pp. 674–679, 1981.

[7] Y. Ma, S. Soatto, J. Kostecka, S. Sastry. *An Invitation to 3D Vision*. Springer-Verlag, New York, 2004.

[8] J.L. Mundy, A. Zisserman. *Geometric Invariance in Computer Vision*. MIT Press, 1992.

[9] C. Poelman, T. Kanade. A paraperspective factorization method for shape and motion recovery. *IEEE Trans. Pattern Analysis and Machine Intelligence*, **19**(3): 206–218, 1997.

[10] C. Poelman, T. Kanade. *A Paraperspective Factorization Method for Shape and Motion Recovery*. Technical Report CMU-CS-93-219, Pittsburgh, PA, Dec 1993.

[11] C. Tomasi, T. Kanade. Shape and motion from image streams under orthography: a factorization method. *Int. J. Computer Vision*, **9**(2): 137–154, 1992.

[12] C. Tomasi, T. Kanade. *Shape and Motion from Image Streams: a Factorization Method*. Technical Report CMU-CS-92-104, Carnegie Mellon University, Pittsburgh, PA, Jan 1991.

[13] E. Trucco, A. Verri. *Introductory Techniques for 3D Computer Vision*. Prentice-Hall, 1998.

[14] http://www.vision.caltech.edu/bouguetj/calib_doc/index.html

[15] www.idtvision.com (XS-3 XStremVISION)