ISSN 0005-1144
ATKAFF 50(3–4), 195–204(2009)

Robert Cupec, Emmanuel Karlo Nyarko, Andreja Kitanov, Ivan Petrović

RANSAC-Based Stereo Image Registration with Geometrically Constrained Hypothesis Generation

UDK 004.932
IFAC 5.8.7;4.2.2

Original scientific paper

An approach for registration of sparse feature sets detected in two stereo image pairs taken from two different views is proposed. Analogously to many existing image registration approaches, our method consists of initial matching of features using local descriptors followed by a RANSAC-based procedure. The proposed approach is especially suitable for cases where there is a high percentage of false initial matches. The strategy proposed in this paper is to modify the hypothesis generation step of the basic RANSAC approach by performing a multiple-step procedure which uses geometric constraints in order to reduce the probability of false correspondences in generated hypotheses. The algorithm needs approximate information about the relative camera pose between the two views. However, the uncertainty of this information is allowed to be rather high. The presented technique is evaluated using both synthetic data and real data obtained by a stereo camera system.

Key words: Image registration, Stereo vision, Feature tracking, RANSAC

Registracija stereo slika postupkom zasnovanim na RANSAC strategiji s geometrijskim ograničenjem na generiranje hipoteza. U radu je predložen jedan pristup registraciji skupova značajki detektiranih na dva para stereo slika snimljenih iz dva različita pogleda. Slično mnogim postojećim pristupima registraciji slika, predložena se metoda sastoji od početnog sparivanja značajki na temelju lokalnih deskriptora iza kojeg slijedi postupak temeljen na RANSAC-strategiji. Predloženi je pristup posebno prikladan za slučajeve kada rezultat početnog sparivanja sadrži veliki postotak pogrešno sparenih značajki. Strategija koja se predlaže u ovom članku je da se korak RANSAC-algoritma u kojem se slučajnim uzorkovanjem generiraju hipoteze zamijeni postupkom u kojem se hipoteza generira u više koraka, pri čemu se u svakom koraku, korištenjem odgovarajućih geometrijskih ograničenja, smanjuje vjerojatnost izbora pogrešno sparenih značajki. Algoritam treba približnu informaciju o relativnom položaju kamera između dva pogleda, pri čemu je dopuštena nesigurnost te informacije prilično velika. Predstavljena strategija je provjerena korištenjem sintetičkih podataka te pokusima sa slikama snimljenim pomoću stereo sustava kamera.

Ključne riječi: registracija slika, stereo vizija, praćenje značajki, RANSAC

1 INTRODUCTION

Registration of data obtained by viewing a scene from two or more different views is a fundamental problem in computer vision. The solution to that problem provides a means for motion estimation and 3D scene reconstruction by integration of the data obtained from multiple views. In this article, the registration of the data obtained by stereo vision is addressed. Stereo vision is a powerful tool for obtaining 3D information from camera images. The stereo vision system considered in this paper consists of two calibrated cameras and appropriate software which uses the images taken by both cameras to perform 3D reconstruction of a set of left camera image points. A pair of images taken by a the two calibrated stereo cameras is referred to in this paper as a *stereo image*.

The most often used approach to registration of the data obtained by stereo images taken from two different views is to perform stereo reconstruction for each view thus obtaining two sets of 3D points and then to perform registration of these two feature sets. This can be achieved by extracting from image a sparse set of distinguishable points which can be reliably tracked across a sequence of images. By matching the local descriptors [1] assigned to each extracted point, a set of pairs is obtained, where the first element of the pair is a 3D point from the first stereo image and the second element is a 3D point from the second stereo image. A pair whose both elements represent the same 3D point in the scene is a correct correspondence.

Assuming that the correct correspondences between the 3D point features in two sets are available, the registration of those two sets can be obtained by the closed-form solution presented in [2, 3]. That approach, however, fails to give optimal solutions because it does not consider the uncertainty of the stereo reconstruction caused by non homogenous and non isotropic noise. The methods that deal with non homogenous and non isotropic noise are presented in [4, 5]. The same problem is solved in [6] by a maximum-likelihood estimator used later in [7] for navigation of a mobile robot.

Nevertheless, all these methods assume that correct feature correspondences are available or that the percentage of possible false correspondences is relatively low. However, matching methods based on local descriptors can result in a significant number of false correspondences, referred to in this paper also as *outliers*. Therefore, the set of the point pairs obtained by matching their descriptors, which is referred to in this paper as *initial correspondence set*, must be pruned before using it for estimation of the motion parameters.

One approach which can be used to deal with false correspondences is RANSAC.

RANSAC (RANdom SAmple Consensus) [8] is a very popular method widely used in robotics to fit a model to a set of data corrupted by outliers. Among other applications, it has been used for registration of image data obtained by a single camera [9] or a stereo camera system [10, 11]. The idea of RANSAC is to generate a set of model hypotheses by selecting randomly subsets of the given data set containing the minimum number of data points sufficient to define a model. For each hypothesis a set of data points which fits the hypothesized model within a certain tolerance is determined, called the consensus set. The hypothesis corresponding to the greatest consensus set is considered to be the most probable one. The probability of generating a false hypothesis from a randomly selected data set increases with the number of outliers in the input data set as well as with the size of the selected set.

The strategy proposed in this paper is to modify hypothesis generation step of the basic RANSAC approach by performing a multiple-step procedure which uses geometric constraints in order to reduce the probability of false correspondences in the hypothesis. The procedure starts with a prior information about the relative camera pose between the two views whose uncertainty is allowed to be rather high. This information can be estimated e.g. from the motion commands given to the robot. Alternatively, the initial translation and rotation of the robot can be set to zero, and the uncertainty of the change in robot's pose can be estimated from the known maximum translational and rotational velocity of the robot. This initial camera pose together with its uncertainty is corrected recursively by the information provided by a feature pair randomly selected from the input set of match candidates. The corrected pose and its uncertainty are used to formulate geometric constraint for the selection of the next match candidate. The procedure is repeated until a sufficient number of matches are considered in the hypothesis. Recursive pose estimation from a set of feature pairs, where the selection of the next data is constrained by currently computed pose and its uncertainty, is already applied in the approaches presented in [12] and [13] which use Kalman Filter formalism. In this paper, the pose refinement is performed using a version of the Levenberg-Marquardt algorithm ([14]) described in [15].

