Robust image matching based on SIFT and delaunay triangulation

Jianfang Dou (窦建方)* and Jianxun Li (李建勋)

Department of Automation, Shanghai Jiao Tong University, and Key Laboratory of System Control and Information Processing, Ministry of Education of China, Shanghai 200240, China

**Corresponding author: specialdays_2010@163.com*

Received September 23, 2011; accepted October 20, 2011; posted online April 27, 2012

Image matching is an important question in computer vision, however, due to the large viewpoint and similar regions, there exist false matches. A robust matching method-DelTri is proposed. Based on the initial matching of Scale Invariant Feature Transform, the matched keypoints are respectively triangulated to create the triangulation net, which can express the overlapped physical structure of the objects. The matched triangles can lead to the final matches. Compared with classical RANSAC, experiments show that DelTri can improve the match robustness, including matching accuracy and magnitude efficiency.

OCIS codes: 100.2000, 100.5010.

doi: 10.3788/COL201210.S11001.

As a fundamental problem of computer vision, image matching is to determine point correspondences between two images of the same scene or object[1]. The vast applications involve location recognition, facial recognition, object recognition, motion understanding, change detection, among others.

A wide variety of interest point and corner detectors exist in the literature. They can be divided into three categories: contour based, intensity based and parametric model based methods[2]. Among the intensity based methods, the first one uses non-linear filter, such as SUSAN corner detector; the second one is based on curvature of planar curves, such as Kitchen and Rosenfeld's method. The typical one is Harris and Stephens' method. Lowe[3] proposed a scale invariant feature transform (SIFT), which combines a scale invariant region detector and a descriptor based on the gradient distribution in the detected regions. The descriptor is represented by a 3D histogram of gradient locations and orientations. The main advantage of SIFT is that it is able to detect and describe local features that are invariant to scaling and rotation.

The problem of outliers in structure and motion recovery from images is well known in the literature. The RANSAC(Random Sample Consensus) paradigm proposed by Fischler *et al.*[4] detects outlying data by first randomly selecting samples of the minimum number of data items required to estimate a given entity and then looking for consensus of the estimates among the samples. However, For large viewpoints, the (non-cyclo) rotation of the camera about the optical centre is significant. There may be severe projective distortion due to differing perspective foreshortenings of the plane in each image. Previous works[5] did not consider the perspective distortion caused by large viewpoints. Delaunay triangulations pervade computer vision. They not only provide a convenient and robust neighbourhood representation for Voronoi tessellations of the image plane, but also provide a powerful geometric representation for volumetric information. Based on the uniqueness of Delaunay triangulation[6] for an image and the similarity of triangulations for the same scene in different images, In this letter, a robust matching method-DelTri is proposed.

RANSAC is a robust estimator originally proposed by Fishcler and Bolles in 1981[6] where it was used to derive a usable model from a set of data. In Ref. [7], RANSAC is used to filter out the incorrectly mapped points that come from the imprecision of the SIFT model. RANSAC starts by assuming some transformation model (typically affine or perspective).

The first parameter for RANSAC is the distance threshold t. The common method to determine it is based on statistical theory. Firstly, assume that the distribution of effective point under transformation model according to the distance is known, we calculate the distance threshold t such that the probability of effective point in point set is α. Suppose the distribution satisfies the zero mean and variance σ of the Gaussian distribution, we can compute the value t. In this case, the square distance between points is d^2, which is the square sum of Gaussian variant, is meet the χ^2_m (Chi-square Distribution) that has m degrees of freedom. Based on the integral property of Chi-square Distribution, the probability of random variable that obeys the Chi-square Distribution is lower than the integral upper limit, the formal is as

$$F_m(k^2) = \int_0^{k^2} \chi^2_m(\xi)\mathrm{d}\xi < k^2, \tag{1}$$

the distance threshold t can be caculated by

$$t^2 = F_m^{-1}(\alpha)\sigma^2, \tag{2}$$

Then, we can classify the point set into effective point and invalid point.

$$\begin{cases} \text{effective point} & d^2 < t^2 \\ \text{invalid point} & d^2 \geqslant t^2 \end{cases}. \tag{3}$$

The second parameter for RANSAC is the number of iterations N, is chosen high enough to ensure that the probability p (usually set to 0.99) that at least one of the

sets of random samples does not include an outlier. Let u present the probability that any selected data point is an inlier and $v = 1 - u$ is the probability of observing an outliers. N iteration of the minimum of points denoted are required, where

$$1 - p = (1 - u^m)^N, \tag{4}$$

And thus with some manipulation,

$$N = \frac{\log(1-p)}{\log[1-(1-v)^m]}. \tag{5}$$

The consume time of RANSAC can be calculated as follows:

$$T = N(T_{\mathrm{G}} + MT_{\mathrm{E}}). \tag{6}$$

where T_{G} is the time spent on generating a hypothesis, T_{E} is the time spent on evaluating the hypothesis for each data, M is the number of whole data.

