1

Evolutionary Approach to Epipolar Geometry Estimation

Sergio Taraglio and Stefano Chiesa
ENEA, Robotics Lab, Rome
Italy

1. Introduction

An image is a two dimensional projection of a three dimensional scene. Hence a degeneration is introduced since no information is retained on the distance of a given point in the space. In order to extract information on the three dimensional contents of a scene from a single image it is necessary to exploit some *a priori* knowledge either on the features of the scene, i.e. presence/absence of architectural lines, objects sizes, or on the general behaviour of shades, textures, etc. Everything becomes much simpler if more than a single image is available. Whenever more viewpoints and images are available, several geometric relations can be derived among the three dimensional real points and their projections onto the various two dimensional images. These relations can be mathematically described under the assumption of pinhole cameras and furnish constraints among the various image points. If only two images are considered, this research topic is usually referred to as epipolar geometry. Naturally there is no mathematical difference whether the considered images are taken at the same time by two different cameras (the stereoscopic vision problem) or at different times by a single moving camera (optical flow or structure from motion problem). In Robotics both these cases are of great significance. Stereoscopy yields the knowledge of objects and obstacles positions providing a useful key to obtain the safe navigation of a robot in any environment (Zanela & Taraglio, 2002). On the other hand the estimation of the *ego-motion*, i.e. the measure of camera motion, can be exploited to the end of computing robot odometry and thus spatial position, see e.g. (Caballero et al., 2009). In addition the visual sensing of the environment is becoming ubiquitous out of the ever decreasing costs of both cameras and processors and the cooperative coordination of more cameras can be exploited in many applicative fields such as surveillance or multimedia applications (Arghaian & Cavallaro, 2009). Epipolar geometry is then the geometry of two cameras, i.e. two images, and it is usually represented by a 3 x 3 *fundamental* matrix, from which it is possible to retrieve all the relevant geometrical information, namely the rigid roto-translation between camera positions. The estimation of the fundamental matrix is based on a set of corresponding features present in both the images of the same scene. Naturally the error in the process is directly linked to the accuracy in the computation of these correspondences. In the following a novel genetic approach to epipolar geometry estimation is presented. This algorithm searches an optimal or sub-optimal solution for the rigid roto-translation between two camera positions in a evolutionary framework. The fitness of the tentative solutions is measured against the full set of correspondences through a function that is able to correctly cope with outliers, i.e. the incorrectly matched points usually due to errors in feature detection and/or in matching. Finally the evolution of the

solution is granted through a reproduction and mutation scheme. In Section 2 the relevant geometrical concepts of epipolar geometry are recalled, while in Section 3 a review of some of the algorithms devised for the estimation of epipolar geometry is presented. In Section 4 the details of the proposed epipolar geometry estimation based on evolutionary strategies is given. In Section 5 some experimental data relative to both ego-motion and stereoscopy are shown and in Section 6 discussion and conclusions are presented.

2. Theoretical background

Let us briefly review the relevant geometrical concepts of the pinhole camera model and of epipolar geometry.

2.1 Pinhole camera

A point $M = (X, Y, Z, 1)^T$ in homogeneous coordinates in a world frame reference and the correspondent point $m = (x, y, 1)^T$ on the image plane of a camera are related by a projective transformation matrix:

$$sm = \mathbf{P}M \tag{1}$$

here s is a scale factor and $\mathbf{P}$ is a 3x4 projective matrix that can be decomposed as:

$$\mathbf{P} = \mathbf{A}[\mathbf{R}|\mathbf{t}] \tag{2}$$

where $\mathbf{A}$ is the 3x4 matrix of the internal parameters of the camera:

$$\mathbf{A} = \begin{bmatrix} f\alpha & \gamma & c_x & 0 \\ 0 & f\beta & c_y & 0 \\ 0 & 0 & f & 0 \end{bmatrix} \tag{3}$$

with (c_x, c_y) the optical centre of the camera, f its focal length, α and β take into account the pixel physical dimensions and γ encodes the angle between x and y axis of the CCD (skew) and is usually set at 0, i.e. perpendicular axes. The matrix $[\mathbf{R}|\mathbf{t}]$ is a matrix relating the camera coordinate system with the world coordinate one, i.e. the camera position $\mathbf{t}$ and rotation matrix $\mathbf{R}$:

$$[\mathbf{R}|\mathbf{t}] = \begin{bmatrix} \mathbf{R} & \mathbf{t} \\ \mathbf{0}_3^T & 1 \end{bmatrix}. \tag{4}$$

2.2 Epipolar geometry

Let's now consider two images of the same three dimensional scene as taken by two cameras at two different viewpoints (see Fig. 1). Epipolar geometry defines the imaging geometry of two cameras, either a stereoscopic system or a single moving camera. Given a three dimensional point M and its projections m and m' on the two focal planes of the cameras, the three points define a plane Π which intersects the two image planes at the epipolar lines l_m and $l_{m'}$ while e and e' are the epipoles, i.e. the image point where the optical centre of the other camera projects itself. The key point is the so called epipolar constraint which simply states that if the object point in one of the two images is in m, then its corresponding image point in the other image should lay along the epipolar line $l_{m'}$ Such a constraint can be described in terms of a 3x3 fundamental matrix through the:

$$m'^T \mathbf{F} m = 0. \tag{5}$$

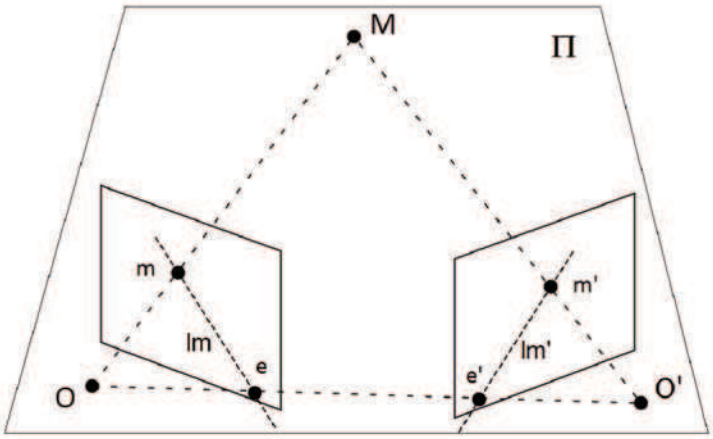


Fig. 1. Epipolar geometry.

