

A fuzzy-logic classifier for estimating the reliability of the self calibration of an embedded stereovision system

Sebastien Kramm*, Pierre Miché**, Abdelaziz Bensrhair***

LITIS, INSA Rouen

Place Emile Blondel

76131 Mt St Aignan, France

Abstract—Estimation of epipolar geometry can be done automatically in on-board stereovision systems, using interest points that are detected and matched. However, image disturbance that can happen in real-life situations can considerably lower the performance. A reliability score computing method is proposed, based on a fuzzy logic classifier. Its input is the data extracted from the estimation process. The classifier is trained with artificial image disturbance, using a set of typical image pairs. Results show that the computed score is indeed related to the performance of estimation.

I. INTRODUCTION

A. Stereovision in consumer vehicles

Implementation of real-time obstacle detection systems based on stereovision onboard of consumer vehicles faces a practical challenge. Most of the real-time algorithms rely on an aligned camera set. This makes matching easier, as the matching step is done in one dimension, separately for each image line. But this constraint is unrealistic in industrial implementations : it is simply impossible to have such a rigid structure between the two cameras if they are, say, in the headlights of the vehicle. Moreover, an automobile is a rough environment, and vibrations will induce mechanical drifting.

In order to overcome this constraint, one approach relies on calibrating the camera rig, that is, determine the two cameras relative orientation. This information is contained in the well known *Fundamental matrix*. Its knowledge can be used to compute planar rectification matrices, that are then used to rectify images before the matching and scene analysis step. Methods to compute rectification matrices from the sole Fundamental matrix with minimal image distortion have been published [8], [6].

This calibration step can be done automatically, using only the image content. The video stream input can be used, using optical flow methods, but it has been showed that a reliable estimation can be done using a single image pair. We will focus here on this latter approach, known as *weak calibration*.

However, another problem remains : the onboard system will have no information about the computed fundamental matrix reliability. This is particularly annoying in an environment with heavy security constraints. We know that

under certain circumstances, the computed matrix can be completely erroneous, thus making the whole system non functional. For example, under bad visibility conditions, or if the current scene does not present sufficient content. It is necessary for the onboard stereovision system to have a reliability information associated with the geometry estimation itself. This way, it can choose to use — or not — the current estimation for image rectification.

We present here an approach based on fuzzy-logic concepts and is aimed at real-world situations. Our method is able to produce such a reliability score for every numerical estimation. The classifier training is based on experimental data, showing that reliability can be related to several criterions extracted from the estimation process.

B. Previous works

Several approaches have been proposed characterising uncertainty of estimated Fundamental matrix.

Csurka *et al.* [1], followed by Zhang [10], propose a general frame for characterizing uncertainty of the estimated Fundamental matrix, based on its covariance matrix computed using the input matches. However, this approach is based on mathematical properties of the used data, and does not seem to be appropriate in real-world situations. It is in fact theoretically possible to have a numerically "perfect" Fundamental matrix that is completely erroneous, if an important part of image is missing. Kanatani [3] considers the case of an estimation computed using the optical flow method, so it can not apply in our case. Leclerc *et al.* [5], [4] propose the idea of algorithm *self-consistency*, and gives a general framework for this concept. However, it is mainly oriented towards the multi-image method, and does not seem to be useful in our case.

II. WEAK CALIBRATION OF CAMERA SET

Weak calibration is the process of estimating relative orientation of the two cameras from the sole image pair. We consider here the now classical scheme : interest point (IP) detection, matching, and Fundamental matrix robust estimation.

A. Interest point detection

Interest point detectors have been a widely covered topic in the past years. They are usually based on gradient methods. Detectors can be divided into two categories :

*sebastien.kramm@univ-rouen.fr

**pierre.miche@insa-rouen.fr

***abdelaziz.bensrhair@insa-rouen.fr

- basic detectors, that only provide coordinates of point. The best known of these is the "Harris" detector,
- local descriptors, that also provide a characterization of the detected points.

