Detection of Planar Regions with Uncalibrated Stereo using Distributions of Feature Points

Yasushi Kanazawa Hiroshi Kawakami
Department of Knowledge-based Information Engineering
Toyohashi University of Technology, Toyohashi 441-8580, JAPAN
kanazawa@tutkie.tut.ac.jp

Abstract

We propose a robust method for detecting local planar regions in a scene with an uncalibrated stereo. Our method is based on random sampling using distributions of feature point locations. For doing RANSAC, we use the distributions for each feature point defined by the distances between the point and the other points. We first choose a correspondence by using an uniform distribution and next choose candidate correspondences by using the distribution of the chosen point. Then, we compute a homography from the chosen correspondences and find the largest consensus set of the homography. We repeat this procedure until all regions are detected. We demonstrate that our method is robust to the outliers in a scene by simulations and real image examples.

1 Introduction

RANSAC [2] and LMedS [12] are very powerful methods for estimating parameters in images and very robust to outliers in data. So, recently, many methods based on them are proposed [13, 14, 15]. In their procedures, we usually use an uniform distribution for sampling data. It is reasonable when we want to estimate global parameters over images. For example, for estimating the fundamental matrix of an image pair, RANSAC and LMedS work very well. For estimating a homography to make a panoramic image, they also work well. Because these matrices are global parameters between the two images.

However, if we want to estimate local parameters in a scene, such as the homography for a small planar region, RANSAC or LMedS with an uniform distribution does not works well. Because the probability of chosen four matches being on the same plane is very small. Therefore, we need many iterations for estimating such local parameters. But we often obtain the homography for a non-existing plane. So, we need some knowledge about the existing planes in the scene.

Dick et al. [1] use the knowledge about the buildings in a scene. They use the perpendicularity of the walls to the ground for finding existing planes. But, this knowledge is not available in a general scene. Kanazawa et. al [9] use the knowledge that the points on the same plane make a cluster in an image. They first put the circumscribed rectangle including all feature points in one image and recursively divide the rectangle into sub-rectangles vertically or horizontally using geometric AICs [5]. They repeat this procedure until each sub-rectangle is regarded as a planar region. It is robust but it detects too many regions. So, they merge the regions by using geometric AICs again. Zucchelli et al. [18] propose

BMVC 2004 doi:10.5244/C.18.27

the method for detecting planar motion flow using LMedS. However, their method does not work well when outliers are in the planar regions. They first choose one point by random sampling and then choose the nearest four points of the first one. Then, they grow the planar regions with the parameters chosen by LMedS. Matas et al. [11] first detect small planar regions called MSER (maximally stable extremal regions) and then find matches between the regions. Trucco et al. [16] find the plane parameter from from disparity space with parallel stereo.

Once we detect two or more planar regions in a scene and compute the homographies for them, we can compute the fundamental matrix from the homographies using compatibility of them [4]. Luong et al. [10] have shown that the fundamental matrix computed from homographies is not accurate, but Kanazawa et al. [9] show that the fundamental matrix computed from the homographies is more accurate than that computed from point matches if we optimally compute the homographies. So, detection of local planar regions in a scene is very important task for reconstructing 3-D structure of the scene.

In this paper, we propose a robust method for detecting local planar regions in a scene with an uncalibrated stereo. Our method is based on random sampling using distributions of feature point locations. For doing RANSAC procedure, we use the distributions for each feature point defined by the distances between the point and the other points. We first choose a correspondence by using an uniform distribution and next choose candidate correspondences by using the distribution of the chosen point. Then, we compute a homography from the chosen correspondences and find the largest consensus set of the homography. We repeat this procedure until sufficient number of regions are detected. We demonstrate that our method is robust to the outliers in a scene by simulations and real image examples.

2 Compatibility of Fundamental Matrix and Homographies

We take the first camera as a reference coordinate system and place the second camera in a position obtained by translating the first camera by vector $\mathbf{t}$ and rotating it around the center of the lens by matrix $\mathbf{R}$. We call $\{\mathbf{t},\mathbf{R}\}$ the *motion parameters*. We assume that the two cameras are modeled by the pinhole camera and their focal lengths are f and f'.

