On-Line Rectification of Sport Sequences with Moving Cameras

Jean-Bernard Hayet[1,⋆] and Justus Piater[2]

[1] CIMAT, A.C., Jalisco S/N
36240 Guanajuato, GTO., Mexico
[2] Institut Montefiore, University of Liege
4000 Liege, Belgium

Abstract. This article proposes a global approach to the rectification of sport sequences, to estimate the mapping from the video images to the terrain in the ground plane without using position sensors on the TV camera. Our strategy relies on three complementary techniques: (1) initial homography estimation using line-feature matching, (2) homography estimation with line-feature tracking, and (3) incremental homography estimation through point-feature tracking. Together, they allow continuous homography estimation over time, even during periods where the video does not contain sufficient line features to determine the homography from scratch. We illustrate the complementarity of the 3 techniques on a set of challenging examples.

1 Introduction

In the current era of mass entertainment, sport broadcasting has become an indispensable ingredient. Given the interests at stake and the huge demand for game analysis, much research has been done over the last decade for enhancing the broadcast video data with meta-data of particular interest to sports fans or coaches, such as player trajectories, off-side lines, etc. Ideally, these data should be produced instantly to help to understand the game as it evolves. Of course, computer vision is at the heart of this research effort, since it can provide the needed automatic procedures, which are usually done in a labor-intensive way. In terms of metric concepts, one needs to transform the relevant data defined in the image coordinate system into a real-world coordinate system, in a frame attached to the terrain field, a process generally referred to as "rectification".

For planar scenes, the image-to-model mapping is a homography, or linear mapping in $\mathbb{P}^2$[1]. It depends on the current internal configuration of the camera and its position with respect to the field. In particular, under camera motion, this transformation evolves. In the case of sport scenes observed by static cameras, one may perform classical calibration beforehand, and then use the computed homography[2]. Now, when the camera moves (as in most outdoor

⋆ This work was sponsored by the Region Wallonne under DGTRE/WIST contract 031/5439.

A. Gelbukh and A.F. Kuri Morales (Eds.): MICAI 2007, LNAI 4827, pp. 736–746, 2007.

sports), one needs to estimate this transformation continuously, which makes the pre-calibrated framework infeasible. Another key requirement of a scene analysis system is the management of uncertainties, as it is important to evaluate the precision of a given position or velocity, e.g. in multi-camera tracking applications. This article proposes a global approach to estimating the image-to-scene homography and its uncertainty in most typical team-sports scenarios, without camera motion sensor.

Most previous work do not fully exploit the temporal continuity of the image sequence and rely on pattern recognition techniques based on line-feature matching[3,4]. In some cases, we may try to calibrate the camera entirely by using the geometric properties of the scene[5]. However, this is often unnecessary, unless the application intrinsically requires 3D information, e.g. for the ball position[6]. In the case of ice hockey, an interesting approach was proposed that involves tracking points within the video over time to estimate inter-image homographies, and using line and circle features for fitting the field model[7]. However, such an approach is difficultly adaptable to soccer, as line features are not always present. In the context of sports, inter-image homographies have been used in mosaicking applications[8]; one of our ideas is to accumulate them across frames to provide estimates of the image-to-model transformation.

The transformation between a plane and its projection is well known as a homography. Here, the image-to-model homography maps points in the model to points in the image through the 3×3 matrix H, that is, up to a scale factor,

$$HP \sim p, \tag{1}$$

where $P = (X, Y, 1)^T$ is a model point and p, image of the point P through the TV camera, is denoted as $(u, v, w)^T$. As an input to our system, we consider a model (e.g., the soccer rules) composed of N line segments $\mathscr{M} = \{S_k = (P_k^b, P_k^e), k = 1, \ldots, N_M\}$, where S_k is a line segment with vertices P_k^b and P_k^e and support line $L_k \in \mathbb{R}^3$, i.e. $(L_k)^T P = 0$ if point P belongs to L_k.

An overview of our rectification approach is depicted in Fig. 1. Three modules are used for initializing and maintaining the homographies. The line-detection (LD) module allows for the initialization of the homographies from scratch, as shown in light gray in Fig. 1. This module is based on the approach presented in [4], which relies on a geometrically-motivated hypothesis generation-verification scheme to match the set of detected lines (i.e., white markings on football fields) to a known model. Although the vanishing points detection is not very precise, we use the approximate results given by this algorithm as an input of the two other modules. The line-tracking (LT) module, shown in medium gray, tracks the lines whose positions can be currently estimated using classical registration techniques. A module for ego-motion estimation, and thus called the visual-odometry (VO) module, shown in dark gray, incrementally computes the motion across images by tracking feature points from image to image. The LT and VO modules are sucessively described in the next sections.