The article is structured as follows. In the next section, the problem of stereo image registration given a set of features and a set of possible feature matches is defined. In Section 3, a geometrically constrained RANSAC algorithm for solving the considered problem is proposed. The presented technique is evaluated using both synthetic data and real data obtained by a stereo camera system. The results of these experiments are reported in Section 4.

2 PROBLEM DEFINITION

A common approach to the problem of registration of two sets of 3D points obtained from two different views of a scene is to determine the correspondences between the features detected in the stereo image pairs taken from these two views and then to compute the relative camera pose between the views.

Let P_i, i = 1, …, N be a set of 3D points observed by a stereo camera pair, as shown in Fig. 1, and let $Q_{L,i}$ and $Q_{R,i}$ be the points representing the projections of P_i onto the left and right camera image respectively. Each camera is assigned a reference coordinate frame centered

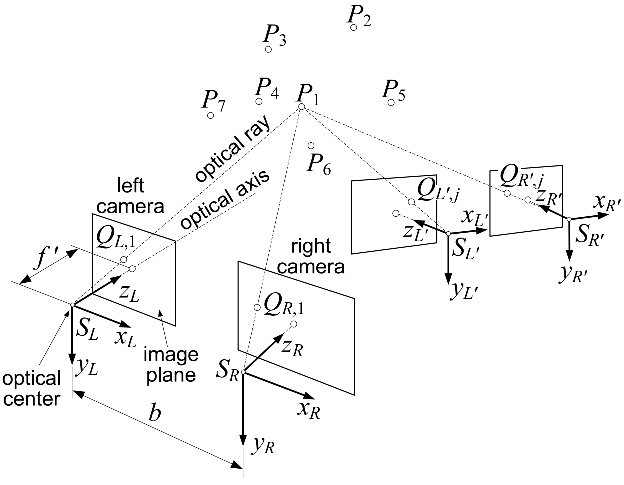


Fig. 1. A set of points observed by a moving stereo camera system

in the optical center of the camera with z-axis identical to the optical axis. Let S_L and S_R be the reference frames of the left and right camera respectively.

The 3D coordinates $\boldsymbol{p}_i = [\ x_i \quad y_i \quad z_i\ ]^T$ of each point P_i relative to the camera frame S_L can be computed from the image coordinates of its projections $Q_{L,i}$ and $Q_{R,i}$ by triangulation.

In this work, the uncertainty of stereo reconstruction is modeled by 3D Gaussian distribution as proposed in [16]. The point coordinates $\boldsymbol{p}_i$ obtained by stereo reconstruction are regarded as random variables with mean $\boldsymbol{p}_i$ and covariance matrix $\boldsymbol{C}_{\boldsymbol{p},i}$. Let $F = \{\boldsymbol{p}_i, i = 1, \ldots, n\}$ represent the set of all features detected in the scene and reconstructed by stereo vision from a stereo image pair. Note that, in general, $n \leq N$ since the vision system can sometimes fail to detect some of the features P_i. Let us now consider the case where the same scene is observed by the stereo camera system from another viewpoint, as shown in Fig. 1. Let S_L' and S_R' be the reference frames of the left and right camera respectively corresponding to the second view and let $F' = \{\boldsymbol{p}_j', j = 1, \ldots, m\}$ be a set of vectors $\boldsymbol{p}_j'$ defining the positions of 3D points reconstructed from the second stereo image pair relative to S_L'. The uncertainties in the coordinates $\boldsymbol{p}_j'$ are described by the covariance matrices $\boldsymbol{C}_{\boldsymbol{p},j}'$. Let the pose of S_L' relative to S_L be described by vector $\boldsymbol{w} = [\ \boldsymbol{\phi}^T \quad \boldsymbol{t}^T\ ]^T$, where $\boldsymbol{\phi} = [\ \alpha \quad \beta \quad \theta\ ]^T$ is a vector of three angles defining the orientation and $\boldsymbol{t} = [\ t_x \quad t_y \quad t_z\ ]^T$ is a vector defining the position of S_L' relative to S_L.

Two features $\boldsymbol{p}_i \in F$ and $\boldsymbol{p}_j' \in F'$ which represent the same point P_i in the scene are referred to herein as *corresponding features*. In the ideal case, for two corresponding features $\boldsymbol{p}_i$ and $\boldsymbol{p}_j'$ the following holds

$$\boldsymbol{p}_i - \boldsymbol{R}(\boldsymbol{\phi})\boldsymbol{p}_j' - \boldsymbol{t} = \boldsymbol{0}, \tag{1}$$

where $\boldsymbol{R}(\boldsymbol{\phi})$ is the rotation matrix whose elements are functions of $\boldsymbol{\phi}$. In reality, however, due to the uncertainty in stereo reconstruction, the term

$$\boldsymbol{e}_{ij}(\boldsymbol{w}) = \boldsymbol{p}_i - \boldsymbol{R}(\boldsymbol{\phi})\boldsymbol{p}_j' - \boldsymbol{t} \tag{2}$$

mostly differs from $\boldsymbol{0}$.

The correspondence between the features detected in the first stereo image pair and the features detected in the second stereo image pair can be defined by a set T of pairs (i, j) where $\boldsymbol{p}_i$ and $\boldsymbol{p}_j'$ are corresponding features. Finding correct feature correspondences is critical for precise estimation of the relative camera pose $\boldsymbol{w}$.

If the features are detected using the Scale Invariant Feature Transform (SIFT) proposed by [17] or Speeded Up Robust Features (SURF) proposed by [18] their correspondences can be obtained by searching for feature pairs with similar local descriptors. Although the correspondences obtained using the descriptors of the detected features provided by SIFT and SURF are very reliable, there is still possibility of false correspondences, especially in the scenes in which a number of features with similar descriptors are detected.

Let T_{ini} be the set of correspondences (i, j) obtained e.g. by comparing the local descriptors and let us assume that T_{ini} contains a significant number of false correspondences. The problem considered in this paper is to prune the set T_{ini} in order to obtain the largest possible set T_{fin} containing only correct correspondences. A set of correspondences T such that for any two pairs $(i, j) \in T$ and $(i', j') \in T$ holds $i' \neq i$ and $j' \neq j$ is referred to herein as a *set of unique correspondences*. The set T_{fin} must be a set of unique correspondences.