The latter category is in the present case much more appropriate, as there is no need for a characterisation step. Among this group, the SIFT detector [7] has shown its performance in providing excellent reliability in the matching phase. It has also been shown that this detector is highly insensitive to rotation and scale variation. For a complete review of such local descriptors, see [9].

B. Matching

The matching strategy is inspired from the one proposed in [7]. It uses the attribute vector associated with interest points, and is very close to a "nearest neighbor" algorithm. For each interest point in one image, the interest point of the other image giving the smallest Euclidian distance between their vectors is searched. This is done by parsing the list of points of a given side, and, for each point p_A , searching in the opposite list the point p_{B1} that has the minimum Euclidian distance with p_A .

Euclidian distance between two points is defined as :

$$d(p_A, p_B) = \sqrt{\sum_{i=1}^n (m_{Ai} - m_{Bi})^2} \quad (1)$$

with m_{Ai} and m_{Bi} being the attribute-vector element i of points p_A and p_B , and n being the size of the attribute vector ($n = 128$ for the original SIFT implementation).

In order to discriminate ambiguous matches, the point p_{B2} corresponding to the second best choice for matching is also searched. The Euclidian distance between the points p_A and p_{B1} is noted d_1 , the distance between p_A and p_{B2} is noted d_2 . A match is validated only if two conditions are met :

- the distance d_1 is sufficiently lower than distance d_2 , explicitly if $d_1/d_2 < r$,
- the point p_{B1} has not been used in a match yet.

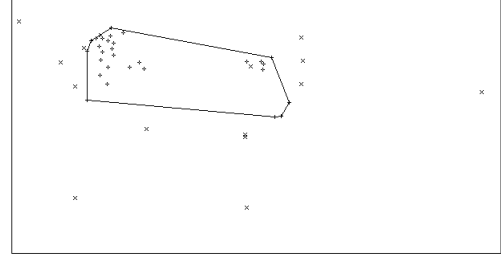
The coefficient r is usually equal to 0.6 : a lower value will induce a more restrictive matching, resulting in a lower number of matches. The second condition above is a unicity constraint : an interest point can only be part of one match.

Search space is limited by a geographical distance constraint. This constraint uses three different values : horizontal distance, vertical distance, and diagonal distance. Indeed, in a common horizontally aligned camera set, the vertical pixel distance of a match will remain small if camera rotations are limited, so the threshold value can be quite smaller than the horizontal value.

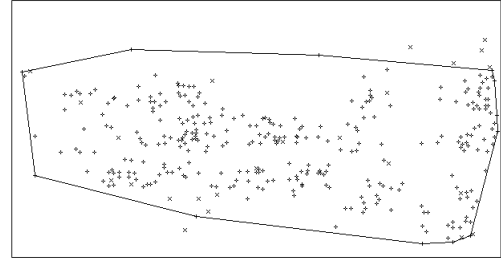
C. Fundamental matrix estimation

We use here classical and easily available robust estimators, RanSaC and LMedS. These estimators have an interesting property : besides giving an numerical estimation, they also split the initial matches set into two subsets, known as inliers and outliers. This information is interesting for reliability characterisation : an estimation based on a set of

matches that covers a large area of the image will have a chance of being more reliable than one based on a small area. To measure this knowledge, we consider the convex hull of used points on each image, as shown on fig. 1. We can assume that the larger this polygon is, the better the estimation will be. We compute this surface S_p , and consider the ratio S_p/S , with S the surface of the used image ROI¹, usually the top two thirds of the image.



(a) Matching example 1



(b) Matching example 2

Fig. 1. An example of matching results with an equivalent number of points. On 1(a), one can see the left image ROI with a low number of matches, and, among these, the ones used by the estimator generate a small surface compared to the ROI. On 1(b), one can see that generated surface is considerably higher, thus corresponding estimation will be of higher reliability

Performance differences between LMedS and RanSaC are minor. Due to space limitations, we only present here results using the latter algorithm, as it showed slightly higher performance.

III. RELIABILITY COMPUTING SCHEME

The idea presented here is to use a fuzzy logic classifier to compute a reliability score associated with an estimation.

