PLANE DETECTION USING AFFINE HOMOGRAPHY

Kelson R. T. Aires*, Hélder de J. Araújo†, Adelardo A. D. de Medeiros*

*Departament of Computing Engeneering and Automation
Federal University of Rio Grande do Norte
Natal, RN, Brazil

†Institute of Systems and Robotics
Department of Electrical and Computer Engeneering
University of Coimbra
Coimbra, Portugal

Emails: kelson@dca.ufrn.br, helder@isr.uc.pt, adelardo@dca.ufrn.br

Abstract— Planes are important geometric features and can be used in a wide range of applications like robot navigation. This work aims to illustrate a homography-based method to detect planes using the affine model. Using two image frames from a monocular sequence, a set of match pairs of points is obtained using Harris corner detector combined with the Scale Invariant Feature Transform (SIFT) as local descriptor. An algorithm was developed to cluster interest points belonging to the same plane. Tests are performed in different sequences of outdoor images and results are shown.

Keywords— Robot Vision, Affine Homography, Plane Detection.

1 Introduction

Robots need dense, accurate and reliable information concerning the environment for safe navigation. There are many ways to provide this type of information to a robot. Images acquired by a camera, laser, sonar and infrared are examples of these. The cheapest way are vision-based systems. Vision-based robot navigation can be performed using one or more cameras. In the last case, a monocular vision-based system provides to the robot an image sequence. The extracted information from this image sequence allows the robot to self-locate and to know about the environment.

Planes are important geometric features and can be used in a wide range of applications like robot navigation (Okada et al., 2001) and camera calibration (Sturm and Maybank, 1999). Planar surfaces present within a scene can provide useful information for safe robot navigation. A pair of images captured by a stereo rig or a single moving camera can be used to extract such information.

Various algorithms for plane detection are found in the literature. Piazzi and Prattichizzo (2006) show how to compute, using stereo images, the normal vector to a plane by using only three corresponding points. Silveira et al. (2006) present a method for detecting multiple planar regions using a progressive voting procedure from the solution of a linear system exploiting the two-view geometry. Rodrigo et al. (2006) use SIFT [1] features to obtain the estimates of the planar homographies which represent the motion of the major planes in the scene. They track a combined Harris and SIFT feature using the prediction by these homographies. The approach of Okada et al. (2001) detects three-dimensional planar surfaces using 3D Hough transformation to extract plane segment candidates. Finally, Agarwal et al. (2005) present a detailed review and performance comparisons of planar homography estimation techniques.

This work presents a homography-based method to detect planes using the affine model. Using two image frames from a monocular sequence, a set of match pairs of points is required in order to estimate all planar homographies. Each homography represents a planar surface present in the scene. A matching process is performed to obtain this set using Harris corner detector combined with a local descriptor. We use the Scale Invariant Feature Transform (SIFT) as local descriptor. An algorithm was developed to cluster interest points belonging to the same plane. Tests are performed in four different sequences of outdoor images. Results are illustrated showing the detected planes.

The paper is organized as follows. Section 2 presents some aspects related to planar homography, highlighting the affine model. In Section 3, the proposed approach is formulated and in Section 4 the results are shown and discussed. Finally, Section 5 summarizes this work and some references are presented.

2 Planar Homography

When a planar object is imaged from multiple view-points the images are related by a unique homography. Given two views of the same plane $\pi = \begin{bmatrix}\mathbf{v}^T & 1\end{bmatrix}^T$, the ray corresponding to a point $\mathbf{x}$ in the image $\mathbf{I}$ meets the plane at a point $\mathbf{X}_\pi$, which projects into **x'** in image **I'**. The map from

[1] Scale Invariant Feature Transform

x to **x'** is the homography induced by the plane π. This is shown in Figure 1. Therefore, the homography is determined uniquely by the plane and vice versa (Hartley and Zisserman, 2004).