One approach to solve the considered problem is to search for the largest set $T \subseteq T_{ini}$ for which a pose $\boldsymbol{w}$ exists such that $\boldsymbol{e}_{ij}(\boldsymbol{w})$ is sufficiently small for all $(i, j) \in T$. The terms $\boldsymbol{e}_{ij}(\boldsymbol{w})$ can be assessed by a measure $r(\boldsymbol{p}_i, \boldsymbol{C}_{\boldsymbol{p},i}; \boldsymbol{p}_j', \boldsymbol{C}_{\boldsymbol{p},j}'; \boldsymbol{w})$ which takes into account the uncertainties of the coordinates $\boldsymbol{p}_i$ and $\boldsymbol{p}_j'$.

Let us define the *consensus set* as the set $W(\boldsymbol{w})$ of unique correspondences (i, j) assigned to a pose $\boldsymbol{w}$ such that for every pair $(i, j) \in W(\boldsymbol{w})$

$$r(\boldsymbol{p}_i, \boldsymbol{C}_{\boldsymbol{p},i}; \boldsymbol{p}_j', \boldsymbol{C}_{\boldsymbol{p},j}'; \boldsymbol{w}) \leq \varepsilon_0 \tag{3}$$

where ε_0 is a predefined threshold. The problem of registration of two 3D point sets can thus be formulated as the search for the pose $\boldsymbol{w}$ with the greatest consensus set $W(\boldsymbol{w}) \subseteq T_{ini}$. This consensus set can be considered as the most probable set of correct correspondences.

3 RANSAC WITH GEOMETRICALLY CONSTRAINED HYPOTHESIS GENERATION

A straightforward approach to apply RANSAC to the problem described in Section 2 would be the following. Let the feature sets F and F' obtained from two views be the input data set and a relative camera pose $\boldsymbol{w}$ between these two views the model which is fitted to that data set. Let us assume that an initial feature correspondence T_{ini} is available, although it is allowed to be corrupted by a significant amount of false correspondences. Since three 3D points are sufficient for constraining the camera pose, a set U of three pairs of corresponding points is randomly selected from T_{ini} for generating a hypothesis. The pose $\boldsymbol{w}$ computed from these three points can then be used to transform all features $\boldsymbol{p}_j' \in F'$ to the reference frame S_L of the first view and the hypothesis can be evaluated by the size of the corresponding consensus set $W(\boldsymbol{w})$. The distance measure used herein to define the consensus set is the *Mahalanobis distance*. Let $\boldsymbol{p}_i$ be the coordinates of a 3D point relative to the reference frame S_L and let $\boldsymbol{p}'_j$ be the coordinates of a 3D point relative to the reference frame S'_L. Furthermore, let $\boldsymbol{w} = [\ \boldsymbol{\phi}^T \quad \boldsymbol{t}^T\ ]^T$ be the pose of S_L relative to S'_L. The Mahalanobis distance between the first point and the second point whose coordinates are transformed to the reference frame S_L are given by

$$r(\boldsymbol{p}_i, \boldsymbol{C}_{\boldsymbol{p},i}; \boldsymbol{p}'_j, \boldsymbol{C}'_{\boldsymbol{p},j}; \boldsymbol{w}) = \boldsymbol{e}_{ij}(\boldsymbol{w})^T \boldsymbol{C}_{ij}^{-1}(\boldsymbol{\phi}) \boldsymbol{e}_{ij}(\boldsymbol{w}), \tag{4}$$

where

$$\boldsymbol{C}_{ij}(\boldsymbol{\phi}) = \boldsymbol{C}_{\boldsymbol{p},i} + \boldsymbol{R}(\boldsymbol{\phi}) \boldsymbol{C}'_{\boldsymbol{p},j} \boldsymbol{R}^T(\boldsymbol{\phi}). \tag{5}$$

In that case, condition (3) becomes

$$\boldsymbol{e}_{ij}(\boldsymbol{w})^T \boldsymbol{C}_{ij}^{-1}(\boldsymbol{\phi}) \boldsymbol{e}_{ij}(\boldsymbol{w}) \leq \varepsilon_0. \tag{6}$$

A pseudocode of an algorithm for registration of 3D point sets based on the standard RANSAC approach (std. RANSAC) is given in the following.

Algorithm 1 RANSAC-based stereo image registration

```
Parameters: $\varepsilon_0$, $k_{max}$
Input: $F$, $F'$, $T_{ini}$
Output: $\boldsymbol{w}_{fin}$, $T_{fin}$
 1. $T_{best} \leftarrow \emptyset$
 2. $k \leftarrow 0$
 3. repeat
 4.     Generate pose hypothesis: Select randomly a subset $U$ of 3 unique correspondences from the set $T_{ini}$ and determine the pose $\boldsymbol{w}$ which minimizes a particular cost function

        $$\Im(\boldsymbol{w}, U) = \sum_{(i,j) \in U} f(\boldsymbol{p}_i, \boldsymbol{C}_{\boldsymbol{p},i}; \boldsymbol{p}'_j, \boldsymbol{C}'_{\boldsymbol{p},j}; \boldsymbol{w}) \tag{7}$$

        where $f$ is a function defining the contribution of one feature pair to the overall cost. If the obtained pose $\boldsymbol{w}$ is invalid, repeat this step.
 5.     Determine the consensus set $W(\boldsymbol{w})$ according to the condition (3).
 6.     if $|W(\boldsymbol{w})| > |T_{best}|$ then
 7.         $T_{best} \leftarrow W(\boldsymbol{w})$
 8.         $\boldsymbol{w}_{best} \leftarrow \boldsymbol{w}$
 9.     end if
10.     $k \leftarrow k + 1$
11. until $k = k_{max}$
12. Determine the pose $\boldsymbol{w}_{fin}$ which minimizes (7) over the set $T_{best}$.
13. Determine the consensus set $W(\boldsymbol{w}_{fin})$.
14. $T_{fin} \leftarrow W(\boldsymbol{w}_{fin})$
15. return $\boldsymbol{w}_{fin}$, $T_{fin}$
```

