

FAST 2D MODEL-TO-IMAGE REGISTRATION USING VANISHING POINTS FOR SPORTS VIDEO ANALYSIS

Jean-Bernard Hayet, Justus H. Piater, and Jacques G. Verly

Department of Electrical Engineering and Computer Science (Institut Montefiore)
University of Liège, B28, B-4000 Liège 1, Belgium

ABSTRACT

This article focuses on the field-to-image registration for TV sports-related image processing through basic geometrical reasoning. We propose a robust automatic line-based method that can cope with imprecisions in the line detection steps through a hypothesis generation/pruning paradigm where the number of hypotheses is reduced by using the vanishing points. The results are satisfying and illustrated with examples from sport.

1. INTRODUCTION

It has been several decades now that the TV industry has developed a major interest in sports and seeks cheap, simple technology to rejuvenate broadcasting through augmented reality or on-the-fly event detection, e.g., offside detection, or distance measurement. As most of these operations require teams of people to track players or to mark lines throughout the broadcast, substantial research has been done in the last few years towards automated sports game analysis, some of which is being used in commercial systems. One of the keys for the computer to understand a game is to develop algorithms to (1) track the players and (2) to maintain an estimate of the geometry of the scene, continuously through the game. We address aspects of the second point here.

The geometry of the problem is well known. The transformation that maps points in the game field to points in the image is a homography, i.e., a linear mapping from a 2D projective space to another 2D projective space, whose estimation requires that matches be established between field features and image features. Here, we propose a new algorithm for estimating this mapping by matching line segments detected in the image to line segments in the model.

Most of the existing reports on image rectification in sports sequences focus on the case of static cameras [1, 2], where calibration may be done once by hand. The problem we address involves moving cameras (translation, pan, tilt,

zoom), so the challenge is to come up with a fast, robust, and automatic approach. Related work has been done in the context of ice hockey [3], where point features are detected and tracked by a KLT process. Recently, homography-based registration has been investigated in a scheme using assumptions regarding the camera model [4]. The contributions we make here are (1) to rely on geometrical reasoning only and to avoid any assumptions regarding the camera model, and (2) a rapid and robust system capable of coping with different, imperfect line detectors. Section 2 describes the global approach we propose for registration. Section 3 addresses the generation of hypothetical model-to-image mappings. Section 4 tackles the problem of validating these mappings. Section 5 discusses results obtained with our new method.

2. OVERVIEW

2.1. Model-to-image registration

The first step to perform registration is to detect image features to be matched to the model features. Since we focus on sport fields, we mainly use straight line segments as features. Many methods exist for detecting straight line segments, e.g., Hough methods [4] or edge segmentation methods [5]. Note that, in our strategy, it is not critical that the line detectors be perfect. Incomplete segments, missing segments, and spurious segments will be taken care of by the approach to robust segment matching described here.

In this paper, we specifically focus on the case where all lines can be classified as belonging to at most two distinct vanishing points p_v and p_h . This is the case with most markings on sports fields.

Following the extraction of the segments from the image, we group the segments as corresponding to p_v or p_h and we simultaneously compute p_v and p_h , e.g., as described in [5]. These two sets of segments are illustrated in Fig. 1, where we use a specific gray level for each set.

The second step is to match these two sets of image segments to the segments in the model.

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

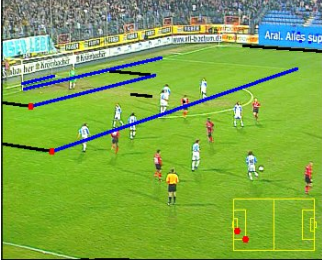


Figure 1: Example of line extraction, grouping, and image/model point matches (shown in lower-right corner).

2.2. Matching image segments to model segments

A simple minded approach would be to use brute force matching of all image segments (many) to all model segments (generally fewer). With enough information to avoid ambiguity, we could even just match independently the segments corresponding to each class. However the matching would require information about the camera model [4].

Here, we consider all pairs of segments corresponding to the Cartesian product of the segments in the set corresponding to p_v and in the set corresponding to p_h . We find them both in the image (in real time, online) and in the model (once, offline).

The support lines of the segments in each pair cross at some point. Therefore, instead of matching pairs of line segments, we match the corresponding intersection points, taking into account their particular configuration. One can show that, once p_v and p_h are known, only two such point matches are required to perform the registration. This is illustrated in Fig.1 by the pairs of dots in the image and the overlaid model. We use a hypothesis generation/pruning paradigm to produce a set of reliable point matches. This is explained in Sect. 3 and 4, respectively.

