Adaptive and Hybrid Genetic Approaches for Estimating the Camera Motion from Image Point Correspondences

Francisco Vasconcelos
DEEC/ISR
University of Coimbra
Coimbra, Portugal
fpv@isr.uc.pt

Carlos Henggeler Antunes
DEEC/INESC
University of Coimbra
Coimbra, Portugal
ch@deec.uc.pt

João P. Barreto
DEEC/ISR
University of Coimbra
Coimbra, Portugal
jpbar@deec.uc.pt

ABSTRACT

Rigid motion estimation from image point correspondences is an overconstrained problem that can be solved by minimizing an adequate cost function. Given the unreliable nature of image point correspondences, they must be divided into two categories: *inliers* and *outliers*. Finding the correct camera motion and discarding the outliers is a coupled problem usually solved by a random search of the solution space. This article proposes adaptive and hybrid genetic approaches to improve the efficiency of this search. We build on top of the GASAC algorithm that has been recently presented for solving problems in geometric computer vision. GASAC is modified to address the specific issues of camera motion estimation such as outlier ratios above 50% due to wide-baseline image acquisition and an adequate choice of a fitness function. In order to avoid local minima, we propose three adaptive strategies: varying the mutation probability, resampling the lowest ranked individuals, and using a hybrid approach that combines GASAC with simulated annealing. Results are validated on publicly available benchmark images, and it is shown that the proposed genetic approaches outperform the standard RANSAC search used among computer vision practitioners.

Categories and Subject Descriptors

I.4.8 [**Image Processing and Computer Vision**]: Scene Analysis—*motion, stereo*; I.2.8 [**Artificial Intelligence**]: Problem Solving, Control Methods, and Search—*heuristic methods*

General Terms

Algorithms, Performance

GECCO'11, July 12–16, 2011, Dublin, Ireland.

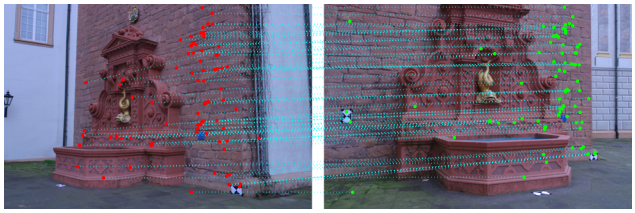


Figure 1: The scene is imaged from two distinct viewpoints, and points are matched across views. The image correspondences are used as input for estimating the camera motion.

Keywords

Hybrid genetic algorithm, Adaptive genetic algorithm, Genetic based sample consensus, RANSAC, Stereo camera calibration

1. INTRODUCTION

This work focuses on the estimation of the relative rigid displacement (rotation and translation) between two camera views using solely image information. This is an important problem in computer and robot vision with applications that range from localization of autonomous vehicles [9] to 3D modeling of cities [12], passing by calibration of camera networks [1], just to name a few examples.

The input for the estimation algorithm is a set of point correspondences between the two views, which is obtained using image processing techniques. Figure 1 shows an example of two images of the same scene acquired from distinct viewpoints. The SIFT algorithm, proposed in [5], is used to establish point matches that are overlaid. It can be shown that each pair of corresponding points defines a bilinear constraint on the camera motion parameters known as the *epipolar constraint* [6]. Thus, and taking into account that the SIFT algorithm usually establishes hundreds of matches between two images of the same scene, a possible solution for the estimation consists in stacking all the epipolar constraints, build a suitable cost function on the rotation and translation parameters, and finally find the correct solution using standard a minimization procedure (e. g. gradient descent).

Unfortunately, this straightforward estimation approach typically leads to poor results because many of the point correspondences are erroneous and their inclusion on the

global cost function compromises the search. This means that the global minimization must be preceded by a robust estimation step, that finds a suitable initialization for the camera motion parameters, and divides the set of input correspondences into two subsets: the *inliers*, that are considered for the estimation of the rigid displacement; and the *outliers* that are point matches that do not fit the epipolar geometry of the two views.

This robust estimation is typically performed by running the so called *five-point-algorithm* [8] within a Random Sample Consensus (RANSAC) procedure [2]. Nister showed in [8] how to obtain solutions for the camera motion by analytically solving the system of equations composed by the epipolar constraints arising from five point correspondences. The robust search is carried by grouping the set of all point matches in five-tuples and randomly sampling the resulting population. For each sampling iteration the selected tuple is used as an input in Nister's five-point-algorithm for generating a candidate hypothesis for the camera motion. The hypothesis are ranked according to a certain criterion and the set of point correspondences is divided into inliers and outliers based on how well they fit the epipolar geometry defined by the winning candidate.

RANSAC is broadly used in geometric computer vision for tasks other than estimating the relative camera motion. In the last years different flavors of the basic procedure have been proposed with the objective of either improving the robustness and accuracy [13], or decreasing the computational overhead [7]. A recent development that is particularly interesting is the *Genetic Algorithm Sample Consensus* (GASAC) [10]. Rodehorst et al. suggest in their paper to replace the random sampling of the population by a genetic oriented search, and show that GASAC outperforms RANSAC in determining the fundamental matrix [6] when the ratio of outlier correspondences is around 45%.