A pose hypothesis is generated by determining the pose $\boldsymbol{w}$ which minimizes a particular cost function. If this cost function is

$$\Im(\boldsymbol{w}, U) = \sum_{(i,j) \in U} \|\boldsymbol{e}_{ij}(\boldsymbol{w})\|^2, \tag{8}$$

a closed-form solution to this problem exists [2, 3], which enables very efficient computation of the pose $\boldsymbol{w}$. This approach, however, does not consider the uncertainty of the stereo reconstruction caused by non homogenous and non isotropic noise. Alternatively, the cost function

$$\Im(\boldsymbol{w}, U) = \sum_{(i,j) \in U} \boldsymbol{e}_{ij}(\boldsymbol{w})^T \boldsymbol{C}_{ij}^{-1}(\boldsymbol{\phi}) \boldsymbol{e}_{ij}(\boldsymbol{w}), \tag{9}$$

can be used which takes into account the uncertainty of $\boldsymbol{p}_i$ and $\boldsymbol{p}'_j$. Nevertheless, to the best of our knowledge, the closed-form solution to minimization of (9) does not exist, which means that $\boldsymbol{w}$ must be computed by an iterative approach such as Levenberg-Marquardt algorithm. Therefore, the hypothesis generation using (9) is much slower than the one based on minimization of (8).

The probability of generating a false hypothesis from a randomly selected set U of feature pairs increases with the percentage of outliers in the input data set as well as with the number of pairs in U. In the case of a high percentage of false correspondences in T_{ini}, the number of random samplings for a reliable performance can be large.

The strategy proposed in this paper is to perform a multiple-step hypothesis generation, which uses geometric constraints in order to reduce the probability of false correspondences in the hypothesis. Instead of randomly selecting a minimum data set U needed to define camera pose (see step 4 of Algorithm 1), set U is formed sequentially by taking u pairs from T_{ini} one by one in such a way that the information provided by the feature pairs currently contained in U is used to formulate a geometric constraint for selection of the next pair.

The proposed algorithm needs a prior information about the relative camera pose between the two views. However, this information is allowed to have a rather high uncertainty. Let $\boldsymbol{w}$ be the pose of S'_L relative to S_L and let the uncertainty of this information be described by covariance matrix $\boldsymbol{C}_{\boldsymbol{w}}$. This information can be used to reject false correspondences in T_{ini}. Assuming that $\boldsymbol{w}$, $\boldsymbol{p}_i$ and $\boldsymbol{p}'_j$ are statistically independent, selection of the pairs from T_{ini} can be reduced to only those pairs (i, j) which satisfy the following condition

$$\boldsymbol{e}_{ij}(\boldsymbol{w})^T \Sigma_{ij}^{-1}(\boldsymbol{\phi}) \boldsymbol{e}_{ij}(\boldsymbol{w}) \leq \varepsilon_0, \tag{10}$$

where

$$\Sigma_{ij}(\boldsymbol{\phi}) = \boldsymbol{C}_{ij}(\boldsymbol{\phi}) + \boldsymbol{J}_j(\boldsymbol{\phi}) \boldsymbol{C}_{\boldsymbol{w}} \boldsymbol{J}_j^T(\boldsymbol{\phi}), \tag{11}$$

$$\boldsymbol{J}_j(\boldsymbol{\phi}) = \left[\left. \frac{\partial \left(\boldsymbol{R}(\psi) \boldsymbol{p}'_j \right)}{\partial \psi} \right|_{\psi = \boldsymbol{\phi}} \quad \boldsymbol{I}^{3 \times 3} \right]. \qquad (12)$$

Let the prior pose be denoted by $\boldsymbol{w}_{ini}$ and let its uncertainty be described by covariance matrix $\boldsymbol{C}_{\boldsymbol{w},ini}$. First pair in U can be selected from T_{ini} according to the constraint (10), where $\boldsymbol{w} = \boldsymbol{w}_{ini}$ and $\boldsymbol{C}_{\boldsymbol{w}} = \boldsymbol{C}_{\boldsymbol{w},ini}$. Assuming that this pair (i, j) represents a correct match, the coordinates $\boldsymbol{p}_i$ and $\boldsymbol{p}'_j$ as well as the initial pose $\boldsymbol{w}_{ini}$ can be used to update the pose information. The pose consistent with this information can be determined by minimizing the following cost function

$$\begin{aligned} \Im(\boldsymbol{w}) \quad = \quad & \sum_{(i,j) \in U} \boldsymbol{e}_{ij}(\boldsymbol{w})^T \boldsymbol{C}_{ij}^{-1}(\boldsymbol{\phi}) \boldsymbol{e}_{ij}(\boldsymbol{w}) + \\ & + \left(\boldsymbol{w}_{ini} - \boldsymbol{w} \right)^T \boldsymbol{C}_{\boldsymbol{w},ini}^{-1} \left(\boldsymbol{w}_{ini} - \boldsymbol{w} \right). \end{aligned} \qquad (13)$$

and the uncertainty of the obtained pose $\boldsymbol{w}$ can be described by the covariance matrix

$$\boldsymbol{C}_{\boldsymbol{w}} = \left(\sum_{(i,j) \in U} \boldsymbol{J}_j^T(\boldsymbol{\phi}) \boldsymbol{C}_{ij}^{-1}(\boldsymbol{\phi}) \boldsymbol{J}_j(\boldsymbol{\phi}) + \boldsymbol{C}_{\boldsymbol{w},ini}^{-1} \right)^{-1}. \qquad (14)$$

as proposed in [6]. Selection of the next pair of the set U is constrained to the pairs $(i, j) \in T_{ini}$ which satisfy (10), where $\boldsymbol{w}$ is the pose obtained in the previous step and the covariance matrix $\boldsymbol{C}_{\boldsymbol{w}}$ used in computation of Σ_{ij} is given by (14). This geometrically constrained hypothesis generation reduces the probability of selecting a false correspondence in the set U. Algorithm 1 with Algorithm 2 used instead of step 4 is referred to in this paper as geometrically constrained RANSAC (GCRANSAC).