The vanishing points are generally not very precise, as there are not a lot of lines in the type of scene we consider, so the registration will be a rough one. However, it is straightforward to refine this rough registration by locally warping the backprojected features [5].

3. GENERATION OF HYPOTHESES

A hypothesis consists in the collection of a first pairing of an intersection point p_1 in the image to an intersection point P_1 in the model and a second, similar pairing of p_2 to P_2 . It is denoted by $h = (p_1, P_1; p_2, P_2)$. Since it is out of question to consider all pairs of line crossings present in the image, we select a subset of them by using the relative positions of line segments.

3.1. Characterization of pairs of image/model segments

We characterize each pair of image and model segments according to the relation between the segments. We define 17 such relations, 3 of which are illustrated in Fig. 2. These 17 relations depend upon the relative positions of the segments with respect to (a) each other, (b) the intersection of the support lines, and (c) the position of the vanishing points. In Fig. 2, relations Q_1 , Q_2 and Q_3 are differentiated by the position of one segment w.r.t. the other. The other 14 relations result from the characteristics of the intersection point (i.e., whether it belongs to none, one, or both segments) and from the segments relative positions w.r.t the vanishing points. The set of 17 relations is denoted by \mathcal{Q} .

Since the support lines of the pairs of segments of interest intersect, we denote these pairs of segments by their points of intersection, say p_i (in the image) and P_j (in the model). The relation associated with each pair is denoted by $q(x)$, where x is either p_i or P_j .

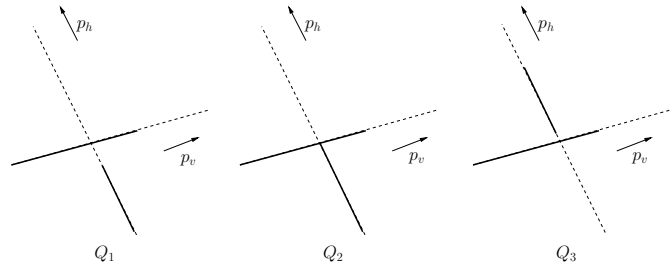


Figure 2: Qualifying a line crossing by using the vanishing points: a few examples of elements of \mathcal{Q} .

The similarity measure $\gamma(p_i, P_j)$ between p_i and P_j is then defined in terms of $q(p_i)$ and $q(P_j)$, i.e.,

$$\gamma(p_i, P_j) = c(q(p_i), q(P_j)).$$

The function c is defined empirically to penalize topologically incorrect correspondences. This is illustrated in the following table for the 3 relations shown in Fig. 2. Of course, in actuality, we work with a 17×17 table:

	Q_1	Q_2	Q_3
Q_1	1	0.5	-1
Q_2	0.5	1	-1
Q_3	-1	-1	1

3.2. Generation of hypotheses

Point matches are denoted by (p_k, P_k) . A naive idea would be to generate hypotheses from all pairs of matches such that $\gamma(p_k, P_k) > 0$. However, the combinatorial burden could be significant. To alleviate it, we proceed as follows.

First, we build two lists. The first, l^+ , includes all candidate point matches (p_k, P_k) such that $\gamma(p_k, P_k) = 1$ (i.e.,

the best topological similarity). The second, l^- , consists of the matches satisfying $\gamma(p_k, P_k) > 0$ (note that $l^+ \subset l^-$).

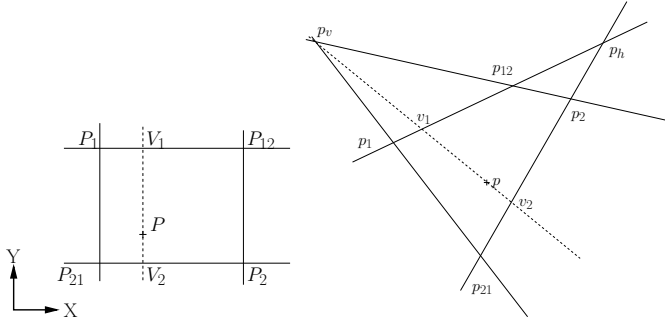
Second, we build the list of hypotheses $(p_k, P_k; p_l, P_l)$ from all pairwise combination in $l^+ \times l^-$, without degenerate cases (i.e., the same support line can only appear once).

4. PRUNING OF POINT MATCHES HYPOTHESES

Let us consider the hypothesis $h = (p_1, P_1; p_2, P_2)$ generated as described above.