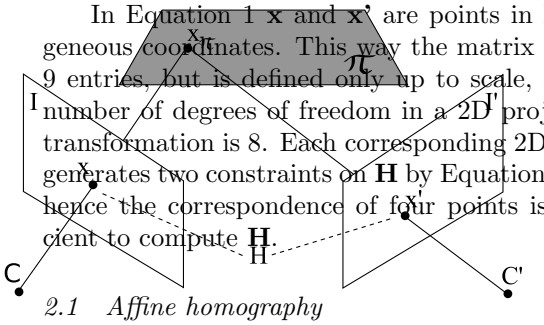

Figure 1: Given a planar object imaged from two view-points, the planar homography **H** is a mapping from points **x** of image **I** to points **x'** of image **I'**.

Given the projection matrices $\mathbf{P} = \begin{bmatrix}\mathbf{I}|0\end{bmatrix}^T$ and $\mathbf{P'} = \begin{bmatrix}\mathbf{A}|\mathbf{a}\end{bmatrix}^T$ for the two views, the homography induced by the plane is given by Equation 1.

$$\mathbf{x'} = \mathbf{H}\mathbf{x}, \tag{1}$$

with

$$\mathbf{H} = \mathbf{A} - \mathbf{a}\mathbf{v}^T$$

In Equation 1 **x** and **x'** are points in homogeneous coordinates. This way the matrix **H** has 9 entries, but is defined only up to scale, so the number of degrees of freedom in a 2D projective transformation is 8. Each corresponding 2D point generates two constraints on **H** by Equation 1 and hence the correspondence of four points is sufficient to compute **H**.

2.1 Affine homography

An affine homography $\mathbf{H}_A$ can be defined as non-singular linear transformation followed by a translation. The matrix representation is formulated as in Equation 2

$$\begin{bmatrix}x'\\y'\\1\end{bmatrix} = \begin{bmatrix}a_{11} & a12 & t_x\\a_{21} & a22 & t_y\\0 & 0 & 1\end{bmatrix} \cdot \begin{bmatrix}x\\y\\1\end{bmatrix} \tag{2}$$

or in block form as in Equation 3

$$\mathbf{x'} = \mathbf{H}_A\mathbf{x} = \begin{bmatrix}\mathbf{A} & \mathbf{t}\\\mathbf{0}^T & 1\end{bmatrix} \cdot \mathbf{x} \tag{3}$$

A planar affine homography has 6 degrees of freedom corresponding to the 6 matrix entries. Thus, the transformation can be computed from three point correspondences, i. e., the homography $\mathbf{H}_A$ can be computed from 3 matched pairs of points obtained using a pair of images captured by a stereo rig or a moving camera (monocular vision).

A modified version of *Direct Linear Transformation* method is used to estimate the homography. The last row of an affine homography $\mathbf{H}_A$ is equal to $\begin{bmatrix}0 & 0 & 1\end{bmatrix}$, thus the Equation 1 can be formulated in terms of an inhomogeneous set of linear equations as Equation 4

$$\begin{bmatrix}x_1 & y_1 & 1 & 0 & 0 & 0\\0 & 0 & 0 & x_1 & y_1 & 1\\x_2 & y_2 & 1 & 0 & 0 & 0\\0 & 0 & 0 & x_2 & y_2 & 1\\x_3 & y_3 & 1 & 0 & 0 & 0\\0 & 0 & 0 & x_3 & y_3 & 1\end{bmatrix} \cdot \begin{bmatrix}h_1\\h_2\\h_3\\h_4\\h_5\\h_6\end{bmatrix} = \begin{bmatrix}x'_1\\y'_1\\x'_2\\y'_2\\x'_3\\y'_3\end{bmatrix}, \tag{4}$$

or in block form

$$\mathbf{X} \cdot \mathbf{H}_A = \mathbf{X'}, \tag{5}$$

where $\mathbf{x}_i = \begin{bmatrix}x_i & y_i\end{bmatrix}$ and $\mathbf{x'}_i = \begin{bmatrix}x'_i & y'_i\end{bmatrix}$ are the match pairs of points in inhomogeneous coordinates and $\mathbf{H}_A$ has the form of Equation 6

$$\mathbf{H}_A = \begin{bmatrix}h_1 & h_2 & h_3\\h_4 & h_5 & h_6\\0 & 0 & 1\end{bmatrix}. \tag{6}$$

