Robust Homography Optimization in Soccer Scenes

Afshin Bozorgpour, Mehran Fotouhi, Shohreh Kasaei
Department of Computer Engineering
Sharif University of Technology
Tehran, Iran
bozorgpour@ce.sharif.edu, fotouhi@ce.sharif.edu, skasaei@sharif.edu

***Abstract*—Soccer content-based analysis is still a challenging process because of lack of information about the soccer scenes. An accurate homography is one of the most important information that is used to map the frame coordinate to the world coordinate, and vice versa. In this paper, a novel approach is presented to extract the frame line mask and tune the initial homography. The proposed method utilizes lines and arcs in the frame to find the best solution. The proposed method can also detect the invalid solutions. The obtained results are compared to the state-of-the-art method that uses the *Levenberg-Marquardt* technique in its optimization stage. The experimental results show that the proposed method is robust enough against strong shadows, lighting conditions, noise, and motion blurring.**

Keywords—homography optimization; line detection; soccer field registration; sports video processing.

I. INTRODUCTION

In recent years, the availability of high quality cameras as well as the increase of the processing speed has led to increase the focus of researchers on the processing of sports videos. In general, text, audio, still images, graphics, and video are processed for analyzing sports videos. Event detection and summarization, player tracking and localization, inserting virtual advertisements (augmented reality), ball detection and tracking, tactical and team statistical analysis, and computer-assisted refereeing are common processings of sports videos.

One of the main stages in many approaches to solve these mentioned applications is finding the homography map between the camera and the soccer field. In fact, it is a map between the two planes. In the literature, several approaches have been introduced for finding the homography. Almost all of them use segments of lines and points of intersection that characterize the match field. Obviously, the accuracy of line detection affects the accuracy of the homography. The quality and resolution of captured video can also affect the precision of line detection step. As such, an optimization process is needed to find an accurate homography.

Computation of homography map is a difficult task, especially with moving cameras. This is because the location of points and lines changes in the frame coordinate by camera motions. Hence, a new homography map should be calculated. The public broadcast videos lie in this category. In fixed cameras, the homography is obtained just once and there is no need to update it.

The aim of this paper is to design a novel and robust approach to obtain an optimized homography map for soccer scenes. The inputs of our algorithm are the initial homography (that is a coarse estimation of it) and the soccer field model. First, the match field mask is generated based on the difference of pixel values. The most dominant hue value in the HSV color space is selected as a match field hue color. Afterward, the field lines are detected to generate the line mask. Then, the line mask is back projected to the field model by using the initial homography. The optimization step is started by sampling the model lines. Then, for each sample, the nearest point on the back projected line mask is determined as a corresponding match.

We use *random sample consensus* (RANSAC) to remove outliers based on the homography model. The new homography is estimated based on corresponding points by the *direct linear transform* (DLT). Homography optimization is performed until the homography error converges during several successive iterations. Experimental results show the robustness and effectiveness of the proposed method.

The rest of this paper is organized as follows. Section II presents related works for finding homography in sports scenes. In Section III, the proposed method is introduced. Experimental results are reported in Section IV. We conclude by highlighting some issues and further work in Section V.

II. RELATED WORKS

Many sports applications require the homography map between the frame coordinate and the world coordinate. Some approaches have been presented in the literature for calculating homography. In [2], four predefined points on a tennis court is detected for calibration. However, the method has to be initialized manually. It is not also robust against occlusions of the court lines connecting these four points. In [4, 5], an exhaustive search through the parameter space is introduced to initialize their soccer field detection algorithm. Battikh and Jabri [3] considered a field model, too. Hough transform is performed to detect segment lines of the model. The position, orientation, and focal length of a camera are estimated to overlay graphics on a soccer pitch.