The fundamental matrix **F** contains the intrinsic parameters of both cameras and the rigid transform of one camera with respect to the other and thus describes the relation between correspondences in terms of pixel coordinates. A similar relation can be found for the so called essential matrix where the intrinsic parameters of the cameras are not considered and the relation between correspondences is in terms of homogeneous coordinates. The algorithms for the estimation of epipolar geometry deal with actual pixel positions as produced by actual lenses and cameras. Therefore the interest of such algorithms is in the fundamental matrix rather than in the essential one. The standard approach for the computation of the fundamental matrix is based on the solution of a homogeneous system of equations in terms of the nine unknowns of the matrix **F**:

$$\mathbf{Zf} = 0 \tag{6}$$

where

$$\mathbf{f} = (f_1, f_2, f_3, f_4, f_5, f_6, f_7, f_8, f_9)^T \tag{7}$$

and

$$Z = \begin{pmatrix} x'_1x_1 & x'_1y_1 & x'_1 & y'_1x_1 & y'_1y_1 & y'_1 & x_1 & y_1 & 1 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ x'_nx_n & x'_ny_n & x'_n & y'_nx_n & y'_ny_n & y'_n & x_n & y_n & 1 \end{pmatrix}. \tag{8}$$

If nine or more correspondences are known the system is overdetermined and a solution can be sought in a least square sense; in a subsequent step, from the found fundamental matrix, the geometrical information is derived, exploiting the knowledge about the two camera matrices (equation 3). The number of independent unknowns varies among the different approaches employed for the computation. Some approaches don't take into account the additional rank-two constraint on the fundamental matrix (8 point algorithms) and some do (7 point algorithms). Naturally the former considers the rank constraint in a subsequent phase; finally the solution is derived with an unknown scale factor. Let us now suppose that the rigid motion of one camera with respect to the other is *a-priori* known, it is then possible to build directly the fundamental matrix. Let us consider the essential matrix **E** defined as (Huang & Faugeras, 1989):

$$\mathbf{E} = \mathbf{TR} \tag{9}$$

and

$$\mathbf{T} = \begin{pmatrix} 0 & -t_z & t_y \\ t_z & 0 & -t_x \\ -t_y & t_x & 0 \end{pmatrix} \tag{10}$$

here (t_x, t_y, t_z) is the translation vector and $\mathbf{R}$ the rotation matrix. It is possible to rewrite the equation for the fundamental matrix using $\mathbf{E}$ and taking into account the intrinsic parameters matrix $\mathbf{A}$, as

$$\mathbf{F} = (\mathbf{A}^T)^{-1}\mathbf{T}\mathbf{R}\mathbf{A}^{-1}. \tag{11}$$

This equation holds if the intrinsic parameters matrix is equal for both cameras, i.e. a single moving camera is considered as in the ego-motion problem. In the case of stereoscopic imaging, there will be two different matrices ($\mathbf{A}_L$, $\mathbf{A}_R$), one for each of the cameras considered and the equation would be:

$$\mathbf{F} = (\mathbf{A}_L^T)^{-1}\mathbf{T}\mathbf{R}\mathbf{A}_R^{-1}. \tag{12}$$

In practical terms the relation linking the correspondences to the epipolar lines can be considered from two different perspectives. On one side it helps in correctly finding a correspondence to a given pixel since it will lay along the relative epipolar line, on the other side if a given match, obtained with some matching algorithm, is far from the epipolar line relative to the considered pixel, it will be incorrect, i.e. an outlier.

3. State of the art

As described in Section 2 the starting point for epipolar geometry estimation is represented by a set of correspondences between two images of the same scene as taken from different viewpoints. The existing techniques to exploit this pairwise information for fundamental matrix estimation can be classified in three broad classes: linear, iterative and robust. Longuet-Higgins in 1981 (Longuet-Higgins, 1981) opened the way to the computation of scene reconstruction from epipolar geometry through a linear approach. The basic procedure is the so called Eight-Point Algorithm, an algorithm of low complexity but prone to great sensitivity to noise in the data, i.e. error in the pixel position of the correspondences, and to the possible presence of outliers, i.e. incorrectly matched points. The outliers are usually due to error in feature detection and in matching and are in large disagreement with the inliers, i.e. the correctly matched points. Further refinement by Hartley (Hartley, 1995) allowed a sensible amelioration of the original algorithm through a simple normalization of image data. The linear approach solves a set of linear equations relating the correspondences through the fundamental matrix, i.e. solves equation (6). If a large number of correspondences is available, the solution is sought in a least square sense or through eigen analysis determining the fundamental matrix through eigen values and vectors, see (Torr & Murray, 1997). The iterative methods basically try to minimize some kind of error signal and can be classified in two groups: those minimizing a geometrical distance between points, and their is corresponding epipolar lines and those based on the gradient. The most widely used geometrical distances are the Euclidean distance and the Sampson one. They both measure with slightly different means the distance between a correspondence and its relative epipolar line in a symmetric way. Since two are the correspondences, the distance from the first one to the epipolar line originating from the other is computed and then the positions are reversed and the distance of the second from the epipolar line originating from the first one is computed and added to the former. Finally all the contribution are added up and considered in an average value. The minimization can be carried out with different approaches: classical