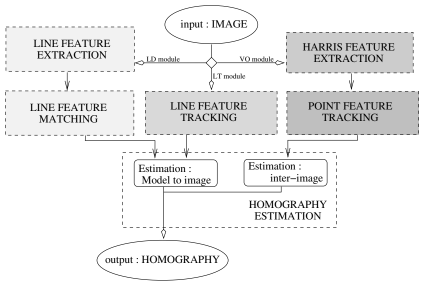


Fig. 1. Overview of our approach

2 Line Tracking (LT): Updating of Homographies

When a sufficient number of line features are visible, efficient and precise line tracking from image to image is critical for ensuring fast re-estimation of the homography H^t, i.e. the image-to-model homography at time t, supposing we have an estimation of H^{t-1}. This section briefly describes how line tracking is performed and explains how its results are used to estimate H^t.

Suppose that we have estimated H^{t-1} at time $t-1$. Then, all the line segments that form the field model $\mathscr{M} = \{S_k = (P_k^b, P_k^e), k = 1..N_M\}$ can be reprojected onto the image in a set $\{s_k = (p_k^b, p_k^e), k = 1..n_M\}$, where $n_M \leq N_M$. Each of these segments s_k is warped at regular samples along the direction that is orthogonal to the line. We use for that a traditional correlation technique on color profiles [9], i.e., at every point $\pi_{k,l}$, we look for a point $\pi_{k,l}^*$ such that the correlation with a reference color profile is maximum in the search direction.

These maxima are used as candidate points for the estimation of the new line parameters for the corrected version s_k^* of s_k through a robust RANSAC procedure on the set of $\{\pi_{k,l}^*\}$. It allows to get rid of outliers due to spurious local warpings, e.g. coming from players. The result of this process is a set of line segments s_k^* together with the corresponding inliers $\pi_{k,l}^*$.

If a sufficient number of matches are available between the current image and the field model, then we can use them to estimate the homography H.

In the classical approach, one would use line correspondences between s_k^* (in the image) and their counterpart in the model, the line segments S_k^*. Two constraints on the homography H would be established for each correspondence; however, the line parameters do not constitute very reliable measures, as a first estimation of them is needed. By contrast, we use the points $\pi_{k,l}^*$ that have been successively warped to express contraints on the homography H:

$$L_k^T (H \pi_{k,l}^*) = 0. \tag{2}$$

This equation is asymmetric and differs from traditionnal approaches that use direct correspondences (points-to-points or lines-to-lines). Here, each of the n_s warped point gives rise to one linear equation (instead of two) in the homography, so that we obtain a linear equation in the 9×1 vector h containing the elements of the matrix H, i.e. $h = (H_{11}, H_{12}, H_{13}, \ldots, H_{33})^T$.

The complete system has the form $Sh = 0$, where S is a $n_s \times 9$ matrix. Each row s_i of the matrix S is given by

$$s_i = \big(\sin\alpha_{\iota(i)}u_i \; \sin\alpha_{\iota(i)}v_i \; \sin\alpha_{\iota(i)} \; \cos\alpha_{\iota(i)}u_i \; \cos\alpha_{\iota(i)}v_i \; \cos\alpha_{\iota(i)}u_i \; c_{\iota(i)}u_i \; c_{\iota(i)}v_i \; c_{\iota(i)})\big)$$

where (u_i, v_i) are the coordinates of one of the warped point, and $L_k = \big(\sin\alpha_k \; \cos\alpha_k \; c_k\big)^T$ are the $k-th$ line parameters. We denote by $\iota(i)$ the mapping that associates an index among the rows of matrix S with an index among the set of lines, i.e., $\iota(i) = \iota(j)$ when the equations coming from rows i and j arise from points belonging to the same line. The solution of this system is clearly the right singular vector of matrix S associated to the smallest singular value. Note that, since the points have been filtered through local RANSAC on each warped line, there is no need for a robust strategy here; just one linear system is solved.

