# Fundamental Matrix estimation and photo sequencing

Marc Carné
Universitat Autnoma de Barcelona
Bellaterra, Catalonia/Spain
marc.carne.herrera@gmail.com

Alejandro Nespereira
Universitat Autnoma de Barcelona
Bellaterra, Catalonia/Spain
alejandronespereira@gmail.com

Gonzalo Benito
Universitat Autnoma de Barcelona
Bellaterra, Catalonia/Spain
gonzabenito@gmail.com

## Abstract

*An image is the projection of the real world on an image plane. Projecting from a 3D world to a 2D world we lose spatial information. Besides, a projective transformation is bound to create some distortions in the image plane. Projective transformation have inherently a lose of lengths, angles metrics or parallelisms. Despite this information loss we can reconstruct scenes in 3D or build mosaics from recovered images. In this work we focus on homographies and their uses. We present one of the methods available to compute them and finally we apply this knowledge to a set of images that comes from different scene views in order to create a single mosaic. Additionally, we explore one of the methods used for auto-calibrating cameras with easily available materials.*

## 1. Motivation

3D reconstruction is a growing field in both academia and industry. The appearance of virtual and enhanced reality has given rise to new technologies that can be used in our everyday life, such as entertainment, design and maintenance. However, the requirements to accurately reconstruct a 3D scene are high, at least on the case of multi camera geometry. Other approaches such as pattern light, time of flight or laser triangulation, offer similar results with different limitations, but they are not in the scope of this report.

Working on camera images with no additional depth information requires an accurate understanding of the scene from the different camera views. This is achieved studying the spatial relationship between the centers of the cameras, which is conveniently compressed in the fundamental matrix **F**. This matrix plays a key role in triangulation, as it can reduce the search of point correspondence between images from a 2D plane, our image, to a single 1D line, the epipolar line.

To robustly estimate **F**, a set of matching correspondences between image points is needed. These matches are obtained through SIFT descriptors, but other descriptors or Optical Flow information could have been used as well. We study a robust way to construct our matrix.

## 2. Method

### 2.1. The Fundamental Matrix

The fundamental matrix captures the information about the epipolar geometry of 2 views. It is important to note that this matrix does not relate the pixel positions in two images of an object, as it only gives constrains about how these pixel positions change under a point of view transformation. This is not sufficient to establish a pixel to pixel relation between images, but it enormously reduces the search space for a correspondence. Moreover, as it only relates changes under view point transformations, any moving object that presents a different world position in each image can not be described with this matrix.

The fundamental matrix between two cameras is directly derived from the epipolar constraints. With two image planes and the projection of a world object into two image points **x** and **x'**, we can see the relation between the camera centers as a translation vector **T** and a rotation matrix **R**. Setting a plane that has both our camera centers and our image points **x** and **x'**, it can be deduced that the vector $T' \times Rx$ is perpendicular to this plane. As **x'** is in the plane this leads to

$$x'^T[T'_x]Rx = 0 \tag{1}$$

$[T'_x]R$ is called the Essential matrix, and it relates point to point correspondences: $x'Ex = 0$.
As **x** and **x'** are both in camera coordinate system, we can obtain **p** and **p'** as:

$$p = Kx \rightarrow x = K^{-1}p \tag{2}$$
$$p' = K'x' \rightarrow x' = K'^{-1}p' \tag{3}$$

Replacing in (1) we obtain:

$$(K'^{-1}p')^T[T'_x]RK^{-1}p = 0 p'K'^{-T}[T'_x]RK^{-1}p = 0 \tag{4}$$

The fundamental matrix F is then defined as $F = K'^{-T}[T'_x]RK^{-1}$ and it relates point to point correspondences similarly to the essential matrix: $p'Fp = 0$, where **p** and **p'** are expressed in pixels. For the unlikely case where both camera intrinsic matrices are equal to the identity, the fundamental matrix is in fact, the same as the essential matrix.