Although GASAC was designed for estimation tasks in multi-view geometry, there is no report in applying it to the case of rigid displacement between two views. We show that a genetic approach is particularly useful when dealing with wide-baseline stereo, for which the ratio of outlier matches is high due to the fact that images are acquired from different viewpoints. In this situation the random search used by RANSAC requires an excessive number of iterations to find a sufficiently accurate solution. Therefore we propose to take advantage of the GASAC guided search and introduce the necessary modifications for further improving the results taking into account the problem specifics. Our contributions can be summarized as follows:

- We use for the first time the GASAC framework for estimating rigid camera motion, addressing the problem of choosing an adequate fitness function and using wide-baseline pairs of real images with outlier ratios that can significantly exceed 50 %.
- We propose modifications to GASAC, namely by changing the strategy whenever the search becomes stagnated. These modifications enable to cover a wider region of the solution space while keeping the computation tractable.

2. NOTATION

Vectors are denoted by lowercase bold symbols, e.g. $\mathbf{p}$, scalars are written as plain symbols, e.g. p_1, matrices are represented by letters in sans serif font, e.g. E, and sets are denoted by letters in cal font, e.g. $\mathcal{S}$.

A 2D image point with coordinates (p_1, p_2) is usually represented by a 3×1 homogeneous vector:

$$\mathbf{p} \sim \begin{pmatrix} p_1 & p_2 & 1 \end{pmatrix}^{\mathsf{T}}$$

with $\sim$ denoting the equality up to a scale factor. The vector product between $\mathbf{p}$ and $\mathbf{q}$ is carried out as

$$\mathbf{p} \times \mathbf{q} = [\mathbf{p}]_{\times}\, \mathbf{q}$$

where $[\mathbf{p}]_{\times}$ is the skew symmetric matrix defined by $\mathbf{p}$

$$[\mathbf{p}]_{\times} = \begin{pmatrix} 0 & -1 & p_2 \\ 1 & 0 & -p_1 \\ -p_2 & p_1 & 0 \end{pmatrix}.$$

The operator $(\cdot)_n$ is used whenever convenient and returns the n^{th} entry of a vector, e.g. $(\mathbf{p})_2 = p_2$. The rotations in the 3D space are represented by a 3×3 orthonormal matrix R that satisfies the following properties: $\det(\mathsf{R}) = 1$ and $\mathsf{R}^{-1} = \mathsf{R}^{\mathsf{T}}$

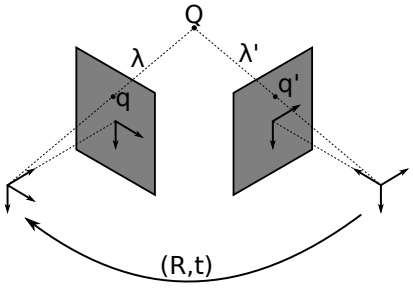


Figure 2: The 3D point Q is projected onto the image points q and q′ in the two views

3. BACKGROUND CONCEPTS

Figure 2 shows a 3D point $\mathbf{Q}$ that is projected into two distinct views as $\mathbf{q}$ and $\mathbf{q}'$. Let $\mathbf{c} = (\mathbf{q}, \mathbf{q}')$ be a correspondence between two views of the point $\mathbf{Q}$. In order to estimate the relative rigid displacement an input set $\mathcal{C}_N = \{\mathbf{c}_1, \mathbf{c}_2, ..., \mathbf{c}_N\}$ containing N correspondences is used to compute a rotation R and a translation $\mathbf{t}$.

It can be shown that, if we use exclusively image information, the translation motion can only be determined up to a scale factor [6]. Thus, the stated estimation problem has a total of 5 degrees of freedom: 3 unknowns for rotation, and 2 unknowns for translation.

Let λ denote the depth of the 3D point in the reference frame of the first camera (see Fig. 2). It follows that its non-homogeneous 3D coordinates are

$$\mathbf{Q} = \lambda\, \mathbf{q}$$

In a similar manner, and if λ' denotes the depth with respect to the second camera, the 3D coordinates of the scene point in the right reference frame are

$$\mathbf{Q}' = \lambda'\, \mathbf{q}'$$

Taking into account the rigid displacement between the two cameras, it follows from a simple change of coordinates that

$$\lambda\, \mathbf{q} = \lambda'\, \mathsf{R}\, \mathbf{q}' + \mathbf{t}$$

Multiplying both sides of the equation by $[\mathbf{t}]_\times$, and knowing that $[\mathbf{t}]_\times \mathbf{t} = 0$, it yields

$$\lambda[\mathbf{t}]_\times \mathbf{q} = \lambda' [\mathbf{t}]_\times \mathsf{R}\, \mathbf{q}'$$

Since $\mathbf{q}$ and $[\mathbf{t}]_\times \mathbf{q}'$ are necessarily orthogonal, we multiply both sides by $\mathbf{q}^\mathsf{T}$ and finally obtain

$$\mathbf{q}^\mathsf{T} \underbrace{[\mathbf{t}]_\times \mathsf{R}}_{\mathsf{E}} \mathbf{q}' = 0 \tag{1}$$