Moreover, to improve the precision of the linear rectification method presented above, we perform a few iterations of non-linear optimization in a Levenberg-Marquardt scheme. The geometric criterion C we minimize is the sum of the squared distances between warped points and the projection of their corresponding line in the model, i.e.

$$C = \sum_{k,l} (n(H^{tT} L_k).\pi^*_{k,l})^2$$

where $n(l) = \frac{1}{\sqrt{l_1^2+l_2^2}}(l_1, l_2, l_3)^T$ for $l \in \mathbb{R}^3$, $(l_1, l_2) \neq (0, 0)$. Figure 2 shows examples where the optimization of C in $\mathbb{R}^9$ improves the estimation of H.

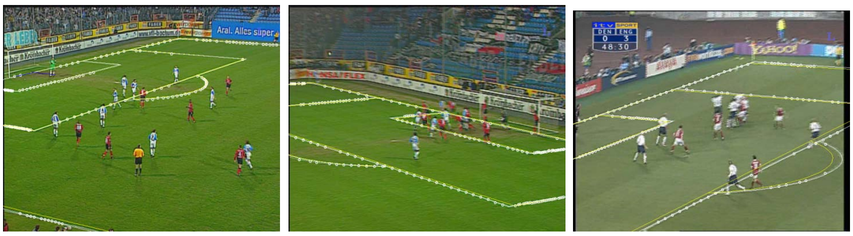
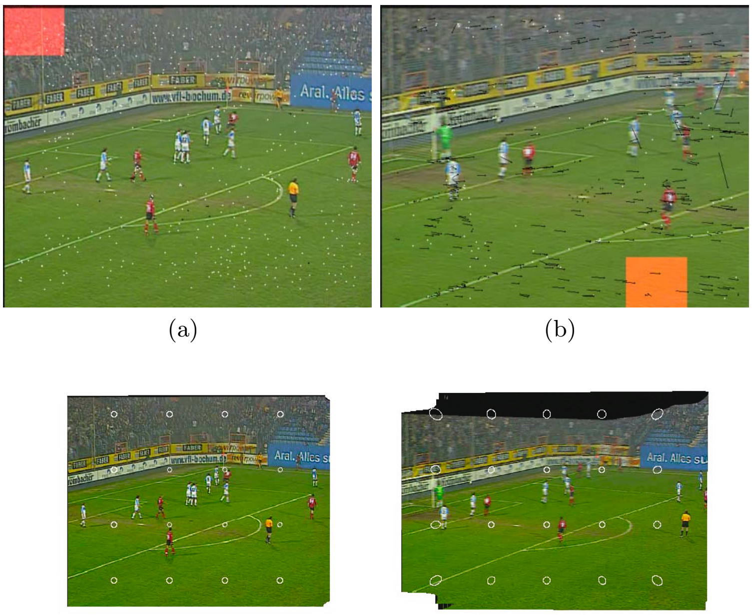


Fig. 2. Effects of non-linear refinement: grey lines are projected using the estimate of H computed by the linear method; white lines are projected using the refined estimate. The white circles are the inlier warped points $\pi^*_{k,l}$.

3 Visual Odometry (VO): Incremental Updating

The VO module aims essentially at helping in those situations where, knowing an estimate H, there are not enough known line features visible in the image. To handle this situation, we permanently track a set of locally salient features in the image, and, when needed, we use frame-to-frame matches to determine ΔH, which is the homography from image $t-1$ to image t, i.e. $\Delta H^t p^{t-1} \sim p^t$ for all points p^{t-1} in the image $t-1$ with counterparts p^t at time t.

These points are detected through a Harris detector and tracked by the KLT tracker. One of the main problems here is to modify the Harris detector so that the points are spatially spread over the largest extent possible. For this purpose, we use an adaptive algorithm for point selection[10].



(a) (b) (c) (d)

Fig. 3. Motion estimation (VO module): tracked points and frame projection into mosaic. Inlier and outlier points are respectively in white and black in the first row. Highlighted rectangular areas are areas where new feature points are searched.

Figure 3 illustrates the process involved in the VO module. It shows motion estimation and the corresponding mosaicking over 150 frames of a short sequence with fast camera rotation. In Fig.3(a) and 3(b), the tracked points are displayed, in white if they correspond to inliers of the ΔH^t estimation process, in black otherwise. Note that the number of outliers may be very large in proportion (as in Fig.3(b)), especially under strong blurring situation, so that establishing the correspondences through KLT is meaningful. Figures 3(c) and 3(c) show the corresponding mosaics. Note that whenever the image blur becomes too large, a failure is possible, which is the main limit of this module.