4.1. Fast projection of a model point onto the image

Assuming the vanishing points p_v and p_h are known, we consider the geometrical setup of Fig. 3, i.e., the points P_{12}, P_{21}, V_1 and V_2 (in Fig. 3(a)) and their projections (in Fig. 3(b)). We want to express the projection p onto the image of any model point P given the hypothesis h . Therefore, we want to find the relation Π_h such that $p = \Pi_h(P)$.



(a) Setup for model (b) Setup for image

Figure 3: Geometrical setup for fast projection.

To determine Π_h , we avoid a full homography estimation, which requires inverting matrices. Instead, we use the fact that cross-ratios remain invariant under perspective projection. To derive an expression for the position vector \underline{v}_1 corresponding to point v_1 , which is the projection of point V_1 obtained as shown in Fig. 3, we first write

$$[p_1, v_1, p_h, p_{12}] = \frac{X - X_1}{X_2 - X_1}, \quad (1)$$

where $[a, b, c, d] = \frac{(l_b - l_a)(l_d - l_c)}{(l_d - l_a)(l_b - l_c)}$, l_p being the abscissa of point p , measured on the straight line defined by points a, b, c and d w.r.t. any reference point on this line. X is the abscissa of point P in the model frame as indicated in Fig. 3. Points P_1 and P_2 , with respective coordinates $(X_1, Y_1)^T$ and $(X_2, Y_2)^T$ define the hypothesis under consideration.

Equation 1 can be rewritten as

$$\frac{(l_{v_1} - l_{p_1})(l_{p_{12}} - l_{p_h})}{(l_{p_{12}} - l_{p_1})(l_{v_1} - l_{p_h})} = \frac{X - X_1}{X_2 - X_1}.$$

Then, we easily find

$$\underline{v}_1 = \frac{1}{1 - \lambda_1} \underline{p}_1 - \frac{\lambda_1}{1 - \lambda_1} \underline{p}_h, \quad (2)$$

where $\lambda_1 = \frac{X - X_1}{X_2 - X_1} \frac{y_{p_{12}} - y_{p_1}}{y_{p_{12}} - y_{p_h}}$, since, by using Thales theorem, $\frac{l_{p_{12}} - l_{p_1}}{l_{p_{12}} - l_{p_h}} = \frac{y_{p_{12}} - y_{p_1}}{y_{p_{12}} - y_{p_h}}$. A similar expression can be easily obtained for \underline{v}_2 .

Using the cross-ratio $[v_1, p, p_h, v_2] = \frac{Y - Y_1}{Y_2 - Y_1}$ and Eq. 2, we obtain

$$\underline{p} = \Pi_h(P) = \frac{1}{1 - \mu} \underline{v}_1 - \frac{\mu}{1 - \mu} \underline{p}_v, \quad (3)$$

where $\mu = \frac{Y - Y_1}{Y_2 - Y_1} \frac{x_{v_2} - x_{v_1}}{x_{v_2} - x_{p_v}}$.

We now explain how this result is used to evaluate each point match hypothesis $h = (p_1, P_1; p_2, P_2)$.

4.2. Evaluation of a hypothesis

Evaluating a hypothesis means quantifying to what extent it explains the current observations, i.e., the detected lines. We consider the set $\{p_i\}$ of ‘‘intersections’’ p_i in the image, and the set $\{P_j\}$ of ‘‘intersections’’ P_j in the model. A hypothesis $h = (p_1, P_1; p_2, P_2)$ is evaluated by comparing the set $\{p_i\}$ with the set $\{\Pi_h(P_j)\}$.

To this end, we compute the quantity

$$s(h) = \sum_j \max_{\{i | p_i \in B(\Pi_h(P_j), \epsilon)\}} \gamma(p_i, P_j),$$

where the sum is taken over the model points whose projections lie in the image. $B(p, \epsilon)$ is the circle centered at p and with radius ϵ (chosen as a tolerance on the position error). We calculate $s(h)$ for all hypotheses h generated. The hypothesis corresponding to the maximum value s_{max} of $s(h)$ is retained provided s_{max} is above a given threshold. Otherwise, no hypothesis is considered satisfactory.

We can then use the retained hypothesis (if any) for projecting any model point onto the image.

5. RESULTS

Our method has been tested on various sports video sequences. Each row in Fig. 4 gives some results for one of these videos. The first column shows the lines that have been detected and correctly associated with their vanishing points. The second column presents the resulting point matches, where the points forming the selected hypothesis are shown as black dots and the point matches consistent with this hypothesis are shown as white dots. The last column shows the backprojection of the model after the mapping has been refined. Note that for the second video, we applied a radial distortion compensation to all detected lines.