Let (x,y) be the image coordinates of a 3-D point P projected onto the image plane of the first camera, and (x',y') be those for the second camera. We use the following three-dimensional vectors to represent them (the superscript $\top$ denotes transpose):

$$\mathbf{x} = \left(x/f_0, y/f_0, 1\right)^\top, \quad \mathbf{x}' = \left(x'/f_0, y'/f_0, 1\right)^\top. \tag{1}$$

Here, f_0 is a scale factor for stabilizing computation.

We regard vectors $\mathbf{x}_\alpha$ and $\mathbf{x}'_\alpha$ as the projected images of 3-D points P_α, $\alpha = 1$, ..., N. Here, the vectors $\mathbf{x}_\alpha$, $\mathbf{R}\mathbf{x}'_\alpha$, and $\mathbf{t}$ are coplanar, so they satisfy the following *epipolar equation* [4, 5]:

$$(\mathbf{x}_\alpha, \mathbf{F}\mathbf{x}'_\alpha) = 0. \tag{2}$$

Here, $(\mathbf{a},\mathbf{b})$ denotes the inner product of vectors $\mathbf{a}$ and $\mathbf{b}$. The matrix $\mathbf{F}$ is a singular matrix of rank 2 and is called the *fundamental matrix*.

When 3-D points P_β, $\beta = 1$, ..., M, lie on a plane Π, the vectors $\mathbf{x}_\beta$ and $\mathbf{x}'_\beta$ are related in the following form [4, 5]:

$$\mathbf{x}'_\beta = Z[\mathbf{H}\mathbf{x}_\beta]. \tag{3}$$

Here, $Z[\cdot]$ designates a scale normalization to make the third component 1. The matrix $\mathbf{H}$ is a nonsingular matrix and is called the *homography*. The vectors $\mathbf{x}_\beta$ and $\mathbf{x}'_\beta$ must satisfy Eqs. (2) simultaneously, so we obtain the following equation:

$$(\mathbf{x}_\alpha, \mathbf{F}\mathbf{H}\mathbf{x}_\alpha) = 0. \tag{4}$$

This equation means that the matrix product $\mathbf{FH}$ must be a skew-symmetric matrix, which satisfies

$$\mathbf{F}\mathbf{H} + \mathbf{H}^\top \mathbf{F}^\top = \mathbf{O}. \tag{5}$$

This time, the homography $\mathbf{H}$ is *compatible* to the fundamental matrix $\mathbf{F}$ [4].

When we detect two or more planar regions in a scene, we can compute the fundamental matrix $\mathbf{F}$ from the homographies for them using the compatibility (5). If we compute the homographies by an optimal method [7], the fundamental matrix obtained from the homographies are more accurately than that computed from the point matches [9].

3 Detection of Planar Regions

We first detect feature points in each image by a feature detector, such as Harris operator [3], and then find point matches over two image by an automatic matching program, such as [8] or [17].

In order to compute the homographies for local planar regions in a scene, we take a double random sampling scheme; the first sampling with the uniform distribution is used for choosing a local area in the image; the second sampling with the probabilities we here defined is used for choosing four points almost in the local area. By introducing this scheme, we can easily choose four points in the same plane.

3.1 Definition of Probabilities for Feature Points

Let P_λ, $\lambda = 1, ..., N$, be the feature points in an image I and (x_λ, y_λ) be their coordinates. The Euclidean distance $d_{\alpha\beta}$ between two points P_α and P_β is defined by the following equation.

$$d_{\alpha\beta} = \sqrt{(x_\alpha - x_\beta)^2 + (y_\alpha - y_\beta)^2}. \tag{6}$$

Using the distance $d_{\alpha\beta}$, we define the conditional probability $p(\beta|\alpha)$ by

$$p(\beta|\alpha) = \begin{cases} \frac{1}{Z_\alpha} e^{-s_\alpha d_{\alpha\beta}^2}, & \alpha \neq \beta \\ 0, & \alpha = \beta \end{cases}, \tag{7}$$

which indicates the probability of the point P_α under the point P_β being chosen. Here,

$$Z_\alpha = \sum_{\beta \neq \alpha} e^{-s_\alpha d_{\alpha\beta}^2}. \tag{8}$$

The parameter s_α is determined by solving the following equation.