gradient descent, Levenberg-Marquardt or more refined ones such as the Newton-Raphson technique (Chojnacki et al., 2000; 2004) The main drawback of iterative methods is represented by their incapability to correctly treat outliers. Moreover the iterative methods compute in a intensive way, even if with more accuracy than the linear methods. Finally robust methods are those methods able to cope with outliers and noise, maintaining a relative good accuracy. Most of them are based on statistical methods used to pick a subset of all the available correspondences yielding the best linear or iterative estimation of the fundamental matrix. The basic idea is that if a sufficient number of random extractions of correspondences subsets is performed, eventually a good one, i.e. one composed of inliers only with a limited noise, will be picked out. This implies a large number of linear or iterative estimations, but on a limited number of correspondences. Following (Hu et al., 2008) the best known robust methods are LMedS (Least Median of Squares) (Zhang, 1998), RANSAC (RANdom SAmple Consensus) (Torr & Murray, 1997), MLESAC (Maximum Likelihood Estimation SAmple Consensus) (Torr & Zisserman, 2000) and MAPSAC (Maximum A Posteriori SAmple Consensus) (Torr, 2002). LMedS and RANSAC randomly sample some subset of seven matching points in order to estimate, with a linear approach, the model parameters and use additional statistical methods to derive a minimal number of samples needed since all possible subset can not be considered to save time; the difference between the two is the technique used to determine the best result: on one side the median distance between points and epipolar lines on the other the number of inliers. MLESAC is an improvement over RANSAC and MAPSAC is a further improvement over MLESAC. It must be here noted that there is an important drawback in robust methods: they are usually not repeatable. Since they aleatorily select points there is no certainty that any of these algorithms on a given pair of images will yield the same result if made run more than once. A side effect of this is that it sometimes happens that even if accurate from a numerical point of view (error value) a robust algorithm does not always properly model the epipolar geometry. In (Armangué & Salvi, 2003) a full comparison both theoretical and experimental among many approaches of the three different categories is presented in depth. Besides these classical algorithms, more recently a philosophically different approach has been proposed. Several authors (Chai & De Ma, 1998; Hu et al., 2002; 2004; 2008) have employed a genetic computing paradigm to estimate epipolar geometry. The main idea is to employ an evolutionary approach in order to choose, among the available correspondences, the optimal, or sub-optimal, set of eight points by which the epipolar geometry estimation can be carried out with minimal error. In these genetic approaches each individual is represented by a set of pairs and the algorithm is able to change a subset of these during the temporal evolution, measuring the fitness of each individual with the already mentioned geometrical distance functions. Therefore these evolutionary approaches can be considered part of the robust algorithms family. In conclusion all of the algorithms in literature start with the the available correspondences and try to estimate the fundamental matrix solving the epipolar constraint equation, trying to avoid with different means the faulty matches. The roto-translation between the cameras is then computed with a single value decomposition (SVD) method. It is interesting to note that no constraints are placed on the fundamental matrix from geometrical considerations on the final roto-translation. In other words the fundamental matrix is computed regardless to the possibility that the resulting roto-translation between the cameras may be physically incorrect or even impossible.

4. Epipolar geometry estimation using a genetic approach

As briefly reminded in Section 3 the genetic approaches found in the literature evolve their characteristics in order to pin out the set of correspondences data point able to perform best in a standard computation of the geometric parameters. The idea underlying the present algorithm is different, the evolutionary approach is exploited in a more *natural* way. A set of solutions for the epipolar geometry estimation (i.e. a set of roto-translations) is hypothesized, then it is tested against the available experimental data points, genetically evolving the initial hypotheses into better and better ones until an optimal or near optimal solution is reached. The evolutionary algorithm goes through the standard logical steps of any genetic approach. An individual is defined and a fitness function is designed in order to measure the individual ability to solve the task. Finally a reproduction phase is implemented inserting some kind of variability in the genetic pool.

4.1 The individual

Each individual of the population of N estimators of the epipolar geometry is implicitly a possible fundamental matrix and is implemented by a vector of six real values representing the chromosomes:

$$i_i = [\theta, \phi, \psi, t_\alpha, t_\beta, \sigma] \tag{13}$$

these chromosomes are the three angles of a three dimensional rotation (the pitch, roll and yaw angles), two direction cosines for the translation and a sixth value that is the standard deviation of a normal distribution used for the chromosome mutation, as it will be explained in more detail later. The translation is here described with the two direction cosine only, since the solution for any epipolar geometry estimation is always found with an unknown scale factor. In other words the epipolar geometry is insensitive to scale, a scene can be viewed either at a close distance with a short translation between cameras or at far with a large translation, with no difference on the fundamental matrix. In order to superimpose a metric to the environment a simple calibration step considering a known distance measurement about the image must be added.

4.2 The fitness function

Each individual is used to compute a fundamental matrix following equation (11) if considering an ego-motion case or equation (12) for stereoscopy. Each individual must be tested against the environment, i.e. the correspondences, employing a fitness function, whose design is critical for the success of the algorithm. The implemented function takes into account the following two aspects of the problem: on one side the interest in having a low geometric error between correspondences and their relative epipolar lines and on the other in maximizing the number of correct correspondences, i.e. the inliers. Thus the fitness function has been defined as the ratio between the number of inliers found and the total symmetric transfer error:

$$f = \frac{n_{inliers}}{E} \tag{14}$$

where

$$E = \sum_{i=0}^{N} [d(x_i', Fx_i) + d(x_i, F^T x_i')] \tag{15}$$

is the symmetric transfer error and $d(x, y)$ is the standard Euclidean distance. This error is the sum of the distances between a given point and the epipolar line relative to its corresponding point plus the symmetric distance obtained switching the points; finally it is summed over the

full set of available correspondences. Naturally it may be convenient to use a relative error measure in order to weight in an opportune way those correspondences which are very close to each other and that may be less reliable in the geometry estimation, using:

$$E = \sum_{i=0}^{N} \frac{d(x_i', Fx_i) + d(x_i, F^T x_i')}{d(x_i, x_i')}. \tag{16}$$

The number of inliers is defined as the number of correspondences for which

$$[d(x_i', Fx_i) + d(x_i, F^T x_i')] < \lambda \tag{17}$$

where λ is a threshold, empirically determined. As an alternative it can be also employed the Sampson distance, defined as:

$$E = \sum_{i=0}^{N} \frac{(x_i' F x_i)^2}{(Fx_i')_1^2 + (Fx_i)_2^2 + (F^T x_i')_1^2 + (F^T x_i)_2^2} \tag{18}$$

here the subscripts indicate the k-th entry of the vector. The experiments have shown that the results in the algorithm performance are practically insensitive to the used error distance.

