

RESEARCH

Open Access



A geometric method for contour extraction of *Drosophila* embryos

Qi Li^{1*} and Yongyi Gong²

From IEEE BIBM International Conference on Bioinformatics & Biomedicine (BIBM) 2016
Shenzhen, China. 15-18 December 2016

Abstract

Background: High resolution images of *Drosophila* embryos in their developmental stages contain rich spatial and temporal information of gene expression. Automatic extraction of the contour of an embryo of interest in an embryonic image is a critical step of a computational system used to discover gene-gene interaction on *Drosophila*.

Results: We propose a geometric method for contour extraction of *Drosophila* embryos. The key of the proposed geometric method is k-dominant point extraction that is a generalization of 3-dominant point extraction proposed in our previous work. Based on k-dominant point extraction, we can approximate a connected component of edge pixels by a polygon that can be either convex or concave. The test on BDGP data shows that the proposed method outperforms two existing methods designed for contour extraction of *Drosophila* embryos.

Conclusions: The main advantage of the proposed geometric method in the context of contour extraction of *Drosophila* embryos is its ability of segmenting embryos touching each other. The proposed geometric method can also be applied to applications relevant to contour extraction.

Keywords: *Drosophila* embryo, Contour extraction, Concave shape, Dominant point, Geometric sequence

Background

High resolution images of *Drosophila* embryos in their developmental stages contain rich spatial and temporal information of gene expression. They have become a valuable instrument for micro-biologists to discover gene-gene interaction [1]. Automatic extraction of the contour of an interest embryo in an image is a critical step of a computational system for the discovery of gene-gene interaction on *Drosophila* [2].

In general, *Drosophila* embryonic images contain substantial amount of variations [3, 4]: i) imaging conditions, such as contrasts, scale, orientation, and neighboring embryos, ii) gene expression patterns, and iii) developmental stages. Most existing methods were developed upon low-level image features, such as edge pixels or pixels with a high deviation of grayvalues in a local window

[3–10]. Peng and Myers [5] proposed a method that uses the standard deviation of the local windows of a pixel to classify the pixel as a foreground or background pixel. Their method applies a 8-neighbor-connectivity region-growing method to extract the contour of an embryo. Pan et al. [6] applied a variant of Marquardt-Levenberg algorithm to estimate an optimal affine transformation to register localized embryos into an ellipsoidal region. Puniyani et al. [7] proposed an edge detection based method that assumes a number of heuristic constraints, including object size, convexity, shape features (e.g., ratio of the major over minor axis of an object), and the percentage of overlapping regions. Frise et al. [8] developed the method of Peng and Myers [5] by adding three morphological operations on a binary image: i) removal of isolated pixels, ii) dilation, and iii) majority processing. Furthermore, Frise et al. [8] proposed a heuristic algorithm to separate the embryo of interest from multiple touching embryos, with the assumption that the center of the embryo of interest is the image center. Mace et al. [9] proposed an eigen-embryo method to extract

*Correspondence: qi.li@wku.edu

¹Western Kentucky University, 1906 College Blvd., 42101 Bowling Green, KY, USA

Full list of author information is available at the end of the article

the contour of embryos, where a particle swarm optimizer was used to reduce the computational cost of searching optimal eigen parameters. Li and Kambhmettu [3] proposed a quadratic curve model to initialize the contour of the embryo of interest based on edge pixels, and applied an active contour model to refine embryo contours. Bessinger et al. [10] proposed criteria to select the optimal connected component of edge pixels in the scale space of an input image. Li [4] proposed algorithms to detect and restore deficiencies and faults of primal sketch tokens that occur when a targeting object is surrounded by a complex background.

In this paper, we propose a geometric method for contour extraction of *Drosophila* embryos, and the key of the proposed method is k -dominant point extraction. In the context of rectangular shape detection [11], we have proposed 3-dominant point extraction that can be used to analyze the geometric structure of licence plates. Note that the contour of a licence plate is relatively simple—first of all, it is piecewise linear; second, it is convex. We propose to generalize 3-dominant point extraction to k -dominant point extraction in order to adapt to the complex geometric structure of the contour of a *Drosophila* embryo. The complexity of these geometric structures mainly lies in the following two aspects: i) the contour of a *Drosophila* embryo may be concave and ii) the contour of two *Drosophila* embryos that touch to each other may be concave (see Fig. 1). The proposed method is able to segment embryos touching to each other. Note that many methods on contour extraction, including an active contour model, can not segment two objects that touch to each other. The proposed geometric method can also be applied to other tasks relevant to contour extraction.

Methods

3-dominant point extraction

Given a set of points, such as a connected component of edge pixels, Li et al. [12] proposed a recursive method to extract three dominant points v_1, v_2 and v_3 . The first two points (v_1 and v_2) maximize the Euclidean distance of an arbitrary pair of points in C , and the third point v_3 maximizes the sum of distances between a $p \in C$ and $v_i, i = 1, 2$, i.e.,

$$\max_{p \in C} (\|p - v_1\| + \|p - v_2\|). \quad (1)$$

Given a point and a line segment $v_1 v_2$, we call $\|p - v_1\| + \|p - v_2\|$ a *point-line-segment* distance to distinguish the common *point-line* distance that is defined by the distance between p and its vertical intersection with a line passing v_1 and v_2 .

Based on v_1, v_2 , and v_3 , the piecewise linearity of C is then verified, i.e., where each point $p \in C$ falls on either the line segment $v_1 v_3$ or $v_2 v_3$. If yes, the recursion is

