

Optimal Surface Normal from Affine Transformation

Barath Daniel^{1,2}, Jozsef Molnar² and Levente Hajder^{1,2}

¹*Geometric Modelling and Computer Vision Laboratory, MTA SZTAKI, H-1111 Kende utca 13-17, Budapest, Hungary*

²*Department of Image Processing and Computer Graphics, Institute of Informatics, University of Szeged, Árpád tér 2., H-6720 Szeged, Hungary*

Keywords: 3D Reconstruction, Normal Estimation, Calibrated Stereo Images.

Abstract: This paper deals with surface normal estimation from calibrated stereo images. We show here how the affine transformation between two projections defines the surface normal of a 3D planar patch. We give a formula that describes the relationship of surface normals, camera projections, and affine transformations. This formula is general since it works for every kind of cameras. We propose novel methods for estimating the normal of a surface patch if the affine transformation is known between two perspective images. We show here that the normal vector can be optimally estimated if the projective depth of the patch is known. Other non-optimal methods are also introduced for the problem. The proposed methods are tested both on synthesized data and images of real-world 3D objects.

1 INTRODUCTION

Although computer vision has been an intensively researched area in computer science from many decades, several unsolved problems exist in the field. This paper proposes a novel optimal method for estimating the normal vector of a planar surface patch if the affine transformation of the patch between two calibrated (stereo) images is known.

The normal vector estimation problem itself can most accurately be solved by photometric stereo (PS) that was introduced many decades ago (Woodham, 1978). The main drawback of this method is that it can only be used in laboratories where light conditions are totally controlled. PS usually assumes that the object to be reconstructed is illuminated by directional light source(s) (Woodham, 1978), but point-light sources can also be applied. (Fodor et al., 2014).

The image-based normal vector estimation is usually carried out by decomposing the homography of the tangent plane between stereo images (Faugeras and Lustman, 1988; Malis and Vargas, 2007). Unfortunately, these methods are ambiguous as it was shown in several studies (e.g. in (Liu, 2012)). To the best of our knowledge, the problem of image-based normal vector estimation from affine transformation has not been solved yet. The first similar work was published in two papers by Habbecke and Kobbelt in (Habbecke and Kobbelt, 2006; Habbecke and Kobbelt, 2007).

They estimate the parameters of a flat patch based on photo-consistency. The plane is parameterized in 3D by the implicit parameters of a general plane. (These are three real values as the implicit parameters of a 3D plane are defined up to an arbitrary scale.) Our method only concentrates on the estimation of the spatial normal vector since the point of the plane can be estimated in 3D by triangulation if corresponding projections on two patches are known (Hartley and Sturm, 1997; Hartley and Zisserman, 2003).

The closest work to our study is that of Megyesi et al. (Z. Megyesi and D. Chetverikov, 2006). They compute a dense 3D reconstruction using normal vectors. The normal vectors themselves are calculated from the affine parameters between a rectified stereo image pair. For this reason, only two parameters of the affine transformation have to be estimated. The drawback of this work is that the rectification itself cannot be perfect due to noise and computational inaccuracy. Our method proposing here is more general as it works on arbitrary stereo image pairs. The only restriction is that the stereo images have to be taken by perspective cameras. (Or the non-perspective distortion of the images has to be undistorted.)

To the best of our knowledge, this is the first study that deals with surface normal computation from affine transformation using calibrated stereo images. *The main contribution of this paper is twofold:* (i) We show here the general relationship among sur-

face normal vector, affine transformation and camera parameters. *The formulas proposing here is valid for every kind of cameras.* (ii) Different surface normal estimators are proposed here including an optimal one that *finds the optimal normal vector in the least squares sense if the affine parameters are contaminated with noise.*

2 GEOMETRIC BACKGROUND

Two projections of a 3D surface are given in stereo images. If the neighboring pixels are selected around image locations, these pixels form the so-called patches. The affine transformation between two corresponding patches is assumed to be known. The goal of this study is to show how the surface normal n can be estimated if the images are calibrated. The problem is visualized in Figure 1.