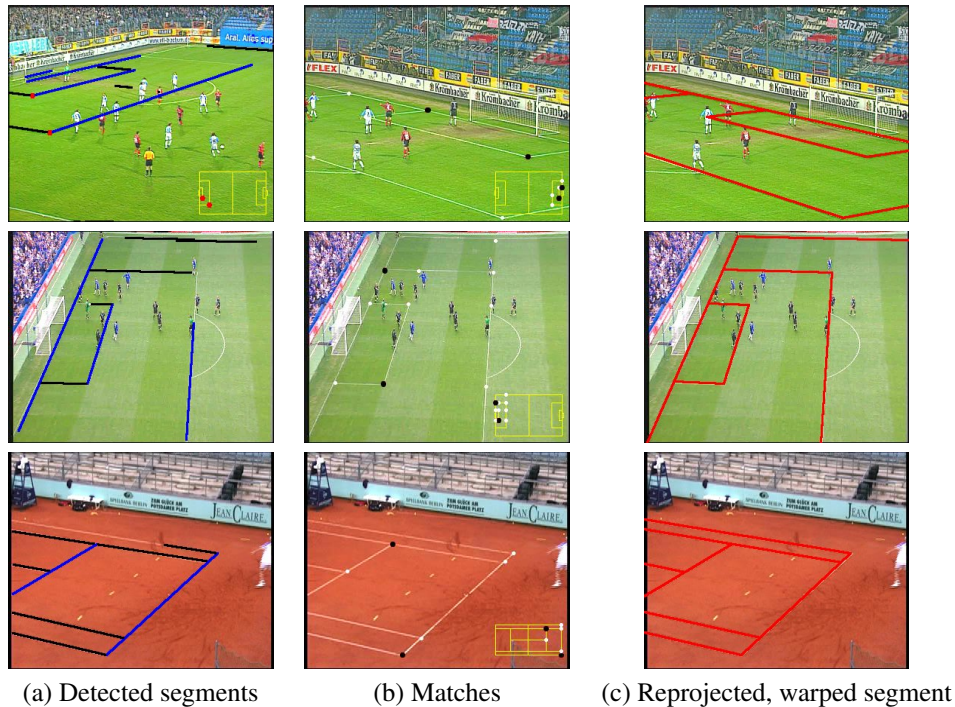


Figure 4: Results examples: each row corresponds to a different video sequence.

The system can cope with poorly segmented lines or false positives in the line detection process (see the first image, in particular). This demonstrates its overall robustness. Of course, a sufficient number of segments of reasonable quality must be extractable from the images for the registration system to work. The geometrical framework reduces the number of hypotheses to be tested to a reasonable value, which makes this approach suitable for video-rate model-to-image registration. For example, in our experiments on soccer video sequences with a Centrino 1.6 MHz processor, without taking into account line detection and vanishing point determination, matching is performed in about $10ms$ in typical situations and in at most $50ms$ in the most complex situations, i.e., with a substantial number of spurious segments.

6. CONCLUSION

We have proposed a novel method for performing image registration in sports video sequences in the specific cases where a set of segments of a planar model have to be matched to detected segments in an image. The originality of the approach is that we rely almost exclusively on geometrical reasoning to search for a best match and that our hypothesis generation/pruning scheme simplifies the pruning step by avoiding the explicit computation of the homography thanks to the use of the scene vanishing points. The method is fast

and performs reliably, even if line detection is erratic. This makes it suitable for online, i.e., real-time, (re)initialization of homographies in soccer game analysis.

7. REFERENCES

- [1] J. Ren, J. Orwell, G.A. Jones, and M. Xu, "Real-time 3D soccer ball tracking from multiple cameras," in *Proc. of the British Machine Vision Conference (BMVC'04)*, 2004, pp. 829–838.
- [2] J. Pers and S. Kovacic, "Tracking people in sport: Making use of partially controlled environment," in *Proc. of International Conference Computer Analysis of Images and Patterns (CAIP'01)*, 2001, pp. 374–382.
- [3] K. Okuma, Little J. J., and D. G. Lowe, "Automatic rectification of long image sequences," in *Proc. of the Asian Conf. on Computer Vision (ACCV'04)*, 2004.
- [4] D. Farin, S. Krabbe, P.H.N. de With, and W. Effelsberg, "Robust camera calibration for sport videos using court models," *SPIE Electronic Imaging*, vol. 5307, 2004.
- [5] J.B. Hayet, Piater J., and J. G. Verly, "Incremental rectification of sports fields in video streams with application to soccer," in *Proc. of the Advanced Concepts in Intelligent Vision Systems (ACIVS'04)*, 2004.