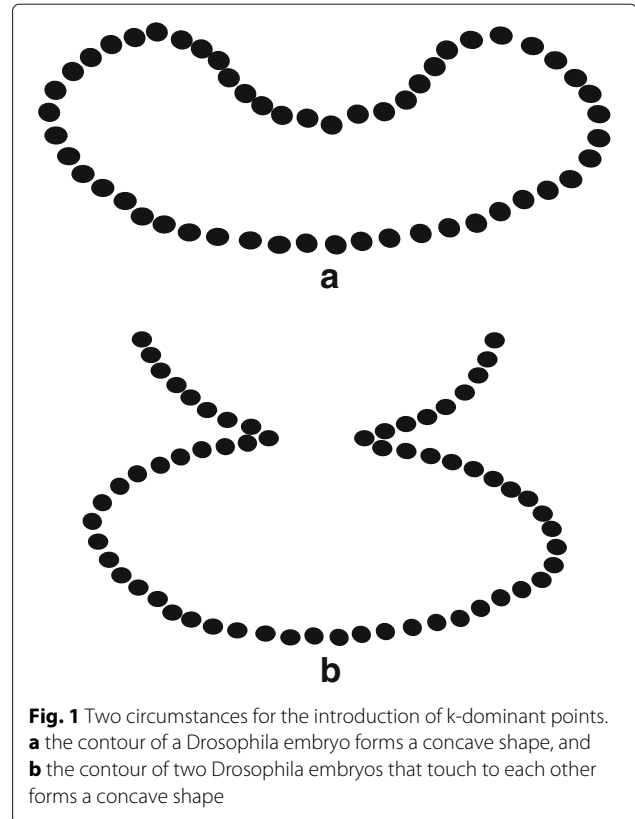


Fig. 1 Two circumstances for the introduction of k -dominant points. **a** the contour of a *Drosophila* embryo forms a concave shape, and **b** the contour of two *Drosophila* embryos that touch to each other forms a concave shape

stop. Otherwise, C is partitioned into two subsets, and the above 3-dominant point extraction method is then applied to the two subsets recursively.

K -dominant point extraction: basic concepts

In this section, we first generalize the approach for locating the 3rd dominant point, given two dominant points, from a set of unordered points P to a formula for locating the i -th dominant point, given i number of dominant points, from P . Then, we propose a simple method to insert a new dominant point into a sequence of geometrically-ordered dominant points so that dominant points are ordered geometrically. Last, we proposed a solution to address the challenge of concave polygons.

The basic idea of k -dominant point extraction is to iteratively insert a new dominant point into a set of $k - 1$ dominant points that have been found, under certain geometric constraint. For the convenience of illustration, we now introduce several basic concepts. First of all, the $k - 1$ dominant points are expected to be a geometric sequence that is consistent with a given set of 2D points.

Denote $\langle r(1), \dots, r(k) \rangle$ is a permutation of $1, \dots, k$. A *geometric sequence* of $k - 1$ dominant points is denoted as $S_{k-1} = \langle v_{r(1)}, v_{r(2)}, \dots, v_{r(k-1)} \rangle$. A valid geometric sequence is expected to be a polyline, i.e., $v_{r(1)} v_{r(2)}$ is the first line segment passing a subset of points, and $v_{r(2)} v_{r(3)}$

is the second line segment passing a subset of points, etc. The initial geometric sequence contains two dominant points, ideally representing a line segment (also called *1-piece polyline*).

A *closed* tag is introduced with respect to a consecutive pair of dominant points $(v_{r(i)}, v_{r(i+1)})$ in a geometric sequence with the motivation of speeding up the insertion of a new dominant point. A consecutive pair with $(v_{r(i)}, v_{r(i+1)})$ a *closed* tag indicates that a new dominant point is not allowed to be inserted between $v_{r(i)}$ and $v_{r(i+1)}$ in the associated geometric sequence.

Insertability is introduced with respect to a new dominant point in order to tell whether the new dominant point is allowed to insert to a geometric sequence of dominant points. Insertability of a point is essentially introduced as a condition to stop the “global” search of dominant points. Imagine that we have a set of points forming a rectangle. After we find out four dominant points associated with the four vertices of a rectangle, the insertability of the fifth dominant point is expected to be NO in order to avoid inserting a non-vertex point into the sequence. Given a geometric sequence $S = \langle v_1, v_2, \dots, v_{k-1} \rangle$ and a point v_k , the point v_k is called (S, ϵ) -*insertable* if the point-line-segment distance between v_k to every pair (v_i, v_{i+1}) is less than or equal to $(1 + \epsilon)$ times the length of the line segment $v_i v_{i+1}$, i.e.,

$$\|v_k - v_i\| + \|v_k - v_{i+1}\| \leq (1 + \epsilon) \|v_i - v_{i+1}\|, \quad \forall i, \quad (2)$$

where ϵ is a parameter to tolerate the distortion of a straight line. ϵ is not a sensitive parameter, and it can be set from 0.01 to 0.05. In this paper, we fix it to be 0.02. Thus, we sometime simply call a point v_k *S*-insertable, or just insertable.

Initialization

Similar to 3-dominant point extraction, *k*-dominant point extraction starts from locating two points v_1 and v_2 , given a point set, such that their Euclidean distance is a maximal distance among distances of all pairs of points in P , i.e.,

$$(v_1, v_2) = \operatorname{argmax}_{p_1 \in P, p_2 \in P} \|p_1 - p_2\|. \quad (3)$$

The closed tags for v_1 and v_2 are both initialized as 0 (i.e., false).

Searching a new dominant point

Given a set of points P and $k - 1$ dominant points v_1, \dots, v_{k-1} , $k > 2$, we propose the following formula to search the k -th dominant point:

$$\max_{p \in P} \sum_{i=1}^{k-1} \|p - v_i\|. \quad (4)$$

For $k = 3$, we have an initial geometry sequence $S_2 = \langle v_1, v_2 \rangle$. Based on Eq. 2, we can test the insertability of the

new dominant point v_3 . For $k > 3$, we can assume that a geometry sequence S_{k-1} has been iteratively built, as described in the following section.

It is worth noting that the time complexity of searching a new dominant point depends on the point set P , i.e., $\Theta(|P|)$. However, testing the insertability of the new dominant point depends on the geometry sequence S only, i.e., in the cost of $\Theta(|S|)$. With closed tags, the computational cost can be further reduced.

Upon an insertable dominant point: growing S_{k-1}

Given an *S*-insertable dominant point, we will insert the point into an open “slot” of the geometric sequence S . Given a sequence S_{k-1} of $k - 1$ dominant points with a geometric order, i.e., $S_{k-1} = \langle v_{r(1)}, v_{r(2)}, \dots, v_{r(k-1)} \rangle$, each open pair $(v_{r(i)}, v_{r(i+1)})$ of S offers a space for an insertable v_k to insert, as follows:

$$S_{k-1,i} = \langle v_{r(1)}, \dots, v_{r(i)}, \square, v_{r(i+1)}, \dots, v_{r(k-1)} \rangle, \quad (5)$$

where \square indicates a possible space where v_k may be inserted. Here the notation $S_{k-1,i}$ represents an *abstract* geometric sequence that contains a placeholder \square between the pair (v_i, v_{i+1}) . Furthermore, we denote $\langle S_{k-1,i}, v_k \rangle$ a *concrete* geometric sequence by replacing the placeholder \square in the abstract sequence $S_{k-1,i}$ by v_k .

