Vision, Video, and Graphics (2003), pp. 1–8
P. Hall, P. Willis (Editors)

Iterative Multi-planar Camera Calibration: Improving Stability using Model Selection

Abstract

Tracking, or camera pose determination, is the main technical challenge in numerous applications in computer vision and especially in Augmented Reality. However, pose computation processes commonly exhibit some fluctuations and lack of precision in the estimation of the parameters. This leads to unpleasant visual impressions when augmented scenes are considered. In this paper, we propose an efficient and reliable method for real time camera tracking which avoid unpleasant statistical fluctuations. This method is based on the knowledge of a piecewise planar structure in the scene and makes use of model selection to reduce fluctuations. Videos are attached to this paper which proved the effectiveness of our approach (hallTrack.mpg, hallCamera.mpg, hallAugmented.mpg, roomWithoutMS.mpg and roomWithMS.mpg).

1. Introduction

Augmenting real video sequences of a scene with computer generated objects is one of the main goals of many applications such as virtual museums, interactive interior design or architectural design, computer-aided repair and learning systems [15]. All these interactive applications require that the augmented scene is continually updated as the user moves about the real scene.Hence, one of the most basic challenge to overcome is the registration problem: the objects in the real and the virtual world must be properly aligned with respect to each other or the illusion that the two worlds coexist will be compromised.

In this paper, we address the registration problem for interactive AR applications. Such applications require sequential and real-time registration process. Though the registration problem has received a lot of attention in the computer vision community, the problem of real time registration is still far from a solved problem. Ideally, an AR system should work in all environments without the need to prepare the scene ahead of time and the user should walk anywhere he pleases. In the past, several AR systems have achieved accurate and fast tracking and registration, putting dots over objects and tracking the dots with a camera [7, 8]. However, such methods restrict the flexibility of the system. Hence, there is a need to investigate registration methods which work in unprepared environments and which reduce the need to know the geometry of the objects in the scene.

1.1. State of the art

Today, the approaches to sequential viewpoint computation can be divided in two main categories: model-based approach or move-matching approach. Model-based techniques rely on the identification in the images of features from the object model. Hence, a direct correspondence between the 3D object-coordinate system and each image is set up [8, 10]. This capability of treating each image independently makes such methods more appropriate for real time implementations. Another consequence of model-based tracking is the absence of drift. However, it is commonly true that few features are available for registration. Moreover, noise in the image measurements hampers their accurate detection and consequently corrupts the estimated pose. As a result, the tracking suffers from high-frequency jitter. More importantly, such methods require significant manual intervention to construct the model.

On the other hand, new move matching methods [11] attempt to compute the relative motion between two successive frames using planar structures. If the position of the camera in the first frame is known, the absolute position of the camera is obtained by compositing each relative motion. These systems are attractive because they do not require any knowledge on the scene. However, they can suffer from drift because errors accumulate over time.

In interactive real time applications, a good way to assess the viewpoint accuracy is to consider the visual impression of the augmented scene. Today, it appears that statistical fluctuations in the viewpoint computations lead to unpleasant jittering effects or to sliding effects in the scene. This problems are particularly conspicuous when the motion of the camera is small because noise of the extracted features lead to large fluctuations in the viewpoint computation. The problem of stabilization was addressed in [4]. The idea is to classify the typical movements of the camera into models (stationary, panoramic, general,zoom in) in order to fix some of the parameters assuming that the variations at these parameters are due to statistical fluctuations. Of course, stability and accuracy over the remaining parameters are better because the degree of freedom of the function to be optimized is smaller. In this paper, we investigate further this idea with the following contributions: (i) we propose a method for viewpoint computation which is based on the observation of a multiplanar structure in the scene. Such structures are quite common both for indoor or outdoor scenes (ii) Following Kanatani [4], we investigate the use of model selection to improve the stability of the computed viewpoint. Various model selection criteria are considered and tested in this study. We prove that the ones which make use of the covariance on the estimated parameters give better results than the classical criteria (iii) the effectiveness of our pose algorithm is assessed on various sequences.

The method for multiplanar viewpoint computation is given in section 2. Section 3 exhibit results and strategies for model selections. Finally, various snapshots of augmented scenes are provided.

2. Multiplanar viewpoint computation

2.1. Overview

This section gives an overview of our registration method. The equations of the planes used by the registration process are given by the user. In our approach, the intrinsic parameters are supposed constant and are computed beforehand. The first camera pose is also estimated. Often, this estimation is obtained by using a poster in the scene.

Once the preprocessing stage has been achieved, the registration follows a four step loop: key-points are extracted and matched from frame to frame. Then, for each model, the projection matrix is computed using constraints induced by the homographies. Finally the right motion model is selected and the viewpoint is computed accordingly. In the following, the main steps of this algorithm are described with further details (Fig. 1).

