3D Corner Detection and Matching for Manmade Scene/Object Structure Cognition

Jiao Tian and Derek Molloy

Centre for Image Processing and Analysis, City University, Dublin, Ireland

Keywords: 3D Corner, Plane Decomposition, Feature Grouping, Structure Cognition.

Abstract: In this paper, we describe a novel framework for 3D corner detection and matching. The proposed method is based on the assumption that the viewed scene contains definite planar surfaces. The contribution of our method is the integration of constraints imposed by the existing planes and the local feature matches to achieve improved plane decomposition and also optimal feature grouping. We describe the foundation of the framework and show how it can be employed in applications including 3D reconstruction, plane extraction and robot navigation. The effectiveness of our framework is validated through experimentations on synthetic 3D object and real architecture images.

1 INTRODUCTION

In 2D images, 3D corners usually have similar visual patterns as 3-junctions (the number of wedges, such as *T* or *Y* junctions) and they locate on the intersection region between planes. In existing junction/corner detection algorithms, none of them can appropriately tell which particular junction is a 3D corner among the detected junctions. In feature matching algorithms, 3D corners are often eliminated from the potential feature sets for matching because their visual appearances usually change a lot as the viewpoint shifts. Although the determination and matching of 3D corners is difficult, it is still an issue worth investigating as the 3D corners contain extra structure information which is useful in structure analysis related problems, such as geometric reasoning in image spatial layout analysis (Lee et al., 2009), and structure and motion estimation (Liu et al., 2003) in image feature based applications.

2 BACKGROUND

Our approach is a joint approach which combines spatial layout analysis, local feature grouping, and 3D corner detection and matching. In the proposed approach, these three parts are designed and expected to complement each other.

2.1 Image Spatial Layout Analysis

The image spatial layout is very useful for many computer vision tasks, including recognition, navigation and single view 3D reconstruction, etc. The determination of the orientation that relates to different planes of the scene/object is an important step in spatial layout analysis. In the literature, there are two branches dealing with this issue: one is based on a *priori* learning procedures (Hedau et al., 2009) or fixed templates about the indoor spatial layout (Lee et al., 2009); the other is reliant on the information inferred from local features, such as the co-planarity (Yu et al., 2008). For the latter approaches, a batch of features indicating the same structure information will make the results of spatial layout analysis more convincing.

2.2 Feature Grouping Methods

Feature grouping/clustering is often related to multi-model-fitting. Many methods have been proposed for multi-model-fitting, such as the least square methods, Hough transform, PEARL (Isack and Boykov, 2012) and the most standard RANSAC like algorithms: multi-RANSAC (Zuliani et al., 2005) and J-linkage (Toldo and Fusiello, 2008**).** All of these methods can be used for planar surface detection. However, when the feature matches are unevenly distributed on the views of the same scene, in other words, if there is a dominant plane existing

in the scene, the deduced geometry model will be inaccurate. For this situation, if the number of models (rough plane decomposition) is known in advance, a good result can be achieved.

2.3 Corner/Junction Detection

Usually, corners/junctions are classified by their visual patterns, i.e. L-Junction, Y-Junction, T-Junction, Arrow-Junction, and X-Junction. Another corner categorisation, proposed by Trajkvoc (Trajkovic and Hedley, 1998), considers only two separate corner categories - geometric and texture. Geometric corners belong to the boundaries of objects in the image, where texture corners come from the textures of objects in the image. In our case, the 3D corners are geometric corners which appear as T or Y junction.

In the literature, only two approaches related to 3D corners are proposed. Liu et al. (Liu et al., 2003) illustrated an experimental method for the 3D corner matching based on line matching and geometric constraints (the edges intersect at one common point). Ding et al. (Ding et al, 2008) labels an end point of a line segments as a 3D corner if there are two sufficiently long lines converging to the other two vanishing points in a region near this point.

3 OUR APPROACH

The algorithm is as follows:

Algorithm.

1. Image pre-processing to reduce the number of textured corners while retaining the geometric corners
2. Junction detection
3. Plane decomposition:
   A) Image spatial layout analysis to get a rough plane decomposition, and a narrowed down search area for the 3D corner
   B) Detect feature matching between views by the ASIFT feature detection and matching algorithm. The multi model-fitting algorithm is then used for planar surface detection based on the obtained feature matches.
4. Run the ASIFT feature matching algorithm again on the decomposed plane pairs, and calculate the homography transformation between the views.
5. 3D corner selection and matching.
5.1. Screen out the 2D 3-junctions by the plane intersection areas information (calculated in step 3).
5.2. Eliminate the junctions only satisfying one homography.

3.1 Pre-processing

Trajkovic believes that geometric corners are more stable than texture corners (Trajkovic and Hedley, 1998), though this claim seems incorrect in recent feature detection and matching algorithms, such as SIFT (Lowe, 2004) and MSER (Matas et al., 2002), his idea about reducing the number of texture corners while retaining the majority of geometric corners detected in the image reminds us that a suitable image smoothing step is necessary to help remove texture corners before the junction detection step.