4.3 Reproduction

At each time step a subset of individuals, represented by the best 20%, is allowed to reproduce. This subset gives rise to a new generation of full 100% individuals through a five fold replication of the chosen subset, affected by a mutation mechanism in order to search further in the solution space. This mechanism is implemented with a random extraction of a number from a normal distribution as the displacement around the current value of a single chromosome of the individual. The standard deviation of the normal distribution employed is that of the sixth chromosome in the individual, see equation (13). Thus this mutation amplitude becomes itself part of the genetic algorithm optimization strategy. In more detail, of the six chromosomes one of the five geometrically meaningful is chosen with equal probability. This is mutated adding a value randomly extracted with a normal distribution of zero mean value and standard deviation as indicated by the sixth chromosome,

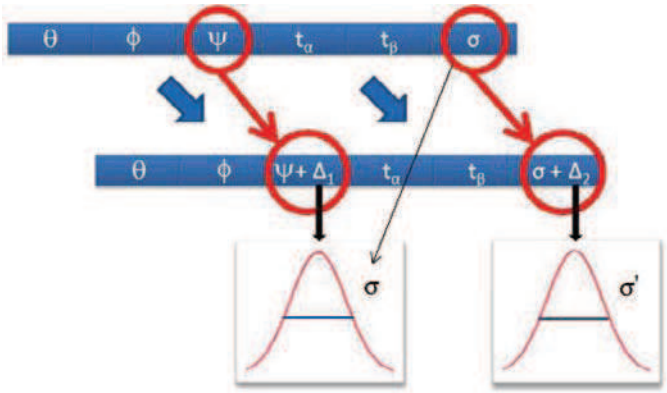


Fig. 2. Mutation implementation.

see Figure 2. The sixth chromosome is then itself updated through a similar mutation with a random extraction of normal distribution with zero mean and a fixed $\sigma' = 0.4$. The individual with the overall best fitness is always retained at each time step.

4.4 Outliers detection and exclusion

A very important issue to discuss here is relative to the outliers. As seen in Section 3 one of the main concerns of the algorithms in literature is the individuation of outliers and their elimination for an accurate estimation of epipolar geometry. These outliers originate from inaccurate performance of the image processing algorithms resulting in errors in feature detection and in matching. In the presented approach the fitness function computation (equation (14)) easily and naturally shows which of the point pairs are outliers, as it will be experimentally presented in Section 5. After a few iterations the error distribution over the experimental pairs relative to the best individual, computed with equation (15), clearly shows which of the points are inliers and which outliers, permitting the limitation, in the following time steps, of the number of pairs used for the fitness function computation. In other words the algorithm is capable to perform the detection and exclusion of outliers in a fully automatic way. This detection is performed through a threshold value to isolate those correspondences yielding a too large error in the best individual. The cutoff value can be chosen as three times the standard deviation in the error distribution, since the expected value for the error is null.

4.5 Algorithm flow

The algorithm flow is as follows.

Initialization: a population of 100 individuals is created with random values for θ, ϕ, ψ, t_α, t_β and $\sigma = 10$.

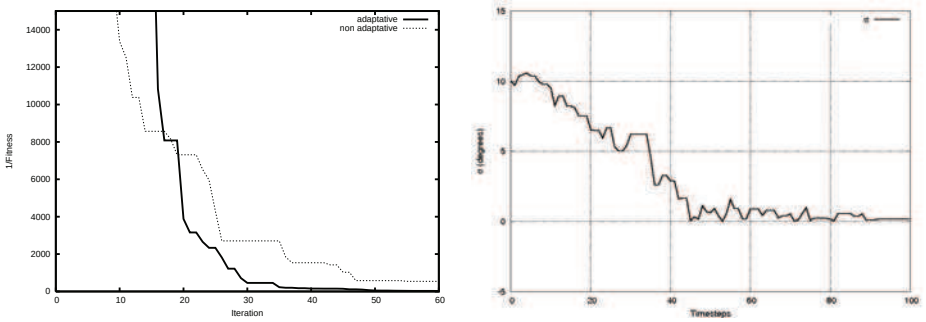
Main loop:

1. For each individual
   (a) compute **F** (equation (11) or (12))
   (b) compute fitness on the available number of N correspondences (equation (14))
2. order the population with ascending fitness
3. take the 20 best individuals and reproduce with mutation, keeping the best individual
4. goto 1

After a given number of iterations the outliers are removed (K) and the set of used correspondences reduced to $M = N - K$. The genetic algorithm then restarts with this reduced set of correspondences. Presently the genetic algorithm removes its outliers only once and is stopped after a given number of iterations has passed without an improvement in error.

5. Experimental data

In the following experimental data in the two cases of ego-motion computation and stereoscopy are provided. The proposed algorithm has been tested on both synthetic and real images. The synthetic data have been prepared projecting a grid of three dimensional points onto two image planes displaced one with respect to the other via equations (1) and (2), inserting given amounts of Gaussian noise when needed. Also for the purpose of



(a) Adaptative vs non adaptative. (b) Mutation amplitude (σ) as a function of time.

Fig. 3. Algorithm convergence and mutation amplitude.

testing, different quantities of inliers have been added. The real images data come from different sources: some of them derive from a paper (Armangué & Salvi, 2003) presenting a comparative study on epipolar geometry estimation, some have been acquired in the ENEA Robotics Lab using different kinds of calibrated cameras and of moving robots, some are from publicly available data-sets and finally one has been shot in everyday life.