For convenience, we introduce $r(k) = r(1)$ and augment the sequence $\langle v_{r(1)}, v_{r(2)}, \dots, v_{r(k-1)} \rangle$ to $\langle v_{r(1)}, v_{r(2)}, \dots, v_{r(k-1)}, v_{r(k)} \rangle$.

We propose two criteria to measure the confidence of a sequence of dominant points $S_k = \langle v_{r(1)}, v_{r(2)}, \dots, v_{r(k)} \rangle$. The first one is:

$$\operatorname{conf}(S_k) = \left| \{p \in C : p \in v_{r(i)} v_{r(i+1)}\} \right|, \quad (6)$$

where $\|\cdot\|$ denotes the cardinality of a set.

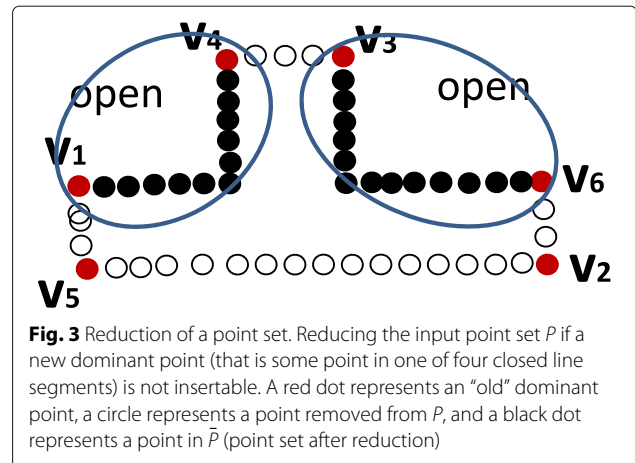
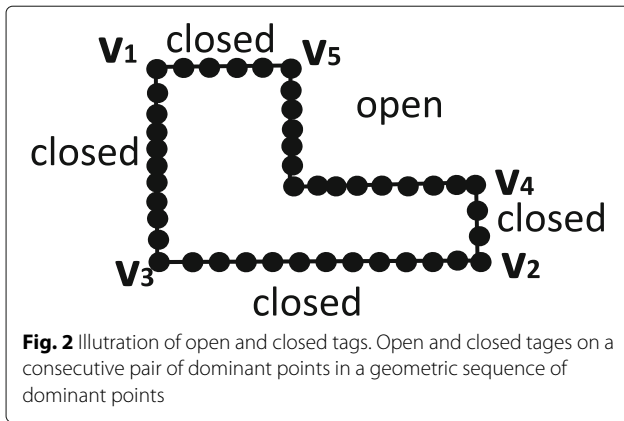
The second criterion is based on overall length of the polyline formed by the sequence S_k , i.e.,

$$\operatorname{conf}(S_k) = \sum_{i=1}^k \|v_{r(i)} - v_{r(i+1)}\|. \quad (7)$$

The second criterion can be applied if C forms a simple curve (no self intersection). It is also easy to see that the computational cost of the second type of confidence is much lower than the first type.

By maximizing the confidence of each derived sequence, we can decide the optimal insertion of a new dominant point in order to maintain the geometric order.

Figure 2 illustrates an example of assigning a geometric order of k dominant points. Given the first dominant points v_1 and v_2 . Without lose of generality, we start from the geometric sequence $\langle v_1, v_2 \rangle$. After v_3 is computed by Eq. 4, there are two possible options to insert v_3 : i) $\langle v_1, \square, v_2 \rangle$ and $\langle v_1, v_2, \square \rangle$. The first option brings us the sequence $\langle v_1, v_3, v_2 \rangle$, where closed tag assignment



is $(1, 1, 0)$, meaning that v_1v_3 and v_3v_2 are closed, and v_2v_1 is open. The second option brings us the sequence $\langle v_1, v_2, v_3 \rangle$, where closed tag assignment is $(0, 1, 1)$, meaning that v_1v_2 is open, and v_2v_3, v_3v_1 are both closed. Based on the confidence measure, these two options are both optimal. Without loss of generality, we choose the first option for following illustration. Consecutively, the proposed method grows the geometric sequence as follows: i) $\langle v_1, v_3, v_2 \rangle$ with closed tags $(1, 1, 0)$; ii) $\langle v_1, v_3, v_2, v_4 \rangle$ with closed tags $(1, 1, 1, 0)$; iii) $\langle v_1, v_3, v_2, v_4, v_5 \rangle$ with closed tags $(1, 1, 1, 0, 1)$.

Upon a non-insertable dominant point: reduce P

If a new dominant point v_i computed by Eq. 4 is non-insertable, we will stop growing the geometric sequence and start to reduce the input point set P . The basic idea of reducing P is to remove all points that lie in one of closed line segments such that we obtain a subset of points that have a simpler topology. A point remained in P must be associated with a certain open pair of dominant points. So the above method can be recursively applied to each subset of points, and in turn each output (a sequence of dominant points from a subset of points can be correctly inserted into a higher level output). Simply speaking, a non-insertable dominant point activates a divide-and-conquer strategy that can handle the concavity of a data set shaped by a concave polygon.

Figure 3 illustrates the basic idea on how to reduce P when a new dominant point is tested to be non-insertable. After reduction, there are two subsets of points since there are two open pairs of dominant points in the geometric sequence. K-dominant point extraction method is then recursively applied to these two point sets, respectively.

Algorithm 1 summarizes the procedure of testing the closedness of a pair of dominant points (v, w) , given a point set P . Note that if closedness is true, P will be updated by removing all points near the line segment vw . The time complexity of the algorithm is dominated by Step 5 (sorting) that is $O(|P| \log(|P|))$. Algorithm 2

summarizes the recursive implementation for k-dominant point extraction. The time complexity of the algorithm is dominated by two non-recursive steps: i) Step 4 (the initialization of S) that is $O(|P|^2)$, and ii) Step 13 (closedness) that is $O(|P| \log(|P|))$, in addition to the recursive step, i.e., Step 16. In many cases, the concavity is not very complicated and the depth of recursion won't be over 2. Therefore, the overall time complexity of the Algorithm is $O(|P|^2(1 + \log(|P|)^2))$ with an assumption that the number of vertices of a polygon is a constant and points in P can be uniformly projected to S .

