# A Robust Ellipse Fitting Algorithm Based on Sparsity of Outliers

Elaheh Sobhani[*], Mostafa Sadeghi[*], Massoud Babaie-Zadeh[*], Christian Jutten[†]

[*]Electrical Engineering Department, Sharif University of Technology, Tehran, Iran.

Email: sobhani.es@gmail.com, m.saadeghii@gmail.com, mbzadeh@yahoo.com

[†]GIPSA-Lab, Grenoble, and Institut Universitaire de France, France.

Email: christian.jutten@gipsa-lab.grenoble-inp.fr

*Abstract*—**Ellipse fitting is widely used in computer vision and pattern recognition algorithms such as object segmentation and pupil/eye tracking. Generally, ellipse fitting is finding the best ellipse parameters that can be fitted on a set of data points, which are usually noisy and contain outliers. The algorithms of fitting the best ellipse should be both suitable for real-time applications and robust against noise and outliers. In this paper, we introduce a new method of ellipse fitting which is based on sparsity of outliers and robust Huber's data fitting measure. We will see that firstly this approach is theoretically better justified than a state-of-the-art ellipse fitting algorithm based on sparse representation. Secondly, simulation results show that it provides a better robustness against outliers compared to some previous ellipse fitting approaches, while being even faster.**

## I. INTRODUCTION

Ellipse fitting solutions are followed in two main classes: clustering-like methods [1] and least squares (LS) methods [2], [3]. Since an ellipse is determined by five parameters, clustering-like methods, *e.g.* Hough transform [1], [4] and its variants, will require heavy computations because of this five dimensional search space [5]. LS techniques focus on finding the best ellipse parameters, which minimize the sum of the squares of the errors, that is, the distances of data points from the estimated ellipse. Depending on the definition of the distance, these methods are divided into two approaches: geometric methods and algebraic methods [5]. Geometric methods are mainly based on the geometric distance, which is defined as the orthogonal distance of a point from an ellipse. These methods require non-linear optimization and heavy computations [6].

Algebraic methods are based on minimizing "algebraic distances" of data points from the ellipse. The "algebraic distance" is described in Section II in details, but in brief, it is the conic equation value for each data point. Some of the algebraic methods such as direct least squares (DLS) [3] intend to find the best ellipse parameters based on using all data points. On the other hand, some other methods select specific data points, and use them to estimate the parameters. Recently, a new method of this group, which is based on the sparse representation theory [7], is proposed in [5]. However, as will be explained in Section II-B, we believe that ellipse parameters cannot be represented as a sparse combination of data points. To verify this point, in addition to the theoretical

discussions of Section II-B, we design some simulations in Section IV.

Then in this paper, we develop a new ellipse fitting approach, which is based on algebraic distance minimization. Our new algorithm uses the well-known robust Huber function as a data fidelity measure [8]. In [9], it has been shown that minimizing the sum of Huber functions evaluated on residuals is equivalent to adding variables for modeling outliers. These variables are responsible for distinguishing outliers from inliers. Thus, ellipse parameters can be determined from the inliers detected through the algorithm. Our proposed algorithm is a least squares approach based on sparse representation theory. It will be seen that our algorithm is highly faster than Hough Transform approach of [1], [4] and the algorithm of [5]. Moreover, our algorithm is more robust than [5] to the number of data points or their locations.

The rest of this paper is organized as follows. The problem formulation of ellipse fitting and a brief review on the previous methods are stated in Section II. Our proposed method is presented in Section III, while Section IV is devoted to experimental results.

## II. ELLIPSE FITTING

### A. Problem Formulation

Consider an ellipse in the $xy$-plane, centered at $(c_x, c_y)$ having semi-axes with lengths $r_x$ and $r_y$ and a counterclockwise rotation, $\theta$. In the rest of the paper, we refer to $[c_x, c_y, r_x, r_y, \theta]$ as *ellipse characteristics vector*.