Frame noise, motion blur, and other artifacts affect the line and corner point detection algorithms. Consequently, these methods are not good enough to find an exact homography. On the other hand, insufficient corresponding points are visible in many frames. For the same reason, arc detection is often used in addition to line detection. Arc detection methods are widely discussed in some papers. Le Troter *et al.* [6], suggest an approach for detecting the ellipse in the frame formed by the center circle. Obviously, it only works when more than half of the circle is visible. In that approach, other line features in the frame are used. Wang *et al*. in [7] suggest an approach that works entirely based on the detection of arcs, as their system detects the arc of the penalty box as well.

One of the well-known papers in this area belongs to Farin *et al.* [1]. In that work, frame pixels are classified as line pixels based on their color and texture. Then, the Hough transform is performed to extract straight lines. The detected lines in input frame are corresponded to the model field lines by a search algorithm. Homography tracking and optimization is done for the succeeding input frames. Our homography optimization method is compared to homography optimization of that work in the experimental results section.

III. PROPOSED METHOD

Most of previous works rely on finding the correspondence between the intersection of field lines and the intersection of frame lines. Therefore, by having four or more corresponding points, the calculation of homography is straightforward by applying the DLT method. The rest of the reported work is based on extraction method of camera parameters. Detecting lines in the input frame is an important step of most approaches. Frame noise and other artifacts exist in broadcasting videos. Camera motions cause motion blurring. All these artifacts effect the line detection step as well as the quality and resolution of captured videos. On the other hand, it is impossible to detect enough corresponding points in many frames. As such, calculating an exact homography is a difficult task in most frames. In fact, in real systems, an approximated homography can be calculated; because the error of the other steps is considerable.

The goal of this paper is to introduce an efficient solution to compute an accurate homography map. The inputs of our algorithm are the model of soccer field and a rough estimation of the homography between the frame coordinate and the world coordinate. The proposed method has two main stages. In the first stage, the soccer field is detected based on color information (to extract the frame line mask). The second stage is related to the optimization block. In this stage, the frame line mask is overlaid onto the soccer field model. Then, the nearest point of the projected line mask is found from some taken sample points on the field model. RANSAC algorithm is applied to find the correct correspondences. Based on corresponding points, the new homography is calculated. This algorithm is performed iteratively until the homography map converges. Fig. 1 shows the overall block diagram of the proposed method. Each stage is explained in more details next.

A. *Preprocessing stage*

In the preprocessing stage, line pixels should be extracted. This stage consists of two steps.

1) *Soccer field detection*

The aim of this step is to detect the soccer field to increase the precision and speed of the method. Therefore, pixels should be classified as “match field” and “out of match field” pixels. The color information is used for this purpose. The HSV color space is an appropriate color space that separates the hue value from other channels. We suppose that hue values of the soccer field pixels are more frequent than other pixels. For example, the color of match field in most soccer fields is green. Computing the frame histogram on the H channel and determining the most frequent bin is used for detecting match field. Obviously, some other pixels that have a similar value of hue are selected too. On the other hand, it is possible that some match field pixels are not selected because of some artifacts. Therefore, morphological operations are utilized to fill holes and remove noise. Fig. 2 shows these steps for an instance frame.