Let $P = \{p_1, \cdots, p_n\}$ be a set of points in R^d. The Voronoi cell associated to a point p_i, denoted by $V(p_i)$, is the region of space that is closer from p_i than from other points in P[8]:

$$V(p_i) = \{p \in R^d : \forall j \neq i, ||p - p_i|| \leqslant ||p - p_j||, \tag{7}$$

where $V(p_i)$ is the intersection of $n-1$ half-spaces bounded by the bisector planes of segments $[p_i p_j]$, $j \neq i$. Therefore, $V(p_i)$ is a convex polytope, possibly unbounded. The Voronoi diagram of P, denoted by $V\mathrm{or}(P)$, is the partition of space induced by the Voronoi cells $V(p_i)$.

Triangulation is a process that takes a region of space and divides it into subregions. The space may be of any dimension, however, a 2D space is considered here since we are dealing with 2D points. In this case, the subregions are simply triangles. Euler formula of Triangulation is

$$f - e + v = 1, \tag{8}$$

where f is the number of facet; e is the number of edges, v is the number of vertex. The complexity of n points P constructed triangulation has N_{tri} triangles and N_{edge} edges. In this case, compared with (8), $e = N_{\mathrm{edge}}$.

$$N_{\mathrm{tri}} = 2n - 2 - k, \tag{9}$$

$$N_{\mathrm{edge}} = 3n - 3 - k, \tag{10}$$

where k is the number of points P in on the convex hull of P.

The Delaunay triangulation $\mathrm{Del}(P)$ of P is defined as the geometric dual of the Voronoi diagram: there is an edge between two points p_i and p_j in the Delaunay triangulation if and only if their Voronoi cells $V(p_i)$ and $V(p_j)$ have a non-empty intersection. It yields a triangulation of P, that is to say a partition of the convex hull of P into d-dimensional simplices (i.e. into triangles in 2D, into tetrahedra in 3D, and so on). The formula of $\mathrm{Del}(P)$ is (11). Figure 1(a) displays an example of a Voronoi diagram and its associated Delaunay triangulation in the plane.

$$\mathrm{Del}(P) = \{T(p_i, p_j, p_k) | p_i \in P, p_j \in P, p_k \in P, C(p_i, p_j, p_k) \cap P \backslash (p_i, p_j, p_k) = \phi\} \tag{11}$$

where $C(p_i, p_j, p_k)$ is the circle circumscribed by three vertices p_i, p_j, p_k, which form a Delaunay Triangle $T(p_i, p_j, p_k)$

The algorithmic complexity of the Delaunay triangulation of n points is $O(n \log n)$ in 2D[9]. Figures 1(b) and (c) Show the created Delaunay Triangulations using 20 discrete points.

Compared with DelTri algorithm, it can be seen that RANSAC algorithm selects samples randomly. The random sampling has some disadvantages: it increases the number of iterations. Above all, there are three problems for RANSAC: firstly, there is no upper bound on the time it takes to compute the transformation model parameters; secondly, the number of iterations N computed is limited the solution obtained may not be optimal, and it may not even be one that fits the data in a good way; finally, it requires the setting of problem-specific thresholds.

The Delaunay triangulation has many known properties that make it the most widely-used triangulation. Our choice of Delaunay triangulation as a space subdivision for image matching is motivated by the following remarkable property: under some assumptions, and especially if P is a "sufficiently dense" sample of a surface, in some sense defined, then a good approximation of the surface is "contained" in $\mathrm{Del}(P)$, in the sense that the surface can be accurately represented by selecting an adequate subset of the triangular facets of the Delaunay triangulation.