Algorithm 2 Geometrically constrained hypothesis generation

Parameters: ε_0, u

Input: F, F', T_{ini}, $\boldsymbol{w}_{ini}$

Output: $\boldsymbol{w}$

```
1.  repeat
2.     $\boldsymbol{w} \leftarrow \boldsymbol{w}_{ini}$
3.     $V \leftarrow F$
4.     $V' \leftarrow F'$
5.     $U \leftarrow \emptyset$
6.     repeat
7.        Select randomly $\boldsymbol{p}'_j \in V'$ and remove it from $V'$.
8.        Form the set $X \subseteq V$ containing all $\boldsymbol{p}_i$ such that
          $(i, j) \in T_{ini}$ and the condition (10) is satisfied.
9.        if $X \neq \emptyset$ then
10.          Select randomly $\boldsymbol{p}_i \in X$ and remove it from
             $V$.
11.          Insert $(i, j)$ into $U$.
12.          Determine the pose $\boldsymbol{w}$ which minimizes (13).
13.          Compute $\boldsymbol{C}_{\boldsymbol{w}}$ by (14).
14.          if $U \nsubseteq W(\boldsymbol{w})$ then continue from line 1.
15.       end if
16.    until $|U| = u$ or $V' = \emptyset$
17. until $|W(\boldsymbol{w}) \cap U| = u$
18. return $\boldsymbol{w}$
```

4 TEST RESULTS

The proposed GCRANSAC has been tested on both synthetic and experimental data. The benefits of the introduced geometrically constrained hypothesis generation have been shown by the performance of GCRANSAC to the performance of std. RANSAC. All experiments presented herein are performed using a 3.40GHz Intel Pentium 4 Dual Core CPU with 2GB of RAM.

In the experiments reported in the following, the number u of pairs used to generate a pose hypothesis in GCRANSAC is chosen to be 5. The reason for using exactly that number of pairs is that five pairs of point features detected in the images taken from two distinct viewpoints by a single calibrated camera is sufficient for determining the relative camera pose [19].

The optimization in Step 12 of Algorithm 2 and Step 12 of Algorithm 1 is performed by a method based on the equations proposed in [6]. We have slightly modified those equations and reformulated them in the form of the Levenberg-Marquardt (LM) algorithm [14]. Since the optimization is performed u times for each hypothesis, its efficiency influences the performance of the GCRANSAC algorithm significantly. Efficient execution of Step 12 is achieved by stopping the optimization process when the condition (6) is satisfied. Only after the last pair is inserted into the set U, the optimization is allowed to proceed until the vicinity of a local minimum of the cost function is reached.

4.1 Evaluation on Synthetic Data

In order to quantitatively evaluate the proposed method, a set of synthetic data with appropriate properties was generated and used as input to both std. RANSAC and GCRANSAC. A set of 3D points $Z = \{\boldsymbol{z}_i, i = 1, \ldots, 3n\}$ was generated using a pseudorandom number generator. All points were inside the field of view of the left camera of a virtual stereo camera system. The resolution of the camera was 320×240 pixels. The camera was assigned a reference frame S_L centered in the optical center of the camera with z-axis identical to the optical axis. The z-coordinate of the points $\boldsymbol{z}_i$ with respect to the reference frame S_L was uniformly distributed over the range of 2 to 6 meters. The first $2n$ points from the set Z were then projected onto the stereo image. The image coordinates of the obtained image projections were perturbed by Gaussian noise with zero mean and variance $\delta = 1\,\text{pix}^2$. After performing stereo reconstruction, a set of 3D point features $F = \{\boldsymbol{p}_i, i = 1, \ldots, 2n\}$ was obtained, where $\boldsymbol{p}_i$ was computed from the perturbed stereo image projection of $\boldsymbol{z}_i$ by triangulation. Each feature $\boldsymbol{p}_i$ was assigned a covariance matrix $\boldsymbol{C}_{\boldsymbol{p},i}$ describing its uncertainty due to the influence of the noise to the stereo reconstruction. This uncertainty was determined using the approach proposed in [16].

The pseudorandom number generator was also used to generate a vector $\boldsymbol{w} = [\ \boldsymbol{\phi}^T \quad \boldsymbol{t}^T\ ]^T$ representing the pose of the second view camera reference frame S'_L relative to the reference frame S_L. The discussed algorithm assumes that the pose of S'_L relative to S_L is a random variable having Gaussian distribution with mean $\boldsymbol{w}_{ini}$ and covariance $\boldsymbol{C}_{\boldsymbol{w},ini}$. If this was the case, the probability that the condition

$$(\boldsymbol{w} - \boldsymbol{w}_{ini})^T \boldsymbol{C}_{\boldsymbol{w},ini}^{-1} (\boldsymbol{w} - \boldsymbol{w}_{ini}) \leq \chi^2_{.99}, \qquad (15)$$

is satisfied would be 99%, where $\chi^2_{.99} = 16.81$ is the value of the Chi-Square distribution with 6 degrees of freedom [20]. In order to demonstrate that the performance of the proposed approach does not depend significantly on that assumption, in the experiments with synthetic data a uniform distribution of the pose $\boldsymbol{w}$ was used. Poses were randomly selected from the set of vectors $\boldsymbol{w} \in \mathbb{R}^6$ which satisfy (15) where $\boldsymbol{w}_{ini} = \boldsymbol{0}$ and

$$\boldsymbol{C}_{\boldsymbol{w},ini} = \begin{bmatrix} (4^\circ)^2 \cdot \boldsymbol{I}^{3\times3} & \boldsymbol{0} \\ \boldsymbol{0} & (0.2\,m)^2 \cdot \boldsymbol{I}^{3\times3} \end{bmatrix}.$$

By transforming the first n and the last n points from Z to the reference frame S'_L, projecting them onto the second stereo image and performing stereo reconstruction using the procedure described above, a second view feature set F' was obtained.

Notice that the first n points from the total of $3n$ points in Z were projected onto the stereo images of both views, the second n points were projected only onto the first stereo image and the last n points were projected only onto the second stereo image. This way, we wanted to simulate situations in which 50% of the features detected in the first view are not detected in the second view and vice versa. In the experiments reported in this paper, $n = 100$.