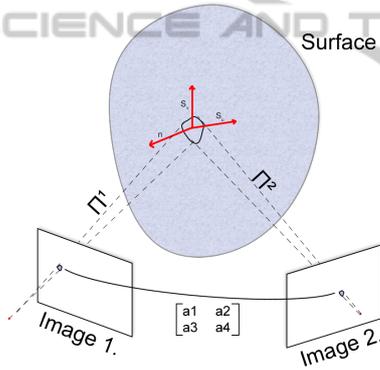


Figure 1: 3D patch perspective projected to stereo images.

$$x = \Pi_x(X, Y, Z) \quad y = \Pi_y(X, Y, Z)$$

The surface point $[X, Y, Z]^T$ is written in parametric form

$$X = X(u, v), \quad Y = Y(u, v), \quad Z = Z(u, v).$$

As it is well-known in differential geometry (Kreyszig, 1991), the tangent vectors of the plane are written by the partial derivatives of the spatial coordinates, while the surface normal is given by the cross product of the tangent vectors.

$$S_u = \begin{bmatrix} \frac{\partial X(u,v)}{\partial u} \\ \frac{\partial Y(u,v)}{\partial u} \\ \frac{\partial Z(u,v)}{\partial u} \end{bmatrix}$$

$$S_v = \begin{bmatrix} \frac{\partial X(u,v)}{\partial v} \\ \frac{\partial Y(u,v)}{\partial v} \\ \frac{\partial Z(u,v)}{\partial v} \end{bmatrix} \quad n = S_u \times S_v$$

It is known that the 3D point $[X, Y, Z]^T$, and tangent vectors S_u and S_v determine the tangent plane. Locally, the surface can be approximated by its tangent plane. We assumed that we have images taken from the object. Now, a point of the surface close to the given 3D location $[X, Y, Z]^T$ is approximated by the first order Taylor-series:

$$\begin{bmatrix} x + \Delta x \\ y + \Delta y \end{bmatrix} \approx \begin{bmatrix} \Pi_x(X, Y, Z) \\ \Pi_y(X, Y, Z) \end{bmatrix} + \begin{bmatrix} \frac{\partial \Pi_x(X, Y, Z)}{\partial u} & \frac{\partial \Pi_x(X, Y, Z)}{\partial v} \\ \frac{\partial \Pi_y(X, Y, Z)}{\partial u} & \frac{\partial \Pi_y(X, Y, Z)}{\partial v} \end{bmatrix} \begin{bmatrix} \Delta u \\ \Delta v \end{bmatrix}$$

Let us see that the partial derivatives of the projection functions give the affine transformation between 3D and 2D surface patches.

$$\begin{bmatrix} \Delta x \\ \Delta y \end{bmatrix} \approx A \begin{bmatrix} \Delta u \\ \Delta v \end{bmatrix} \quad A = \begin{bmatrix} \frac{\partial \Pi_x(X, Y, Z)}{\partial u} & \frac{\partial \Pi_x(X, Y, Z)}{\partial v} \\ \frac{\partial \Pi_y(X, Y, Z)}{\partial u} & \frac{\partial \Pi_y(X, Y, Z)}{\partial v} \end{bmatrix}$$

The partial derivatives can be reformulated using the chain rule. For instance,

$$\frac{\partial \Pi_x(X, Y, Z)}{\partial u} = \frac{\partial \Pi_x(X, Y, Z)}{\partial X} \frac{X}{\partial u} + \frac{\partial \Pi_x(X, Y, Z)}{\partial Y} \frac{Y}{\partial u} + \frac{\partial \Pi_x(X, Y, Z)}{\partial Z} \frac{Z}{\partial u} = \nabla \Pi_x^T S_u,$$

where $\nabla \Pi_x$ is the gradient vector of the projection function w.r.t. the spatial coordinates X , Y , and Z of the surface patch. Similarly,

$$\frac{\partial \Pi_x}{\partial v} = \nabla \Pi_x^T S_v \quad \frac{\partial \Pi_y}{\partial u} = \nabla \Pi_y^T S_u \quad \frac{\partial \Pi_y}{\partial v} = \nabla \Pi_y^T S_v.$$

Therefore, the affine matrix can be written as

$$A = \begin{bmatrix} \nabla \Pi_x^T \\ \nabla \Pi_y^T \end{bmatrix} \begin{bmatrix} S_u & S_v \end{bmatrix}.$$

In stereo vision, two images are given. The affine transformation between the image patches is obtained by concatenating the inverse of affine transformation A_1 (between the patches of image #1 and the spatial patch), and the affine transformation A_2 (between 3D patch and that in image #2). Formally, it can be written as

$$\begin{bmatrix} \Delta x_2 & \Delta y_2 \end{bmatrix}^T = A_2 A_1^{-1} \begin{bmatrix} \Delta x_1 & \Delta y_1 \end{bmatrix}^T$$