When the viewpoints are increasing, the number of correct matches will go down quickly for the SIFT algorithm. The reason is that due to large viewpoint, the similarity of the SIFT descriptor will become small. In order to enhance the resist the changes caused by large viewpoint, in this letter, we exploit the structure of the overlapped regions for different images, if we can use a certain number of discrete points to express the similar

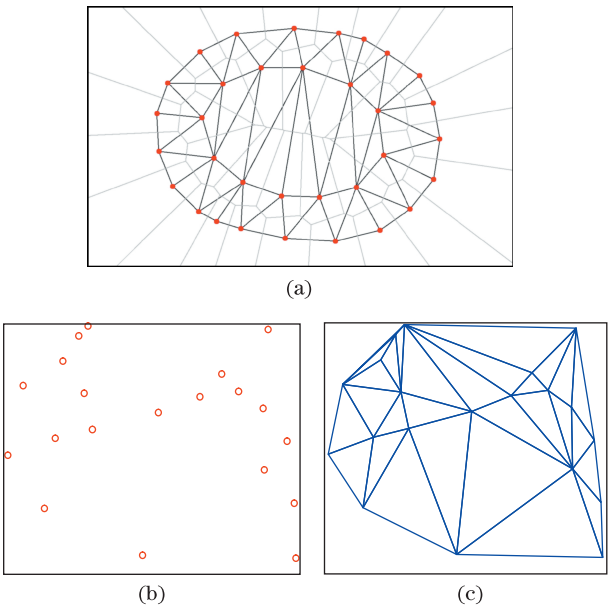


Fig. 1. (Color on line) (a) Voronoi diagram (gray edges) of a set of 2D points (red dots) and its associated delaunay triangulation (black edges). (b) Delaunay triangulation discrete points (20). (c)Triangulated meshes.

scene such as by certain methods, we will see that, the structure is unique. Based on this fact, we introduce the Delaunay triangulation of discrete points. The Delaunay triangulation has the property of uniqueness which is fit for the work. Because the triangle has three sides, we can choose the correct matches based on triangle or lines. Through this, we can achieve the robust matching based on Delaunay Triangulation. Figures 2(a) and (b) show six of detected keypoints. The two source images are acquired from a very large viewpoint, due to uniqueness of Delaunay triangulation, the matches can be stably tracked. Figures 2(c) and (d) show the triangulation net created by the Delaunay Triangulation. For RANSAC method, the number of matches is zero which is shown Fig. 2(f).

The proposed algorithm can be expressed as follows:

1) SIFT descriptors $\mathbf{X} = \{x_1, x_2, \cdots, x_n\}^{\mathrm{T}}$ for the extracted feature points from the input images I_1 and I_2 according to the SIFT algorithm.

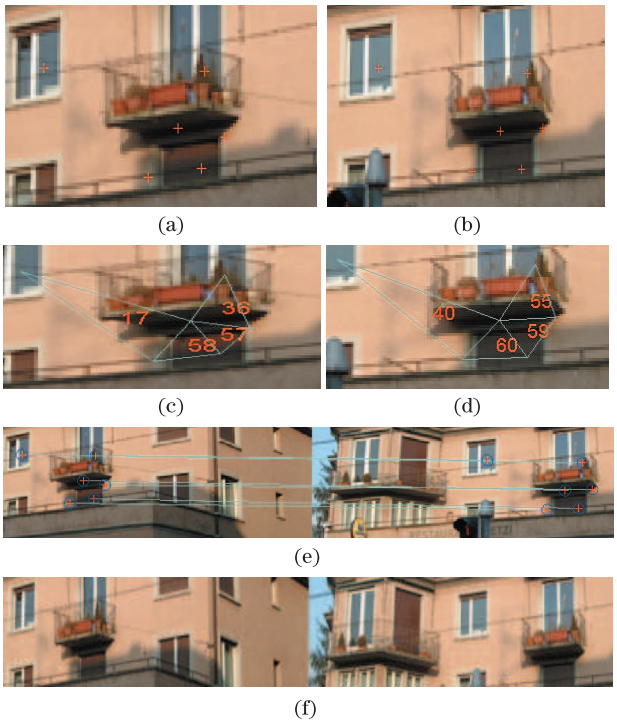
2) The Euclidian distance between SIFT descriptors is employed to determine the initial corresponding feature point pairs.

3) Delaunay Triangulation between the input images I_1 and I_2 according to the initial corresponding feature point pairs by step 2).

4) Classify the inliers based on triangles of the Delaunay triangulation and finally get the correct corresponding feature point pairs.