Equation 1 is bilinear in the image points coordinates and E is the essential matrix that encodes the rotation and translation. It may be shown that E can be easily factorized into the motion parameters, which means that solving for E is the same as solving for R and $\mathbf{t}$ [6]. Since E has size 3×3 and is defined up to a scale factor, its estimation without further constraints requires determining a total of 8 parameters. However, an essential matrix must always satisfy the relation of equation 2, that puts three additional constraints on the entries of E, leaving only a total of five degrees of freedom to be determined. This explains why it is possible to achieve an estimation using a minimum of five point correspondences. More details on the properties of the essential matrix can be found in [6].

$$\mathsf{E}\,\mathsf{E}^\mathsf{T}\,\mathsf{E} - \frac{1}{2}\mathrm{trace}(\mathsf{E}\mathsf{E}^\mathsf{T})\mathsf{E} = 0 \tag{2}$$

Nister [8] provides an algorithm for analytically solving the system arising from stacking together five instances of equation 1 with equation 2, returning up to ten distinct solutions for E. The choice of the correct solution can be performed by testing consistency with additional point correspondences or using any other heuristic (e. g., visibility constraints, prior knowledge about the camera pose, etc). This method, which is called the *five point algorithm*, maps a 5-tuple $\mathcal{C}_5$ containing 5 correspondences into a rigid displacement $(\mathsf{R}, \mathbf{t})$.

4. CAMERA MOTION ESTIMATION AS AN OPTIMIZATION PROBLEM

From the stated above 5 point correspondences are enough to compute a camera motion. However, and since in general more than five correspondences are available, we are interested in using additional information to refine the estimation. Given an input set $\mathcal{C}_N$ containing $N > 5$ correspondences, a possible strategy consists in choosing a 5-tuple $\mathcal{C}_5 \subset \mathcal{C}_N$, compute an initial solution, and then refine it by minimizing a cost function that takes into account the remaining elements from $\mathcal{C}$. Unfortunately, this method only works if all correspondences are correct up to some moderate noise term, which is rarely the case in a real scenario, specially in wide-baseline stereo, where correspondence detection results can be corrupted with more than 50% of mismatches. This puts two problems:

1. the choice of a subset $\mathcal{C}_5$ that provides a good initialization. If a single erroneous correspondence is included in this initial choice, the estimation will most likely not converge to the correct result;

2. the classification of point correspondences as inliers or outliers, considering only the former in building the cost function.

In this work we consider the set $\mathcal{S}_M$ containing all M combinations of $\mathcal{C}_5 \subset \mathcal{C}_N$, develop strategies to search in $\mathcal{S}_M$ for the 5-tuple that provides the most accurate solution, and use it to discard possible outliers.

4.1 RANSAC

RANSAC iteratively samples random 5-tuples from $\mathcal{S}_M$ that are used to compute a candidate essential matrix, and evaluates how well they fit into the remaining correspondences. This is achieved by defining a suitable cost function. Until a stopping condition is verified, it keeps storing the 5-tuple with the lowest cost. The algorithm is as follows:

ALGORITHM 1. *RANSAC*

1: $C_{min} = \inf$
2: ***while not*** *Stopping criteria* ***do***
3: $\quad \mathcal{C}_5 =$ *Select a random element from* $\mathcal{S}_M$
4: $\quad \mathsf{E} =$ *Compute essential matrix from* $\mathcal{C}_5$
5: $\quad C = costFunction(\mathsf{E})$
6: $\quad$ ***if*** $C < C_{min}$ ***then***
7: $\quad\quad C_{min} = C$
8: $\quad\quad \mathsf{E}_{best} = \mathsf{E}$
9: $\quad$ ***end if***
10: ***end while***
11: *Return* E_{best}

RANSAC assumes that the set of correspondences is composed by inliers and outliers, and its goal is to find a 5-tuple whose correspondences are all inliers.

It also assumes that randomly selected correspondences have an independent and equal probability of being an inlier. Therefore if we have a dataset with an inlier ratio of ϵ, the probability of a set $\mathcal{C}_n$ with n different correspondences having at least one outlier will be approximately $1 - \epsilon^n$ (assuming that the total number of correspondences $N \gg n$). Then if P_{clean} is the probability of selecting a set $\mathcal{C}_n$ containing no outliers after J trials we have:

$$1 - P_{clean} = (1 - \epsilon^n)^J \tag{3}$$

Another way to interpret equation 3 is to state that we need to sample at least J random sets $\mathcal{C}_n$ until we find a hypothesis not contaminated by outliers, being:

$$J = \frac{\log(1 - P_{clean})}{\log(1 - \epsilon^n)} \tag{4}$$

From equation 4 it can be concluded that a lower value for n requires a lower number of trials J until we find a non-contaminated set. This fact makes advantageous to generate hypotheses from the lowest possible number of correspondences, which is five in the context of our problem.

An online estimate for J, in which ϵ is updated according to the current minimum cost, is usually used to define a stopping criteria for the RANSAC algorithm.

