Motion-based 3D Object Reconstruction under Weak-Perspective

Levente Hajder

PhD Theses

Supervisors:
Dr. Dmitry Chetverikov (MTA SZTAKI)
Dr. István Vajk (BME AAIT)

Computer and Autmation Research Institute
Hungarian Academy of Sciences

Department of Automation and Applied Informatics
Budapest University of Technology and Economics

Budapest 2007.

1. Introduction

Motion-based object reconstruction is an interesting and challenging research area of computer vision. This area has rapidly progressed for the past 20 years. However, this does not effect that every problem has become solved. To the contrary, a lot of new problem to be solved are found by applying the methods published in the computer vision comunity's forums.

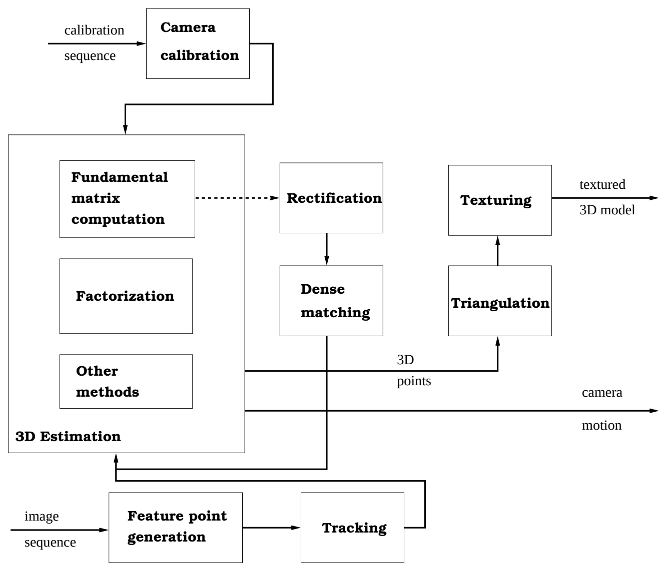
The results of object reconstruction are used in a lot of real application, but their speed, accuracy and automation level can be significally adjusted.

One of the most important application area of 3D computer vision is the robot vision, because the presence of the third dimension gives more information to the reconstruction methods. Another important applicaton is entertaining industry, especially film industry: since the millenium, a lot of world famous movie contains effects which use the results of the published reconstruction methods. The proposed methods are used from survaillance industry to the motorway patrol as well.

The researcher who want to develop reconstruction method has to have knowledge in the following topics: mathematics (optimization methods), optics (projections, lenses), digital image processing and computer graphics.

The main goal of my research work was to develop a modular reconstruction system with the help of my colleagues and university graduate students. This goal has been reached.

The developed system is visualised as follows.



Due to modularity, the system is under continuous development, components can be replaced by new ones. In my hteses, the proposed methods can be inserted into this sytem (except the algoritmhs proposed in the last thesis).

Structure of the dissertation

My dissertation can be divided into two main parts: in the first part, I deal with the reconstruction problem itself and with the geometry of the projection models, while the novel scientific results are given in the second part.

The component of the developed reconstruction system is discussed in Section 2 (after the introduction in Section 1). This section overviews the most important publications in the field. Section 3 reviews three projection model: perspective projection, weak-perspective projection and orthography.

The theses are given in the second part of the dissertation: the first one (Section 4) deals with the reconstruction of 3D points and 3D camera motion. Segmentation algorithms are proposed in the second thesis (Section 5), while Section 6 gives novel grouping algoritmhs of articulated objects.

2. Methodology

Factorization

The goal of the factorization is to compute the 3D coordinates of the object to be reconstructed and the camera motion from the moving feature points tracked in 2D images. The problem can written under weak perspective as follows:

If P feauture points of a rigid, moving object are detected and tracked over F frames, then the 2D coordinates of the p^{th} point in the f^{th} frame can be written under orthography as follows:

$$x_{fp} = R_f s_p + t_f, \tag{1}$$

where $R_f = [r_{f1}, r_{f2}]^T$ is the first two rows of an orthogonal matrix, s_p is the 3D coordinates of the point, t_f is the 2D place of the 3D origin in the image. If the weak perspective model is used insteaad of orthography, the equation is modified as follows:

$$x_{fp} = q_f R_f s_p + t_f, \tag{2}$$

where q_f is the scale factor of the weak-perspective camera model.

t_f can be eliminated if the location of the 3D origo is known in 2D. In order to simplify the problem, the origin should be selected as the center of gravity of the 3D object. This simplification modifies the projection equation as follows:

$$x_{fp} = q_f R_f s_p. \tag{3}$$

If all the points are considered, a matrix equation is obtained:

$$\underbrace{W_f}_{2\times P} = (x_{f1} \dots x_{fP}) = \underbrace{M_f}_{2\times 3} \cdot \underbrace{S}_{3\times P} \tag{4}$$

where M_f called *motion matrix*, $S = (s_1, \dots, s_P)$ *structure matrix*. Under orthography, $M_f = R_f$, while under weak-perspective $M_f = q_f R_f$.