## 2.2. 8 point Algorithm

More often than not, however, the spatial relationship between cameras or the camera intrinsic matrices are not available. However, we can estimate this with point correspondences, as we know that $p'Fp = 0$ for every pair of matching points. As both points are in homogeneous pixel coordinates we can write it as

$$\begin{pmatrix} u & v & 1 \end{pmatrix} \begin{pmatrix} F_{11} & F_{12} & F_{13} \\ F_{21} & F_{22} & F_{23} \\ F_{31} & F_{32} & F_{33} \end{pmatrix} \begin{pmatrix} u' \\ v' \\ 1 \end{pmatrix} = 0 \tag{5}$$

This can be rewritten as:

$$\begin{pmatrix} uu' & vu' & u' & uv' & vv' & v' & u & v & 1 \end{pmatrix} \begin{pmatrix} F_{11} & F_{12} & F_{13} & F_{21} & F_{22} & F_{23} & F_{31} & F_{32} & F_{33} \end{pmatrix}^T = 0 \tag{6}$$

Adding 7 additional point correspondences creates a linear system $Wf = 0$, solvable with Linear Least Square, decomposing W into its SVD matrices we can obtain the solution as the last column of V, and then use it to compose F from the f values. However, this gives a rank 3 fundamental matrix, which make the epipolar lines not coincide in the epipole. To solve this, we need to decompose again using SVD and force F to be rank 2, by removing the last singular value in the diagonal of the matrix D. With this we can recompute a F that is assured to have rank 2.

## 2.3. Normalized 8-point algorithm

Reliable as it is, the Linear Least Square method is very sensible to noise and highly unstable, performing poorly when faced against data of varying magnitude. This can be addressed by previously transforming the data to a normalized space. Data in this space should be centered around the origin and have a mean square distance to it of 2 pixels. This can be achieved with an homography consisting of a translation and a scale factor. While it is simple to see that the translation is the centroid of the original data, the scale factor is more complex to obtain. The mean square distance from each point to the centroid is

$$d^2 = \frac{1}{N}\sum \|\frac{p_i}{s} - \frac{c}{s}\|^2 = \frac{1}{s^2 N}\sum \|p_i - c\|^2 \tag{7}$$

where $p_i$ are the points in the original space and $c$ is their centroid. The term $s$ is the scale that we want to apply to the points. It can be proven that the scale is:

$$s = (\frac{1}{d^2 N}\sum \|p_i - c\|^2)^{\frac{1}{2}} = (\frac{1}{2N}\sum \|p_i - c\|^2)^{\frac{1}{2}} \tag{8}$$

With our centroids **c** and **c'** and the scales **s** and **s'** , we can build our pair of homographies **H** and **H'** that normalize our two input set of points:

$$H = \begin{bmatrix} s^{-1} & 0 & s^{-1}c_x \\ 0 & s^{-1} & s^{-1}c_y \\ 0 & 0 & 1 \end{bmatrix} \tag{9}$$

H' is constructed identically using **c'** and **s'**. We calculate $q = Hp$ and $q' = H'p'$, our normalized inputs and use them to obtain the fundamental matrix with the 8 point algorithm described above. The resulting $F_q$, however, still has to be de-normalized in order to be valid for our points **p** and **p'**:

$$F = H'^T F_q H \tag{10}$$

## 2.4. RANSAC Fundamental Matrix

Although normalizing the input data shields the fundamental matrix from errors due to numerical instability, errors in the points selected will result in a wrong fundamental matrix. This is due to the fact that the fundamental matrix estimation is naive in the sense that it does not perform any validation of the selected matches. To solve this, an iterative RANSAC approach can be used to estimate a matrix that provides a good result while being robust to outliers.

The idea is very similar to the one presented in the past session for homography estimation. From our set of point matches we select a subset of N points (8 for fundamental matrix estimation), from which we compute **F**. We test this matrix against all our points and count the number of inliers. We repeat this until we are sure, in a probabilistic way, that we have picked at least one subset free of outliers. As with the homography case, we need to provide a method to compute the number of inliers for a given model. We have applied a threshold over the first order of the geometric error to determine if a point correspondence is an inlier. This approximation is also known as the Sampson distance:

$$d_i = \frac{(x_i^T F x_i)^2}{(Fx_i)_1^2 + (Fx_i)_2^2 + (Fx_i')_1^2 + (Fx_i')_2^2} \tag{11}$$

where the subindex denotes the dimension.

## 2.5. Epipolar geometry

Through the session we have used several notions of epipolar geometry that we think is necessary to add here.

### 2.5.1 Epipoles

The epipoles present very useful relationships with the rest of the epipolar geometrical elements. The epipoles are the points where the baseline intersect with the image planes, and therefore $Fe = 0$ and $F^t e' = 0$. Moreover, the epipoles are the points in which all the epipolar lines intersect, and thus it can be written that $l \times e = 0$ for all epipolar lines l.

