

ROBUST SPARSE MATCHING AND MOTION ESTIMATION USING GENETIC ALGORITHMS

M. Shahbazi^{a,*}, G. Sohn^b, J. Théau^a, P. Ménard^c

^a Dept. of Applied Geomatics, Université de Sherbrooke, Boul. de l'Université, Sherbrooke, Québec, Canada - (mozhdeh.shahbazi, jerome.theau)@usherbrooke.ca

^b Dept. of Geomatics Engineering, York University, Keele Street, Toronto, Ontario, Canada- gsohn@yorku.ca

^c Centre de géomatique du Québec, Saguenay, Québec, Canada- pmenard@cgq.qc.ca

Commission III, WG III/4

KEY WORDS: Genetic Algorithm, Structure from Motion, Epipolar Geometry, Image Matching, Outlier Detection

ABSTRACT:

In this paper, we propose a robust technique using genetic algorithm for detecting inliers and estimating accurate motion parameters from putative correspondences containing any percentage of outliers. The proposed technique aims to increase computational efficiency and modelling accuracy in comparison with the state-of-the-art via the following contributions: i) guided generation of initial populations for both avoiding degenerate solutions and increasing the rate of useful hypotheses, ii) replacing random search with evolutionary search, iii) possibility of evaluating the individuals of every population by parallel computation, iv) being performable on images with unknown internal orientation parameters, iv) estimating the motion model via detecting a minimum, however more than enough, set of inliers, v) ensuring the robustness of the motion model against outliers, degeneracy and poor-perspective camera models, vi) making no assumptions about the probability distribution of inliers and/or outliers residuals from the estimated motion model, vii) detecting all the inliers by setting the threshold on their residuals adaptively with regard to the uncertainty of the estimated motion model and the position of the matches. The proposed method was evaluated both on synthetic data and real images. The results were compared with the most popular techniques from the state-of-the-art, including RANSAC, MSAC, MLESAC, Least Trimmed Squares and Least Median of Squares. Experimental results proved that the proposed approach perform better than others in terms of accuracy of motion estimation, accuracy of inlier detection and the computational efficiency.

1. INTRODUCTION

Sparse image matching is the main stage in many computer vision applications, including structure from motion (SfM) and robot navigation. Sparse matching is performed only on parts of stereo images, in which sufficient amount of information is available to achieve a meaningful comparison between images. Recent researches apply sparse matching to address a variety of problems. For instance, feature-based techniques of simultaneous localization and mapping (SLAM) rely on sparse matching (Hartmann et al., 2013). Also, in the indirect dense SLAM approaches, the tracking task is still based on sparse matching (Wendel et al., 2012). Besides, using sparse matches for external camera calibration provides the basis for almost all the dense matching techniques, which are based on rectified images (Wöhler, 2013). Real-time mosaicing of aerial images, particularly the images acquired from unmanned aerial vehicles, is also another important application, which applies the sparse matches for computing the sequential homographies between the images (Kekec et al., 2014).

Since the sparse matching plays an important role in the mentioned applications, special care should be taken to control the accuracy of matching. Sparse matching techniques can be divided to two main categories of correlation-based and feature-based approaches. The feature-based approach is recently gaining more popularity for addressing the sparse matching problem, because the recently-developed feature descriptors are

invariant to many sorts of geometric and photometric transformations. However, even the results of feature-based matching are usually contaminated with a considerable percentage of outliers. From now on, the term "putative correspondences" is used to refer to the raw results of sparse matching, since they are the presumed matches possibly containing many outliers. The presence of outliers among putative correspondences is due to several factors. These factors include noisy measurements, insufficiency of local descriptors, and lack of distinctive texture patterns in the scene and existence of repetitive textures that cause high ambiguity in matching. In addition, the threshold value, which decides whether two points are similar enough to be matched, has an important impact on the amount of the outliers (Li and Allinson, 2008). As a result, outlier detection should naturally be integrated into the sparse matching procedure to avoid the outliers affecting the accuracy of motion estimation and scene reconstruction.

Generally, outlier detection techniques are based on the fact that correct matches (inliers) have some spatial characteristics in common. Therefore, the matches which are not consistent with such spatial characteristics can be classified as outliers. Consequently, the techniques of outlier detection can be divided to two categories based on the spatial characteristic that they use for detecting the inliers: i) the techniques based on epipolar geometry, ii) the techniques based on other spatial characteristics. The state-of-the-art with respect to these two

* Corresponding author

categories is discussed in subsections 1.1 and 1.2. Accordingly, the main focus and contributions of this paper are presented in subsection 1.3.

1.1 Inlier Detection Based on Epipolar Geometry

The idea of using epipolar geometry as the constraint to detect the inliers has been proposed in several studies. With this regard, the matching problem turns into two problems of i) robust estimation of epipolar geometry (motion model) from putative correspondences, ii) detecting the entire set of inliers using the estimated motion model as the constraint; i.e. the matches not fitting to the estimated motion model are the outliers. The techniques of robust epipolar geometry estimation can themselves be divided to two categories: i) the techniques that search the discrete space of matches to find the minimum number of inliers required for calculating the epipolar geometry, ii) the techniques that directly search the continuous space of the orientation parameters. The related work with respect to each of these categories is discussed in the following paragraphs.

Random Sample Consensus (RANSAC) techniques are popular hypothesize-and-verify approaches in the field of robust estimation. Basically, RANSAC aims to determine the optimal motion model from an uncontaminated sample set by maximizing its support size. Then, the inliers can be detected as matches whose residuals from the estimated motion model are less than a given threshold. To find an uncontaminated sample set, successive random sampling from the putative correspondences is performed. By this definition, five major questions are involved in RANSAC-like techniques. i) Is maximizing the support cardinality a reasonable objective function when no information about the amount of the outliers is available? ii) How to handle the large number of random samples (number of RANSAC iterations) in cases where the inlier ratio is either unknown or very small? iii) How to ensure the robustness of the estimated motion model against the influence of the noise since it is relying on a minimal, just enough, subset of inliers? iv) How to control the effect of the degenerate solutions which naturally maximize the support cardinality? v) Does the threshold used to detect all the inliers reflect the uncertainty of estimated motion model as well? Some of these questions are partly answered by different variants of RANSAC, none of which actually try to address all these questions.

