

# Two-View Orthographic Epipolar Geometry: Minimal and Optimal Solvers

Magnus Oskarsson<sup>1</sup>

Received: 27 February 2017 / Accepted: 12 July 2017  
© The Author(s) 2017. This article is an open access publication

**Abstract** We will in this paper present methods and algorithms for estimating two-view geometry based on an orthographic camera model. We use a previously neglected nonlinear criterion on rigidity to estimate the calibrated essential matrix. We give efficient algorithms for estimating it minimally (using only three point correspondences), in a least squares sense (using four or more point correspondences), and optimally with respect to the number of inliers. The inlier-optimal algorithm is based on a three-point solver and gives a fourth-order polynomial time algorithm. These methods can be used as building blocks to robustly find inlier correspondences in the presence of high degrees of outliers. We show experimentally that our methods can be used in many instances, where the orthographic camera model isn't generally used. A case of special interest is situations with repetitive structures, which give high amounts of outliers in the initial feature point matching.

## 1 Introduction

The geometry underlying two views of a rigid scene has been studied for a very long time. Using different projection models, the *structure from motion problem*, i.e., the problem of estimating both the scene geometry and the camera geometry from only image data, has been solved for many cases [8]. In his seminal paper from 1979, Ullman used the orthographic projection model to formulate and solve a number of structure from motion problems, for a small number of views and points [27]. During the following years, these theories were developed and refined. The main concern was

to develop methods and algorithms to accurately estimate the geometry, in the presence of image noise [7, 9, 10, 14]. During recent years, more focus has been given to robustly estimating geometry in the presence of outliers in the data. To enable this, a number of approaches have been followed. A classic way to handle outliers is through robust estimation schemes based on RANSAC. In these frameworks, one needs methods for estimating model parameters given a small or minimal number of data points. For calibrated two-view projective geometry, the five-point algorithm estimates the essential matrix minimally, using five image point correspondences. In another direction, algorithms for global inlier maximization have been developed. These are often based on relaxing an initial non-convex problem into a more tractable problem, using different error norms or branch-and-bound methods [3, 12, 13, 16]. Arguably, in many cases it is more beneficial to instead relax the underlying camera models (e.g., by using an orthographic camera model), to get tractable problems. Using approximate models to estimate epipolar geometry has been investigated previously; in [6], Goshen and Shimshoni used two matching SIFT descriptors, with dominant directions and scale, to construct eight correspondences. Perdoch et al. [20] used a fixed partial calibration to estimate the fundamental matrix from two local affine frames. In [21], it was shown that the affine fundamental matrix is the first-order Taylor approximation of the full perspective fundamental matrix. Using two matching MSER feature points, an approximate fundamental matrix could be estimated. Lately, methods that optimally find models that maximize the number of inliers, in polynomial time, have been developed, [5, 25]. We will in this paper investigate a number of specific geometric problems that have received little or no attention earlier, related to orthographic projections of rigid scenes in two views. We believe that this work gives some additional theoretic insights and at the same time also

✉ Magnus Oskarsson  
magnuso@maths.lth.se

<sup>1</sup> Lund University, PO Box 118, 221 00 Lund, Sweden

gives new powerful tools to robustly establish image point correspondences in the presence of outliers.

Our main contribution is three algorithms for estimating orthographic epipolar geometry for two views:

- A *minimal three-point solver* that uses only three points and yields only two solutions, making it very suitable for RANSAC-based estimation.
- An *optimal solver, maximizing the number of inliers* using recent methods for optimal estimation and based on three specialized minimal solvers.
- A *least squares solver* that minimizes a relevant reprojection error, by finding all stationary points of the corresponding Lagrangian.

## 2 Essential Matrix Estimation

In [18] the general form of the fundamental matrix for two affine cameras was given,

$$F = \begin{bmatrix} 0 & 0 & a \\ 0 & 0 & b \\ c & d & e \end{bmatrix}. \quad (1)$$

This matrix has five degrees of freedom, but it is only defined up to scale. It can also be described using the directions and offsets of the corresponding epipolar lines [17]. It can be linearly estimated from at least four point correspondences [22] or two ellipses [1].

The corresponding calibrated entity—the essential matrix corresponding to two orthographic projections—will fulfill an additional nonlinear constraint. This fact has been mentioned in passing earlier [9, 15], but the implications have not been pursued in any depth. The extra constraint means that the essential matrix in this case only has three degrees of freedom. This low dimension will give us powerful tools for estimating the epipolar geometry. We will state the constraint in terms of the essential matrix.

**Theorem 1** *The essential matrix,*

$$E = \begin{bmatrix} 0 & 0 & a \\ 0 & 0 & b \\ c & d & e \end{bmatrix}, \quad (2)$$

*corresponding to two orthographic views, will fulfill*

$$a^2 + b^2 = c^2 + d^2. \quad (3)$$

*Proof* Two orthographic views can be represented by the camera matrices

$$P_1 = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad P_2 = \begin{bmatrix} r_{11} & r_{12} & r_{13} & t_1 \\ r_{21} & r_{22} & r_{23} & t_2 \\ 0 & 0 & 0 & 1 \end{bmatrix}, \quad (4)$$

where  $\mathbf{r}_1 = [r_{11}, r_{12}, r_{13}]$  and  $\mathbf{r}_2 = [r_{21}, r_{22}, r_{23}]$  are orthonormal vectors. The corresponding essential matrix is then given by

$$E = \begin{bmatrix} 0 & 0 & a \\ 0 & 0 & b \\ c & d & e \end{bmatrix} = \begin{bmatrix} 0 & 0 & r_{13}r_{21} - r_{11}r_{23} \\ 0 & 0 & r_{13}r_{22} - r_{12}r_{23} \\ r_{23} & -r_{13} & r_{13}t_2 - r_{23}t_1 \end{bmatrix}. \quad (5)$$

Let

$$\begin{aligned} \mathbf{r}_3 &= [r_{31}, r_{32}, r_{33}] = \mathbf{r}_1 \times \mathbf{r}_2 \\ &= [r_{12}r_{23} - r_{13}r_{22}, r_{13}r_{21} - r_{11}r_{23}, r_{11}r_{22} - r_{12}r_{21}]. \end{aligned} \quad (6)$$

(7)

