International Journal of Engineering Sciences & Emerging Technologies, Dec. 2014.
ISSN: 2231 – 6604 Volume 7, Issue 3, pp: 680-686 ©IJESET

OVERVIEW OF THE PROCESS OF 3D MODELLING FROM VIDEO

Svetlana Mijakovska, Igor Nedelkovski, Filip Popovski
Faculty of Technical Sciences, St. Kliment Ohridski University - Bitola, Macedonia

ABSTRACT

In this paper we give overview of the four main steps in the process of 3D modelling from video. We describe all the steps in detail to explain the basic concepts of the process of 3D modelling from video, especially the second step (structure and motion recovery.) The process is to find features in different frame of the video and match them, in order to create a model from them. The goal is to find the best algorithm for finding and fitting features to create a 3D model from video. In this paper, the three algorithms (RANSAC, MLESAC, MSAC) are described, tested and compared. These algorithms are important for process of 3D modelling from video, because they estimate parameters of models with outliers (outlier is an observation point that is distant from other observations, and may be due to variability in the measurement or it may indicate experimental error) and also resolve the correspondence problem in this process. We use these algorithms and create 3D model of a cube.

KEYWORDS: *3D Modelling, 3D Model, Video, Epipolar geometry, Structure from Motion (SfM), RANSAC, MLESAC, MSAC, Least Square.*

I. INTRODUCTION

The recovery of three dimensional structures from video is basically problem of Structure from Motion (SfM) and this process involves trying to recover, the 3D structure of a scene also the orientation and position of the camera at the moment the video is taken. Technics for structure from motion are used in many applications like photogrammetry [1], reconstruction of virtual reality models [2], estimating the camera motion [3].

Applications for SfM can be split into two categories, those that require geometric accuracy and those that require photorealism:

Applications that require geometric accuracy are less concerned with the visual appearance of the model and require high degree of accuracy for reconstruction of the scene structure and camera motion. For example, robot navigation, the reverse engineering of existing objects for use in CAD, film special effects that place computer–generated objects into the film and other 'augmented reality' applications thigh–accuracy models, require the camera motion to be very accurately reconstructed but the appearance of the structure is irrelevant as it is never seen in the finished product.

Applications that require photorealism are concerned about visual appearance of the model, and the reconstruction of the camera motion and scene structure are less concerned. This kind of applications is virtual reality, simulators, computer games and special effects that require a virtual set based on a real scene. [4]

In computer vision there are many applications for 3D modelling from images. But the process of 3D modelling is more interesting to research because the most important advantage of using a video sequence as input is high quality which can be obtained. Geometric accuracy and visual quality can be improved through the use of redundant data.

This paper is organized in sections as follows. In section 2 – "Related work", the previous work and basics of the structure from motion is presented. In section 3 – "Overview of 3D modelling from video sequences" we describe in detail every step of the process of 3D modelling from video. Section

4 – "Feature detection and matching" discuss about the first step of finding features and their matching. Also in this section the basics of epipolar geometry are given. Section 5 – "Structure and motion recovery" describe algorithms (RANSAC, Least Squares, MSAC and MLESAC) for finding corresponding points. The comparison, critical analysis of these algorithms and their parameters is made using the program Voodoo Camera Tracker, and the practise example of creating a 3D cube from video is described in Section 6 – "Practice example of generating cube from video". Section 7 – "Conclusion" summarises the observation made about the process of 3D modelling from video, and the obtained results from the practice example. Also in this section is give directions for further research. Section 8 lists the papers and other materials referenced.

II. RELATED WORK

Algorithms for structure from motion apply the principles of multi view geometry in order to match features across the sequences of images and to recover the structure of the scene and motion of the camera. These features are often points [5], but they can be lines [6] and primitives [7].
The first research of structure from motion was in 1980 by Longuet-Higgins [8], who made reconstruction of a scene from two views using eight point correspondences (this introduced the concept of the Essential matrix). In 1990 was found that the essential matrix could also be generalised to the case of uncalibrated cameras and the Fundamental matrix [9] was introduced. Then the trifocal tensor [10], quadrifocal tensor [11] and approaches for view from N-views were conceived [12].
Techniques of structure from motion are very unstable because the real data include outliers and noise. Torr [13] and Zhang [14] proposed using techniques like RANSAC and LMedS for increasing the robustness of finding and matching the features. The similar approach for increasing the robustness in case of using of trifocal tensor made Torr and Zisserman [15].