Let us first consider the features of the present approach. The rapid convergence of the genetic algorithm is plotted in Fig. 3(a) where the inverse of the average fitness of the entire population is shown as a function of time for a typical run over synthetic data. Here two different modes are compared: one plot is relative to an adaptive σ in the individual, i.e. undergoing a genetic optimisazion as described in Section 4, while the second is relative to a fixed σ value. It is evident that the adaptivity is an advantage in terms of speed of convergence. The reason for such a behaviour can be understood examining Figure 3(b). Here the time evolution of the mutation amplitude σ is plotted as a function of time for the former of the two modes described. The decreasing behaviour shows how the algorithm searches the solution space at first in a coarse way, with large *jumps* and, as the error decreases, the search becomes finer and finer. This variability accounts for a more efficient search in the solution space, while a constant amplitude in the mutation algorithm forces the algorithm to *jump away* from a good solution, rendering the convergence much longer. In (Armangué & Salvi, 2003) a very interesting comparison is presented among most of the epipolar geometry estimators. Since the relative experimental data are freely available over the Internet a direct comparison of the presented algorithm is possible. In Fig. 4 is shown the robustness of the algorithm against the adding of Gaussian noise to the data point location. The error of the best individual increases linearly with the amount of added noise. These data are compared to the available results relative to the two best performing robust algorithms reviewed in (Armangué & Salvi, 2003), i.e. MAPSAC and LMedS. The presented approach performs better. The capability of the algorithm to identify and remove outliers is evidenced in Figure 5. Here it is plotted the ordered error value for the best individual as a function of the correspondence pair. The plot can be easily divided in two parts: on the left the inliers, with a limited amount of error, on the right the outliers, with a large error. From this plot appears evident how it is possible to separate the two sets via an opportune threshold. If it is assumed that the expected value for the error should be null and that a Gaussian distribution may describe the behaviour, a tentative threshold can be placed at three times the standard deviation in the error distribution. Naturally a limited number of outliers must be assumed. The robustness

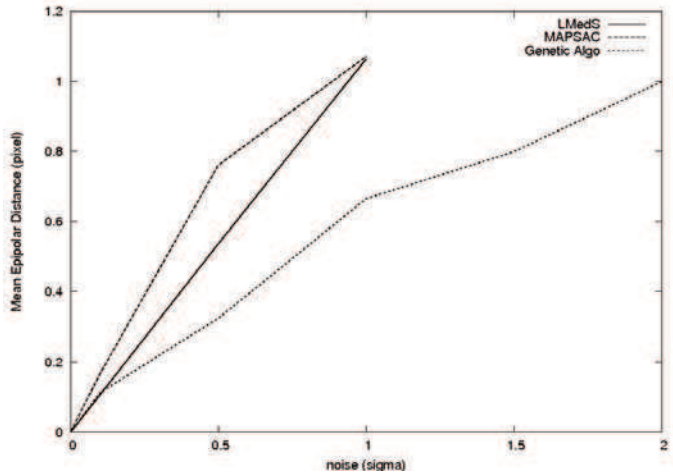


Fig. 4. Performance as a function of noise.

against the presence of outliers is further shown in Table 1 where the algorithm insensitivity is evident. In Figure 6 is shown the relevant data to assess the repeatability of the presented approach. The graphs are relative to 100 runs of the algorithm with different random seeds, i.e. initial conditions, in a real image case (the *kitchen* one, Figure 7(f)). They present the distributions of differences from the average value for the five physically meaningful values of the approach, namely θ, ϕ, ψ, t_α and t_β. Even if the number of runs is not large, the distributions can be considered Gaussians with a null expected value, moreover the spread of the differences is very limited, less than one hundredth of degree for the rotation angles and half a degree for the direction cosines of the translation vector. These data show that the found

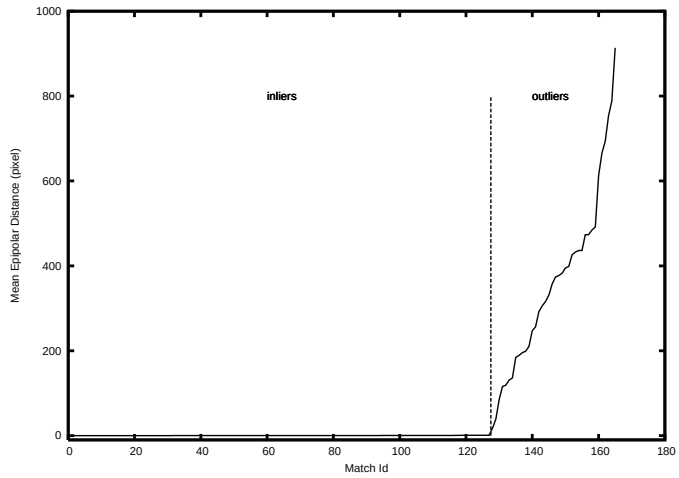
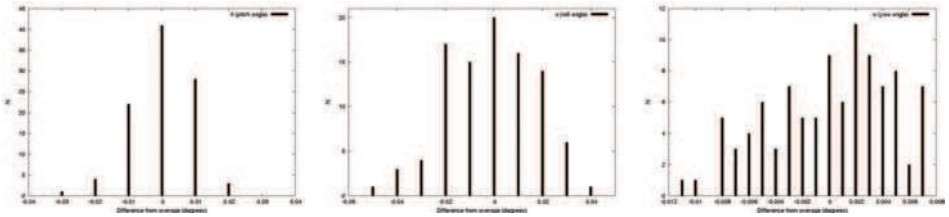


Fig. 5. Error as a function of correspondences for the best individual. The outliers are clearly visible.