As explained in Section 3, feature-based image registration considered herein starts from preliminary information about the correspondence between the features detected in two views. The feature correspondence is in this article represented by a set T_{ini} of pairs (i, j) where $\boldsymbol{p}_i$ and $\boldsymbol{p}'_j$ are corresponding features. In order to compare the performances of the considered algorithms in the case of highly ambiguous feature correspondences, the experiments were performed with T_{ini} containing intentionally introduced false feature correspondences. All false correspondences were selected among the pairs (i, j) for which the condition (10) is satisfied where $\boldsymbol{w} = \boldsymbol{w}_{ini}$ and $\boldsymbol{C}_{\boldsymbol{w}} = \boldsymbol{C}_{\boldsymbol{w},ini}$. Hence, the prior pose information did not help in selecting the first feature pair in hypothesis generation procedure. In the experiments reported below, $\boldsymbol{w}_{ini} = \boldsymbol{0}$.

In the case of GCRANSAC, the threshold for evaluation of the conditions (6) and (10) was set to $\varepsilon_0 = 11.34$ which is the 99% value of the Chi-Square distribution with 3 degrees of freedom [20]. Since computation of the pose in the step 4 in the case of std. RANSAC is performed by minimization of a criterion which does not consider directional uncertainty of 3D points obtained by stereo reconstruction, the uncertainty of the obtained pose is expected to be relatively high. A proper way of dealing with this uncertainty would be to use the condition (10) instead of (6), where $\boldsymbol{C}_{\boldsymbol{w}}$ is computed from the uncertainties of the positions of the considered points. However, the computation of Σ_{ij} would increase the computational cost of the step 5 of Algorithm 1. Instead, we chose to relax the condition (6) by multiplying $\boldsymbol{C}_{ij}$ by a factor of 4. This modification has shown to improve the performance of std. RANSAC significantly.

Let r be the percentage of false correspondences in T_{ini}. GCRANSAC is expected to have better performance than std. RANSAC for higher values of r. In order to estimate for which r is advantageous to use GCRANSAC, the comparison of the two algorithms was performed for the values of r between 20% and 80% with step of 10%. For each of these values, 1100 input data sets were generated using random number generator as explained above. The first 1000 of these 1100 data sets were used for the evaluation of the algorithm performance and the last 100 were used to determine the execution time of the compared algorithms, as explained in the following.

The hypothesis generation step of GCRANSAC is much more computationally expensive then the hypothesis generation step of std. RANSAC. On the other hand, the probability of making a correct hypothesis by the proposed geometrically constrained procedure is much higher then by purely random sampling. In order to make a 'fair' comparison of GCRANSAC and std. RANSAC, Algorithm 1 is modified so that the main loop (steps 3 - 11) is exited after a limited time t_{loop} has elapsed. That time is set to the same value for both algorithms. Determining the optimal t_{loop} by a theoretical analysis would be a rather difficult task. Hence, we performed an experimental estimation of this parameter. During the experiments with GCRANSAC it has been noticed that in most cases it gives a good result when 3 or more correct hypotheses, i.e. the hypothesis based on 5 correct matches, are posed. The time needed

Table 1. Results of the experiments on synthetic data

$r(\%)$	$e_{\phi,99.5\%}(\circ)$		$e_{t,99.5\%}$(m)		$n_{99.5\%}$		$t_{loop}(s)$
	STD	GC	STD	GC	STD	GC	
20	0.8	0.6	0.05	0.04	95	95	0.020
30	0.8	0.6	0.06	0.05	95	95	0.025
40	1.2	0.6	0.08	0.04	90	96	0.032
50	7.2	0.6	0.44	0.04	40	95	0.043
60	2.9	0.6	0.22	0.04	76	95	0.087
70	19.2	0.6	1.35	0.05	18	94	0.084
80	37.7	0.7	2.59	0.07	5	91	0.160

to obtain a correct hypothesis increases with the percentage of false matches r. Therefore, for each considered r, GCRANSAC was applied to the last 100 of the total of 1100 generated data sets and the maximum time needed to obtain 3 correct hypotheses is considered to be a suitable choice for t_{loop}.

The performances of the evaluated algorithms were assessed by considering the orientation error e_{ϕ} and the position error e_{t} defined as follows. The orientation error represents the angle for which the reference frame $\hat{S}'_L$, whose pose relative to S_L is computed by the considered algorithm, must be rotated around a particular axis in order to fit the true reference frame S'_L. The position error represents the Euclidean distance between the estimated and the true reference frame S'_L. Another performance index used in this evaluation is the number of correct matches. Tab. 1 contains the error limits and the minimum number of correct matches for the best 99.5% results of each algorithm. For example, the values in the row denoted by $r = 60\%$ and the columns denoted by $e_{t,99.5\%}$ indicate that for $r = 60\%$ the position error obtained by GCRANSAC was below 0.04 m and the position error obtained by std. RANSAC was below 0.22 m in 99.5% of trials. Analogously, the values in the same row and the columns denoted by $n_{99.5\%}$ indicate that for the same percentage of false correspondences, GCRANSAC provided at least 95 and std. RANSAC at least 76 out of 100 possible correct matches in 99.5% of trials. From the experimental results presented in Tab. 1 it can be concluded that, under the conditions considered in the conducted experiments, GCRANSAC performs noticeably better than std. RANSAC when $r > 50\%$.

4.2 Experiments with a Stereo Vision System

Two experiments with images taken by a camera system Videre design STH-MDCS2-VAR were performed. In the first experiment, the aforementioned stereo camera system was mounted on a mobile robot P3DX navigating in indoor environment and in the second experiment, the images were taken by the hand-held camera system in an outdoor environment. Two sample images, one from each experiment, are shown in Fig. 2. A total of 395 samples were acquired during the first experiment and 97 during the second experiment. Features were detected by the SIFT-algorithm [17] and a Small Vision System (SVS) [21] was used to compute the disparity map of the first and the second stereo image. The disparities assigned to the anchor points of the features detected by SIFT were used to determine their 3D coordinates. This way, the feature sets F and F' were obtained for each two consecutive stereo image pairs. The initial feature correspondences T_{ini} were obtained by comparing the local descriptors assigned to the detected features by the SIFT-algorithm. This method showed to be very reliable and resulted in a relatively low percentage of false correspondences. Nevertheless, its computational cost was rather high.

