CAMERA CALIBRATION OF LONG IMAGE SEQUENCES WITH THE PRESENCE OF OCCLUSIONS

IBMR Lab
Image Based Modeling and Rendering Lab
Electrical Eng. & Computer Science Dept.
University of California, Irvine

Miguel Sainz, Antonio Susin and Nader Bagherzadeh

LABSID
Dynamic Simulation Lab
Applied Mathematics Dept.
Universitat Politècnica de Catalunya

For more information: *ibmrlab.ece.uci.edu* or *www-ma1.upc.es/~susin/contingut/labsid.html*

The Calibration Pipeline

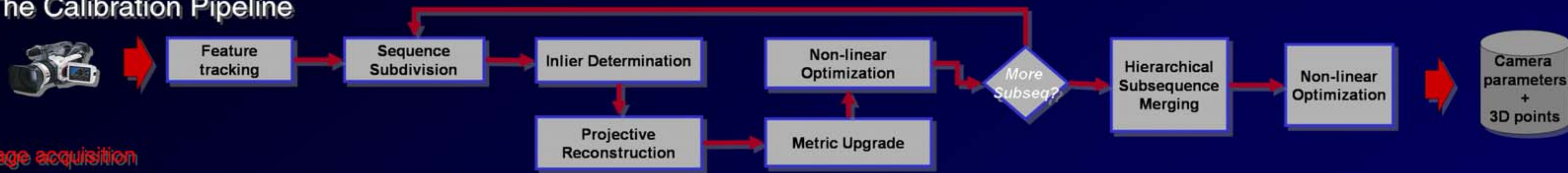

Image acquisition

We use a standard off-the-shelf camcorder and walk around the object to be reconstructed assuming no restrictions on the path, number of frames and camera parameters. Using a 2D tracking system, a set of significant features are extracted from the images.

$$\begin{pmatrix} \lambda_{11}\begin{bmatrix}u_{11}\\ v_{11}\\ 1\end{bmatrix} & \cdots & \lambda_{1n}\begin{bmatrix}u_{1n}\\ v_{1n}\\ 1\end{bmatrix} \\ \vdots & \vdots & \vdots \\ \lambda_{m1}\begin{bmatrix}u_{m1}\\ v_{m1}\\ 1\end{bmatrix} & \cdots & \lambda_{mn}\begin{bmatrix}u_{mn}\\ v_{mn}\\ 1\end{bmatrix}\end{pmatrix} = \begin{pmatrix} K_1\cdot[R_1|T_1] \\ \vdots \\ K_m\cdot[R_m|T_m]\end{pmatrix}\cdot(X_1 \cdots X_n)$$

n features
m views

$$W = P \cdot X$$

Image of a sparse measurement matrix

Sequence subdivision

The sequence is subdivided when one of the following is satisfied:

- A minimum of key frames has been selected.
- The end of the sequence is reached.
- More than a user selected percentage of the original features of the initial frame are lost in the actual one.

When this occurs, a new subsequence is created as the set of key-frames starting from the first frame to the last key-frame before the ending condition has been triggered. We use a RANSAC based approach to robustly determine the key-frames present in the subsequence by calculating a planar homography between the last key-frame and the actual one. If the percentage of inliers is small, there exists some significant camera motion not modeled by the 2D homography, and the actual frame is marked as a key-frame.

Camera calibration

Inlier determination

In order to improve the calibration, we determine the best set of inliers to perform a projective factorization by calculating the maximal set of features that fit in an affine reconstruction. This is performed using a RANSAC approach and a efficient and closed-form affine reconstruction solution.

Projective reconstruction

We perform a projective reconstruction over the set of inliers by means of an iterative factorization process that determines the projective depths that provide us with a proper rank-4 decomposition of the measurement matrix of the subsequence.

$$\underset{3m\times n}{W} = \underset{3m\times 4}{\hat{P}} \cdot \underset{4\times n}{\hat{X}} \qquad rank(W) = 4$$

Metric upgrade

The goal is to calculate the *projective distortion matrix* (PDM) that upgrades the projective reconstruction to a metric one.

$$W = \hat{P}\cdot\hat{X} = \hat{P}\cdot H\cdot H^{-1}\cdot\hat{X} \quad \text{such that} \quad P_i\cdot H = \mu_i K_i\cdot(R_i|T_i) \quad i = 1,\ldots,m$$

Assuming only the focal length as the single internal parameter, a closed form solution for the metric upgrade can be obtained by carefully analyzing the rank of the matrices involved in the upgrade and the relation of the PDM and *the absolute dual quadric*.

Non-linear optimization

A final non-linear optimization process based on the Levenberg-Marquardt algorithm is performed in order to reduce the re-projection error accounting for all the nonlinearities not recovered in the metric solution. The same method is applied after the subsequences have been merged.

Hierarchical merging

The merging of the subsequences is performed in metric space by determining the set of common points, between the last frame of one sub-sequence and the first of the next one, that by construction correspond to the same camera and measurements.

Results

The following example, the dataset head, consists of a video sequence with 370 frames taken with a handheld camcorder by moving it around a person's head. A total of 500 features per frame with automatic replacement have been tracked. The final reprojection error is 0.7 pixels with a standard deviation 0.6 pixels.

Recovered camera path

This research was partially supported by the National Science Foundation under contract CCR-0083080 and by the Comissio Interdepartamental de Recerca i Innovacio Tecnologica, Gaspar de Portola grant C02-03.