### 2.5.2 Relationship between a point and its epipolar line

From the fundamental matrix defined as $x'Fx = 0$, we can define the relationship between a point in the first image **x** and its corresponding epipolar line in the second image as: $l' = Fx$. The relationship between a point in the second image **x'** and its epipolar line in the first image is $l' = F^T x$.

# 3. Results

## 3.1. Compute the fundamental matrix

In the first exercise we have computed the fundamental matrix with 8 given points. We have then compared this fundamental matrix with the one obtained applying the formal definition of the fundamental matrix. The comparison has been done by evaluating the square difference of both matrices divided by their norm. We have obtained a difference of 0.27622, being the absolute differences between elements:

$$Diff(F_{gt}, F_{es}) = \begin{bmatrix} 0.0125 & 0.0466 & 0.5104 \\ 0.0466 & 0.0125 & 0.0724 \\ 0.0046 & 0.0762 & 0.0000 \end{bmatrix} \tag{12}$$

## 3.2. Robustly fit fundamental matrix

The second exercise is a more complex case, where we look for the fundamental matrix between two images without prior knowledge or point correspondences.

The firs step of the process is extract keypoints and their descriptions that we can use to estimate the fundamental matrix. We extract these points using SIFT feature detector and descriptor, which can be seen at figure 3. The two sets of points are then matched using the matcher function provided to us.

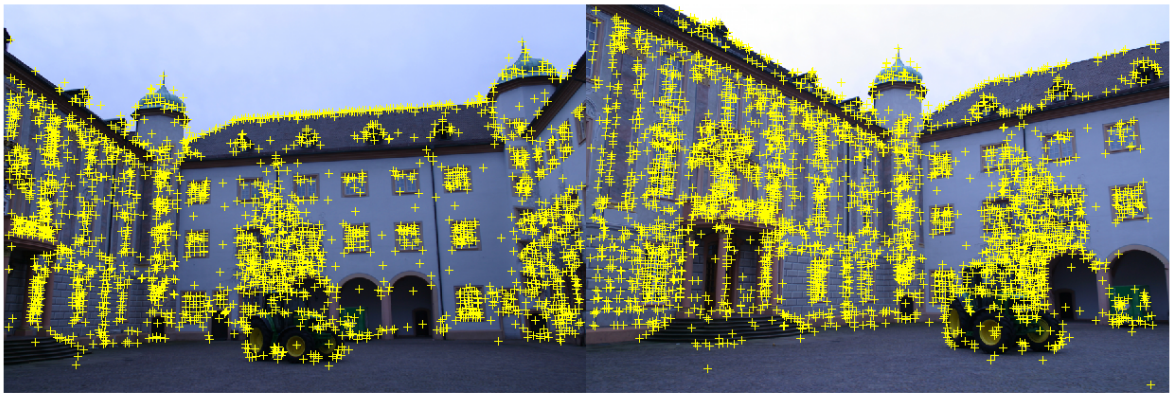

Figure 1: The sift keypoints found at the two images. Due to the movement of the camera, some points of each image cannot be matched against the other image.

Once we have the matches we can estimate the fundamental matrix **F** using the RANSAC approach discussed in section 2.4. In our case, and as can be seen at figure **??**, the keypoints have been very well matched with few errors.

After an average of 36 iterations (tested 20 times), we obtain the fundamental matrix **F**:

$$F = \begin{bmatrix} 0.0000 & 0.0000 & 0.0021 \\ 0.0000 & 0.0000 & -0.0068 \\ 0.0030 & 0.0072 & 0.4911 \end{bmatrix} \tag{13}$$

We then proceed to evaluate visually the results. To do that, we compute all the epipolar lines from the two sets of keypoints. Then, selecting three keypoints from the inliers obtained when computing F at random we plot them along with the epipolar lines of their matches.