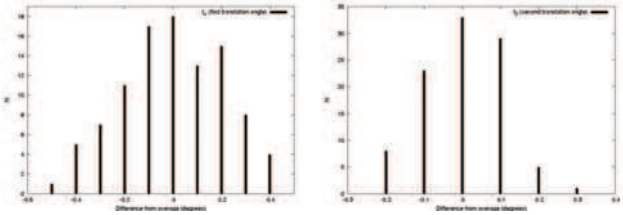
Outliers	GeneticAlgo
10%	0.193 0.123
20%	0.236 0.120
30%	0.169 0.116

Table 1. Performance as a function of percentage of outliers. Every cell show the mean and standard deviation of the error between points and epipolar lines in pixels.

solution is always the same, i.e. that repeatability is no concern for this approach. While robust approaches may found different subsets of correspondences for a given error limit, yielding different fundamental matrices, this algorithm when repeated simply changes the starting point for the search for an extremisation of the same error function (or fitness function), always yielding a near optimal solution in a limited neighbourhood of the actual optimum, properly modelling epipolar geometry. In Table 2 it is presented the algorithm performance on the real images shown in Figure 7. The data are compared to those of the LMeds and MAPSAC robust algorithms. The performance of the presented approach is similar in most cases with that of the published algorithms and with markedly better performances on the *mobile robot* image (Figure 7(b)). The reason of the choice of these two algorithms among those in (Armangué & Salvi, 2003) is the combination of two positive features: they performed well and they yielded the correct epipolar geometry for the used images. As above pointed out, it may actually happen that a robust algorithm gives a good performance in terms of error but with a mistaken geometry. In Figure 8 an example on real images taken by one of the surface robots is shown together with the capability of the algorithm to remove outliers. In this figure most of them are in the central part of the image. In Figure 9 an example of an everyday life, large baseline stereogram is shown. In Figure 9(a) and 9(b) are visible the epipolar lines. In

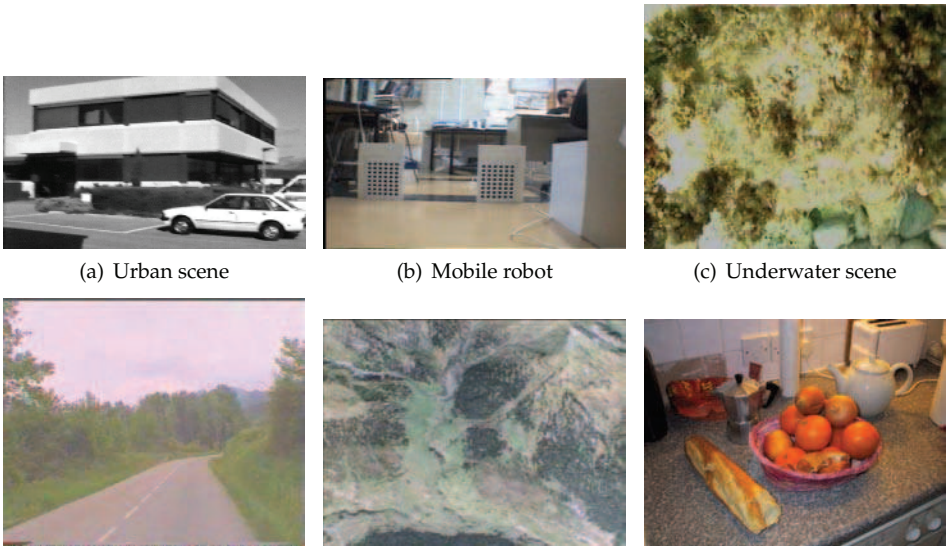


(a) θ(pitch angle) (b) ϕ(roll angle) (c) ψ(yaw angle)



(d) t_α (first direction cosine) (e) t_β (second direction cosine)

Fig. 6. The distributions of distances in pixels from average values over 100 runs.



(a) Urban scene (b) Mobile robot (c) Underwater scene

(d) Road scene (e) Aerial view (f) Kitchen scene

Fig. 7. The real images set from (Armangué & Salvi, 2003)

the second one is also visible the epipole, i.e. the actual location of the centre of projection of the other camera. In Figure 9(c) are shown the displacements of the correspondences from one image to the other. In Figure 10 two classical stereograms and their epipolar lines are presented as computed by the presented approach. The algorithm computed data for Figure 10(b) are: $\theta = 0.0000$, $\phi = 0.0049$, $\psi = 0.0000$, $t_\alpha = 90.0373$ and $t_\beta = -0.0003$ with an average symmetric transfer error of $E = 0.1819$ pixels, in optimal agreement with the actual values, representing a perfect lateral shift.

Image	LMedS	MAPSAC	GeneticAlgo
urban	0.319	0.440	0.393
	0.269	0.348	0.314
mobile robot	1.559	1.274	0.490
	2.715	2.036	0.715
underwater	0.847	1.000	0.848
	0.740	0.761	0.792
road	0.609	0.471	0.433
	0.734	0.403	0.491
aerial	0.149	0.257	0.432
	0.142	0.197	0.308
kitchen	0.545	0.582	0.543
	0.686	0.717	0.571

Table 2. Real images results. Comparison with Lmeds and MAPSAC from (Armangué & Salvi, 2003). Every cell show the mean and standard deviation of the average discrepancy between points and epipolar lines in pixels.