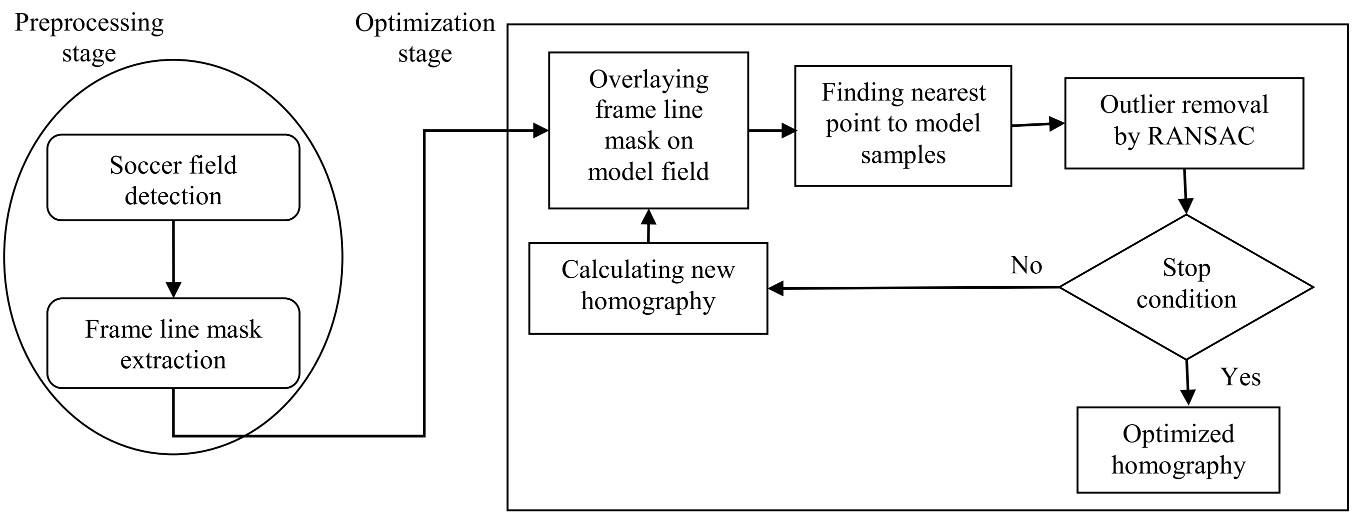


Fig. 1. Block diagram of proposed method.

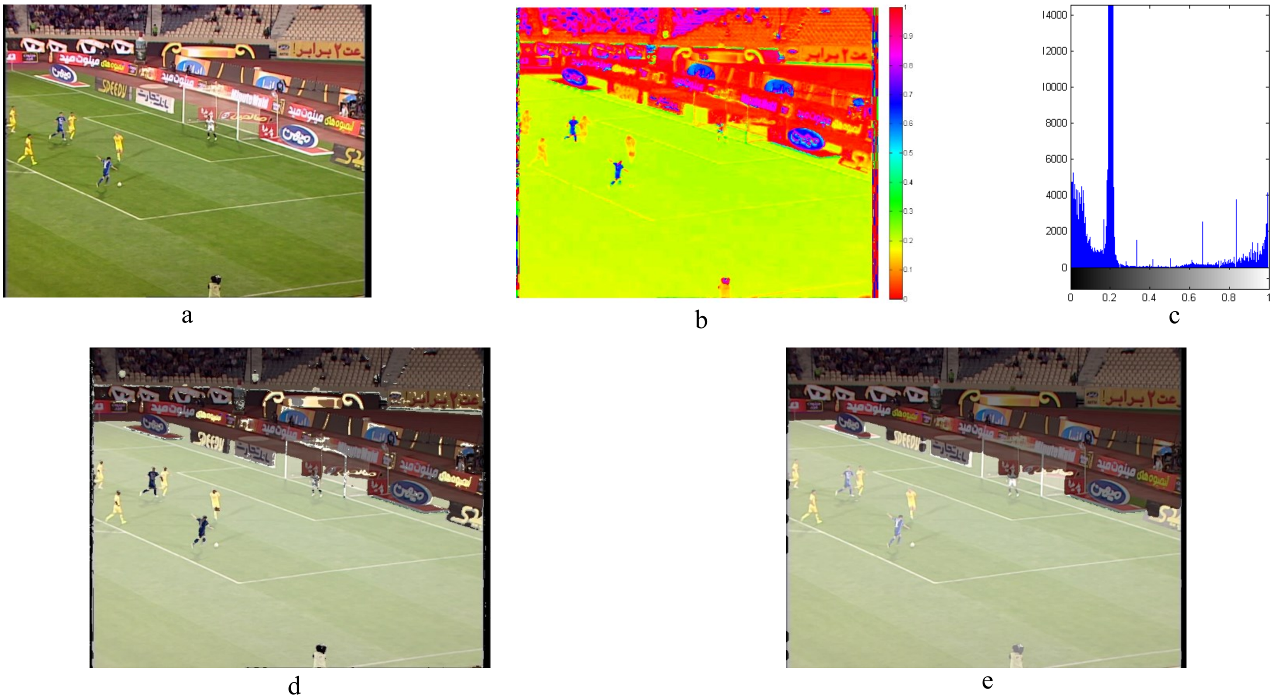


Fig. 2. a) Input frame. b) Color map in hue channel. c) Hue channel histogram. d) Result of frequent hue selection. e) Final field mask.