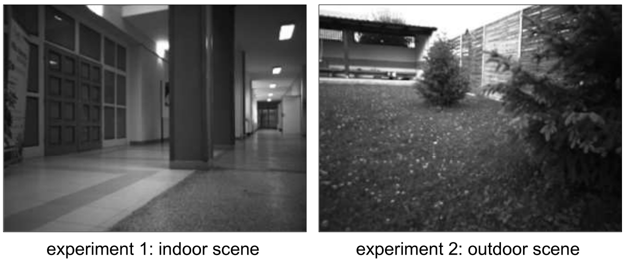


Fig. 2. Sample images from the indoor and outdoor experiment

In order to obtain reference data to be used as the ground truth, std. RANSAC with t_{loop} set to 5 seconds was applied to the sets F, F' and T_{ini}. Such a long execution time was expected to provide highly reliable stereo image registration. The feature correspondences obtained by this procedure were considered to be correct, since they were confirmed by both SIFT matching and mutual geometrical consistency. This way a reference set of feature correspondences T_{ref} is created for each two consecutive stereo images.

The evaluation of GCRANSAC algorithm as well as std. RANSAC was performed by applying these algorithms to the sets F, F' and T_{ini} obtained from all considered stereo images and comparing the obtained feature correspondences to the ground truth. A correspondence obtained by one of these two algorithms was considered correct if it was contained in the reference set T_{ref}.

The performance index used to compare the considered algorithms is the percentage of missed correspondences computed by

$$\mu = 100\frac{|T_{ref}| - |T_{fin} \cap T_{ref}|}{|T_{ref}|},$$

where $|\cdot|$ denotes the number of elements in a set and T_{fin} is the set of correspondences obtained by the evaluated algorithm. Lower value μ indicates better matching result. If $\mu = 0$, then all correct feature pairs are successfully detected. The number of false matches was insignificantly low. Time t_{loop} was set to 20 ms.

The experimental results are presented by the normalized cumulative histograms shown in Fig. 3, where each value μ_0 on the x-asix is assigned the percentage of samples for which $\mu \leq \mu_0$. In the first experiment, both algorithms had similar performance. In this case, GCRANSAC showed no improvement over std. RANSAC. This can be explained by a very low percentage of false correspondences in T_{ini}. On the other hand, GCRANSAC gave slightly better results in the second experiment. It can be seen that in the case of GCRANSAC for more than 97% of samples μ was at most 10%, while in the case of std. RANSAC, the same upper bound of μ was achieved for approximately 91% of samples.

It should be mentioned, that the false feature pairs were not always the result of the limitations of the applied SIFT-based matching. It can also happen that one of two features which are correctly matched according to their SIFT-descriptors is assigned a false disparity by the applied stereo reconstruction method and therefore this feature pair does not satisfy the geometric constraint.

5 CONCLUSION

In this paper, an approach for registration of sparse feature sets detected in stereo images taken from two different views is proposed. The special focus of this work is robustness in presence of high ambiguity in feature correspondences in the input data set which is achieved by applying a geometrically constrained form of the RANSAC procedure.

In the experiments with a stereo vision system, the feature detection and matching was implemented by SIFT resulting in a relatively low percentage of false correspondences and therefore the benefits of the proposed strategy could not be clearly demonstrated. Nevertheless, although in most cases the performance of the proposed method was similar to the performance of the standard RANSAC, for particular samples, where the input data contained a relatively high percentage of false feature correspondences, the proposed approach showed to be more reliable. The significant advantage of the proposed geometrically constrained hypothesis generation over purely random sampling in the case of multiple possible correspondences and very high percentage of false feature correspondences was clearly demonstrated by the experiments on synthetic data. This indicates the possibility of using the proposed geometrically constrained RANSAC approach for tracking of features which are less distinctive in comparison to those obtained by SIFT, but which can be extracted from the image much faster.

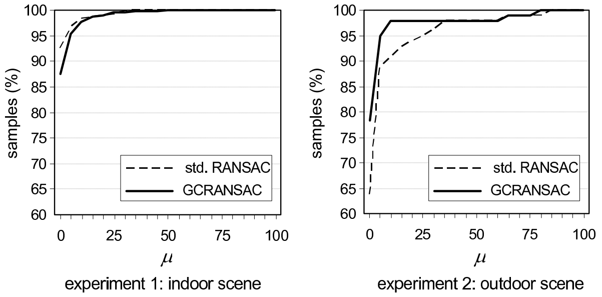


Fig. 3. Normalized cumulative histograms of the percentage of missed correspondences for the indoor and outdoor experiment

ACKNOWLEDGMENT

This work was supported by the Ministry of Science, Education and Sports of the Republic of Croatia under grant No. 036-0363078-3018

REFERENCES

[1] K. Mikolajczyk and C. Schmid, **A Performance Evaluation of Local Descriptors**, IEEE Trans. Pattern Anal. Machine Intell., vol. 27, no. 10, pp. 1615–1630, 2005.

[2] K. S. Arun, T. S. Huang, and S. D. Blostein, **Least-Squares Fitting of Two 3-D Point Sets**, IEEE Trans. Pattern Anal. Machine Intell., vol. 9, no. 5, pp. 698–700, September 1987.

[3] B. K. P. Horn, **Closed-Form Solution of Absolute Orientation using Unit Quaternions**, Journal of the Optical Society of America, vol. 4, pp. 629–642, April 1987.

[4] N. Ohta and K. Kanatani, **Optimal Estimation of Three-Dimensional Rotation and Reliability Evaluation**, in Proc. European Conference on Computer Vision (ECCV), pp. 175–187, Freiburg, Germany, June 1998.

[5] B. Matei and P. Meer, **Optimal Rigid Motion Estimation and Performance Evaluation with Bootstrap**, in Proc. IEEE International Conference on Computer Vision and Pattern Recognition (CVPR), pp. 339–345, Fort Collins, Colorado, June 1999.

[6] L. H. Matthies, **Dynamic Stereo Vision**, PhD thesis, Carnegie Mellon University, USA, October 1989.

[7] C. F. Olson, L. H. Matthies, M. Schoppers, and M. W. Maimone, **Rover navigation using stereo ego-motion**, Robotics and Autonomous Systems, vol. 43, pp. 215–229, 2003.

[8] M. A. Fischler and R. C. Bolles, **Random Sample Consensus: A Paradigm for Model Fitting with Applications to Image Analysis and Automated Cartography**, Graphics and Image Processing, vol. 24, no. 6, pp. 381–395, 1981.