From (5) and the orthonormality constraints, it is clear that

$$c^2 + d^2 = r_{23}^2 + r_{13}^2 = 1 - r_{33}^2. \quad (8)$$

From (5) and (7), we have that  $a^2 + b^2 = r_{32}^2 + r_{31}^2$ . The orthonormality constraint now gives that

$$a^2 + b^2 = r_{32}^2 + r_{31}^2 = 1 - r_{33}^2, \quad (9)$$

and hence  $a^2 + b^2 = c^2 + d^2$ .  $\square$

For two corresponding points  $\mathbf{u} = [x, y, 1]$  and  $\mathbf{u}' = [x', y', 1]$ , we have that  $\mathbf{u}^T E \mathbf{u}' = 0 \Leftrightarrow e + ax + by + cx' + dy' = 0$ . The essential matrix is only determined up to scale. To get appropriate and symmetric epipolar line constraints, we will fix the scale by setting  $a^2 + b^2 = 1 (= c^2 + d^2)$ . This will lead to that  $(a, b)$  and  $(c, d)$  represent the unit norm normal directions for the two epipolar lines. The epipolar line distances to the origin are then given by  $e + cx' + dy'$ , respectively,  $e + ax + by$ . This gives us a natural way to define a symmetric and geometrically valid error. The mean perpendicular distance to the epipolar lines in the two images is given by

$$D = \frac{1}{2}(|ax + by + e + cx' + dy'| \quad (10)$$

$$+ |cx' + dy' + e + ax + by|) \quad (11)$$

$$= |e + ax + by + cx' + dy'|. \quad (12)$$

In [22], three different cost functions, for affine geometry estimation, were discussed. Our choice of normalization makes these three different cost functions collapse into the

one described above. For a thorough discussion on model selection functions for two-view geometry see [26].

It is known that given two orthographic views of a rigid scene, the full camera geometry cannot be established [9, 10]. But the image data must fulfill the epipolar constraint in order for the geometry to be valid. We will use this rigidity constraint to enable estimates of the geometry, based on a low number of tentative point correspondences. These estimates in turn can be used to find image feature point correspondences in robust manner.

We will in the next sections show how the essential matrix can be estimated. In Sect. 3, we solve the minimal case of three point correspondences. This gives a very fast solver that can be used in a standard RANSAC framework. In Sect. 4, we derive solvers that can be used to find the globally optimal essential matrix that maximizes the number of inliers, given a set of point correspondences corrupted by outliers. Finally in Sect. 5, we show how to find the nonlinear least squares solution, given four or more point correspondences.

### 3 A Minimal Three-Point Solver

The essential matrix has three degrees of freedom, and each point correspondence gives one constraint on the parameters. We should hence be able to minimally estimate the essential matrix using three point correspondences. We will now show how this can be done. We have according to (12) the three constraints  $D_i = 0$ ,  $i = 1, 2, 3$ . These are linear in the five parameters  $(a, b, c, d, e)$ . In addition to this, we have the two quadratic constraints  $a^2 + b^2 = 1$  and  $c^2 + d^2 = 1$ . We can start by eliminating  $e$  by taking  $D_2 - D_1$  and  $D_3 - D_1$ . This gives two linear constraints in  $(a, b, c, d)$ . We can use these to express  $(a, b)$  in terms of  $(c, d)$ . Substituting these expressions into  $a^2 + b^2 = 1$  gives a polynomial in  $c$  and  $d$ . We can use the second quadratic constraint  $c^2 = 1 - d^2$  to eliminate all factors of  $c$  of higher degree than one. This gives a polynomial of the form  $p_1 = k_1cd + k_2d^2 + k_3$ , where  $k_i$  only depends on image point measurements. Multiplying this polynomial by  $c$  and again eliminating terms of higher degree in  $c$  gives a polynomial of the form  $p_2 = q_1cd^2 + q_2c + q_3d^3 + q_4d$ . We can write these two polynomial constraints as

$$B \begin{bmatrix} c \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \quad (13)$$

with

$$B = \begin{bmatrix} k_1 & k_2d^2 + k_3 \\ q_1d^2 + q_2 & q_3d^3 + q_4d \end{bmatrix}. \quad (14)$$

In order for (13) to have a solution, the determinant of  $B$  must vanish. Taking the determinant of  $B$  gives a fourth degree

polynomial in  $d$  but with only even degrees. It is clear that if we have a solution  $(a, b, c, d, e)$  to our three-point problem then  $(-a, -b, -c, -d, -e)$  will also be a solution, but this corresponds to the same essential matrix. This means that we only need to consider two of the solutions for  $d$ . The solver will be extremely fast since we only need to solve a second-degree polynomial in the end.

### 4 Maximizing the Number of Inliers

In the presence of outliers, finding accurate correspondences is difficult, and robust methods are highly desirable. A common approach is to estimate a model that optimizes the number of inliers. Measuring reprojection errors is normally the preferred choice, as this accurately models the limited precision of feature detection techniques. Although such formulations lead to challenging optimization problems, using recent advances in robust estimation it is sometimes possible to develop tractable methods. In [5], it was shown how the number of inliers can be maximized in polynomial time, for a fixed-dimensional model, where the computational complexity follows directly as a consequence of the theory of optimization. One requirement is that the parameter space is a differentiable manifold embedded in  $\mathbb{R}^m$  with a set of equality constraints. The authors used this to produce algorithms for optimal image stitching and 2D-registration. In [24, 25], the authors used similar methods to perform large-scale image-based localization. We will here describe how these ideas can be applied to orthographic essential matrix estimation, resulting in an optimal method.

