**ISSN**: 1231-4005
**e-ISSN:** 2354-0133
**DOI:** 10.5604/01.3001.0010.2814

# VISION BASED NAVIGATION FOR SATELLITE DOCKING

**Mariusz Jacewicz, Robert Głębocki**

*Warsaw University of Technology, Faculty of Power and Aeronautical Engineering*
*Nowowiejska Street 24, 00-665 Warsaw, Poland*
*tel.:+48 022 234 7184, +48 022 234 5933*
*e-mail: mjacewicz@meil.pw.edu.pl, rgleb@meil.pw.edu.pl*

***Abstract***

*This article deals with the situation of a space debris or not working satellite in an unidentified pose with respect to the master satellite. Feature based monocular pose estimation vision system was presented. The results of numerical simulation were described. The results of implementation and testing of simulation intended for vision-based navigation applications such as rendezvous of satellites and formation flying is shown. In this document markerless local features based navigation system has been studied. The proposed vision navigation system satellites are able to determine the position and orientation of a target in relation to the coordinate system of the camera. It relates from the time when the satellite is visible as a small object until docking with the chaser. A modified algorithm soft Position Iterations was used to estimate the pose of the target. Visual navigation system uses a single camera. The impact of changes in illumination of the object was analysed. In order to reproduce the space conditions the laboratory stand was built. The developed method was tested experimentally for different scenarios approach satellites to each other. Comparing the ground truth position and orientation and the results obtained with the aim of vision navigation system it is worth nothing to observe accuracy of the developed method. Achieved satisfactory performance of the algorithm.*

***Keywords:*** *transport, satellite, vision navigation, autonomous docking*

## 1. Introduction

This article refers to the results of experiments proposed for vision-based rendezvous of satellites. In this document monocular navigation system has been developed. It was done under project conducted by Warsaw University of Technology. A servicing satellite is sending to capture target object and execute servicing tasks [1]. Described method focus on final phases of rendezvous. It was assumed an image of an object taken by a calibrated camera in each step of time is known, and was assumed a 3D representation of an object is recognized. It was proposed a solution for tracking 3D rigid objects that is based on local features and promises improved computational performance. Proposed method is robust on the tracking failures. Point-to-point correspondences are used to estimate the pose of the target. The target is passive but the inspection vehicle has knowledge of the target geometry. The target object cannot assist in a process of rendezvous or docking. It does not have actuators or its actuators are deactivated (Fig. 1).

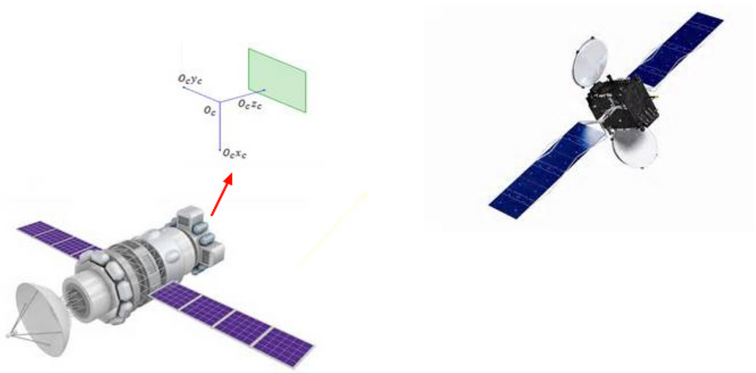

*Fig. 1. Satellite Rendezvous and Docking with pose estimation problem*

It does not have visual markers. The task of visual tracking is present in different fields, of which a few ones will now be briefly summarized. Autonomous Rendezvous and Docking is receiving attention from the space community. With TriDAR, a solution of the pose calculation problem is offered, combining the LiDAR [2] approach with triangulation. It joins the long-range capabilities of a LiDAR sensor and the precision of a triangulation approach in the near range. The resulting sensor system is unfortunately expensive when related to camera-based systems. Good idea is the use of a scanning LiDAR sensor for estimating the pose of the target. The target is detected and range is estimated. A 3D model is fitted, after which the full pose is obtainable [3]. In the Orbital Express mission, the Automated Video Guidance Sensor is primarily intended for the vehicle [4]. Many spacecrafts use laser-based sensors and lasers are used to illuminate the object. This is an inefficient solution, which requires mirrors at the target. Often, markers must be present on the surface of the object [5]. Monocular vision is rarely used for 6-DoF pose estimation, because of accuracy problems. Stereo vision is seldom proposed in this situation but LiDAR sensors are dominating in this field [2]. The possibility of directly measuring the distance is important, so the method proposed in this work looks promising. It may give an alternative to close-range laser sensors. Rendezvous phases are defined as phasing, far range (10-100 km), close range, and final approach (100-500 m to contact). Different sensor types are currently chosen according to the operation range and performance. Radar or ground based navigation is reasonable for long-range rendezvous phases. On the other hand, only a visual sensor type will be suitable for navigation during close range rendezvous phases with non-cooperating spacecraft. In fact, either a camera type sensor or a laser range finder sensors dot not requires a cooperating target. Relative navigation using laser range finder would be interesting since distance should be small between the satellites. It is an alternative if combined with a camera type sensor. In this case, the laser range finder can be additionally used after identification of the target during the last phase of the Rendezvous procedure [6]. The laser range finder can be employed in proximity operation phases. The visual system task is used to identify the relative position and orientation of the client satellite. The required hardware shall be available even for small spacecrafts with a small power budget. Some advantages of a camera sensor are listed: cost benefits compared to an active sensor, no need of a cooperating target and antennas attached on target, the measurement accuracy increase with decreasing range, because of increasing resolution, the measurement of pose parameters can be obtained in single step procedure, camera sensor has no moving parts and is then less sensitive to orbital environment [6].