Unlike the standard RANSAC, there are improved variants which use robust objective functions to determine the support cardinality. In MSAC, a version of RANSAC inspired by M-estimators, inliers are scored based on their fitness and outliers are scored with a non-zero constant penalty. MLESAC algorithm maximizes the log-likelihood of the solution via RANSAC process by assuming that the outliers are distributed uniformly and the residuals distribute a Gaussian function over inliers. MAPSAC is also a refined version of MLESAC with Bayesian model estimation. Such objective functions make motion models with similar inlier scores more distinguishable. However, their main drawback is that they make certain strict assumptions about the distribution of the residuals either for inliers or outliers.

The sampling strategy in RANSAC is very crucial as it influences the efficiency of the algorithm with respect to the number of RANSAC iterations and degeneracy of the estimated model. To control the first factor (speed of the algorithm), two

strategies can be taken. The first one is to enforce initial consistency check on the samples for improving the quality of the hypotheses. PROSAC method by Chum and Matas (2005) and GroupSAC method by Ni et al. (2009) are the examples of such a strategy. The second strategy proposed in the literature to reduce the number of RANSAC iterations is to reduce the solution space by only verifying the hypotheses with higher probability of being optimal. These high-probable hypotheses can be selected by $T_{d,d}$ test (Matas and Chum, 2004) or bail-out test (Capel, 2005). In addition, the hypothesis verification can be performed preemptively in a breadth-first manner only for a fixed number of sample sets (Nistér, 2005). The main strategy to control the degeneracy of the solutions is to investigate the support of the current best solution either locally or globally. Lo-RANSAC method by Chum et al. (2003) and QDEGSAC proposed by Frahm and Pollefeys (2006) are the examples of such techniques. The main issue concerned with these techniques is that, besides adding operations to the original RANSAC algorithm, they mostly require supplementary information of the scene or the matches.

As mentioned earlier, there are studies in the literature for robust epipolar geometry estimation that propose to search the continuous space of orientation parameters directly. The main issue with such techniques is that the space of orientation parameters are not bounded, specifically talking about translations. To overcome this drawback, the search space can be re-parameterized as a rotation space. Taraglio and Chiesa (2011) propose a genetic approach for estimating the epipolar geometry by directly searching the space of five relative orientation parameters. However, the continuous, 5-dimensional search-space reduces the chance of reaching the optimal solution with evolutionary search, specifically when high accuracy is required. Branch-and-bound technique has been used in the literature to efficiently search such a 5-dimensional rotation space (Hartley and Kahl, 2007). The main issue concerning such techniques is their vulnerability to large percentage of outliers (Yang et al., 2014). Besides, these techniques require that the cameras are intrinsically calibrated. This is due to the fact that they are only able to estimate the essential matrix and not the fundamental matrix. As a result, they are not suitable for image collections from unknown sources (e.g., images from internet).

1.2 Inlier Detection Based on Other Spatial Characteristics

In addition to the techniques of inlier detection using epipolar geometry, there are several methods which are based on other spatial criteria. Here, the related work reflecting the main ideas of such methods is reviewed, and the reasons why they are not practical in general cases are discussed. Accordingly, the inlier detection methods based on robust epipolar geometry estimation are more general than the methods discussed in this subsection; because they can be used for various types of images without being concerned about their photometric characteristics.

Radhika et al. (2009) suggest detecting the outliers as the matches that have disparity values noticeably far from the mode of the disparity values of their local neighbours. Determining a reliable threshold and size of the local neighborhood is a concerning factor in this approach. Hasler et al. (2003) believe that outlying matches follow a pattern similar to the one that is generated by comparing two random regions of the images. This pattern is used to predict the error distribution among the outliers. However, this method may fail to define a proper error

distribution in images with relatively high homogeneity in one region and high diversity in another region. Adam et al. (2001) suggest that the lines between correct matches follow a typical orientation if the feature points in one image are rotated by a certain angle. The matches whose joining lines do not make such an orientation are recognized as outliers. However, the mentioned assumption is true for images taken by narrow-angle cameras. Likewise, Wang et al. (2010) suggest using the global parallax histograms of distances and angles between the matches to detect outliers. They assume that the peaks of the histograms correspond to the average distance and angle between correct matches. Therefore, outliers produce distances and angles, which are considerably far from these peaks. However, such methods may fail when the scene objects are located at considerably varying depths and orientations relative to the cameras. In such situations, neither the distance histogram nor the angle histogram does provide a unique peak.

1.3 The Power of Genetic Algorithm

In this paper, we focus on the problem of detecting inliers from putative correspondences based on the robust estimation of the epipolar geometry. To this end, we adapt the integer-coded genetic algorithm (GA) to accurate motion estimation and inlier detection. The proposed technique can be considered as an extension of RANSAC-like methods where random search is replaced with evolutionary search. The evolutionary search has two major advantages. First, the search is performed on a population of feasible solutions instead of a single initialization. Second, new samples are produced considering the results obtained via the previous samples. This causes the search algorithm both to locally improve the previous solutions (exploitation) and to globally search for new solutions (exploration) (Haupt and Haupt, 2004). The genetic operators considered in this paper are i) integer crossover, which combines the position of the matches in the previous samples to reproduce new samples, ii) mutation, which modifies each solution locally to test new samples at the neighbourhood of previous samples, iii) random exploration, which explore the entire search space without considering the previous solutions to avoid local optima, and iv) replacement, which determines a new population of hypotheses from both the best solutions at the previous population (elites) and the genetically reproduced solutions. Each operator is specifically adapted to the requirements of the inlier detection process.