$$\sum_{\beta=1}^{N} (d_{\alpha\beta} - \bar{d_\alpha}) e^{-s_\alpha d_{\alpha\beta}^2} = 0, \tag{9}$$

where

$$\bar{d_\alpha} = \frac{1}{N}\sum_{\beta=1}^{N} d_{\alpha\beta}. \tag{10}$$

This means that we determine s_α to satisfy the identity $\sum_\beta^N p(\beta|\alpha)d_{\alpha\beta} = \bar{d_\alpha}$.

If a point P_β is near a point P_α, the conditional probability $p(\beta|\alpha)$ is high. If not, $p(\beta|\alpha)$ is low. So, if we first choose the point P_α by an uniform distribution, we can choose a close point P_β to P_α efficiently by using $p(\beta|\alpha)$.

Here, in order to use the probabilities $p(\beta|\alpha)$ in RANSAC procedure, we introduce an index array $k_\alpha(\mu)$ and initialize them by

$$k_\alpha(\mu) = \mu, \quad \mu = 1, ..., N. \tag{11}$$

We sort $p(k_\alpha(\beta)|\alpha)$ in descending-order for each P_α. Then, by the following equation, we define the conditional cumulative probabilities $q(\beta|\alpha)$ when P_α is chosen.

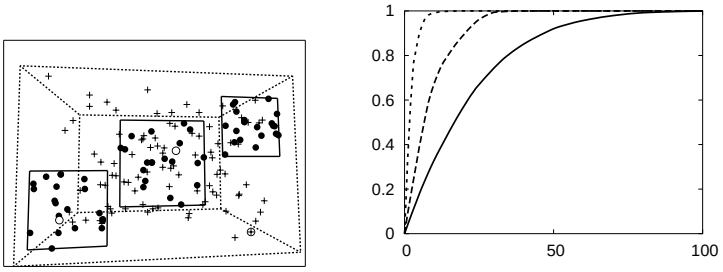
$$q(\beta|\alpha) = \sum_{\mu=1}^{\beta} p(k_\alpha(\mu)|\alpha). \tag{12}$$

3.2 RANSAC using $p(\beta|\alpha)$

Let P'_α in the image I' be the feature point that correspond to P_α in the image I. The RANSAC using the conditional probabilities (7) is done by the following procedure.

1. Let α be a pair chosen by using the uniform distribution from the set of the corresponding feature points $\mathscr{S}$=$\{P_\lambda, P'_\lambda\}$, λ=1, ..., N.
2. Initialize $\mathscr{S}_\alpha^{\max}$=$\phi$, $M_{\max}^\alpha$=0 and $\mathbf{H}_{\max}^\alpha$ = $\mathbf{O}$.
3. Using the conditional probabilities $p(\beta|\alpha)$, choose four pairs β_1, β_2, β_3, β_4. In order to do random sampling using $p(\beta|\alpha)$, first generate one random number r in the range $[0,1)$ using the uniform distribution, then, increase β from 1 and find β satisfied
$$r \le q(\beta|\alpha). \tag{13}$$
4. If the chosen four pairs are not-skewed, compute the homography $\mathbf{H}_\alpha$ from them. If the pairs are skewed, return the step 3 and choose four pairs again.
5. Compute the reprojection errors for all pairs. Let $\mathscr{S}_\alpha$ be the set of the pairs which errors are less than the threshold d and M_α be the number of the elements of $\mathscr{S}_\alpha$.
6. If $M_\alpha \ge M_\alpha^{\max}$, update
$$\mathscr{S}_\alpha^{\max} = \mathscr{S}_\alpha, \quad M_\alpha^{\max} = M_\alpha, \quad \mathbf{H}_\alpha^{\max} = \mathbf{H}_\alpha. \tag{14}$$
Otherwise, return the step 1.
7. Repeat the above computation until $M_\alpha^{\max}$ reaches its maximum[1].
8. If $M_\alpha^{\max}$ is sufficiently large[2], compute the homography $\hat{\mathbf{H}}_\alpha$ from $\mathscr{S}_\alpha^{\max}$ optimally [7]. Then, update $\mathscr{S} - \mathscr{S}_\alpha \to \mathscr{S}$ and return the step 1.