# 4. Optionals

In this section we have had the chance to sequencing algorithm based on the work of Dekel, Moses and Abidan in 2013. The aim of it is to incorporate temporal information to a set of still images which already hold dynamic context. The proposal takes a set of images taken by uncalibrated cameras and allows to establish the chronological order of the captures. The first

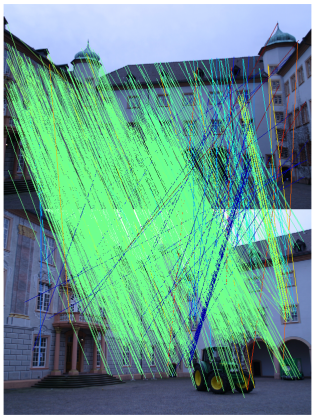

Figure 2: The matches for our SIFT keypoints. To ease the viewing, each match line has been coded depending on angle between both image pixel positions. It can be seen that most of them share a very similar angle, which points to a robust matching. The few exceptions are caused by the keypoints not being present in both images.

Figure 3: Keypoints in both images as well as their corresponding epipolar lines. Judging visually, it seems that the estimated fundamental matrix F is precise, as all the points are located in their corresponding matching points epipolar lines. Colors of points and lines from matching points match to ease visualization .

step involves the extraction of static features, which in this case are taken by implementing a SIFT detector. These features help to compute the epipolar geometry between images. Once, obtained the SIFT points, a match between image 1 and each of the other images is done so that we get a set of correspondent points in every pair of pictures to compute the fundamental matrix associated to them. Figure 4 shows the match between sift points on a pair of images.

The SIFT matches are then introduced to a RANSAC-scheme algorithm to compute the fundamental matrix. This matrix will be needed further on to compute epipolar lines between images. Even though the publication talks about thresholding to separate static feature points from the dynamic ones, we make use of a given dynamic point for simplification reasons in order to perform the rest of the steps for the algorithm. Instead of making use of the matches output directly, we chose to compute the correspondent feature points of the van on each image by an auxiliary function which minimized the norm of the difference between feature points on each image, and the given feature point in image 1. This solution proved to more stable against small changes in sift detections throughout the tests. These correspondences found are the $idx\_car\_I2, I3 and I4$ respectively, and their location can be appreciated in Fig. 5.

Following this, comes the compute of the trajectory line for the van in image 1. This line is obtained as the cross product between 2 points belonging to it, one is the $idx\_car\_I1$ point and the other one is given as data. Figure 6 shows the trajectory