III. OVERVIEW OF 3D MODELLING FROM VIDEO SEQUENCES

The four main tasks of 3D reconstruction are:

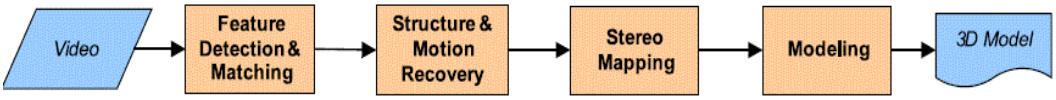


Figure 1: Main tasks of 3D reconstruction

The 3D reconstruction can be divided into 4 main tasks (Figure 1), which are discussed in the following sections:
1. Feature detection and matching. In this step using appropriate detector the features information is obtained, and using the descriptors the initial matching is made. The result from this step is correspondences.

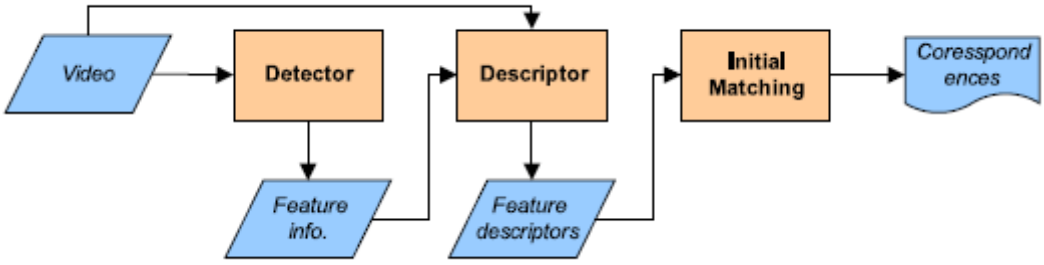


Figure 2: Feature detection and matching

2. Structure and Motion Recovery. Obtained correspondences from the first step are input in projective reconstruction which gave the projective structure (point cloud) and projection matrices. Using this information as input in metric reconstruction, metric structure (point cloud) and metric motion are obtained.

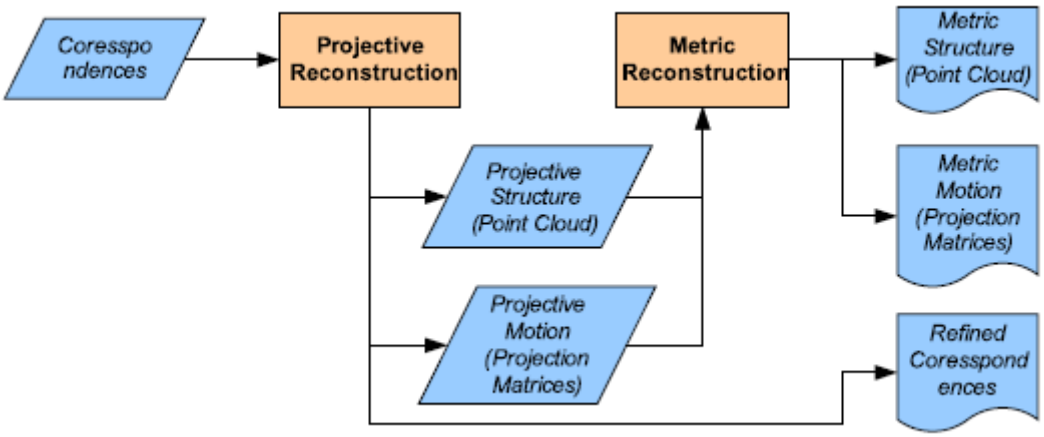


Figure 3: Structure and Motion Recovery

3. Stereo Mapping. Video frames and metric projection matrices are used in process of rectification and the result are rectified frames that together with correspondences are exposed in stereo mapping. The result of this process is dense matching map.

