

# Classifying Estimated Stereo Correspondences based on Delaunay Triangulation

Cristina Bustos, Elizabeth Vargas, and Maria Trujillo

Multimedia and Computer Vision Group, Universidad del Valle, Cali, Colombia  
{cristina.bustos, elizabeth.vargas, maria.trujillo}@correounivalle.edu.co

**Abstract.** The stereo vision problem gives rise to two sub-problems: the correspondence and the stereo calibration. Once the sub-problems have been solved, depth is estimated by triangulation. Depth estimation, using stereo images, is based on disparity – the relative displacement between corresponding points – and the camera geometry. Small errors in disparity may produce large errors in depth estimates due to the ill-posedness nature of the stereo vision problem. Moreover, if the camera geometry is unknown, it is estimated using corresponding points. A solution of stereo vision problem may be improved by identifying corresponding points which are inaccurately estimated. In this paper, the classification of a set of estimated corresponding points uses Delaunay triangulation by restricting it to a given subset of estimated corresponding points. Delaunay edges among estimated corresponding points are used to build undirected graphs. Classification criteria based on adjacencies are defined in order to decide whether or not a corresponding point is correctly estimated. Experimental evaluation, using ground truth image sets for quantitative analysis, shown values of specificity around 70% while sensitivity up to 96%.

**Keywords:** Corresponding points, Delaunay triangulation, 3D-reconstruction.

## 1 Introduction

Stereo vision refers to the ability to infer 3D structure, such as depth, of a scene using two or more images taken from different viewpoints. The correspondence problem has to be solved in order to infer the 3D structure of a scene. The correspondence problem consists in determining which point on the left image corresponds to which point on the right one [20]. Depth estimation has application in different sectors of industry, such as: autonomous robots [9], self-driving cars [7], intelligent cars [26], immersive 3D video conferencing [8], 3D-TV [13] and urban planning [14], among others.

The stereo correspondence problem is not solved yet, since it is an inverse and ill-posed problem. Different techniques have been employed in order to find corresponding points given two images [18,2,4]. When the correspondence problem is addressed, corresponding points may be inaccurately estimated: falsely matched or badly located. Small inaccuracies, in estimated corresponding points,

may have a large impact on depth estimation [20], due to the ill-posedness nature of the stereo vision problem. The ill-posedness nature gives rise to a set of constraints and rules in order to keep only disparities of reliable points by rejecting points that are mismatched. However, a solution to the inverse problem does not always exist. Thus, there is a room for proposals on identifying corresponding points which are inaccurately estimated. A set of correctly estimated corresponding points may be used as anchors for estimating dense disparity maps along with the camera geometry.

Delaunay triangulation has been used in feature correspondence [19], region corresponding [22] and block matching [23]. In [10], it is presented a method for establishing dense correspondence maps between two images, in a video sequence or in a stereo pair, by combining feature matching with Delaunay triangulation. This approach has advantages, such as: it allows estimating large displacements and subsequently taking into account disparity discontinuities. However, the results depend on the presence of texture. The method works well in images with strong textures, where a large number of small triangles are built. In untextured areas, the results are not accurate due to homogeneous regions. In [27], it is presented an improved SUSAN (Smallest Univalued Segment Assimilating Nucleus) corner detection algorithm and also an effective corners correspondence algorithm based on Delaunay Triangulation. This algorithm is robust to scaling, rotation and translation with short baseline, since interior angles and local structural similarity are robust to noise and motion transform. Dense disparity maps are calculated using Delaunay triangulation in [24], where image pairs are triangulated and triangles are classified into matched and unmatched triangles using a matching criterion: matched triangle and unmatched one, and disparity values in two types of triangle region are solved, respectively.

A generative probabilistic model to binocular stereo for fast matching is proposed in [6]. A Bayesian approach that uses a 2D mesh via Delaunay triangulation for stereo matching is presented, which is able to compute accurate disparity maps of high resolution images at frame rates close to real time without requiring global optimisation. The algorithm reduces the search space and can be parallelised. In [3], it is presented a method for stabilising the computation of stereo correspondences using Delaunay triangulation. The input images are partitioned into small and localised regions. Adjacent triangles are merged into larger polygonal patches and the planarity assumption is verified in order to achieve the planarity test robustly. Point correspondences are established by planar homographies, and are used as control points in the final matching process using dynamic programming. The approach was tested on three different types of data, including medical data. The results showed that it can speed up the stereo matching process. In [25], it is proposed a Delaunay-based stereo matching method, which relies on Canny edge detector to extract key-points. Using those points, Delaunay triangulation is performed and an initial disparity estimation is calculated, based on the histogram of each triangular area. This matching is further refined by considering that depth values of adjacent triangles should have similar depth values.

