STATISTICAL FEATURE POINT MATCHING METHOD

Lucile Martin[1,2], Christophe Leroux[1], Edwige Pissaloux[2]

[(1)] CEA/LIST, BP 6, 92 256 Fontenay-aux-Roses, France {lucile.martin,, christophe.leroux}@cea.fr
[(2)] Université Paris 6, Laboratoire de Robotique de Paris, CNRS/FRE 2507, BP 61, 92 256 Fontenay-aux-Roses, France {martin, pissaloux}@robot.jussieu.fr

KEYWORDS: matching, point of interest, statistics, epipolar geometry.

ABSTRACT :

This paper presents a statistical method to match feature points from stereo pairs of images. The proposed method is evaluated in terms of effectiveness, robustness and computational speed. The evaluation was performed on several pairs of real stereo images of natural scenes taken onboard an unmanned aerial vehicle. The results show that the proposed method reduces the number of incorrect matches and is fast.
Cet article décrit une méthode de mise en correspondance de points d'intérêts extraits d'images stéréoscopiques. Cette méthode a été évaluée en termes d'efficacité, de robustesse et de temps de calcul. L'évaluation a porté sur plusieurs paires d'images prises dans un environnement naturel à partir d'un banc stéréoscopique embarqué sur un drone d'intérieur. Les résultats montrent que la méthode proposée est très rapide et réduit considérablement le nombre de mauvais appariements.

1. INTRODUCTION

3D scene reconstruction is one of the most important basic operations for intelligent vision systems and especially for autonomous robotic systems such as unmanned aerial vehicle (UAV).Moreover, the precision of 3D scene reconstruction is fundamental for an autonomous robot to behave properly in its nearest environment.
Many of existing 3D reconstruction processes use active sensors (telemeter, sonar etc); but passive sensors, such as optical cameras are often more suitable in terms of price, accuracy, calculation speed, non invasivity of the environment etc...
The performance of the vision depth sensing estimators strongly depends on the matching process accuracy and reliability.
Matching interest points is the process of identifying the 2D image points corresponding to a same 3D scene point in a pair of stereo images representing that scene.
Much work on matching points' methods has been done (E. Vincent 2004), starting from simple correlation methods up to more sophisticated method such as the RANSAC iterative process (Fischler, Bolles 1981). Almost all of them are based on 3D scene local data.
The RANSAC method (and the ones derived from it i.e. MLESAC (Torr, 1996)) can be very efficient and reliable but it often leads to long computational time making it unrealistic to use onboard a moving robot and incompatible with video processing rate.
For instance, to self-localize while navigating in 3D scene it is useful to have a fast effective and reliable matching method in order to be able to process up to 24 pairs of images a minute. For such reason, this paper proposes an approach which allows matching feature points of at least 2 pairs of images a second.
This proposition is being validated in the frame of the RobVolInt project, a prototype of UAV using vision to self-localize being developed by the French Atomic Energy Commission (CEA), in collaboration with the IRISA, the I3S (Nice university) and the LRP (University of Paris 6).
Subsequent sections outline the theory and context of validity of the statistical matching inliers filter proposed (section 2), the context of experimentation (section 3), a comparative study of the statistical method and major matching methods (section 4), and a short conclusion (section 5).

2. MATCHING FEATURE POINTS : A STATISTICAL METHOD

2.1 Criteria for image matching method selection

The main issue in matching properly image feature points is the depth reconstruction it allows. If the feature points are matched with their real homologous point, it is then possible to localize the stereo rig relatively to its environment (assuming the cameras are calibrated accurately). Thus, the percentage of mismatches left by a matching method is of high important.
A second major criterion to choose a matching method is the computing time it takes to produce pairs of homologous points.
Moreover, local methods are prone to an important number of wrong matches whereas global methods are frequently time consuming.
A solution would then be a mixed method, like the statistical one outlined in this paper : global as it uses statistical data and local as it eliminates redundant feature points' matches.

2.2 Context of validity

The proposed method is valid under constraints that will be described in this paragraph.
To use this method the camera system should be either a quasi epipolar stereo rig (figure 2) or a mono-camera system equivalent to a stereo rig (figure 1), ie: a camera which movements in between image capture are limited to translations only (not necessary known).

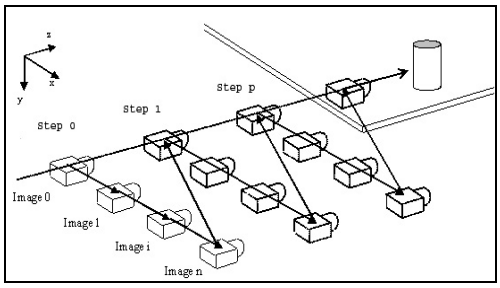

Figure1 : Mono-camera system equivalent to a stereo rig

Indeed, according to (Horaud 1995), it is possible to consider a slightly translated camera is equivalent to a stereo rig.
In case of a stereo rig, the rotation between the optical axes of the cameras should be negligible: the optical axes should be considered collinear. Otherwise, a calibration of the stereo rig would be necessary in order to express all feature points of all images in collinear image frames.

