# A Factorization Method for Structure from Planar Motion 

Jian Li and Rama Chellappa *<br>Center for Automation Research (CfAR) and Department of Electrical and Computer Engineering<br>University of Maryland, College Park, MD 20742<br>$\{$ lij, rama $\} @ c f a r . u m d . e d u$


#### Abstract

We propose a factorization method for structure from planar motion using a stationary perspective camera. Compared with [8] for general motion, our work has three major differences: a different measurement matrix specialized for planar motion is formed. The measurement matrix has a rank of at most 3 , instead of 4; the measurement matrix needs similar scalings, but estimation of fundamental matrices or epipoles is not needed; we have an Euclidean reconstruction instead of a projective reconstruction. The camera is not required to be calibrated. A simple semiautomatic calibration method using vanishing points and lines is sufficient. Experimental results show that the algorithm is accurate and fairly robust to noise and inaccurate calibration.


## 1. Introduction

Since it was first introduced in [11], the factorization approach has been very popular for solving structure from motion (SfM) problems. It stands out among other batch processing methods because the reconstruction can be easily carried out through singular value decomposition (SVD). Generally, the observation matrix is factorized as a bilinear product of motion and shape matrices. SVD is then used to find the factorized components, at the same time denoising the data.

Factorization approach for SfM has been studied extensively in the last decade [4]. The essence of factorization approach lies in finding the right rank constraint, which corresponds to the lowest rank among all the factor matrices. The main goal is to find a lower rank condition by exploiting the special structures or constraints. Rank constraints have been found for different camera models, or different types of objects such as planar objects, rigid or non-rigid objects. Another branch of study has focused on the rank constraints for certain kinds of motion, such as linear and planar motions.

[^0] 2004H80200000.

The study of rank constraints for certain kind of motion can be found in [2] [6] [7] [12] etc. In [2], Han et al. study the case of linear motion with constant speed for multiple objects, and develop a method of scene reconstruction through factorization. In [7], Quan et al. develop a method for decomposing a 2D image into two 1D images, so that the structure can be reconstructed from images captured under constrained planar motion, but they restrict the image plane to be perpendicular to the motion plane. Both of these methods have very restrictive assumptions. In [6] and [12], matrices formed by displacements of feature points are shown to have a lower rank condition with certain approximations under constrained motion. Then iterative estimate of the motion and depths can be carried out. Their methods can be used to recover the motion plane, but lack the flavor of a direct factorization without iteration. We propose a generic factorization method for structure from planar motion under perspective projection. A measurement matrix specialized for planar motion is formed in order to lower the rank condition. We exploit the constraint of planar motion to find a simple scaling method instead of using the fundamental matrix and epipoles as in [8], where the authors propose a factorization approach for the perspective camera.

Study of planar motion is motivated by applications such as parking lot surveillance, traffic scene analysis etc. Planar motion is arguably the most dominant type of motion in surveillance videos, simply because most objects move on the typically planar ground plane. In this paper, we focus on a monocular sequence captured by a stationary perspective camera, in which a rigid object moves on the ground plane.

Many methods for structure from planar motion have been proposed such as [6][10][12]. Many of them need nonlinear optimizations or iterative schemes and sometimes cannot guarantee a global optimal solution. Our method only requires SVD and linear operations with no iterations. Our formulation is similar to [10], where the authors consider the reconstruction of perspective depths from correspondences across two frames. They find quadratic equations of depths by using two facts: corresponding feature points lie on a plane parallel to the ground plane, and the

3D displacement between them is constant for every pair of feature points because of the rigidity constraints. But since they only consider a pair of frames and there is no denoising process, the method is very sensitive to noise. Our method makes use of multiple frames and takes the factorization approach, and therefore, is more robust to noise and incorrect calibration.

In our method, the camera does not have to be calibrated. For an uncalibrated sequence, our method requires the estimation of the focal length and ground plane constraints (GPC), so that the motion plane is known. In our experiment, we use vanishing points and line [3][5][13] to find the GPC.

The rest of the paper is organized as follows. Section 2 first gives a derivation and detailed description of our method, followed by comparisons with the factorization method for general motion. Section 3 explains the necessary preprocessing steps for an uncalibrated sequence, then gives both quantitative and qualitative analysis on the experimental results. Section 4 concludes the paper.