An approach for classifying corresponding points based on the Delaunay triangulation is presented in [21]. However, we did not consider that corresponding points are constrained by image content. Thus, a triangulation may be built using points belonging to different objects. In this paper, given a set of estimated corresponding points, a classification approach using Delaunay triangulation is presented. Delaunay edges are used to build undirected graphs, points belonging to the right image built one graph and points belonging to the left image built other graph. Classification criteria based on adjacencies are defined in order to decide whether or not a corresponding point is correctly estimated. Without loss of generality, the propose approach is validated using ground truth image sets.

## 2 Corresponding Points Classification

Once corresponding points are estimated, some of them may be inaccurately estimated. Thus, the classification of corresponding points consists in deciding whether or not a matching point is reliable. Estimated corresponding points are mapped into two graphs, points belonging to the right image are mapped into one graph and points belonging into the left image are mapped into the other graph. Thus, corresponding points are classified as correctly estimated if and only if the two graphs are isomorphic. Additional conditions are imposed in order to deal with the inverse and ill-posed characteristics of this classification.

### 2.1 Fundamentals

Suppose that  $G = (V, E)$  is an undirected graph where  $|V| = n$ . The adjacency matrix  $A$ , with respect to an ordered set of vertices, is the  $n \times n$  zero-one matrix with one at the  $(i, j)$ th entry when  $v_i$  and  $v_j$  are adjacent and zero at the  $(i, j)$ th entry when  $v_i$  and  $v_j$  are not adjacent. A bijective function  $\phi : V \leftrightarrow V$  is a graph isomorphism if:

$$v_i, v_j \in E \Leftrightarrow \phi(v_i), \phi(v_j) \in E'. \quad (1)$$

Moreover, a bijective function  $\phi$  preserves the adjacency between vertices [5].

A Delaunay triangulation is a triangulation of the vertex set with the property that the circumcircle of every triangle is empty, that is, there is no point from the vertex set in its interior.

### 2.2 Classification based on Delaunay Triangulation

The proposed classification approach is presented, without lost of generality, under the assumption of the stereo images are rectified. Given a set of estimated corresponding points, two Delaunay triangulation are build, one with points from the right image and the other one with points from the left image. After the triangulation, points from each image are labelled with the same number in order to identify the correspondence. Points are vertices and Delaunay edges are edges in an undirected graph.

In the proposed approach, the bijective function  $\phi$  is represented by estimated corresponding points and a set of adjacent vertices may have cardinality larger than 3. We will use the term correctly estimated to indicate that the points represent the same 3D point in both images. The classification of estimated corresponding points is based on the following restrictions:

**Restriction 1** (Adjacency). *Given two matching points, they are correctly estimated if and only if the set of adjacent vertices in the right image is equal to the set of adjacent vertices in the left image.*

**Restriction 2** (Cardinality). *Given two matching points and two sets of adjacent vertices of cardinalities  $n$  and  $m$  with  $n < m$ , the matching points are correctly estimated if and only if the set of vertices with low-cardinality is a proper subset of the set of vertices with high-cardinality.*

Bearing in mind that corresponding points are representing as disparity vectors and disparity is inversely proportional to depth, corresponding points are constrained by image content. In this way, object features may provide information to the Delaunay triangulation indicating a way in which feature points are related.