Besides the evolutionary search, we have used the power of genetic encoding scheme for encoding the matches as a meaningful array of integers, instead of a random permutation of integers. With this regard, each match is labelled with two integers representing its position. Therefore, the genetic operators become efficiently meaningful. For example, the mutation operator modifies the horizontal and vertical position of the matches locally. In addition, instead of random sampling for exploring the search space, a guided sampling scheme is proposed. This scheme is based on the spatial distribution of the matches; thus, it is useful to avoid degenerate and locally optimum solutions.

Furthermore, we have defined the cost function in genetic algorithm not only based on the residuals at the minimal sample set of inliers (from which the epipolar geometry is initially estimated), but also based on the fitness of the estimated model to the least number of inliers in the dataset, which we call the inlier-set of minimum cardinality. We have also considered a simple technique to ensure that the robust estimation of the

epipolar geometry is performable for poor-perspective camera models as well.

Once the genetic algorithm is terminated, the inlier-set of minimum cardinality and the re-estimated motion model using this set are determined. So, the last task is to calculate a threshold value to detect all the inliers among putative correspondences based on their residuals from the estimated motion model. To this end, the uncertainty of the estimated motion model and spatial configuration of the inlier-set are taken into account. Therefore, the threshold is determined adaptively without enforcing any probability distribution on the inliers residuals.

The rest of the paper is organized as follows. The main problem of simultaneous inlier-detection and motion estimation is formulated in section 2. Section 3 describes the solution using the genetic algorithm, which is followed by the method of detecting all the inliers in section 4. The experimental results are discussed in section 5, and the conclusion is presented in section 6.

2. PROBLEM FORMULATION

In this study, the motion model is represented by the fundamental matrix. Therefore, the problem is formulated as estimating the accurate fundamental matrix by minimizing the sum of squared Sampson errors over an inlier-set of minimum cardinality. This minimum cardinality can be easily hypothesized without loss of generality. For instance, considering a data set with 60% outliers in reality, one who has no knowledge of the error scale can assume that at least 10% of the data are inliers. Generally, the least number of inliers would be a more relaxed assumption than the approximate ratio of outliers in the dataset. Therefore, the problem is first to find a minimal set of inliers from which the fundamental matrix is calculated. Then, the next step is to evaluate the optimality of the estimated fundamental matrix based on the squared Sampson distances that are calculated for all the matches. A fundamental matrix is optimal only and only if it minimizes the sum of squared Sampson errors over an inlier-set of minimum cardinality, for example over the best 10% of the dataset.

The fundamental matrix (F) is defined by equation (1), assuming that the world-coordinate system coincides with the first camera coordinate system:

$$F = K_2^T [t]_{\times} R K_1, \quad (1)$$

where K_2 and K_1 are the calibration matrices of the second and the first camera respectively. (R, t) represent the rigid body transformation (rotation and translation) from the second camera coordinate system to the first one, and $[t]_{\times}$ is the anti-symmetric matrix defined from t . Considering the fact that $[t]_{\times}$ is of rank 2, F is also a rank-2 matrix with zero determinant. Considering a fundamental matrix F , the Sampson distance, d , can be calculated for the corresponding points \mathbf{x}, \mathbf{x}' as the measure of the re-projection error (Torr and Murray, 1997).

$$d^2(F) = (\mathbf{x}'^T F \mathbf{x})^2 / (l_1^2 + l_2^2 + l_3^2 + l_4^2) \quad (2)$$

where:

$$[l_1, l_2, l_3]^T = F \mathbf{x} \quad \text{and} \quad [l_1', l_2', l_3']^T = \mathbf{x}'^T F$$

Within the scheme of this study, each match is labelled with two integers based on its position on the overlapping area of images, one integer for its horizontal and the other for its vertical position relative to other points. This labelling fashion

is justified later in section 3. Therefore, the whole set of matches can be represented by a label-set S as $S = \{k(i, j) | 1 \leq k \leq N, 1 \leq i \leq W, 1 \leq j \leq H\}$, where W and H represent the vertical and horizontal extent of the matches, and N is the total number of matches in the dataset. Accordingly, a feasible solution for estimating the fundamental matrix can be represented as a set Ω with the following definition and constraints:

$$\Omega = \{m_1, \dots, m_{12} \mid \begin{matrix} m_i \in S \\ \forall i \in \{1:12\} \end{matrix} \text{ and } \begin{matrix} m_i \neq m_j \\ \forall i \in \{1:12\} \forall j \in \{1:12\}, j \neq i \end{matrix} \} \quad (3)$$

Every corresponding pair, m_i , is composed of two points, the point x_{m_i} on the left image and its corresponding x'_{m_i} on the right image. The matches in a feasible solution set are substituted in equation (4) to determine the fundamental matrix. In equation (4), the coordinates u_{m_i} and u'_{m_i} are the normalized homogeneous coordinates calculated from x_{m_i} and x'_{m_i} in the normalization process using transformation matrices T and T' respectively (Hartley, 1997).

$$U(\bar{F}, \Omega) = \begin{cases} u'_{m_i}{}^T \bar{F} u_{m_i} = 0 & \text{for } i \in \{1:12\} \\ \det(\bar{F}) = 0 \end{cases} \quad (4)$$

Where:

$$\bar{F} = \begin{bmatrix} \bar{F}_{11} & \bar{F}_{12} & \bar{F}_{13} \\ \bar{F}_{21} & \bar{F}_{22} & \bar{F}_{23} \\ \bar{F}_{31} & \bar{F}_{32} & \bar{F}_{33} \end{bmatrix} = T'^{-T} F T^{-1}$$