Limitation of max-sum-distance measure

Figure 4 shows three scenarios of point sets after the first three dominant points have been located. Figure 4a shows a scenario where the fourth dominant point is non-insertable with respect to the sequence $\langle v_1, v_2, v_3 \rangle$, however a point set is non-reducible. More specifically,

Algorithm 1 Closedness of a pair of dominant points

- | | | |
|---------|-----------|---------------------------|
| input: | (v, w) | a pair of dominant points |
| | P | a point set |
| output: | closedTag | true/false |
| | P | updated P |
1. $P' = \{p \in P : \|p - v\| + \|p - w\| \leq (1 + \epsilon)\|v - w\|\}$
 2. $d = \frac{w-v}{\|w-v\|}$ // unit directional vector
 3. $P_{proj} = \{(p - v) \cdot d : p \in P'\}$
 4. $P_p = \{p \in P_{proj} : p \geq 0 \wedge p \leq \|v - w\|\}$
 5. $P_s = \text{sort } P_p \text{ in an increasing order}$
 6. $Gap = \{p_{i+1} - p_i : p_i \in P_s\}$
 7. if $\max(Gap) \leq \sqrt{2}$ // based on 8-connectivity
 8. closedTag = true
 9. $P = P \setminus P'$ // remove points on line segment vw
 10. else
 11. closedTag = false

Algorithm 2 k-dominant point extraction (kdp)

```

input:   a 2D point set  $P = \{p_i\}_{i=1,\dots,n}$ 
output:  a geometric sequence  $S$ 
1. if  $P$  is empty // base case
2.   return  $S = \emptyset$ 
3. else // recursive case
4.   initiate  $S = \langle v_1, v_2 \rangle$  by Eq. 3
5.    $i = 3$ ; // start from the third dominant point
6.   while  $P$  is not empty
7.     compute  $v_i$ 
8.     if  $v_i$  is insertable for  $S$ 
9.        $S = \text{insert}(v_i, S)$ 
10.     $i++$ 
11.   else
12.     break; // stop growing sequence
13.   for each consecutive pair  $(v_i, v_{i+1}) \in S$ 
14.      $(\text{closedTag}_i, P) = \text{closedness}((v_i, v_{i+1}), P)$ 
15.     clustering  $P$  based on open pairs, i.e.,
         $P = P_1 \cup \dots \cup P_m, \text{ where } P_i \cap P_m = \emptyset, \forall i \neq j$ 
16.   return  $S = \langle S_1, \text{kdp}(P_1), \dots, S_m, \text{kdp}(P_m) \rangle$ 

```

given a point set (a number of black dots) as illustrated in the figure, assume that the first three dominant points have already been located according to the proposed kdp method. Since the max function in Eq. 4 is a convex function, the fourth dominant point is expected to be near to the boundary of the convex hull of the point set (which is equivalent to the triangle with vertices v_1, v_2 and v_3). In other words, the fourth dominant point must be a point nearest to either v_1, v_2 or v_3 . Based on the definition of “insertible”, the fourth point is non-insertible with respect to the geomtric sequence v_1, v_2 and v_3 . According to the above algorithm, we now stop growing the sequence, and

start to reduce the given point set P . However, it turns out that P is not reducible. This scenario shows a limitation of using max-sum-distance in computing k-dominant points. Note that this scenario is a representative formulation of two objects who have smooth boundaries and are touching to each other.

Figure 4b shows a scenario where the fourth dominant point is non-insertible and the point set is reducible with respect to $v_i, i = 1, \dots, 3$. The fourth dominant point is expected to be a point on the line segment v_2v_3 . Precisely speaking, the fourth dominant point is the point nearest to v_2 as $\|v_2 - v_1\| > \|v_3 - v_1\|$. Note that for any point p on the line segment, the sum of distances $v_2v_3, \|p - v_2\| + \|p - v_3\|$ is a constant. If $\|v_2 - v_1\| = \|v_3 - v_1\|$, the fourth dominant point can be an arbitrary point on the line segment v_2v_3 . Unlike the scenario Fig. 4a where all pairs of dominant points are open, the point set in Fig. 4b is reducible because (v_2, v_3) forms a closed pair.

Figure 4c shows a scenario where the fourth dominant point is insertible. Therefore, the algorithm will continue to locate the fifth dominant point. Reducibility is thus not applicable in this scenario.

The rationale of introducing a min-sum-distance can also be found by a Fermat point and its generalization called a geometric median. Given three points in a plane, the *Fermat point* is the point in the plane that minimizes the sum of distances from itself to the three points. The Fermat point can be computed analytically, as illustrated in Fig. 5. Specifically, we first construct an equilateral triangle (in dash lines) for each edge of the given triangle (in solid lines), and then connect the outmost vertex of each equilateral triangle to a given vertex outside the equilateral triangle by a red line. The Fermat point must be the intersection of three red lines. Given m points in a plane, the *geometric median* is the point in the plane that minimizes the sum of distances from itself to the m points.

A subtle difference between computing a geometric median and computing a dominant point under the min-sum-distance scheme is that the search space in the former problem is a continuous and infinite plane, while the search space in the latter problem is a discrete and finite point set. For convenience, we call the problem of computing a dominant point under a min-sum-distance scheme *discrete geometric median*.

Max/min-sum-distance measure

The three scenarios illustrated in Fig. 4 show that a max-sum-distance scheme is not able to handle various structures of a point set to extract dominant points. A min-sum-distance scheme is expected to be integrated with a max-sum-distance scheme adaptively. In other words, we have to compare both schemes, a max-sum-distance and a min-sum-distance, and select a better scheme under some criterion to extract the next dominant