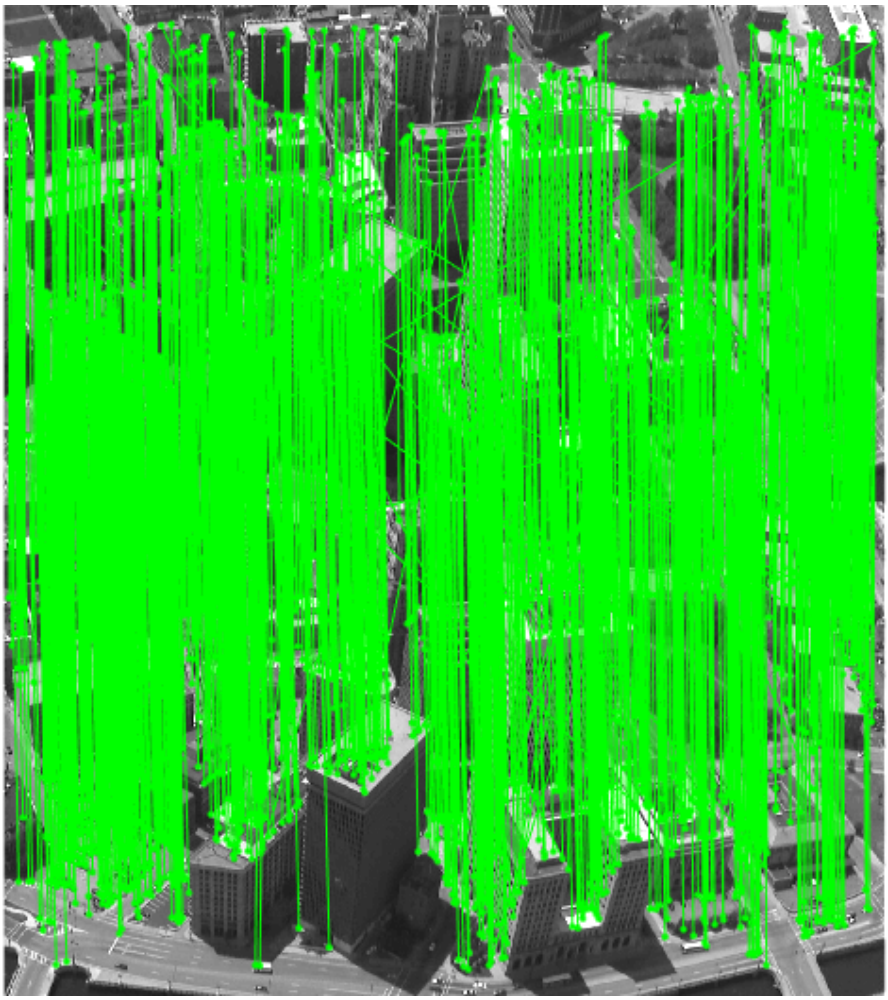

Figure 4: Matching of SIFT points between 2 images.

of the van.

Once defined this, we proceed to take the homogeneous coordinates of the $idx\_car\_I\#$ points –dynamic reference correspondences–, and compute the epipolar lines $l_i$ of each point $i$ in the reference image 1. The intersection between these lines and the trajectory line of the van, define the points which give temporal order to the image sequence as can be seen in the resulting plot of Fig. 7

It is worth of mentioning that in this case, the epipolar lines and the trajectory lines are too collinear. As stated in the reference paper, results are degenerated by this effect, and minor changes in the sift matches due to uncertainty may prevent the arrival to a correct sequence ordering.

## 5. Available code

All the code implemented in this project can be found in the following github repository: https://github.com/marccarne/M6project

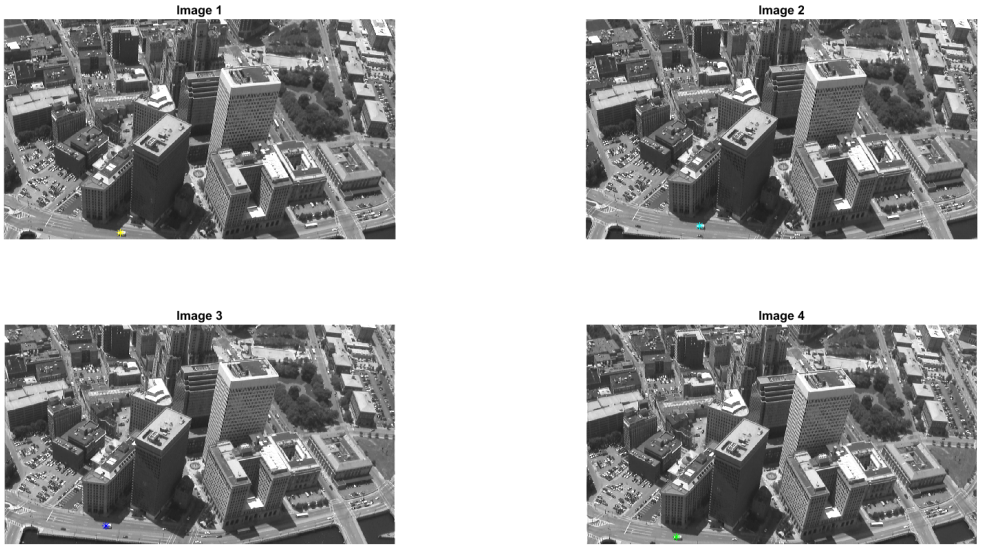


Figure 5: location of $idx$ points in each image.