The coefficients of the fundamental matrix are defined only up to an unknown scale. Therefore, a constraint should be added to the elements of the fundamental matrix to prevent the trivial solution $F=0$. Several methods exist to impose such a constraint. Here, the eight-point algorithm is used, which suggests to divide all the elements by F_{33} , leaving $F_{33}=1$. However, this assumption ($F_{33}=1$) is not realistic in case of poor-perspective camera models. That is, the rotation of the second camera coordinate system with respect to the first one is mainly planar (kappa rotation), and the omega and phi rotations are not considerable. In this case, F_{33} approaches zero, and the above-mentioned assumption fails. In this paper, we propose to use an arbitrary fundamental-matrix transformer in order to keep the conventional assumption of the eight-point algorithm ($F_{33}=1$), and to avoid exceptions in case of poor-perspective camera models ($F_{33} \rightarrow 0$). In other words, we define an arbitrary, non-zero, full-rank 3×3 transformation matrix (T_f) to bring the points expressed in the coordinates system of the second camera to an arbitrary coordinate system, which certainly forms a perspective model with the first camera. Therefore, we change the equation (4) as follows:

$$U(\tilde{F}, \Omega) = \begin{cases} \tilde{u}'_{m_i}{}^T \tilde{F} u_{m_i} = 0 & \text{for } i \in \{1:12\} \\ \det(\tilde{F}) = 0 \\ \tilde{F}_{33} - 1 = 0 \end{cases} \quad (5)$$

Where:

$$\tilde{u}'_{m_i} = T_f u'_{m_i}$$

Thus, the final fundamental matrix (F) can be recovered using the following equation.

$$F = T'^T T_f^T \tilde{F} T \quad (6)$$

As mentioned earlier, we assume that at least J matches in the dataset of size N are certainly inliers. Indeed, we set the breakdown point of the inlier detection technique to $(N - J)/N$. Then, the optimality of the fundamental matrix is

measured as the sum of squared Sampson distances over an inlier-set of size J . Therefore, the optimal fundamental matrix, \hat{F} , can be defined as in equation (7), where d is the Sampson distance computed using equation (2).

$$\hat{F} = \arg \min \left(\sum_{j=1}^J d_j^2(F) \right) \text{ s.t. } d_1^2 < d_2^2 < \dots < d_J^2 < d_{J+1}^2 < \dots < d_N^2 \quad (7)$$

In summary, the main problem is to find the minimal set Ω of matches to find an optimal F , which minimizes the sum of the J shortest Sampson squared distances in the dataset. In theory, the solution space of this problem contains all possible combinations of twelve out of N putative matches (${}^N C_{12}$).

3. SPARCE MATCHING VIA GENETIC ALGORITHM

The concluded problem can be considered as an integer, nonlinear problem where the variables are sets of integer pairs (equation (3)), and the objective function is complex and nonlinear (equation (7)). There are several ways to solve such a problem, including dynamic programming, mixed-integer programming and heuristic optimization. Although dynamic programming and mixed-integer, nonlinear programming provide good tools to address such problems, they still require considerable computational effort (Elbeltagi et al., 2005). In contrast, heuristic and meta-heuristic optimization techniques are capable of finding good-enough solutions within a reasonable time by applying probabilistic transition rules rather than deterministic ones. Accordingly, in this paper, a modified version of the integer-coded genetic algorithm originally proposed by Deep et al. (2009) is applied. Although GA is conventionally known to be slower than similar optimization methods, using its evolutionary nature with properly-defined operators and sampling strategies can make it much efficient. Besides, genetic algorithms have the advantage that they can be simply parallelized (Haupt and Haupt, 2004).

Input	Putative correspondences
Output	Estimated epipolar geometry (fundamental matrix) and the entire set of inlier matches
a) Genetic Algorithm	<ul style="list-style-type: none"> - Input: The label-set of matches - Objective: Estimate the accurate fundamental matrix via finding an inlier-set of minimum cardinality <ol style="list-style-type: none"> 1. Initialize the first population by guided sampling While the best solution is not improved anymore 2. Compute F from each individual 3. Evaluate each individual by computing the sum of the J smallest squared Sampson residuals (Equation (7)) 4. Perform genetic operators on the individuals of the current population and reproduce the next population 5. Save the best solution so far with the inlier-set of minimum cardinality (size=J) associated with it End while 6. Re-estimate the fundamental matrix (\hat{F}) using the inlier-set of minimum cardinality from the best solution (\hat{I})
b)	Estimate the uncertainty of the motion model ($\hat{\Sigma}_{\hat{F}}$)
c)	Estimate the average and uncertainty of the Sampson residuals (M_d, S_d) for \hat{I} to determine the outlying threshold
d)	Compute and threshold the Sampson residuals on other matches to determine the entire set of inlier

Table 1. The pseudo-code of the proposed sparse matching technique via genetic algorithm

Basically, GA encodes the decision variables (chromosomes) into a cellular environment (e.g. binary strings, integer sets or real-value arrays). Then, several sub-sets of the chromosomes

are sampled; each combination is called an individual. A set of individuals form the population. In every iteration of the evolution, the parent individuals in the population are evaluated against the objective function. They keep being crossed-over, mutated and/or randomized based on their fitness to produce a new population. These iterations go on the same way until reaching an optimal solution, which cannot be improved anymore by younger generations.

As mentioned earlier in section 2, each gene is a correspondence that is labelled with two integers from the label-set S . This labelling fashion accelerates the hypothesis generation step and makes the integer mutation and cross-over procedures geometrically meaningful. Correspondingly, each individual denotes a feasible subset of matches, as a variable-set Ω , for estimating the correct fundamental matrix.

First, GA is initialized by generating a random population of candidate solutions sampled across the search space of size ${}^N C_{12}$. If the sampling is done in a completely random way, then there is big risk to produce more degenerate solutions than near-optimal ones. See the example given in section 5. To avoid such degenerate solutions (local optima), guided sampling is proposed in this paper. To guide the sampling, the minimal rectangle containing the model area on the first image is divided to 12 sub-regions of equal area as in Figure 1. The density of each region is calculated as the number of matches enclosed by it, normalized by the total number of matches. For the first half of the population, every individual (every set Ω in the population) is made of twelve matches that are randomly picked up from twelve different regions. For the other half of the population, the matches are picked up from twelve regions, which are selected successively in the roulette-wheel selection. That is the density of a region determines its probability to participate in sampling. Following this sampling scheme, we can assure that the samples are well distributed on the whole model area. Figure 2 illustrates an example of putative matches distributed on the minimal rectangle and the roulette wheel corresponding to it.