[1] In our experiment, we stopped the search when no update occurred 100 times consecutively.
[2] In our experiment, we use the threshold is 10.



(a) (b)

Figure 1: Examples of conditional cumulative probabilities $q(\beta|\alpha)$ for (a): the horizontal axes are the index number of the sorted points and the vertical axes is probabilities. (a) Example image. (b) Solid line: probabilities for the point $\bigcirc$ in the central plane. Dashed line: probabilities for the point $\bigcirc$ in the left plane. Dotted line: probabilities for the point $\oplus$ in left bottom.

4 Simulations and Real Image Examples

4.1 Conditional cumulative probabilities

We first illustrate the curves of some conditional cumulative probabilities by simulation. We take three planes in a space and place 20 points randomly on each plane. They are marked by circles (filled and not filled) in Fig. 1(a). We also place 90 points that are not on these planes randomly in the space as outliers. These are marked by plus in Fig. 1(a).

The conditional cumulative probabilities of three points to the other points are shown in Fig. 1(b). In this figure, the solid curve indicates the conditional cumulative probability of the point marked by circle in the center plane; the dashed curve indicates that of the point marked by circle in the left plane; the dotted curve indicates that of the point marked by plus with circle. We can see that the probability of the point in dense area has gently slope and the probability of the point in sparse area has steeply slope. So, the slope becomes steeply as the density decreasing. We see that we can efficiently choose close points to the first chosen point through the conditional probability for the chosen point.

In addition, we see that we can also apply the probability for choosing the first point; we define a probability for choosing the first point by the angle of the curve of the conditional cumulative probability. Generally, the number of sampling times in RANSAC procedure is determined by the least number of data to compute a model and the percentage of outliers in data [4]. If we define the probability for choosing the first point using the angle of its conditional probabilities, we can reduce the sampling times in RANSAC and then reduce the cost of the computation.

4.2 Simulations

We next illustrate the effectiveness of our method by doing detection of planar regions in some simulated scenes. We do simulations in various number of planes and various size of planes in the scene. The results are shown in Fig.2.

We first simulate scenes that have from one to five planes. We place 150 points in the scene; 20 points are on each plane and the others are in the space as outliers. After projecting the points onto the image planes of two cameras, we add Gaussian random noise with standard deviation 0.0 to 0.4 (pixels) to each coordinates independently. Fig. 2(a) shows an example image of the scenes and Fig. 2(c) shows the number of the detected