**Restriction 3** (Object boundaries). *Given two matching points and two sets of adjacent vertices of cardinalities  $n$  and  $m$  with  $n < m$ , also edges represent object boundaries, matching points must be triangulated in groups belonging to the same object. The matching points are correctly estimated if and only if they belong to the same 3D object.*

### 3 Experimental Evaluation

The proposed approach is validated using existing feature detection and matching schemes. An initial set of key-points is obtained using algorithms from the state of the art. In particular, the Scale Invariant Feature Transform (SIFT) [12], the Features from Accelerated Segment Test (FAST) [15], the Speeded Up Robust Features (SURF) [1] and Good Features to Track (GFTT) [17]. The matching SIFT key-points algorithm is provided in [11]. The FAST, SURF and GFTT key-points are matched using Sum of Squared Differences (SSD). The obtained matching points in the matching stage are classified using the proposed approach based on the Delaunay triangulation and the restrictions.

In order to enforce Restriction 3, a pre-processing step is conducted to obtain a binary image where the main contours could be identified. In this step, binarisation is obtained by applying a threshold using the Otsu algorithm. A closing and an opening morphological operations are applied to the resulting binary image, independently. Contours are obtained by subtracting resultant images after opening and closing operations. Every key-point is assigned to one contour, in order to triangulate points belonging to the same object by finding whether the key-point is inside or outside or on the contour. This process is independently conducted using key-points from the left and the right images.

The proposed approach is quantitatively evaluated using a "gold standard". The Middlebury Stereo Page (<http://vision.middlebury.edu/stereo/>) provides stereo images and each dataset includes the true disparity map [16]. In total, 8 sets of images are selected: Art, Books, Cones, Dolls, Laundry, Moebius, Reindeer and Teddy. The metrics employed to evaluate the proposed approach performance are Sensitivity, Specificity, Positive Predictive Value (PPV) and Negative Predictive Value (NPV). The ground-truth information is used for validating an identified correct or incorrect estimated corresponding point. Initial estimated corresponding points are compared with the ground-truth using the absolute error equation. The error at the  $(i, j)$  position is calculated using the following formula:

$$error_{i,j} = |DT_{i,j} - d_{i,j}|, \quad (2)$$

where  $DT_{i,j}$  represents the ground-truth disparity value at  $(i, j)$  and  $d_{i,j}$  is the estimated disparity value at  $(i, j)$ . An estimated corresponding point is labelled as estimate incorrectly with an error value bigger than 1, otherwise is considered as estimate correctly. In this way, the performance metrics are built based on estimate incorrectly and estimate correctly.

Tables 1, 2, 3 and 4 contain the performance metrics obtained using SIFT, FAST, SURF and GFTT key-points. In general, the sensitivity decreased for most datasets while the specificity increased. The reason, is that there are some edges missing in comparison with the unconstrained triangulation in [21], and some of them were not supposed to be there, decreasing the number of false positives (which increases the specificity) and some had to be there, increasing the number of false negatives (which decreases the sensitivity).

**Table 1.** Performance metrics of the proposed classification using SIFT keypoints.

Dataset	Sensitivity(%)	Specificity(%)	PPV(%)	NPV(%)
Art	75,76	55,23	24,51	92,23
Books	68,63	70,86	25,55	93,94
Cones	42,62	71,67	18,44	89,26
Dolls	58,26	60,54	27,57	84,91
Laundry	64,21	54,76	44,53	73,02
Moebius	46,43	62,21	20,80	84,46
Reindeer	49,00	40,91	79,03	15,00
Teddy	64,41	55,76	24,20	87,72

In general, key-points calculated using SURF and GFTT were not useful for the proposed classification. Using SIFT and FAST, while the specificity value incremented for most datasets (for books there was no change), the sensitivity decremented for some of them, and for others keep equal.

Algorithms' run-time is used as a measure of the complexity. Table 5 shows run-times of the proposed approach using keypoints calculated with the SIFT, the FAST, the SURF and the GFTT. In general, the FAST took fewer seconds than the others except for the Reindeer dataset.