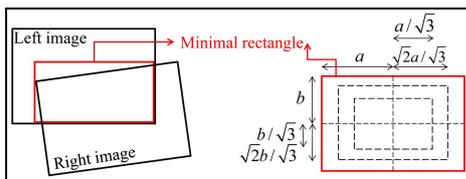


Figure 1. Guided sampling- finding the minimal rectangle and dividing it to sub-regions of equal area

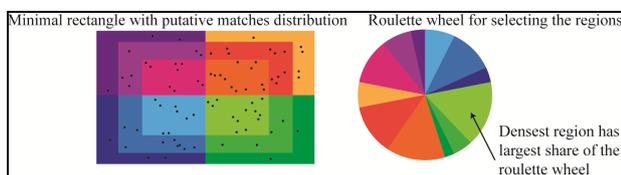


Figure 2. Guided sampling based on the density of the regions

Once the population is formed, for each individual the cost function is calculated as in equation (7), and a fitness value is assigned to the chromosome; the lower the value of the cost function, the fitter the solution. A selection operator is then applied to the population to allocate the instances of fitter solutions for entering a mating pool as parents to reproduce a new generation of solutions. The tournament selection is used here due to its higher computational efficiency over other selection techniques (Deep et al., 2009). Afterwards, new

individuals are reproduced from the selected parents by crossover and mutation operators.

The crossover operator combines the chromosomes to explore the search space. Two offsprings, $\Omega_{\text{offspring1}}$ and $\Omega_{\text{offspring2}}$, are created by combining two selected parents, $\Omega_{\text{parent1}} = \{m_i^{\text{par1}} \mid i \in \{1:12\}\}$ and $\Omega_{\text{parent2}} = \{m_i^{\text{par2}} \mid i \in \{1:12\}\}$, using the crossover operator as in equation (8).

$$\begin{aligned} \Omega_{\text{offspring1}} &= \{m_i^{\text{offs1}} \mid i \in \{1:12\}\} \\ \Omega_{\text{offspring2}} &= \{m_i^{\text{offs2}} \mid i \in \{1:12\}\} \end{aligned} \quad (8)$$

where:

$$\begin{aligned} m_i^{\text{offs1}} &= \text{round}(m_i^{\text{par1}} + \beta_i |m_i^{\text{par1}} - m_i^{\text{par2}}|) \\ m_i^{\text{offs2}} &= \text{round}(m_i^{\text{par2}} + \beta_i |m_i^{\text{par1}} - m_i^{\text{par2}}|) \end{aligned}$$

Instead of the originally proposed Laplace-distributed parameter by Deep et al. (2009) for β_i , we define it differently via equation (9) in order to preserve the constraints of the decision variables (equation (3)). Therefore, constraint-violation validation is not required for the produced offsprings. In equation (9), $u_i \in [0,1]$ is a uniformly distributed random number.

$$\beta_i = \max\left(\frac{1 - m_i^{\text{par1}}}{|m_i^{\text{par1}} - m_i^{\text{par2}}|}, \frac{1 - m_i^{\text{par2}}}{|m_i^{\text{par1}} - m_i^{\text{par2}}|}\right) + \left(\frac{\min\left(\frac{(W, H) - m_i^{\text{par1}}}{|m_i^{\text{par1}} - m_i^{\text{par2}}|}, \frac{N - m_i^{\text{par2}}}{|m_i^{\text{par1}} - m_i^{\text{par2}}|}\right) - \max\left(\frac{1 - m_i^{\text{par1}}}{|m_i^{\text{par1}} - m_i^{\text{par2}}|}, \frac{1 - m_i^{\text{par2}}}{|m_i^{\text{par1}} - m_i^{\text{par2}}|}\right)}{u_i}\right) \quad (9)$$

While crossover is conducted on two parental individuals, mutation is carried out to modify one solution randomly but locally via mutating an individual's genes. A mutated solution, Ω^{mut} , is created from a parent solution Ω^{par} by the following equation (Deep et al., 2009).

$$\Omega^{\text{mut}} = \begin{cases} \left[\Omega^{\text{par}} - s(\Omega^{\text{par}} - \min(\Omega^{\text{par}}))\right], & t < u \\ \left[\Omega^{\text{par}} + s(\max(\Omega^{\text{par}}) - \Omega^{\text{par}})\right], & t \geq u \end{cases} \quad (10)$$

In equation (10), $t = \frac{\Omega - \min(\Omega)}{\max(\Omega) - \min(\Omega)}$, u is a random uniformly

distributed number between zero and one, and $s = (s_1)^2$ follows a power distribution with s_1 having a random value between zero and one. At any iteration, random exploration is also performed by generating a fixed number of individuals based on the sampling strategy explained earlier. Generating random solutions as one portion of the population reduces the chance of converging to local optima.

The replacement strategy applied in this study can be considered as a combination of steady-state and elitist replacement methods. It helps to keep the best solutions of older generations and to maintain the population diversity to avoid premature convergence. Assume that P^- is the population of the last generation, P^+ is the population of the selected parents from P^- , and P^{++} is the population of the reproduced offsprings. Accordingly, q^- demonstrates the least fitness value among the best third quartile of individuals in P^- . Therefore, the new population P starts forming by the fittest individuals of P^- (elites). Among the elite individuals with similar fitness, those whose variable-sets are formed by matches coming from

more distinct regions of the minimal rectangle have priority in the replacement. The rest of the places at P will be occupied by the following replacement condition.

$$P_i = \begin{cases} P_i^{++} & \text{if fitness}(P_i^{++}) > q^- \\ P_i^+ & \text{if fitness}(P_i^{++}) < q^- \end{cases} \quad (11)$$

Equation (11) implies that an offspring whose quality is worse than 75% of the previous solutions cannot replace its parents. The genetic algorithm iterates the procedures mentioned above until there is no improvement in the average of the elites' fitness values for a specified number of generations.