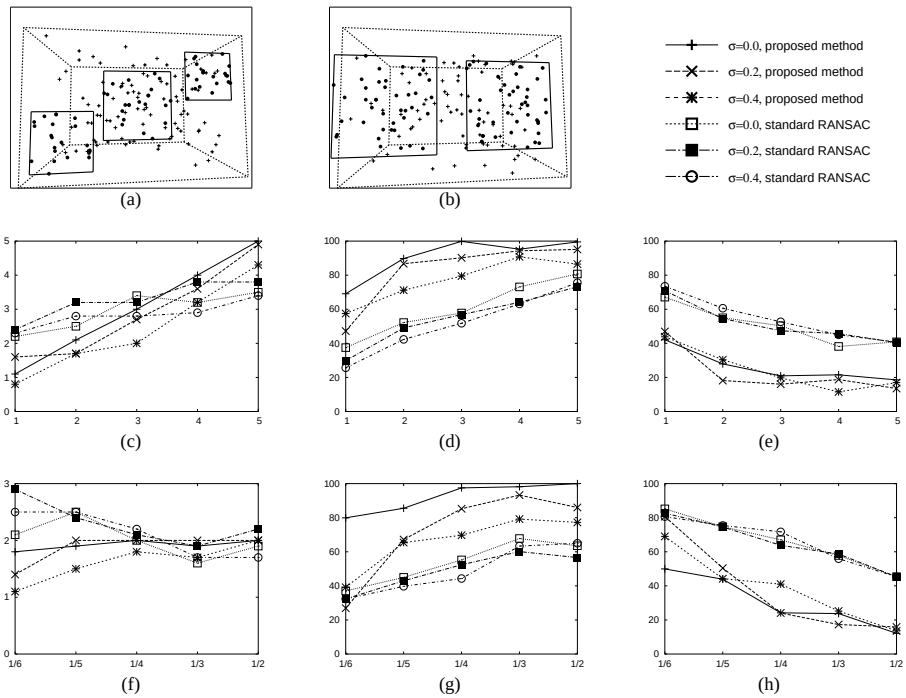


Figure 2: Simulation for detecting planar regions. Upper: example images of simulated images. (a) Example image at varying in the number of planes. (b) Example image at varying in heights of planes. Middle: results at varying in the number of planes. (c) Number of detected planes. (d) Detection rates of correct points. (e) Detection rates of wrong points. Lower: results at varying in lengths. (f) Number of detected planes. (g) Detection rates of correct points. (h) Detection rates of wrong points.

planes by our method and the standard RANSAC, for comparison. In this graph, the horizontal axis indicates the number of planes in the scenes and the vertical axis indicates the number of detected planes. Here, when 10 or more points on a plane are detected, we regard that plane is detected successfully. We can see that the standard RANSAC detects non-existing planes when the number of planes is small; for example, when the number of planes is 2, but the number of the detected planes is larger than 2. However, our method detects almost the correct number of planes. The detection rate of the correct points P_c and the detection rate of the wrong points P_e are defined by the following equations.

$$P_c = \frac{\text{\# of detected correct points}}{\text{\# of correct points}}, \quad P_e = \frac{\text{\# of detected error points}}{\text{\# of detected points}} \tag{15}$$

In Fig. 2(d) and (e), the horizontal axes indicate the number of the planes and the vertical axes indicate these rates. We can see that our method can detect more correct points on the planes than the standard RANSAC.

We next simulate scenes that have two planes which sizes are 1/6 to 1/2 of the height of the image. Fig. 2(b) shows an example image of the scenes and Fig. 2(f), (g), and (h) show the number of the detected planes, the rate of the detected correct points, and the rate of the detected error points respectively. We can also see that the standard RANSAC de-

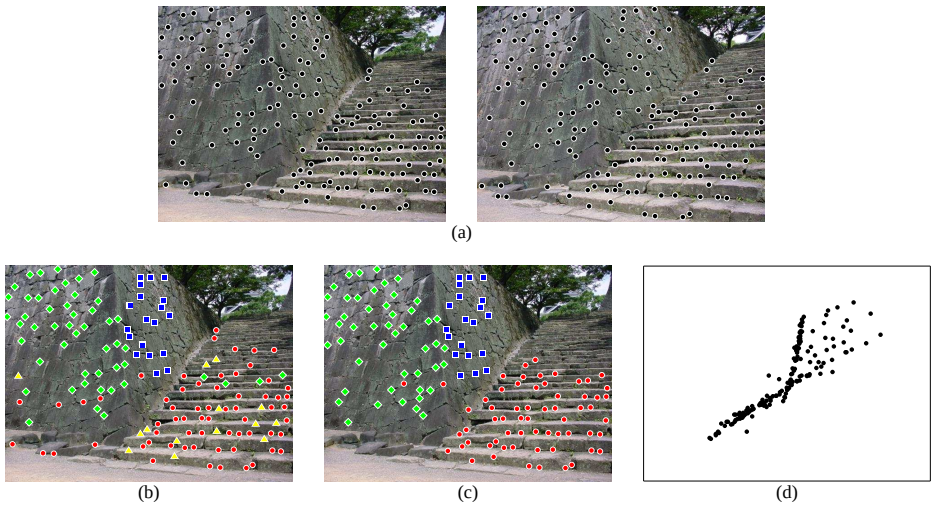


Figure 3: Real image examples for detecting planar regions. (a) Real images. (b) Planes detected by the standard RANSAC. (c) Planes detected by our method. (b) 3-D reconstructed shape (side view) using the fundamental matrix computed from homographies of the planes detected by our method.

tects non-existing planes when the planes are small but our method detects almost correct planes in every size.