So, in the $xy$-coordinate system a general ellipse has the following form

$$\frac{((x - c_x)\cos\theta + (y - c_y)\sin\theta)^2}{r_x{}^2} + \frac{(-(x - c_x)\sin\theta + (y - c_y)\cos\theta)^2}{r_y{}^2} = 1. \tag{1}$$

A general conic equation in inhomogeneous coordinates is $F(\mathbf{a}, \mathbf{x}) = \mathbf{a}^T \cdot \mathbf{x} = ax^2 + bxy + cy^2 + dx + ey + f = 0$, where $\mathbf{a} \triangleq [a, b, c, d, e, f]^T$ and $\mathbf{x} \triangleq [x^2, xy, y^2, x, y, 1]^T$. By defining $\mathbf{x}_i \triangleq [x_i^2, x_iy_i, y_i^2, x_i, y_i, 1]^T$, the value of $F(\mathbf{a}, \mathbf{x}_i)$ is called the "algebraic distance" of the point $(x_i, y_i)$ from the conic $F(\mathbf{a}, \mathbf{x}) = 0$ [10]. The use of algebraic distances for conic-fitting was first introduced by Bookstein [10]. Bookstein

showed that although the conic-fitting based on minimizing the algebraic distances is not geometrically or statistically intuitive, this approach with an appropriate choice of normalization can give suitable results.

### B. Previous Methods

In this subsection, the methods that use the minimization of algebraic distances to fit the best ellipse on the 2D data points are briefly reviewed. Mathematically, the underlying problem is [3]

$$\min_{\mathbf{a}} \sum_{i=1}^{I} F^2(\mathbf{a}, \mathbf{x}_i), \tag{2}$$

where $I$ is the total number of data points.

In order to ensure that any multiple of the optimized vector $\mathbf{a}$ corresponds to the same conic, and to avoid the trivial solution $[a, b, c, d, e, f]^T = \mathbf{0}_{6\times1}$, a normalization constraint such as $\|\mathbf{a}\|^2 = 1$, $a+c = 1$ [11], [12], $a^2 + \frac{1}{2}b^2 + c^2 = 0$ [10], $f = 1$ [11], or $f = -1$ [5] have been proposed to be added to (2).

In this paper, for the normalization constraint, we use $f = -1$ as in [5]. Furthermore, similar to [5], the vectors $\mathbf{a}$ and $\mathbf{x}$ are re-defined as

$$\mathbf{a} \triangleq [a, b, c, d, e]^T, \tag{3}$$
$$\mathbf{x} \triangleq [x^2, xy, y^2, x, y]^T. \tag{4}$$

Moreover, we refer to $\mathbf{a}$ as *ellipse parameters vector*.

In matrix form, (2) is re-written as

$$\min_{\mathbf{a}} \|\mathbf{X}\mathbf{a} - \mathbf{1}_{I\times1}\|_2^2, \tag{5}$$

where $\mathbf{X} \triangleq [\mathbf{x}_1 \mathbf{x}_2 \ldots \mathbf{x}_I]^T$, $\mathbf{x}_i \triangleq [x_i^2, x_i y_i, y_i^2, x_i, y_i]^T$, and $\mathbf{1}_{I\times1}$ is a vector whose elements are all equal to one. The optimal solution of (5) is given as

$$\mathbf{a} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{1}_{I\times1}. \tag{6}$$

Sensitivity to outliers is a common disadvantage of (5) and its variants (LS methods) [2], [5]. Since (6) implies that the ellipse parameters vector is the combination of all data points, Liang *et. al.* [5] proposed that the ellipse parameters vector is obtained through linear combination of "more accurate" data points. In other words the authors of [5] proposed to write

$$\mathbf{a} = \sum_{i=1}^{I} b_i \mathbf{x}_i = \mathbf{X}^T\mathbf{b}, \tag{7}$$

where $\mathbf{b}$ is a sparse vector. Then by substituting (7) in (5), they obtained

$$\min_{\mathbf{b}} \|\overline{\mathbf{X}}\mathbf{b} - \mathbf{1}_{I\times1}\|_2^2,$$

where

$$\overline{\mathbf{X}} \triangleq \mathbf{X}\mathbf{X}^T. \tag{8}$$

So, taking into account the sparsity constraint of $\mathbf{b}$, the authors of [5] proposed the minimization problem

$$\min_{\mathbf{b}} \|\overline{\mathbf{X}}\mathbf{b} - \mathbf{1}_{I\times1}\|_2^2 + \lambda\|\mathbf{b}\|_1. \tag{9}$$