We assume that the transformation between the input images is projective transformation which can be defined as:

$$\begin{bmatrix} xh \\ yh \\ k \end{bmatrix} = \mathbf{H} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix}, \tag{12}$$



(a) (b) (c) (d) (e) (f)

Fig. 2. Six keypoints for image 1(a) and image 2(b). Triangulation net from six keypoints for image 1(c) and image 2(d). Match result by our method (e) and by RANSAC (f).

$$\mathbf{H} = \begin{bmatrix} h_{11} & h_{12} & h_{13} \\ h_{21} & h_{22} & h_{23} \\ h_{31} & h_{32} & h_{33} \end{bmatrix}. \tag{13}$$

The non-homogeneous coordinates (x', y') are computed as

$$x' = \frac{xh}{k} = \frac{h_{11}x + h_{12}y + h_{13}}{h_{31}x + h_{32}y + h_{33}},$$
$$y' = \frac{yh}{k} = \frac{h_{21}x + h_{22}y + h_{23}}{h_{31}x + h_{32}y + h_{33}}, \tag{14}$$

where $(x', y') \leftrightarrow (x, y)$ are pixel point correspondences and $\mathbf{H}$ is homography transformation matrix. Using the transformation matrix, the symmetric transfer error $d(x, \mathbf{H}^{-1}x')^2 + d(x', \mathbf{H}x)^2$ is calculated for every matching point, and the inliers that are less than the threshold value are counted. Here $d(x, y)$ is the Euclidean distance between points x and y.

For the three image pair, we use the epipolar constraints and considered the match (X, Y) was a correct match using the evaluation metrics as follows:

$$d(X, l_X) + d(Y, l_Y) \leqslant d_t, \tag{15}$$
$$l_X = F^{\mathrm{T}}Y, \tag{16}$$
$$l_Y = FX, \tag{17}$$

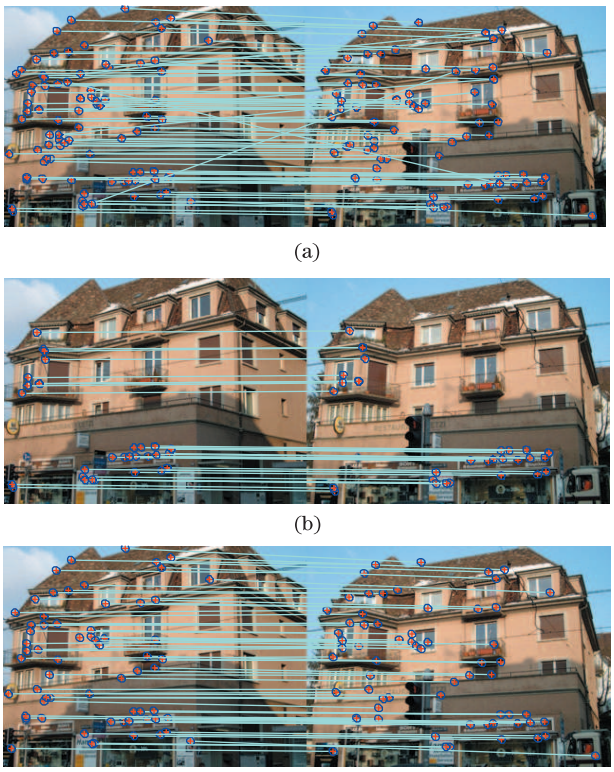
where $d(X, l_X)$ was the distance from point X to the epipolar line l_X which obtained by Y and fundamental matrix F, accurate fundamental matrix F between each image pair was calculated from some handpicked control point pairs, so is $d(Y, l_Y)$, and choose $d_{\mathrm{t}} = 2$.

In order to verify the effectiveness of the proposed algorithm, experimental results of sets of images from Ref. [10] are given below. ZuBuD is a database of color images of 201 buildings in Zurich city. Each building is represented by five snapshot taken from five different viewpoints. Illumination conditions vary for different buildings. Tests were conducted using some of 550 image pairs from the Zurich Building Database. The tested image sets are resized into 320×240. The threshold of SIFT initial matching is 0.6, the distance threshold for deciding outliers for RANSAC is $t = 0.004$, Maximum number of iterations is 1 000.