Recently, an effective image smoothing algorithm, based on L_0 gradient minimisation (Xu et al., 2011), is proposed for extracting prominent structures inside images. Their algorithm is exploited in our image pre-processing step to remove low-amplitude structures and globally preserve salient edges. By this means, a considerable rate of 2D non-geometrical corners will be excluded. As shown in Figure 1, the detected junction number is reduced in the smoothed image, particularly on the right side of the image.

3.2 Junction Detection

We modified the algorithm developed by Rimon Elias and Robert Laganière (Elias and Laganière, 2012) for junction detection. First, create two binary edge maps, one thick edge map and one thin edge map. The thick map is obtained by imposing a threshold on the gradient magnitude image, and the thin map is obtained by a non-maxima suppression process on the thick map. Circular masks are centred at the potential junction, and the radial lines are scanned in the masks to determine the presence of the junctions. Because 3D corners appear as 3-junctions, we modified the algorithm to only detect the junctions where three edges meet (Figure 1).

Figure 1: Result of 3-edge junction detection with the same parameter. (Left) junction detection on the original image; (Right) junction detection on the smoothed image.

3.3 Plane Decomposition

There are two approaches for plane decomposition

in our approach. One is based on the analysis of the line segments/edges in the image. The other is derived from the point features clustering on different planar surfaces. The former one is based the geometric constraints self-contained in the line segments, i.e. the line segments in the scene can be grouped into different categories with their associated vanishing points. The latter one relies on the assumption that the feature matches clustering on the same plane will satisfy a planar transformation (homography) between different views.

3.3.1 Spatial Layout Analysis from Edges

Figure 2 illustrates the steps of the spatial layout analysis in our approach. Firstly, we detected the line segments with the Canny edge detector and kept the line segments with lengths greater than 30 pixels. Secondly, we used the Vanishing Points detection algorithm proposed by Hedau et al. (Hedau et al., 2009), which calculates vanishing points from line segments and thus grouped the line segments according to their respective vanishing points. Then, we simply created an x coordinate range histogram, which indicates the coordinate range for each group of the line segments. As shown in Figure 2c, we can estimate that there are two main consecutive clusters existing in the group, as well as one consecutive cluster existing in Figure 2b. Finally, we created a rough plane decomposition by checking the x-coordinate ranges of these two line segment groups: there are three planes in total existing in this view, the first plane's x coordinate range is around 0 -100, the second one's is 100-400, and the third 450-750.
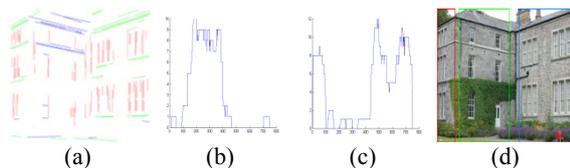

(a)　　　(b)　　　(c)　　　(d)

Figure 2: Spatial analysis results. (a) Grouped line segments according to their associated vanishing points represented by three different colours; (b) The x-coordinate range for the blue line segments; (c) The x-coordinate range for the green line segments; (d) the rough plane decomposition.

3.3.2 Planar Cues from ASIFT Matches

In our approach, we chose Affine SIFT (ASFIT) (Morel, J. M. and Yu. G. S., 2009) for feature detection and matching, as our approach is targeted at sparse image sets, where the viewpoint changes between different views are bigger than usual image sequences. ASIFT's accuracy on viewpoint changes

outperforms the other SIFT-based algorithms since ASIFT is a fully affine invariant method that simulates all image views obtainable by varying the two camera axis orientation parameters, namely the latitude and the longitude angles with SIFT. In other words, ASIFT simulates three parameters: the scale, the camera longitude angle and the latitude angle, and normalises the other three (translation and rotation) where SIFT only considers the zoom, rotation and translation.

We found that the features detected by ASIFT are mainly clustering on the texture abundant areas of different planes, so, we used the sequential RANSAC method to group these features with the different homographies they satisfy (Figure 3).
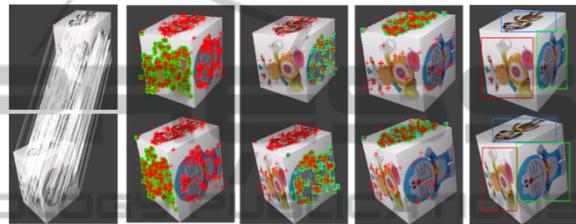

Figure 3: Plane decomposition from grouping features.

3.4 3D Corner Matching

After the rough plane decomposition, the probable range of different planes is obtained and the corresponding plane pairs are indicated by ASIFT feature matches. Then, we run the ASIFT algorithm again on these plane pairs and estimated the associated planar homographies from the ASIFT feature matches. Since feature match clustering on the same plane will satisfy a homography transformation between different views, 3D corners which locate on the intersection areas of different planes will satisfy 2 or more such transformations. At the same time, the 2D junctions can be screened out in such process (Figure 4).