To estimate $\mathbf{H}_A$ from 3 matched pairs of points $\mathbf{x}_i \leftrightarrow \mathbf{x'}_i$, $i = 1, 2, 3$, we can solve the Equation 5 using the pseudo-inverse as

$$\mathbf{H}_A = (\mathbf{X}^T\mathbf{X})^{-1} \cdot \mathbf{X}^T\mathbf{X'} \tag{7}$$

whose computation is fast due to the low matrix dimensions.

2.2 Reprojection error

After an affine homography $\mathbf{H}_A$ has been computed, an error measure can be considered in order to verify whether a given matched pair of points belongs to the plane represented by $\mathbf{H}_A$. We use the *reprojection error* which can be defined as

$$e_i = |\mathbf{x}_i - \hat{\mathbf{x}}_i|^2 + |\mathbf{x'}_i - \hat{\mathbf{x}'}_i|^2, \tag{8}$$

where $\hat{\mathbf{x}'}_i = \mathbf{H}_A\mathbf{x}$ and $\hat{\mathbf{x}}_i = \mathbf{H}_A^{-1}\mathbf{x'}$.

When the reprojection error e_i is below a certain threshold, we consider that the given matched pair $\mathbf{x}_i \leftrightarrow \mathbf{x'}_i$ belongs to the plane represented by $\mathbf{H}_A$.

3 Plane Detection

A plane detection method is presented in this section. Interest points are detected using the Harris corner detector (Harris and Stephens, 1988) in two acquired images (stereo rig or monocular sequence), and a set of corresponding points between these images are created using a local descriptor like SIFT (Lowe, 2003). After this step the Delaunay triangulation is performed on the set of interest points of the first image. Using the set of triangles resulting from Delaunay method and

the set of matched pairs of points from combined Harris detector and SIFT descriptor, we calculate homographies and perform a clustering scheme to detect planes present in the imaged scene.

The entire methodology used is described in the following sections.

3.1 Interest points detection

Different primitives to detect and match image points exist in the literature (Mikolajczyk and Schmid, 2004). The Harris corner detector was chosen since it is fast, reliable and provides good repeatability under varying rotation and illumination. The Harris' method lies on the calculation of a matrix, M, using the partial derivatives of the intensity function

$$M = w \otimes \begin{pmatrix} \left(\frac{\partial I}{\partial x}\right)^2 & \left(\frac{\partial I}{\partial x}\right) \cdot \left(\frac{\partial I}{\partial y}\right) \\ \left(\frac{\partial I}{\partial x}\right) \cdot \left(\frac{\partial I}{\partial y}\right) & \left(\frac{\partial I}{\partial y}\right)^2 \end{pmatrix}, \quad (9)$$

where w specifies a gaussian window.

A corner is detected thresholding a measure based on the determinant and trace of the matrix M (equation 9).

$$\begin{aligned} M &= \begin{pmatrix} A & C \\ C & B \end{pmatrix} \\ det(M) &= AB - C^2 \\ trace(M) &= A + B \\ R &= det(M) - k(trace(M))^2 \end{aligned} \quad (10)$$

where k is constant and empirically adjustable, and R is called by *corner response*. Points with R above a certain threshold T_h are considered a corner.

In order to eliminate weak corners a non-maximal suppression procedure is performed in the neighborhood of a detected corner.

3.2 Local descriptor

Following the detection of interest points in the two consecutive images, a local descriptor must be used to establish a measure of correlation between the possible candidates to a matched pair of corners. This work uses the Scale Invariant Feature Transform descriptor. The combination of Harris detector with the SIFT descriptor can be considered as a good choice because it produces good and fast results (Mikolajczyk and Schmid, 2005).