2) Line mask extraction

A line can be defined as a set of directed pixels with limited width with a different color from their adjacent pixels. There are several strategies to detect frame lines. Some approaches use edge detection for determining line pixels in the frame. The main drawback of the mentioned approaches is that they are very sensitive to image noise. On the other hand, our approach for extracting the frame line mask is very similar to foreground detection techniques and is based on line properties.

If a median filter is applied whose window size is greater than twice the line width, the color of line pixels is wiped out and the background area is detected. Here, the median filter is performed on the L channel in the Lab color space. The lines in soccer scenes can be categorized as horizontal and vertical. Therefore, two filters are utilized to wipe out the lines.

Fig. 3 shows the output of vertical and horizontal median filters on an input frame. It is reasonable to suppose that the width of a line is less than a player body size. As shown in Fig. 3, horizontal median filter wipes out vertical line pixels and vertical median filter wipes out horizontal line pixels. Consequently, each directional median filter generates a background image.

Each category of lines can be retrieved by subtracting the related background from the original frame. Then, histogram thresholding is applied for binarization of each extracted line category. The result of binarization operation contains some lines and as well as some noise. The noise can be detected by its area. This is because if its area is large, it could not be wiped out by the horizontal and vertical median filters and vice versa. Therefore, if its area is large enough, it is a line area. Thus, the noise is removed based on its area.

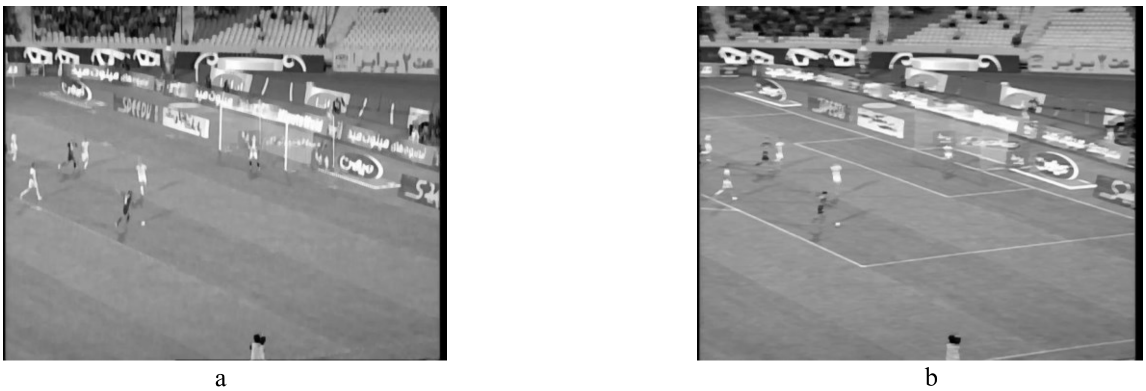


Fig. 3. a) Result of vertical median filter. b) Result of horizontal median filter.