Initialization stage:
1. Give the equation of the observed planes used for registration ,
2. Compute the projective matrix for the first frame P^0,

Computation of the projective matrix P^i for $i > 0$:
1. Compute the set of matched key-points between images $i-1$ and i for each observed plane.
2. For each motion model, compute P_i from P_{i-1} using the constraints induced by the homographies
3. Select the best model according to the selection criterion which is a tradeoff between accuracy and simplicity of the model
4. Compute the motion using the selected model

Figure 1: *Overview of the multi-planar tracking method*

2.2. Planar viewpoint computation

We assume that the position, orientation, and the internal parameters of the camera are known for the first image. Then all the images of the sequence may be related to the preceding one by setting up correspondences between points.

We know that given two projection matrices $P_1 = [I|0]$ and $P_2 = [A|a]$ and a plane defined by the vector v such that $v^T X + 1 = 0$, the corresponding homography matrix [5] can be expressed as:

$$H = K_2(A - av^T)K_1^{-1} \qquad (1)$$

where $K_i = \begin{pmatrix} k_u & k_u & u_0 \\ 0 & k_v & v_0 \\ 0 & 0 & 1 \end{pmatrix}$ is the matrix of intrinsic parameters for the image i.

Given the intrinsic parameters and a set of matched points (x_j, x'_j) on the considered plane between two images, a classical procedure to get the viewpoint parameters is to minimize the mean error of the matched points with respect to the transformation [4], thus:

$$A, a = argMin(J(\mathbf{A}, \mathbf{a}) = \frac{1}{N}\sum_{j=1}^{N} ||x'_j - Z(Hx_j)||^2)$$

where $Z[]$ denotes normalization to make the third component 1.

2.3. Multi-planar calibration

Our experiments proved that the accuracy of the single plane registration method is not sufficient to obtain a good visual impression of the augmented scene. Indeed the accuracy depends on the relative position of the camera and of the observed plane and also on the number of matched points. Moreover, as sequential viewpoint computation is considered, errors on viewpoint accumulate over time and the viewpoint parameters tend to diverge from the real ones especially when large sequences are considered.

That is the reason why we suggest to use several planes because it brings more information about the tridimensional space and reduces considerably the variability of the estimated calibration parameters. It will also help us to handle large environments for AR applications.

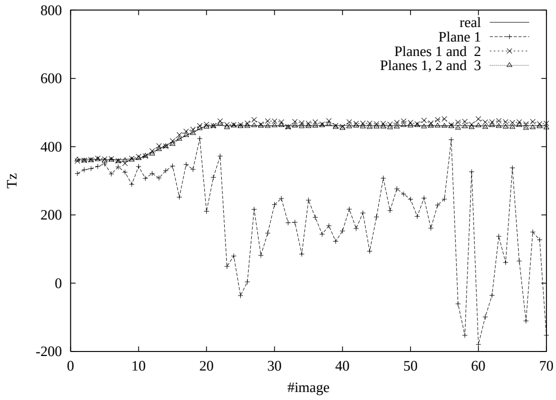


Figure 2: *The computed Z-translation through the sequence using one, two or three planes.*

When several planes are considered, the function to be optimised is:

$$J(\mathbf{A},\mathbf{a}) = \frac{1}{N_1+\cdots+N_n}\sum_{k=1}^{n}\sum_{j=1}^{N_k}||x'_{kj} - Z(H_k x_{kj})||^2$$

where n is the number of planes, N_k the number of points belonging to the plane k, H_k the corresponding homography.

A typical method used to minimize this non-linear function is Newton iterations but it is very sensitive to the initial estimation. Thus we use the Levenberg-Marquardt method that is more stable and almost as fast as the Newton method.

2.4. Results

To prove the effectiveness of the approach, we considered a image sequence using the model of our three-plane calibration target. Various motions were considered: x and y translation, panoramic motion, stationary motion. Various noise were added to the image points.

In Figure 2, we compare the actual translation coordinates T_z (Figure 2), and the computed coordinates when a single plane, 2 planes, and 3 planes are used. In these images, gaussian noise with covariance matrix $\sigma^2\mathbf{I}$, with $\sigma = 0.5$, was added to the image points. The viewpoint is found computing the 6 extrinsic parameters and fixing all the intrinsics to a pre-calibrated value. These results show that using a single plane, the estimated viewpoint is very unstable and the estimated coordinates are lacking of precision. By adding a second and a third plane both, precision and regularity, are improved considerably.