## 2. Problem formulation

Pose estimation is a problem, which has its origin in photogrammetry, and it is known as space resection [7]. Orientation of a camera with a set of n 2D-to-3D point correspondences is a well-studied problem. The problem of defining the absolute position and orientation of a camera is known as the Perspective-n-Point (PnP) problem. The minimal number of correspondences to solve PnP is three. Fischler found that P4P problem with non-coplanar points had many solutions and with coplanar points had only one solution [8]. For P5P problem, there were two solutions. For more than 6 correspondences, it became Direct Linear Transformation (DLT) problem.

In this section, it was assumed that a 3D model of the spacecraft is obtainable or can be estimated on-line. The pose should be estimated knowing the correspondences between 2D measurements in the images and 3D features of the model. It will be consider here that these features are 3D points and their 2D projections in the image. Let us denote $F_c$ − the camera frame and $\boldsymbol{T}_w^c$ the transformation that defines the position of $F_w$ with reference to $F_c$. $\boldsymbol{T}_w^c$ is a matrix:

$$\boldsymbol{T}_w^c = \begin{bmatrix} \boldsymbol{R}_w^c & \boldsymbol{t}_w^c \\ \boldsymbol{0}_{3x1} & \boldsymbol{1} \end{bmatrix}, \tag{1}$$

where $\boldsymbol{R}_w^c$ and $\boldsymbol{t}_w^c$ are the rotation matrix and translation vector that define the position of the camera in the world frame.

The perspective projection $\boldsymbol{x} = (u, v, 1)^T$ of a point $\boldsymbol{X} = (X, Y, Z, 1)^T$ will be given by:

$$\boldsymbol{x} = \boldsymbol{K}\boldsymbol{\Pi}\boldsymbol{T}_w^c\boldsymbol{X}, \tag{2}$$

where $\boldsymbol{x}$ are the coordinates, expressed in pixel, of the point in the image; $\boldsymbol{K}$ is the camera intrinsic parameters matrix and is defined by:

$$\boldsymbol{K} = \begin{bmatrix} \boldsymbol{p_x} & \boldsymbol{0} & \boldsymbol{u_0} \\ \boldsymbol{0} & \boldsymbol{p_y} & \boldsymbol{v_0} \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{1} \end{bmatrix}, \tag{3}$$

where $(u_0, v_0, 1)$ are the coordinates of the principal point (the intersection of the optical axes with the image plane) and $p_x$ (resp $p_y$) is the ratio between the focal length of the lens $f$ and the size of the pixel $l_x$: $p_x = f/l_x$ ($l_y$ is the height of a pixel, $p_y = f/l_y$). $\boldsymbol{\Pi}$ the projection matrix is given, in the case of a perspective projection model, by:

$$\boldsymbol{\Pi} = \begin{bmatrix} \boldsymbol{1} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{1} & \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{1} & \boldsymbol{0} \end{bmatrix}. \tag{4}$$

The intrinsic parameters were obtained through an off-line calibration stage. It is consider here a perspective projection model. It will be consider that the camera is calibrated and that the coordinates are expressed in the normalized space. If someone has $N$ points $\boldsymbol{X}_i, i = 1 \dots N$ whose coordinates expressed in $F_w$ are given by $\boldsymbol{X}_i = (X_i, Y_i, Z_i, 1)^T$, the projection $\boldsymbol{x}_i = (x_i, y_i, 1)^T$ of these points in the image plane is then given by:

$$\boldsymbol{x_i} = \boldsymbol{\Pi}\boldsymbol{T}_w^c\boldsymbol{X_i}. \tag{5}$$

Knowing 2D-3D point correspondences, $x_i$ and $X_i$, pose estimation consists in solving the system given by the set of equations (5) for $\boldsymbol{T}_c^w$.