We choose one group of images to test the proposed methods. Figure 3 shows the image set taken at different viewpoint and Fig. 3(a) is a reference image and Figs. 3(b)–(d) are real images. Image pairs 1 is referred to Fig. 3(a) and (b); Image pairs 2 is referred to Fig. 3(a) and (c); Image pairs 3 is referred to Fig. 3(a) and (d). Figures 4 and 5 show the matching result of image pairs 3 for RANSAC method and DelTri approach. We choose image pairs 3 to test the method because the two images have large viewpoints which show great projective distortion and result in mismatches. Figure 4(a) show the initial matched key points by SIFT. From the match result, we can see that due to large viewpoints or repetitive patterns on the scene, there exist a certain number of false matches. We used RANSAC to filter the wrong matches. Figure 4(b) shows the result by RANSAC. Figure 4(c) shows the result by DelTri method. Figures 5(a) and (b) show the Delaunay Triangulation Net of the two images of Figs. 3(a) and (d) from

(a) (b) (c) (d)

Fig. 3. One image sets from Ref. [10].



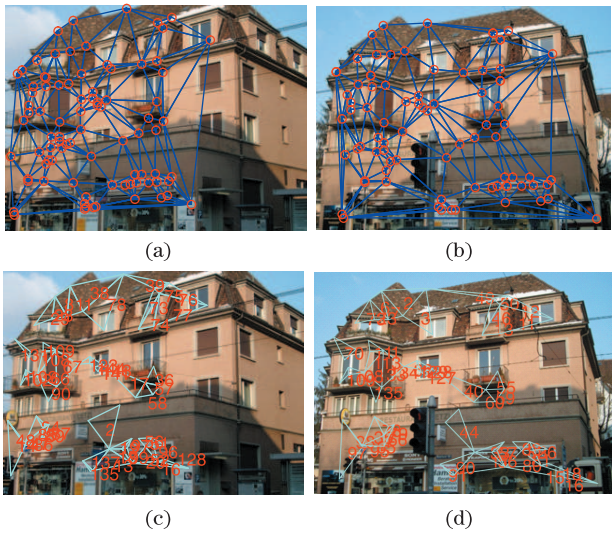
(a) (b) (c)

Fig. 4. (a) Initial matching by SIFT (34/91); (b) fine matching by RANSAC (20/31); and (c) matching by our proposed method (30/64).

initial matched points from Fig. 5(c). Due to the uniqueness of Delaunay Triangulation on overlapped region of the Image pairs 3, our method can overcome the changes caused by different viewpoint at a certain degree. The number of final matched triangles are shown in Fig. 5(c) and (d). From table 2, we can see that the final correct ratio for RANSAC is 58.8% while the DelTri is 88.2%.

Because RANSAC algorithm selects samples randomly, it can not make sure that the selected samples are all correct and can calculate the homography correctly. In order to test how the randomly selected samples affect the matching accuracy, compared with DelTri method, we both run RANSAC and DelTri 40 times based on initial matching of SIFT for image pairs 3 and at each time to calculate the correct ratio of result matches for both method. We also calculate the computation time at each time. Figure 6(a) shows that the correct ratio of DelTri is higher than RANSAC, while Fig. 6(b) shows the computation time for each time, the average consume time of DelTri is 0.14s and RANSAC 1.10 s. The distance threshold t for RANSAC is changed from 0.001 to 0.01, the correct ratio of result matches and computation time versus distance threshold are shown in Figs. 6(c) and (d). The reason is that the disadvantage of random samples for RANSAC and the uniqueness of Delaunay Triangulation for the same scene or object.

The final result for image pairs1-3 was shown in Tables 1 and 2. Table 1 show the Initial matches obtained by



(a) (b) (c) (d)

Fig. 5. (a) Delaunay triangulation net of Fig. 4(a) (160 triangles); (b) delaunay triangulation net of Fig. 4(d) (143 triangles); (c) matched triangles (51) of Fig. 5(a); (d) matched triangles (51) of Fig. 5(b).

