

# GENERATING 3D MODEL FROM VIDEO

Svetlana Mijakovska<sup>1</sup>, Igor Nedelkovski<sup>2</sup> and Filip Popovski<sup>3</sup>

<sup>1</sup>Faculty of Technical Sciences, St. Kliment Ohridski University - Bitola, Macedonia  
[svetlanamijakovska@gmail.com](mailto:svetlanamijakovska@gmail.com)

<sup>2</sup>Faculty of Technical Sciences, St. Kliment Ohridski University - Bitola, Macedonia  
[igor.nedelkovski@uklo.edu.mk](mailto:igor.nedelkovski@uklo.edu.mk)

<sup>3</sup>Faculty of Technical Sciences, St. Kliment Ohridski University - Bitola, Macedonia  
[filip.popovski@tfb.uklo.edu.mk](mailto:filip.popovski@tfb.uklo.edu.mk)

## **ABSTRACT**

*In this paper the process of 3D modelling from video is presented. Analysed previous research related to this process, and specifically described algorithms for detecting and matching key points. We described their advantages and disadvantages, and made a critical analysis of algorithms. In this paper, the three detectors (SUSAN, Plessey and Förstner) are tested and compare. We used video taken with hand held camera of a cube and compare these detectors on it (taking into account their parameters of accuracy and repeatability). In conclusion, we practically made 3D model of the cube from video used these detectors in the first step of the process and three algorithms (RANSAC, MSAC and MLESAC) for matching data.*

## **KEYWORDS**

*3D modelling, structure from motion, corner detector, key points (features), feature detection, matching.*

## **1. INTRODUCTION**

Complex, simple and large-scale 3D models are commonly used in areas such as computer games, animation and virtual reality. The generating three dimensional structure from video of scenes is interesting area for research and this problem is known as Structure from Motion (SfM). This process recovers the 3D structure of a scene and the orientation and position of the camera when each image was captured. In the years backward algorithms for structure from motion (Structure from Motion) refer to points so that can detect 3D structure of a scene from video sequences. These structures are mostly in the form of sparse 3D point cloud, the additional negative impact of so-called outliers (points that do not belong to the structure). The system is able to create 3D models from video, if it is getting successively network of these clouds of points, accurately and automatically [1], [2], [3], [4]. Obtaining 3D models from images is known and there are more applications for this. But getting the 3D models of the video is a new challenge for research and the goal is generating 3D models from amateur video made with a digital camera, mobile phone or tablet. First step in this process is obtaining a cloud of points, then matching the key points, obtaining the network and convert the network in polygonal model.

## **2. RELATED WORK**

Because the first step in the process of generating 3D models from video is main, and therefore the quality of resulting model depends from this step, the related work is refers to corner

detectors. From 1970's when the first corner detector was developed in the late, many corner detectors have been proposed.

The first operators of interest points was developed by Moravec (1977), Beaudet (1978), Kitchen & Rosenfeld (1982), Forstner (1987), Plessey (1988), Deriche (1990), Wang & Brady (1992), SUSAN (1997), CSS, Trajkovic & Hedley (1998), Zheng & Wang (1999). All corner detectors belong to one of these three categories: contour based, intensity based and parametric model based detectors. If detector extract contour from the image and find interest point, than that is contour based detector. Detector is intensity based if it find points by examine the intensity change around the point. If from the matching models the points are found, than we are talking about model based detector. In following text we cover intensity based detectors (SUSAN, Plessey and Förstner), because we take point as a features.

### **3. 3D MODELLING FROM VIDEO**

3D modelling is the process that, by using the appropriate software, provides mathematical representation of three-dimensional surface of an object. As a result of this process the 3D model is obtained. If the model is render, than it is shown as two-dimensional image.

The process of 3D modelling from video consists of four main tasks:

1. Feature detection and matching.
2. Structure and Motion Recovery.
3. Stereo Mapping.
4. Modelling.

In the first step the same features are found in different images, and these features are match. The second step the parameters of camera (orientation, position) are recovered, and thus the 3D coordinates of the features are also calculated. In the third step the dense map is obtained, using the recover structure from the previous step. And the last step, modelling, gives realistic of the model, adding textures.

### **4. FEATURE DETECTION AND MATCHING**

This process consists of detecting and matching the same features in different images (frames of the video). Usually, the features used in structure recovery processes are points ([5, 6]) and lines ([7]).