## 3. Literature study

In this paragraph, we review approaches allowing solving the set of equations (5) for the pose $\boldsymbol{T}_w^c$. First methodologies to visual tracking were based on tracking of the outer shape of a target. Contour-based trackers gained attention on the early vision based navigation systems. There were proposed algorithms, which were based on fiducial markers on the target object to make simpler the registration task. The 3D positions of the markers in the world coordinate system are assumed to be recognisable at all times. When one or more blobs are missing, these methods are not able to get right results of pose. This method in case of spacecrafts is impractical because many existing ones have not markers [9]. The other approach is based on three-dimensional edge models. In this case, pose, computation is achieved by minimization the space between 3D model edges and the corresponding edge features in the image. The most weak point of method based on 3D model is reliance on full geometric model. Satellites are made with cylindrical, spherical and complex shapes. This makes the system fragile. When the object is complex there are achieved low frame per second rates. To reject outliers algorithms such as Random Sample Consensus (RANSAC) can achieve robustness to illumination conditions in space [10]. Light Detection and Ranging (LIDAR) was used for pose estimation. Camera sensors are cheaper and can provide accurate capabilities to obtain relative pose. The use of the interface circle used to attach the satellite to additional vehicle has been proposed for capturing the satellite. This has a disadvantage because it is limited for proximity operations where the target satellite is visible from the interface side. Capturing should be performed autonomously because there are communication delays between on-orbit systems and earth. Feature matching computer vision approaches have been developed but they are computational intensive and cannot be used during entire mission. Critical sensitivity to illumination and occlusions of spacecraft had been obtained [3]. Learned database is used on Orbital Express and the algorithms are based on edges in this case. Lepetit suggested for tracking objects in 3D by using corner features with a single camera. This approach was robust to camera displacements and partial occlusion [9]. Drawback of this method was camera should be close enough to one of key frames and there is a real problem when then tracking must be initialized after tracking failure. Limitation of described methods lies in their execution time. The existing approaches are computationally expensive [1]. In many cases, system is based on artificial features. This makes tracking far more difficult because finding and following feature points or edges can be hard because there are too few of them on many typical objects. It is much more better to rely on naturally present features, such as edges, corners, or texture [11]. In this work, authors assumed that satellite has been detected. Proposed system is planned to deals with finally stages of satellite rendezvous from far proximity operations, when the satellite is about 2 km from satellite to the contact of satellites. Someone can divide PnP algorithms to non-iterative and iterative [10]. Non-iterative formulate the problem as system of equations, which is solved in a sequence of operations. Iterative approaches search in a parameter space and optimize some objective function, for example distance between reprojected 3D points and measurements. Iterative approaches usually minimize an error function but may fall into local minimum and result in pose ambiguity. Among iterative approaches, Dementhon presented Pos with Iteration (POSIT) algorithm to solve PnP problem for more than four non-coplanar correspondences [6]. Lu introduced iterative algorithm, which minimized a 3D space error and was faster than other ones [12]. Someone can split pose estimation algorithms into minimal, which use the smallest possible set of point correspondences between 2D and 3D space to compute camera pose, and non-minimal, which use more point correspondences to linearize the task or to return the more accurate result. Minimal algorithms are usually used to filter out incorrect correspondences [8]. Once precise correspondences are recognised one can use non-minimal algorithms to improve the final result. To find correct correspondences, it is important to have fast algorithm, which use smallest number of measurements to calculate the camera pose. It is because such an algorithm is implemented inside the RANSAC loop many times and the number of iterations is proportional to the size of the points set. It is possible to expansive solution to more points, but this is not practical since the number of possible triplets grows exponentially with the number of points. It is known that three point correspondences are adequate to recover the camera rotation and translation in the case of calibrated camera. Direct Linear Method is impractical since it produces a rigid transformation in the cases when the image is free of noise. When measurements are noisy, this method does not give the correct solution. Several methods solve the problem analytically when few measurements are given and when the model points are in a specific configuration. The assumption that the matching is known is not acceptable in all pose estimation problems. One of the approaches suggested to solve the matching is to calculate the pose and the matching simultaneous during the iterative process. Pose can be calculated from various kinds of image measurements like a set of 2D projections of 3D points or 3D lines, from a mixture of points and lines, projections of known objects like chessboards, coplanar circles or intersections of lines. The estimation from objects, like lines or circles, may be more precise, but it is needed to solve tasks such as detection, and to compensate the fact that image is affected by distortion, leaving lines curved or circles deformed. The aim is to estimate the pose of a passive object using its known three-dimensional model and single two-dimensional images collected on-board the chaser. Star trackers can be used for increased dynamic range accurately to navigate from zero to several tens of kilometres. The unknown correspondences between image and model points may lead to serious problems and to a large computational load. Someone can formulate requirements for monocular vision systems [13]. Pose estimator should rely on a small number of image features. It should be robust to pose solutions and to image noise. It cannot rely on the uncooperative satellite dynamics. The definition of a good spacecraft model is a central step of the pose estimation strategy. The vehicle model has to be as simple as possible to reduce the system complexity. The estimate of the initial pose is the most difficult task of the pose estimation. Many authors assume a priori knowledge of the relative position to aid the vision navigation system. Using a single monocular image, and utilizing knowledge of the target spacecraft, estimation of the target's relative pose parameters with respect to the camera are found. Sensors like ultrasonic, magnetic, inertial devices, compass, and optical cameras have been considered. Indeed, with respect to other sensors, a camera combined with a display is an appealing configuration. Until the early 2000 s, almost all the vision-based registration techniques relied on markers [14]. Markerless model based tracking techniques improve marker-based methods. The introduction of Simultaneous Localization and Mapping (SLAM), as a result of structure from motion methods, allows to get rid of a template of the spacecraft. Mission as ETS-VII or Orbital Express has used cameras for navigation during close range rendezvous operations. The visual marker was attached on Orihime target satellite to make possible the capture process. The Orbital Express mission has used a visual marker for the last phase [4]. A three-dimensional gold dots against a flat black background was mounted on the NextSat. Visual markers are useful features on client vehicles, particularly those that are designed to be serviced or for which the ability to be repaired is considered important. Many missions cannot rely on special vision markers. However, other well-characterized surface features may be substituted to accomplish the same task. Considering that, most satellites have a V-flange structure that shall be a suitable mechanical interface for capturing and a natural marker for the relative visual navigation system [15]. SUMO is designed to service different types of spacecraft without requiring aids; the visual system identifies fiducial points in order to establish the pose estimation. The representation of the object pose plays an important role in the estimation process. The correspondence problem deals with the question, which object feature belongs to which image feature. Basically, either an iterative softassign algorithm can be used. The ability of the navigation system to deal with different light condition is important over the close-range operation. Lighting conditions in space is a major factor to be considered during visual inspection and can change rapidly and dynamically. The following sources have to be considered according to the importance of their effects: the Sun, reflections of sunlight on target surface, reflections of sensor illuminator light on target surface, direct light or reflections of other light sources. In many cases, the 3D geometry of the spacecraft is available a priori and the pose estimation uses model-based techniques. A short description of model-based techniques can be found in [10]. A review on point-based and higher order entities pose estimation methods are presented in [15]. Model-based techniques shall be considered as an alternative for the navigation system during inspection of a non-cooperating satellite. The knowledge of the customer satellite geometric is provided by CAD models. Cropp use such techniques for pose estimation applied to a microsatellite [16]. The satellite in this case was modelled by lines. The simple box satellite allows detecting easily the lines in the image. The approach requires many iterations have reduced convergence results. Algebraic methods are based on retroreflective tape mounted on the satellite. An autonomous navigation system based upon point-based models and softposit is developed in [17], in which the vision system is tested in a testbed, which uses a scaled satellite model and a robotic manipulator capable of simulating 6 degrees of freedom motion of a satellite in orbit. A visual system for replacement of orbital replace unit is developed in [18], in which the vision system uses a grasping interface mock up, and a stereo vision system is employed to compute the pose. Natural features, like corners, of the spacecraft can be extracted from the image, and they can be matched with the correspondent spacecraft model. Hough transform algorithms are used for feature extraction. Numerous visual navigation systems employing landmarks attached on target satellite [11]. A real-time visual system is described in [5], in which the optical system identifies infrared light-emitting diodes attached on the target and estimates its pose. The pose calculation is based on four-point coplanar algorithm, which is provided in [12]. The experimental results show that this system can provide reliable and robust measurements of the target. The visual system would provide the required navigation solution for others GNC functions during medium and close approach phases. A visual system based on markers would be restricted to final approaching and docking/berthing phases and for satellites with three-axis attitude control. The majority of the stated works on autonomous RVD process uses either point or lines based methods. Point based models have been more widely developed than others model-based methods. Originally, this problem is usually referred to as PnP problem to designate the problem of determining the pose of the object with reference to the camera, given a set of n correspondences between points in the image and points in the 3D model of the object of interest. Methods of computing the pose by line-based schemes can be found in [19]. Pose estimation methods using higher entities like kinematic chains and free-form contours are addressed in. Real-time performance of free-form solutions is worse than point based solutions. This drawback would limit the application of such methods in system with reduced computational burden. A Kalman filter approach represents an alternative and attractive solution for referred problem. Goddard describes the use of the extended Kalman filter for approximation of pose and velocities [20]. The approach achieves good results for numerical experiments. In his work, translational and rotational velocities are assumed constant during the motion of the object. Pose estimation with the smallest subset of correspondences P3P is an important and old problem for which many solutions have been proposed. Theoretically, since the pose can be represented by six independent parameters, three points should be sufficient to solve this problem. The rotation represented by quaternions can be obtained using a close form solution. Alternatively, least squares solution that uses the Singular Value Decomposition (SVD) can be considered. It is then necessary to have at least a fourth point to disambiguate the obtained results. Kneip propose a closed-form solution that directly computes the rigid transformation between the camera and world frames $\boldsymbol{T}_w^c$. The proposed algorithm is much faster than the other solutions since it avoids the estimation of the 3D points depth in the camera frame and the estimation of the 3D-3D registration step [21]. Although P3P is a well-known solution to the pose estimation problem, other PnP approaches that use more points were usually preferred. Indeed pose accuracy usually increases with the number of points. Nevertheless, with in an outliers rejection process such as RANSAC, being fast to compute and requiring only three points correspondences. One can consider multi-stage methods that estimate the coordinates of the points or of virtual points in the camera frame and then achieve a 3D-3D registration process. Direct or one stage minimization approaches have been proposed. In the EPnP approach, the 3D point coordinates are expressed as a weighted sum of virtual control points. The problem is then reduced to the estimation of the coordinates of these control points in the camera frame. The main advantage of this approach is its reduced complexity, which is linear in the function of the number of points. Within the latter one-step approaches, the Direct Linear Transform (DLT) is certainly the oldest one but this solution is sensitive to noise. To reject outliers, methods like RANSAC or the use of M-Estimators such as the Tukey estimator are common trends to make the algorithm robust to occlusions and illumination variations. RANSAC uses the smallest set of possible correspondences and proceeds iteratively to enlarge this set with consistent data. M-estimators are a generalization of maximum likelihood estimation and least squares. Therefore, they are well suited to detect and reject outliers in a least square or iterative least square approach. With respect to RANSAC, which aggregates a set of inliers from a minimal number of correspondences, M-estimators use as many data as possible to obtain an initial solution and then iterate to reject outliers. M-estimators are more common than least squares because they allow using different minimization functions [22]. Many functions have been proposed in the literature that allow uncertain measures to have less influence on the final result and in some cases completely to reject the measures. Many M-estimator (Beaton Tuckey, Cauchy) can be considered leading to various ways to compute the confidence. RANSAC or M-estimators are two ways to ensure robust estimation. For space applications, rather than computational efficiency, accuracy is the key criterion. Although markers were often used, keypoints have been widely considered in the literature. A non-linear minimization technique is for example considered in [23] using SIFT. Robust process using RANSAC is usually considered. Considering keypoints, these methods are used for localization issue [6] in environments that have been reconstructed using a SLAM approach. Numerical nonlinear optimization techniques like