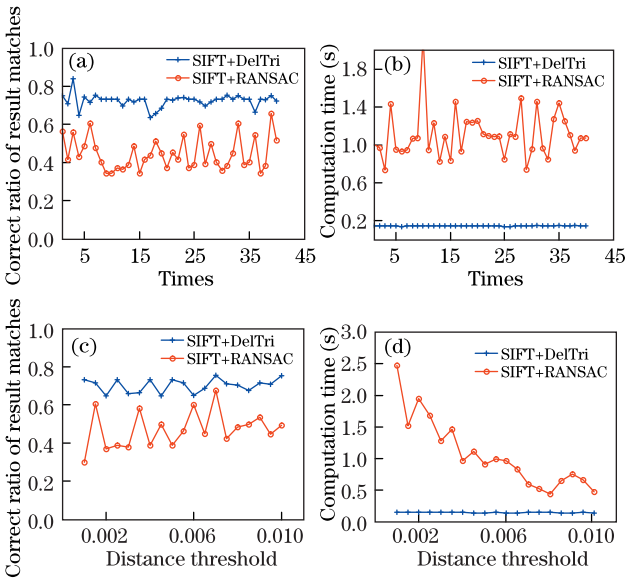


Fig. 6. Comparison of accuracy and computation time for RANSAC and DelTri.

the SIFT algorithms and the correct ratio of matches decided by epipolar constraints. Table 2 compares the computation time and matching accuracy of the two methods. We can see that the high correct ratio matches were efficiently obtained by proposed method, and also the computation time can decreased for some image pairs 1 and image pairs 2. From the experiment results we can see that our proposed method is effective and can improve the matching accuracy .

In conclusion, this letter presents a robust matching-DelTri approach to deal with the false matching problem caused by large viewpoints and similar region. Our approach detects inliers and outliers based on the Delaunay triangulation. By exploiting the Delaunay triangulation uniqueness property we are able to create a robust efficient matching algorithm without the inefficiencies inherent in fit-and-test approaches. Compared with classical RANSAC, experiment shows the effectiveness of proposed method.

Table 1. SIFT Initial and Correct Matches for Image Pairs1-3 (Fig. 4)

Image Pairs	Image Pairs 1	Image Pairs 2	Image Pairs 3
Key Points	781/835	781/957	781/925
Num of Initial Matches	339	118	91
Num of Correct Matches	292	69	34
Correct Ratio of Initial Matches	86.1%	58.5%	37.3%

Table 2. Comparison of Matching Methods

Image Pair		Image Pairs 1	Image Pairs 2	Image Pairs 3
SIFT+ RANSAC	Num of Correct Matches	220(201)	36(24)	31(20)
	Computation Time	5.32	7.07	6.65
	Correct Ratio of Result Matches	68.8%	34.7%	58.8%
SIFT+ Delaunay	Num of Correct Matches	270(205)	85(49)	64(30)
	Computation Time	6.96	6.21	5.95
	Correct Ratio of Result Matches	70.2%	71.0%	88.2%

This work was supported by the National Natural Science Foundation of China (Nos. 61175008, 60935001); the "973" Project of China (Nos. 2009CB824900 and 2010CB734103), and the Space Foundation of Supporting-Technology (No. 2011-HT-SHJD002).

References

1. K. Grauman and T. Darrell, in *Proceedings of International Conference on Computer Vision and Pattern Recognition*(San Diego, USA, 2005) 627.
2. C. Schmid, R. Mohr, and C. Bauckhage, Int. J. Comput. Vision. **37,** 151 (2000).
3. D. G. Lowe, Int. J. Comput. Vision. **60,** 91 (2004).
4. M. A. Fischler and R. C. Bolles, Commun. Acm **24,** 381 (1981).
5. D. Fleck and Z. Duric, in *Proceeding of International Conference on Image Analysis and Recognition* (Halifax, Canada, 2009) 268.
6. C. B. Barber, D. P. Dobkin, and H. Huhdanpaa, ACM Trans. on Mathematical Software (TOMS) **22,** 469 (1996).
7. Z. Yuan, P. Yan, and S. Li, in *Proceeding of Audio, Language and Image Processing* (Shanghai, China, 2008) 1550.
8. J.-D. Boissonnat and M. Yvinec, *Algorithmic Geometry, Chapter Voronoi Diagrams: Euclidian Metric, Delaunay Complexes* (Cambridge University Press, UK, 1998).
9. D. Attali, J.-D. Boissonnat, and A. Lieutier, in *Proceedings of 19th Annual Symposium on Computational Geometry* (San Diego, USA, 2003) 201.
10. H. Shao, T. Svoboda, and L. Van Gool. Zubud-Zurich building database for image based recognition. Technical Report TR-260, Computer Vision Lab, Swiss Federal Institute of Technology, Switzerland 2003.