**Table 2.** Performance metrics of the proposed classification using FAST keypoints.

Dataset	Sensitivity(%)	Specificity(%)	PPV(%)	NPV(%)
Art	63,89	31,52	42,20	52,73
Books	93,29	23,38	54,09	78,26
Cones	71,79	51,05	37,58	81,51
Dolls	67,06	62,05	43,51	81,21
Laundry	95,88	39,39	75,61	82,98
Moebius	68,25	49,33	36,13	78,72
Reindeer	77,59	35,66	32,85	79,69
Teddy	78,86	36,79	44,29	73,20

**Table 3.** Performance metrics of the proposed classification using SURF keypoints.

Dataset	Sensitivity(%)	Specificity(%)	PPV(%)	NPV(%)
Art	63,16	50,41	37,11	74,70
Books	61,61	51,60	27,49	81,86
Cones	43,88	61,12	17,62	85,18
Dolls	60,87	59,71	38,89	78,37
Laundry	67,67	53,37	48,13	72,08
Moebius	52,63	57,92	22,90	83,73
Reindeer	52,24	63,13	34,65	77,93
Teddy	62,50	56,04	29,57	83,50

## 4 Conclusions

In this paper, we presented an approach for classifying estimated corresponding points based on the constrained Delaunay triangulation and considering that corresponding points are constrained by image content. The proposed approach uses graph isomorphism to formulate three restrictions: adjacency, cardinality and object boundaries, in order to classify estimated corresponding points as correctly estimated if and only if the two graphs are isomorphic.

Regarding the keypoints, the SURF and the GFTT algorithms did not yield key-points useful for the proposed approach. The proposed approach exhibits the best performance with the FAST corners, where an initial set of corresponding points contains more than 50% estimate incorrectly. However, the proposed approach produced error propagations since a corresponding point incorrectly estimated affects directly the neighbour in the graphs.

In comparison to the unconstrained Delaunay triangulation presented in [21], the use of constrains contribute to the increment of specificity, given that not all the matching points are triangulated together, reducing the number of false positives. However, it also decrements the sensitivity, because for some points that are "estimate incorrectly", there are missing edges, generating a false negative.

**Table 4.** Performance metrics of the proposed classification using GFTT keypoints.

Dataset	Sensitivity(%)	Specificity(%)	PPV(%)	NPV(%)
Art	60,98	61,25	28,74	85,96
Books	65,00	50,69	10,79	94,04
Cones	45,95	64,68	9,94	93,38
Dolls	61,90	53,80	13,68	92,27
Laundry	73,48	57,29	61,86	69,62
Moebius	82,72	56,59	33,17	92,63
Reindeer	65,71	45,84	10,22	93,44
Teddy	67,74	49,75	17,43	90,78

**Table 5.** Run-time using keypoints calculated with the SIFT, the FAST, the SURF and the GFTT (in seconds).

Dataset	SIFT	SURF	FAST	GFTT
Art	0.43	0.37	0.25	0.31
Books	0.36	0.43	0.23	0.32
Cones	0.46	0.49	0.24	0.32
Dolls	0.52	0.45	0.27	0.35
Laundry	0.43	0.40	0.25	0.32
Moebius	0.39	0.33	0.26	0.33
Reindeer	0.33	0.37	0.38	0.33
Teddy	0.37	0.34	0.24	0.43

## References

1. Baya, H., Essa, A., Tuytelaarsb, T., Goo, L.V.: Speeded-Up Robust Features (SURF). *Computer Vision and Image Understanding* 110(3), 346–359 (2008)
2. Bleyer, M., Gelautz, M.: Graph-cut-based stereo matching using image segmentation with symmetrical treatment of occlusions. *Signal Processing: Image Communication* 22, 127–143 (2007)
3. Chen, C.i., Sargent, D., Tsai, C.m., Wang, Y.f., Koppel, D.: Stabilizing Stereo Correspondence Computation Using Delaunay Triangulation and Planar Homography. *Lecture Notes in Computer Science* 5358, 836–845 (2008)
4. Cho, N.: Stereo Matching using Multi-directional Dynamic Programming and Edge Orientation. *Korea-Japan Joint Workshop on Frontiers of Computer ... 1*, 233–236 (2007)
5. Gallardo, P.: *Notas de matematica discreta* (2002), [http://www.uam.es/personal\\_pdi/ciencias/gallardo/capitulo8a.pdf](http://www.uam.es/personal_pdi/ciencias/gallardo/capitulo8a.pdf)
6. Geiger, A., Roser, M., Urtasun, R.: Efficient large-scale stereo matching. *Computer Vision/ACCV 2010* pp. 25–38 (2011)
7. Google: German scientists unveil self-driving car (2012), [http://www.google.com/hostednews/afp/article/ALeqM5izCRPtPhsc2k6kCY2\\_-Y0eUs7r\\_w?docId=CNG.c668be9320e3376a10d767deb0b0649f.a91](http://www.google.com/hostednews/afp/article/ALeqM5izCRPtPhsc2k6kCY2_-Y0eUs7r_w?docId=CNG.c668be9320e3376a10d767deb0b0649f.a91)
8. Ho, Y.S., Jang, W.S.: Gaze correction using 3D video processing for videoconferencing. In: *IEEE China Summit and International Conference on Signal and Information Processing (ChinaSIP)*. pp. 496–499 (2015)