## 2. Problem formulation

In this section, we derive the factorization method for structure from planar motion and present the detailed algorithm. Then we compare our method with the factorization method for general motion under perspective projection.

In our work, the camera is not necessarily calibrated. The calibration and ground plane constraints can be estimated through a number of ways. In the next section, we show a simple semi-automatic calibration scheme using vanishing points and line. Once the calibration and GPC is estimated, a matrix is constructed from the observations, which after properly scaling has a rank of at most 3 .

### 2.1. Derivation of our method

Consider the selection of CCS and WCS in Figure 1, where the $x-y$ plane in WCS lies on the ground plane. In this paper, we focus on the typical case that the image plane is not parallel to the ground plane.


Figure 1: Selection of world coordinate and camera coordinate systems. O is the camera center.

In CCS, denote the direction of $\mathrm{x}-\mathrm{y}$ - and z -axis of WCS as $g_{x}, g_{y}$ and $g_{z}$ respectively, and the origin of WCS as $g_{0}$. $f$ is the focus. As mentioned above, all those vectors can be estimated.

Suppose a point $p$ is the image of a point $P$ in the 3D space. Its representation is $p_{c}=(x, y, f)^{T}$ in the CCS. Then $P_{c}=\lambda(x, y, f)^{T}$, where $\lambda$ is the perspective depth. Its representation in WCS is,

$$
P_{w}=\left[\begin{array}{c}
g_{x}^{T}  \tag{1}\\
g_{y}^{T} \\
g_{z}^{T}
\end{array}\right]\left(\lambda\left[\begin{array}{l}
x \\
y \\
f
\end{array}\right]-g_{0}\right)
$$

According to our assumption, the only unknown in the above equation is $\lambda$. Here are some key observations about the factorization method. If the feature points are from a rigid object moving on the ground plane, then each of them will move on a plane parallel to the ground plane, and hence its z-coordinates in WCS will remain constant across all frames. Because of rigid motion, its $x$ - and $y$-coordinates in WCS will take a 2D Euclidean Transform. Note in [10], similar constraints are used, but they only consider two frames and the algorithm is more sensitive to noise.

Suppose we have an image sequence from a stationary perspective camera, in which a rigid object moves on the ground. N visible image points on the object have been tracked over M frames. We will use factorization to estimate the structure of the vehicle.

Denote $\lambda_{i t}$ the perspective depth of point $i$ in frame $t$. And from Eq. 1, all points should satisfy
$P_{w}^{t i}=\left[\begin{array}{c}g_{x}^{T} \\ g_{y}^{T} \\ g_{z}^{T}\end{array}\right]\left(\lambda_{t i}\left[\begin{array}{c}x_{t i} \\ y_{t i} \\ f\end{array}\right]-g_{0}\right) \equiv \lambda_{t i}\left[\begin{array}{c}u_{t i} \\ v_{t i} \\ w_{t i}\end{array}\right]-\mathbf{s}$
$\mathbf{s}$ should be the same for all points across all frames. Then if we form a matrix from the $x$ and $y$ components of $P_{w}^{t i}$, and using the fact that they take a 2D Euclidean transform, we will have

$$
\begin{aligned}
\boldsymbol{W} & \equiv\left[\begin{array}{cccc}
\lambda_{11} u_{11} & \lambda_{12} u_{12} & \ldots & \lambda_{1 N} u_{1 N} \\
\lambda_{11} v_{11} & \lambda_{12} v_{12} & \ldots & \lambda_{1 N} v_{1 N} \\
\ldots & \ldots & & \ldots \\
\ldots & \ldots & & \ldots \\
\lambda_{M 1} u_{M 1} & \lambda_{M 2} u_{M 2} & \ldots & \lambda_{M N} u_{M N} \\
\lambda_{M 1} v_{M 1} & \lambda_{M 2} v_{M 2} & \ldots & \lambda_{M N} v_{M N}
\end{array}\right] \\
& =\left[\begin{array}{ccc}
r_{11}^{1} & r_{12}^{1} & t_{x}^{1} \\
r_{21}^{1} & r_{22}^{1} & t_{y}^{1} \\
\ldots & & \\
r_{11}^{M} & r_{12}^{M} & t_{x}^{M} \\
r_{21}^{M} & r_{22}^{M} & t_{y}^{M}
\end{array}\right]\left[\begin{array}{cccc}
x_{1} & x_{2} & \ldots & x_{N} \\
y_{1} & y_{2} & \ldots & y_{N} \\
1 & 1 & \ldots & 1
\end{array}\right]
\end{aligned}
$$