Levenberg-Marquardt are considered. The robustness declines when ambiguities between different edges occur. Other solutions have considered multiple hypotheses for potential edge-locations in the image. One of the drawbacks of these methods is that the 3D model may be made of segments, which leads to dealing with the CAD model. The previous approaches require a 3D model of the object or of the environment. Since a comprehensive or even a sparse 3D knowledge is not always easily available, the development of pose estimation methods that involve less constraining knowledge about the observed scene has been considered. The idea is then to perform the estimation of the scene structure and the camera localization within the same framework. This problem originally known as the structure from motion issue was handled off-line due to the high computational complication of the solution. This leads to vision-based SLAM that received much attention in robotics. Monocular SLAM methodology is based on Bayesian filtering approaches. It was proposed to integrate data thanks to an Extended Kalman Filter. Measurements are sequentially integrated within the filter, updating the probability density associated with the camera position and the scene structure. All past poses being marginalized; the number of parameters to be estimated only grows with the size of the map. The latter approach is based on the minimization of reprojection errors, have clearly demonstrated the feasibility of a deterministic SLAM system. Nevertheless, such SLAM based approaches lack absolute localization. It is computationally expensive in large environments. To achieve real-time requirement it has been proposed to decouple the localization and the mapping step. In SLAM methods comparable to PTAM, only few pixels contribute to the pose and structure estimation process. Dense or direct approaches such as DTAM allow each pixel contributing to the registration process. This is the case for LSD-SLAM. This latter approach is a keyframe method that builds a semi-dense map, which provides more information about the scene than feature-based approaches. Another interesting feature of LSD-SLAM is that it does not estimate a rigid transformation between two camera positions but a similarity transform which allows solving the scale estimation issue thanks to a scale-drift aware image alignment process. It demonstrated good results showing that scene reconstruction and camera localization can be achieved in real-time. Considering only pixel intensities, these approaches do not need feature extraction and matching process and provide a dense or semi-dense map of the environment. Nevertheless, the underlying illumination model assumes photometric consistency, which is not always realistic in real scenes and imposes small baselines between frames. Over the years, EKF based SLAM has been progressively replaced by keyframe methods. This was certainly due to PTAM, which demonstrated that a real-time implementation of BA was possible. So far, we considered a 2D-3D registration process. With some devices (e.g., multiple cameras systems), it is possible merging various cues to improve localization. It has been noted that it could be interesting to merge 2D-3D registration methods along with 2D-2D ones. Indeed, approaches, which directly compute the pose, are intrinsically mono image processes and can be subject to jitter, whereas motion based methods consist in multi-view processes that are subject to drift. Fiducial markers allow achieving simultaneously both target identification and camera localization. To simplify the detection process and the underlying image-processing algorithm, their design is ultimately simplified. The texture of the marker is individually designed for marker identification.