In the Fundamental matrix estimation process, shown on fig. 2, we can access several criterions that we will show to be related to overall performance. For example, the number of interest points detected is a valuable information : if it is really lower than a normal value, then this might mean that the scene is not suitable. For each of these criterions, we define a fuzzy membership function to determine its correctness. All the fuzzy variables are then combined, in order to achieve a score, defined here as a fuzzy variable ($\in [0, 1]$). Defuzzification (i.e. obtaining an answer to the question "Estimation of F is good enough") can be achieved using a threshold. Its value can be used as an external

¹Region Of Interest

constraint parameter: high values will be uncommon, but will guarantee a high level of confidence in the numerical estimation. Typical values can be between 0.6 to 0.8.

A. Classifier training

Like most classifiers, ours relies on a training step. It is achieved here experimentally, on real image pairs that are progressively perturbed. We consider here three type of image perturbations, that we believe to be representative of real-world situations.

First, we consider what would happen with an optical lens with bad focus. This is simulated here through gaussian blurring, with 15 σ values from 0.2 to 3.0 (see example on fig. 3).



Fig. 3. Example of blurred images with $\sigma = 3.0$ (only left image is shown)

Secondly, we consider the case of occlusion. We simulate what would happen if one of the cameras field of view is covered by an object. This is done by generating 15 occluded images, from 10 to 80% of S (see example on fig. 4). Thirdly, we consider brightness changes. This is particularly



(a) 30%

(b) 70%

Fig. 4. Example of occluded images (only left image is shown)

important, as in real world implementation, lighting is a difficult parameter to handle. A car can frequently and abruptly change from shade to full sunlight, or both cameras can have different shutter settings inducing different overall brightness of image. We simulated this by generating 15 images, with brightness from 40 to 300% (see example on fig. 5).



(a) 60%

(b) 280%

Fig. 5. Example of image brightness change (only left image is shown)

B. Distance Criteria

In order to characterise estimation performance, necessary for the training phase, we need a distance criterion between the estimated fundamental matrix and a ground truth.

For the ground truth, as we have no information about the real geometry, we manually define matches in both images, for each pair of images used here (see fig. 7). These points are used to compute a reference fundamental matrix, using the normalised 8-points method [2]. As no particular optical distortion correction is done, this matrix might not necessarily have a high accuracy, but experiments show that it is sufficient for the application described here. What we need is a static matrix that will be used as a reference when the image is modified, so low accuracy is not essential. We are mainly interested by the evolution of the distance between this reference matrix and the estimated one.

For the distance between fundamental matrices, we use the distance defined in [10], which gives a value in pixels. It is based on epipolar geometry and a Monte Carlo technique, by randomly sampling image space. For each sample, associated epipolar lines in other image are computed, using both matrices. A random point is selected on these lines, and distance between points and lines are computed. This is done symetrically for the whole image space. By averaging computed distance, this gives us a value d (pixels), that is a measure of distance between the two fundamental matrices.

We used here 200000 random points, which is roughly the area of the considered images. Those were of "VGA" size, using as ROI the top two thirds of the image. This choice is related to the considered application (embedded scene-analysis systems), where the bottom of the image doesn't contain much information.

C. Information used for the reliability computing scheme

Several indicators are available during the estimation process. We will show that some will indeed have their value directly related to global performance, others will not, or less. The items are :

- Image mean gray level (left and right) : M_L, M_R
- Number of IP in image (left and right) : N_L, N_R
- Number of matches : N_M
- Number of matches used for estimation of fundamental matrix : N_F
- Surface of the hull of the points used for estimation : S_p

However, we can not use these values directly, as they are closely related to image size or an other parameter. For example, the number of matches is meaningless on its own, as it is related to the initial number of detected IP. Also, the number of IP is related to the detector sensitivity and to the size of the image or ROI used. It is much more meaningful to consider relative values. We also have to consider differences between extracted data in both images. For example, if 800 IP are extracted in one image and only 400 in the other one, then this pair is unlikely to produce a reliable estimation, although the number of IP, taken alone, is correct. So we choose to monitor the following derived items :