where $\boldsymbol{W}$ is the rescaled observation matrix. $\quad \boldsymbol{R}_{t}=$ $\left[\begin{array}{ll}r_{11}^{t} & r_{12}^{t} \\ r_{21}^{t} & r_{22}^{t}\end{array}\right]$ is an orthogonal matrix and corresponds to
the rotation matrix for frame $t$. [ $\left.\begin{array}{cc}t_{x}^{k} & t_{y}^{k}\end{array}\right]^{T}$ corresponds to the translation vector for frame $k .\left[\begin{array}{ll}x_{i} & y_{i}\end{array}\right]^{T}$ is the $\mathrm{x}-\mathrm{y}$ coordinate of point $i$, but because of the ambiguity in selecting the reference coordinates for motion and structure, they are not necessarily defined in WCS. Also note that the contribution from $\mathbf{s}$ is absorbed into $\left[\begin{array}{cc}t_{x}^{k} & t_{y}^{k}\end{array}\right]^{T}$.

Therefore the rescaled matrix $\boldsymbol{W}$ will have a rank of at most 3. To find the scale, we use the fact that the $z$ component of each feature point in WCS remain constant across all frames, i.e. $z\left(P_{w}^{t i}\right)=z\left(P_{w}^{s i}\right)$ for $s=1, \ldots, M$. Then we can obtain $\lambda_{t i} w_{t i}=\lambda_{s i} w_{s i}$. Thus the ratio of $\lambda \mathrm{s}$ can be recovered along each column in $\boldsymbol{W}$.

Thus using the equality, we can set $\lambda_{t i} w_{t i}=c_{i}$, for $t=1, \ldots, M$, and $i=1, \ldots, N$. Typically $w_{t i} \neq 0$ and $c_{i} \neq 0$ unless the feature point lies on the vanishing line of the ground plane. So we substitute $\lambda_{t i}$ and move all the unknowns to the right hand side, which will produce,

$$
\begin{align*}
\tilde{\boldsymbol{W}} & \equiv\left[\begin{array}{cccc}
u_{11} / w_{11} & u_{12} / w_{12} & \ldots & u_{1 N} / w_{1 N} \\
v_{11} / w_{11} & v_{12} / w_{12} & \ldots & v_{1 N} / w_{1 N} \\
\ldots & \ldots & & \ldots \\
u_{M 1} / w_{M 1} & u_{M 2} / w_{M 2} & \ldots & u_{M N} / w_{M N} \\
v_{M 1} / w_{M 1} & v_{M 2} / w_{M 2} & \ldots & v_{M N} / w_{M N}
\end{array}\right]  \tag{3}\\
& =\left[\begin{array}{cccc}
r_{11}^{1} & r_{12}^{1} & t_{x}^{1} \\
r_{21}^{1} & r_{22}^{1} & t_{y}^{1} \\
\ldots \\
r_{11}^{M} & r_{12}^{M} & t_{x}^{M} \\
r_{21}^{M} & r_{22}^{M} & t_{y}^{M}
\end{array}\right]\left[\begin{array}{cccc}
x_{1} / c_{1} & x_{2} / c_{2} & \ldots & x_{N} / c_{N} \\
y_{1} / c_{1} & y_{2} / c_{2} & \ldots & y_{N} / c_{N} \\
1 / c_{1} & 1 / c_{2} & \ldots & 1 / c_{N}
\end{array}\right] \\
& \equiv \tilde{\boldsymbol{M} \tilde{\boldsymbol{S}}} \tag{4}
\end{align*}
$$

The matrix $\tilde{\boldsymbol{W}}$ can be directly calculated from the observations, calibration parameters and GPC. It is a bilinear product of motion and shape matrices on the $x-y$ plane in WCS. Its rank is at most 3 .

Once we have the rank constraints, similar procedure for other factorization methods can be applied. SVD can be applied to find the rank 3 matrix which is closest to $\tilde{\boldsymbol{W}}$ in the sense of Frobenius norm.