The main theorem from [5] shows that one can find the optimal solution with respect to the number of inliers by enumerating a finite set of so called *critical points*, essentially being the Karush–Kuhn–Tucker (KKT) points. These critical points divide the solution space into regions that contain different combinations of inliers and outliers, and the optimal solution with respect to the number of inliers will be found in one of the critical points. Our parameter space  $(a, b, c, d, e)$  is embedded in  $\mathbb{R}^5$  with the constraints  $h_1 = a^2 + b^2 = 1$  and  $h_2 = c^2 + d^2 = 1$ . The critical points satisfy the KKT-conditions for local optimality to the optimization problems

$$\min_{(a,b,c,d,e)} f(a, b, c, d, e) \quad (15)$$

$$h_j(a, b, c, d) = 1, \quad j = 1, 2, \quad (16)$$

$$D_i(a, b, c, d, e) \leq \epsilon \text{ for all } i \in C, \quad (17)$$

where  $C$  runs over all subsets of correspondences of size  $|C| \leq 3$  (the number of degrees of freedom of our problem) and  $D_i$  are the epipolar distances (12). The function  $f$  is an auxiliary goal function, which can be chosen arbitrarily, as long as the KKT points constitute a finite set. Most often a

linear function will yield simple equations, and we will show that choosing the goal function

$$f(a, b, c, d, e) = a + d, \quad (18)$$

will give us a finite set of KKT points. For an inlier bound of  $\epsilon$ , the KKT points are then found by looking at points where a number of constraints using (17) are active, i.e.,

1.  $D_i^2 = \epsilon^2, i = i_1, i_2, i_3$ .
2.  $D_i^2 = \epsilon^2, i = i_1, i_2$  and  $(\nabla f, \nabla h_j, \nabla D_i^2)$  are linearly dependent.
3.  $D_i^2 = \epsilon^2, i = i_1$  and  $(\nabla f, \nabla h_j, \nabla D_i^2)$  are linearly dependent.
4.  $\nabla f = 0$ .

In order to find the solution, we need solvers to the first three different cases above (the last case gives only a trivial solution), and we can then find the optimal solutions by evaluating all KKT points. The time complexity of the algorithm is determined by step 1, where three constraints are active. Going through all triplets of points and then for each solution checking how many inliers we have gives a total complexity of  $\mathcal{O}(n^4)$  for  $n$  tentative correspondences. In this way, it is possible to maximize the number of inliers in  $\mathcal{O}(n^4)$  time. The steps of the method are summarized in Algorithm 1. In Sect. 4.1, we show how we construct the main optimal three-point solver. The optimal two- and one-point solvers are discussed in Sect. 4.2.

---

**Algorithm 1** Optimal inlier maximization
 

---

- 1: Given a set of corresponding points  $\mathcal{C}$  in two images, and an inlier bound  $\epsilon$ .
  - 2: For all possible combinations  $C3_k \subset \mathcal{C}$  with  $|C3_k| = 3$ , find the corresponding essential matrices  $E3_k$  as described in Section 4.1 and count the total number of corresponding points  $N3_k$  with error less than  $\epsilon$ .
  - 3: For all possible combinations  $C2_k \subset \mathcal{C}$  with  $|C2_k| = 2$ , find the corresponding essential matrices  $E2_k$  as described in Section 4.2 and count the total number of corresponding points  $N2_k$  with error less than  $\epsilon$ .
  - 4: For all  $C1_k \subset \mathcal{C}$  correspondences, find the corresponding essential matrices  $E1_k$  as described in Section 4.2 and count the total number of corresponding points  $N1_k$  with error less than  $\epsilon$ .
  - 5: The maximal number of inliers is the maximum of all  $N3_k, N2_k$  and  $N1_k$  for all  $k$ . The globally optimal essential matrix is the corresponding  $E3_k, E2_k$  or  $E1_k$  for the optimal  $k$ .
- 

#### 4.1 The Three-Point Solver for Inlier Maximization

Given an inlier bound  $\epsilon$  and three point correspondences, we want to find an essential matrix  $E$  such that

$$D_i^2 = \epsilon^2, \quad i = 1, 2, 3. \quad (19)$$

These equations are quadratic in the unknowns  $(a, b, c, d, e)$ , but we can simplify them by considering that the solution must fulfill

$$D_i = \pm\epsilon, \quad i = 1, 2, 3. \quad (20)$$

This gives in total eight combinations of solutions. We linearly eliminate  $e$ , giving for each combination a system on the form

$$D_i - D_1 = w_i, \quad i = 2, 3, \quad (21)$$

where  $w_i$  can take the values of  $-2\epsilon, 0$  or  $2\epsilon$ . This gives a very similar system to the one we solved in Sect. 3, but slightly more complicated due to the constant factors. We will solve it by explicitly constructing a Gröbner basis [4]. We start in the same way and write  $(a, b)$  in  $(c, d)$  using the linear constraints. Resubstitution into  $a^2 + b^2 = 1$  gives an equation on the form

$$k'_1 c^2 + k'_2 cd + k'_3 c + k'_4 d^2 + k'_5 d + k'_6 = 0, \quad (22)$$

where  $k'_i$  only depends on image measurements. We can eliminate the  $c^2$ -term in favor for  $d^2$  using the constraint  $c^2 + d^2 = 1$ . We then get

$$p_1 = cd + k_1 + k_2 c + k_3 d + k_4 d^2, \quad (23)$$

where each  $k_i$  depends on  $k'$ . Multiplying this equation with  $c$  and  $d$  gives two new equations. Again we eliminate all  $c^2$ -terms. Taking a linear combination of these two polynomials gives us a polynomial with monomials  $\{cd, c, d^3, d^2, d, 1\}$ . We can eliminate the  $cd$ -term by solving for  $cd$  in  $p_1 = 0$  and resubstituting. This gives us

$$p_2 = d^3 + q_1 + q_2 c + q_3 d + q_4 d^2, \quad (24)$$

where again  $q_i$  only depends on the image coordinates. The polynomials  $p_1$  and  $p_2$  together with  $p_0 = c^2 + d^2 - 1$  constitute a Gröbner basis for our problem. We can now construct our solver using the action matrix method [4]. We use the linear basis  $\{1, c, d, d^2\}$  and  $d$  as action variable to construct the action matrix  $A$ ,

$$d \begin{bmatrix} 1 \\ c \\ d \\ d^2 \end{bmatrix} = \begin{bmatrix} d \\ cd \\ d^2 \\ d^3 \end{bmatrix} = A \begin{bmatrix} 1 \\ c \\ d \\ d^2 \end{bmatrix}, \quad (25)$$

with

$$A = \begin{bmatrix} 0 & 0 & 1 & 0 \\ -k_1 & -k_2 & -k_3 & -k_4 \\ 0 & 0 & 0 & 1 \\ -q_1 & -q_2 & -q_3 & -q_4 \end{bmatrix}. \quad (26)$$

We can now solve for  $c$  and  $d$  by finding the eigenvectors of  $A$ , giving four solutions. In total, we get  $8 \cdot 4 = 32$  solutions to the three-point problem.