If all frames are considered, equation (4) is modified and the so-called *measurement matrix* is expressed:

$$\underbrace{W}_{2F\times P} = \underbrace{M}_{2F\times 3} \cdot \underbrace{S}_{3\times P}, \tag{5}$$

where $W^T = [W_1^T, W_2^T, \dots, W_F^T]$ and $M^T = [M_1^T, M_2^T, \dots, M_F^T]$.

If the origin is not eliminated, the factorization can be writen in the following compact form:

$$W = \left[\begin{array}{c|c} M & t \end{array}\right] \left[\begin{array}{c} S \\ \hline 1^T \end{array}\right]. \tag{6}$$

where vector $t = [t_1^T t_2^T \dots t_F^T]^T$ contains the location of the origo in 2D.

The factorization task is to estimate structure matrix M and motion matrix M from measurement matrix W.

Tomasi-Kanade factorization

In order to compute 3D structure and camera motion, the Tomasi-Kanade factorization method consists of two main substep : (1) rank reduction is carried out by Singular Value Decomposition and affine structure and motion matrices are obtained: $W = M_{aff} S_{aff}$ where M_{aff} and S_{aff} are affine structure and motion matrices, respectively. In the computed subspace, a Q linear transformation should be calculated. Then the real structure and motion matrices are obtained:

$$W = MS \tag{7}$$

where $M = M_{aff} Q$ and $S = Q^{-1} S_{aff}$.

It is proved that at least 3 frames are needed to carry out the weak-perspective factorization.

Testing envirinment on sytnhesizes data

My novel methods proposed in the theses were tested on synthesized data which was generated on the same way. The 2D test coordinates were generated as follows:

1. 3D coordinates of a rigid object were generated by Gaussian random number generator. Its mean was zero and its deviation was a parameter.
2. For each frame, 3 rotation angles were generated by a random number generator as well. The 3D structure points were rotated by the generated angles around the corresponding axes.

3. Points were projected to the image planes by weak-perspective projection model.
4. 2D Gaussian noise was added to the projected point. (2D noise was generated a zero-mean Gaussian random number generator as well.)

The algorithms proposed in the theses were run on the generated test data. The results can be easily examined, because both the original 3D structure points and the camera motion are known.

Testing envirinment on real data

On real data, the first task is to detect corner points which are easy to track. This task is solved by the widely used KLT corner detector. A correlation-based tracker was used to match the feature points from frame to frame. Tipically, until 5...10 frames a feature point can be tracked.

Then the feature points are used as the input of the proposed methods. Practically, the segmentation method proposed by the second thesis should be applied first, because it can segment the moving feature points of a moving and compact rigid objects.

3. New scientific results

The new scientific results given by the dissertation are grouped into three theses:

First thesis. Iterative improvement of the Tomasi-Kanade factorization [2,3,8,16,17]

I proposed a new method which is the improvement of the widely used Tomasi-Kanade factorization.

The proposed method is an iteration. The algorithm finishes its run in finite time as it is proved in the dissertation.

Each cycle of the iteration consists of 3 steps:

S-step The S-step recomputes the 3D structure. The element of the structure matrix can be arbitrary, therefore the computation can be carried out optimally by the so-called pseudo-inverse operator, because the problem is linear.

$$S^{(k)} = M^{(k-1)^\dagger} W \tag{8}$$

M-step The M-step recomputes the motion matrix which is composed of the base vectors of the image planes. Under weak perspective, these base vectors are written as functions of 4 parameters. The motion matrix computation is solved optimally, because the problem can be transformed to registration problem of two 3D pointset. This problem is already solved except the scale parameter for which the optimal solution is given in the appendix of the dissertation.

Completion-step The completion step completes the measurement matrix by inserting the third cordinates of the moving objects' feature points in all frames by estimating them from the projection.

It is demonstrated on synthesized data that the improved method serves better results than the original one. It is true especially when the object to be reconstructed consists of a few point. The improved method was tested on real data as well.

Second thesis. Segmentation of moving rigid objects under weak-perspective [2,4,6,9,10]

I proposed novel methods which segment the feature points of moving objects. These methods can separate the feature points of a rigid object from outliers (badly tracked points) and from feature points of other moving objects.

2.a. Outlier rejection by Monte-Carlo-type methods In this subthesis, the Trajković-Kurata method was improved which can separate the outliers from the dataset containing the moving objects' feature points.