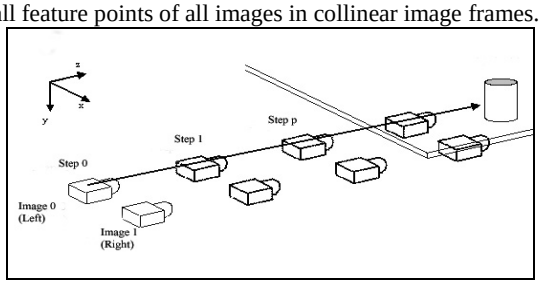

Figure 2 : Stereo rig with coplanar optical axes

It is not necessary to calibrate the system but the calibration of the cameras could be used, as it will be explained in the following paragraph.
To summarize the valid experimental conditions necessary to use the statistical method bellow, one should be able to express all feature points of all images in image frames only distinct from a translation.
Finally, the illumination conditions are supposed constant.

2.3 The proposed statistical method

The whole process is composed of 2 steps:

- generation of a set of feature points' pairs
- estimation of the stereo images apparent movement direction.

The first step of the method is the generation of a set of feature points' pairs. It can be done using a correlation similarity measure or a KLT tracker estimation of homologous points. The method just has to be of a low computational time.
In case of non-calibrated cameras, the second step estimates the orientation of each line defined by a pair of points of the set generated at first step (figure 5). This is equivalent to the estimation of the global apparent movement between two images.

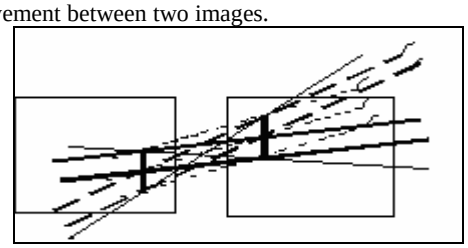

Figure 3 : Estimating homologous pair's orientations.

The most frequent orientation of the matches is considered the good one, and only the matches having that orientation are kept (figure 3 and figure 4).

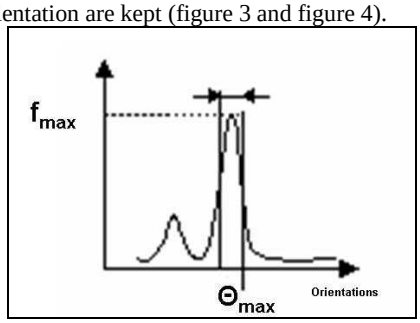

Figure 4 : frequencies of the orientations

This assumption is possible because the vision system used is equivalent to a stereo rig considered quasi stereo rectified, which means the only non negligible transformation between the left and right cameras is the translation in the common image plan. All points move following the same direction representing the 3D real translation in the 2D image frame. One should notice the feature points do not move following the same exact translation in terms of direction and distance. Objects close to cameras and objects far from cameras do not move with the same speed in images.
In case of calibrated systems, the calculation of the most frequent orientation of the feature points matches does not occur: the image dominant (thus global) orientation is the orientation of the line defined by the principal points of the cameras. The matches to be kept are the ones parallel to that orientation. In fact the translation between the principal points' is the same as the projection in the 2D image frame of the 3D translation between the cameras. Therefore, if the stereo rig was exactly stereo-rectified, all matches would have to be horizontal.
To optimize this statistical method execution time, it can be noticed that the image global orientation should be computed once, from a first pair of images which gives the reference matching orientation. On subsequent pairs of images a simple check of parallelism should be performed.

2.4 Interest of the proposed method

The information obtained by this method is the global orientation, in 2D image frame, of all homologous points' pairs.
Consequently, all feature points' pairs not parallel to the reference orientation should be eliminated.
Considering the case of a stereo rig, the obtained orientation is the same as the one between the principal points of both cameras. This leads to important information on the estimation of the principal points' coordinates of both cameras
In fact, the estimation of the principal point's coordinates is a crucial issue in terms of depth computation. In the case of a monocular camera system, the error of estimation of these coordinates is of less importance as it is compensated while being propagated from an image to the other (Horaud 1995).
In the case of a stereo rig, an error of estimation is done on each camera's principal point. The depth computation comes from the comparison of the images taken by these cameras. As the errors aren't the same for both cameras, they don't compensate each other like in the monocular

case. In fact, it is not easily possible to know witch error part is compensated or increased in the final error. Knowledge of the orientation of the line defined by both principal points gives the sum of the errors done on the principal points' estimation, and thus the depth computation accuracy can be highly improved.