2.5. Improving the robustness of the viewpoint computation

It is well known that false matches $x_j \leftrightarrow x'_j$ can severaly disturb the viewpoint computation process. To cope with this problem, RANSAC algorithm is classically used for each visible plane in order to discard false matchings on each plane. However such an approach considers the planes independently and does not take into account the multi-planar model of the scene. We propose here two methods to cope with this problem:

Method 1 We first use an iterative method to refine the set of inliers. As the multi-planar model of the scene is available, the homography induced by each plane can be deduced from the projection matrix computed with the multi-planar algorithm using (1). This allows us to compute the set of inliers compatible with the computed homographies. The projection matrix is then computed from this new set of inliers, which is in better agreement with the scene geometry, and the process is iterated until convergence. This way, false matchings are removed and new ones can be added.

However, this approach may fail if a small number of points is available on a given plane. In this case, the RANSAC algorithm is not always able to select the right points and the first multi-planar estimate of the viewpoint is erroneous and the iterated process too. To cope with this problem, we suggest an approach which fully integrates the muti-planar model in the robust estimation process:

Method 2

1. Four point correspondences are randomly chosen in the full set of matched points (the union of the matched points for all visible planes).
2. the viewpoint is computed these four correspondences using the multi-planar method
3. The homography induced by each plane is computed from the projection matrix and from the known plane equations using (1).
4. A new set of inliers is computed for each plane. This is the union of the correspondences which are in agreement with the computed homograghy in each plane.
5. repeat 1 to 4 L times (The number L of samples is choosen according to the law $L = log(1-p)/log(1-(1-\varepsilon)^4)$ were $p = .99$ and the proportion of outliers is $\varepsilon = .3$ in our experiments.)
6. Select the parameters which correspond to the biggest set of inliers.

To prove the efficiency of these two methods, we consider a **turntable sequence**. This sequence was considered to assess the accuracy of the viewpoint algorithm. The camera was fixed to the turnable and we consider a closed sequence which is described in Fig. 3.a and b. Fig. 3.c exhibits the computed translation along the Z axis when the three described methods are used. As the sequence is closed, a good way to asses the accuracy is to check if the final position is the same as the initial one. Fig. 3.c clearly shows that using method 1 and 2, the final position is very close to the initial one. For method 2, the difference between the initial and the final position is very small. To consider the influence of the method on the visual impression of the augmented scene, a cube is added to the scene and is shown in the first image of the sequence in Fig. 4.a. The three other images show the augmented scene in the final position, which is the same as the initial one, when the three matching methods are uses: classical, method 1 and method 2 (Fig. 4.b,c and d). These snapshots proved that sliding effects occurred when the classical matching method is used. The use of methods 1 and 2 clearly improved the viewpoint accuracy with noticeable better results for method 2. However, as the computational cost of method 2 is very high, we prefer to use the method 1 which is a good compromise between computational time and accuracy.

3. Use of stabilisation methods

3.1. Aims

Even when the precision of the viewpoints is improved by considering several planes, fluctuations in the parameters are often observed and may lead to unpleasant visual impressions such as jittering or sliding when augmented scenes are considered. These fluctuations are especially conspicuous when the camera motion is small because of noise and imprecision in computing the points coordinates. In the past, several papers used Kalman filtering for prediction and stabilization task. However, the use of a Kalman filter is not always advantageous for AR. This is because a low order dynamical model of human motion may not be always appropriate except under very constraint scenarios.

Following Matsunaga and Kanatani[4] and [13] we investigate the use of motion model selection to reduce fluctuations of the camera parameters and to improve the visual impression of the augmented scene. The underlying idea in model selection is as follows: a higher order motion model fits any data set more accurately than a lower order model. However, high order models fit part of the random noise they are supposed to remove. Thus, a high order model, although accurate, is less stable to random perturbations in the data. A good motion model must strike the right balance between accuracy and stability. The model selection principle demands that the model should explain the data very well and at the same time have a simple structure.

3.2. State of the art

Many model selection criteria for balancing the residual and the degree of freedom of the model have been proposed in the literature [3]. All of them are the sum of an accuracy criterion and of a term which is a measure of the complexity of the model. Most of them are based on statistical and information-theoretic criteria. Among them, the most widely used criterion are the geometric Akaike's criterion and the minimum description length (MDL) criterion. The AIC criterion can be viewed as an approximation of an entropy criterion (the Kullbak-Leibler distance) whereas the MDL criterion try to choose the model that minimizes the number of bits required to express the model:

$$G_{AIC} = \hat{J} + 2k\varepsilon^2$$
$$G_{MDL} = \hat{J} - k\varepsilon^2 log\varepsilon^2$$

where k is the degree of freedom of the motion. The square noise level ε^2 can be estimated from the residuals $\hat{J}$ (the one corresponding to the highest order model [42]). Whatever the considered criterion, the use of a too complex model is penalised with respect to simpler model.