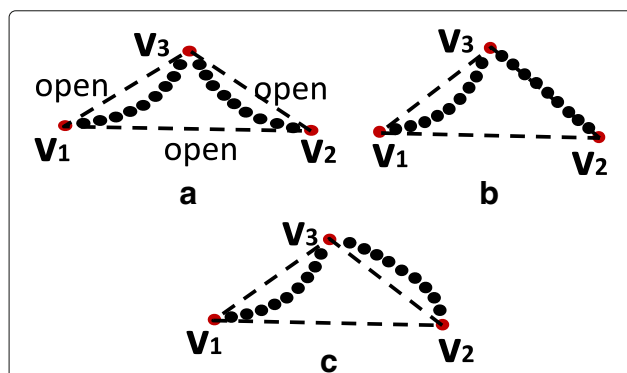
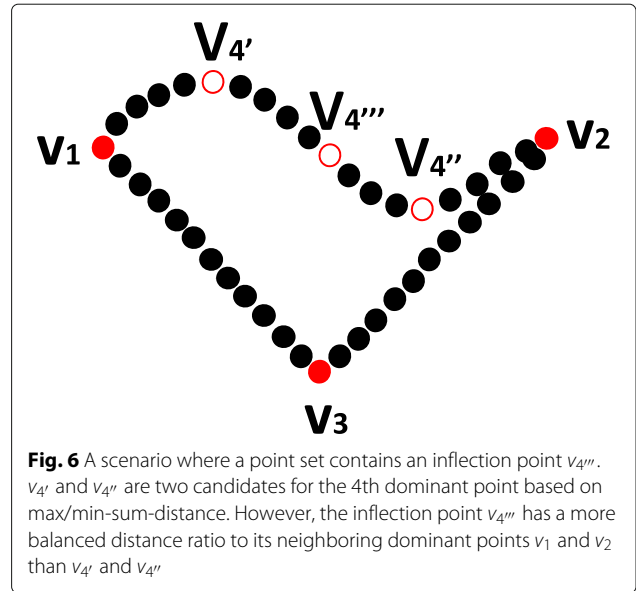
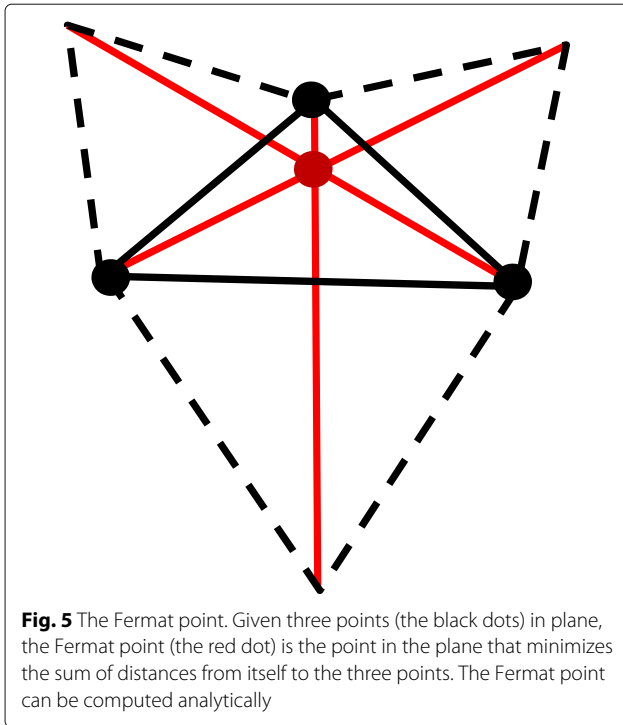


Fig. 4 Different scenarios on dominant points and point sets. **a** A scenario where the fourth dominant point is non-insertible, and the point set is non-reducible with respect to $v_i, i = 1, \dots, 3$; **b** A scenario where the fourth dominant point is non-insertible, and the point set is reducible with respect to $v_i, i = 1, \dots, 3$; **c** A scenario where the fourth dominant point is insertible with respect to $v_i, i = 1, \dots, 3$, and the algorithm will continue to locate the fifth dominant point



point during the stage of growing the geometric sequence of dominant points. For convenience, we call the problem as *max/min-sum-distance measure problem*.

To solve the max/min-sum-distance problem, we propose a balance-oriented criterion to select an optimal measure between max-sum-distance and min-sum-distance measure as follows: *balance the distance from a new dominant point to its two neighboring dominant points in the new geometric sequence*. Specifically, given a point set P and a geometric sequence of dominant points $S_{k-1} = \langle v_1, \dots, v_{k-1} \rangle$, we have two candidates of a new dominant point: i) v_k^{\max} according to the max-sum-distance measure, and ii) v_k^{\min} according to the min-sum-distance measure. For each candidate, v_k^{\max} or v_k^{\min} , we first compute the optimal location to insert by maximizing the confidence of a derived sequence, i.e.,

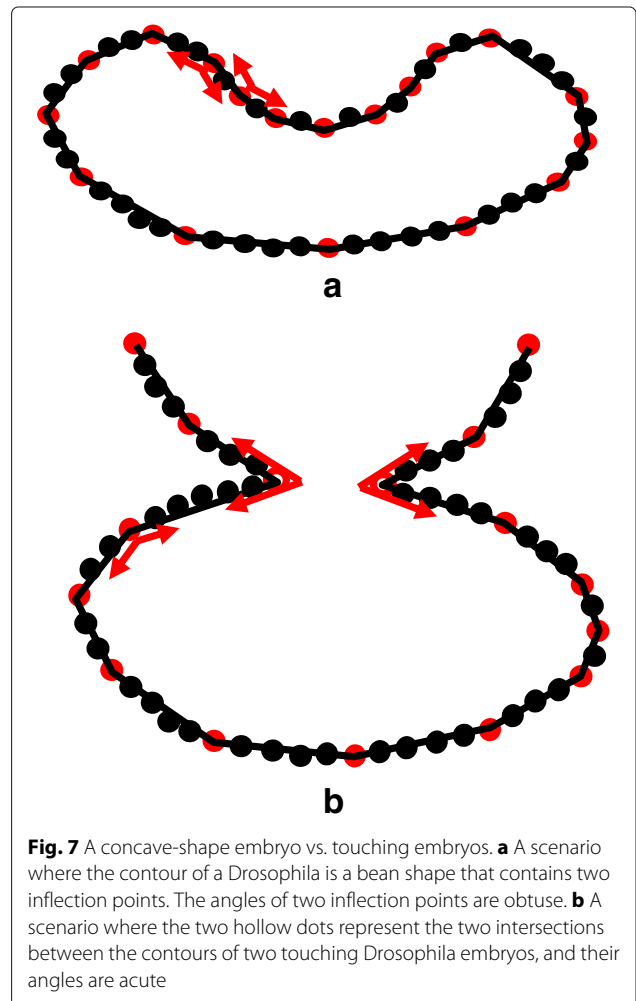
$$i^{*,\max} = \operatorname{argmax}_i \operatorname{conf} (\langle S_{k-1,i}, v_k^{\max} \rangle)$$

$$i^{*,\min} = \operatorname{argmax}_i \operatorname{conf} (\langle S_{k-1,i}, v_k^{\min} \rangle)$$