For this step of feature detection and matching two concepts are important: detectors and descriptors. Feature detector is process to detect features from the image. The most important information a detector gives is the location of features. A detector is good if it has repeatability (the same feature can be detected in different images) and reliability (the detected point should be distinctive enough so that the number of its matching candidates is small).

The second concept is descriptor: if we have two images (from two different views) of a scene and already have extracted some features of them, we need descriptors to find corresponding pairs of features. This is a process that takes information of features and image to produce descriptive information i.e. features description, which are usually presented in form of vectors and then matched a feature from one to another image. A good descriptor should be invariant to scaling, rotation, and affine transformation.

The following text is an overview and comparison of different corner detectors [8] - [13].

Table 1. Comparison of characteristics of different corner detectors

Corner detector	Detection Rate	Localization	Repeatability Rate	Robustness to Noise	Speed
Beaudet	Fair	Fair	Poor for scaling, good for affine transformations	Poor	Good
Moravec	Fair	Good	Fair	Fair	Good
Kitchen & Rosenfeld	Fair	Fair	Fair	Fair	Poor
Forstner	Good	Good	Excellent for affine transformations, fair for scaling	Good	Poor
Plessey	Good	Good for L-junctions, poor for all other types	Excellent for affine transformations if isotropic gradient calculations are used, fair for scaling	Fair	Poor
Deriche	Left	Good	Good	Poor	Good
Wang & Brady	Good	Good	Good	Fair	Good
SUSAN	Good	Bad for blurred images, very good otherwise	Good for scaling, poor for affine transformations	Excellent	Good
CSS	Good	Good	Excellent	Good	Highly dependent on edge detector used
Trajkovic & Hedley (4-neighbours)	Poor	Good	Fair (not rotationally invariant)	Poor	Excellent
Trajkovic & Hedley (8-neighbours)	Fair	Good	Fair (not rotationally invariant)	Good	Excellent
Zheng & Wang	Good	Good for L-junctions, fair for all other types	Excellent for affine transformations, fair for scaling	Fair	Fair

In order to find the best corner detector which we will use in our practice example of 3D modelling from video, using Voodoo Camera Tracker [13] program, the three corner detectors are compared: SUSAN, Plessey and Förstner, and we received the following results:

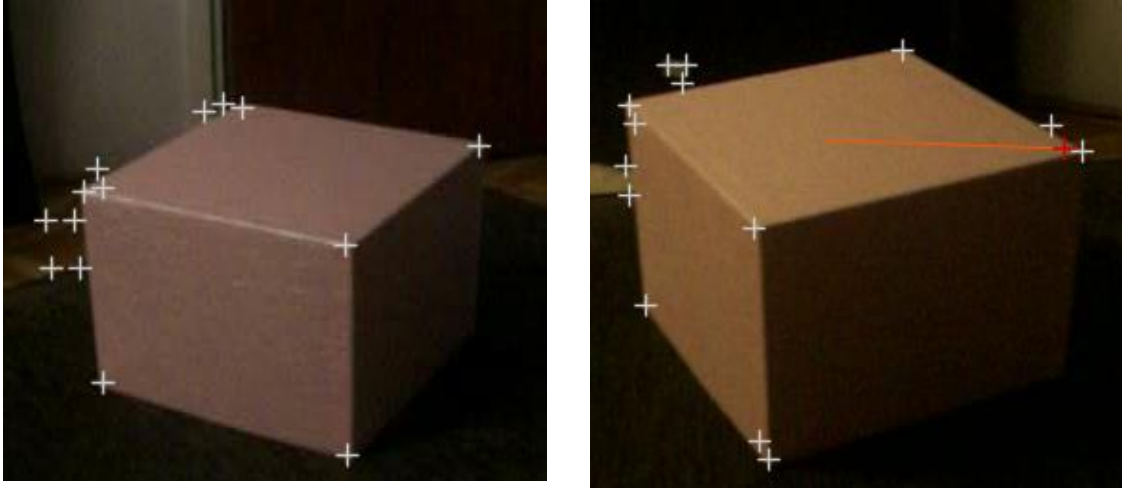


Figure 1. SUSAN corner detector

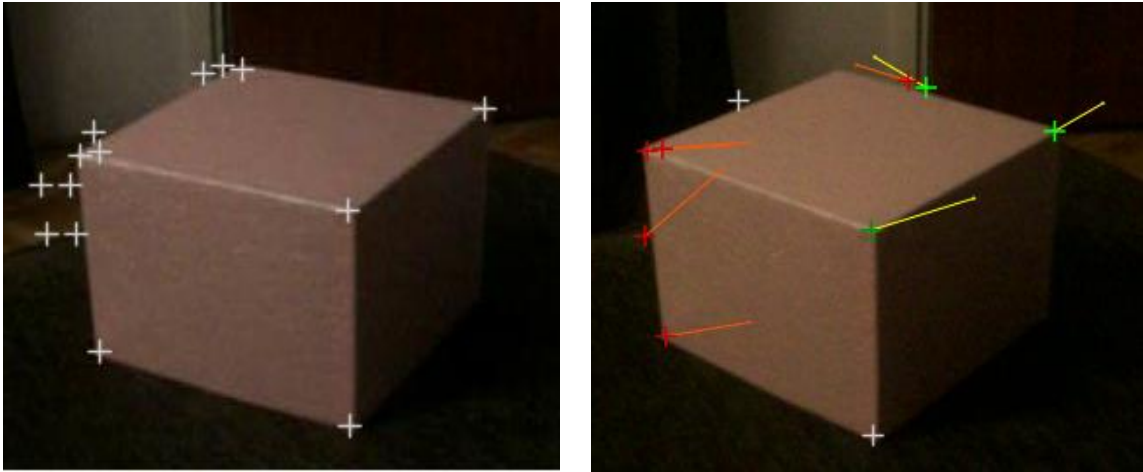


Figure 2. Plessey corner detector

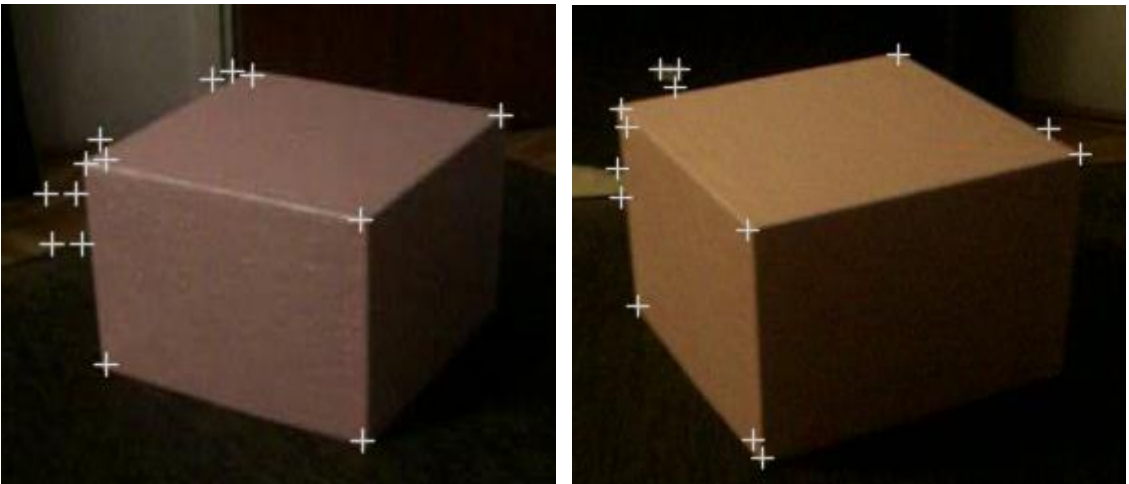


Figure 3. Förstner corner detector

From the results it can be concluded that the best results give Förstner corner detector, and in view of the detection of interesting points, good speed and lowest value of RMSE (root-mean-square error) i.e. difference between the values of the model and the values obtained in a classical way (Table 2). The Root Mean Square Error (RMSE) is error measure. A model predict values and RMSE make differences between them and observed values from the environment that is being modelled. RMSE collect these differences (residuals) into a single measure of predictive power.

If  $X_{obs}$  is observed values and  $X_{model}$  is modelled values at time/place  $i$ , the RMSE of a model is defined as the square root of the mean squared error:

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (X_{obs,i} - X_{model,i})^2}{n}}$$

The RMSE values can be used to compare the individual model performance to that of other predictive models.

Table 2. Comparison of RMSE and speed of different corner detector

<b>Detector</b>	<b>Speed</b>	<b>RMSE</b>
SUSAN	5s	0,343508
Plessey	6s	0,441709
Förstner	5,4s	0,321272

For the purposes of this paper, as a first step in the process of 3D modeling from video, we will use Förstner corner detector, because it is the most stable with respect to changes in contrast and noise, unlike the other two detectors.