Kanatani [4] previously applied this technique to the calibration problem by using a single plane, specifically, a calibration pattern. In this seminal work, Kanatani classifies the movements in six types, specifically those with fixed focal length are four:

Movement	Known parameters	Variables
stationary	$\mathbf{A}_i = \mathbf{A}_{i-1}, \mathbf{a}_i = \mathbf{a}_{i-1}$	$-$
panoramic	$\mathbf{a}_i = \mathbf{a}_{i-1}$	$\mathbf{A}_i$
$t-$predicted	$t_i = 2t_{i-1} - t_{i-2}, \mathbf{a}_i = \mathbf{a}_{i-1}$	$\mathbf{A}_i$
general model	$-$	$\mathbf{A}_i, \mathbf{a}_i$

In the $t-$predicted model, the camera position is linearly extrapolated as $t_i = 2t_{i-1} - t_{i-2}$. The optimisation is only performed on the rotation.

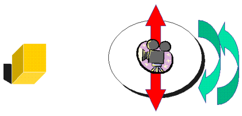
In Kanatani's approach, only two criteria are considered G_{AIC} and G_{MDL} using one single plane. However, there are many other criteria, especially those which make use of the covariance matrix or the information matrix on the estimated parameters.

3.3. Our approach to model selection

We suggest to use together the model selection strategy and the multi-planar calibration in order to improve the stability and the accuracy of the estimated the parameters. There are different branches using model selection, but there is no such successful criterion in general, as can be seen in some reviews comparing some of them for different problems: finding the polynom's degree [2], surface merging [3], type of motion [12], detection of geometric primitives [6].

For this reason, we compare different model selection criteria (table 1). These criteria use the same accuracy measure: the residual error evaluated at the maximum likelihood parameters. But, the complexity term is different for each model and depends on different assumptions over the parameters and their distribution. In this work, we especially considered criteria which involve the covariance matrix on the estimated parameters $(V(\theta_k))$ and the Fisher information matrix $(I(\theta_k) = V(\theta_k)^{-1})$. Indeed, often, criteria such AIC are only asymptotic approximations of a criterion which includes the covariance or the information matrix. So, we hope that such criteria will improve the model selection.

Frame	*Movement*
$0-5$	*Stationary*
$5-25$	$Rotation + 10\,\text{deg}$
$25-65$	$Rotation - 20\,\text{deg}$
$65-75$	$Translation 10cm.(left)$
$75-85$	*Stationary*
$85-115$	$Rotation + 15\,\text{deg}$
$115-125$	$Translation 10cm.(right)$
$125-135$	$Rotation - 5\,\text{deg}$
$135-140$	*Stationary*



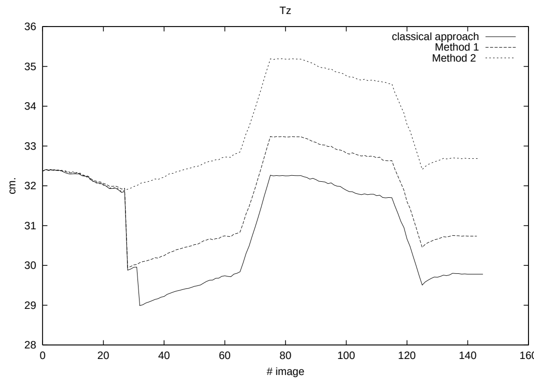


Figure 3: *Turnable sequence: (a) and (b): actual motion of the camera, (c) comparison of the three robust matching methods on the turntable sequence.*

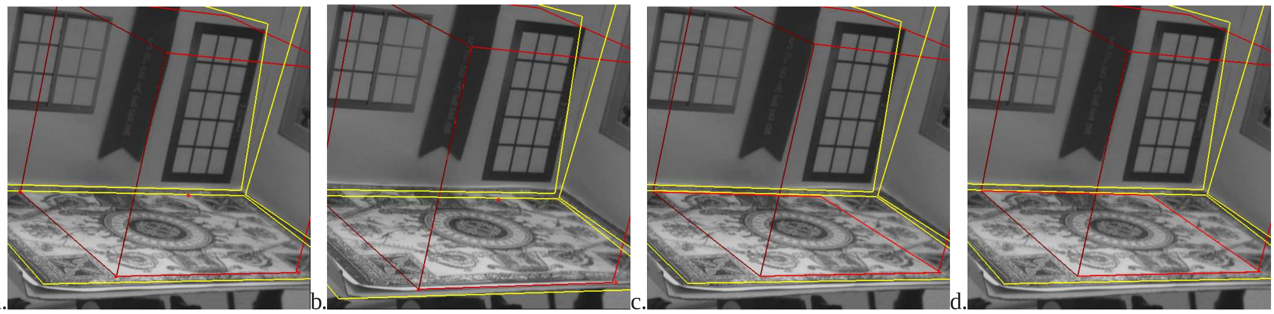
a. b. c. d.