## 4.2 The Two- and One-Point Solvers for Inlier Maximization

We will here give an outline to our two and one point solvers, i.e., the solutions to  $D_i^2 = \epsilon^2, i = i_1, i_2$ , and  $(\nabla f, \nabla h_j, \nabla D_i^2)$  are linearly dependent, respectively,  $D_i^2 = \epsilon^2, i = i_1$  and  $(\nabla f, \nabla h_j, \nabla D_i^2)$  are linearly dependent. To get as simple expressions as possible, we have chosen the goal function  $f = a + c$ . That the gradients should be linearly dependent can be expressed using the determinant of the corresponding stacked gradients, yielding for the two-point case an expression on the form

$$s_1 ad + s_2 bc + s_3 bd = 0, \quad (27)$$

where each  $s_i$  only depends on image measurements. Taking the difference of the two epipolar line distance equations eliminates  $e$  and gives a linear constraint on  $(a, b, c, d)$ . This gives together with the two embedding criteria  $a^2 + b^2 = c^2 + d^2 = 1$  a system of four polynomial equations in the four variables  $(a, b, c, d)$ . Using similar techniques as previously, this system can be solved, yielding at most eight real solutions. Since we, for each of the two points, have a distance of  $\pm\epsilon$  we get in total  $4 \cdot 8 = 32$  solutions.

For the one point case, this point will be used to express  $e$  in terms of the unknowns. The linearly dependence of the gradients will in this case give the constraints  $bd = 0, ad = 0, bc = 0$  and  $s'_1 ad + s'_2 bc + s_3 bd = 0$ . This together with the embedding constraints only gives the solutions  $(a, b, c, d) = (1, 0, 1, 0)$  and  $(a, b, c, d) = (-1, 0, 1, 0)$ . In total, we will get four solutions since we get two solutions for  $e$  corresponding to  $\pm\epsilon$ .

## 5 A Least Squares Solver

Having more than three point correspondences will lead to an overdetermined system of equations if we want to solve  $D_i = 0$ . We will solve this in a least squares sense, i.e., given  $n$  point correspondences,

$$\min_E \sum_{i=1}^n D_i^2, \quad s.t. \quad a^2 + b^2 = 1, \quad c^2 + d^2 = 1. \quad (28)$$

Actually in the very first ECCV, Harris described the same problem [7], and it was noted that the solution can be found as one of the roots of an 8-degree polynomial. However, no details as to how to construct the final polynomial were given. Here we go through the solution in detail and construct a

solver that is both robust to noise and efficient. We will solve the least squares problem by finding all the stationary points of (28) by differentiating the corresponding Lagrangian function,

$$L(a, b, c, d, e, \lambda, \gamma) = \sum_{i=1}^n D_i^2 + \lambda(1 - a^2 - b^2) + \gamma(1 - c^2 - d^2), \quad (29)$$

with

$$\nabla L = 2 \begin{bmatrix} \sum_{i=1}^n x_i D_i - a\lambda \\ \sum_{i=1}^n y_i D_i - b\lambda \\ \sum_{i=1}^n x'_i D_i - c\gamma \\ \sum_{i=1}^n y'_i D_i - d\gamma \\ \sum_{i=1}^n D_i \\ 1 - a^2 - b^2 \\ 1 - c^2 - d^2 \end{bmatrix}. \quad (30)$$

From  $\nabla L = 0$  it is clear that

$$e = \frac{1}{n} \sum_{i=1}^n [x_i \ y_i \ x'_i \ y'_i] \begin{bmatrix} a \\ b \\ c \\ d \end{bmatrix}. \quad (31)$$

Thus  $e$  can be eliminated by removing the centroid of each of the image point sets, i.e.,  $\tilde{\mathbf{u}}_i = \mathbf{u}_i - \bar{\mathbf{u}}$  and  $\tilde{\mathbf{u}}'_i = \mathbf{u}'_i - \bar{\mathbf{u}}'$ . The vanishing of the gradient of  $L$  (without the normalization constraints) can then be written

$$Mv = Sv, \quad (32)$$

with  $v = [a, b, c, d]^T$  and

$$M_{4 \times 4} = \begin{bmatrix} \tilde{\mathbf{u}} \\ \tilde{\mathbf{u}}' \end{bmatrix} [\tilde{\mathbf{u}}^T \ \tilde{\mathbf{u}}'^T], \quad S = \begin{bmatrix} \lambda & 0 & 0 & 0 \\ 0 & \lambda & 0 & 0 \\ 0 & 0 & \gamma & 0 \\ 0 & 0 & 0 & \gamma \end{bmatrix}. \quad (33)$$

This is almost an eigenvalue problem, but the different  $\lambda$  and  $\gamma$  make it slightly more complicated to solve. One can see that (32) is homogeneous in  $(a, b, c, d)$  so we can scale the equations with  $1/d$  giving equations in  $(a/d, b/d, c/d, 1) = (\hat{a}, \hat{b}, \hat{c}, 1)$ . We can then use three of the four equations to linearly express  $(\hat{a}, \hat{b}, \hat{c})$  in terms of  $\lambda$  and  $\gamma$ . We reinsert these expressions into the fourth equation of (32) and the normalization constraints  $\hat{a}^2 + \hat{b}^2 = \hat{c}^2 + 1$ . This gives two equations of degree four in  $\lambda$  and degree two in  $\gamma$ . Multiplying these two equations with  $\gamma$  gives two new equations. We can now write these equations as



$$B_{4 \times 4}(\lambda) \begin{bmatrix} \gamma^3 \\ \gamma^2 \\ \gamma \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}. \quad (34)$$

Since this should have a solution, the determinant of  $B$  should be zero. Expressing this determinant in  $\lambda$  gives a 12 degree polynomial. This polynomial can be factorized into an 8-degree polynomial and a 4-degree polynomial. The 4-degree polynomial [containing spurious solutions due to the denominator from the linear solution of (32)] can be factored out, leaving an 8-degree polynomial to solve. The rest of the unknowns can be found by back substitution. The least squares solution should be one of the eight solutions, and we simply choose the one with smallest error.

## 6 Experimental Validation

We have conducted a number of experiments, on both synthetic and real data, to show the performance of our solvers. We have implemented all solvers in MATLAB, and they are publicly released at <https://github.com/hamburgerlady/ortho-gem>. In Table 1, the running times for our MATLAB implementations are shown. All tests were conducted on a desktop computer running Ubuntu, with an Intel Core i7 3.6 GHz processor.