Because $\tilde{\boldsymbol{W}}=\tilde{\boldsymbol{M}} \tilde{\boldsymbol{S}}=(\tilde{\boldsymbol{M}} \boldsymbol{T})\left(\boldsymbol{T}^{-1} \tilde{\boldsymbol{S}}\right)$ for any $3 \times 3$ nonsingular matrix $\boldsymbol{T}$, we need additional constraints to eliminate the ambiguity. Rewrite $\boldsymbol{T}=\left[\boldsymbol{T}_{1} \boldsymbol{T}_{2}\right]$, where $\boldsymbol{T}_{1}$ and $\boldsymbol{T}_{2}$ are $3 \times 2$ and $3 \times 1$ matrix respectively. Suppose the correct motion matrix $\boldsymbol{M}=\tilde{\boldsymbol{M}} \boldsymbol{T}$, then

$$
\tilde{\boldsymbol{M}} \boldsymbol{T}_{1}=\left[\begin{array}{c}
\boldsymbol{R}_{1}  \tag{5}\\
\boldsymbol{R}_{2} \\
\ldots \\
\boldsymbol{R}_{M}
\end{array}\right]
$$

Using the orthogonal property of $\boldsymbol{R}_{t}$, certain elements along the tri-diagonals in $\left(\tilde{\boldsymbol{M}} \boldsymbol{T}_{1}\right)\left(\tilde{\boldsymbol{M}} \boldsymbol{T}_{1}\right)^{T}=\tilde{\boldsymbol{M}}\left(\boldsymbol{T}_{1} \boldsymbol{T}_{1}^{T}\right) \tilde{\boldsymbol{M}}^{T}$ are known. So the linear least square estimate (LLSE) can be
used to recover $\boldsymbol{Q} \equiv \boldsymbol{T}_{1} \boldsymbol{T}_{1}^{T}$. Then $\boldsymbol{T}_{1}$ can be estimated through SVD of $Q$ and it is unique up to a $2 \times 2$ rotational matrix. That ambiguity comes from the selection of the reference coordinate system. If we select the first frame as the reference frame, namely $\boldsymbol{R}_{1}=\boldsymbol{I}_{2 \times 2}$ and $t_{x}^{1}=t_{y}^{1}=0$, then $\boldsymbol{T}_{1}$ can be uniquely recovered and $\boldsymbol{T}_{2}$ can be recovered up to a scale. $\boldsymbol{T}_{2}$ can be solved using the known elements in $\tilde{\boldsymbol{M}} \boldsymbol{T}_{2}$. The scale can be arbitrarily chosen and that is an inherent ambiguity when a monocular sequence is used.

After $\boldsymbol{T}$ is found, the desired shape matrix is $\boldsymbol{S}=\boldsymbol{T}^{-1} \tilde{\boldsymbol{S}}$. Using the the last row of $\boldsymbol{S}$, we can easily find the perspective depth of each feature points, thus the 3D structure can be recovered. It is only recovered up to a scale because of ambiguity in $\boldsymbol{T}_{2}$.

The algorithm can be summarized as follows,

## Algorithm:

a. Calculate the matrix $\tilde{\boldsymbol{W}}$ from Eq. (2) and (3), using the observations, estimated calibration and GPC.
b. Use SVD to get $\tilde{\boldsymbol{W}}=\boldsymbol{U} \operatorname{diag}\left(\sigma_{1}, \sigma_{2}, \ldots, \sigma_{s}\right) \boldsymbol{V}^{T}$, where $s=\min (2 M, N)$. Use the 3 largest singular values and the corresponding singular vectors to get an initial estimate of motion and shape matrices, i.e., set $\tilde{\boldsymbol{M}}=\boldsymbol{W}_{2 M \times 3} \operatorname{diag}\left(\sigma_{1}^{0.5}, \sigma_{3}^{0.5}, \sigma_{3}^{0.5}\right)$ and $\tilde{\boldsymbol{S}}=$ $\operatorname{diag}\left(\sigma_{1}^{0.5}, \sigma_{3}^{0.5}, \sigma_{3}^{0.5}\right) V_{N \times 3}$.
c. Find $\boldsymbol{T}=\left[\boldsymbol{T}_{1} \boldsymbol{T}_{2}\right]$ and eliminate the ambiguity. First find LLSE of $\boldsymbol{Q} \equiv \boldsymbol{T}_{1} \boldsymbol{T}_{1}^{T}$ from known elements in $\tilde{\boldsymbol{M}}\left(\boldsymbol{T}_{1} \boldsymbol{T}_{1}^{T}\right) \tilde{\boldsymbol{M}}^{T}$ according to Eq. (5). Set the first frame as the reference frame, then use $\boldsymbol{R}_{1}=\boldsymbol{I}_{2 x 2}$ and $t_{x}^{1}=t_{y}^{1}=0$ to further remove the ambiguity in $\boldsymbol{T}_{1}$ and find $\boldsymbol{T}_{2}$ up to a scale. Set an arbitrary scale for $\boldsymbol{T}_{2}$.
d. Reconstructed shape and motion matrices are $\boldsymbol{S}=$ $\boldsymbol{T}^{-1} \tilde{\boldsymbol{S}}$ and $\boldsymbol{M}=\tilde{\boldsymbol{M}} \boldsymbol{T}$ respectively. Then $1 / s_{3 i}=\lambda_{t i} w_{t i}$ are used to find $\lambda_{t i}$, the perspective depth, and hence the structure is reconstructed.