The SIFT descriptor is a local descriptor highly distinctive and invariant to changes in illumination and 3D viewpoint. The descriptor is based on the gradient magnitude and orientation of all pixels in a region around the keypoint. These are weighted by a gaussian window and accumulated into orientation histograms summarizing the contents over subregions, as shown in Figure 2. The length of each arrow corresponds to the sum of the gradient magnitudes near that direction within the region.

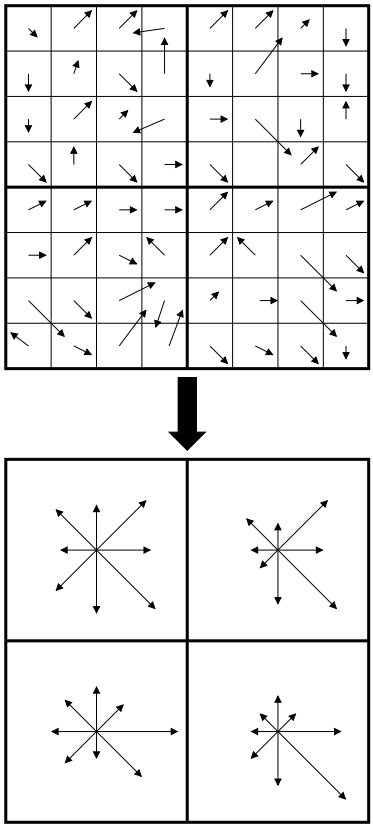

Figure 2: This figure shows a 2×2 descriptor array computed from an 8 × 8 set of samples. Each descriptor has 8 bins. The descriptor is represented by a vector of size 32 ($2 \times 2 \times 8$).

The orientation histogram entries corresponding to the lengths of the arrows in the bottom of Figure 2. The descriptor is formed from a vector containing all these entries.

The distance between histograms (SIFT descriptor) is used as measure of correlation. A simple Euclidean measure of distance is used. If the distance is below a certain threshold then a possible matched pair is detected.

3.3 Regressive corner filtering and matching

A regressive filtering scheme is used to establish a correspondence between the detected corners and finding matched pairs.

For each detected corner in the first image, a correspondent corner is searched for in a neighborhood in the second image satisfying a correlation measure described in Section 3.2. All corners in the first image, which do not have a correspondent in the second image are discarded. In the next step, a regressive filtering is performed to prevent that two or more corners of the first image have the same match in the second image. Now, for

every corner in the second image 2, only the best correlated corner in the first image is maintained. At the end of the filtering, a set of matching pairs of corners M is obtained from the two images.

Given M, a Delaunay triangulation is performed only on the detected corners of the first image. Since the set of triangles has been obtained, a clustering scheme must be applied to join triangles in the same plane and discard triangles belonging to virtual planes in the image.

3.4 Clustering points

The clustering scheme used is based on affine homography and reprojection error that allows obtaining planes and joining points in the same plan.

Given the set of matched points M and the set of Delaunay triangles T, we define H_p as the set of all p homographies existing between the two images. Each homography in H_p defines a plane in the image. Initially H_p is defined as empty. The first affine homography $\mathbf{H}_A$ is computed using the three points of the first triangle $T(1)$ and their matched pairs into the set M according to the Equation 7. The homography $\mathbf{H}_A$ obtained is included in H_p, and all used points are marked as *visited* and assigned to the homography $H_p(1)$, i.e., the first plane.

In the next step, the next triangle of T and their matched points in M is taken. For each $H_p(i)$ in H_p, all points of the triangle are verified if they belong to any of the existing planes in H_p. If the point had been marked as *not-visited* and the reprojection error for H_i is below a certain threshold, the point is marked as *visited* and assigned to the homography H_i. If the point had been marked as *visited* and if the new reprojection error for H_i is smaller than the old one, the point is assigned to the plane H_i. In the case where all points of the triangle do not belong to any existing plane, a new affine homography $\mathbf{H}_A$ is computed with those points. The new homography represents a new plane and is included in H_p. This loop is performed until there are no unvisited matched pairs of points.

The clustering method described before can be presented in the Algorithm 1.

At the end of clustering stage, only planes with a number of points above a certain threshold are considered.