Figure 4: *Comparison of the final position when the three matching methods are used: (1): initial position (b,c,d): final position when the classical method, method 1 and method 2 are used.*

Criterion	**Complexity term**
Akaike's AIC [1]	$2k$
Bozdogan's CAIC [2]	$k(\log n + 1)$
Bozdogan's CAICF [2]	$k(\log n + 2) + \log\lvert \mathbf{I}(\theta_k)\rvert$
Schwarz's BIC [9]	$2k \log n$
MDL	1/2 klog(n)
Kanatani's gMDL [4]	$-k \log \varepsilon^2$
BMSC-RISS [3]	$k/2log_2 * (\theta_k^t \mathbf{I}(\theta_k)\theta_k) + log_2(V_k)$

Table 1: *The criteria considered in our study (k is the size of the model and n is the number of data.*

3.4. Experimental results

Experiments were conducted both on real and synthetic images. In our experiments, we assign an integer to each type of movement : 0 for stationary; 1 for panoramic; 2 for translation and rotation (general); and 3 for predicted translation and variable rotation.

3.4.1. Predicted model

The possibility of using linearly predicted models was described in [4]. The main problem of sequential calibration is that some variations are very small, and some subsequences may seem piecewise linear. The consequence is that for some noise level, the predicted model is often preferred to the general one because its complexity is simpler than the real model. This may lead to divergence after some images as exhibited in figure 5. In this case, we use the CAICF criterion, but the behaviour is similar for the other ones. In our approach, it is important both to select always the right model and to compute the parameters accurately. This is the reason why we decide to use just three models for the fixed focal length calibration.

3.4.2. Comparing selection criteria

In order to compare the selection criteria, we use a synthetic sequence corrupted with various noise levels. For each image i we consider the model selected between frame $i-1$ and i. As an iterative procedure is used, we use as initialisation the actual viewpoint for frame $i-1$ in order to avoid drift problem. As the actual model is known, we show in Table

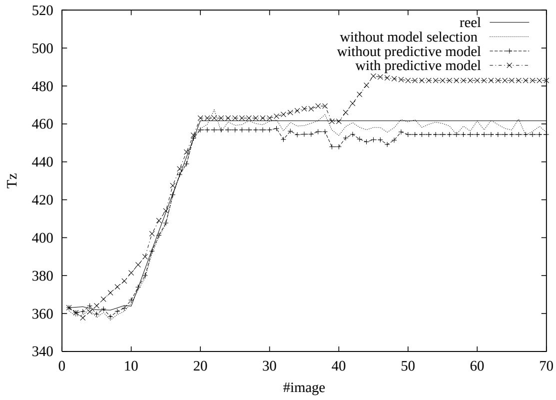


Figure 5: *The computed Z-translation using model selection with and without $t-$predicted model*

Motion	Criterion	$\sigma = 0.3$		
		Underfit	**correct**	Overfit
Static	AIC	-	83.1%	16.9%
	CAIC	-	98.7%	1.3%
	CAICF	-	100.0%	0.0%
	BIC	-	100.0%	0.0%
	gMDL	-	77.5%	22.5%
	MDL	-	83.2%	16.8%
	BMSC-RISS	-	86.3%	13.7%
Pan	AIC	0.0%	85.3%	14.7%
	CAIC	0.0%	99.3%	0.7%
	CAICF	0.0%	98.7%	1.3%
	BIC	0.0%	100.0%	0.0%
	gMDL	0.0%	84.7%	15.3%
	MDL	0.0%	85.4%	14.6%
	BMSC-RISS	0.0%	100.0%	0.0%
General	AIC	0.0%	100.0%	-
	CAIC	1.5%	98.5%	-
	CAICF	1.3%	98.7%	-
	BIC	5.4%	94.6%	-
	gMDL	0.0%	100.0%	-
	MDL	0.0%	100.0%	-
	BMSC-RISS	5.8%	94.2%	-

Table 2: *Percentage of good model selection for various criteria, noise level = 0.3 .*

Motion	Criterion	$\sigma = 1.0$		
		Underfit	**correct**	Overfit
Static	AIC	-	83.7%	16.3%
	CAIC	-	100.0%	0.0%
	CAICF	-	100.0%	0.0%
	BIC	-	100.0%	0.0%
	gMDL	-	0.0%	100.0%
	MDL	-	83.8%	16.2%
	BMSC-RISS	-	100.0%	0.0%
Pan	AIC	0.0%	86.7%	13.3%
	CAIC	0.0%	100.0%	0.0%
	CAICF	0.0%	97.3%	2.7%
	BIC	0.0%	100.0%	0.0%
	gMDL	0.0%	0.0%	100.0%
	MDL	0.0%	88.0%	12.0%
	BMSC-RISS	0.0%	100.0%	0.0%
General	AIC	11.5%	88.5%	-
	CAIC	24.1%	75.9%	-
	CAICF	20.3%	79.7%	-
	BIC	33.6%	66.4%	-
	gMDL	0.0%	100.0%	-
	MDL	11.5%	88.5%	-
	BMSC-RISS	36.6%	63.4%	-

Table 3: *Percentage of good model selection for various criteria, noise level= 1.*

2 and 3 the percentage of correct model choice obtained for each criterion and for ech model.