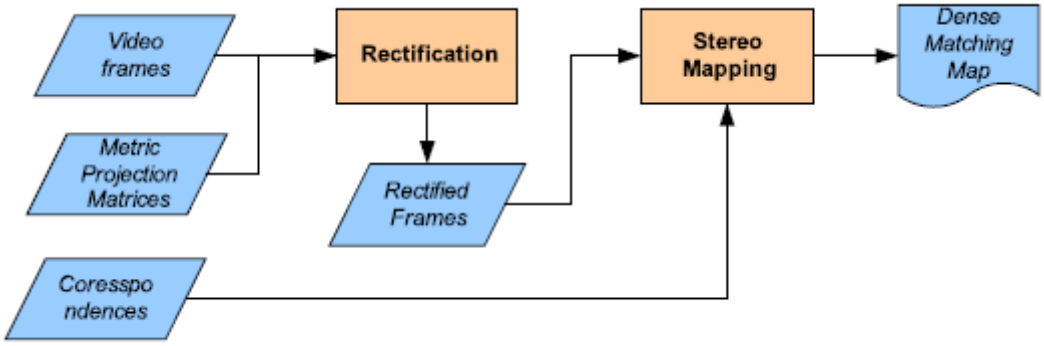


Figure 4: Stereo Mapping process

4. Modelling. The last make a realistic model of the scene (e.g. building mesh models, mapping textures). Using triangulation of the cloud point, the mesh is building, and with maps the texture we got a realistic 3D model.

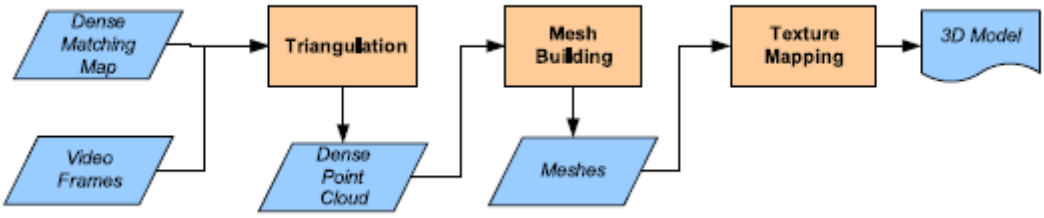


Figure 5: Modelling process

IV. FEATURE DETECTION AND MATCHING

Feature detection and matching (Fig.2) is process that detects and match features in different images. Video sequence is created of more images so in this step we must find interested points (point feature), i.e. detectors and descriptors.

The most important information given by detector is the location of features, but other characteristics such as the scale can also be detected. Two characteristics that a good detector needs are repeatability

and reliability. Repeatability means that the same feature can be detected in different images and should be distinctive enough so that the number of its matching candidates is small.
The descriptions are used to match a feature to one in another image. A descriptor should be invariant to rotation, scaling, and affine transformation.
The second task Structure and motion recovery recovers the structure of the scene and the motion information (position, orientation, and intrinsic parameters of the camera at the captured views) of the camera. The structure information is captured by the 3D coordinates of features. Because the fact that video sequence is created of more images, for this step we must research 3D reconstruction from multiple views i.e. multiple view geometry (Fig.6).
For the calibrated case, the essential matrix E [16] is used to represent the constraints between two normalized views. Given the calibration matrix K, the view is normalized by transforming all points by the inverse of K: $\hat{x} = K^{-1}x$, in which x is the 2D coordinate of a point in the image. K is a 3x3 matrix that includes the information of focal length, ratio, and skew of the camera. The new calibration matrix of the view is now the identity. Then with a corresponding pair of points (x, x') in homogeneous coordinates, E is defined by a simple equation: $\hat{x}'^{T} E\hat{x}$ = 0. Later the research to the uncalibrated case has been extended. During the 1990s, Faugeras [17] and Hartley [18] introduced concept of fundamental matrix F. The F matrix is the generalization of E and the defining equation is very similar: $x'^{T}Fx = 0$.