Algorithm 1

```
1 Create an empty set of homographies H_p
2 M: set of matched points between two images
4 T: set of triangles of the first image
5 t: number of triangles in T
6 err: reprojection error
7 T_e: reprojection error threshold
Take M and T(1), compute H_A and put in H_p
Assigned the three points of the triangle T(1)
to H_p(1), and mark them as visited
for j = 2 : t do
  for i = each one in H_p do
    for each point p of T(j) do
      if p = not-visited then
        if err < T_e then
          Assign p to H_p(i)
          Mark p as visited
        end if
      else if p = visited then
        if err < err_old then
          Update p to H_p(i)
        end if
      end if
    end for
  end for
  if All p of T(j) = not-visited then
    Compute new H_A and put in H_p
    Assign all p to the new H_A and mark them
    as visited
  end if
end for
```

4 Experimental Results

In this section we report experiments that demonstrate accurate plane detection from two consecutive frames of an image sequence. Four image sequences of different outdoor scenes were acquired by a camera. All used images in the experiments are in gray level with size 640×480.

Different parameters were empirically adjusted at each stage of the entire process of plane detection. The following sections discuss each one of them.

4.1 Interest point detection

At this stage, the Harris detector was implemented using $k = 0.13$ and a gaussian window with standard deviation $\sigma = 1.5$ and size of $6\sigma \times 6\sigma$. A threshold value T_h was used to distinguish between corners and non-corners. T_h must be set high enough to avoid the detection of false corners which may have a relatively large corner response R (Equation 10) due to noise. The value of T_h is based on the maximum corner response computed, R_{max}. It was used $T_h = 0.01 \cdot R_{max}$. After all corners had been detected, a non-maximum suppression scheme was performed with a window of size 10×10, in order to eliminate weak corners.

4.2 Local descriptor

The best results were obtained using SIFT with 4×4 descriptors computed from a 16×16 sample array. The weighting gaussian window has size 16×16 and $\sigma = 1.5$. Each descriptor is described by an orientation histogram with 8 bins, each one

of size $\pi/4$. The total size of SIFT descriptor is 128 ($4 \times 4 \times 8$).

A simple Euclidean distance d was used to compare descriptors. If $d < T_m$ then we have a possible matched pair of interest points. It was initialized $T_m = 10$.

4.3 Regressive corner filtering and matching

In the regressive corner filtering and matching stage, we must perform a search in the second image for a corner that best matches a specific corner in the first image. This should be done for all detected corners in the first image. The size of the region of search can be constrained to improve the computational efficiency of the algorithm. This region can be defined based on the largest displacement of pixel, i.e., based on the camera's movement during the acquisition process. In our case a region around the interest point of size 25×25 was used.

The value of T_m is dynamically adjusted during the process of matching. When a possible match is found, the value of the threshold T_m is updated with the computed distance d. This was done to allow that the best candidate is chosen as a match.

4.4 Clustering points

Only one parameter is adjustable at this stage. A reprojection error threshold was empirically chosen as $T_e = 5.0$. The value of T_e is dynamically updated during the clustering process. Thus, each matched pair of points in the set M is assigned to the best homography in the set H_p based on the computed reprojection error.

4.5 Results

Four different sequences were used in the experiments. Sequences with planes at different positions and orientations in the scene were chosen. Figure 3 shows two frames of a sequence.

(a) Frame 01 (b) Frame 02

Figure 3: Two consecutive frames of a sequence used in experiments.

The planes detected are shown using colored markers in the first frame of each sequence. The results for the all sequences are shown in Figures 4, 5, 6 and 7.

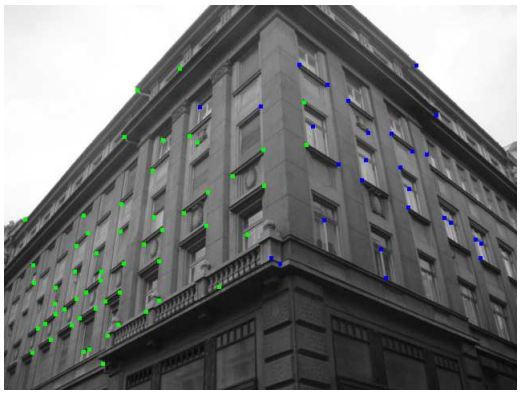

Figure 4: Colored markers specify the detected planes.