These tables proved that for a moderate noise level ($\sigma = 0.3$), most of the criteria perform well. We see that some criteria prefer the more general model (AIC, gMDL) while some others always chose the simpler one (BIC). It can also be noticed that some criteria are more sensitive to noise and select a wrong model often than others. In general, the criteria based on Akaike's Information Criterion (AIC, CAIC and CAICF) performs well. However, AIC tend to admit more overfitting than the CAIC or CAIF and the stabilisation performance is then reduced. The experiments we performed on Bayesian criteria such as BMSC-RISS are not convincing. First, this criterion tend to admit underfitting for the general motion when the noise level is relatively high. Second, the results seem to depend tightly on the a priori probability on the various motion models.

That is the reason why the experiments in the following are done using the CAICF as selection criterion, because it performs well and it considers also that the nature of the parameters is different by including the Fisher's information matrix in the complexity term.

Figure 6 shows the performance of the criteria in the sequence compared with the real motion model. The X axis indicates the number of the image, and the Y axis the model of movement.

If the noise level increases, some criteria tend to select a simpler model, specially the panoramic model even when the real one varies in translation and in rotation. Often, small

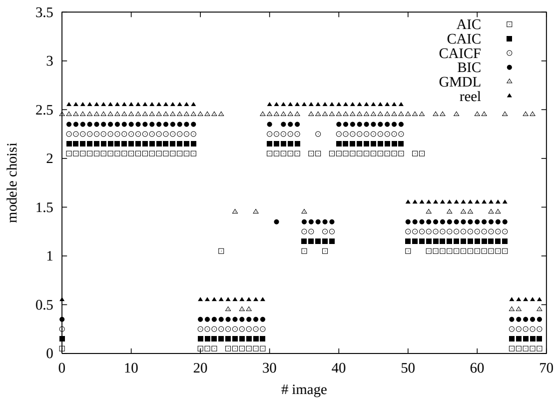


Figure 6: *Comparison of model selection criteria on a synthetic sequence,* $\sigma = 1.$.

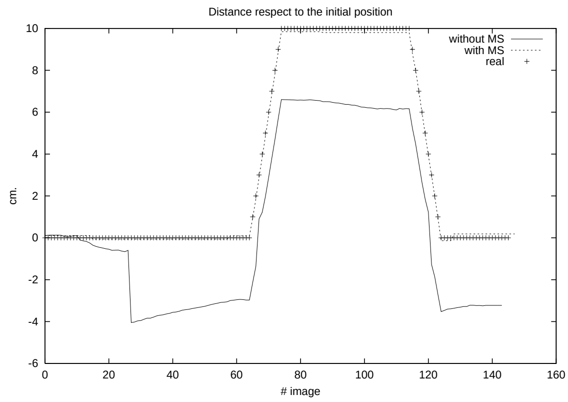


Figure 7: *The distance between the current viewpoint and the initial one for the turntable sequence*

translations are mistaken by panoramic motions because the direction of motion of the points is the same and the residual error is also similar, but the complexity of the model to detect a translation is bigger (because it is just included in the general model) than to detect a panoramic movement, the first model has 6 degrees of freedom, rather than 3 for the second one.

3.4.3. The Turntable Sequence

We demonstrate the effectiveness of the approach on the turntable sequence, which was described in section 2. Fig. 7 shows the distance from the current camera position to the initial camera position computed with and without model selection. We can notice that when model selection is used, the trajectory is more stable. As the total translation of the turntable is perfectly known (10cm), we can also estimate the accuracy of the process by comparing the estimated translation with and without model selection. When model selection is used the estimated total translation is 9.82 cm, whereas it is around 6.56cm without model selection. In addition, as the sequence is closed, the drift can be used to assess the two methods. The total drift of the camera position when model selection is used is 0.14cm. Without model selection, the drift is 3.26 cm. During the stationnary and the rotating motion, the distance between the current and the initial position is constant. When model selection is used, this distance is really constant, whereas it is not without model selection. Two videos using the same room with more abrupt motions are attached to this paper. The sequence roomWithoutMS.mpg exhibits the augmented scene and the selected model when no motion selection is used The video roomWithMS.mpg exhibits results when motion selection is used. In these videos, the symbol in the upper-left corner of the images indicates the selected model. The red cross indicates the stationary model, the green circle corresponds to the panoramic rotation, and the blue square to the general model.