As mentioned in [5], (9) works well under the assumption that the noise of the data points follows a Gaussian distribution, but it is sensitive to random outliers with extreme noise values. Hence, they proposed to use $\ell_1$ term instead of $\ell_2$ term to reduce this sensitivity:

$$\min_{\mathbf{b}} \|\overline{\mathbf{X}}\mathbf{b} - \mathbf{1}_{I\times1}\|_1 + \lambda\|\mathbf{b}\|_1. \tag{10}$$

We call this method, that is, using (10) where $\overline{\mathbf{X}}$ is defined as (8), *"Liang's method"* throughout the rest of the paper.

We believe that the underlying justification of (7) in this method is not relevant. Indeed, even forgetting the sparsity of b, (7) implies that the ellipse parameters vector (defined in (3)) is stated as a linear combination of the data points $\mathbf{x}_i$ (defined in (4)). However, such a property does not necessarily hold for an ellipse. Actually, we believe that (7) does not inherently use the "data" matrix $\mathbf{X}^T$ as the sparsifying dictionary, and that the dictionary $\mathbf{X}^T$ can be replaced by almost any other overcomplete dictionary $\mathbf{D}$ of the same size. In other words, we claim that one can use the following equation instead of (7) without meaningful changes in final results:

$$\mathbf{a} = \mathbf{D}\mathbf{b}, \tag{11}$$

where $\mathbf{D}$ is an almost arbitrary matrix with the same size as $\mathbf{X}^T$. This results in the same optimization problem as in (10), with the re-definition of $\overline{\mathbf{X}}$ as

$$\overline{\mathbf{X}} \triangleq \mathbf{X}\mathbf{D}. \tag{12}$$

Besides the above theoretical justification, this claim will be experimentally studied by simulation results in Section IV-C.

## III. OUR PROPOSED METHOD

One of the well-known robust functions against outliers and large residuals is the robust least squares or Huber function [13]:

$$\phi_{hub}(u) \triangleq \begin{cases} u^2 & |u| \leq M \\ M(2|u| - M) & |u| > M, \end{cases} \tag{13}$$

where $M$ is a breaking point.

Therefore, we propose using the robust Huber function in estimating an ellipse parameters vector. In other words, we propose the following cost function to estimate the vector $\mathbf{a}$:

$$\min_{\mathbf{a}} \sum_{i=1}^{I} \phi_{hub}(r_i), \tag{14}$$

where $r_i \triangleq \mathbf{x}_i^T\mathbf{a} - 1$. Problem (14) is convex, and iterative re-weighted methods can be used to solve it [13]. In addition, [9] has shown that utilizing the Huber function in a way that is used in (14) is equivalent to adding variables for modeling outliers. Thus, (14) is equivalent to the following minimization:

$$\min_{\mathbf{a},\mathbf{o}} \sum_{i=1}^{I} \frac{1}{2}(\mathbf{x}_i^T\mathbf{a} - 1 - o_i)^2 + \lambda|o_i|, \tag{15}$$

where $o_i$ is a variable to model outliers, and $\lambda$ is a parameter to tune the sparseness of $\mathbf{o} \triangleq [o_1, o_2, \ldots, o_I]^T$. Note that the coefficient $\frac{1}{2}$ was added to simplify subsequent equations.

In matrix form, (15) can be re-written as

$$\min_{\mathbf{a},\mathbf{o}} \frac{1}{2}\|\mathbf{X}\mathbf{a} - \mathbf{1}_{I\times 1} - \mathbf{o}\|_2^2 + \lambda\|\mathbf{o}\|_1, \qquad (16)$$

where $\mathbf{X} \triangleq [\mathbf{x}_1\mathbf{x}_2\ldots\mathbf{x}_I]^T$. Since $\mathbf{o}$ should be a sparse vector, (16) is based on both the sparsity of outliers and the least squares of the residuals. Problem (16) is convex, and can be solved using several effective numerical methods [13]. For example, CVX, a package for specifying and solving convex programs [14], [15], can solve (16). However, since solving (16) with the CVX package is relatively slow (as will be seen in our numerical simulations of SectionIV-B), we propose to use an alternating minimization approach (also known as block coordinate descent[1]) for solving it [17].