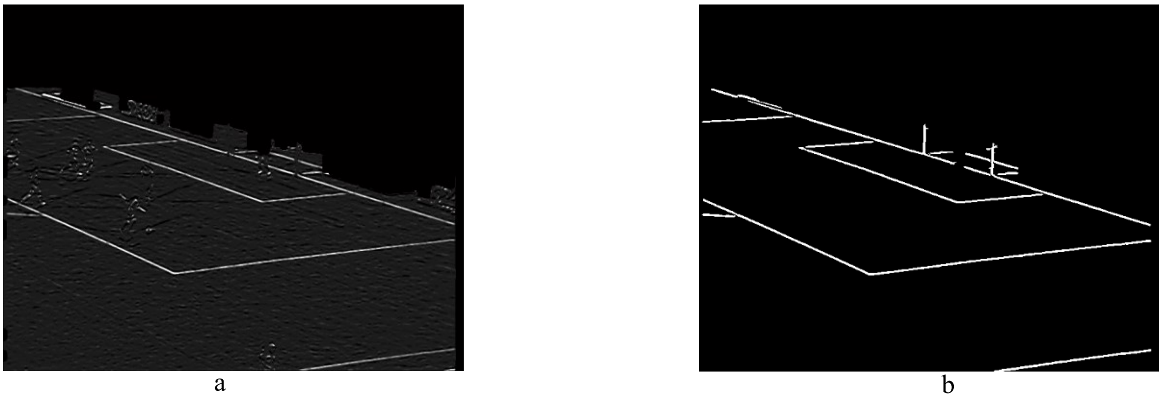


Fig. 4. a) Result of subtracted background related to vertical median filter. b) Final frame line mask.

The final operation for extracting the frame line mask is a combination of each binary category of that line. Fig. 4 shows the result of mentioned operations on a typical input frame. To increase the speed and accuracy of overlaying the frame line mask on the soccer field model, a thinning operation is performed by using the skeleton algorithm.

B. Optimization stage

The goal of this stage is to converge the initial homography to an optimized one. For obtaining the optimized homography, several steps are performed iteratively.

At first, the pixels of the thinned line mask are projected as the points in the field model by using an initial homography. Some samples are selected from the field model for which the distance of each adjacent pair of samples on the lines and arcs are the same. Note that the sampling rate on arcs are more than the lines. Then, the nearest line mask point is determined for each sample of the model. It is possible that several points be mapped to one model sample. In this situation, the closest point is considered as the correct correspondence.

To remove outliers, the RANSAC algorithm is applied to corresponding points. To increase the RANSAC efficiency, obvious outliers like the correspondences which are related to the opposite side are detected and removed. The mean error of homography map is measured based on the distance between corresponding points. The mean error in primary iterations is not accurate and depends on the initial homography. In our tests, after 30-40 iterations the homography error is computed, correctly. If the stop condition is not met, the new homography is calculated based on corresponding points by using the DLT algorithm. These steps are carried out iteratively until the algorithm converges. Several iterations are shown in Fig. 5.

For stop condition criteria, several criteria are investigated. Finally, the robust results are concluded based on: i) the homography error acceptability (we chose 30 cm for maximum mean error), ii) low enough variance of mean error in the last K iterations, iii) low enough slope of the fitted line to mean error in the last K iterations. If all these conditions are met, the homography is converged to the optimized homography, otherwise, the suggested approach has a problem in the nearest point finding step.

