Real-time Modeling and Simulation of Tire-Soil Interaction: A 3D Reconstruction Approach

Wei Hao, David L. Page and Mongi A. Abidi

Abstract—The modeling of tire-soil interaction attracts people from different engineering and basic science communities. In this project, we develop a system which can capture the tire-soil interaction using state of art computer vision techniques in real time. Rather than to build a numerical model which simulating the interaction between tires and deformable surfaces, we will build 3D textured model of the niche of tire-soil friction which will demonstrate the interaction between them. Both stereo and motion information will be utilized to generate the 3D model. Efforts will be made to enhance the accuracy of the surface texture and to accelerate the modeling process thus a sequence of 3-D model can be reconstructed and be displayed.

Index Terms—Tire-Soil Interaction, Spatial-temporal Stereo, Motion, Dynamic Scene Analysis, Dynamic Scene Reconstruction.

I. INTRODUCTION

Tire-soil interaction is a fast changing process which are in the time scale of us\(\left(10^{-6}\right)\) second or ms\(\left(10^{-3}\right)\) second thus involves complex physics. From the application perspective, it is very interesting for manufacturers of average vehicles and specialized vehicles such as HMMWV to know the detail of the tire-soil interaction, especially in the cases of off-road traffic.

There are many efforts of building kinetic or numerical models of tire-soil interaction since 1950’s(for instance, [1][2][3] and related publications), but it is not an easy task because the modeling involves the kinetics of soil, and the soil profile, which generates the vibration of the vehicle, is significantly modified by the vehicle itself. Simulation of off-road interaction is made more difficult by the fact that cohesion and friction are not true constants for a given soil.

To the best of our knowledge, the establishment of an accurate numerical tire-soil model for off-road traffic has not been achieved up to now because of the difficulties mention above.

To avoid building numerical physics or kinetic model, researchers have developed methods utilizing sensors like ultra-wide band radar in modeling [4].

In this research, we use state of arts techniques in computer vision to build real-time 4D movie of the tire-soil interaction for off-road cases. The 4D movie is a temporal sequence of 3-D models of the whole processing of the interaction, which can be acquired by using the stereo and motion information collected using stereo video cameras.

The advantage of our methods is obvious. To mention some of them, first, the data acquisition is relatively simpler compare to the cases of using radar or other mechanical sensors. Second, the hardware
design is relatively easier, the budget of the system is lower and the mounting of the system is easier than in other cases. Third, the 3-D reconstruction based on visual stereo information is comparatively matured.

On the other hand, building 3-D models of dynamic scene from stereo and motion clues remains an intensive research interest for decades in the computer vision community[5]. It will be a contribution to the computer vision research if 4-D movies can be generated by utilizing temporal and spatial clues in visual information.

The arrangement of the rest of the paper is as followed, in section II, we present the design of data acquisition system and show some results collected using the system. In section III, a short discussion of the data processing is conducted and related techniques are surveyed. In section IV, we present the algorithms and results of 3-D reconstruction based on stereo clues. In section V, possible enhancements of the proposed methods are discussed.

II. FAST DIGITIZING OF THE TIRE-SOIL INTERACTION

Using high-speed video cameras, the tire-soil interaction can be captured. The system diagram is as followed,

![Diagram of data acquisition system](image)

Fig. 2. The pipeline of data acquisition system, which can digitize the fast changing process of tire-soil interaction.

The cameras are allocated according to some sensor placement scenario. A tentative sensor placement is as shown in Fig. 3. The advantage of this setup is that the software of 3-D reconstruction can be made easier under this condition.

![Tentative sensor placement scenario](image)

(a) Tentative sensor placement scenario.

![Close view of the stereo camera system](image)

(b) Close view of the stereo camera system.

Fig. 3. The sensor placement scenario.

Here are some sample frames collected. The whole video sequences can be retrieved from the web[6].

![Sample frame from the left camera on soft soil](image)

(a) Sample frame from the left camera on soft soil.

![Sample frame from the right camera on soft soil](image)

(b) Sample frame from the right camera on soft soil.
Fig. 4. Sample frame from the video capture for tire running on soft soil.

Sample frames on grassy soil captured by the stereo video cameras are shown in Fig. 5.

Fig. 5. Sample frame from the video capture for tire running on soft soil.

Sample frames on road captured by the stereo video cameras are shown in Fig. 6.

Fig. 6. Sample frame from the video capture for tire running on soft soil.

Sample frames on gravel road captured by the stereo video cameras are shown in Fig. 7.