4.3 Real Image Examples

We finally show some real image examples.

Fig. 3(a) shows a stereo image pair of an outdoor scene. We determine point matches by the automatic matching method [8]. Fig. 3(b) shows the result obtained by the standard RANSAC and Fig. 3(c) shows the result obtained by our method. In these figures, we plot the points on the same plane by the same marks, for example, circles plotted in the image are regarded be on the same plane. We can see that the standard RANSAC detects a wrong planes (stone steps). However, our method detects correct three planes and the points on each plane are almost correct. We compute the fundamental matrix from the obtained homographies using the compatibilities (5) and decompose it into the camera parameters by the method of Kanatani and Matsunaga [6]. Using the homographies obtained by our method, we obtained the 3-D reconstruction shown by Fig. 3(d). We can see it is very accurate. However, we could not decompose the fundamental matrix from the homographies detected by the standard RANSAC. We could not decompose the fundamental matrix computed from point matches either. So, for obtaining an accurate 3-D reconstruction, it is very important to detect planar regions accurately in the scene.

Fig. 4(a) shows an example for a building scene. Fig. 4(b) and (c) show the planes detected by the standard RANSAC and our method, respectively. In this example, we can see that the standard RANSAC detects wrong planes and points, but our method detects almost correct planes and points. Fig. 4(d) shows the 3-D shape reconstructed from the fundamental matrix computed from point matches. Fig. 4(e) shows the 3-D shape reconstructed from the fundamental matrix computed from the homographies detected by the standard RANSAC. Fig. 4(f) shows the 3-D shape reconstructed from the fundamental

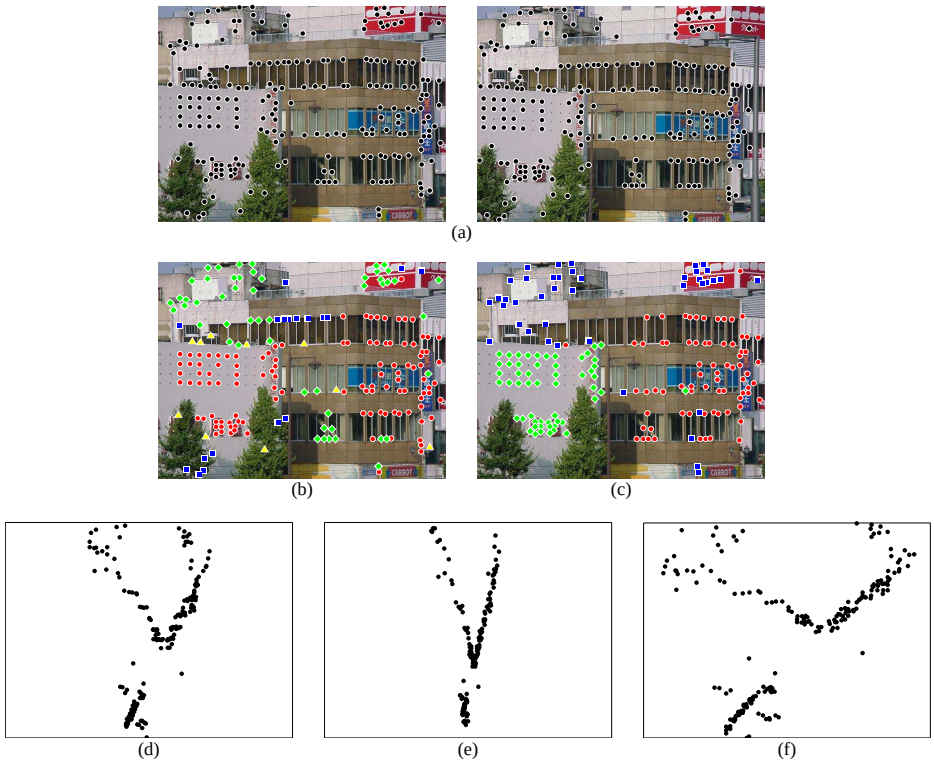


Figure 4: Real image examples for detecting planar regions. Top: (a) Real images. Middle: Detected planes. (b) The standard RANSAC. (c) Our method. Bottom: 3-D reconstructed shape (top view). (d) Using the fundamental matrix computed from point matches. (e) Using the fundamental matrix computed from homographies of the planes detected by the standard RANSAC. (f) Using the fundamental matrix computed from homographies of the planes detected by our method.