## 4. Methods

In this part, the projected method was described. Presented method is applicable to monocular camera systems. There are given images of known rigid object, which is seen, from different camera locations at every next frame. The features used to describe the target object are in 3D, while the projection of a feature found in the image is in 2D. The pinhole camera model is used to done the projection of the 3D coordinates of the object features with respect to the camera frame to the 2D coordinates found in the image. This transformation from the 3D coordination to the 2D coordinate is called a perspective projection. There are 2 sets of points: a 3D set representing the model denoted by, and a 2D set detected from the image denoted by. Assume it is known that each point in the 2D set is the image local feature, which is in the 3D set. Unfortunately, the correspondence between the 2D features and 3D is usually not known. This leads to a closely related to the pose estimation problem known as the correspondence problem. It is the process of finding out, which features in a set correspond to a feature in another set. If the position of the target is approximately known, the correspondence problem becomes simpler because one can project the geometric representation onto the image plane and associate each projected model feature to the closest image feature to obtain the correspondence. A feature representation of an object is effective to the pose determination. Every entity is locally evaluated. In addition, it represents a small part of the object. Examples for local features are corners. In this case at least 3 model points and three corresponding image points are required in order to determine the pose of an object. However in order to reduce the effects of measurement noise on the precision of the solution several studies including presented work use a larger number of points. The advantage in using point features in the pose estimation solution is the relative ease of extracting these features. Unknown are the six parameters that can describe relative pose of two objects: three coordinates which describes the linear translation of object in relation to camera and three angles of rotation (roll, pitch, and yaw) which describes the mutual angular orientation of two objects in space.

There were defined two coordinate of this coordinate system is located in optical centre. The $x_C$ axis of this coordinate system is oriented downward, the $y_C$ axis is oriented on right and the $z_C$ axis completes the right-handed coordinate system. The position of each pixel on photo is given in camera coordinate system. The second one is model coordinate system and the origin of this coordinate system is located on the surface of object. $X_W$ axis is oriented in front of the target, the $Y_W$ axis on right and the $Z_W$ axis completed the coordinated system. Origin of this coordinate system is translated from origin of camera coordinate system by vector $\boldsymbol{t}_w^c$. Down subscripts, $W$ means that coordinates are in model of coordinate frame. The image plane is parallel to the axes of camera coordinate system at distance $f$ from the optical center. Object is rotated and displaced with respect to camera coordinate system. The camera internal parameters are known. This was designated in camera calibration process. In most 3D tracking methods, the internal parameters are assumed to be known. For additional details about camera models, the interested reader is referred to the photogrammetric literature. The transformation of two object consists of two components: a component that describes the rotation of the model coordinate frame relative to the camera coordinates and a component that describes the relative translation between the two coordinate frames. The coordinates of points of object are referred in model coordinate frame so before mapping on image plane they must be transferred to camera coordinate system. Object is rotated and displaced with respect to camera coordinate system. Unknown are the Euclidean transformation from a world coordinate system to the camera coordinate system. The rotational component can be described in several methods, which influence the choice of computational method to be used for the estimation process. The angular orientation of object was parametrized by using of Euler angles. These three angles form three free parameters that describe any rotation transformation. There is singularity when the coordinate frames are rotated mutual by pitch angle equal. The proposed method work in such manner as described downwards. At first, a photo of object is taken. Next on this photograph, there are detected local features. Local features of object are usually associated with a change of image properties simultaneously. To handle as wide as possible a range of viewing conditions, feature point extraction should be insensitive to scale, viewpoint, and illumination changes. The local features of the object are extracted by using of Scale Invariant Feature Transform (SIFT) detector and descriptor proposed by Lowe. Algorithm extracts features and is for object recognition based on local 3D extrema in the scale-space pyramid build with difference-of-Gaussian filters. First, the location and scale of the keypoints are determined precisely by interpolating the pyramid of Difference-of-Gaussians used for the detection. The input image is successively smoothed with a Gaussian kernel and sampled. The difference of Gaussian representation is obtained by subtracting two successive smoothed images.