---

**Algorithm 1** Proposed Method

---

**Input:** 2D data point set, $\lambda$

**Output:** Ellipse parameters vector, vector $\mathbf{o}$

1: **procedure**
2:     Form data vector as follows: $\mathbf{x}_i = [x_i^2, x_i y_i, y_i^2, x_i, y_i]^T$
3:     Construct the matrix $\mathbf{X}$ like the following form: $\mathbf{X} = [\mathbf{x}_1\mathbf{x}_2\ldots\mathbf{x}_I]^T$ ($I$ is the total number of points)
4:     Initialize $\mathbf{o}$ vector with *e.g.* the zero vector
5:     **repeat**
6:         $\mathbf{a} = \mathbf{X}^{\dagger}(\mathbf{o} + \mathbf{1}_{I\times 1})$
7:         $o_i = \mathbf{x}_i^T\mathbf{a} - 1 - \lambda\,\text{sign}(o_i), \mathbf{o} = [o_1,\ldots,o_I]$
8:     **until** convergence
9: **end procedure**

---

## IV. Experimental Results

### A. Generating test samples

In order to form this kind of data, an ellipse with known characteristics should be considered. After producing data points on the ellipse, a white Gaussian noise with zero mean and variance $\sigma^2$ is added to them. Then, in order to create outliers, some of the data points are selected randomly, and a white Gaussian noise, which has zero mean and a variance highly larger than $\sigma^2$, is added to them. Fig. 1 shows an example of the first set of test samples. It contains $I = 40$ data points of an ellipse with the ellipse characteristics vector $[2, 12, 5, 2, 60°]$ which corresponds to an ellipse with $r_x = 2$, $r_y = 12$, $(c_x, c_y) = (5, 2)$, $\theta = 60°$, corrupted by a Gaussian noise with zero mean and variance 0.05. Finally, eight points of data are selected randomly, and are converted into outliers by adding a Gaussian noise with zero mean and variance 20 to them.

### B. Experiment I: Running time

As a rough measure of the complexities of the algorithms, their running times are compared in this subsection. In this experiment, 28 data points including 20 inliers and 8 outliers,

---

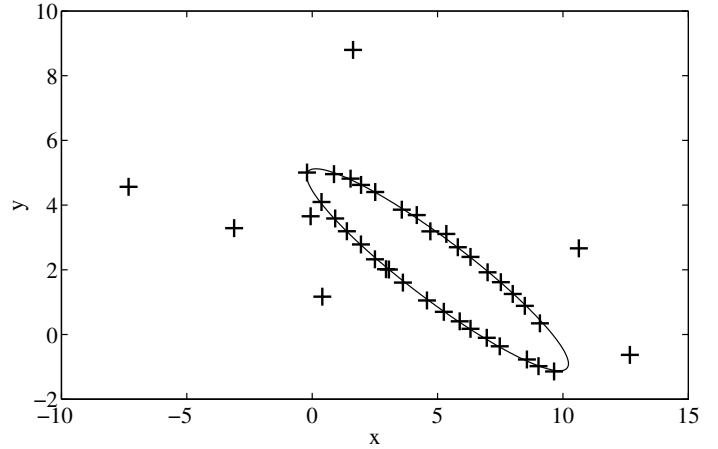[1]The convergence of block coordinate descent approaches has been discussed in [16].



Fig. 1. An example of the first set of test samples. It has 40 data points corrupted by an additive Gaussian noise with zero mean and variance 0.05. Eight of these data points are randomly selected, and are converted to outliers by adding a Gaussian noise with zero mean and variance 20.

TABLE I
COMPARISON OF THE AVERAGE OF COMPUTATION TIMES

| Method | Our method with alternating minimization | Our method with the CVX package | Liang's method |
|---|---|---|---|
| Average of running time (sec) | 0.0223 | 0.9029 | 0.3513 |

of a known ellipse are considered. Then, a white Gaussian noise with zero mean is added to them. The variance of the noise is gradually increased from 0.0005 to 0.03 during 30 trials of this experiment, and the average of computation times is given in Table I.