matrix using the homographies detected by our method. We can see that the 3-D shape obtained from homographies detected by our method is more accurate than the others.

Fig.5 shows an example for another building scene. We can see that our method detect almost correct planar regions, but the standard RANSAC detect wrong planes. So, we can see the resulted 3-D shape from the fundamental matrix using the homographies by the standard RANSAC is wrong, for example, the red signboard on the upper left has been reconstructed in back of the building on the right. We can also see the 3-D reconstruction obtained from the fundamental matrix using point matches is also wrong.

5 Conclusion

We proposed a robust method for detecting local planar regions in a scene with an uncalibrated stereo. Our method is based on random sampling using distributions of feature point locations. For doing RANSAC procedure, we use the distributions for each feature point defined by the distances between the point and the other points. Using these probabilities we can detect planar regions robustly and efficiently. We demonstrated that our method is robust to the outliers in the scene by simulations and real image examples.

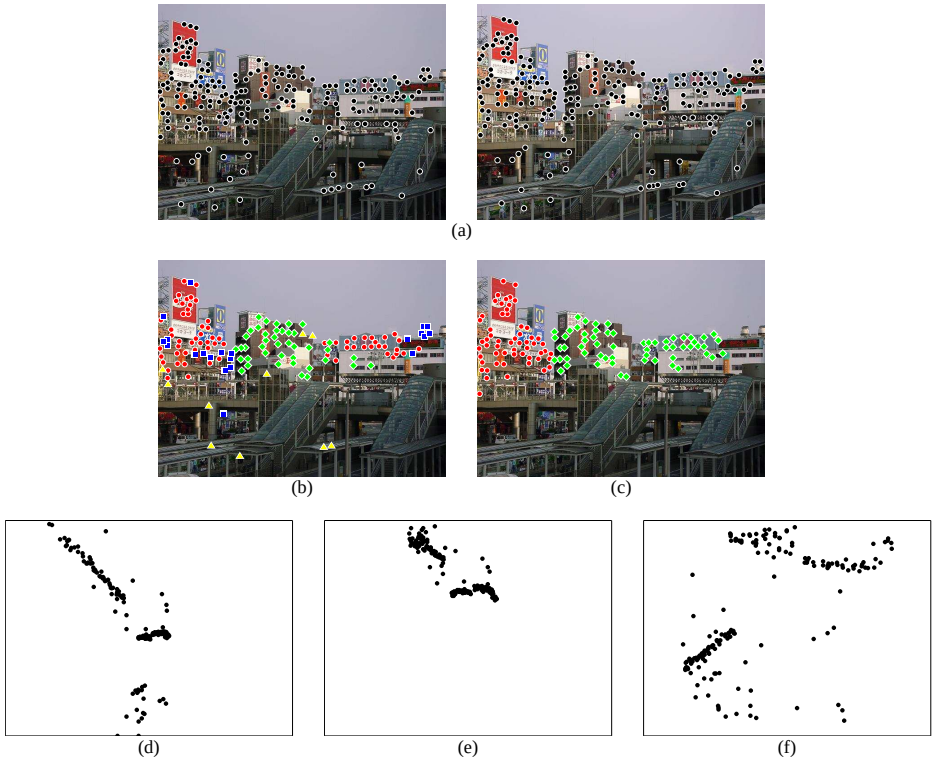


Figure 5: Real image examples for detecting planar regions. Top: (a) Real images. Middle: Detected planes. (b) The standard RANSAC. (c) Our method. Bottom: 3-D reconstructed shape (top view). (d) Using the fundamental matrix computed from point matches. (e) Using the fundamental matrix computed from homographies of the planes detected by the standard RANSAC. (f) Using the fundamental matrix computed from homographies of the planes detected by our method.