4.2 Cost evaluation metrics

Equation 1 is a good starting point to build an evaluation function, as we expect it to return decreasing values (ideally zero) with increasingly better estimates for E. However, since the magnitude of correspondence coordinates has impact on its residue, we should use a normalized metric instead, such as the Sampson distance [3]:

$$S_d = \frac{(\mathbf{q}^\mathsf{T}\,\mathsf{E}\mathbf{q}')^2}{(\mathsf{E}\mathbf{q}')_1^2 + (\mathsf{E}\mathbf{q}')_2^2 + (\mathsf{E}^\mathsf{T}\,\mathbf{q})_1^2 + (\mathsf{E}^\mathsf{T}\,\mathbf{q})_2^2} \tag{5}$$

If a set $\mathcal{I}$ containing all inlier correspondences $\mathbf{c}_i = (\mathbf{p}_i, \mathbf{p}'_i)$ was known beforehand, our optimization problem would take the form (note that Sampson distances are always positive):

$$\min_{\mathsf{E}} C = \sum_{\mathbf{c}_i \in \mathcal{I}} S_d(\mathsf{E}, \mathbf{c}_i) \tag{6}$$

Since $\mathcal{I}$ is not known, we must apply a robust evaluation function $costFunction(\cdot)$ to the set $\mathcal{C}_N = \{\mathbf{c}_1, \mathbf{c}_2, ...\mathbf{c}_N\}$ containing all correspondences, in order to filter the effect of outliers:

$$C = costFunction(S_d(\mathsf{E}, \mathbf{c}_1), S_d(\mathsf{E}, \mathbf{c}_2), ..., S_d(\mathsf{E}, \mathbf{c}_N)) \tag{7}$$

A common robust operator is the median, which remains completely unaffected by outliers if they represent less than 50% of all correspondences.

Another possibility is using a sample consensus, that consists in counting the number of outlier correspondences. A threshold parameter τ is compared against the Sampson distance of each correspondence, and values exceeding τ are considered outliers (equation 9). The best solution is the set $\mathcal{C}_5$ that generates the lowest number of outliers among all elements of $\mathcal{C}_N$. The original RANSAC relies on this function to provide robust estimations.

$$C = \sum_{\mathbf{c}_i \in \mathcal{C}_N} out(S_d(\mathsf{E}, \mathbf{c}_i)) \tag{8}$$

$$out(S_d) = \begin{cases} 1 & \text{if } S_d > \tau \\ 0 & \text{if } S_d \leq \tau \end{cases} \tag{9}$$

A final approach is to use a metric that simultaneously takes into account the number of outliers and the Sampson distance of inliers. Such cost function can be defined by equation 11, which will be called bounded distance. This function is used in a modified version of RANSAC [13].

$$C = \sum_{\mathbf{c}_i \in \mathcal{C}_N} bound(S_d(\mathsf{E}, \mathbf{c}_i)) \tag{10}$$

$$bound(S_d) = \begin{cases} \tau & \text{if } S_d > \tau \\ S_d & \text{if } S_d \leq \tau \end{cases} \tag{11}$$

5. A GENETIC APPROACH

If we analyse the structure of the 5-tuples sampled by RANSAC, it becomes clear that they are not probabilistically independent. Samples sharing some correspondences are more likely to have a similar cost evaluation than samples containing completely different correspondences. Although a single outlier can significantly change the hypothesis generated, there are still some conservative boundaries we can impose on the correlation between the number of common correspondences and the cost evaluation.

Consider that from the set $\mathcal{C}_N$, K solution hypotheses were generated and ordered in a list according to increasing cost. Let $\mathcal{C}_B$ be the set containing the $B < K$ best ranked elements from the list. Being ϵ the probability of randomly chosing an inlier from $\mathcal{C}_N$, and α the probability of selecting an inlier from $\mathcal{C}_B$, then $\alpha > \epsilon$. The difference between these two values will become greater as K increases.

Let two 5-tuples $\mathcal{C}_5$ and $\mathcal{G}_5$ be two sets. $\mathcal{C}_5$ contains five random correspondences from $\mathcal{C}_N$ and $\mathcal{G}_5$ contains $k < 5$ random correspondences from $\mathcal{C}_B$ and $5-k$ correspondences from $\mathcal{C}_N$. If P_C and P_G are the probabilities of $\mathcal{C}_5$ and $\mathcal{G}_5$ containing only inliers, respectively, then it follows:

$$P_C = \epsilon^5 \tag{12}$$

$$P_G = \alpha^k + \epsilon^{5-k} \tag{13}$$

And we can conclude that

$$P_r > \epsilon \Rightarrow P_2 > P_1 \tag{14}$$

It is clear that there is an advantage in generating new hypotheses that share correspondences with previously evaluated good solutions. Particularly if k is high, P_G increases as well, providing a background to define the genetic operators mutation ($k = 4$, that is 4 correspondences are kept) and crossover ($k = 5$, that is 5 correspondences resulting from two differnet individuals are kept).

The $mutation(\cdot)$ operator consists in generating a new 5-tuple by randomly selecting one correspondence from $\mathcal{C}_N$ and keeping four from the input solution.

The $crossover(\cdot, \cdot)$ operator consists in generating two new 5-tuples that inherit all their five correspondences from two other samples. In a first step input samples are randomly sorted, then a section point is randomly defined to divide both samples in two parts. Correspondences placed before the section point switch between samples and the others remain in place.

In both operators, it is ensured that output samples do not have repeated correspondences.