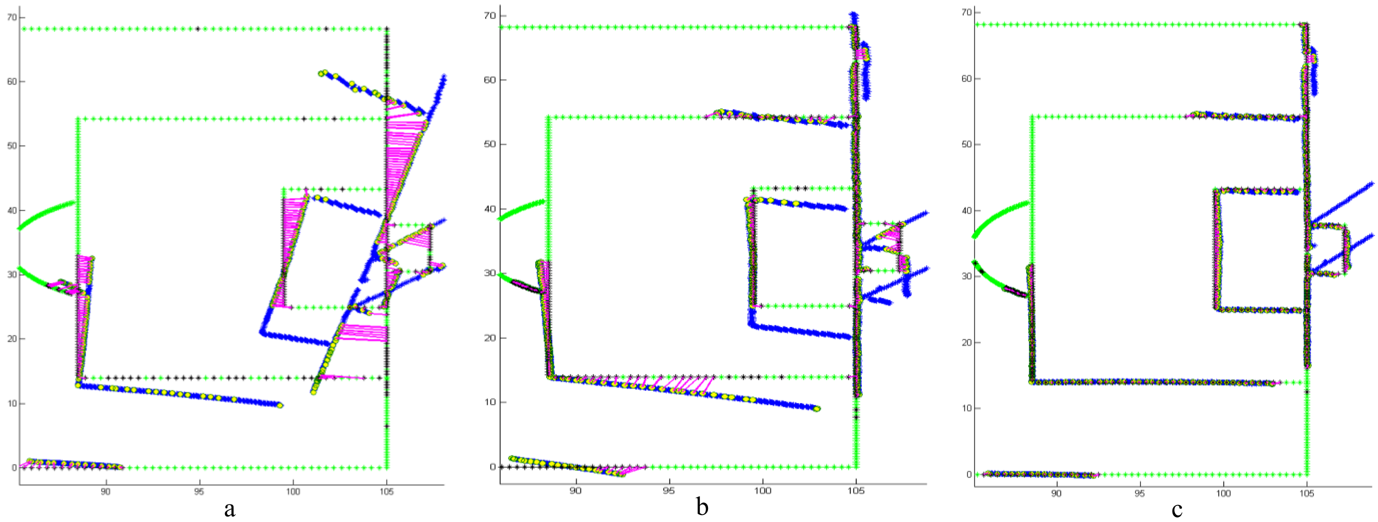


Fig. 5. Effect of optimization steps on initial homography: a) first iteration, b) 61th iteration, c) 200th iteration. [Green points are the field model samples, blue line is the projected line mask, and magenta lines show the distance of valid corresponding points.]

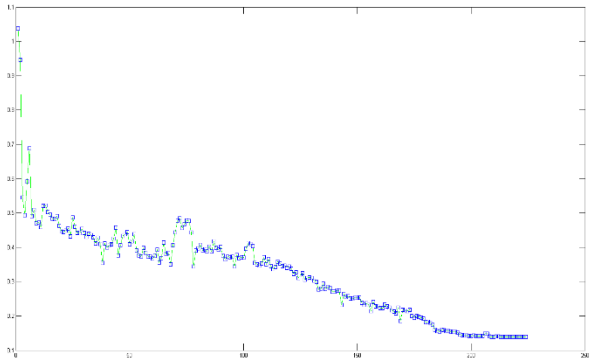


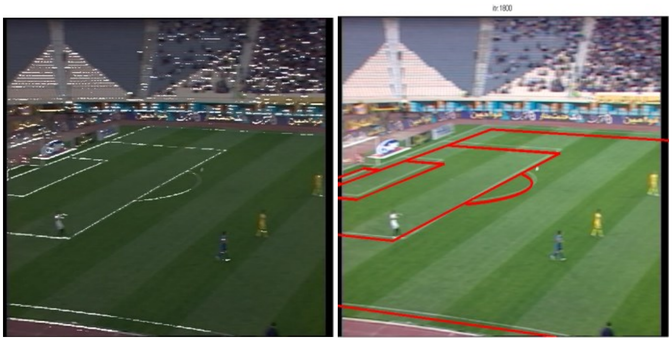
Fig. 6. Behavior of homography mean error for input frame. Mean error for optimized homography is about 13 cm.

In our tests, K is set to 30, the maximum variance of the mean error is considered as 10^{-4}, and the slop of the fitted line to the mean error is considered as 10^{-3}. The homography which is related to the minimum mean error is selected as the solution. Fig. 6 shows the mean error behavior in all iterations.

IV. EXPERIMENTAL RESULTS

To evaluate the proposed method, Azadi dataset that is prepared by *image processing lab* (IPL), at Sharif University of Technology is used. This dataset has several soccer videos in different views. The frame rate of video sequences is 25 frames per second and size of each frame is 720×576. Within this dataset, 3000 sequences are selected for test. The used dataset is very challenging and it contains real problems including noise, low resolution, lighting changes, shadows, lens distortions, and other capturing issues. The evaluation is performed on 75 random sequences that cover almost all different situations in soccer scenes. We could not reach to the optimum solution in frames that do not have any arcs or more than four straight lines. But, in other frames the result of the proposed method was promising (by using not much deviated initial homography).