In future work, for obtaining more accurate correspondences between two images, we will build our method into image matching.

Acknowledgement

This work was supported in part by the Ministry of Education, Culture, Sports, Science and Technology, Japan, under the Grant for 21st Century COE Program “Intelligent Human Sensing”.

References

[1] A. Dick, P. Torr, and R. Cipolla. Automatic 3d modeling of architecture. In *Proc. 11th British Machine Vision Conf.*, pages 372–381, Bristol, U.K., September 2000.

[2] M.A. Fischler and R.C. Bolles. Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography. *Comm. ACM*, 24(6):381–395, 1981.

[3] C. Harris and M. Stephens. A combined corner and edge detector. In *Proc. 4th Alvey Vision Conf.*, pages 147–151, Manchester, U.K., August 1988.

[4] R. Hartley and A. Zisserman. *Multiple View Geometry*. Cambridge University press, Cambridge, 2000.

[5] K. Kanatani. *Statistical Optimization for Geometric Computation: Theory and Practice*. Elsevier Science, Amsterdam, 1996.

[6] K. Kanatani and C. Matsunaga. Closed-form expression for focal lengths from the fundamental matrix. In *Proc. 4th Asian Conf. Comput. Vision*, pages 128–133, Taipei, Taiwan, January 2000.

[7] K. Kanatani, N. Ohta, and Y. Kanazawa. Optimal homography computation with a reliability measure. *IEICE trans. Inf. & Syst.*, E83-D(7):1369–1374, June 2000.

[8] Y. Kanazawa and K. Kanatani. Robust image matching preserving global consistency. In *Proc. 6th Asian Conf. Comput. Vision*, pages 1128–1133, Jeju Island, Korea, January 2004.

[9] Y. Kanazawa, T. Sakamoto, and H. Kawakami. Robust 3-d reconstruction using one or more homographies with uncalibrated stereo. In *Proc. 6th Asian Conf. Comput. Vision*, pages 503–508, Jeju Island, Korea, January 2004.

[10] Q.-T. Luong and O. Faugeras. Determining the fundamental matrix with planes: unstability and new algorithms. In *Proc. Comput. Vision Patt. Recog.*, pages 489–494, New York, U.S.A., June 1993.

[11] J. Matas, O. Chum, M. Urban, and T. Pajdla. Robust wide baseline stereo from maximally stable extremal regions. In *Proc. 13th British Machine Vision Conf.*, pages 384–393, Cardiff, U.K., September 2002.

[12] P. J. Rousseeuw and A.M. Leroy. *Robust Regression and Outlier Detection*. Wiley, New York, 1987.

[13] P.H.S. Torr and C. Davidson. Impsac: Synthesis of importance sampling and random sample consensus. *IEEE Trans. Patt. Anal. Mach. Intel.*, 25(3):354–364, 2003.

[14] P.H.S. Torr and A. Zisserman. Robust computation and parameterization of multiple view geometry. In *Proc. 6th Int. Conf. Computer Vision*, pages 727–732, Bombay, India, January 1998.

[15] P.H.S. Torr and A. Zisserman. Mlesac: A new robust estimator with application to estimating image geometry. *Comput. Vis. Image. Understand.*, 78:138–156, 2000.

[16] E. Trucco, F. Isgrò, and F. Bracchi. Plane detection in disparity space. In *Proc. IEE Int. Conf. Visual Information Engineering*, pages 73–76, Guildford, U.K., July 2003.

[17] Z. Zhang, R. Deriche, O. Faugeras, and Q.-T. Luong. A robust technique for matching two uncalibrated images through the recovery of the unknown epipolar geometry. *Artif. Intell.*, 78:87–119, 1995.

[18] M. Zucchelli, J. Santos-Victor, and H.I. Christensen. Multiple plane segmentation using optical flow. In *Proc. 13th British Machine Vision Conf.*, pages 313–322, Cardiff, U.K., September 2002.