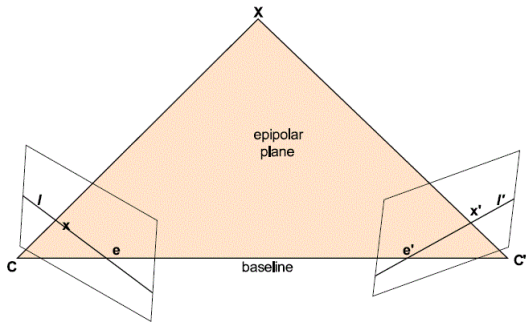


Figure 6: Two-view geometry

Three-view geometry is also developed during the 1990s. The geometry constraints are presented by trifocal tensors that capture relation among projections of a line on three views. The trifocal tensor defines a richer set of constraints over images (Fig.7). Apart of a line-line-line correspondence, it also defines point-line-line, point-line-point, point-point-line, and point-point-point constraints.

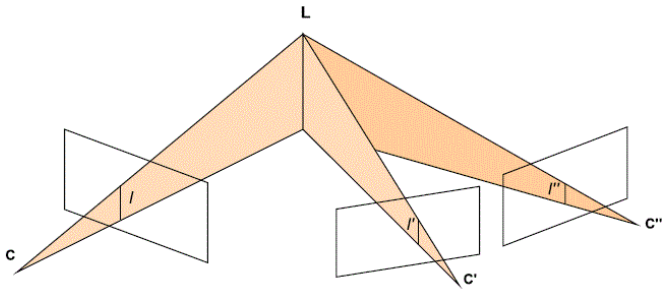
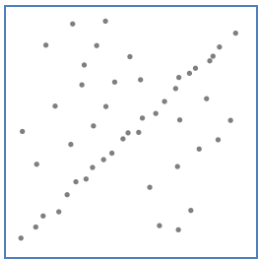


Figure 7: Line correspondence among three views - basis to define trifocal tensors

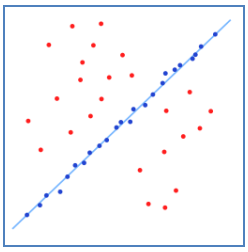
Projective reconstruction allows reconstruction knowledge of feature correspondences and there are many ways to obtain projection matrices from a geometry constraint, i.e. a fundamental matrix or a focal tensor. Hartley and Zisserman in [19] gave good review of the methods, implementation hints, and evaluations. If the input, i.e. feature correspondences, includes outliers, robust methods such as RANSAC, MLESAC, MSAC, LMS can be employed to reject them.

V. STRUCTURE AND MOTION RECOVERY

This step is actually the main step in 3D modeling from video, because in this step we must choose which algorithms to be used for find corresponding points of two images or more images with moving cameras at different points in time, with moving objects using different methods such as feature matching and block matching. We are research RANSAC, Least Squares, MSAC and MLESAC.
RANSAC algorithm is method to estimate the parameters of a certain model starting from a set of data contaminated by large amounts of outliers (than 50% of outliers). [20]
In this paper we describe the RANSAC algorithm and it two repeated steps hypothesize and test:
• Hypothesize. From the input dataset, minimal sample sets (MSSs) are randomly selected. Using this element of this set the model parameters are computed.
• Test. In the second step RANSAC produce called consensus set (CS) via checking which elements of the entire dataset are consistent with the parameters estimated in the first step.



Data with outliers



Line obtained with RANSAC, no influence of the outliers.

Figure 8. Example of line obtained with RANSAC algorithm without influence of outliers.