3. EXPERIMENTAL EVALUATION OF THE PROPOSED STATISTICAL METHOD

The proposed method (as well as all methods presented in the next section) has been tested with real indoor 640 x 480 size images, one example is shown in figure 5.
Images were taken under usual inside lighting conditions. It wasn't necessary to turn the light on if there were windows in the room as the used cameras were 1 lux sensitive. The lighting conditions were considered constant.
The images were taken in grey scale format.
The distance between the rig and the observed objects from the scene was no more than 3 meters.
The observed objects were at least 50 cm from cameras.
There could be occultation of parts of the 3D scene, but the image processing didn't deal with it.
All algorithms have been developed using C++ with Qt graphical user interfaces on a usual PC platform (Pentium 4 2GHz processor).

4. RESULTS

The efficiency of the proposed statistical method has been compared with most popular matching methods: correlation, RANSAC, KLT tracker, epipolar geometry constraint.

4.1 Correlation

Correlation is the most common method by which feature points in different images are compared. It measures the similarity between image points' neighborhoods. Several similarity methods exist. The one shown here is the zero mean sum of absolute deviation, which gave the best results with the images obtained from the UAV(figure 5).
Without knowing the structure of the scene and without the use of any criteria of unicity or scene symmetries etc (Horaud 1995, Vincent 2004)…correlation based matching method will produced some mismatches (table 1) : 21 percent of the matches obtained with this method were mismatches.

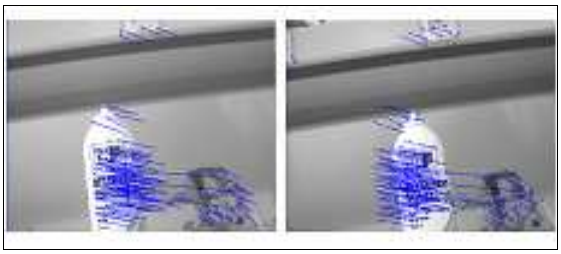

Figure 5: Correlation methods (Zero mean Sum of Absolute Deviation)

This is why constraints such as epipolar geometry of the system are required.

4.2 Epipolar geometry constraint

The epipolar geometry constraint is illustrated in figure 6.

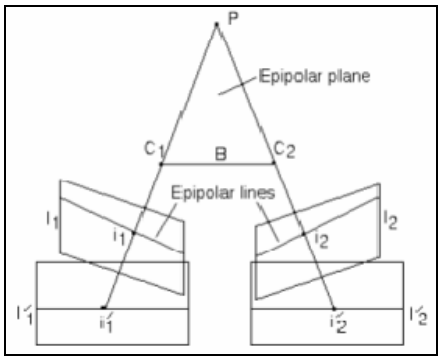

Figure 6: Stereo rig epipolar rectification

It shows the homologous point of a point from the left image will be found on a line in the right image frame which equation parameters can be computed from the cameras calibration parameters (Horaud 1995, Vincent 2004)].
Therefore, using this method requires an accurate calibration method, especially for the optical centers coordinates' estimation. Experiments revealed the Tsaï calibration method implemented in the OpenCV library is not accurate enough to consider using epipolar constraint. Moreover, it is very difficult to use with images taken onboard a flying robot: UAV movements are not stable enough.

4.3 RANSAC

The RANSAC was introduced by Fischler and Bolles in 1981.
It is commonly used to estimate the fundamental matrix from a set of feature point pairs determined by a correlation (Fischler 1981, Vincent 2004).
In the RANSAC fundamental matrix estimation scheme, 7 (or 8) pairs of feature points (or 8) only are randomly selected at each iteration. A fundamental matrix is computed from theses 7 (or 8) pairs, and is tested against all candidate matches. The cardinality of the set of matches which fits with this matrix is a measure of the accuracy of the fundamental matrix.
In practice, fundamental matrices are computed this way until a number of iteration predefined with regards to the allowed computational time or until one agrees with a given minimum number of pairs of feature points.
It results in a high computing time not compatible with a frequency of 2 to 24 pairs of images processed per minute. Moreover 10 percent (table1) mismatches (figure 7) still exist.

Figure 7 : RANSAC filtering method

4.4 KLT Tracker

The KLT Tracker is one of the most popular method used to track feature points from one image to the other in robotic applications using vision to self localize. The

main advantage of the method is the fact the feature points are extracted (with an extractor such as SUSAN or Harris & Stephens) only once, the extracted are then tracked in the following images using a pyramidal approach. (Tomasi, Kanade 1991)
The pyramidal processing reduces significantly processing temporal complexity, thus computational time, but a significant number of mismatches occur (fig 8): 15 percent of the obtained pairs of feature points are wrong (table1).