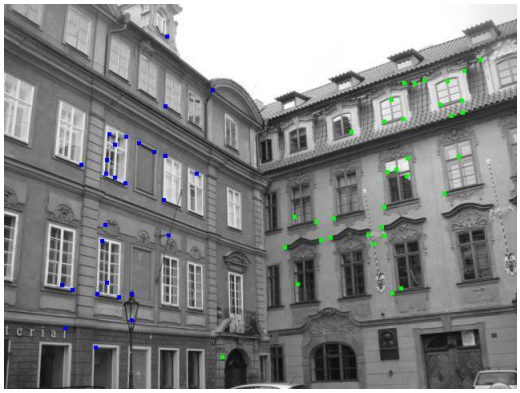

Figure 5: Colored markers specify the detected planes.

4.6 Discussion

The sequences were acquired under different conditions of illumination and different positions of the camera relative to the planes in the scene. The best results were obtained using Harris corner detector combined with SIFT descriptor.

The matching process is one of the most important stages of the algorithm. An important parameter is the size of the region in which a match is searched for.

Each one of the four sequences presents two major planes of a scene. The results clearly show that the algorithm performs well using outdoor images.

5 Conclusion

In this paper we present a homography-based method for detecting planes. The affine model was chosen for simplicity. In order to obtain a set of matched points between two frames of an image sequence, the Harris corner detector combined with SIFT descriptor and a regressive filtering method was used. The algorithm was tested using outdoor image sequences and presented good results.

Figure 6: Colored markers specify the detected planes.

Figure 7: Colored markers specify the detected planes.

Future work will be concentrated at joining affine homography and affine optical flow to improve the precision and reliability under a wide range of conditions.

References

Agarwal, A., Jawahar, C. V. and Narayanan, P. J. (2005). A survey of planar homography estimation techniques, *Technical Report IIIT/TR/2005/12*, Centre for Visual Information Technology.

Harris, C. and Stephens, M. (1988). A combined corner and edge detection, *Proceedings of The Fourth Alvey Vision Conference*, pp. 147–152.

Hartley, R. and Zisserman, A. (2004). *Multiple View Geometry in Computer Vision*, second edn, Cambridge University Press.

Lowe, D. (2003). Distinctive image features from scale-invariant keypoints, *International Journal of Computer Vision*, Vol. 20, pp. 91–110.

Mikolajczyk, K. and Schmid, C. (2004). Scale and affine invariant interest point detectors, *International Journal of Computer Vision* **60**(1): 63–86.

Mikolajczyk, K. and Schmid, C. (2005). A performance evaluation of local descriptors, *IEEE Transactions on Pattern Analysis & Machine Intelligence* **27**(10): 1615–1630.

Okada, K., Kagami, S., Inaba, M. and Inoue, H. (2001). Plane segment finder: Algorithm, implementation and applications, *IEEE International Conference on Robotics and Automation*, Vol. 2, IEEE Computer Society, pp. 2120–2125.

Piazzi, J. and Prattichizzo, D. (2006). Plane detection with stereo images, *IEEE International Conference on Robotics and Automation*, IEEE Computer Society, pp. 922–927.

Rodrigo, R., Chen, Z. and Samarabandu, J. (2006). Feature motion for monocular robot navigation, *Information and Automation, 2006. ICIA 2006. International Conference on*, pp. 201–205.

Silveira, G., Malis, E. and Rives, P. (2006). Real-time robust detection of planar regions in a pair of images, *International Conference on Intelligent Robots and Systems*, IEEE Computer Society, pp. 49–54.

Sturm, P. F. and Maybank, S. J. (1999). On plane-based camera calibration: A general algorithm, singularities, applications, Vol. 1, pp. 432–437.