The proposed method is implemented by MATLAB 8.3.0 (R2014a), using RAM of 8GB DDR3L and CPU of 4700HQ under Windows 8.1. In our tests, the runtime of preprocessing stage is about 460ms for each frame and the optimization stage runtime depends on the initial homography. It takes about 90ms, on average, in each iteration. The obtained results are compared to the state-of-the-art method [1]. We have implemented that method based on our best knowledge on that work. In [1], the *Levenberg-Marquardt* technique is used for homography optimization. For fair comparisons, the extracted line mask is used in both approaches; because Azadi sequences are noisy and suggested line detection in [1] is not suitable for this dataset because the *Levenberg-Marquardt* method is more sensitive to outliers and may lead to local minima. Fig. 7 shows the output of the used line detection in [1] and its optimization homography method for a sample sequence of Azadi dataset. For initial homography of it, we considered an optimized homography which is obtained by applying our proposed method. To evaluate the robustness of our algorithm, the initial homography is computed based on some manual corresponding points.



a b

Fig. 7. a) Line pixels for a motion blurred frame. b) Homography optimization approach introduced in [1]. [An optimized homography is used as an Initial homography.]

Some other obtained results are shown in Fig. 8. The second and third rows are from offside view cameras and other images are from the center viewed camera. The two last rows show very difficult samples that many approaches fail in finding acceptable homography for them. The fifth row shows that if the initial homography is good enough, the introduced approach in [1] can also reach to an optimized homography. In the last row, although the suggested approach in [1] could not reach to an appropriate homography even by using an initial homography map, the proposed method was able to obtain an appropriate homography map.

V. CONCLUSION AND FUTURE WORK

The aim of this paper was to introduce an efficient and robust method to obtain an optimized homography map from an initial homography matrix. We proposed an iterative algorithm to find the optimized homography. A different technique is proposed to extract the frame line mask. The proposed method is compared to the state-of-the-art method to show its superiority. The future work will be related to tracking the optimized homography in consecutive frames.

REFERENCES

[1] D. Farin, S. Krabbe, and W. Effelsberg, "Robust camera calibration for sport videos using court models," In SPIE Storage and Retrieval Methods and Applications for Multimedia, Vol. 5307, pp. 80–91, 2004.

[2] G. Sudhir, J. C. M. Lee, and A. K. Jain, "Automatic classification of tennis video for high-level content-based retrieval," IEEE International Workshop on Content Based Access of Image and Video Databases, p.p 81–90, 1998.