The Gaussian kernel and its derivatives are the most efficient smoothing kernels for scale space analysis. To achieve image rotation invariance, an orientation is assigned to the keypoint. It is taken to be the one corresponding to a peak in the histogram of the gradient orientations within a region around the keypoint. All dig levels are constructed by combining smoothing and subsampling. This method is quite stable under viewpoint changes, and achieves an accuracy of a few degrees. An image is transformed into a group a local features. On the exit of this algorithm there is known the two dimensional vector of coordinates of each feature and the second vector which contains the radius of each feature and the angle of orientation in radians. To further explore, the methods of solving the pose estimation problem, one must be able to model the target object. It is usually described by a set of features. Next, similar to the previous step, the local features of the object from model of target are extracted. During an offline training stage, a database of interest object points was build. Their positions on the object surface are known. Images in which the object has been manually registered were used for this purpose. At runtime, SIFT features are extracted from the current frame, matched against the database, resulting in a set of 2D-3D correspondences. It was mentioned that point correspondences should be available beforehand. In 1999 SIFT was proposed and later many types of keypoint detectors and descriptors have been considered. The common framework for 2D matching usually considers steps like keypoints extraction, description and matching. First, among all the pixels in the image, a subset of pixels of interest is selected. For each selected pixel, its local texture is then converted into a descriptor, which is a vector that intends to encode, in a unique way, the keypoint and its local neighbourhood. Finally, these descriptors extracted in two images are matched to create correspondences. As far as pose estimation is concerned, keypoint descriptors on a reference model are first computed offline and stored in a descriptor database. Then, on-line, keypoints are extracted from each image and matched, in the descriptor space, with those in the template. Finally, camera pose or displacement can be computed from these correspondences. Such features should be extracted from perspectively transformed images. This process should be highly repeatable and performed in real time. Keypoint detectors are designed to feature invariance properties with respect to geometric transformation such as translation, rotation, affine transformation and scale change. Harris detector is a widely used corner detector that computes the cornerness score of each pixel from gradients of an image patch. SUSAN is an alternative approach that selects a pixel as a corner if it is not self-similar within a local image patch. The similarity is computed between a pixel and its surrounding pixels. FAST is similar to SUSAN approach and considers only pixels on a circle for fast calculation. FAST is computationally efficient because it computes similarity with pixels selected with a machine learning technique. Since the keypoints are not scale-invariant, an image pyramid can be considered so that keypoints can be detected under scale changes. The next step usually consists in computing a feature vector that fully describes the keypoint and its local neighbourhood. For robustness issue, the resulting descriptor should be made invariant to photometric variations. Rotation invariance is normally achieved by assigning orientation to extracted keypoints for orientation normalization. Orientation is computed as the peak of histogram of gradient. For each oriented keypoint, a feature vector is then computed. A local keypoint descriptor can be classified into two approaches: histogram of oriented gradients or intensity comparisons. Histogram of oriented gradients used in SIFT is computed such that an image patch is segmented into small regions, and histogram of oriented gradients in each region is computed. This preserves the intensity information of an image patch. A similar framework is used in SURF. Since feature descriptors from the methods above have floating point values, they can be compacted into a binary string. Computational cost for matching is then reduced.

The next task is the pose estimation of the object. It was assumed earlier a set of point correspondences with model points and image points. The object pose can then be estimated from such correspondences. The unknowns are the translation vector $\boldsymbol{t}_w^c$ and rotational vector $\boldsymbol{R}_w^c$. They have been found iteratively by using POSIT algorithm. This algorithm needs focal length, which is assumed same for $x$ and $y$, 4 or more non-coplanar 2D-3D correspondences. This algorithm estimates pose uses a scaled orthographic projection, which resembles the real perspective projection at convergence. Such approximation leads to a linear equation system. This gives the rotation and translation directly and there is no the needs of a starting pose. A scale value is introduced for each correspondence, which is iteratively updated. What is known, the distribution of the feature points on the object and the images of these points by perspective projection. If someone could build SOP images of the object feature points from a perspective image, someone could apply the POS algorithm to these Scale Orthographic Projection (SOP) images and we would obtain an exact object pose. Computing exact SOP images requires knowing the exact pose of the object. However, someone can apply POS to the actual image points we then obtain an approximate depth for each feature point and we position the feature points on the lines of sight at these depths. Then we can compute an SOP image. At the next step, we apply POS to the SOP image to obtain an improved SOP image. Repeating these steps, someone converge toward an accurate SOP image and an accurate pose. Facts about POSIT algorithm can be found in [3].

## 5. Experiments

This section described experiments, which were completed. Experiments were tested in Matlab software. It was needed to perform a simulation of space environment on Earth. Experiments were tested as follow. Satellite was modelled as rigid box of dimension 130x70x80 mm (Fig. 4). Object was suspended under mounting stand. Camera was mounted on mobile robot, which can be precisely translated and rotated in relation to object coordinate frame (Fig. 2). A set of photos of object was used. It was measured ground truth position of robot. It was used inertial navigation sensor Honeywell HMR-3500 and GX3. The source of light was fixed. Uniform illumination for the experiments was provided.

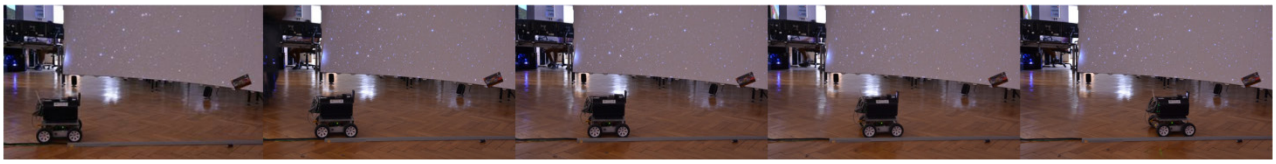

*Fig. 2. First experiment*

*Fig. 3. Second experiment*

There were measured of position and three rotational degrees of freedom. Next, the result of ground truth measurement was compared with calculated results. The main goal of experiments was to check how accurate algorithm is. Others goals were to confirm the robustness to varying occluding conditions. It was expected that calculated results should be similar to ground truth measurements. Internal parameters were estimated during an offline camera calibration phase. The six plots present results for a first chosen example. The camera has a distance of approximately 1600 mm to the object. At the beginning, the mobile robot is not moving. Next, the robot is moving and takes photos.

*Fig. 4. Image sequence obtained from one of the experiments*

On the horizontal axes of first three plots (left side) there is given time in seconds and on vertical axes the measured translations in millimetres (Fig. 5). On the next plots (right side) on horizontal axes there are given, similar as upper figures, time and on vertical orientation in degrees. Green line shows ground truth using HMR3500 sensor, red GX3 and blue line shows the vision system based measurements.