Two improvement were proposed for two subproblems: (i) the LMedS robust statistics was replaced to LTS (Least Trimmed Squares) (ii) the usage of real motion matrix was proposed instead of affine one. Improved methods serves better segmentation performance. With the help of LTS statistics, the methods can cope with more tha 50% percent of outliers.

A novel RANSAC-based segmentation was also proposed where RANSAC (RANdom SAmpling Consensus) is another reliable robust statistics. Ransac-based method can work with affine motion matrix as well as with real one.

The proposed methods were compared to the original one. It was shown that the outlier-ratio can exceed 50% if the LTS or RANSAC robust statistics are used. It is also demonstrated that the proposed algorithms give better separation performance than the original one.

The proposed algoritmhs are successfully carried out on real sequences as well.

2.b. Region-based segmentation

I have proposed a novel, region-based 3D segmentation method. This method uses the outlier rejection method proposed in thesis 2.a. The novel method can segment the moving feature points of a compact, rigid object. To my best knowledge, this is the first region-based 3D segmentation algorithm.

I proved that the reprojection error, which bases the proposed segmentation algorithm is invariant to the ambiguity of the factorization.

The proposed alhorithm is tested on real data and the results show that the segmentation works well in all test cases.

Third thesis. Reconstruction of articulated objects under weak perspective [11,12,13,14,18]

I have proposed two novel methods which can decide whether two rigid moving objects forms an articulated object. The first method deals with two objects connected by a joint, while the second one deals with the connection with an axis. The first method gives an estimation to the position of the joint in both 2D and 3D.

3.a. Estimation of the joint. It is assumed that the motion and structure matrices of the two rigid objects are computed as follows.

$$W_1 = \left[\begin{array}{c|c} M_1 & t_1 \end{array}\right] \left[\begin{array}{c} S_1 \\ \hline 1_1^T \end{array}\right], \tag{9}$$

$$W_2 = \left[\begin{array}{c|c} M_2 & t_2 \end{array}\right] \left[\begin{array}{c} S_2 \\ \hline 1_2^T \end{array}\right]. \tag{10}$$

The estimation of the joint is done by minimizing the following cost function:

$$J = \left\| W_1 - \left[\begin{array}{c|c} M_1 & t \end{array}\right] \left[\begin{array}{c} S'_1 \\ \hline 1_1^T \end{array}\right] \right\|_F^2 + \left\| W_2 - \left[\begin{array}{c|c} M_2 & t \end{array}\right] \left[\begin{array}{c} S'_2 \\ \hline 1_2^T \end{array}\right] \right\|_F^2. \tag{11}$$

The solution can be written by closed-form formula, because

the problem is linear:

$$\left[\begin{array}{c} \hat{o}_1 \\ \hat{o}_2 \end{array}\right] = [M_1| - M_2]^{\dagger}[t_2 - t_1], \tag{12}$$

where o_1 and o_2 denotes the joint in 3D, while t_1 és t_2 is the projections of the joint in 2D images.

3.b. Grouping of articulared objects connected by a joint. The solution of the grouping problem is solved by the estimation error of the joint given in thesis 3.a. as it can be written as follows:

$$\frac{1}{F}||M_1\hat{d}_1 + t_1 - M_1\hat{d}_2 - t_2||_2. \tag{13}$$

A threshold value should be given. This can be easily done, because the threshold is given in pixels.

3.c. Grouping of articulared objects connected by an axis. I transformed the grouping problem to the estimation of caxial circles. Then, the following algorithm is proposed to the coaxial circle fitting problem:

1. Parallel plane fitting to the given 3D points.
2. Extimation of the coaxial circles in the fitted planes.

The decision is done by a threshold value which should be compared with the estimation error of the coaxial circles fitting.

Every subthesis is tested on synthetic data as well as on real data.

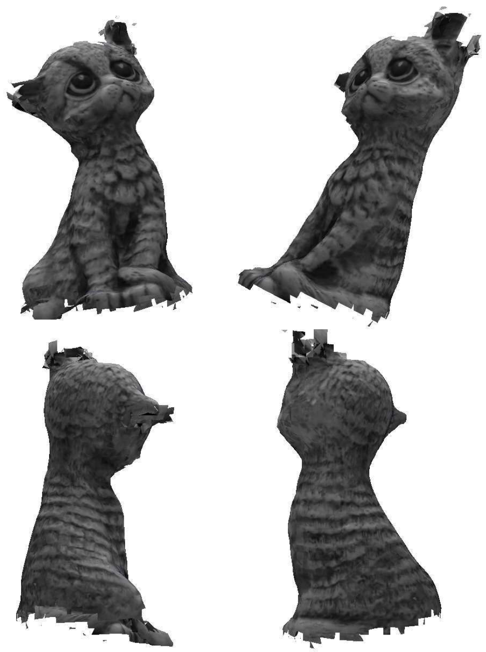
4. Applications

As it is mentiond in the introduction, the goal of my research work was to develop a full reconstruction system. Every thesis in the dissertation can be inserted into this system.