5.1 GASAC

GASAC applies the previously defined operators to individuals $\mathcal{G}_5$ in order to search for the optimal solution. Offspring are created using crossover, followed by the possibility of occuring a mutation with a probability P_{mut}. The population is initialized with K individuals, and in each generation $2B$ new offspring are produced. The population is then ranked and reduced to its best K elements, maintaining a constant size. The fitness function is one of the evaluation functions mentioned above. GASAC is applied to our problem as follows:

ALGORITHM 2. *GASAC*

for i=1 to K individuals **do**
  $\mathcal{G}_5(i)$ = *random sample from* $\mathcal{S}_M$
  $C(i)$ = *fitnessFunction*($\mathcal{G}_5(i)$)
end for
rank individuals according to their fitness
while not *Stopping criteria* **do**
  for $b = 1$ *to* B *best individuals* **do**
    $\mathcal{R}_5$ = *random individual from population*
    $\{\mathcal{G}_5(offspr1), \mathcal{G}_5(offspr2)\} = crossover(\mathcal{G}_5(b), \mathcal{R}_5)$
    for $j = \{$ *offspr1, offspr2* $\}$ **do**
      if *probability* P_{mut} **then**
        $\mathcal{G}_5(j) = mutation(\mathcal{G}_5(j))$
      end if
      $C(j)$ = *fitnessFunction*($\mathcal{G}_5(j)$)
    end for
  end for
  Reduce population to the best K individuals
end while

6. ADAPTIVE GASAC

The five correspondences that constitute an individual in GASAC may be a very small subset of $\mathcal{C}_N$, which can amount to thousands of correspondences. This implies that a population usually contains only a fraction of $\mathcal{C}_N$ in each generation. This may present a problem in terms of convergence to the optimal solution since we can inadvertently ignore crucial samples, which can only be incorporated into the population by the very small chance of occuring a mutation in the right way.

One straightforward solution would be to set the population size to a very high number, assuring that it contains at least the great majority of correspondences from $\mathcal{C}_N$. However, since one of the obectives of GASAC is to achieve low computational costs this would be counterproductive.

A second path that could be taken is to increase the mutation probability to achieve more diversity. However, if not done carefully this can produce a slow convergence as well, since a high mutation probability will dissipate a significant part of the intelligent search provided by crossover.

An eventual alternative that revealed more productive is to change the optimization strategy once GASAC stagnates in a local optimum. This can be done by altering GASAC parameters, such as the mutation probability, or combining it with other methods to circumvent its drawbacks.

6.1 GASAC with variable mutation probability (GA-M)

In the first proposed version of an adaptive GASAC, if the best solution does not improve in a predefined number of generations, the mutation probability is switched to a higher value. This strategy aims at introducing new correspondences at a higher rate, to increase the exploration abilities of the genetic algorithm.

ALGORITHM 3. *Adaptive GA-M*

Initialize GASAC
$C_{previous} = \infty$
$counter = 0$
while not *Stopping criteria* **do**
    Compute a GASAC iteration
    $C_{current}$ = *current best cost*
    if $C_{current} == C_{previous}$ **then**
        $counter = counter + 1$
    else
        $counter = 0$
        reset mutation probability to default value
    end if
    if $counter > counter\_threshold$ **then**
        Switch mutation probability to a higher value
    end if
end while

6.2 GASAC with population resets (GA-P)

In another approach, to increase diversity of correspondences in a more drastic way, whenever the best solution does not improve after a predefined number of generations the worst half of the population is removed and substituted by completely new random samples.

ALGORITHM 4. *Adaptive GA-P*

Initialize GASAC
$C_{previous} = \infty$
$counter = 0$
while not *Stopping criteria* **do**
    Compute a GASAC iteration
    $C_{current}$ = *current best cost*
    if $C_{current} == C_{previous}$ **then**
        $counter = counter + 1$
    else
        $counter = 0$
    end if
    if $counter > counter\_threshold$ **then**
        for $w = 1$ *to* $K/2$ *worst individuals* **do**
            $\mathcal{G}_5(w)$ = *random sample from* $\mathcal{S}_M$
        end for
    end if
end while

6.3 Hybrid GASAC and Simulated Annealing (GA+SA)

As our final proposed stategy, we combine GASAC with Simulated annealing (SA). Hybrid approaches of Genetic Algorithms and Simulated Annealing already proved to be a successful strategy in many problems. In previous works (e. g. [14]), this is usually done by applying SA to elements of the population after being modified by genetic operators in each generation. Instead we apply the two unaltered algorithms sequencially.

We use GASAC in a first step to quickly find a solution close to the optimum, and then apply a Simulated Annealing algorithm (SA) to exploit its neighborhood space. The SA algorithm is presented in the next subsection.