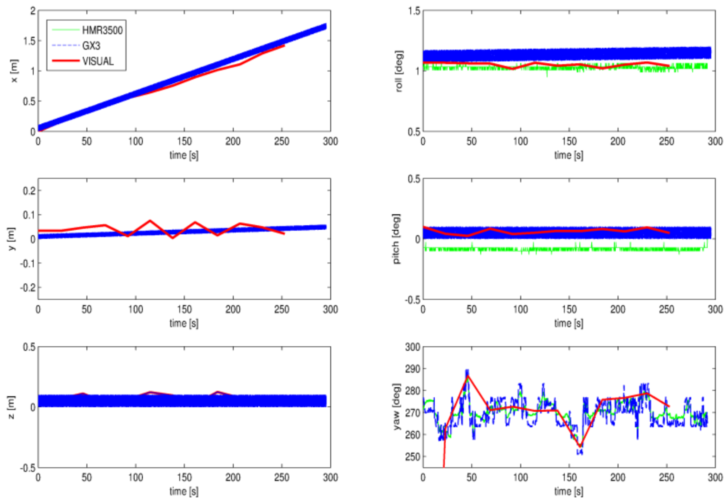


*Fig. 5. Results of the first experiment*

Ground truth (green and blue lines) should be close to measured results. On upper left there was presented the linear translation of object along x-axis of the camera coordinate system. On the second plot there was presented linear translation along y-axis. In reality, there was no translational motion along y but from vision system measurements one can see that maximum difference for y-axis is about 9 mm. Possible cause of this errors is nature of present's method. There is possible to try reducing the errors if better correspondence generation algorithm will be obtained. On third plot error for z-axis is about 85 mm, which is much more than for x and y-axis. This is measured in direction perpendicular for image plane and it is distance from camera coordinate system and the object. It was expected that this error should be smaller. Next three plots presents rotations around three axes of object coordinate system. In ground truth measurements there were no rotation about x and y-axes. Four plot presents that there was a rotation about 3 deg. On the fifth plot, there is small error 3 deg between both ground truth rotation and vision based measurement. In the case of rotation around z axis there was quite small error. The results were as expected. Translational errors were under few millimetres for x and y-axes and bigger for z-axis.

Afterward, there was conducted second experiment (Fig. 3). Mobile robot was moved in other way as in first experiment. Similar as in the first case there are presented six plots (Fig. 6). First three presents linear translations along axes of camera coordinate system and three presents angular orientation of object.

First plot presents linear translation along x-axis. There is significant error between both measurements. At the end of simulation, difference is about 100 mm. There were small motions caused by imperfectly camera to robot. On the second plot both lines green and blue are close because object was constrained on y-axis and could not pitch and yaw. On the next three plots there are presented measurements for angular orientation of object. For roll motion there is error about 2 degrees. Similar results were obtained for pitch. Ground truth error varies around 1 mm in space. The sixth plot presents yaw. In this case, error is 5 deg. There was presented linear translation along y-axis. From vision system measurements one can see that maximum difference for y-axis is about 10 mm. Possible cause of this errors is nature of present's method. There is possible to try reducing the errors if better correspondence generation algorithm will be obtained. On third plot error for z-axis is about 2 mm, which is less than for x and y-axis. Next three plots presents rotations around three axes of object coordinate system. Four plot presents that there was a rotation about 3 deg. On the fifth plot, there is small error between both ground truth rotation and vision-based measurement. After 13 second, error is bigger than at the beginning. Pose estimation errors of 4 degrees in orientation are obtained with the testbed experiment. Up to 25 iterations are needed for convergence. It is mentioned that the precision can be improved, however at higher running time.

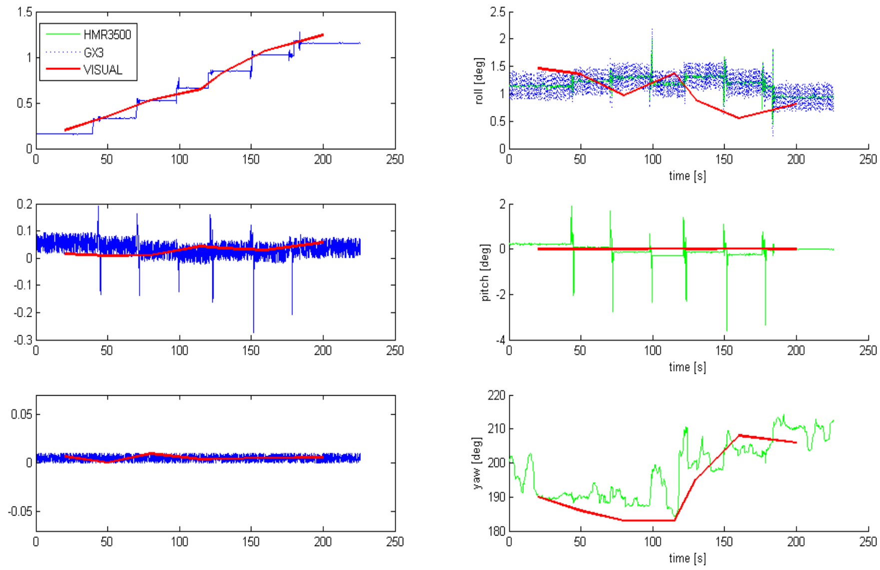


*Fig. 6. Relative position and orientation results obtained in the second experiment*

The experiments take about 13 ms per frame on modern CPU. It was shown that introduced method is able to run in real time. The camera bed should be more precisely screwed because it introduced serious problems. Using a single monocular image and utilising knowledge of the target spacecraft, estimation of the target's six relative rotation and translation parameters with respect to the camera are found. Simulation results for a microsatellite 2 metres away show errors on the order of 10 mm in position and 1 in rotation. It is planned still to work on proposed method.