This experiment is done on a computer with an Intel(R) Core(TM) i5 CPU, 4GB RAM, Win7, and MATLAB 2013b. Table I compares the average of running times of the two approaches (alternating minimization and the CVX package) for solving (16). Since alternating minimization solves (16) much more faster than the CVX package, we choose this approach for solving our proposed method in the rest of the experiments. In addition, Table I compares the speed of our method and Liang's method[2]. As can be seen in Table I, computing times of our method is much less than that of Liang's method. Since Liang's method is faster than ellipse fitting based on Hough Transform and Bayesian algorithm [5], our method would be faster than those, as well.

### C. Experiment II

As mentioned in Section II-B, Liang's method in (7) implies that an ellipse parameters vector is written as a linear combination of $\mathbf{x}_i$ vectors [5]. In the current subsection, we suggest the following experiment to show that Liang's method does not inherently utilize the set of the data points as a sparsifying dictionary in (7), and actually almost any dictionary $\mathbf{D}$ can be used in (7) instead of $\mathbf{X}^T$.

---

[2]The authors would like to thank Mr. J. Liang for sending us the MATLAB implementation of his method.

TABLE II
COMPARISON OF ESTIMATED ELLIPSE PARAMETERS USING (10) WITH
DEFINITIONS IN (8) AND (12)

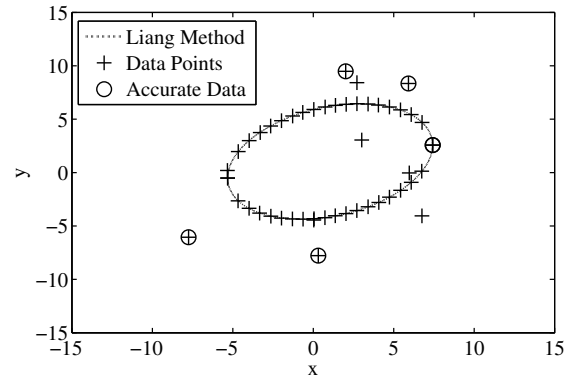| Parameters | Exact values | with $\overline{\mathbf{X}}$ as defined in (8) | with $\overline{\mathbf{X}}$ as defined in (12) |
|---|---|---|---|
| a | −0.0157 | −0.0155 | −0.0154 |
| b | −0.0489 | −0.0486 | −0.0486 |
| c | −0.0440 | −0.0437 | −0.0438 |
| d | 0.2551 | 0.2530 | 0.2527 |
| e | 0.4203 | 0.4199 | 0.4201 |



Fig. 3. Data points marked with circles are recognized as accurate points by Liang's method. Although Liang's algorithm call these points accurate, they are mostly outliers.



Fig. 2. The results of solving (10), once with $\overline{\mathbf{X}}$ as defined in (8) corresponding to Liang's method; and once with $\overline{\mathbf{X}}$ as defined in (12) in which $\mathbf{D}$ is a random matrix.
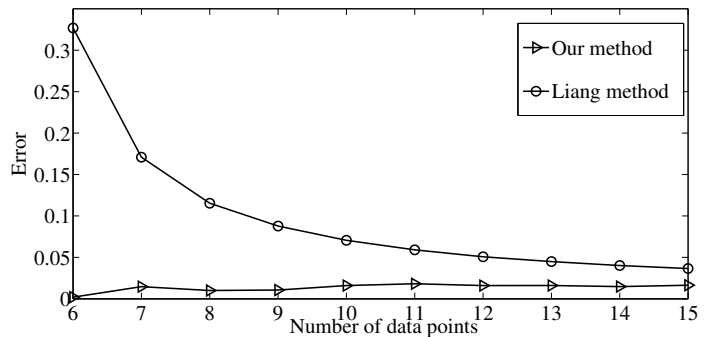


Fig. 4. The error of the estimation of the ellipse parameters vector versus the number of data points. Data points are contaminated by an additive Gaussian noise with zero mean and variance 0.005. Approximately 20% of data points are outliers.

In other words, given a particular set of data points, we estimate the best fitted ellipse parameters using (10), once with $\overline{\mathbf{X}}$ as defined in (8) which corresponds to Liang's method; and once with $\overline{\mathbf{X}}$ as defined in (12) in which the entries of $\mathbf{D}$ are drawn from a random variable uniformly distributed over $(0, 1)$.