We then compute a distance ratio with respect to v_k^{\max} , v_k^{\min} and their neighboring dominant points as follows:

$$r^{\max} = \frac{\min (\|v_k^{\max} - v_{i^{*,\max}}\|, \|v_k^{\max} - v_{i^{*,\max}+1}\|)}{\max (\|v_k^{\max} - v_{i^{*,\max}}\|, \|v_k^{\max} - v_{i^{*,\max}+1}\|)}$$

$$r^{\min} = \frac{\min (\|v_k^{\min} - v_{i^{*,\min}}\|, \|v_k^{\min} - v_{i^{*,\min}+1}\|)}{\max (\|v_k^{\min} - v_{i^{*,\min}}\|, \|v_k^{\min} - v_{i^{*,\min}+1}\|)}$$



Note that both r^{\max} and r^{\min} range from 0 to 1. A larger value indicates more balanced distance from a candidate to its optimal neighbors. So, if $r^{\max} \geq r^{\min}$, the candidate v_k^{\max} is selected. Otherwise, v_k^{\min} is selected.

It is intuitive that dominant points selected by a balance-oriented criterion have a better description of the global structure of a point set, and thus they can provide more reliable estimation of geometric properties of a point set, such as curvature. In contrast, if a dominant point has an imbalanced distance ratio to its two neighbors, the estimation of a geometric property at this dominant point will be very sensitive to the localisation error of itself and its neighbors.

It is easy to see that the balanced-oriented criterion on distance measure can address the above-mentioned three scenarios very well. It is also not difficult to show that the max/min-sum-distance measure can be consistently selected if a point set forms a convex polygonal shape.

Limitation of max/min-sum-distance measure

Figure 6 shows a scenario where a point set contains an inflection point $v_{4''}$. Since an inflection point is neither a local maximum or a local minimum, the max/min-sum-distance measure fails to extract such a point as a dominant point at an “early” round. (It is possible that an inflection point can be eventually extracted after a few more rounds.) More specifically, three dominant points v_1, v_2 and v_3 are first extracted without any problem. Based on max/min-sum-distance measure, there are two candidates for the 4th dominant point: i) $v_{4'}$ according to max-sum-distance measure, and ii) $v_{4''}$ according to min-sum-distance measure. However, the inflection point

$v_{4''}$ has a more balanced distance ratio to its neighboring dominant points v_1 and v_2 than $v_{4'}$ and $v_{4''}$. Thus, a inflection point, intuitively, can have a better description of the global structure of a point set, say, a sequence of dominant points that can response to inflection points can cover a larger number of points in a given point set than an equal-length sequence of dominant points that can't response to inflection points.

Max/min/median-sum-distance measure

We now generalize the max/min-sum-distance measure to a max/min/median-sum-distance measure. We follow a balance-oriented criterion to select an optimal measure among three measures: i) max-sum-distance, ii) min-sum-distance measure, and iii) median-sum-distance as follows: *balance the distance from a new dominant point to its two neighboring dominant points in the new geometric sequence.* Specifically, given a point set P and a geometric sequence of dominant points $S_{k-1} = \langle v_1, \dots, v_{k-1} \rangle$, we have two candidates of a new dominant point: i) v_k^{\max} according to the max-sum-distance measure, ii) v_k^{\min} according to the min-sum-distance measure, and iii) v_k^{mdn} according to the median-sum-distance measure. For each candidate, v_k^{\max} , v_k^{\min} or v_k^{mdn} , we first compute the optimal location to insert by maximizing the confidence of a derived sequence, i.e.,

$$i^{*,\max} = \operatorname{argmax}_i \operatorname{conf} \left(\langle S_{k-1,i}, v_k^{\max} \rangle \right)$$

$$i^{*,\min} = \operatorname{argmax}_i \operatorname{conf} \left(\langle S_{k-1,i}, v_k^{\min} \rangle \right)$$

$$i^{*,\text{mdn}} = \operatorname{argmax}_i \operatorname{conf} \left(\langle S_{k-1,i}, v_k^{\text{mdn}} \rangle \right)$$

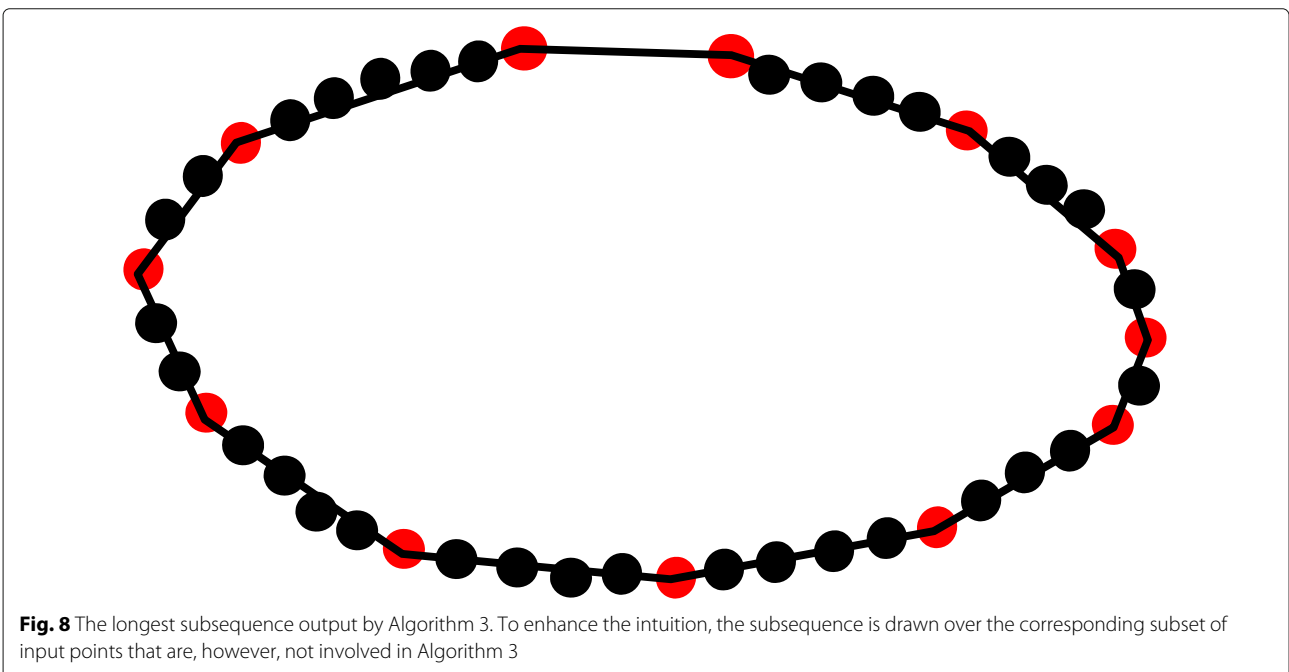


Fig. 8 The longest subsequence output by Algorithm 3. To enhance the intuition, the subsequence is drawn over the corresponding subset of input points that are, however, not involved in Algorithm 3