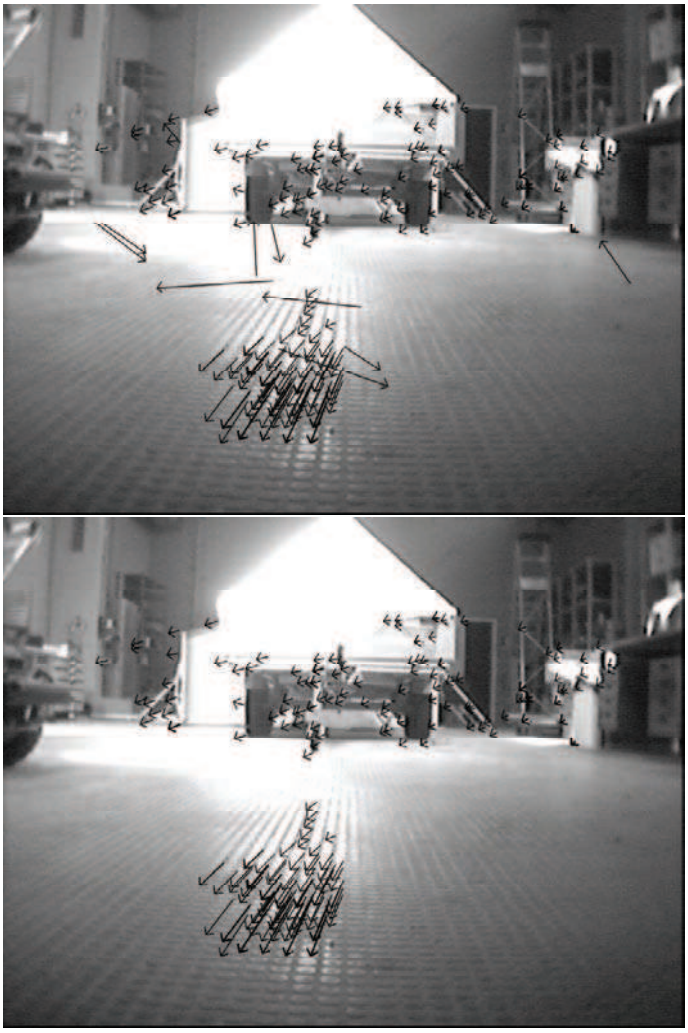
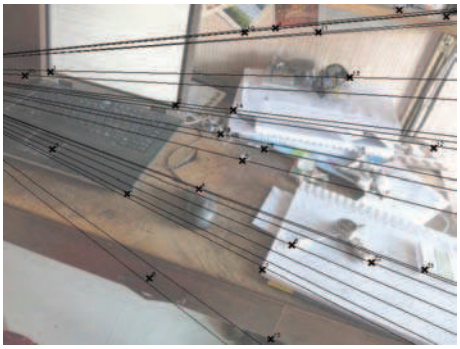


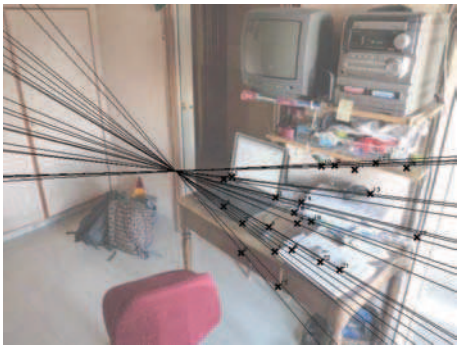
Fig. 8. An example of real image and the outliers removal.

6. Discussion and conclusions

A novel genetic approach for the estimation of epipolar geometry has been here presented. The classical algorithms take as input the whole experimental data, the correspondences, and from them compute the fundamental matrix and then the rigid roto-translation from one camera to the other. The presented approach, instead, tackles the problem in the opposite way, it hypothesize an initial set of random roto-translations and then genetically optimise it against a fitness function computed over the correspondences set. The advantages of the described approach are represented by the following points. The algorithm is sensibly simpler than the ones in literature allowing a more limited computing intensiveness. The convergence is quickly reached. This is especially true if considering the ego-motion problem. If the



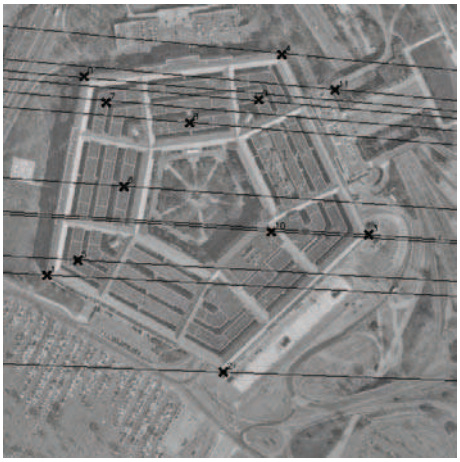
(a) Left image



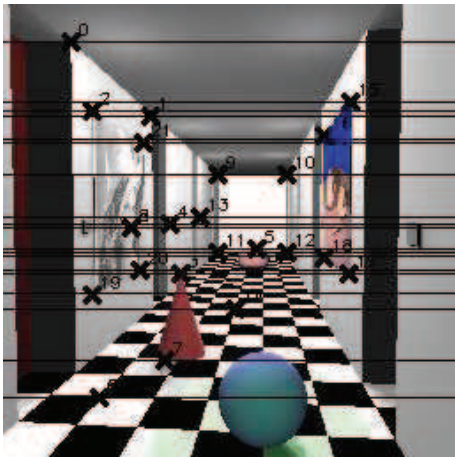
(b) Right image

(c) Optical flow

Fig. 9. An example of large baseline stereoscopy on real images.



(a) Aerial view of Pentagon.



(b) A synthetic corridor.

Fig. 10. Two classical examples of stereograms.

epipolar geometry is exploited to measure the motion of a camera in the three dimensional space, the solution search space of the genetic algorithm can be dramatically reduced. If the camera at the preceding time step was heading in a given direction with a given pose, at the following one will head toward a direction and with a pose not much dissimilar from the preceding ones, with a very high probability. This means that the initial population of roto-translation can be tuned on the basis of the near motion history of the camera and the algorithm will quickly converge to the right solution. With a standard approach the motion information would have been of no use, since the epipolar geometry estimation would have been performed blindly, processing all the data points. In different words, using the present approach it is possible to add constraints on the geometry of the camera system. The genetic design of the algorithm is responsible for the capability to automatically ignore outliers during the computation. The fitness computation is based on the geometric distance between points and epipolar lines relative to their correspondences. If a given roto-translation is the one accurately describing the geometry, then all the correct correspondences will have a quasi null distance and the outliers will yield a large one. This mechanism can be exploited to automatically pin out the outliers from the set of available experimental correspondences also implying the possibility to reduce the data set to the inliers only, gaining in precision and on the quantity of data to process. The results in terms of performance are good. In Section 5 it has been shown that the error is not larger than that of the best performing algorithm presently available in literature, that the robustness against noise in the data is good and also the insensitivity against the presence of outliers among the data point is high naturally. The algorithm has a high degree of repeatability, reaching the same solution even if started from different initial conditions. This feature must be compared to some difficulties in the robust approaches to the same end. Future work will be carried out in the direction of tuning the genetic components of the algorithm and of code optimisation in order to obtain real time performances in the processing of video streams from PAL cameras. This in order to compute visual odometry for underwater robotic platforms, where such a solution may be of help since the number of exploitable sensors is limited. Under water the electromagnetic radio signals are rapidly damped and GPS can't be used, accelerometers or gyroscopes need numerical integration and thus present large errors, unless using expensive ones. Naturally also underwater video sequences are blurred and noisy and need some kind of pre processing, but still this may be a viable solution. Further work will be dedicated to the implementation of a full visual odometry system as a standalone device composed of a camera of the webcam class and of an embeddable computer of the genre of the Gumstix computer-on-module. This means the implementation of a feature extraction stage (e.g. using Harris angles), of an optical flow model (e.g. Lucas Kanade algorithm) and finally of the present algorithm completed with a metric calibration. From the point of view of epipolar geometry estimation in the stereoscopy field of application, further efforts will be devoted towards the exploitation of the computed geometry in order to match targets across different cameras for the purpose of video surveillance and target tracking in cluttered environments.