ALGORITHM 5. *Hybrid GA+SA*

Initialize GASAC
$C_{previous} = \infty$
$counter = 0$
while not *Stopping criteria* **do**
    Compute a GASAC iteration
    $C_{current}$ = *current best cost*
    if $C_{current} == C_{previous}$ **then**
        $counter = counter + 1$
        exit **for** *cycle*
    end if
    if $counter > counter\_threshold$ **then**
        Switch to SA
        Enable stopping criteria
    end if
end while

6.4 Simulated annealing

Simulated Annealing (SA) is based on the analogy of metal cooling into a minimum energy crystalline structure and the search for the optimal solution in a complex optimization problem, namely with combinatorial and/or non-linear characteristics [4]. SA is a non-population based meta-heuristic (in contrast with genetic algorithms or particle swarm optimization approaches, which are population-based) that uses a stochastic scheme to avoid becoming trapped at local optima. In general, SA starts from a randomly initialized candidate solution. A neighborhood, in some metric sense, is defined around the current solution and a competitor solution is selected to be compared with the incumbent. The

(a) Image 1 (b) Image 2 (c) Image 3

Figure 3: Fountain-P11 Dataset

(a) Image 1 (b) Image 2 (c) Image 3

Figure 4: Entry-P10 Dataset

new solution becomes the current solution not just if it provides a better value for the evaluation function but it may be also accepted with a given probability even if it is worse. This acceptance probability p depends on the difference of performance Δ according to the evaluation function and also on a temperature parameter T, hence the analogy with the annealing process (see equation 15). The temperature parameter decreases along the search process, for instance following a negative exponential function, making increasingly difficult that a worse neighbor solution replaces the current solution (see equation 16 in which r is the cooling rate and t is the iteration counter). The stopping condition may be reaching a minimum temperature.

$$p = e^{-\frac{\Delta}{T}} \tag{15}$$

$$T(t) = T_{max} e^{-rt} \tag{16}$$

SA may be faced as a hill climber procedure with a stochastic backup plan to avoid local optima. In the first iterations (high temperature) it typically allows the current solution to freely navigate throughout the solution space, thus enabling an exploration (diversification) strategy, but the moves gradually become more restricted not allowing high increases in the cost function as temperature decreases, thus enabling an exploitation (intensification) strategy. In the final iterations it becomes almost impossible to increase the cost of the current solution, and the algorithm behaves as a traditional hill-climber.

The definition of a candidate solution is done the same way as in RANSAC, as a 5-tuple of feature correspondences. Neighbors are defined as every solution with four common correspondences, following the reasoning of the previous section.

7. EXPERIMENTAL RESULTS

Performance evaluation is achieved using publicly available datasets presented in [11], from which a set of six images were selected. Three of them are from the fountain-P11 (figure 3) dataset and another three from entry-P10 (figure 4).

The SIFT detector [5] was used to generate four sets of correspondences between image pairs (1,2) and (1,3) from both Foutain-P11 and entry-P10 (table 1).

Since this process is only based on image appearance, several geometrically inconsistent correspondences are established. Correctly removing them, while estimating an accurate pose solution for image pairs is the aim of this test.

Table 1: Input datasets

image pairs	outl. ratio (%)	correspondences
Fountain-P11 (1,2)	47	1111
Fountain-P11 (1,3)	72	864
Entry-P10 (1,2)	56	3163
Entry-P10 (1,3)	70	2316

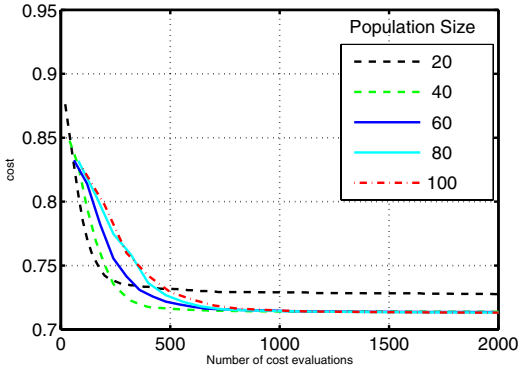


Figure 5: Bounded distance evolution on image pair (1,3) from Entry Dataset during population size tuning

8. GASAC PERFORMANCE

As a first experiment we tested RANSAC, GASAC, and a randomly initialized SA.

For each algorithm, 50 trials were run in each of the four datasets, using two different evaluation functions: number of outliers and bounded distance.

We set RANSAC and SA to compute 300 iterations on image pairs (1,2) and 2000 on image pairs (1,3). For GASAC we divided these numbers by the population size to make the computational effort equivalent to the previous methods.

The population size in GASAC was set to 40. As it can be seen in figure 5 this value presents a balance between a low final cost evaluation and a fast convergence to its value. Although the population tuning results are only shown for one particular case, almost identical plots can be obtained both by using the other datasets or the bounded distance as a cost function.

In [10], Rodehorst used also median squared distance as a cost function with good results, but this was only due to using datasets with an outlier ratio below 50%; otherwise the cost evaluation would become itself corrupted by outliers, leading to a breakdown in performance. In that paper the performance tests were run in the context of fundamental matrix estimation, which requires a minimum of seven samples instead of five. According to equation 4, this means that for the same outlier ratio, the fundamental matrix estimation requires more computational effort than the essential matrix estimation. In the context of our problem, outlier ratios around 50% are easily tackled by RANSAC in within around 100 iterations, as it is shown on figure 6 (almost identical plots could be obtained for Fountain-P10

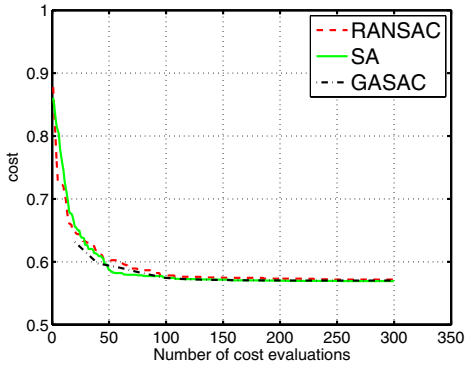


Figure 6: Bounded distance evolution on image pair (1,2) from Entry-P10

and/or number of inliers). The really challenging situations that require a faster search method are when the great majority of samples are corrupted. Therefore we choose not to include the median operator as a viable cost function.