### 2.2. Comparisons with the factorization method for general motion

Factorization method is generalized for perspective projection in [8]. The authors form a measurement matrix of rank at most 4 . The key factor in factorization under perspective projection lies in the scaling of the measurement matrix according to the perspective depths. In [8], the authors use epipolar constraints to recover the ratio of the scalings. However, the estimation of the fundamental matrix and epipoles is not an easy job.

The general method can be used to handle the special case of planar motion. However, we find a simpler and more efficient formulation which is tailored to deal with the constrained case. In summary, our method is different from the general method in the following ways,

1. We form the measurement matrix using only the $x$ $y$ coordinates in WCS instead of CCS. The measure-
ment matrix is shown to come from points taking a 2D Euclidean transform, which has a rank of at most 3. In [8], the homogeneous coordinates of projections scaled by the depths are gathered together to form a matrix of rank at most 4 . So the rank condition is reduced for planar motion.
2. We use the property that the z-component for each feature point is constant in WCS to find the right scaling, which is a lot easier than estimating the fundamental matrix and epipoles. However, we do need one-time calibration and estimation of the motion plane, which can be done fairly easily using vanishing points and lines.
3. We have an Euclidean reconstruction instead of a projective reconstruction, because more ambiguities are removed in our formulation. The structure can be recovered up to a scale, while in [8], it is recovered up to a $4 x 4$ non-singular matrix.

## 3. Experiments

In this section, we first explain the simple calibration method we use in our experiment, then we do quantitative and qualitative analysis of our method by applying it to some real and synthetic sequences.

### 3.1. Calibration through vanishing points and line

Many methods are available for automatic or semiautomatic calibration and recovery of the ground plane constraints (GPC). We use vanishing points and lines [3][5][13][14]. We make use of parallel and perpendicular lines, which are very often seen in man-made environments.

In perspective camera model, the images of parallel lines typically will intersect at one point, which is the vanishing point corresponding to those lines. Geometrically, it is the intersection of the image plane with a ray passing through the camera center and parallel to those lines. Algebraically, the direction of the ray $\mathbf{d}$ and the vanishing point $\mathbf{v}$ will be related by $\mathbf{v}=\boldsymbol{K} \mathbf{d}$ [3], where $\boldsymbol{K}$ is the calibration matrix.

The angle between two such rays is,

$$
\begin{equation*}
\cos \theta=\frac{\mathbf{v}_{1}^{T}\left(\boldsymbol{K}^{-T} \boldsymbol{K}^{-1}\right) \mathbf{v}_{2}}{\sqrt{\mathbf{v}_{1}^{T}\left(\boldsymbol{K}^{-T} \boldsymbol{K}^{-1}\right) \mathbf{v}_{1}} \sqrt{\mathbf{v}_{2}^{T}\left(\boldsymbol{K}^{-T} \boldsymbol{K}^{-1}\right) \mathbf{v}_{2}}} \tag{6}
\end{equation*}
$$

The vanishing line for the ground plane corresponds to the intersection of image plane with the plane passing through the camera center and parallel to the ground plane. Once calibration is done and the vanishing line is known, the ground plane normal can be calculated.


Figure 2: Illustration of the calibration results. Left: Locating vanishing points by marking parallel lines. Right: The WCS we choose.

We assume a pinhole camera model, and the image center to be the central projection point. Then the only unknown in $\boldsymbol{K}$ is focus $f$.