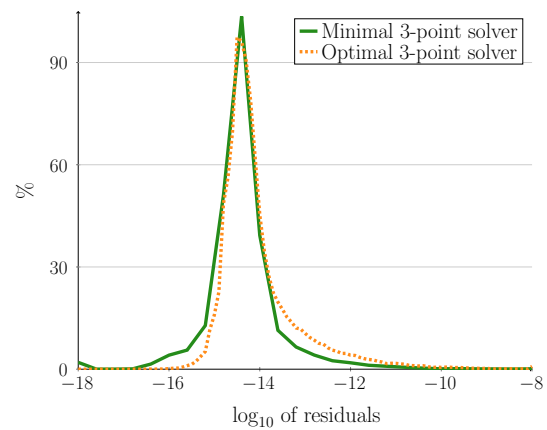
### 6.1 Synthetic Data

In order to test the numerical stability of our minimal solvers, we did some simple tests.

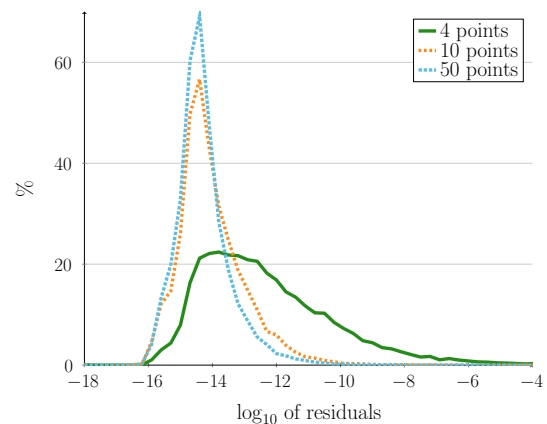
We randomly generated true orthographic epipolar geometry, with corresponding essential matrices and image data (without added noise, 10,000 instances). We then ran our solvers: the minimal three-point solver and the three-point solver used in the optimal inlier solver. For the optimal inlier solver we also used a random bound. We then calculated the equation residuals for all solutions ((3) and (12) for the minimal three-point solver and (3) and (19) for the optimal three-point solver). The resulting histogram of the logarithm of the errors is shown in Fig. 1, where one can see that we get errors close to machine-precision.

**Table 1** Running times of our minimal and least squares solvers. The solvers were implemented in MATLAB and run on a desktop computer with an Intel Core i7 3.6 GHz processor

	Minimal 3pt solver	Least squares solver	Optimal 3pt solver
Time (ms)	0.001	0.45	0.35
Solutions	2	1	32

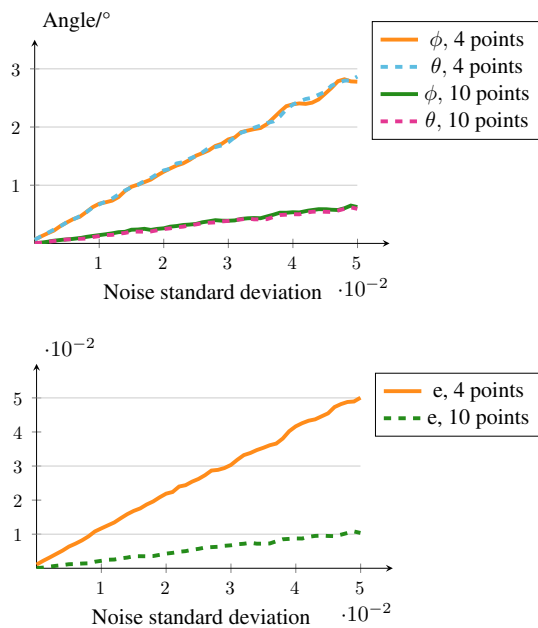


**Fig. 1** Evaluation of the numerical performance of our minimal solvers. The graph shows the histograms of the errors for the minimal three-point solver and the three-point solver for the optimal inlier solver, respectively



**Fig. 2** Graph shows histograms of the logarithm of error norms between the estimated essential matrix and the ground truth, using our least square solver. We get small errors in general, with slightly larger errors for the almost minimal case of four points, which is to be expected

In order to test our least squares solver, we again randomly generated true orthographic epipolar geometry, without noise. We then ran our least squares solver on a large number of examples (10,000), and for each solution we recorded the Frobenius norm of the difference of the estimated essential matrix and the ground truth essential matrix. The distributions of errors are shown in Fig. 2. We get slightly different behavior for different number of points, but overall we get results with very small errors. To test the dependence on image point errors, for the least square solver, we ran a similar experiment as the previous one. We added Gaussian noise to the image measurements and ran our least square solver. In Fig. 3, the resulting errors in the final estimates are shown. Top shows the mean absolute angular error between the estimated and ground truth epipolar line directions for the two images. Bottom shows the corresponding absolute



**Fig. 3** Graphs show the least square estimates of the essential matrix using four and ten points. *Top* The angle deviations from the ground truth epipolar line directions, measured in degrees ( $\theta$  in the first image, and  $\phi$  in the second). *Bottom* The deviation from the ground truth epipolar line offset. Note that since we are working with normalized coordinates, noise with standard deviation equal to 0.05 corresponds to a large pixel error

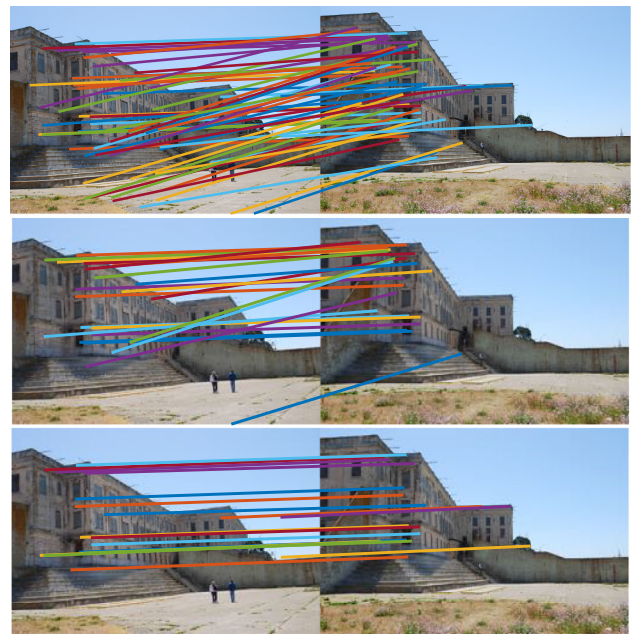
error in line offset (the last entry in the essential matrix). We show the results using 4 and 10 correspondences.