The final results are given in Table II and Fig. 2. It is seen that the fitted ellipses using these two approaches are almost the same, which emphasizes that (7) does not inherently use the "data" matrix $\mathbf{X}^T$ as the sparsifying dictionary.

As stated before (7), another justification behind (7) is that the $\mathbf{a}$ vector is stated as a linear combination of "more accurate" data points. To verify this justification, in another ellipse fitting simulation using Liang's method which is shown in Fig. 3, the data points that correspond to large values of the $\mathbf{b}$ vector are marked by circles. It is seen that these data points are not "more accurate", and in fact, they are mostly outliers in this simulation.

### D. Experiment III

To reach a desired accuracy, if a fitting algorithm requires less number of data points compared to other methods, that algorithm is considered stronger and more efficient. In the current subsection, we compare other methods with ours from this perspective. To this end, the ellipse parameters vector is obtained for each number of data points, and the norm of the error vector (*i.e.* the difference between the estimated and the original vectors) is computed.

Fig. 4 shows the norm of the error vector versus the number of data points for Liang's method and ours. As it is seen in this figure, the estimation error of Liang's method increases when the number of data points decreases. In contrary, our algorithm almost has a constant performance versus the number of data points. Therefore, Fig. 4 emphasizes on the strength of our method as compared to Liang's method when the number of data points is small.

### E. Experiment IV

In this subsection, the performance of ellipse fitting algorithms for long ellipses is evaluated. For this experiment, long ellipses are prepared such that the ratio of semi-major axis to semi-minor one is in the range of 6 to 60. As before, the ellipse parameters vector is obtained, and the norm of the error vector is computed.

Fig. 5 shows the result for this experiment. As can be seen, the estimation error of Liang's method is larger when the ratio of semi-axes is high (corresponding to longer ellipses). In fact, as the high ratio of semi-axes is equivalent to a large condition number of matrix $\mathbf{X}^T$, Liang's method is sensitive to the condition number of $\mathbf{X}^T$, which is composed of data
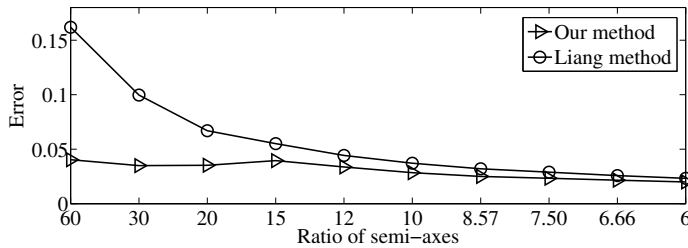
Fig. 5. The error of the estimation of the ellipse parameters vector versus the ratio of semi-axes. Data points are contaminated by an additive Gaussian noise with zero mean and variance 0.005. Approximately 20% of data points are outliers.
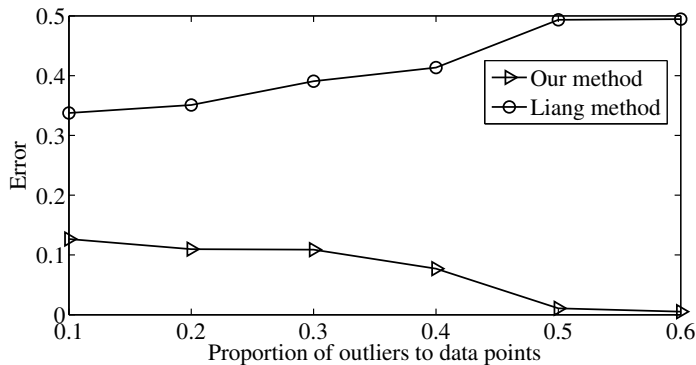


Fig. 6. The error of the estimation of the ellipse parameters vector versus the proportion of outliers to data points. The proportion of outliers to data points varies from 10% to 60%.

points. On the other hand, our method is almost not sensitive to the ratio of the semi-axes.

*F. Experiment V*

This experiment is to compare the performance of the algorithms according to the amount of outliers. There are 10 data points contaminated by an additive Gaussian noise with zero mean and variance 0.005 and considered as inliers. In addition, there are some other data points corrupted by an additive Gaussian noise with zero mean and variance 20 to be considered as outliers. During the experiment, the number of outliers varies from 1 to 6. Fig. 6 shows the result of this experiment. It is seen that when the number of outliers increases, the estimation error of Liang's method increases. Indeed, when the number of inliers is small, Liang's method is more affected by the outliers than ours.