Rotation and translation estimates from image pairs (1,3) are compared against groundtruth to obtain a relative translation error e_t and a rotation error e_R for each of the 50 trials. Mean and maximum errors are displayed in tables 2 and 3, showing that the bounded distance metric leads to considerably more accurate estimations than the number of outliers.

Results show that a very high outlier ratio (figures 7(a) and 7(b)) enhances the advantage of SA and GASAC over RANSAC becomes more evident. The guided search of SA and GASAC provide a faster convergence

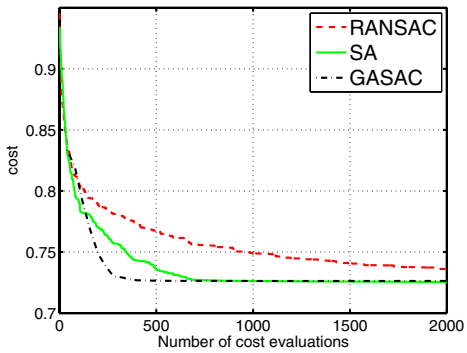
In the particular test that uses image pair (1,3) from entry-P10, the high outlier ratio is combined with a large solution space. While the number of possible solutions is almost irrelevant for RANSAC performance, it negatively affects GASAC when the population size is fixed. In this situation we can observe that GASAC can still be improved, as it gets outperformed by SA. The parallell search of GASAC seems to be more useful for an initial exploration, when there is no prior knowledge on the data, but as soon as it finds a sufficiently good sub-optimal solution, it does not exploit its neighborhood as intensively as SA.

9. ADAPTIVE GASAC PERFORMANCE

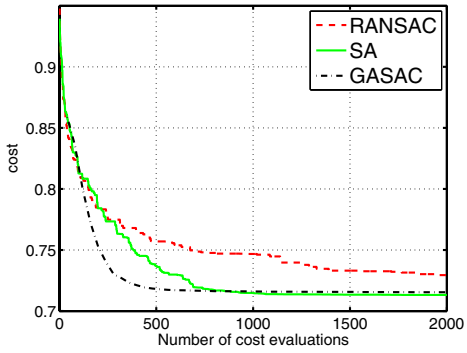
In a second set of experiments we compare the original version of GASAC against its adaptive modifications: GA-M, GA-P and GA+SA. To focus on the most relevant cases only image pairs (1,3) were used, and the cost function was only set to the bounded distance metric.

In their first iterations the new algorithms are identical to GASAC, thus having a similar performance. However, when GASAC becomes stagnated in a local minimum, the adaptive and combined approaches show that there is still margin for reaching lower costs. In figures 8(a) and 8(b), after the first 200 cost evaluations, it is demonstrated that GASAC is outperformed.

As it can be observed in tables 2 and 3, the new genetic approaches do not estimate rigid motion with significantly lower mean errors than GASAC. However, they provide more stable solutions, as it is shown by their lower



(a) Fountain P-11



(b) Entry P-10

Figure 7: Bounded distance evolution on image pairs (1,3)

maximum errors. Particularly GA-P and GA+SA return consistently more stable solutions in both datasets.

10. CONCLUSIONS

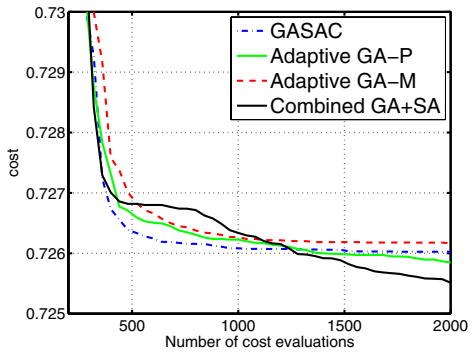
A genetic approach proved to be competitive in estimating the rigid motion between two cameras, with the experiments showing that a guided search is more efficient than using a random sampling. The methods proposed managed to correctly estimate the motion for the case of wide-baseline stereo pairs with a ratio of outlier up to 70 %, converging significantly faster than RANSAC.

The major drawback is the limited searching range when the number of correspondences is high. However, the adaptive and hybrid modifications proposed for the GASAC enhanced the search process when the solution becomes stagated, by increasing its searching range without compromising the computational efficiency. It has been experimentaly demonstrated that they can push the estimated solution closer to its optimum value.

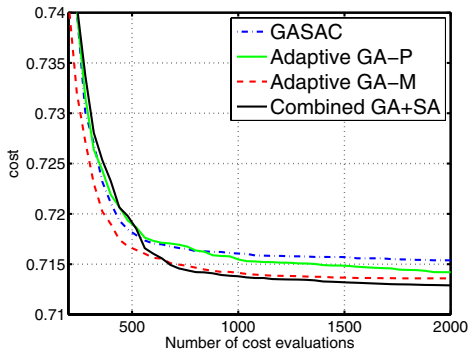
The faster convergence of the proposed methods makes them more suitable for real-time applications, where it must be guaranteed that a sufficiently good solution is found within a limited number of iterations.