## 6.2 Semisynthetic Data

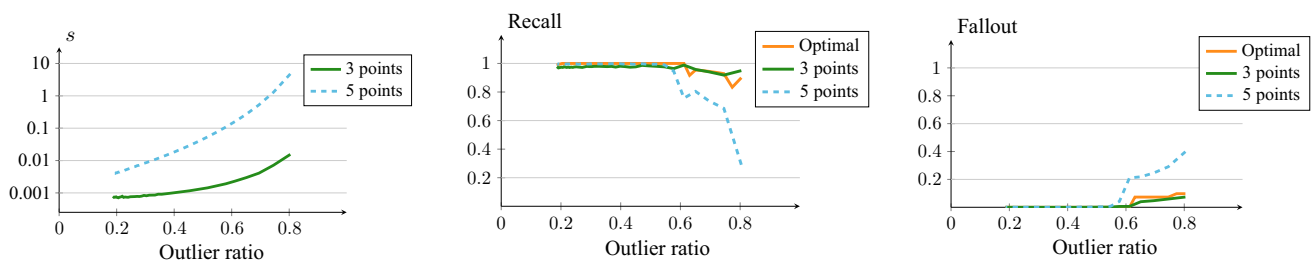
To test our algorithms' behavior in the presence of outliers and real errors, we have conducted a number of tests based on real data.

*Alcatraz* In a first experiment, we used the *Alcatraz* dataset, as described in [19]. We used two images (image number 1 and 39) with SIFT correspondences, including outliers. Since

we had access to the ground truth, we knew which correspondences were inliers and outliers, respectively. We have compared our minimal three-point solver to the calibrated five-point solver, using the MATLAB-mex implementation from [23], that runs in around 0.15 ms. We used 41 outliers in the correspondences and varied the number of inliers between 10 and 180. This gave us a number of sets with outlier ratio varying between 19 and 80%. The input for a large outlier ratio is shown in Fig. 5. Using a standard RANSAC loop—with the number of iterations set so that we would find an inlier set with probability 99.9%—we compared the five-point solver to our minimal three-point solver. The results are



**Fig. 5** One example of the output of the semisynthetic RANSAC experiment, for a high outlier ratio (70%). *Top* input correspondences. *Middle* the five-point algorithm. *Bottom* the minimal three-point solver. We have in this case few inliers, and the five-point solver finds an erroneous solution with a larger number of inliers due to the relatively large number of degrees of freedom in the fitted model. The more restricted orthographic model gives in this case a better estimate of the true epipolar geometry



**Fig. 4** Graphs show the result of the semisynthetic RANSAC experiment, comparing our minimal three-point solver to the calibrated five-point solver. Also included are the results of running our optimal method, that maximizes the number of inliers. *Left* The average run-

ning times for the two methods as functions of the outlier ratio. *Middle* and *right* The recall and fallout, respectively, for the three methods, as functions of the outlier ratio

shown in Fig. 4. The left graph shows the running times as function of outlier ratio, on a logarithmic scale. Our method is faster due to three factors. Firstly, the minimal solver is faster. Secondly, since our solver maximally gives two solutions, we get in general fewer real solutions that need evaluating on the whole point set. And thirdly, since we only need to sample three points, the likelihood of finding an outlier free hypothesis set is higher. We get these benefits at the cost of having a more restrictive camera model, which might not capture the complete geometry. This will of course be highly scenario dependent, but as shown in Fig. 4, where the recall and fallout of the estimated inlier set are shown, we match the five-point solver well in terms of finding the true inlier set. In Sect. 6.4, we give more results on how well the orthographic model approximates the true epipolar geometry. We have also compared with our optimal method that maximizes the number of inliers. The running time of this method doesn't depend on the inlier ratio, but only on the number of initial correspondences, but it is in general much slower than the RANSAC methods.

In the middle of Fig. 5, the resulting inlier set for the five-point solver is shown. Below this, the result of the minimal three-point solver is shown. In this case, there are very few inliers, and the more restricted orthographic camera model serves as a regularization. On a systems level, this would of course be easy to handle using the five-point solver, since the full 3D geometry probably wouldn't fit, but we still believe that this shows the benefit of using a simpler model for establishing correspondences.

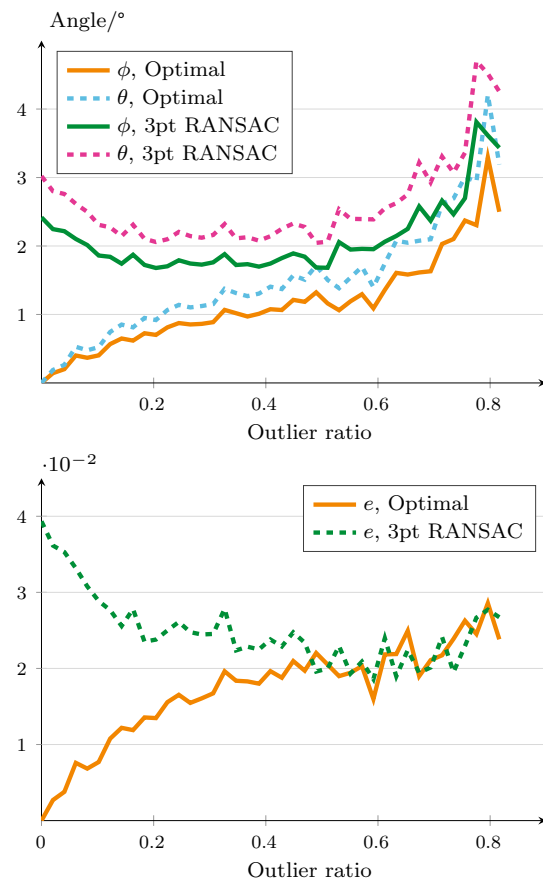
**Dinosaur** To test our optimal inlier maximization algorithm, as described in Sect. 4, we used the well-known dinosaur sequence. This sequence contains very little outliers, and the camera geometry has been shown to be well approximated by an affine model, see, e.g., [11].