[9] M. Sarkis, K. Diepold, and K. Hüper, **A Fast and Robust Solution to the Five-Point Relative Pose Problem Using Gauss-Newton Optimization on a Manifold**, in Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP), pp. 681–684, Honolulu, Hawaii, USA, April 2007.

[10] D. Nistér, O. Naroditsky, and J. Bergen, **Visual Odometry for Ground Vehicle Applications**, Journal of Field Robotics, vol. 23, no. 1, January 2006.

[11] N. Sünderhauf and P. Protzel, **Towards Using Sparse Bundle Adjustment for Robust Stereo Odometry in Outdoor Terrain**, in Proc. Towards Autonomous Robotic Systems (TAROS), pp. 206–Ű213, Guildford, UK, 2006.

[12] N. Ayache and O. D. Faugeras, **Maintaining Representations of the Environment of a Mobile Robot**, IEEE Trans. on Robotics and Automation vol. 5, no. 6, pp. :804–819, December 1989.

[13] A. Kosaka and A. C. Kak, **Fast Vision-Guided Mobile Robot Navigation Using Model-Based Reasoning and Prediction of Uncertainties**, CVGIP: Image Understanding, vol. 56, no. 3, pp. 271–329, November 1992.

[14] D. W. Marquardt, **An Algorithm for Least-Squares Estimation of Nonlinear Parameters**, SIAM Journal of Applied Mathematics, vol. 11, no. 2, pp. 431–441, 1963.

[15] R. Cupec, A. Kitanov, and I. Petrović, **Stereo Image Registration Based on RANSAC with Selective Hypothesis Generation**, in Proceedings of 40th International Symposium on Robotics, pp. 265–270, Barcelona, Spain, 2009.

[16] L. Matthies and S. A. Shafer, **Error Modeling in Stereo Navigation**, IEEE Journal of Robotics and Automation, vol. 3, no. 3, pp. 239–248, June 1987.

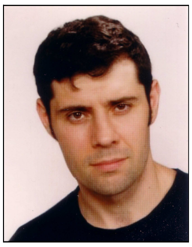
[17] D. G. Lowe, **Distinctive Image Features from Scale-Invariant Keypoints**, The International Journal of Computer Vision, vol. 60, no. 2, pp. 91–110, 2004.

[18] H. Bay, T. Tuytelaars, and L. Van Gool, **SURF: Speeded Up Robust Features**, in Proc. European Conference on Computer Vision (ECCV), pp. 404–417, Graz, Austria, May 2006.

[19] D. Nistér, **An Efficient Solution to the Five-Point Relative Pose Problem**, IEEE Trans. Pattern Anal. Machine Intell., vol. 26, no. 6, pp. 756–770, June 2004.

[20] M. R. Spiegel, J. J. Schiller, and R. A. Srinivasan, **Schaum's Outline of Theory and Problems of Probability and Statistics**, Second edition, The McGraw-Hill, USA, 2000.

[21] K. Konolige, **Small Vision Systems: Hardware and Implementation**, In Robotics Research: The 8th International Symposium, pp. 203–212, Springer-Verlag, 1997.



Robert Cupec graduated at the Faculty of Electrical Engineering, University of Zagreb in 1995, where he received the M.Sc. degree in 1999. Upon graduation he joined the Department of Control and Computer Engineering in Automation, University of Zagreb, where he was employed until 2000. From 2000 to 2004 he is working at the Institute of Automatic Control Engineering, Technische Universität München where he received the Ph.D. degree in 2005. Currently, he is employed as an assistant professor at the Faculty of Electrical Engineering, University of Osijek, Croatia. His main research interest is in robot vision.

Emmanuel Karlo Nyarko received his B.Sc. degree in 2001 and M.Sc. degree in 2005, all in Electrical Engineering from the Faculty of Electrical Engineering (ETF Osijek), University of Osijek, Croatia. He was employed as a research and teaching assistant at the Faculty of Electrical Engineering, University of Osijek from 2001 to 2005. From 2005 until 2007, he was employed as a software development engineer at Mono Ltd., a privately owned software company in Osijek. Since 2008, he is a Ph.D. student at the Faculty of Electrical Engineering, University of Osijek and is currently employed at the same faculty as a teaching assistant in undergraduate courses in the field of control systems and robotics.

Andreja Kitanov received B.Sc. degree in 2004. in electrical engineering from the Faculty of Electrical Engineering and Computing (FER Zagreb), University of Zagreb, Croatia. Currently, she is a research assistant at Department of Control and Computer Engineering, FER, Zagreb, where she is doing a doctoral thesis. She has been a member of IEEE since 2004. Her research interests include localization and mapping, stereo computer vision and autonomous mobile robots operating in unstructured environments.

Ivan Petrović received B.Sc. degree in 1983, M.Sc. degree in 1989 and Ph.D. degree in 1998, all in Electrical Engineering from the Faculty of Electrical Engineering and Computing (FER Zagreb), University of Zagreb, Croatia. He had been employed as an R&D engineer at the Institute of Electrical Engineering of the Konèar Corporation in Zagreb from 1985 to 1994. Since 1994 he has been with FER Zagreb, where he is currently the head of the Department of Control and Computer Engineering. He teaches a number of undergraduate and graduate courses in the field of control systems and mobile robotics. His research interests include various advanced control strategies and their applications to control of complex systems and mobile robots navigation. Results of his research effort have been implemented in several industrial products. He is a member of IEEE, IFAC - TC on Robotics and FIRA - Executive committee. He is a collaborating member of the Croatian Academy of Engineering.

AUTHORS' ADDRESSES

Asst. Prof. Robert Cupec, Ph.D.
Emmanuel Karlo Nyarko, M.Sc.
Faculty of Electrical Engineering,
University of Osijek,
HR-31000 Osijek, Croatia
emails: robert.cupec@etfos.hr,
nyarko@etfos.hr

Andreja Kitanov, M.Sc.
Prof. Ivan Petrović, Ph.D.
Department of Control and Computer Engineering,
Faculty of Electrical Engineering and Computing,
University of Zagreb,
Unska 3, HR-10000 Zagreb, Croatia
emails: andreja.kitanov@fer.hr,
ivan.petrovic@fer.hr

Received: 2009-11-02
Accepted: 2009-11-27