4. INLIER DETECTION

Once the genetic algorithm terminates, the inlier-set of minimum cardinality with size J is found out, \hat{I} , and the accurate fundamental matrix, \hat{F} , is re-estimated using these points by performing least squares adjustment on the following observation equations.

$$U(X, L) = \begin{cases} \hat{u}_i^T \hat{F} u_i = 0 & \text{for } i \in \hat{I} \\ \det(\hat{F}) = 0 \\ \hat{F}_{33} - 1 = 0 \end{cases} \quad (12)$$

where:

$$\hat{u}_i = T_j T' x'_i, u_i = T x_i$$

$$X = [\hat{F}_{11} \quad \hat{F}_{12} \quad \hat{F}_{13} \quad \hat{F}_{21} \quad \hat{F}_{22} \quad \hat{F}_{23} \quad \hat{F}_{31} \quad \hat{F}_{32}]$$

$$L = [(\hat{u}_i^T, u_i^T) \text{ for } i \in \hat{I}]$$

Geometric outliers can, now, be classified as the points that their residuals from the estimated fundamental matrix are more than a threshold. The important issue would be the way to determine this threshold. Conventionally, it is common to assume that the residuals for the inliers follow a normal distribution with known standard deviation σ . Standard RANSAC algorithms also estimate σ using maximum likelihood estimation based on the median of the residuals issued by inliers of the minimal sample set (Raguram, 2013). In this paper, instead of using a fixed threshold for outlier classification scheme, we use the uncertainty of the estimated model (\hat{F}) and the position of the inliers on the images to calculate an adaptive threshold.

Referring to equation (12), the covariance matrix of the estimated parameters can be derived using the covariance law as in equation (13), where $A = \frac{\partial U}{\partial X}$, $B = \frac{\partial U}{\partial L}$ and \hat{V} is the vector of estimated residuals.

$$\hat{\Sigma}_{\hat{X}} = \frac{\hat{V}^T \hat{V}}{J-7} (A^T (B B^T)^{-1} A)^{-1} \quad (13)$$

Knowing from equation (6) that $\hat{F} = T'^T T_j^T \tilde{F} T \triangleq G(\hat{X})$, we can determine the uncertainty of the estimated fundamental matrix using the rules of error propagation as follows.

$$\hat{\Sigma}_{\hat{F}} = \left(\frac{\partial G}{\partial \hat{X}} \right) \hat{\Sigma}_{\hat{X}} \left(\frac{\partial G}{\partial \hat{X}} \right)^T \quad (14)$$

Assuming that matches are influenced by a normal measurement noise with a standard deviation bounded to σ_{\max} pixels, we can calculate the uncertainty of Sampson distances for any corresponding pair located at position $x_i = (x_i, y_i)$ on the left image and $x'_i = (x'_i, y'_i)$ on the right image as follows.

$$\sigma_{d_i} = \left(\frac{\partial d_i}{\partial Y} \right) \Sigma_Y \left(\frac{\partial d_i}{\partial Y} \right)^T \text{ for } i \in \hat{I} \quad (15)$$

where:

$$d_i = (x'_i \hat{F} x_i)^2 / (l_1^2 + l_2^2 + l_1'^2 + l_2'^2)$$

$$Y = [\hat{F}_{11}, \dots, \hat{F}_{32}, x_i, y_i, x'_i, y'_i]$$

$$\Sigma_Y = \begin{bmatrix} \hat{\Sigma}_{\hat{F}} & 0_{9 \times 4} \\ 0_{4 \times 9} & \sigma_{\max} I_4 \end{bmatrix}$$

Applying equation (15) to the inlier-set \hat{I} results in a set of Sampson distances. The average of these distances, $M_d = \text{mean}_{i \in \hat{I}}(d_i)$, and the average of the distance uncertainties,

$S_d = \text{mean}_{i \in \hat{I}}(\sigma_{d_i})$, can represent the distribution of the Sampson

distances for the inliers. Considering Tchebyscheff's inequality, at least 90% of the population is within 3.16 times the standard deviation from the mean, no matter what kind of probability distribution they are following. Therefore, every match k is an outlier by confidence of 90% if its residual expressed as Sampson squared distance, d_k , is greater than $(M_d + 3.16 \times S_d)$.

5. EXPERIMENTS AND RESULTS

Different aspects of the proposed approach based on GA were evaluated on a synthetic dataset. The dataset contained 30 synthetic images simulated using a 3D point cloud (Figure 3). Gaussian noise with standard deviation of two pixels was also added to the image points. Using these image points, three sets of contaminated data with outlier ratios of 20%, 45% and 70% were generated. The outliers were simulated as random gross errors from both normal and uniform probability distributions.

The results of inlier detection and motion estimation were compared both against the ground-truth and several other algorithms, including RANSAC, MSAC, MLESAC, Least Trimmed Squares (LTS) and Least Median of Squares (LMedS). For all the methods, the maximum number of iterations was set to 4000, except for the GA, LTS and LMedS. The GA algorithm converges itself when there is no improvement to the results. The fixed number of iterations for LTS and LMedS was set to 300, since generally they didn't improve the residuals after one or two hundred iterations. Our method is implemented in MATLAB directly without using the optimization toolbox, and all the experiments were performed on a standard PC with Intel-i7 2.4GHz CPU and 8GB memory. Parallel computing is not applied at these experiments, though the GA algorithm can be parallelized simply. The efficiency of each algorithm was evaluated in terms of the following factors:

- Time of processing in seconds (t) and number of iterations ($Iter$) before converging to a solution (for GA, the number of iterations is the number of evolution iterations multiplied by the population size which is set to 25),
- Ratio of inliers (ρ) that is the percentage of the detected inliers from the total number of matches,
- Accuracy of inlier detection (α), which is the proportion of correctly identified outliers/inliers (true positive and true negative) among all the points,
- Sensitivity or true-positive rate (TPR), which is the proportion of correctly identified inliers among all inlier ground-truths,
- Specificity or true-negative rate (TNR), which is the proportion of correctly identified outliers among all outlier ground-truths,

- Average of the squared Sampson distances for detected inliers (μ_d), which shows how well the estimated motion parameters fit to the detected inliers,
- Distance between the estimated fundamental matrix and the reference one (D_F), which is measured using the method described by Zhang (1998),
- Frobenius norm of the difference between the normalized estimated fundamental matrix and the reference one ($\|\Delta F\|$).