These results clearly demonstrate that model selection produces smoother trajectory and better visual impression. They also prove that the use of model selection improve the accuracy of the viewpoints and reduces noticeably the drift problems that are common when long sequences are considered.

In order to quantify the time needed for viewpoint computation, table 4 gives the times needed for the different steps of viewpoint computation for one image of the calibration target: extraction and matching steps (we use the MIC algorithm[14] to extract the key points), robust matching using the RANSAC algorithm with method 1 and 2, and viewpoint computation using model selection . About 500 key-points were extracted from each image . After retaining only the points which belongs to three planes, only 100 points are considered in the viewpoint computation process.The full process is about 64 ms when method 1 is used and 124 ms for method 2. This means that we can handle 16 images per second with method 1 and 8 images with method 2.

MIC	15 ms
Matching	9 ms
RANSAC Method 1	25 ms
RANSAC Method 2	85 ms
Viewpoint estimation with model Selection	15 ms
Total Method 1	64 ms
Total Method 2	124 ms

Table 4: *Computation rates obtained on a Pentium IV 2. Ghz*

3.4.4. The snooker sequence

This large sequence was taken using a hand held camera in the hall of our laboratory. The user was free to move anywhere he wanted. Due to the brightness of the floor, some sheets of paper were put down on it to make easier the tracking process. During the sequence, two panoramic motions were realized (see hallCamera.mpg), one with a tripod and the other without a tripod. Both are correctly labeled by the model selection process as can be seen on the video (hallTrack.mpg). The set of inliers is also visible on this video. Finally, some snapshots of the scene augmented with a snooker are shown in Fig. 8 and in video hallAugmented.mpg and prove the effectiveness of our method.

4. Conclusion

We proposed in this paper several improvements to viewpoint computation for multi-planar environments. The use of model selection with various criterion proved that criterion involving information on the covariance of the estimated parameters are well suited to stabilization. Videos attached to this paper proved that this method significantly improved the visual impression of the augmented scene. We now investigate if these criteria are well suited when varying focal lens are considered. Our first experiments proved that the use of non nested models is more difficult to handle. We also plan to investigate the influence of the accuracy on the first viewpoint on the whole process. Indeed, it appears that good registration results can only be obtained if good intrinsic camera parameters are available.

References

1. H. Akaike. A new look at the statistical model identification. *IEEE Trans Aut Ctrl*, 19(6):716–723, 1974. 5
2. H. Bozdogan. Model Selection and Akaike's Information Criterion (AIC); The General Theory and its Analytical Extensions . *Psychometrika*, 52(3):345–370, 1987. 4, 5
3. K. Bubna et C.V. Stewart. Model selection and surface merging in reconstruction algorithms. In *Proceedings of 6th International Conference on Computer Vision, Bombay (India)*, pages 895–902, 1998. 4, 5
4. K. Kanatani C. Matsunaga. Calibration of a Moving Camera Using a Planar Pattern: Optimal Computation, Reliability Evaluation and Stabilization by Model Selection. In *Proceedings of 6th European Conference on Computer Vision, Trinity College Dublin (Ireland)*, pages 595–609, 2000. 2, 4, 5
5. R. I. Hartley et A. Zisserman. *Multiple View Geometry in Computer Vision*. Cambridge University Press, ISBN: 0521623049, 2000. 2
6. K. Kanatani. Model Selection for Geometric Inference. In *Proceedings of 5th Asian Conference on Computer Vision, Melbourne, Australia*, pages 23–25, 2002. 4
7. J. P. Mellor. Realtime Camera Calibration for Enhanced Reality Visualization. In *Proceedings of Computer Vision, Virtual Reality, and Robotics in Medicine'95 (CVRMed'95)*, pages 471–475, April 1995. 1
8. U. Neumann et Y. Cho. A selftracking augmented reality system. In *Proceedings of the ACM Symposium on Virtual Reality Software and Technology*, pages 109–115, 1996. 1
9. G. Schwarz. Estimating the Dimension of a Model. *The Annals of Statistics*, 6(2):461–464, 1978. 5
10. G. Simon et M.-O. Berger. A Two-stage Robust Statistical Method for Temporal Registration from Features of Various Type. In *ICCV'98, Bombay (India)*, pages 261–266, January 1998. 1
11. G. Simon, A. Fitzgibbon, and A. Zisserman. Markerless tracking using planar structures in the s cene. In *Proc. International Symposium on Augmented R eality*, pages 137–146, October 2000. 1
12. P.H.S. Torr. An Assessment of Information Criteria for Motion Model Selection. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, Puerto Rico, PR (USA)*, pages 47–52, June 1997. 4
13. P.H.S. Torr, A.W. Fitzgibbon, et A. Zisserman. Maintaining multiple motion model hypotheses over many views to recover matching and structure. In *ICCV'98, Bombay (India)*, pages 485–491, 1998. 4
14. M. Trajkovic et Mark Hedley. Fast corner detection. *Image and Vision Computing*, (16):75–87, 1998. 7
15. J. Vallino. *Interactive Augmented Reality*. Thèse de doctorat, University of Rochester, December 1998. 1