## 5. STRUCTURE AND MOTION RECOVERY

The goal of the second task Structure and motion recovery is to recover the motion information of the camera and the structure of the scene. Reconstruction with knowledge of only a few features is possible only with projective reconstruction and there are more ways to get the projection matrix of geometrical constraint, from basic matrix or focal tensor. Therefore, the projective reconstruction is a major detection or fundamental matrices or focal tensor. For data matching process one or more of listed algorithms are used: RANSAC, MSAC and MLESAC [14].

In practical examples using the program Voodoo Camera Tracker [13] has been compared the following algorithms for increase the robustness: RANSAC, MSAC and MLESAC.

Information obtained from Voodoo Camera Tracker (selected corner detector and algorithm for matching) are used in the program Video Trace [15] (program for 3D modeling from video) and obtained the following 3D models:

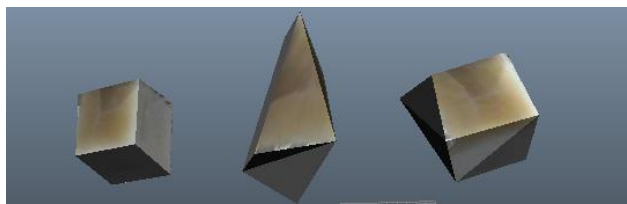


Figure 4. Comparison of MLESAC, MSAC and RANSAC algorithms in a practical example

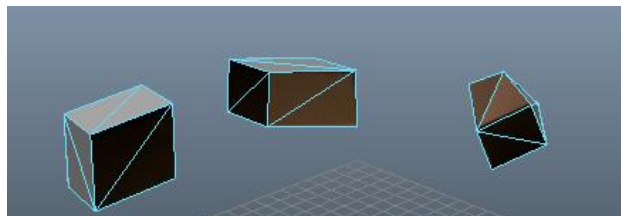


Figure 5. Comparison of MLESAC, MSAC and RANSAC algorithms in a practical example (triangulation)

## 6. CONCLUSIONS

This paper outlines the basic concepts on which basis the process of 3D modelling from video, i.e. obtaining structure from motion. The process contain four main tasks, but we pay attention of the first and second task (Feature detection and matching and Structure and Motion Recovery). The chronology of algorithms for finding interesting points and their matching and comparing are referred. The practical example shows that the best results given MLESAC algorithm using Förstner corner detector (in view of the detection of interesting points, good speed and lowest value of RMSE). The practical example concerns the cube, so as directions for further research can be indicated 3D modelling from video of complex objects that contain curves.

## REFERENCES

- [1] J. O'Rourke, (1998) *Computational geometry*, C. Cambridge University Press.
- [2] W. Press, B. Flannery, S. Teukolsky, and W. Vetterling, (1993) *Numerical Recipes in C: The Art of Scientific Computing*, Cambridge University Press.
- [3] R. Hartley and A. Zisserman (2004) *Multiple View Geometry in Computer Vision*, Cambridge University Press.
- [4] O. Faugeras, Q. Luong and T. Papdopoulou (2001) *The Geometry of Multiple Images: The Laws That Govern The Formation of Images of A Scene and Some of Their Applications*, MIT Press.
- [5] R.I. Hartley (1997) *In defence of the eight-point algorithm*, IEEE Transactions on Pattern Analysis and Machine Intelligence.
- [6] M. Pollefeys (2004) *Visual modelling with a hand-held camera*, International Journal of Computer Vision, 59:207-232.
- [7] R.I. Harley (1995) *A linear method for reconstruction from lines and points*, 5th International Conference on Computer Vision, page 882.
- [8] M. Trajtkovic and M. Hedley (1998) *Fast Corner Detection*. *Image and Vision Computing*, Vol. 16(2), pp. 75-87.
- [9] C. Schmid, R. Mohr, and C. Bauckhage. (1998) *Comparing and Evaluating Interesting Points*, International Conference on Computer Vision, pp. 230-235.

- [10] F. Mohanna and F. Mokhtarian (2001) *Performance Evaluation of Corner Detection Algorithms under Similarity and Affine Transforms*, BMVC.
- [11] Z.Zheng, H.Wang and EKTeoh (1999) *Analysis of Gray Level Corner Detection*, Pattern Recognition Letters, Vol. 20, pp. 149-162.
- [12] D.Parks, J.P. Gravel (2004) *Corner Detection*.
- [13] Stefano Branco (2012) *RANSAC / MLESAC - Estimating parameters of models with outliers*, University of Basel.
- [14] <http://www.viscoda.com/en/products/non-commercial/voodoo-camera-tracker>
- [15] <http://punchcard.com.au/>

### Authors

Ms. 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.



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



Mr. 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.