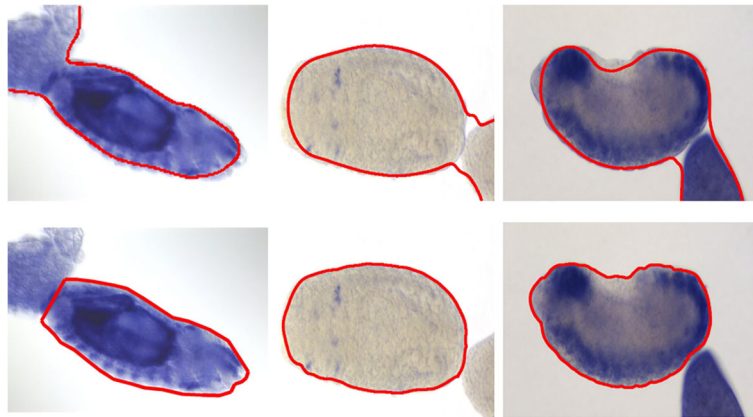


Fig. 9 A comparison between Li's method [4] (the first row) and the proposed method (the second row)

We then compute a distance ratio with respect to v_k^{\max} , v_k^{\min} , v_k^{mdn} , and their neighboring dominant points as follows:

$$r^{\max} = \frac{\min(\|v_k^{\max} - v_{i^*,\max}\|, \|v_k^{\max} - v_{i^*,\max+1}\|)}{\max(\|v_k^{\max} - v_{i^*,\max}\|, \|v_k^{\max} - v_{i^*,\max+1}\|)}$$

$$r^{\min} = \frac{\min(\|v_k^{\min} - v_{i^*,\min}\|, \|v_k^{\min} - v_{i^*,\min+1}\|)}{\max(\|v_k^{\min} - v_{i^*,\min}\|, \|v_k^{\min} - v_{i^*,\min+1}\|)}$$

$$r^{\text{mdn}} = \frac{\min(\|v_k^{\text{mdn}} - v_{i^*,\text{mdn}}\|, \|v_k^{\text{mdn}} - v_{i^*,\text{mdn}+1}\|)}{\max(\|v_k^{\text{mdn}} - v_{i^*,\text{mdn}}\|, \|v_k^{\text{mdn}} - v_{i^*,\text{mdn}+1}\|)}$$

Note that r^{\max} , r^{\min} and r^{mdn} range from 0 to 1. The largest value indicates the most balanced distance from a candidate to its optimal neighbors. For example, if r^{\max} is the largest value, the candidate v_k^{\max} will be selected.

Sequence subdivision

We propose a sequence subdivision method for the segmentation of touching embryos. The basic idea of our method is based on the detection of nonsmooth dominant points from a given geometric sequence. To build up an intuition, let us start from Fig. 7 that shows a comparison between a scenario of a concave-shape embryo and a scenario of touching embryos. Given a dominant point v_i , denote $v_{i-1}v_i = v_{i-1} - v_i$ and $v_{i+1}v_i = v_{i+1} - v_i$ are two directional vectors centered at v_i . The cross angle between these two vectors that can be computed by $\arccos \frac{v_{i-1}v_i \cdot v_{i+1}v_i}{\|v_{i-1}v_i\| \|v_{i+1}v_i\|}$ can be used to tell whether a dominant point is smooth or not. For simplicity, we call the cross angle of the two directional vectors centered at v_i the *angle* of v_i . Figure 7a shows a scenario where the two hollow dots represent the two intersections between the contours of two touching *Drosophila* embryos, and their angles are acute. Figure 7b shows a scenario where the contour of a *Drosophila* is a bean

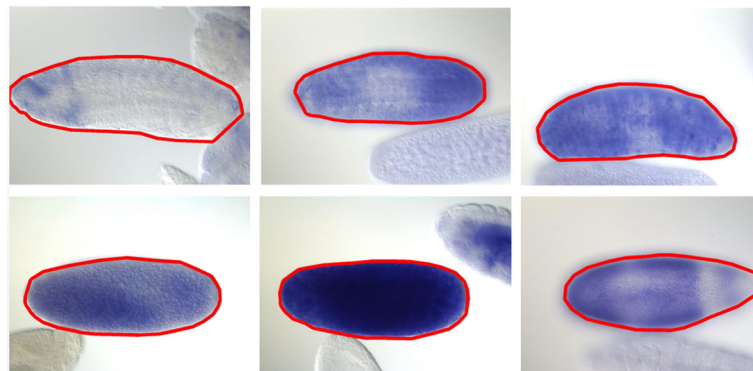


Fig. 10 More positive examples of the proposed algorithm on images of touching *Drosophila* embryos

shape that contains two inflection points. The angles of two inflection points are obtuse, i.e., similar to the cross angle of other dominant points given a smooth contour.

Figure 8 shows the longest subsequence output by the Algorithm 3 being applied to the geometric sequence presented in Fig. 1b.

Results

We tested the proposed method on BDGP (Berkley Drosophila Genome Project) [1]. BDGP images are available at the following public webpage: http://www.fruitfly.org/insituimages/insitu_images/.

We first give a visual evaluation on the proposed method. Figure 9 shows a comparison between Li's method [4] and the proposed method on BDGP images of touching Drosophila embryos. We can observe that Li's method fails in all three cases, while the proposed method works well. Figure 10 shows more positive results of the proposed method.