4 RESULTS OF EXPERIMENTS

We carried out a series of experiments on synthetically generated images and real architecture images to test the performance of our approach.

Comparing the junction detection on the original and the smoothed images, the detected number of 3-junctions is decreased after image smoothing where most textured corners are removed (Figure 1). However, the pre-smoothing will eliminate some potential geometric corners (3D corners) if the edge gradient of the 3-junction is small.

Following our selection criteria (section 3.4) that 3D corners which locate on the intersection areas of different planes will satisfy 2 or more planar transformations – homographies, 3D corner matches are found between different views (Figure 4).

As Figure 3 shows, our first plane decomposition from feature grouping is rough as there are few feature matches near object boundaries; also, lines crossing the plane range but not in the plane will adversely affect the plane range analysis, such as the lines belonging to the ground floor or lines induced from the shades. However, with the help of the 3D corners, an improvement on segmentation is attainable (Figure 5).

Figure 4: 3D corner matching results.

Figure 5: Plane decomposition results.

Table 1: The total number of good feature matches vs. total matches on different data sets before and after plane decomposition.

Data Set	Match no. of whole view before plane extraction (Positive / Total)	Match no. of one plane after plane extraction (Positive / Total)
	630/697	580/637
	70/84	50/60

After the rough plane decomposition, we ran ASIFT again on the cropped plane pair. The number of the feature matches is increased (Figure 8), the number of matches on one single plane pair after plane decomposition is close to the total match number for the whole scene (Table 1).

5 CONCLUSIONS

We have proposed a framework for 3D corner detection and matching which combines local features (ASIFT features) and global geometric information for plane decomposition and feature grouping. With the information provided by detected 3D corner matches, the accuracy of the plane segmentation and feature grouping can be improved. At this stage, the 3D corner detection and matching scheme is immature. Sometimes, potential 3D

corners will be eliminated due to one edge having a low gradient, and the predicted 3D corner locations obtained by affine homographies associated with different planes are not precisely the same (meet at the same location). A possible future work about the 3D corner detection and matching is to separate the 3-junction into several 2-junction, and analysis the appearance of them and then combine with the self-contained structure information.

REFERENCES

Lowe D. G. (2004). Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*, vol. 60, pp. 91-110.

Ding, M., Lyngbaek, K., and Zakhor, A. (2008). Automatic registration of aerial imagery with untextured 3d lidar models. In Proc. of *CVPR*, pp. 1-8.

Elias, R., and Laganière, R. (2012). JUDOCA: JUnction Detection Operator Based on Circumferential Anchors. IEEE Transactions on Image Processing, vol. 21(4), pp. 2109 – 2118.

Hedau, V., Hoiem, D., and Forsyth, D. (2009). Recovering the spatial layout of cluttered rooms. In *Proc. ICCV*, pp. 1849–1856.

Isack, H. and Boykov, Y. (2012). Energy-based geometric multi-model fitting. *International Journal of Computer Vision*, vol. 97, pp.123–147.

Lee, D., Hebert, M., and Kanade, T. (2009). Geometric reasoning for single image structure recovery. In *Proc.CVPR*, pp. 2136–2143.

Liu, Y., Zhang, X., and Huang, T. (2003). Estimation of 3D structure and motion from image corners. *Pattern Recognition*, vol. 36(6), pp.1269 – 1277.

Matas, J., Chum, O., Urban, M. and Pajdla, T. (2002). Robust wide baseline stereo from maximally stable extremal regions. In *Proc. BMVC*, vol. 22, pp. 384-396.

Morel, J. M. and Yu. G. S. (2009) Asift: A new framework for fully affine invariant image comparison. *SIAM Journal on Imaging Sciences*, vol. 2, pp. 438-469.

Toldo, R. and Fusiello, A. (2008). Robust Multiple Structure Estimation with J-Linkage. In *Proc.ECCV,* vol. pages 537–547.

Trajkovic, M. and Hedley, M. (1998). Fast Corner Detection. *Image and Vision Computing*, vol. 16(2), pp. 75-87.

Xu, L. and Lu, C. W., Xu, Y., and Jia, J.Y. (2011). Image Smoothing via L0 Gradient Minimization. In *Proc. ACM SIGGRAPH Asia,* vol. 30, article No. 174.

Yu, S., Zhang, H., and Malik, J. (2008). Inferring spatial layout from a single image via depth-ordered grouping. In *CVPR Workshop*.

Zuliani, M., Kenney, C. S., and Manjunath, B. S. (2005). The multiRANSAC algorithm and its application to detect planar homographies. *In Proc. ICIP*, vol. 3, pp. 153–156.