We used two calibrated images (image 20 and 22) with a subset of 51 true correspondences. Using the full inlier set, we estimated the essential matrix using our least square solver. This was used to define the ground truth of our experiment. We then randomly corrupted a subset of the correspondences to simulate gross outliers and ran our optimal inlier maximization. As a comparison, we also ran our minimal three-point solver using RANSAC exhaustively.

The result of the error between the estimated essential matrices and the ground truth, for different rates of outliers, is shown in Fig. 6. We show the error in angle and offset in a similar way as in Sect. 6.1. As can be seen in the figure, the errors degrade gracefully as functions of the outlier ratio.

### 6.3 Real Data and Repetitive Structures

One scenario where the matching of features is difficult is when the depicted scene contains repetitive structures. This



**Fig. 6** Top absolute error between the ground truth and estimated epipolar line angles as functions of outlier ratio, for the semisynthetic optimal estimation experiment. Bottom absolute error between the ground truth and estimated offset element of the essential matrix, as function of the outlier ratio

is quite common and can occur due to similar textures of objects. To test the use of our minimal solver in such a setting, we conducted a small experiment. The setup of our *Heinz* experiment is shown in Fig. 7. We took four images of a number of cans, with the same texture on them. We extracted SURF features and matched these based on their feature vectors, for all pairs of the four images. A subset of the initial matches can be seen to the left in Fig. 7. For the six pairs, we got between 330 and 530 tentative correspondences. We then ran 5,000 iterations of our minimal three-point solver to estimate the inlier set. We then used this inlier set to estimate the full projective epipolar geometry, with subsequent bundle adjustment. The original points and reprojected points are shown in the middle of the figure. The corresponding rms-value for all views was 1.14 in unnormalized coordinates. We have also run our optimal inlier method, giving a corresponding rms-value for all views of 1.58 after bundle adjustment. The reprojection points can be seen to the right of Fig. 7. The number of inliers for all pairs is given in Table 2.





**Fig. 7** To the left, the four input images from the *Heinz* experiment are shown, with the initial feature point matches for all pairs of images—indicated by *lines*—overlaid. To help visibility, only 50 random correspondences for each pair are shown. As can be seen, there are

multiple mismatches due to the repetitive scene. To the right, the initial points (*red*) and reprojected points (*yellow*) are shown, after running a RANSAC loop with our minimal three-point, respectively, optimal inlier solver and subsequent bundle adjustment (Color figure online)

**Table 2** Number of initial correspondences and the number of estimated inliers for the two repetitive structures experiments

	1	2	3	4	5	6	rms
<i>Heinz</i>							
Init.	382	343	333	385	337	536	
3pt	47	36	34	42	25	65	1.14
Opt.	57	52	44	52	27	92	1.58
<i>Facade</i>							
Init.	668	318	530	230	560	332	
3pt	307	124	207	88	326	84	2.04
Opt.	354	149	230	90	359	116	2.63

The numbers are given for all combinations of the four respective images. Also shown are the final rms-values of the reprojection errors after bundle adjustment

In a second experiment, *facade*, we used four images of a building facade. These images contain a number of repeating structures, such as windows, roof texture, dormer windows, and chimneys. We again extracted SURF features and matched these, for all pairs of the four images. The estimated inliers set using 10,000 iterations of RANSAC with our minimal three-point solver is shown in Fig. 8. We have also run the optimal inlier maximization on this dataset—the details are given in Table 2. In both these experiments, running the optimal solver is orders of magnitude slower than RANSAC, so it is not directly practical to use it for these amounts of initial correspondences. In order for it to be tractable, some form of initial pruning would be needed, but this is left for future work.

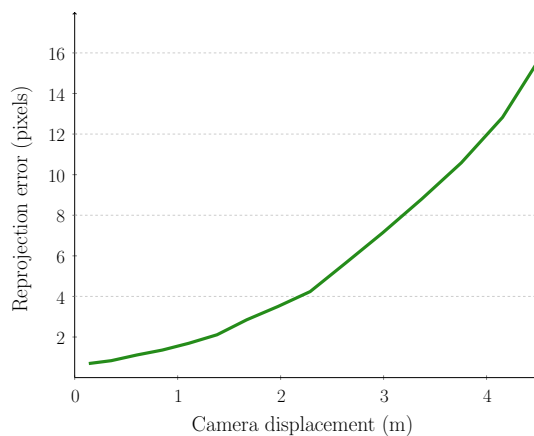
#### 6.4 Orthographic Model Fit

In order to investigate how well the orthographic model approximates the true epipolar geometry, we made the fol-



**Fig. 8** Inlier image points (*red*) and reprojected points (*yellow*), after running a RANSAC loop with our minimal three-point solver and subsequent bundle adjustment on the *facade* dataset (Color figure online)

lowing test on the *Alcatraz* dataset. We used the true inlier set for all pairs of images and used our least squares solver to find the orthographic essential matrix. We then calculated the mean reprojection error (12) for all pairs. We looked at how this error varies with different parameters such as the medium depth of all scene points, the variance of the depth of the scene points, and the baseline between the two cameras. Our conclusion for this test set was that the important parameter was the baseline between the two cameras. In Fig. 9, we show the mean error as a function of the distance between the two corresponding camera centers. The metric scale for the cameras was manually estimated from the images. The average depth of the scene points was 8.2m for this dataset (with a standard deviation of 1.4m).



**Fig. 9** Mean reprojection error as a function of the *baseline* between the two cameras. The essential matrix was estimated using our least squares solver using a validated inlier correspondence set. One can see that for this image sequence we get reasonable errors for a camera baseline below 2.5 m

## 7 Conclusion

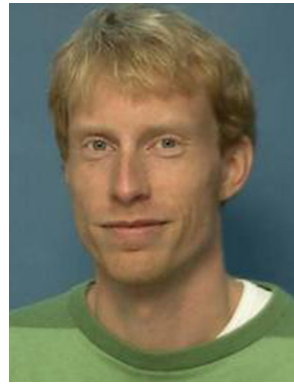
We have in this paper given methods and algorithms for estimating two-view geometry for orthographic cameras. We have shown how to estimate the corresponding essential matrix minimally (using three point correspondences), in a least squares sense, or optimally with respect to the number of inliers. These methods can be used to robustly find inlier correspondences in the presence of high degrees of outliers. They depend on an orthographic camera model, but we indicate in the experimental section that in many cases this model is a very good approximation. Our low-dimensional solvers give many benefits over the full projective estimates. Due to the simplicity of our minimal solver, we get a faster solver that also gives fewer solutions than, e.g., the five-point calibrated solver, which leads to faster validation in a RANSAC loop. Of course, these benefits come at the cost of assuming a more restrictive camera model. This model might not capture the complete geometry and may be biased toward affine geometry. This caveat aside, we believe that we get a very fast framework for robust two-view correspondence estimation. Even though our optimal inlier estimation is based on only three point correspondences, it is in many cases not tractable to check all combinations of three points. Future work will include investigating recent methods that can reduce the number of initial correspondences without sacrificing optimality [2,25].