4 Using Multiple Homography-Estimation Modules

This section describes the process of combining the homography estimates we obtain from the techniques we described previously.

4.1 Estimating the Covariance on Estimated Homographies

A key to the success of any application that seeks to recover geometric information is a proper evaluation of the error associated with the estimated results. We model all the noisy quantities we use (feature positions, estimated objects...) as zero-mean Gaussian random processes. In the computer vision area, there have been two seminal contributions to evaluation and propagation of the covariance matrices, all relying on perturbation theory, i.e. a generic, second-order method[11], and first-order methods that apply to all linear, overdetermined optimization problems solved through matrix spectral decomposition [12,13].

Our procedure is adapted from the aforementioned works, but the nature of the equations is quite different. We propagate uncertainties on points detected in image and uncertainties on model line features (straight lines L_i) up to the homography H. Let us consider that the variance on point $\pi^*_{k,l}$ is isotropic with value σ^2. The covariance on the parameters h is denoted as Λ_h. Starting from the Jacobian expressions [13], we get

$$\Lambda_h = J^T(\sum_{l=1}^{n_s}\sum_{m=1}^{n_s}(h^T E(\delta s_l^T \delta s_m)h)e_l e_m^T)J = J^T \Lambda J \tag{3}$$

where the $n_s \times 9$ matrix J is given by $J_{ij} = -\sum_{l=2}^{9} \frac{U_{il}V_{jl}}{\lambda_l}$, U, V and $\{\lambda_l\}$ being the result of the singular value decomposition of S (left, right singular vectors and corresponding singular values). Vectors e_l are such that $e_l(i) = \delta_{li}$, where δ_{lm} is the Kronecker symbol. Assuming noise variances σ_α^2 and σ_c^2 on the angle, constant line parameters (assumed uncorrelated), and noise variance σ^2 on image points coordinates, we finally obtain

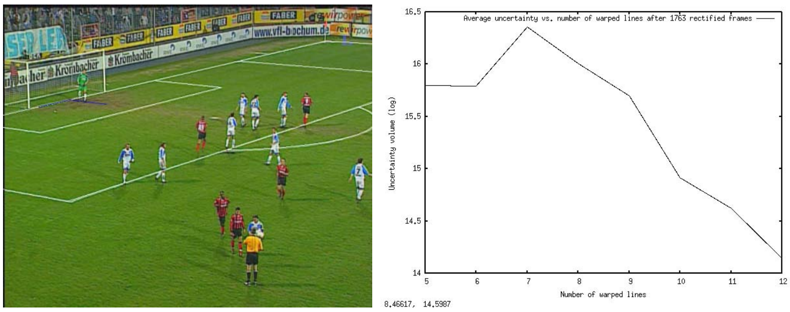
$$\Lambda(i,j) = h^T(\delta_{ij}A_{ij} + \delta_{\iota(i)\iota(j)}B_{ij})h \tag{4}$$

where ι maps row indices in S onto indices in the set of warped lines. The matrices A_{ij} and B_{ij} only depend on the line parameters, point coordinates and their respective uncertainties; they are not detailed here for lack of space.

4.2 Propagating Uncertainty Through Rectification Algorithms

Once the covariance Λ_h has been computed, we can use it to evaluate errors on transformed points. As transformed coordinates of a point $P = (X, Y)$ in the model frame are derived from a multiplication between H and an image point $p = (u, v, 1)^T$, uncertainties on P naturally combines two terms:

$$\Lambda_P = J_W(H\Lambda_p H^T + J_H \Lambda_h J_H^T)J_W^T \tag{5}$$



(a) (b)

Fig. 4. Image rectification: (a) projection of the model lines with estimated h (in white); (b) evolution of the uncertainty volume (determinant of the uncertainty matrix Λ_P) for the image of a given point with respect to the number of warped lines.

where matrices J_W and J_H are Jacobian matrices[12] and matrix Λ_p is the covariance on image points. Typically, this could be the output of a tracking algorithm (uncertainty on the location of the target in the image).

As an illustration, Fig. 5(e) shows the result of error propagation around a homography computed from Fig. 5(a). Red lines in Fig. 5(a) indicate the reprojected model from which the homography has been computed. In Fig. 5(e), at each point from a regular grid, the covariances are represented by a 3σ ellipse.