7. References

Zanela, A. & Taraglio, S. (2002). A Cellular Neural Network Based Optical Range Finder, *Int. J. of Circuit Theory and Applications*, vol 30, pp 271 – 285.

Caballero, F., Merino, L., Ferruz, J., Ollero, A. (2009). "Vision-based odometry and SLAM for medium and high altitude flying UAVs", *J. Intell. Robot Syst.*, vol 54, pp 137 – 161.

Arghaian, H & Cavallaro, A. (2009). "Multi-Camera Networks", *Academic Press*, Burlington

MA.

Longuet-Higgins, H.C. (1981). "A computer algorithm for reconstructing a scene from two projections",*Nature*, vol. 293, pp 133 – 135.

Hartley, R.I. (1995). "In defence of the 8-point algorithm", in *Proc. of the Intl. Conf. on Computer Vision*, pp 1064 – 1070.

Torr, P.H.S. & Murray, D.W. (1997). "The development and comparison of robust methods for estimating the fundamental matrix", *Int. J. Comput. Vision*, vol 24 (3), pp 271 – 300.

Chojnacki, W, Brooks, M.J., van den Hengel, A., Gawley, D. (2000). "On the fitting of surfaces to data with covariances", *IEEE Trans. PAMI*, vol. 22 (11), pp 1294 – 1303.

Chojnacki, W, Brooks, M.J., van den Hengel, A., Gawley, D. (2004). "A new constrained parameter estimator for computer vision applications", *Image Vision Comput.*, vol. 22 (2), pp 85 – 91.

Hu, M., McMenemy, K., Ferguson, S., Dodds, G., Yuan, B. (2008). "Epipolar geometry estimation based on evolutionary agents", *Pattern Recognition*, vol 41, pp 575 – 591.

Zhang, Z. (1998). "Determining the epipolar geometry and its uncertainty: a review", *Int. J. Comput. Vision*, vol 27 (2), pp 161 – 198.

Torr, P.H.S., Zisserman, A. (2000)."MLESAC: a new robust estimator with application to estimating image geometry", *Comput. Vision Image Understanding*, vol 78, pp 138 – 156.

Torr, P.H.S. (2002). "Bayesian model estimation and selection for epipolar geometry and generic manifold fitting", *Int. J. Comput. Vision*, vol 50 (1), pp 35 – 61.

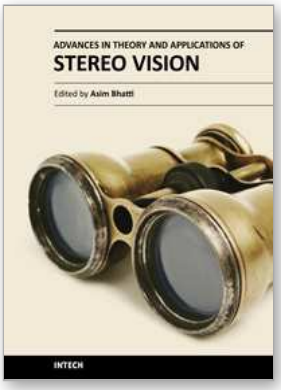
Armangué, X., Salvi, J. (2003). "Overall view regarding fundamental matrix estimation", *Image and Vision Comp.*, vol 21, pp 205 – 220.

Chai, J., De Ma, S. (1998). "Robust epipolar geometry estimation using genetic algorithm", *Pattern Recognition Letters*, Vol 19, pp 829 – 838.

Hu, M., Xing, Q., Yuan, B., Tang, X. (2002). "Epipolar geometry estimation based on genetic algorithm under different strategies", in *Proc. of 6th Intl Conf. on Signal Processing*, pp 885 – 888.

Hu, M., Yuan, B., Dodds, G., Tang, X.(2004). "Robust method of recovering epipolar geometry using messy genetic algorithm", in *Proc. of the First Canadian Conf. on Comp. and Robot Vision*, pp 164 – 171.

Huang, T.S., Faugeras, O.D. (1989). "Some properties of the E matrix in two-view motion estimation", *IEEE Trans. PAMI*, vol 11 (12), pp 1310 – 1312.



Advances in Theory and Applications of Stereo Vision
Edited by Dr Asim Bhatti

ISBN 978-953-307-516-7
Hard cover, 352 pages
Publisher InTech
Published online 08, January, 2011
Published in print edition January, 2011

The book presents a wide range of innovative research ideas and current trends in stereo vision. The topics covered in this book encapsulate research trends from fundamental theoretical aspects of robust stereo correspondence estimation to the establishment of novel and robust algorithms as well as applications in a wide range of disciplines. Particularly interesting theoretical trends presented in this book involve the exploitation of the evolutionary approach, wavelets and multiwavelet theories, Markov random fields and fuzzy sets in addressing the correspondence estimation problem. Novel algorithms utilizing inspiration from biological systems (such as the silicon retina imager and fish eye) and nature (through the exploitation of the refractive index of liquids) make this book an interesting compilation of current research ideas.

InTech Europe
University Campus STeP Ri
Slavka Krautzeka 83/A
51000 Rijeka, Croatia
Phone: +385 (51) 770 447
Fax: +385 (51) 686 166
www.intechopen.com

InTech China
Unit 405, Office Block, Hotel Equatorial Shanghai
No.65, Yan An Road (West), Shanghai, 200040, China
中国上海市延安西路65号上海国际贵都大饭店办公楼405单元
Phone: +86-21-62489820
Fax: +86-21-62489821