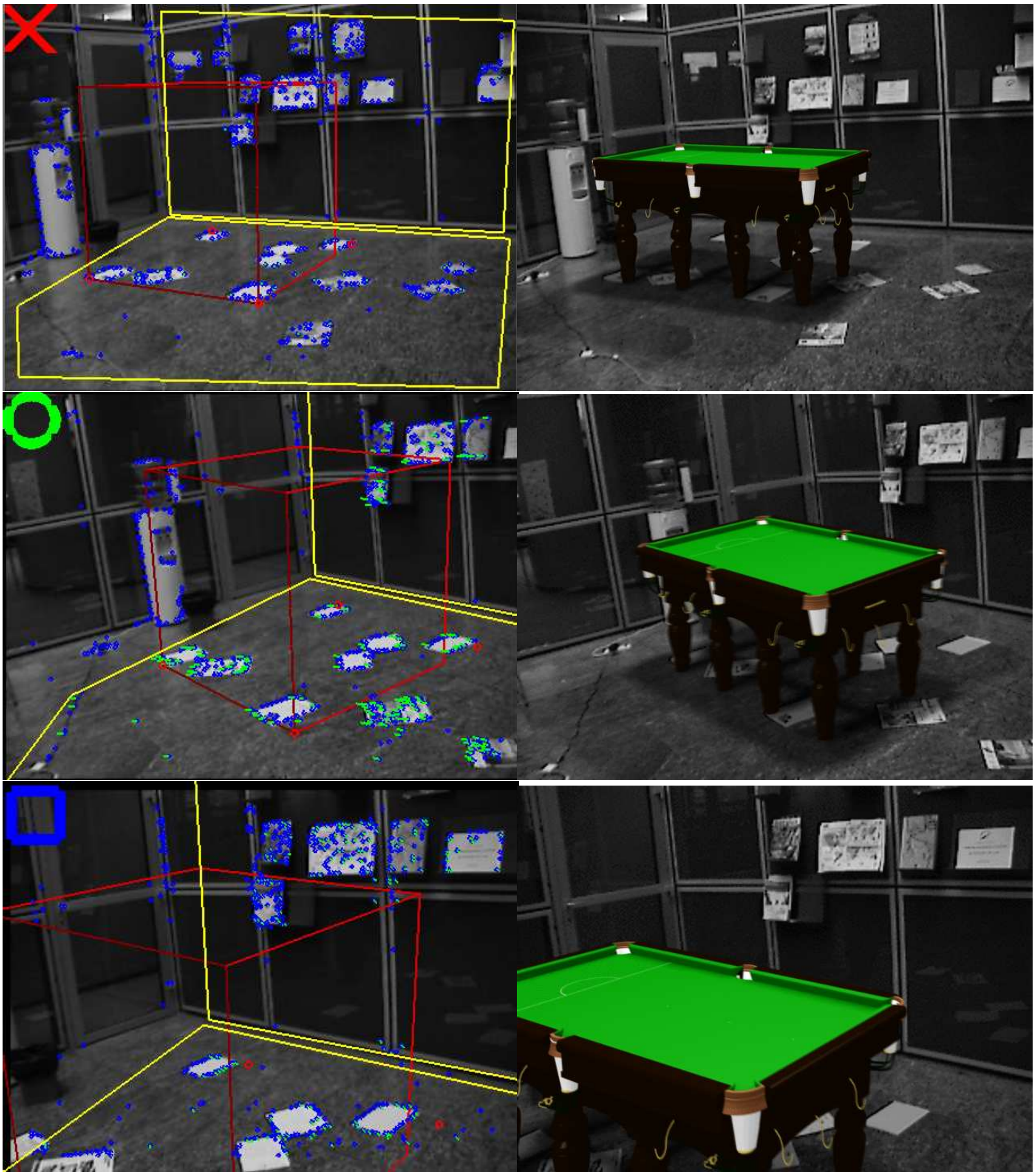


Figure 8: *Snapshots of the augmented scene for the snooker sequence: the set on inliers and the augmented scene.*