**Open Access** This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

## References

1. Arandjelovic, R., Zisserman, A.: Efficient image retrieval for 3d structures. In: British Machine Vision Conference, pp. 1–11 (2010)
2. Ask, E., Enqvist, O., Kahl, F.: Optimal geometric fitting under the truncated  $l_2$ -norm. In: Conference on Computer Vision and Pattern Recognition (2013)
3. Breuel, T.: Implementation techniques for geometric branch-and-bound matching methods. *Comput. Vis. Image Understand.* **90**(3), 258–294 (2003)
4. Cox, D.A., Little, J., O’Shea, D.: Using Algebraic Geometry, vol. 185. Springer, Berlin (2005)
5. Enqvist, O., Ask, E., Kahl, F., Åström, K.: Tractable algorithms for robust model estimation. *Int. J. Comput. Vis.* **112**(1), 115–129 (2015)
6. Goshen, L., Shimshoni, I.: Balanced exploration and exploitation model search for efficient epipolar geometry estimation. *IEEE Trans. Pattern Anal. Mach. Intell.* **30**(7), 1230–1242 (2008)
7. Harris, C.: Structure-from-motion under orthographic projection. In: European Conference on Computer Vision, pp. 118–123. Springer (1990)
8. Hartley, R., Zisserman, A.: Multiple View Geometry in Computer Vision. Cambridge University Press, Cambridge (2004)
9. Hu, X., Ahuja, N.: Motion estimation under orthographic projection. *IEEE Trans. Robot. Autom.* **7**(6), 848–853 (1991)
10. Huang, T.S., Lee, C.: Motion and structure from orthographic projections. *IEEE Trans. Pattern Anal. Mach. Intell.* **11**(5), 536–540 (1989)
11. Jiang, F., Oskarsson, M., Åström, K.: On the minimal problems of low-rank matrix factorization. In: Conference on Computer Vision and Pattern Recognition, pp. 2549–2557. IEEE (2015)
12. Kahl, F., Hartley, R.: Multiple view geometry under the  $L_\infty$ -norm. *IEEE Trans. Pattern Anal. Mach. Intell.* **30**(9), 1603–1617 (2008)
13. Ke, Q., Kanade, T.: Quasiconvex optimization for robust geometric reconstruction. In: International Conference on Computer Vision, pp. 986–993. Beijing, China (2005)
14. Koenderink, J.J., Van Doorn, A.J.: Affine structure from motion. *J. Opt. Soc. Am. A* **8**(2), 377–385 (1991)
15. Lehmann, S., Bradley, A.P., Clarkson, I.V.L., Williams, J., Kootsookos, P.J.: Correspondence-free determination of the affine fundamental matrix. *IEEE Trans. Pattern Anal. Mach. Intell.* **29**(1), 82–97 (2007)
16. Li, H.: A practical algorithm for  $L_\infty$  triangulation with outliers. In: Conference on Computer Vision and Pattern Recognition (2007)
17. Mendonça, P.R., Cipolla, R.: Estimation of epipolar geometry from apparent contours: Affine and circular motion cases. In: Conference on Computer Vision and Pattern Recognition. IEEE (1999)
18. Mundy, J.L., Zisserman, A.: Geometric Invariance in Computer Vision. MIT Press, Cambridge (1992)
19. Olsson, C., Enqvist, O.: Stable structure from motion for unordered image collections. In: Scandinavian Conference on Image Analysis (2011)
20. Perd’och, M., Matas, J., Chum, O.: Epipolar geometry from two correspondences. In: International Conference on Pattern Recognition, vol. 4, pp. 215–219. IEEE (2006)
21. Pritts, J., Chum, O., Matas, J.: Approximate models for fast and accurate epipolar geometry estimation. In: 28th International Conference of Image and Vision Computing New Zealand (IVCNZ), pp. 106–111. IEEE (2013)
22. Shapiro, L.S., Zisserman, A., Brady, M.: 3d motion recovery via affine epipolar geometry. *Int. J. Comput. Vis.* **16**(2), 147–182 (1995)
23. Stewenius, H., Engels, C., Nistér, D.: Recent developments on direct relative orientation. *ISPRS J. Photogramm. Remote Sens.* **60**(4), 284–294 (2006)

24. Svärm, L., Enqvist, O., Kahl, F., Oskarsson, M.: City-scale localization for cameras with known vertical direction. *IEEE Trans. Pattern Anal. Mach. Intell.* **PP**(99), 1–1 (2016). doi:[10.1109/TPAMI.2016.2598331](https://doi.org/10.1109/TPAMI.2016.2598331)
25. Svärm, L., Enqvist, O., Oskarsson, M., Kahl, F.: Accurate localization and pose estimation for large 3d models. In: *Conference on Computer Vision and Pattern Recognition* (2014)
26. Torr, P.H.: Model Selection for Two View Geometry: A Review. In: Forsyth, D.A., Mundy, J.L., di Gesu, V., Cipolla, R. (eds.) *Shape, Contour and Grouping in Computer Vision*, pp. 277–301. Springer, Berlin (1999)
27. Ullman, S.: The interpretation of structure from motion. *Proceed. R. Soc. Lond. B Biol. Sci.* **203**(1153), 405–426 (1979)



**Magnus Oskarsson** received his M.Sc. degree in Engineering Physics in 1997 and Ph.D. in Mathematics in 2002 from the University of Lund, Sweden. His thesis work was devoted to computer vision with applications for autonomous vehicles. He is currently an associate professor at the Centre for Mathematical Sciences, Lund University, where his teachings include undergraduate and graduate courses in mathematics and image analysis.

He is the author and co-author to a number of papers in international journals and conference proceedings within geometry, algebra, and optimization with applications in computer vision, cognitive vision, and image enhancement.