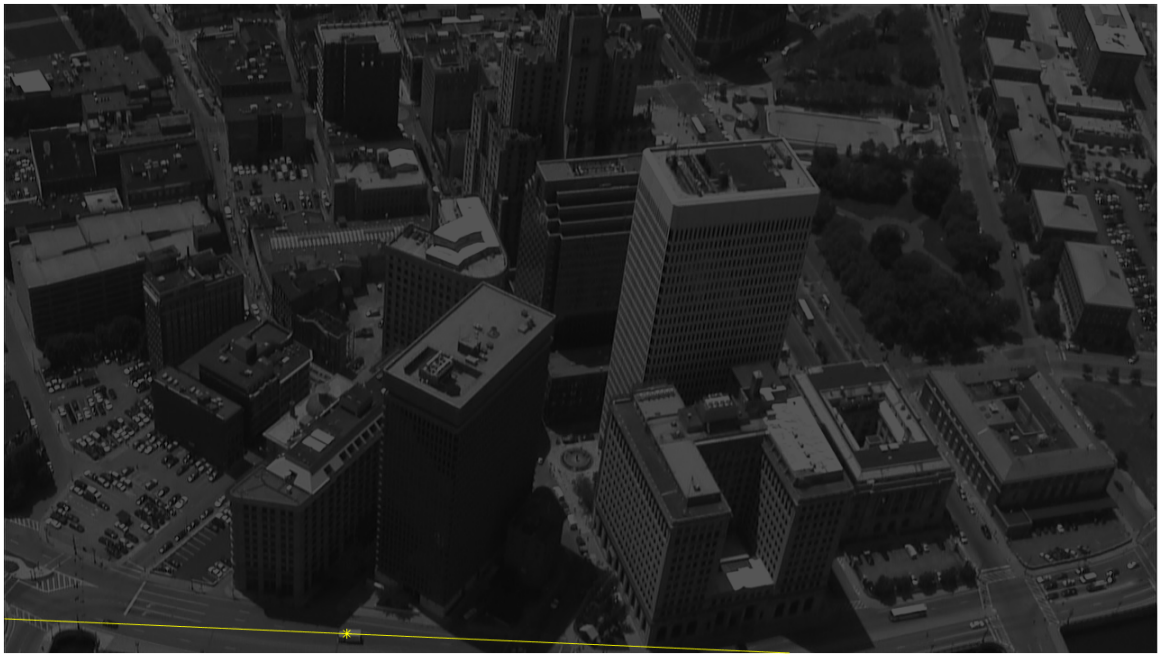

Figure 6: trajectory of the van defined by l1.

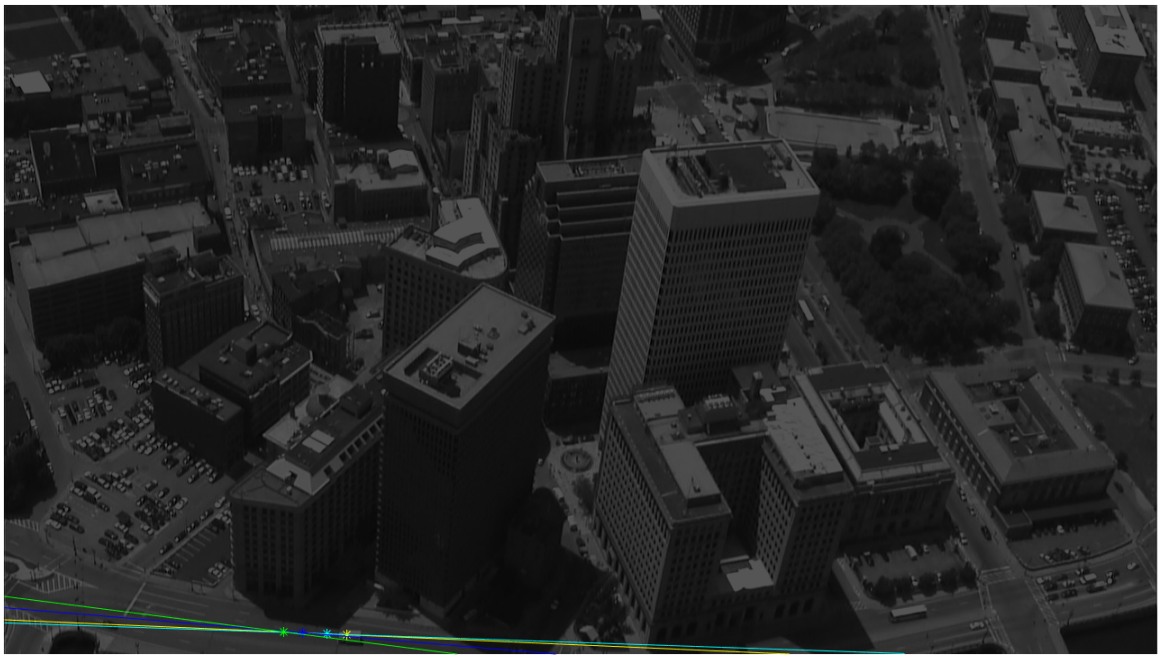

Figure 7: Temporal sequence of image given by the order of the projected points in the van trajectory as seen -yellow, cian, blue and last, green-.