Fig. 7. Sample frame from the video capture for tire running on gravel road.

We noticed that the frame speed of this setup is at 30 frames/second and may be not enough for capturing the dynamic changes take place during the interaction. We propose adoption of higher speed cameras.

An important problem which may be very crucial to the data collection is the synchronization of the stereo cameras. We found many solutions to synchronization units on the internet, such as the module shown in Fig. 8[7].
As shown in Fig. 8, the Pokecope 3D camera remote allows taking 3D photos using two digital cameras. The remote can synchronize image capture to much less than 1 millisecond. This allows capturing action and moving subjects in 3D.

The 3D camera remote uses the power-up of each camera to synchronize the timing signals of each camera. The lag between the timing signals of the two cameras is displayed on the LCD display. Typically, the cameras will be synchronized to within about 0.2 milliseconds.

Over several minutes, the timing signals may drift either toward better or worse synchronization. This allows synchronization as good as about 0.01 milliseconds if extreme synchronization is desired.

III. 3D RECONSTRUCTION PIPELINE OF STATIC TIRE-SOIL SCENE

Based on the matured techniques of stereo 3D reconstruction, we propose the following pipeline for the 3D reconstruction for the static scene. As the first step, software of stereo vision will be applied on the static stereo images. The performance will be evaluated and possible enhancement will be made to integrate the stereo and motion information together.

The camera will be calibrated before hand and thus the camera can be modeled as direction sensor, i.e., after calibration, we can decide the line on which the corresponding 3D scene point of some 2D point on the image. It is easy to locate the real co-ordinates of the point if more than two cameras are used as direction sensors and the two cameras are placed appropriately.
The crucial steps of the reconstruction are camera calibration and stereo matching based on the discussion above and Fig. 10.

To simplify the stereo matching, we adopt the step of image rectification before the stereo matching.

We adopt dense stereo matching for this research and the algorithms in [8] are adopted and the results are compared.

After the camera model is set up by calibration and the stereo matching is established, we can calculate the depth of the object by triangulation with easy.

Textured rending can be applied after the 3D information is recovered.

Textured rending can be applied after the 3D information is recovered.

IV. STEREO CAMERAS CALIBRATION

As shown in the following diagram, the camera model can be built,

The four coordinate systems in Fig. 11 are as followed:

1. The three dimensional coordinate system whose origin is at the center of projection and whose Z axis is along the optical axis, as shown in figure 1. This coordinate system is called the standard coordinate system of the camera.

2. The coordinate system whose origin is at the intersection of the optical axis and the image plane, and whose x and y axes are parallel to the X and Y axes.

3. The actual pixel coordinate system.

4. The three dimensional world coordinates of a point will not be specified in a frame whose origin is at the center of projection and whose Z axis lies along the optical axis. Some other, more convenient frame will more likely be specified, and then we have to include a change of coordinates from this other frame to the standard coordinate system.

For \( M = (X,Y,Z) \) in system 1 and \( m = (x,y) \) in system 2, it is easy to show that

\[
\begin{bmatrix}
    x \\
    s \\
    y \\
    1
\end{bmatrix} =
\begin{bmatrix}
    f & 0 & 0 & 0 \\
    0 & f & 0 & 0 \\
    0 & 0 & 1 & 0 \\
    0 & 0 & 0 & 1
\end{bmatrix}
\begin{bmatrix}
    X \\
    Y \\
    Z \\
    1
\end{bmatrix}
\]

or

\[
s\tilde{m} =
\begin{bmatrix}
    f & 0 & 0 & 0 \\
    0 & f & 0 & 0 \\
    0 & 0 & 1 & 0
\end{bmatrix}
\begin{bmatrix}
    \tilde{M}
\end{bmatrix}
\]

(1)

\( m(x,y) \) in system 2 will have coordinates (u,v) in system 3, and

\[
\begin{bmatrix}
    u \\
    v \\
    1
\end{bmatrix} =
\begin{bmatrix}
    f & 0 & 0 & 0 \\
    0 & f & 0 & 0 \\
    0 & 0 & pixelwidth & 0 \\
    0 & 0 & pixelheight & 0
\end{bmatrix}
\begin{bmatrix}
    X \\
    Y \\
    Z \\
    1
\end{bmatrix}
\]

or

\[
s\tilde{U} = P\tilde{M}
\]

(2)

Where,

\( \tilde{M} \) is a point in the standard coordinate system, i.e., system 1.
P is the perspective projection matrix.
\( \tilde{U} \) represents the homogeneous coordinates in image pixel coordinates system, i.e. system 3.