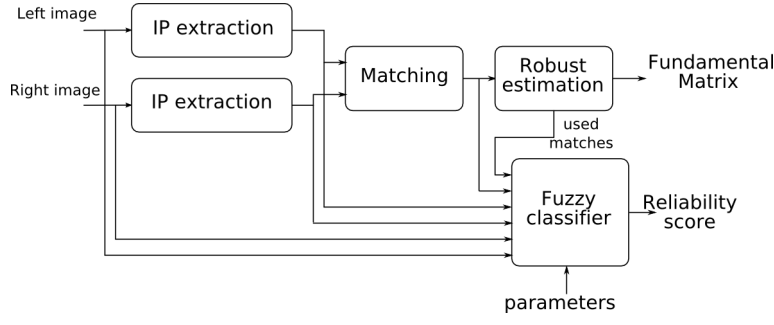


Fig. 2. Fundamental matrix estimation process and reliability score computation. Parameters are obtained through training.

- Image mean gray level :
 $M_0 = 0.5(M_L + M_R)$
- Relative difference of images gray level :
 $M_d = |M_L - M_R|/M_0$
- Mean number of IP in both images per used image surface (ROI) :
 $N_0 = 0.5(N_L + N_R)/S$
- Relative difference of IP in both images
 $N_d = |N_L - N_R|/N_0$
- Ratio of matches per used image surface (ROI)
 $N_{M1} = N_M/S$
- Ratio of matches per mean number of IP
 $N_{M2} = N_M/N_0$
- Mean surface ratio of the polygon of used matches (see II-C) ($R_S = S_p/S$)

For each of these criterions, we use a trapezoidal membership function, shown on fig. 6. The thresholds (x_1, x_2, x_3, x_4) are determined experimentally, using data plots of the distance between estimated and reference Fundamental matrix, related to the considered criterion (see below, section IV).

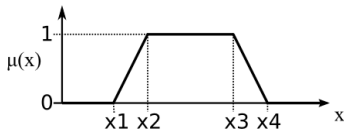


Fig. 6. The fuzzy membership function

D. Computing global reliability score

The inference system used here is quite simple, and can be summarized as : "if all the criterions are good, then the reliability is good". However, the counterpart is that in some cases, one of the criterions will show a "bad" value, although global estimation may be correct. So we cannot use the "AND" fuzzy-operator (min). Instead, we compute a simple weighted sum of all the fuzzy values in order to aggregate them.

IV. EXPERIMENTALS VALIDATION

We present here the experimental setup used to achieve the training of the classifier, using a set of arbitrary stereoscopic

image pairs. For each pair of images, we define a reference Fundamental matrix. We artificially degrade the each pair, process them through the estimation computing process, and log the values of the criterions described in III-C. Validation is done afterwards by processing this set of image pairs through the classifier, and by comparing the computed score with the distance between estimated and reference Fundamental matrix.

A. Used image pairs

The considered data sets are typical road scenes that can be seen on fig. 7. Although the matching algorithm is able to compute a correct estimation for all of them, some of these are more fragile, in the sense that a small disturbance will induce great changes in the final estimation. For example, pair 2 has lots of trees and few distinctive items. We will see on the following figures that it will be the first pair to have its distance diverge.

B. Estimating Fundamental matrix under image modification

We have considered two cases, when only one of the two images is affected by the disturbance, and when the two images are affected in a symetrical way. As expected, it appears that the most critical case is when only one image is affected, so we only present results for this case. On fig. 8, we show the measured distance between estimated and reference Fundamental matrix for the three considered disturbances. The 5 different curves correspond to the 5 different image pairs. It appears clearly that the results are greatly depending on image characteristics.

It must be noted that matching using the SIFT descriptor is highly robust under brightness change, as it can be seen on fig. 8(c).