## V. CONCLUSION

In this paper, we proposed a new method of ellipse fitting, which unlike most of the previous methods such as DLS [3], is robust against outliers. In addition, we discussed that the underlying justification of Liang's method [5] as a state-of-the-art method, is not convincing enough.

Experimental results in [5] had approved the strength of Liang's method in terms of running time and robustness to outliers as compared to previous methods of ellipse fitting. Hence, considering our experimental results which showed superiority of our method over Liang's method, we can assert

that our method would have better performance in the sense of sensitiveness to outliers and complexity comparing to previous methods such as DLS, ellipse fitting based on Hough Transform, and Bayesian algorithm.

## REFERENCES

[1] V. F. Leavers, *Shape detection in computer vision using the hough transform*. New York, USA: Springer-Verlag, 1992.

[2] D. S. Barwick, "Very fast best-fit circular and elliptical boundaries by chord data," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 31, no. 6, pp. 1147–1152, June 2009.

[3] A. Fitzgibbon, M. Pilu, and R. B. Fisher, "Direct least square fitting of ellipses," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 21, no. 5, pp. 477–480, May 1999.

[4] H. Qin, X. Wang, M. Liang, and W. Yan, "A novel pupil detection algorithm for infrared eye image," in *Signal Processing, Communication and Computing (ICSPCC), 2013 IEEE International Conference on.* IEEE, 2013, pp. 1–5.

[5] J. Liang, M. Zhang, D. Liu, X. Zeng, O. O. Jr, K. Zhao, Z. Li, and H. Liu, "Robust ellipse fitting based on sparse combination of data points," *IEEE Trans. Image Process.*, vol. 22, no. 6, pp. 2207–2218, June 2013.

[6] P. Sturm and P. Gargallo, "Conic fitting using the geometric distance," in *Asian Conference on Computer Vision.* Tokyo, Japan: Springer, November 2007, pp. 784–795.

[7] M. Elad, *Sparse and Redundant Representations: From Theory to Applications in Signal and Image Processing*, 1st ed. Springer Publishing Company, Incorporated, 2010.

[8] P. J. Rousseeuw and A. M. Leroy, *Robust regression and outlier detection*. John Wiley & Sons, 2005, vol. 589.

[9] J.-J. Fuchs, "An inverse problem approach to robust regression," in *Acoustics, Speech, and Signal Processing, 1999. Proceedings., 1999 IEEE International Conference on*, vol. 4. IEEE, 1999, pp. 1809–1812.

[10] F. L. Bookstein, "Fitting conic sections to scattered data," *Computer Graphics and Image Processing*, vol. 9, pp. 56–71, 1979.

[11] P. L. Rosin, "A note on the least squares fitting of ellipses," *Pattern Recognition Letters*, vol. 14, no. 10, pp. 799–808, 1993.

[12] W. Gander, G. H. Golub, and R. Strebel, "Least-squares fitting of circles and ellipses," *BIT Numerical Mathematics*, vol. 34, no. 4, pp. 558–578, 1994.

[13] S. Boyd and L. Vandenberghe, *Convex Optimization.* Cambridge university press, 2004.

[14] M. Grant and S. Boyd, "CVX: Matlab software for disciplined convex programming, version 2.1," http://cvxr.com/cvx, mar 2014.

[15] ——, "Graph implementations for nonsmooth convex programs," in *Recent Advances in Learning and Control*, V. Blondel, S. Boyd, and H. Kimura, Eds. Springer-Verlag Limited, 2008, pp. 95–110, http://stanford.edu/~boyd/graph_dcp.html.

[16] A. Beck and L. Tetruashvili, "On the convergence of block coordinate descent type methods," *SIAM Journal on Optimization*, vol. 23, no. 4, pp. 2037–2060, 2013.

[17] C. L. Byrne, "Alternating minimization as sequential unconstrained minimization: a survey," *Journal of Optimization Theory and Applications*, vol. 156, no. 3, pp. 554–566, 2013.