RANSAC only takes into account the number of inliers RANSAC minimizes cost:

$$C = \sum_i p(e_i^2)$$

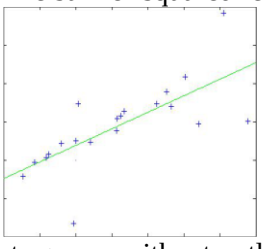
$$p(e_i^2) = \begin{cases} 0 & e^2 < T^2 \\ constant & e^2 \geq T^2 \end{cases}$$

The benefits of RANSAC are:

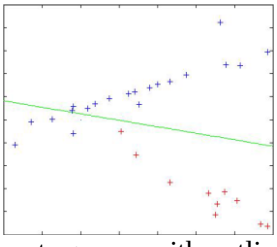
- only takes into account the number of inliers
- RANSAC minimizes cost.

Least Squares

- Calculate parameters of model function
- Over determined data set
- Minimize sum of squared residuals



Least squares without outliers



Least squares with outliers

Figure 9. Example of line obtained with Least Squares

MLESAC - Maximum Likelihood Estimation Sample Consensus
The MLESAC algorithm is an example of RANSAC that uses different cost function than the cardinality of the support. The algorithm was introduced by Torr and Zisserman [21]. Further improvements were made by Tordoff and Murray [22]. Instead of maximizing the support of the

model, the likelihood of the model is maximized. The error distribution is represented as a mixture of inlier and outlier distributions.

$$p(e) = (\gamma(\frac{1}{\sqrt{2\pi\sigma}})^n \, exp\left(-\frac{e^2}{2\sigma^2}\right) + (1-\gamma)\frac{1}{v})$$

$$p_{inlier} = \gamma(\frac{1}{\sqrt{2\pi\sigma}}) \, exp\left(-\frac{e^2}{2\sigma^2}\right)$$

$$p_{outlier} = ((1-\gamma)\frac{1}{v})$$

MSAC – M-estimator Sampling Consensus

$$p(e^2) = \begin{cases} 0 & e^2 < T^2 \\ T^2 & e^2 \geq T^2 \end{cases}$$

VI. PRACTICAL EXAMPLE OF GENERATING CUBE FROM VIDEO

In practical examples using the program Voodoo Camera Tracker [23] has been compared the following algorithms: RANSAC, MSAC and MLESAC. We were setting their parameters (max. and min. repetition, max. error distance, min. support ratio, min. subset size factor), compare the results and used them in the program Video Trace [24,25] (program for 3D modelling from video) and obtained the following 3D models:

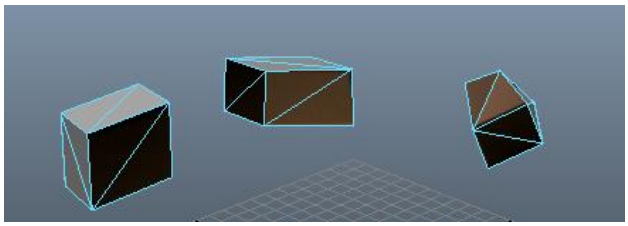


Figure 10. Comparison of MLESAC, MSAC and RANSAC algorithms in a practical example (triangulation)

The practice example of generating 3D model of cube from video shows that the best results give the MLESAC algorithm.

VII. CONCLUSION

The process of 3D modeling over the four main steps: feature extraction and matching, structure and motion recovery, stereo mapping, and modeling. Each step or even sub-step is already a field of research. The goal of this paper is to give overview of 3D modeling from video, especially the second step (structure and motion recovery) and to find the best algorithm for finding and fitting features to create a 3D model from video. We make comparison and testing of MLESAC, MSAC and RANSAC algorithms in a practical example, using Video Trace program for 3D modelling from video, and we get the best result gives MLESAC algorithm. Because our research release on cube, as directions for further research can be indicated 3D modelling from video of complex objects that contain curves and comparing algorithms for finding and fitting features to create a 3D model.

REFERENCES

[1] K. Kraus (1997), *Photogrammetry*, Volumes I and II. D¨ummler.
[2] A. Zisserman, A. Fitzgibbon, and G. Cross (1999*), VHS to VRML: 3D graphical models from video sequences*, In ICMCS.
[3] K. N. Kutulakos and J. R. Vallino (1998), *Calibration-free augmented reality*, IEEE Transactions on Visualization and Computer Graphics.
[4] Oliver Daniel Cooper (2005), *Robust generation of 3D models from video footage of urbane scenes*, University of Bristol.
[5] P. Beardsley, P. Torr, and A. Zisserman (1996), *3D Model Acquisition from Extended Image Sequences*, In ECCV, Cambridge, UK.