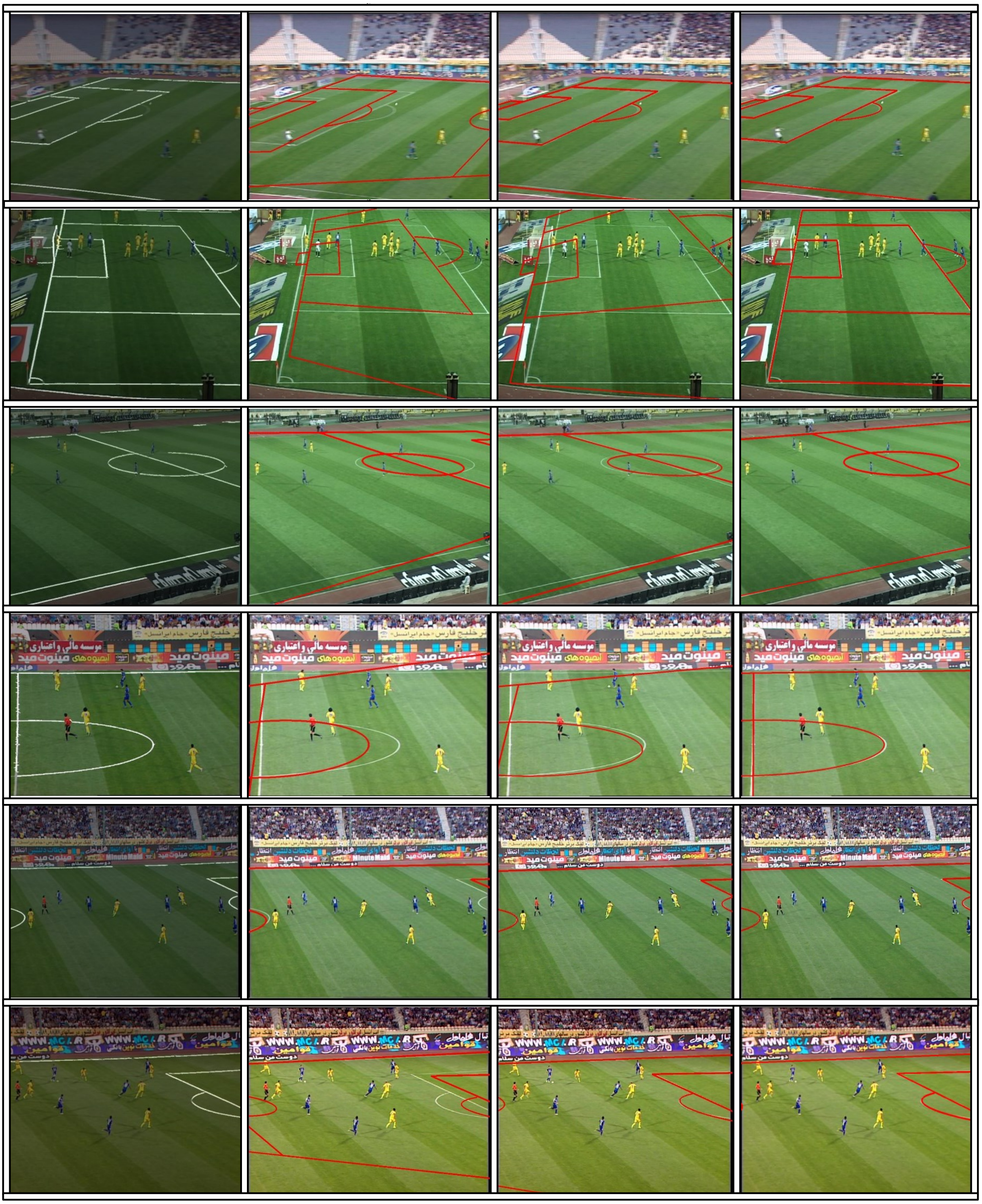
[3] T. battikh and I. Jabri, "Camera calibration using court models for real-time augmenting soccer scenes," In Multimedia Tools and Applications vol. 51, no. 3, pp. 997-1011, 2011.

[4] H. Kim and K. Hong, "Robust image mosaicing of soccer videos using self-calibration and line tracking," In Pattern Analysis & Applications, vol. 4, no. 1, pp. 9–19, 2001.

[5] T. Watanabe, M. Haseyama, and H. Kitajima, "A soccer field tracking method with wire frame model from TV images," IEEE International Conference on Image Processing (ICIP), 2004.

[6] A. L Troter , R. Bulot , J-M. Boï, and J. Sequeira, "Arc of ellipse detection for video image registration," IEEE Workshop on Signal Processing Systems Design and Implementation, 2005.

[7] F. Wang, L. Sun, B. Yang, and S. Yang, "Fast Arc Detection Algorithm for Play Field Registration in Soccer Video Mining," IEEE international Conference on Systems, Man and Cybernetics, vol. 10, pp. 4932–4936, Oct. 2006.



a b c d

Fig. 8. a) Result of line extraction step. b) Initial homography. c) Suggested approach in [1]. d) Proposed method.