The averages of the results for all the images in the synthetic datasets are presented in Table 2, Table 3 and Table 4. As it can be noticed, at presence of high percentage of outliers, the performance of GA was decreased. However, the results (particularly the accuracy and true positive rate) were considerably better than other methods. Besides, the proposed algorithm could deal with high ratio of outliers without dramatically increasing the computational effort.

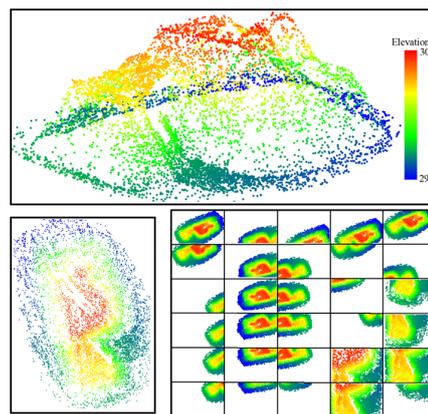


Figure 3. The synthetic data set. The 3D point cloud at the top, the planimetric view of the point cloud at the bottom left and the thumbnails of the simulated images at the bottom right.

	GA	LTS	LMedS	RANSAC	MSAC	MLESAC
t	4.72	0.25	0.26	2.98	3.06	1.43
I_{tr}	3584	300	300	4000	4000	1913
ρ	31.15	9.97	50.03	19.89	16.16	4.29
α	85.49	71.46	67.03	76.75	73.74	70.27
TPR	77.73	19.05	78.43	44.40	33.16	7.59
TNR	88.82	93.92	62.15	90.61	91.13	97.13
μ_d	0.028492	0.011267	0.674036	0.000569	0.000659	0.000736
D_F	54.25	91.91	62.37	63.04	48.96	19.15
$\ \Delta F\ $	0.057683	0.050766	0.055650	0.109889	0.094048	0.066350

Table 2. Results for the synthetic dataset with 70% outlier

	GA	LTS	LMedS	RANSAC	MSAC	MLESAC
t	0.11	0.25	0.25	0.25	0.25	0.14
I_{tr}	1856	300	300	4000	4000	1802
ρ	54.58	9.97	50.03	57.11	57.02	22.71
α	99.11	54.64	90.78	94.71	94.99	65.42
TPR	98.80	17.83	87.09	97.10	97.28	39.21
TNR	99.49	99.63	95.28	91.78	92.19	97.46
μ_d	0.000069	0.000055	0.037973	0.000193	0.000196	0.000554
D_F	10.32	326.22	1105.68	8033.61	5216.04	532.96
$\ \Delta F\ $	0.003664	0.044149	0.020060	0.010616	0.009662	0.048997

Table 3. Results for the synthetic dataset with 45% outlier

	GA	LTS	LMedS	RANSAC	MSAC	MLESAC
t	0.01	0.26	0.26	0.02	0.02	0.00
I_{tr}	2142	300	300	4000	4000	554
ρ	80.01	9.97	50.03	80.08	79.87	67.04
α	99.88	29.97	70.01	99.19	98.73	86.09
TPR	99.93	12.46	62.52	99.55	99.13	83.21
TNR	99.69	99.98	99.94	97.79	97.13	97.61
μ_d	0.000017	0.000000	0.000003	0.000047	0.000036	0.000179
D_F	5.42	19.61	6.75	8.12	788.60	278.71
$\ \Delta F\ $	0.005007	0.014935	0.002542	0.000809	0.004858	0.007914

Table 4. Results for the synthetic dataset with 20% outlier

The experiments also showed the fact the hypothesis generation via the proposed algorithm was more effective than RANSAC-like methods. For these experiments, the useless hypotheses were defined as those samples which even fail to estimate the motion parameters. We defined this failure as the event when the average of 10 percent of the shortest Sampson distances among the whole dataset becomes more than 100. Accordingly, the average rates of useless hypothesis generation for GA, RANSAC, MSAC and MLESAC were 1.1%, 2.8%, 3.4% and 5.1% respectively.

For testing the efficiency of the sampling method proposed in section 3, we considered the two images in Figure 4. Due to the texture pattern on the left wall, many more features are detected over that wall in comparison with other texture-less walls or doors (see the putative correspondences in Figure 4, at left). Therefore, there is a great chance to pick up all the samples from the matches located only on this wall. As a result, the fundamental matrix that is calculated from the matches on this dominant plane will actually be a homography fitting very well to all the inliers on this plane, which is an example of degeneracy. Such a homography does not fit to the matches at the rest of the scene and does not estimate the camera motion correctly. The comparative the results obtained by applying all the methods to the images of Figure 4 are presented in Table 5. To verify the accuracy, some signalized targets were used as control tie points. The average of the squared Sampson distances on the control points ($\mu_{d,CP}$) shows how well the estimated motion model fits to the inliers correctly distributed at the scene. As Figure 5 demonstrates, RANSAC techniques failed to consider the inliers outside the dominant plane for motion estimation. However, the genetic algorithm guided the sampling and evolution process in a way, which considered the distribution of the points. Therefore, the local optima could be avoided more efficiently.