This is the relationship of scene point in system 1 and its image point in system 3. Let \( \alpha_u = f / \text{pixelwidth} \), \( \alpha_v = f / \text{pixellength} \).

\( P \) is determined i.f.f \( \alpha_u, \alpha_v, u, v \) is determined. \( \alpha_u, \alpha_v, u, v \) do not depend on the position and orientation of the camera in space, and are called the intrinsic parameters.

In general case, the 3D coordinate system is not necessarily the standard one, but with a rotation and translation.

Finally we can establish the relationship between scene point \( \tilde{M} \) in system 4 and its image points \( \tilde{U} \) in the image pixel coordinates system.

Finally we will have
\[ s\tilde{U} = P \cdot K \cdot \tilde{M} \] (3)

Where \( K \) is a 4*4 matrix like the following,
\[ K = \begin{bmatrix} R & t \\ 0^T & 1 \end{bmatrix} \] (4)

The top 3*3 corner is a rotation matrix \( R \) and encodes the camera orientation with respect to a given world frame; the final column is a homogeneous vector \( t \) capturing the camera displacement from the world frame origin.

The matrix \( K \) has six degrees of freedom, three for the orientation, and three for the translation of the camera. These parameters are known as the extrinsic camera parameters.

Equation (3) can be rewritten as
\[ s\tilde{U} = P \cdot K \cdot \tilde{M} = CR[I \ | \ R^T] \tilde{M} \] (3')

This equation defines the 3*4 projection matrix from system 4 to system 3.

\( (3') \) can also be re-written as
\[ s\tilde{U} = C[R \ | \ t] \tilde{M} = CR[I \ | \ R^T] \tilde{M} \] (3'')

Camera calibration is estimating the intrinsic and extrinsic parameters of a camera. It is a necessary step in 3D reconstruction in order to extract metric information from 2D images. Much work has been done, starting in the photogrammetry community, and more recently in computer vision. We can classify those techniques roughly into two categories: photogrammetric calibration and self calibration.

**Photogrammetric calibration.** Camera calibration is performed by observing a calibration object whose geometry in 3-D space is known with very good precision. Calibration can be done very efficiently. The calibration object usually consists of two or three planes orthogonal to each other. Sometimes, a plane undergoing a precisely known translation is also used. These approaches require an expensive calibration apparatus, and an elaborate setup.

**Self-calibration.** Techniques in this category do not use any calibration object. Just by moving a camera in a static scene, the rigidity of the scene provides in general two constraints on the cameras’ internal parameters from one camera displacement by using image information alone.

In this research, we fulfill camera calibration using the method of Zhang’s planar pattern [9]. The calibration pipeline is as follows:
1. Print a pattern and attach it to a planar surface;
2. Take a few images of the model plane under different orientations by moving either the plane or the camera;
3. Detect the feature points in the images;
4. Estimate the five intrinsic parameters and all the extrinsic parameters using the closed-form solution as described in Sect. 3.1;
5. Estimate the coefficients of the radial distortion by solving the linear least-squares;
6. Refine all parameters by minimizing certain criterion.
The planar pattern we are using is a chessboard printed using laser printer. Each square in the chessboard are of the same size of 2mm*2mm.

Fig. 12. The planar pattern used for image calibration.

We take a sequence of pictures (16 frames included) of the planar pattern with different orientations. Sample image frames are shown as followed,

Fig. 13. Frame 1, 8, 14, 16 of the different pose of the planar calibration rig of the left camera.

The corners on the chessboard are used as feature points for calibration. To enhance the accuracy, the lines in the chessboard are firstly estimated and after that, the corners are estimated as the intersection of the lines.

Fig. 14. Sample results of corner detection in the enclosed square region on the planar pattern. The corners detected are used as the feature points for calibration.