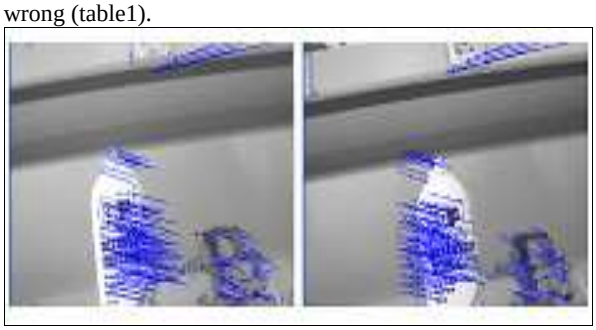

Figure 8 : KLT Tracker method.

4.5 Comparison of the described methods

In order to be able to compare all the previous methods quantitatively, they have all been tested on the same sets of images. For each pair of image, the total number of correct pairs has been calculated manually. This number has then been compared with the number of pairs found by each method (total differential method). Considering the number of pairs found by each method, the percentage of mismatches has been computed for each of these methods (mismatches percentage). The computational time has been recorded for each pair of images processed. Finally, means of the results obtained processing images pairs with each method have been computed and collected in table 1.

Method	Total differential percentage	Mismatch percentage	Computing time
Correlation	+ 2%	21%	1,05212s
Correlation + RANSAC	-15%	10%	1,05212s +6,13219s
Correlation+ Orientation filter	-30%	1%	1,05212s +0,000059s
KLT Tracker	Nonsense (Points are tracked not matched)	15%	0,052s
KLT + Orientation filter	-12%	1%	0,052s +0,000059s

Table 1 : Compared efficiencies of the described methods

The results of this table show the statistic method outlined in this paper is fast and reliable: only 1% of the matches obtained after using the orientation filter are false. The other methods that were tested here left much more mismatches: 10 % after a RANSAC filter. If used after a KLT, it decreases the percentage of mismatches from 15 % to 1%.
But, as shown in the first column of the table 1, while the use of the orientation filter decreases considerably the percentage of mismatches, it also represents an important loss of initial information: only 70% of the real good matches are kept in the case of a correlation, and 88% in the case of the KLT matching initial step.
A compromise has then to be found between the quantity of information required and the quality of the matches found.

5. CONCLUSION

This paper has addressed a new fast, reliable and effective method for stereo rectified rig image matching. The proposed statistical method is a mix of local and global image characteristics : local, because based on feature (interest) points, and global, because based on homologous direction conservation between matched images…
The proposed statistical method to detect mismatches is of very low computation time and produces very few mismatches. A loss of only 15% of the information initially detected has been found.
The obtained results were very satisfying for the UAV depth recovering, and thus the UAV self localization, but the conditions of experimentations and the design of the stereo rig makes it lack of polyvalence.

6. REFERENCES

Fischler, Bolles, 1981 *Random sample consesus :a paradigm for model fitting with application to image analysis and automated cartography*, communications of the ACM, vol. 24, no.6, pp381-395, 1981

Guermeur Ph., 2002, *Vision robotique monoculaire : reconstruction du temps-avant-collision et de l'orientation des surfaces à partir de l'analyse de la déformation*, (PhD) Thèse de Doctorat, Université de Rouen, 2002

Horaud R., Monga O., 1995, *Vision par ordinateur*, Hermes 1995

Lowe, 1999, *Object recognition from local scale invariant features*, proc. Int.conf. on computer vision, vol 2, pp.1150-1157,1999

Pissaloux, E., E., Le Coat, F., Tissot, A., Durbin, F., 2000, *An adaptive parallel system dedicated to projective image matching*, IEEE ICIP 2000 (Int. Conf. on Image Processing), Vancouver, Canada, 10—13 September, 2000 pp. 184—187

P.H.S. Torr, A. Zisserman , 1996, *MLESAC: A new robust estimator with application to estimating image geometry* (1996) Computer Vision and Image Understanding

Tomasi Carlo and Kanade.Takeo, 1991, *Detection and Tracking of Point Features.* Carnegie Mellon University Technical Report CMU-CS-91-132, April 1991.

Etienne Vincent, 2004, *On feature point matching, in the calibrated and uncalibrated contexts, between widely and narrowly separated images*, PhD thesis of The Faculty of Graduate and Postdoctoral Studies of the Ottawa-Carleton Institute for Computer science School of Information Technology and Engineering, 2004

7. ACKNOWLEDGEMENT

We are most grateful for the support from the DGA.
The support of the LISRA is gratefully acknowledged especially M. Laurent ECK, and also the support of M. Jean-Pierre MARTIN for his programming support.