First, to find $f$, two vanishing points $\mathbf{v}_{1}$ and $\mathbf{v}_{2}$ for two perpendicular lines on the ground plane are located semiautomatically. Using Eq. 6 , in which $\cos \theta=0, f$ can be found by solving a quadratic equation. Second, the line passing through the two vanishing points is the vanishing line of the ground plane. Hence the plane normal is known. We can set the WCS as $g_{x}=\boldsymbol{K}^{-1} \mathbf{v}_{1} /\left\|\boldsymbol{K}^{-1} \mathbf{v}_{1}\right\|$, $g_{y}=\boldsymbol{K}^{-1} \mathbf{v}_{2} /\left\|\boldsymbol{K}^{-1} \mathbf{v}_{2}\right\|$, and $g_{z}=g_{x} \otimes g_{y} . g_{0}$ can be set by arbitrarily setting the perspective depth of the image of an point that lies on the ground plane. Results are shown in Figure 2.

### 3.2. Quantitative Analysis on Synthetic Data

Using the calibration data, we synthesize very realistic tracking results that carry characteristics similar to the real data. We do some quantitative analysis on those sequences by adding noise to the tracked feature points and changing the calibration.

Figure 3 shows the example of the synthetic data we generated. 26 points are tracked over 40 frames. Note that the feature points on the vehicle are chosen to be points that can be tracked on a real vehicle, seen in Figure 7.

The noise we add to each feature points is drawn from an i.i.d. isotropic Gaussian distribution. We vary the standard deviation $\sigma$ to test how robust the method is to noise. Two kinds of calibration error are studied here: error from the focal length $f$, and from the angle between $g_{x}$ and the ground plane, denoted as $\phi$. Here we assume that $g_{y}$ can be estimated reliably, i.e., $g_{y}$ lies on the ground plane. That requires very accurate estimation of one vanishing point, which comes from only one set of parallel lines and is very plausible in man-made environments.

First, we consider the case that all the feature points are tracked over all the frames. The reconstruction results without missing data under two conditions are shown in Figure 4. For correct calibration $f=690, \phi=0^{\circ}$, and for wrong calibration $f=690, \phi=-4.6^{\circ}$. The same condition applies to Figure 6 (a) and (b). As can be seen in Figure 4, the
reconstruction does not change much visually when noise and incorrect calibration are considered.

In Figure 6 (a), the relative construction error is plotted as a function of $\sigma$ for both correct and wrong calibrations. For each $\sigma$, the experiments are run for 40 times and the average error is plotted. For correct calibration, the structure can be reconstructed perfectly with small $\sigma$ s. When the noise level $\sigma$ increases to about 7 pixels, performances under correct and incorrect calibrations become almost the same. (c) and (d) show how the reconstruction error changes as the calibration changes. In both cases, noise with $\sigma=2$ is added to all the feature points. The reconstruction error is less than $2 \%$ when $\delta f= \pm 50$ or $\phi= \pm 5^{\circ}$. (b) shows the rank condition by using the ratio of the 3rd to the 4th largest singular value of the scaled measurement matrix. Ideally, the 4th singular value should be zero and the ratio approaches infinity, which is true when $\sigma$ is close to zero with correct calibration. When $\sigma$ is large or calibration is incorrect, the ratio decreases.

As can be seen from these figures, the algorithm degrades gracefully with increasing noise level and incorrect calibration.

Second, we also consider the points occluded in some of the frames. 34 points are tracked in sections of the 40 frames. We apply SVD with missing data algorithm reported in [1] [15] so that we can have a more complete reconstruction. The reconstruction results for the case with missing data is shown in Figure 5. Several views have been fused together to map the texture onto the reconstructed 3D model. For this experiment, the data is noise free and calibration is correct. The reconstructed model correctly captures the structure of the model, such as parallel and perpendicular faces.

### 3.3. Qualitative Analysis on Real Data

The reconstruction for a real sequence is shown in Figure 7. Note that the image of the object is of low resolution and it is taken on a rainy day, which makes the image more blurry. 30 points are tracked over 40 frames on the vehicle with missing data. Because of the low resolution and blurry effects, many of them cannot be reliably tracked even manually. However, our method still correctly captures the relations of most faces. The reconstruction is visually realistic.