Using the methods described in [9], we can estimate the intrinsic parameters of the left and right camera individually, the results of one estimation is as followed,

### Left camera:

- **Focal Length:** \( fc = [1147.60247, 1146.50820] \pm [6.25846, 6.43503] \)
- **Principal point:** \( cc = [511.49949, 397.01074] \pm [8.32481, 9.21247] \)
- **Skew:** \( \alpha_c = [0.00000] \pm [0.00000] \), angle of pixel axes = 90.00000 ± 0.0 degrees
- **Distortion:** \( kc = [-0.10625, 0.04799, -0.00115, 0.00222, 0.00000] \pm [0.02546, 0.11175, 0.00232, 0.00197, 0.00000] \)
- **Pixel error:** \( err = [0.08207, 0.09215] \)

### Right camera:

- **Focal Length:** \( fc = [1149.06986, 1147.29876] \pm [6.78215, 7.01784] \)
- **Principal point:** \( cc = [518.20847, 394.23652] \pm [9.82764, 10.29952] \)
- **Skew:** \( \alpha_c = [0.00000] \pm [0.00000] \), angle of pixel axes = 90.00000 ± 0.0 degrees
- **Distortion:** \( kc = [-0.14847, 0.33853, 0.00076, 0.00110, 0.00000] \pm [0.03173, 0.26917, 0.00281, 0.00213, 0.00000] \)
- **Pixel error:** \( err = [0.09145, 0.10057] \)

Table. 1. The intrinsic camera parameters of the stereo cameras estimated from one test.

From the above results, it is easily to notice that the intrinsic parameters of the two cameras are very similar, which makes sense because we are using two canon powershot A80 CCD cameras for data collection.

We also estimate the relative translation and rotation between the two cameras.

### Extrinsic parameters (position of right camera wrt left camera):

- **Rotation matrix:** \( R = [1.0000, 0.0039, -0.00640, -0.0039, 1.0000, -0.0022, 0.0064, 0.0022, 1.0000] \)

- **Translation vector:** \( T = [104.57154, -0.27825, 1.81454] \)

Table. 2. The relative translation and rotation of the right camera w.r.t the left camera.
V. IMAGE RECTIFICATION

The algorithm we applied here is based on [10]. Some results are shown below. The input frames are the 10, 12, 13, 14 frames from an image sequence of the stereo rig, as shown in Fig. 16.

![Image Rectification](image1.png)

Fig. 15. Image Rectification.

The results of rectification are shown as followed.

![Rectified Image Frame 10, 12, 13, 14](image2.png)

Fig. 17. Rectified Image Frame 10, 12, 13, 14, which are correspondent to the frames shown in Fig. 16.

VI. DENSE MATCHING

Dense matching is very important step for 3D reconstruction and sometimes very difficult. [11] is very useful by providing a taxonomy of various dense matching algorithms. Here is a table of dense matching algorithms we are investigating.

<table>
<thead>
<tr>
<th>Local Matching Methods</th>
<th>Global Methods</th>
</tr>
</thead>
<tbody>
<tr>
<td>- Block Matching;</td>
<td>- Dynamic Programming;</td>
</tr>
<tr>
<td>- Gradient-based Optimization</td>
<td>- Intrinsic Curves</td>
</tr>
<tr>
<td>- Feature Matching</td>
<td>- Graph Cuts</td>
</tr>
<tr>
<td></td>
<td>- Nonlinear Diffusion</td>
</tr>
<tr>
<td></td>
<td>- Belief Propagation</td>
</tr>
<tr>
<td></td>
<td>- Correspondenceless Methods</td>
</tr>
</tbody>
</table>

Table 3. Dense matching algorithms surveyed.

Here are inputs of the experimental algorithm.

![Input frames of a static tire](image3.png)

Fig. 18. Input frames of a static tire.
The results of rectification are shown as followed.

![Image of rectified input frames of a static tire.]

After rectification, the searching of the correspondence is much easier. We only need to consider small stripe of certain epipolar line.

We applied simple SAD method and SSD+ DP method to generate the disparity map as below.

![Image of disparity map generated by SAD and SSD+DP method.]

The result can still be improved by tuning the parameters (like searching window size) and taking into account the constraint that can be consider for the specific scenes and objects.

VII. THE RESULTS OF STATIC TIRE RECONSTRUCTION

The triangulation and visualization of the image will be done soon.

VIII. CONCLUSION AND FUTURE WORK

A data collection system was built to digitize the tire-soil interaction; the 3D reconstruction software based on static stereo is implemented and visualization will be done soon.

Much progress have been achieved in stereo and multiple vision geometry techniques, but not much have been done on combining spatial and temporal clues of stereo image sequences.

Our possible contribution may be:

- Integrate motion and stereo information so that the 3D reconstruction of scene at certain time can be enhanced by introducing the time regularity.
- Use motion evolution information to reduce the amount of work that is necessary for building models for long image sequence. For instance, only the deformation or moving of the correspondent pixels at all time should be considered.
- Apply robust statistics tools such as robust kalman filter to enhance the prediction of deformation or moving along the time dimension.

REFERENCES