C. Performance of estimation related to selected criterions

During the training phase, we log the values of the different criterions described in III-C. One can see on fig. 9 plots of this data, related to the distance d . These plots are used to set the thresholds of the fuzzy membership function for each criterion. These are chosen so most of the points giving a correct distance are "inside", and most of the points giving a bad distance are "outside". Besides this, the support for the fuzzy classifier must take care to include points of all considered disturbance. In some cases, the considered

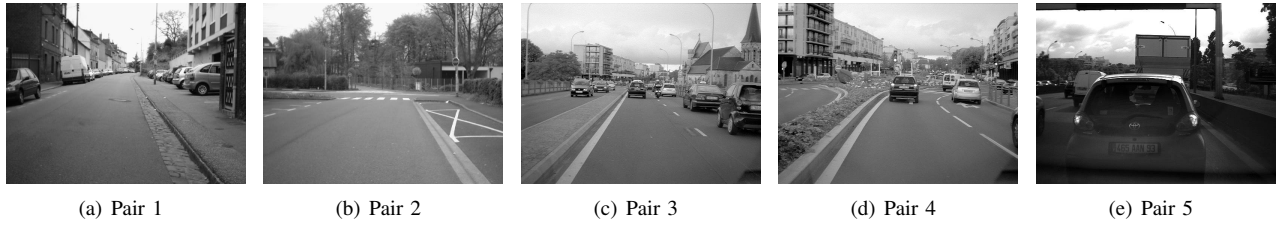


Fig. 7. Image data set used for experiments (only left image is shown)

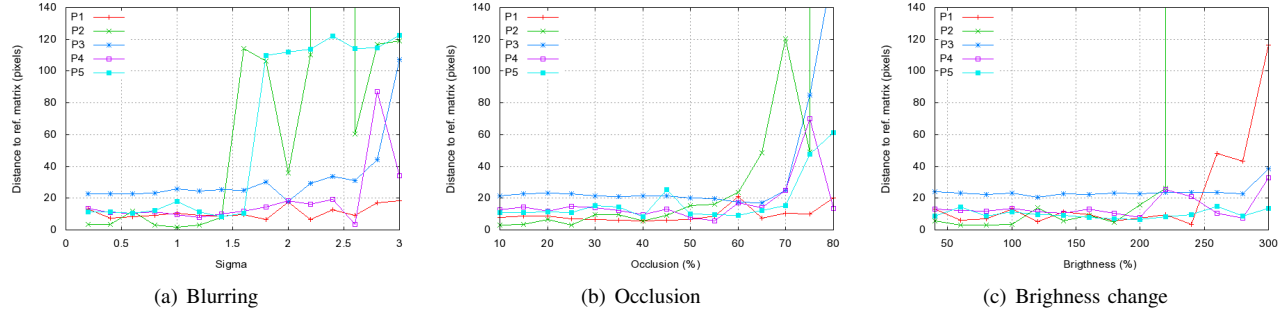


Fig. 8. Distance between estimated and reference Fundamental matrix

disturbance does not change at all the distance, for example blurring does not change at all mean image level, as can be seen on plots 9(a) and 9(b). The fuzzy membership functions with the choosen thresholds are shown in grey on these plots.

The proposed thresholds are summarised in table I.

D. Global validation

In order to validate the method, we process all the data collected during the training step through this classifier, and compare the computed score to the distance between estimated and reference Fundamental matrix. The results are shown fig. 10 : with this small database, one can see that all points with a reliability score higher than 0.6 will guarantee a correct estimation. No particular study has been done on the relative weights, at present the different criterions have all the same. But extended experimenting will probably show some criterions are more relevant than others, thus they could be assigned a heavier weight.

V. CONCLUSIONS AND FUTURE WORKS

We have presented here a general method for computing a reliability score associated with a given estimation of the Fundamental matrix. The method is simple and can be easily implemented. It is based on a fuzzy-logic classifier, using the data computed during the estimation step. It needs a set of parameters, that must be obtained through a training step. This training can be done by comparing estimation with a ground truth, as presented here. A training through a larger database can also be considered, using the first estimation as ground truth. With standard images, and taken into account the high performance of the SIFT detector, this assumption will be correct most of the time.