Finally, as expected, the average error decreases with the number of warped lines, as illustrated in Fig. 4(b). A given point was chosen in the image (the center) and the plot shows how a given uncertainty volume in the image domain is transformed in the model domain throughout a given video sequence (where the viewpoint does not change too much). The more warped lines we have, the smaller uncertainties are, on average over several hundreds of frames.

4.3 Integrating Uncertain Inter-image Homographies

Our policy takes line-based rectification as the first choice. In particular, to save time, inter-image homographies are not computed as long as enough lines are tracked. However, points are detected all over the image at each t. This allows, when the LT module fails for any reason at time t, to compute an estimate ΔH^t and to update H^t through $H^t = H^{t-1}\Delta H^t$.

In a similar way as in 4.1, uncertainties on ΔH^t are computed through Eq. 4 by propagating uncertainties on the matched points, which leads to an estimate of the uncertainty on H^t. As an illustration, the images of points from the first frame of Fig. 3 (left) into the current frame (right) have accumulated uncertainties that are shown in white in the mosaic image.

Note that when the VO module is used, this uncertainty will tend to grow, as we accumulate errors on relative measures. At some moment, this leads to ambiguities while reprojecting the model (one line taken for another one), so that we switch to the LD module whenever the uncertainty level corresponding to the projection of the center of the image point becomes too large.

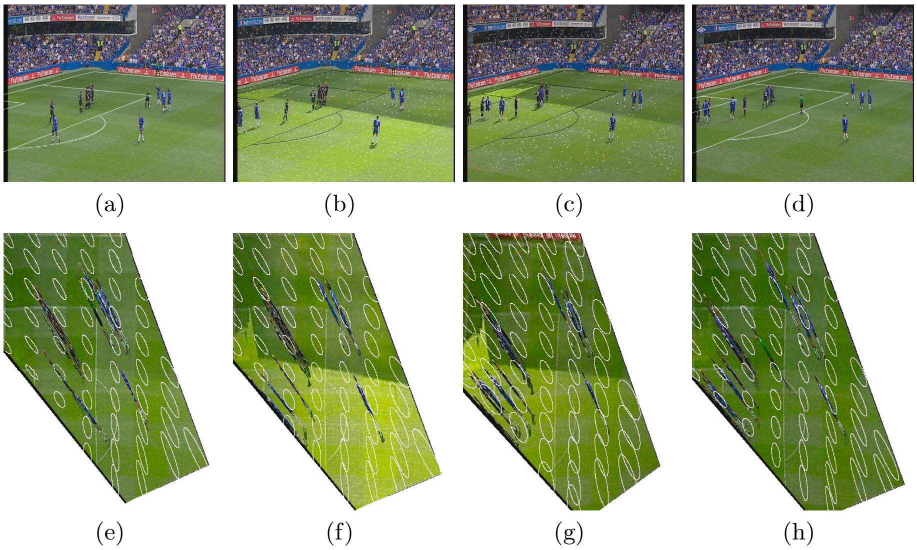


Fig. 5. Sequence 1 : after 500 frames of line-based rectification (a), the system switches to motion estimation because lighting changes render line warping unsuccessful (b). The inlier points (in white) allow motion estimation for more than 400 frames (b,c), until light conditions become normal and the LT module can restart.