Figure 4. Stereo images at the right and putative correspondences illustrated at the left

	GA	LTS	LMedS	RANSAC	MSAC	MLESAC
t	1.14	0.38	0.27	3.15	3.12	2.09
I_{tr}	2800	300	300	4166	4095	3000
ρ	74.14	9.99	50.00	16.91	17.77	15.63
μ_d	0.027023	0.000311	0.012626	0.000828	0.000815	0.000877
$\mu_{d,CP}$	1.65	9695.29	1154.07	6628.88	34350.29	9835.40
D_F	2.63	31475.01	197.93	251.43	2025.65	2810.34
$\ \Delta F\ $	0.000016	0.000318	0.000480	0.000227	0.001977	0.000711

Table 5. Results for the images in Figure 4



Figure 5. Results of RANSAC inlier detection at the left and GA inlier detection at the right

6. CONCLUSION

In this study we proposed a modified version of integer-coded genetic algorithm for the problem of accurate motion estimation and inlier detection from putative matches. The proposed algorithm can be considered as a universal solution to most of the drawbacks involved in conventional robust estimators, specifically RANSAC-like methods. Based on the experiments, the proposed approach showed robustness to high percentage of outlier and, also, to non-uniformly textured images where the chance of degeneracy in motion estimation is high. In general, the method was able to detect the inliers by 85% to 100% accuracy, which is a remarkable success for large data sets.

REFERENCES

- Adam, A., Rivlin, E. and Shimshoni, I., 2001. ROR: Rejection of outliers by rotations. *IEEE T Pattern Anal*, 23(1), pp. 78-84.
- Capel, D., 2005. An effective bail-out test for RANSAC consensus scoring. British Machine Vision Association Conference, <http://www.comp.leeds.ac.uk/bmvc2008/proceedings/2005/papers/224/Capel224.pdf> (20 Dec. 2014).
- Chum, O., Matas, J. and Kittler, J., 2003. Locally optimized RANSAC. *Lect Notes Comput Sc*, Vol. 2781, pp. 236-243.
- Chum, O. and Matas, J., 2005. Matching with PROSAC-progressive sample consensus. In: *Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, Vol. 1, pp. 220-226.
- Deep, K., Singh, K. P., Kansal, M. L. and Mohan, C., 2009. A real coded genetic algorithm for solving integer and mixed integer optimization problems. *Appl Math Comput*, 212(2), pp. 505-518.
- Elbeltagi, E., Hegazy, T., and Grierson, D., 2005. Comparison among five evolutionary-based optimization algorithms. *Adv Eng Inform*, 19(1), pp. 43-53.
- Frahm, J. M. and Pollefeys, M., 2006. RANSAC for quasi-degenerate data (QDEGSAC). In: *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, Vol. 1, pp. 453-460.
- Hartley, R. I., 1997. In defense of the eight-point algorithm. *IEEE T Pattern Anal*, 19(6), pp. 580-593.
- Hartley, R. I. and Kahl, F., 2007. Global optimization through searching rotation space and optimal estimation of the essential matrix. In: *Proceeding of IEEE International Conference on Computer Vision*, Rio de Janeiro, pp. 1-8.
- Hartmann, J., Klussendorff, J. H. and Maehle, E., 2013. A comparison of feature descriptors for visual SLAM. In: *Proceedings of IEEE European Conference on Mobile Robots*, Barcelona, pp. 56-61.
- Hasler, D., Sbaiz, L., Süssstrunk, S. and Vetterli, M., 2003. Outlier modeling in image matching. *IEEE T Pattern Anal*, 25(3), pp. 301-315.
- Haupt, R. L. and Haupt, S. E., 2004. *Practical Genetic Algorithms*. John Wiley & Sons, Inc., New Jersey, pp. 1-24, 137-148.
- Kekec, T., Yildirim, A. and Unel, M., 2014. A new approach to real-time mosaicing of aerial images. *Robot Auton Syst*, 62(12), pp. 1755-1767.
- Li, J. and Allinson, N. M., 2008. A comprehensive review of current local features for computer vision. *Neurocomputing*, 71(10), pp. 1771-1787.
- Matas, J. and Chum, O., 2004. Randomized RANSAC with $T_{d,d}$ test. *Image Vision Comput*, 22(10), pp. 837-842.
- Ni, K., Jin, H. and Dellaert, F., 2009. GroupSAC: Efficient consensus in the presence of groupings. In: *IEEE International Conference on Computer Vision*, Kyoto, pp. 2193-2200.
- Nistér, D., 2005. Preemptive RANSAC for live structure and motion estimation. *Mach Vision Appl*, 16(5), pp. 321-329.
- Radhika, V. N., Kartikeyan, B., Krishna, B.G., Chowdhury, S. and Srivastava, P.K., 2009. Robust stereo image matching for spaceborne imagery. *IEEE T Geosci Remote*, 45(9), pp. 2993-3000.
- Raguram, R., 2013. *Efficient Algorithms for Robust Estimation*. PhD dissertation, University of North Carolina, Chapel Hill.
- Taraglio, S. and Chiesa, S., 2011. Evolutionary approach to epipolar geometry estimation. *Advances in Theory and Applications of Stereo Vision (Ed. Bhatti, A.)*, pp. 1-16.
- Torr, P. and Murray, D. W., 1997. The development and comparison of robust methods for estimating the fundamental matrix. *Int J Comput Vision*, 24(3), pp. 271-300.
- Wang, Y., Huang, J., Liu, J. and Tang, X., 2010. An efficient image-registration method based on probability density and global parallax. *AEU-Int J Electron C*, 64(12), pp. 1148-1156.
- Wendel, A., Maurer, M., Graber, G., Pock, T. and Bischof, H., 2012. Dense reconstruction on-the-fly. In: *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, Providence, RI, pp. 1450-1457.
- Wöhler, C., 2013. *3D Computer Vision: Efficient Methods and Applications (2nd ed.)*. Springer Science & Business Media, pp. 55-73.
- Yang, J., Li, H. and Jia Y., 2014. Optimal Essential Matrix Estimation via Inlier-Set Maximization. *Lect Notes Comput Sc*, Vol. 8689, pp. 111-126.
- Zhang, Z., 1998. Determining the epipolar geometry and its uncertainty: A review. *Int J Comput Vision*, 27(2), pp. 161-195.