On the small database used here, results show that the computed score is a good sign of the reliability of the estimation of the Fundamental matrix.

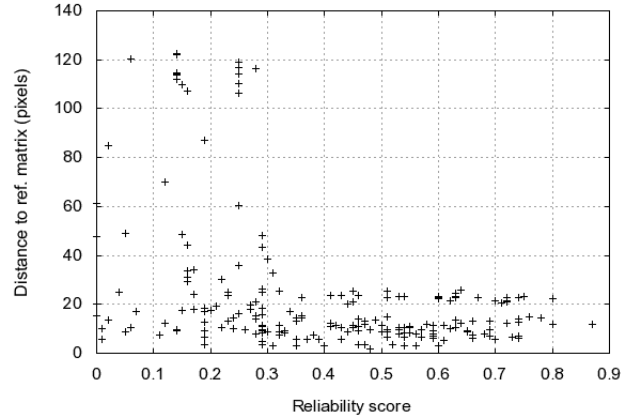


Fig. 10. Distance between estimated and reference Fundamental matrix related to the computed reliability score, using the training data

Others criterions could have been taken into account. For example, image standard deviation would have been an interesting criterion: a really low value will be highly suspect. Future works will focus on validating the method on a large database set, and studying more sophisticated inference system, in order to give a more subtle information.

VI. ACKNOWLEDGMENTS

The authors acknowledge Valeo for their image data, and the reviewers for their comments.

REFERENCES

- [1] G. Csurka, C. Zeller, Z. Zhang, and O. D. Faugeras. Characterizing the uncertainty of the fundamental matrix. *Computer Vision and Image Understanding*, 68(1):18–35, 1997.
- [2] R. Hartley. In defense of the eight-point algorithm. *IEEE Pattern Analysis and Machine Intelligence*, 19(6):580–593, 1997.