In this section two examples are shown which are contained by the dissertation as well. The first one has been reconstructed using three pictures taken from a human face as it is shown in the following:

The reconstruction is made by the weak-perspective algorithm proposed in the first thesis.

The second example is done by full-perspective reconstruction. 80 picture were taken by rotating a cat (sculpt). The computed model is triangulated and textured as it is visualized here:



The system is not perfect. Its advantage is that every subtask has been already solved. I plan to improve the system by inserting better and better novel method. My future methods will be inserted into this system as well.

5. Publications of the author

[1] **Hajder Levente**: Mozgásérzékelés webkamerával. In *Elektrotechnika*, 98:203-205, 2005.

[2] **Levente Hajder**, D. Chetverikov: Weak-perspective Structure from Motion for Stongly Contaminated Data. In *Pattern Recognition Letters*, 27:1581-1589, 2006.

[3] **Levente Hajder**: Shape and Motion from Video. In *Proc. II. Magyar Számítógépes Grafika és Geometria Konferencia*, pp. 98–102, 2003.

[4] **Levente Hajder**: Robust Structure from Motion by Outlier Rejection. In *Proc. Képfeldolgozók és Alakfelismerők Konferenciája*, pp. 30–36, 2004.

[5] **Hajder Levente**, Chetverikov D., Kardos I., Renner G.: Aktív kontúrok és Fast Marching eljárás alkalmazása az orvosi képfeldolgozásban. In *Proc. Képfeldolgozók és Alakfelismerők Konferenciája*, pp. 90–96, 2004.

[6] **Levente Hajder**, Dmitry Chetverikov: Robust Structure from Motion under Weak Perspective. In *Proc.* 2^{nd} *International Symposium on 3D Data Processing, Visualization & Transmission*, pp. 828–835, 2004.

[7] I. Kardos, **Levente Hajder**, D. Chetverikov D, G. Renner: Bone Surface Reconstruction from MR/CT Images using Fast Marching and Level Set Methods. In *Proc. Joint Hungarian–Austrian Conference on Image Processing and Pattern Recognition*, pp. 41–48, 2005.

[8] **Levente Hajder**: An Iterative Improvement of the Tomasi-Kanade Factorization. In *Proc. III. Magyar Számítógépes Grafika és Geometria Konferencia*, pp. 30–36, 2005.

[9] **Levente Hajder**, D. Chetverikov: Robust 3D Segmentation of Multiple Moving Objects under Weak Perspective. In *ICCV Workshop on Dynamical Vision*, CD-ROM, 2005.

[10] **Levente Hajder**, D. Chetverikov: Robust 3D Segmentation of Multiple Moving Objects under Weak Perspective. In *Lecture Notes on Computer Science*, 4358, pp. 48–59, 2007.

[11] **Levente Hajder**: 3D Motion Grouping of Articulated Objects. In *Computer Vision Winter Workshop (oral presentation after double blinded review)* pp. 125–130 ,2006.

[12] **Levente Hajder**: 3D Motion Grouping of Articulated Objects. In *Automation and Applied Computer Science Workshop* pp. 131–142, 2006.

[13] **Hajder Levente**: Tagolt objektumok rekonstrukciója videó alapján *Intelligens Rendszerek Fiatal Kutatók Szimpóziuma* (poszter) 2006.

[14] **Levente Hajder**: Optimal Joint Estimation of Articulated Objects under Weak Perspective *Proc. Képfeldolgozók és Alakfelismerők Konferenciája*, pp. 71–78, 2007.

[15] Csaba Kazó, Ákos Pernek, **Levente Hajder**: Texturing 3D Models from Images *Proc. Képfeldolgozók és Alakfelismerők Konferenciája*, pp. 148–156, 2007.

[16] **Levente Hajder**, Ákos Pernek, Csaba Kazó : Structure from Motion by Fast Alternation. *Proc. Workshop of the Austrian Association for Pattern Recognition (OAGM/AAPR)*, Accepted for publication, 2007.

[17] **Levente Hajder**, Ákos Pernek, Csaba Kazó : Fast and Precise Weak-Perspective Factorization. Submitted to *The 12th International Conference on Computer Analysis of Images and Patterns*, Vienna, Austria, 2007.

[18] **Levente Hajder** : Grouping of Articulated Objects with Common Axis. Submitted to *The 12th International Conference on Computer Analysis of Images and Patterns*, Vienna, Austria, 2007.