## 4. Conclusion and future work

A factorization method for structure from planar motion is proposed and illustrated with experiments. The method fully exploits the constraints, and uses SVD to batch processes the data. Experiments show that our method is fairly robust to noise and incorrect calibration. Future work includes theoretical error analysis along the lines of [9]. Our method can possibly be generalized to projective geometry,
non-rigid motion on the ground plane, and SfM for video captured on moving platforms. Motion constrained on nonplanar surface can also be studied.

## Acknowledgments

We would like to thank Shaohua Zhou for helpful discussions and providing his code for SVD with missing data.

## References

[1] M. Brand. Incremental singular value decomposition of uncertain data with missing values. In $E C C V$, pages 707-720, May 2002.
[2] M. Han and T. Kanade. Reconstruction of a scene with multiple linearly moving objects. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 2000.
[3] R. I. Hartley and A. Zisserman. Multiple View Geometry in Computer Vision. Cambridge University Press, second edition, 2004.
[4] T. Kanade and D. D. Morris. Factorization methods for structure from motion. Philosophical Transactions of the Royal Society of London, Series A, 356(1740):1153-1173, 1998.
[5] S. Li, S. Tsuji, and M. Imai. Determining camera rotation from vanishing points of lines on horizontal planes. In ICCV, pages 499-502, 1990.
[6] J. Oliensis. Structure from linear or planar motions. In CVPR, pages 335-342, 1996.
[7] L. Quan, Y. Wei, L. Lu, and H. Shum. Constrained planar motion analysis by decomposition. Image and vision computing, 22(5):379-389, May 2004.
[8] P. Sturm and B. Triggs. A factorization based algorithm for multi-image projective structure and motion. In $E C C V$, pages II:709-720, 1996.
[9] Z. Sun, A. M. Tekalp, and V. Ramesh. Error characterization of the factorization method. Computer Vision and Image Understanding, 82(2):110-137, 2001.
[10] T. Tan, K. Baker, and G. Sullivan. 3d structure and motion estimation from 2d image sequences. IVC, 11(4):203-210, May 1993.
[11] C. Tomasi and T. Kanade. Shape and motion from image streams under orthography: A factorization method. IJCV, 9(2):137-154, November 1992.
[12] R. Vidal and J. Oliensis. Structure from planar motions with small baselines. In ECCV, pages 383-398., 2002.
[13] L. Wang and W. Tsai. Camera calibration by vanishing lines for 3-d computer vision. IEEE Transactions on PAMI, 13(4):370-376, April 1991.
[14] A. D. Worrall, G. D. Sullivan, and K. D. Baker. A simple, intuitive camera calibration tool for natural images. In Proc. of the 5th British Machine Vision Conference, pages 781790, September 1994.
[15] S. Zhou, R. Chellappa, and D. Jacobs. Characterization of human faces under illumination variations using rank, integrability, and symmetry constraints. In ECCV, 2004.


Figure 3: Sample of a synthetic sequence. Top: The projected model. The car makes a $90^{\circ}$ turn. Bottom: Synthetic tracking results without missing data. Feature points selected according to real vehicles (compared to Figure 7).


Figure 4: Sample of reconstruction results with no missing data. First row, correct calibration with no noise. Second row, wrong calibration with added noise of $\sigma=5$.


Figure 5: Sample of reconstruction results from observations with missing data.


Figure 6: Quantitative analysis of reconstruction error and rank condition. Suppose during estimation of GPC, $g_{y}$ lies on the ground plane, and the angle between $g_{x}$ and the ground plane is $\phi$. For correct calibration $\phi=0$. (a) shows reconstruction error in the case of correct and wrong calibrations. Here for wrong calibration $\phi=-4.6^{\circ}$ and focal length $f$ is correct. (b) The ratio of the 3rd to 4th largest singular values is shown to analyze the rank condition of scaled measurement matrix. Calibration condition same as (a). (c) Reconstruction error as the estimated focal length $f$ deviates. Correct $f=690$. $\phi=0$ in this case. (d) Reconstruction error as estimation of $g_{x}$ changes. $g_{y}$ assumed to be on the ground plane. $f=690$ here. Noise of $\sigma=2$ is added in both (c) and (d).


Figure 7: Reconstruction results for a real sequence. Top row: The real sequence taken on a rainy day and the tracked feature points. Bottom row: Reconstruction with texture mapping.


[^0]:    *Partially funded by the ARDA/VACE program under the contract