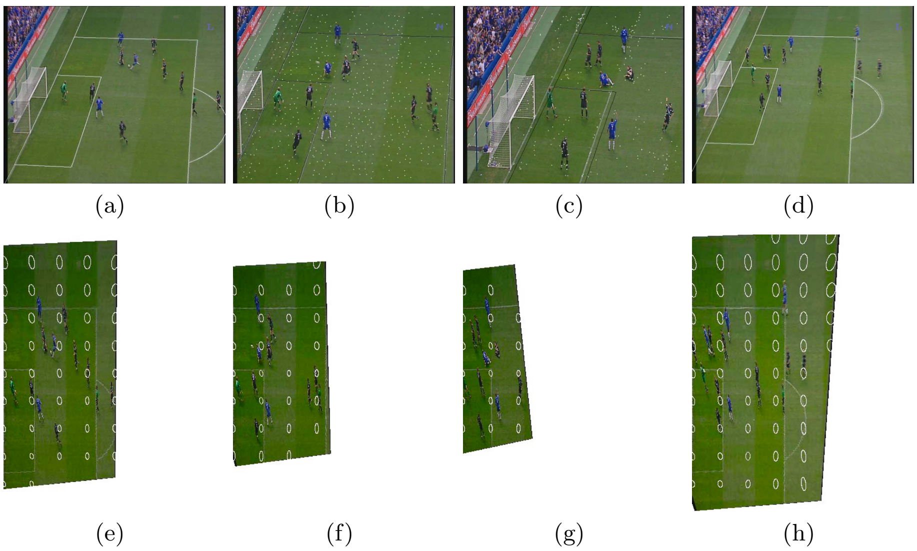
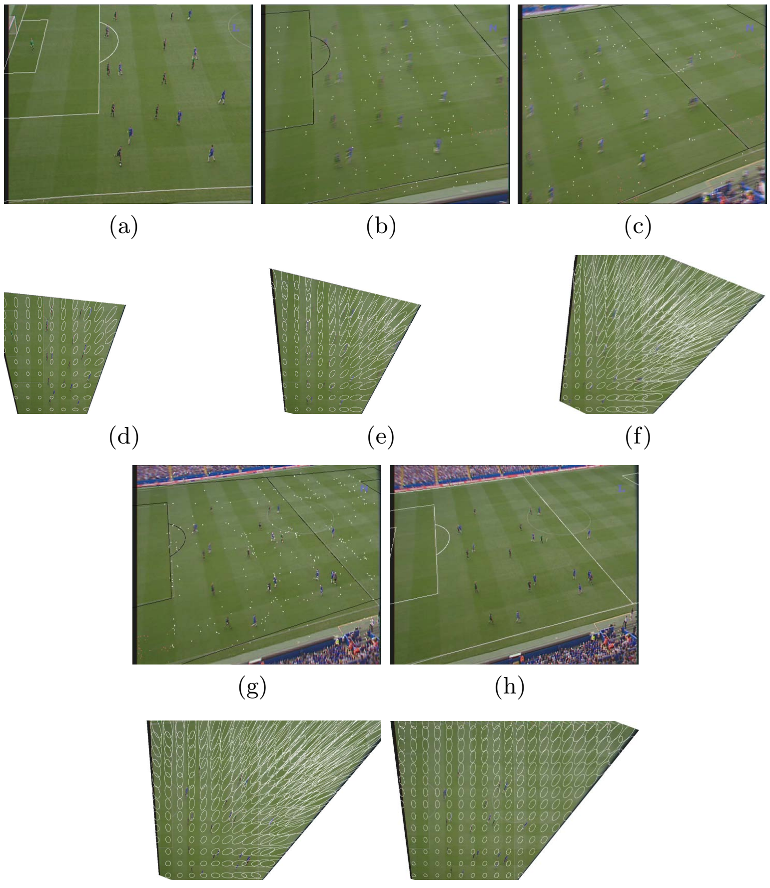


Fig. 6. Sequence 2 : line-based rectification is lost at frame 193 (a), motion estimation is done under fast zooming and rotation (b-c), and lines are recovered a frame 300 (d)

5 Results

This section presents results of our rectification system for different situations, where the complementarity of the 3 modules we described plays a key role in the success of the continuous, automatic estimation of H.

A first example in Fig. 5 illustrates the benefits of using a point-based method to estimate increments of homographies, as the VO module (that projects the model onto black lines) helps to maintain an estimate H while the LT module (that projects the model onto white lines) fails. Indeed, after hundreds of frames rectified by the LT module, sudden, strong lighting changes occur. Starting from frame 529 (Fig. 5(b)), many line segments are strongly reduced in contrast, and



(a) (b) (c) (d) (e) (f) (g) (h) (i) (j)

Fig. 7. Sequence 3: after staying for a while near the goal area, the camera performs an abrupt motion towards the center, where LT is not possible (c). After unzooming (frame 467, f), goal areas are visible again. Motion estimation has accumulated a lot of uncertainty (e-f-i), but LT is successful on frame 471.

the local warpings are unsucessful. However, point tracking remains sufficiently reliable for approximately another 400 frames. Even though uncertainties continue to grow during this period, as seen in Figs. 5(f) and 5(g), line tracking resumes without problems in frame 988 (Fig. 5(d)).