11. ACKNOWLEDGEMENTS

Francisco Vasconcelos and João P. Barreto acknowledge Portuguese Science and Technology Foundation (FCT) by generous funding through grants SFRH/BD/72323/2010 and PTDC/EEA-ACR/68887/2006.



(a) Fountain-P10



(b) Entry-P11

Figure 8: Bounded distance evolution on image pairs (1,3) after 200 cost evaluations

Table 2: Motion estimation on Fountain P-11

Cost function: **number of outliers**				
	$\bar{e}_\mathbf{t}$ (%)	$\max e_\mathbf{t}$ (%)	$\bar{e}_\mathsf{R}$ (deg)	$\max e_\mathsf{R}$ (deg)
Ransac	2.98	8.07	3.67	8.99
SA	3.92	6.51	4.87	7.78
GASAC	3.14	6.83	4.06	7.25
Cost function: **bounded distance**				
	$\bar{e}_\mathbf{t}$ (%)	$\max e_\mathbf{t}$ (%)	$\bar{e}_\mathsf{R}$ (deg)	$\max e_\mathsf{R}$ (deg)
Ransac	2.00	9.31	2.37	13.08
SA	1.68	2.19	2.35	3.23
GASAC	1.47	2.25	2.10	3.25
GA-P	1.58	2.17	2.17	3.13
GA-M	1.54	5.93	2.03	8.04
GA+SA	1.68	2.06	2.43	3.05

Table 3: Motion estimation on Entry P-10

Cost function: **number of outliers**				
	$\bar{e}_\mathbf{t}$ (%)	$\max e_\mathbf{t}$ (%)	$\bar{e}_\mathsf{R}$ (deg)	$\max e_\mathsf{R}$ (deg)
Ransac	6.61	26.21	5.36	21.33
SA	5.66	6.90	4.49	6.14
GASAC	6.38	12.69	5.13	11.01
Cost function: **bounded distance**				
	$\bar{e}_\mathbf{t}$ (%)	$\max e_\mathbf{t}$ (%)	$\bar{e}_\mathsf{R}$ (deg)	$\max e_\mathsf{R}$ (deg)
Ransac	4.79	8.32	3.88	8.82
SA	3.67	4.48	3.14	4.21
GASAC	3.91	15.45	3.15	10.22
GA-P	3.81	5.08	3.09	3.99
GA-M	3.70	4.90	2.91	3.74
GA+SA	3.77	4.67	3.23	3.68

12. REFERENCES

[1] S. Beriault, P. Payeur, and G. Comeau. Flexible multi-camera network calibration for human gesture monitoring. In *International Workshop on Robotic and Sensors Environments, 2007. ROSE 2007*, pages 1–6, 2007.

[2] M. A. Fischler and R. C. Bolles. Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. *Commun. ACM*, 24:381–395, June 1981.

[3] R. I. Hartley and A. Zisserman. *Multiple View Geometry in Computer Vision*. Cambridge University Press, 2000.

[4] S. Kirkpatrick, C. D. Gelatt, and M. P. Vecchi. Optimization by Simulated Annealing. *Science*, 220(4598):671–680, 1983.

[5] D. Lowe. Object recognition from local scale-invariant features. In *Proceedings of the 7th IEEE International Conference on Computer Vision*, volume 2, pages 1150–1157, 1999.

[6] Y. Ma, S. Soatto, J. Kosecka, and S. S. Sastry. *An Invitation to 3-D Vision: From Images to Geometric Models*. SpringerVerlag, 2003.

[7] D. Nister. Preemptive ransac for live structure and motion estimation. In *Proceedings of the 9th IEEE International Conference on Computer Vision, 2003*, volume 1, pages 199–206, 2003.

[8] D. Nister. An efficient solution to the five-point relative pose problem. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 26(6):756–77, June 2004.

[9] D. Nister, O. Naroditsky, and J. Bergen. Visual odometry for ground vehicle applications. *Journal of Field Robotics*, 23(1):3–20, 2006.

[10] V. Rodehorst and O. Hellwich. Genetic algorithm sample consensus (gasac) - a parallel strategy for robust parameter estimation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshop*, CVPRW '06, pages 103–, Washington, DC, USA, 2006. IEEE Computer Society.

[11] C. Strecha, W. von Hansen, L. Van Gool, P. Fua, and U. Thoennessen. On benchmarking camera calibration and multi-view stereo for high resolution imagery. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. CVPR 2008.*, pages 1 –8, 2008.

[12] A. Torii, M. Havlena, and T. Pajdla. From google street view to 3d city models. In *2009 IEEE 12th International Conference on Computer Vision Workshops*, pages 2188–2195, 2009.

[13] P. H. S. Torr and A. Zisserman. Mlesac: a new robust estimator with application to estimating image geometry. *Comput. Vis. Image Underst.*, 78:138–156, April 2000.

[14] Z. G. Wang, Y. S. Wong, and M. Rahman. Development of a parallel optimization method based on genetic simulated annealing algorithm. *Parallel Comput.*, 31:839–857, August 2005.