Algorithm 3 Sequence subdivision

input: $S = \langle v_1, v_2, \dots, v_n \rangle$ with closedTags
 output: $\{S_k\}$: a set of subsequences

1. for $i=1:n$
2. if $\text{angle}(v_i)$ is acute // detection
3. closedTag(v_i) = 0; // tag is set to open
4. $k = 1$; $S_k = \langle \rangle$ // init 1st subsequence
5. $i = 1$;
6. while $i \leq n$
7. insert v_i into S_k until closedTag(v_i)=0;
8. $k++$; $S_k = \langle \rangle$ // init next subsequence
9. $i++$;

Next, we present a quantitative evaluation of different combinations of the parts of the proposed method in terms of detection rates. Given an image, the result is defined as a *successful detection* if the output by Algorithm 2 has larger than 90% overlapping region with the ground truth. The dataset contains 2000 images of BDGP Drosophila embryos. Table 1 shows the quantitative results, including the comparison of two existing methods on contour extraction of Drosophila embryos: i)

Table 1 Quantitative results

Method	Detection rate (%)
Li and Kambhamettu [3]	90
Li [4]	92
Proposed	94

Li and Kambhamettu's method [3] that consists an initialization based on a quadratic curve model, and a refinement based on an active contour model; ii) Li's method [4] that can detect and restore deficiencies and faults of primal sketch tokens occurring when a targeting object is surrounded by a complex background.

Discussion

As we mentioned in the introduction, the proposed framework can be applied to other applications of contour extraction. The main contribution of the proposed framework is k -dominant point extraction based on a specific distance measure. There is a trade off among the three types of distance measures: max-sum, max/min-sum, and max/min/median-sum. The last one is the most sophisticated one, i.e., it can deal with concave contours, and touching scenarios of targeting objects, while the first one is the most efficient one in computation. Therefore, max-sum distance may be used in some circumstance, e.g., contours are convex. K -dominant point extraction may also be applied to more general applications beyond contour extraction, such as data clustering. In other words, the data that k -dominant point extraction is applied can have arbitrary dimensions rather than two.

Conclusions

We have proposed a geometric method for contour extraction of Drosophila embryos. Experiment results show the effectiveness of the proposed method, typically in segmenting two touching embryos in an image. The results also show the superiority of the proposed method over two previous methods. The proposed method advances the theory of control point detection by generalizing 3-dominant points to k -dominant points. The generalization includes strategies to deal with concave shapes, thus the proposed method can be applied to a wide range of applications relevant to contour extraction.

Acknowledgements

We thank editors and reviewers constructive comments and suggestions in improving the quality of the paper.

Funding

This work was supported by National Science Foundation Grant of China 61370160, Guangdong Province Natural Science Foundation Project 2015A030313578, and Guangdong Scientific and Technological Plan Project 2015B010106005. Publication costs were funded by National Science Foundation Grant of China 61370160.

Availability of data and materials

BDGP data is available in the public website: http://www.fruitfly.org/insituimages/insitu_images/.

About this supplement

This article has been published as part of *BMC Systems Biology* Volume 11 Supplement 6, 2017: Selected articles from the IEEE BIBM International Conference on Bioinformatics & Biomedicine (BIBM) 2016: systems biology. The full contents of the supplement are available online at <https://bmcsystbiol.biomedcentral.com/articles/supplements/volume-11-supplement-6>.

Authors' contributions

QL conceived the study, designed algorithms, implemented software, carried out analyses, and wrote paper. YG carried out analyses and wrote paper. Both authors read and approved the final manuscript.

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Author details

¹Western Kentucky University, 1906 College Blvd., 42101 Bowling Green, KY, USA. ²Cisco School of Informatics, Guangdong University of Foreign Studies, 510006 Guangzhou, People's Republic of China.

Published: 14 December 2017

References

- Tomancak P, et al. Systematic determination of patterns of gene expression during drosophila embryogenesis. *Genome Biol.* 2002;3(12):1–14.
- Kumar S, Jayaraman K, Panchanathan S, Gurunathan R, Marti-Subirana A, Newfield SJ. Best: A novel computational approach for comparing gene expression patterns from early stages of drosophila melanogaster development. *Genetics.* 2002;16(4):2037–47.
- Li Q, Kambhamettu C. Contour extraction of drosophila embryos. *IEEE/ACM Trans Comput Biol Bioinforma.* 2011;8(6):1509–21.
- Li Q. A primal sketch based framework for bean-shape contour extraction. *Neurocomputing.* 2014;142:508–19.
- Peng H, Myers EW. Comparing *in situ* mRNA expression patterns of *drosophila* embryos. In: RECOMB. Springer. 2004. p. 157–66.
- Pan JY, Balan AGR, Xing EP, Traina AJM, Faloutsos C. Automatic mining of fruit fly embryo images. In: KDD. ACM. 2006. p. 693–8.
- Puniyani K, Faloutsos C, Xing EP. Spex2: Automated concise extraction of spatial gene expression patterns from fly embryo ish images. *Bioinformatics.* 2010;26(12):47–56.
- Frise E, Hammonds AS, Celniker SE. Systematic image-driven analysis of the spatial drosophila embryonic expression landscape. *Mol Syst Biol.* 2010;6:345. <http://msb.embopress.org/content/6/1/345>.
- Mace DL, Varnado N, Zhang W, Frise E, Ohler U. Extraction and comparison of gene expression patterns from 2d rna in situ hybridization images. *Bioinformatics.* 2010;15(26(6)):761–9.
- Bessinger Z, Xing G, Li Q. Localization of drosophila embryos using connected components in scale space. In: 19th IEEE International Conference on Image Processing, ICIP 2012, Lake Buena Vista, Orlando, FL, USA, September 30 - October 3, 2012. IEEE. 2012. p. 497–500.
- Li Q. A geometric framework for rectangular shape detection. *IEEE Trans Image Process.* 2014;23(9):4139–49.
- Li Q, Liang G, Gong Y. A geometric framework for stop sign detection. In: IEEE China Summit and International Conference on Signal and Information Processing, ChinaSIP 2015, Chengdu, China, July 12-15, 2015. IEEE. 2015. p. 258–62.

Submit your next manuscript to BioMed Central and we will help you at every step:

- We accept pre-submission inquiries
- Our selector tool helps you to find the most relevant journal
- We provide round the clock customer support
- Convenient online submission
- Thorough peer review
- Inclusion in PubMed and all major indexing services
- Maximum visibility for your research

Submit your manuscript at
www.biomedcentral.com/submit