## 6. Conclusions

In last years, autonomous satellite formation flight becomes progressively important. This article addresses the design of a monocular vision-based navigation system for on-orbit-servicing and formation-flying applications. A method for calculating the position and orientation of uncooperative space objects is presented. Markerless local features based navigation system has been studied. Visual navigation system uses a single camera, because satellite services team has defined a resource of energy. The advantages of the developed algorithm are that it does not need manual initialization. A modified softPOSIT algorithm was used to calculate the position and orientation of the object. The algorithm is based on placing the problem of determining the position and orientation of the object as the task of minimizing a function. The algorithm uses a scaled orthographic projection. In order to reproduce the space conditions the laboratory stand was built. The developed method was tested experimentally for different scenarios approach satellites to each other. The impact of changes in illumination of the object was analysed. Comparing the measured position and orientation of the sensors spatial navigation and the navigation system switcher rated accuracy of the developed method. Achieved satisfactory accuracy of the algorithm. It was found that the system is able to determine the position under different lighting conditions.

**Acknowledgments**

*The following research was done within project (agreement No. PBS2/B3/17/2013), sponsored by National Centre for Research and Development.*

**References**

[1] Jacewicz, M., Głębocki, R., *Navigation for satellites*, Warsaw 2016.
[2] Amzajerdian, F., Roback, V., Bulyshev, A., Brewster, P., *Imaging flash LIDAR for safe landing on solar system bodies and spacecraft rendezvous and docking*, *Proceedings of the SPIE Laser Radar Technology and Applications XX and Atmospheric Propagation XII*, Baltimore, MD, USA 2015.
[3] Woods, J., Christian, J., *LIDAR – based relative navigation with respect to non-cooperative objects*, Acta Astronaut, 2016.
[4] Chien, C., Baker, K., *Pose estimation for servicing of orbital replacement units in a cluttered environment*, *Proceedings of the IEEE International Conference on Robotics and Automation*, New Orleans, LA, USA 2004.
[5] Sharma, S., D'Amico, S., *Comparative assessment of techniques for initial pose estimation using monocular*, Acta, Astronaut, 2016.
[6] Mikolajczyk, K., *Scale & Affine Invariant Interest Point Detectors*, International Journal of Computer Vision 60(1), pp. 63-86, 2004.
[7] English, C., Okouneva, G., Saint-Cyr, P., Choudhuri, A., Luu, T., *Real-time dynamic pose estimation systems in space lessons learned for system design and performance evaluation*, Int. J. Intell. Control. Syst., 2011.
[8] Głębocki, R., Kicman, P., Kopyt, A., *Navigation module for mobile robot*.
[9] Opromolla, R., Fasano, G., Rufino, G:, Grassi, M., *Uncooperative pose estimation with a LIDAR-based system*, Acta Astronaut, 2015.
[10] Cropp, A., *Estimating pose of known target satellite*, Electronics Letters, Vol. 36, No. 15, pp. 1331-1332, 2000.
[11] Ho, C. A. M., *Autonomous Spacecraft Docking using a Computer Vision System*, Proc. 31st Conf. on Decision and Control, Tucson Arizona 1992.
[12] Vincent, P., *3D Model-Based Tracking For Space Autonomous Randezvouz*.
[13] Palmerini, G., Sabatini, M., Gasbarri, P., *Analysis and tests of visual based techniques for orbital rendezvous operations*, In Proceedings of the IEEE Aerospace Conference, Big Sky, MT, USA 2013.
[14] *Simulation of the Docking Phase for the SMART-OLEV Satellite Servicing Mission*, 9th International Symposium on Artificial Intelligence, Robotics and Automation in Space (iSAIRAS).
[15] Gasbarri, P., Sabatini, M., Palmerini, G., *Ground tests for vision based determination and control of*, Acta Astronaut, 2014.
[16] Ruel, S., Luu, T., Anctil, M., Gagnon, S., *Target localization from 3D data for on-orbit autonomous rendezvous and docking*, Proceedings of the IEEE Aerospace Conference, Big Sky, MT, USA 2008.
[17] Fischler, M., *Random Sample Consensus: A Paradigm for Model Fitting with Applications toImage Analysis and Automated Cartography*, Comm. of ACM, Vol. 24, No. 6, pp. 381-395, 1988.
[18] Quan, L., Lan, Z-D., *Linear N4-point pose determination*, 6th International Conference on Computer Vision, Bombay 1998.
[19] Rosenhahn, B., *Foundations About 2D-3D Pose Estimation Problem*.
[20] Petersen, T., *A Comparison of 2D-3D Pose Estimation Methods*, Aalborg University, Ballerup 2008.

[21] Philip, N., Kumar, E., Ananthasayanam, M., *Sliding Observer for a Robust Relative Position and Attitude Estimation During the Final Phase of an Autonomous DockingMission*, Proc. 49th Int'l Astronautical Congress, Melbourne, Australia 1998.

[22] Grunert, J., *Das pothenotische Problem in erweiterter Gestalt nebst über seine Anwendungen in Geodäsie*, Grunerts Archiv für Mathematik und Physik, 1841.

[23] Fischler, M., *Random Sample Consensus: A Paradigm for Model Fitting with Applications toImage Analysis and Automated Cartography*, Comm. of ACM*, V*ol. 24, No. 6, pp. 381-395, 1988.

[24] Jasiobedzki, P., *Pose Determination and Tracking for Autonomous Satellite Capture*, Proceeding of the 6th International Symposium on Artificial Intelligence and Robotics & Automation in *Space*, Quebec 2001.

[25] Jörgensen, J., Harr, J., *PRISMA – An Autonomous Formation Flying Mission*, ESA Small Satellite Systems and Services Symposium, Sardinia 2006.