[6] C. Baillard, C. Schmid, A. Zisserman, and A. Fitzgibbon (1999), *Automatic line matching and 3D reconstruction of buildings from multiple views,* IAPRS Conference on Automatic Extraction of GIS Objects from Digital Imagery, Munchen.
[7] J. Alon and S. Sclaroff (2000), *An integrated approach for segmentation and estimation of planar structure*, Technical report, Boston University Computer Science.
[8] H. C. Longuet-Higgins (1981), *A computer algorithm for reconstructing a scene from two projections*, IEEE Transactions on Robotics and Automation.
[9] O. Faugeras, Q. Luong, and S. Maybank (1992), *Camera self-calibration: Theory and experiment*s. In European Conference on Computer Vision.
[10] R. I. Hartley (1995), *A linear method for reconstruction from lines and points*. In Proceedings of the International Conference on Computer Vision.
[11] R. Hartley (1998), *Computation of the quadrifocal tensor,* In Proceedings of the 5th European Conference on Computer Vision.
[12] A. Fitzgibbon and A.Zisserman (1998), *Automatic camera recovery for closed or open image sequences.* In Proceedings of the 5th European Conference on Computer Vision.
[13] P. Torr and D. Murray (1993), *Outlier detection and motion segmentation*. In Proceedings of SPIE Sensor Fusion, Boston.
[14] Z. Zhang, R. Deriche, O. Faugeras, and Q. Luong (1995), *A robust technique for matching two uncalibrated images through the recovery of the unknown epipolar geometry*. Artif. Intell.
[15] P. Torr and A. Zisserman (1997*), Robust parameterization and computation of the trifocal tensor*, Image and Vision Computing.
[16] R. Hartley and A. Zisserman (2004) *Multiple View Geometry in Computer Vision*, Cambridge University Press.
[17] S. Maybank O.D. Faugeras, Q. Luong (1992), *Camera self-calibration: Theory and experiment,* European Conference on Computer Vision.
[18] R.I. Hartley (1992), *Estimation of relative camera positions for uncalibrated camera*, Lecture Notes In Computer Science.
[19] R. Hartley and A. Zisserman (2004), *Multiple view geometry in computer vision second edition*, Cambridge University Press.
[20] Stefano Branco (2012) *RANSAC / MLESAC - Estimating parameters of models with outliers*, University of Basel.
[21] P.H.S. Torr and A. Zisserman, (2000), *MLESAC: A new robust estimator with application to estimating image geometry*, Journal of Computer Vision and Image Understanding.
[22] B. J. Tordo_ and D. W. Murray, (2005), *Guided-MLESAC: Faster image transform estimation by using matching priors*, IEEE Transactions on Pattern Analysis and Machine Intelligence.
[23] http://www.viscoda.com/en/products/non-commercial/voodoo-camera-tracker
[24] http://punchcard.com.au/
[25] A. Van Den Hengel, A. Dick, T. Thormahlen, B. Ward, and P. Torr (2007), *Videotrace: Rapid interactive scene modelling from video*, In ACM SIGGraph.

AUTHORS

Svetlana Mijakovska is an Assistant Professor in Graphic Engineering, Faculty of Technical Sciences in Bitola, Macedonia. She is currently working on her doctoral thesis, and she is interested in computer graphics, 3d modelling, graphic and web design and computer vision.

Igor Nedelkovski is a Doctor of Technical Sciences, Faculty of Technical Sciences in Bitola, Macedonia. He is interested in Computer Aided Engineering.

Filip Popovski is an Assistant Professor in Graphic Engineering, Faculty of Technical Sciences, Bitola, Macedonia .He is currently working on his doctoral thesis, and he is interested in computer graphics, visualization.