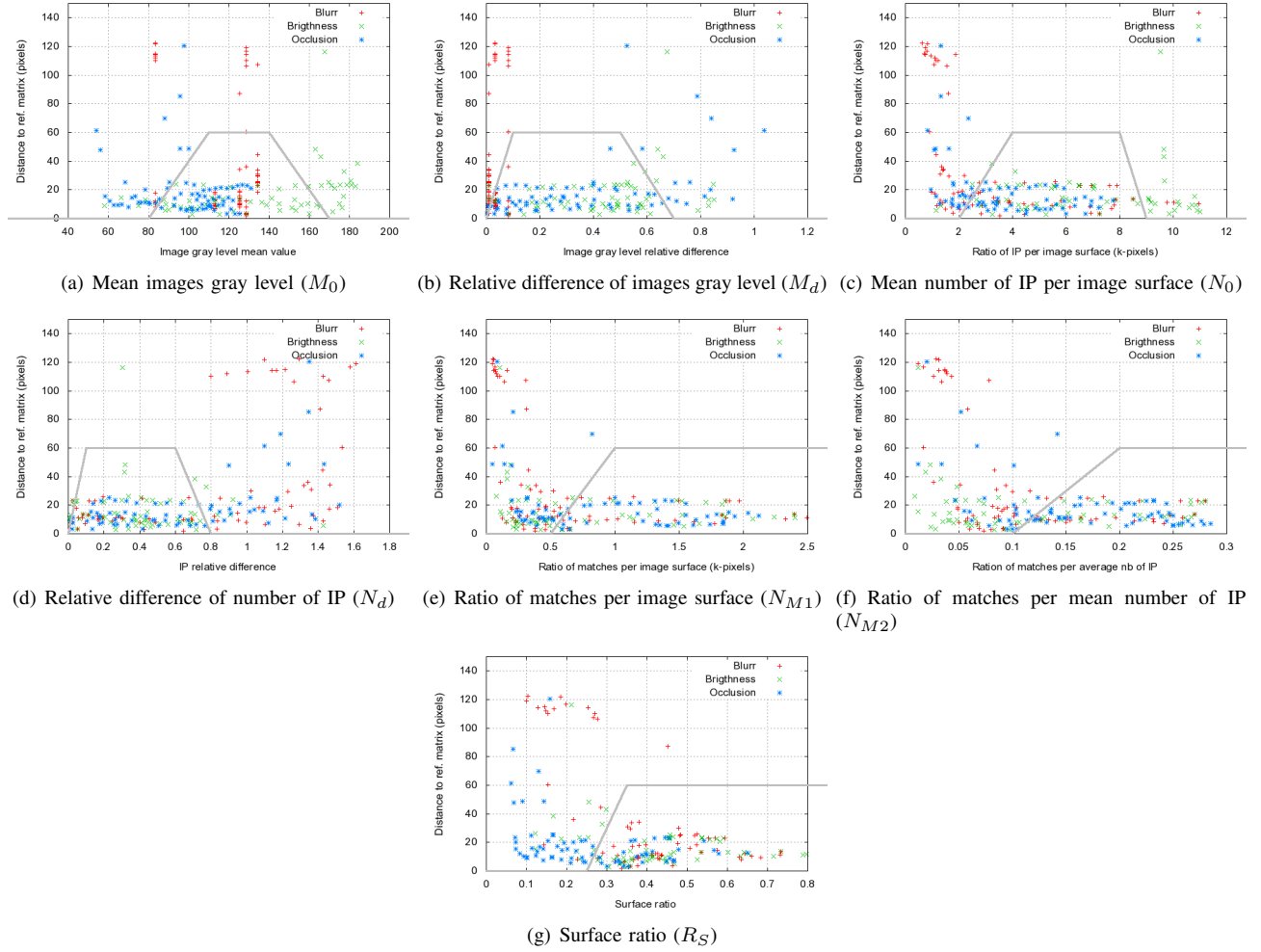


Fig. 9. Distance between fundamental matrix related to measured criterions

Criterion	Symbol	x_1	x_2	x_3	x_4
Image gray level mean value	M_0	80	110	140	170
Relative difference of gray level	M_d	0	0.1	0.5	0.7
Mean number of IP per image surface (k-pixels)	N_0	2	4	8	9
IP relative difference	N_d	0	0.1	.6	.8
Ratio of matches per image surface (k-pixels)	N_{M1}	.5	1	3	4
Ratio of matches per average nb of IP	N_{M2}	.1	.2	.8	.9
Surface ratio	R_S	.25	.35	.8	.9

TABLE I

EXAMPLE OF THRESHOLD VALUES FOR FUZZY MEMBERSHIP FUNCTIONS

- [3] K. Kenichi. Optimal fundamental matrix computation: Algorithm and reliability analysis. In *proceedings of the 6th Symposium on Sensing via Image Information*, pages 291–298, juin 2000.
- [4] Y. G. Leclerc, Q. T. Luong, and P. Fua. Measuring the self-consistency of stereo algorithms. In *proceedings of European Conference on Computer Vision (ECCV'00)*, volume 1, pages 282–298, 2000.
- [5] Y.G. Leclerc, Q.T. Luong, and P. Fua. Self-consistency: A novel approach to characterizing the accuracy and reliability of point-correspondence algorithms. In *ICCV Workshop on Vision Algorithms: Theory and Practice, Corfu, Greece*, 1999.
- [6] C.T. Loop and Z. Zhang. Computing rectifying homographies for stereo vision. In *proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR'99)*, Fort Collins, CO, USA, pages 1125–1131, 1999.
- [7] D.G. Lowe. Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*, 60(2):91–110, 2004.
- [8] J. Mallon and P. F. Whelan. Projective rectification from the fundamental matrix. *Image and Vision Computing*, 23(7):643–650, 2005.
- [9] K. Mikolajczyk, T. Tuytelaars, C. Schmid, A. Zisserman, J. Matas, F. Schaffalitzky, T. Kadir, and L. Van Gool. Mikolajczyka comparison of affine region detectors. *International Journal of Computer Vision*, 65(1-2):43–72, 2005.
- [10] Z. Zhang. Determining the epipolar geometry and its uncertainty : A review. *International Journal of Computer Vision*, 27(2):161–195, 1998.