Complementarity also appears in Fig. 6 and Fig. 7. These examples show that even under fast motion and strong zooming, when white lines are lost (Fig. 6(b)), it remains possible to rely on motion estimation to recover after a while. However, this motion has to remain local and short, in order to avoid accumulating errors in motion estimation. For example, in Fig. 7, the motion is fast and its amplitude is large and so are the accumulated errors and corresponding uncertainties (see Fig. 7(e), 7(f), 7(i)). This last example is an extreme case : the VO module should have been stopped long before for switching to the LD module but has been let running on purpose to illustrate its limits.

Our implementation runs at about 3Hz on a 3.2GHz Bi-Xeon machine, which is a bit slow, even for 720×576 images. Most of the time is spent on point detection and tracking, so our current optimization efforts are focused on it.

6 Conclusion and Future Work

We have presented an approach to estimate the image-to-field homography and its uncertainty continuously over sport sequences : after automatic rectification on a single frame, it tracks line features and re-estimates the homography with the constraints induced by image points warped locally onto the corresponding line feature. Finally, it estimates the incremental homography between two frames when line-based rectification is not possible. By way of several challenging examples, we have demonstrated the added benefit of the two techniques.

Our system still needs to be improved to cope with very long sequences. Very fast camera motion causes our system to lose track until sufficient lines come available to allow reinitialization by the first method. Our current work seeks to improve several aspects of the system. First, we would like to perform better filtering of homography estimates in a lower-dimensional parameter space, which implies on-line determination of internal parameters of the camera. Second, as shown in the experimental results, motion estimation through visual odometry accumulates uncertainty and necessarily leads to drift. Currently, we simply switch to LD module when uncertainty becomes too large, but a better answer to this problem would be to identify particularly salient points of the planar scene and incorporate them into the scene model as landmarks.

References

1. Hartley, R., Zisserman, A.: Multiple View Geometry in Computer Vision. Cambridge University Press, Cambridge (2000)
2. Ren, J., Orwell, J., Jones, G., Xu, M.: Real-time 3D soccer ball tracking from multiple cameras. In: BMVC 2004. Proc. of the British Machine Vision Conf, pp. 829–838 (2004)

3. Farin, D., Krabbe, S., de With, P., Effelsberg, W.: Robust camera calibration for sport videos using court models. SPIE Electronic Imaging 5307 (2004)
4. Hayet, J., Piater, J., Verly, J.: Fast 2D model-to-image registration using vanishing points for sports video analysis. In: ICIP 2005. Proc. of IEEE Int. Conf. on Image Processing, Genoa, Italy, vol. 3, pp. 417–420 (2005)
5. Lihe, Q., Luo, Y.: Automatic camera calibration for images of soccer match. In: Proc. of the Int. Conf. on Computational Intelligence, pp. 482–485 (2004)
6. Reid, I.D., Zisserman, A.: Goal-directed video metrology. In: Buxton, B.F., Cipolla, R. (eds.) ECCV 1996. LNCS, vol. 1065, pp. 647–658. Springer, Heidelberg (1996)
7. Okuma, K., J., L.J., Lowe, D.G.: Automatic rectification of long image sequences. In: Okuma, K. (ed.) ACCV 2004. Proc. of the Asian Conf. on Computer Vision (2004)
8. Kim, H., Hong, K.: Robust image mosaicing of soccer videos using self-calibration and line tracking. Pattern Analysis and Applications 4, 9–19 (2001)
9. Drummond, T., Cipolla, R.: Real-time tracking of complex structures with on-line camera calibration. In: BMVC 1999. Proc. of the British Machine Vision Conf, pp. 574–583 (1999)
10. Brown, M., Szeliski, R., Winder, S.: Multi-image matching using multi-scale oriented patches. In: CVPR 2005. Proc. of IEEE Conf. on Computer Vision and Pattern Recognition, pp. 510–517. IEEE Computer Society Press, Los Alamitos (2005)
11. Haralick, R.: Propagating covariance in computer vision. In: Proc. of ECCV Workshop of Performance Characteristics of Vision Algorithms, pp. 1–12 (1996)
12. Criminisi, A.: Accurate Visual Metrology from Single and Multiple Uncalibrated Images. PhD thesis, University of Oxford, Dept. Engineering Science (1999)
13. Papadopoulo, T., Lourakis, M.: Estimating the jacobian of the singular value decomposition: Theory and applications. In: Vernon, D. (ed.) ECCV 2000. LNCS, vol. 1843, pp. 554–570. Springer, Heidelberg (2000)