9. Jet Propulsion Laboratory, C.I.o.T.: Mars exploration rovers (2012), <http://marsrovers.jpl.nasa.gov/home/index.html>
10. Kardouchi, M., Konrad, J., Vazquez, C.: Estimation of large-amplitude motion and disparity fields: Application to intermediate view reconstruction. Proc. SPIE (2001), [http://proceedings.spiedigitallibrary.org/data/Conferences/SPIEP/34800/340\\_1.pdf](http://proceedings.spiedigitallibrary.org/data/Conferences/SPIEP/34800/340_1.pdf)
11. Lowe, D.: Demo software: Sift keypoint detector (2005), <http://www.cs.ubc.ca/~lowe/keypoints/siftDemoV4.zip>
12. Lowe, D.: Distinctive image features from scale-invariant keypoints. International journal of computer vision pp. 1–28 (2004)
13. Oh, J., Sohn, K.: A Depth-Aware Character Generator for 3DTV. IEEE Transactions on Broadcasting 58(4), 523–532 (2012)
14. Romanoni, A., Matteucci, M.: Incremental reconstruction of urban environments by Edge-Points Delaunay triangulation. In: IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). pp. 4473–4479 (2015)
15. Rosten, E., Drummond, T.: Machine learning for high-speed corner detection. Computer VisionECCV pp. 1–14 (2006)
16. Scharstein, D.: Middlebury stereo datasets (2011), <http://vision.middlebury.edu/stereo/data/>
17. Shi, J., Tomasi, C.: Good features to track. In: Computer Vision and Pattern Recognition, 1994. Proceedings CVPR'94., 1994 IEEE Computer Society Conference on. pp. 593–600. IEEE (1994)
18. Sun, J., Zheng, N., Shum, H.: Stereo matching using belief propagation. Pattern Analysis and Machine Intelligence 25(7), 787–800 (2003)
19. Tan, H., Zhou, F., Zhang, W., WANG, Y.: Automatic correspondence approach for feature points in camera calibration. Journal of Optoelectronics pp. 736–739 (2011)
20. Trucco, E., Verri, A.: Introductory techniques for 3-D computer vision (1998)
21. Vargas, E., Trujillo, M.: A Corresponding Points Classification Approach by Delaunay Triangulation. In: Esteban De La Fuente Rubio (ed.) Proceedings of the 4th Latin American Conference on Networked and Electronic Media. pp. 6–12. Escuela de Informática, Universidad Nacional Andrés Bello, Santiago de Chile (2012)
22. Wang, J., Wang, L., Chan, K., Constable, M.: A linear programming based method for joint object region matching and labeling. Computer VisionACCV pp. 66–78 (2012)
23. Zhang, H., Chen, X.: Effective scene matching for intelligent video surveillance. In: International Conference on Graphic and Image Processing (2013)
24. Zhang, S., Qu, X., Ma, S., Yang, Z., Kong, L.: A dense stereo matching algorithm based on triangulation. J. Comput. Inform. Syst. 8, 283–292 (2012)
25. Zhang, X.h., Li, G., Li, C.l., Zhang, H., Zhao, J., Hou, Z.x.: Stereo matching algorithm based on 2d delaunay triangulation. Mathematical Problems in Engineering 2015 (2015)
26. Zhao, H.: Motion planning for intelligent cars following roads based on feasible neighborhood . In: IEEE International Conference on Control Science and Systems Engineering (CCSSE). pp. 27–31 (2014)
27. Zhou, D., Li, G.: Effective Corner Matching Based on Delaunay Triangulation. Robotics and Automation Proceedings pp. 2730–2735 (2004)