$A_2A_1^{-1}$ is the affine transformation between the images. The inverse of the affine matrix A can be written as

$$A^{-1} = \frac{1}{\det(A)} \begin{bmatrix} \Pi_x^T S_u & -\Pi_y^T S_u \\ -\Pi_x^T S_v & \Pi_y^T S_v \end{bmatrix}$$

where $\det(A) = \Pi_x^T S_u \Pi_y^T S_v - \Pi_x^T S_v \Pi_y^T S_u$. If one makes elementary modification utilizing the fact that $S_v S_u^T - S_u S_v^T = [N]_{\times}$, then the affine transformation $A_2A_1^{-1}$ can be written as

$$A_1^{-1}A_2 = \frac{1}{\Pi_x^T [N]_{\times} \Pi_y^1} \begin{bmatrix} \Pi_x^{2T} [N]_{\times} \Pi_y^1 & \Pi_x^{1T} [N]_{\times} \Pi_x^2 \\ \Pi_y^{2T} [N]_{\times} \Pi_y^1 & \Pi_x^{1T} [N]_{\times} \Pi_y^2 \end{bmatrix}$$

Note that the scale of the normal is arbitrary since both the determinant and the matrix elements are multiplied with the scale of $[N]_{\times}$.

The expression $a^T [N]_{\times} b$ is also called the scalar triple product. Remark that $a^T [n]_{\times} b$ equals to $n^T (b \times a)$. Therefore, the final equation of the affine transformation is written as

$$\begin{bmatrix} a_1 & a_2 \\ a_3 & a_4 \end{bmatrix} = A_1^{-1}A_2 = \frac{1}{n^T w_5} \begin{bmatrix} n^T w_1 & n^T w_2 \\ n^T w_3 & n^T w_4 \end{bmatrix} \quad (1)$$

where $w_1 = \nabla \Pi_y^1 \times \nabla \Pi_x^2$, $w_2 = \nabla \Pi_x^2 \times \nabla \Pi_x^1$, $w_3 = \nabla \Pi_y^1 \times \nabla \Pi_y^2$, $w_4 = \nabla \Pi_y^2 \times \nabla \Pi_x^1$, and $w_5 = \nabla \Pi_y^1 \times \nabla \Pi_x^1$. Equation 1 is a very important formula which shows the relations of the surface normal and the projection of the surface to the stereo image pair. *A very important advantage of this formula is that it is valid for every kind of camera since only the two projective equations must be known.* We show here that the above formula can be used to compute the surface normal if the perspective parameters are calibrated.

2.1 Pin-hole Camera Model

When the standard perspective camera model is used, the projection is written as

$$[x, y, 1]^T = \frac{1}{s} P_{persp} [X, Y, Z, 1]^T, \quad (2)$$

where $[x, y]$ are the projected coordinates in image space, s is the projective depth, P_{persp} is the so called projection matrix with size 3×4 . Let us denote the rows of the projective matrix by p_1^T , p_2^T , and p_3^T . The projection formulas and their gradients can be written as

$$\Pi_x = \frac{p_1^T [X, Y, Z, 1]^T}{s} \quad \Pi_y = \frac{p_2^T [X, Y, Z, 1]^T}{s}$$

$$\nabla \Pi_x = \frac{1}{s} \begin{bmatrix} P_{11} + xP_{31} \\ P_{12} + xP_{32} \\ P_{13} + xP_{33} \end{bmatrix}$$

$$\nabla \Pi_y = \frac{1}{s} \begin{bmatrix} P_{21} + yP_{31} \\ P_{22} + yP_{32} \\ P_{23} + yP_{33} \end{bmatrix}$$

where P_{ij} is the element in the i^{th} row and j^{th} column in projection matrix P_{persp} . The projective depth is obtained as $s = p_3^T [X, Y, Z, 1]^T$. The affine transformation becomes

$$\begin{bmatrix} a_1 & a_2 \\ a_3 & a_4 \end{bmatrix} = \frac{1}{\alpha n^T w_5} \begin{bmatrix} n^T w_1 & n^T w_2 \\ n^T w_3 & n^T w_4 \end{bmatrix} \quad (3)$$

where $\alpha = s^1/s^2$ is the ratio of the projective depths in the first and second images, and $w_1 = s^1 s^2 (\nabla \Pi_y^1 \times \nabla \Pi_x^2)$, $w_2 = s^1 s^2 (\nabla \Pi_x^2 \times \nabla \Pi_x^1)$, $w_3 = s^1 s^2 (\nabla \Pi_y^1 \times \nabla \Pi_y^2)$, $w_4 = s^1 s^2 (\nabla \Pi_y^2 \times \nabla \Pi_x^1)$, and $w_5 = s^2 s^2 (\nabla \Pi_x^1 \times \nabla \Pi_y^1)$.

A very important remark is that if the projective depth s^i is unknown, but the upper left 3×3 submatrices of the projection matrices P_1 and P_2 are known then the gradient vectors can be calculated up to an unknown scale. (This scale is the multiplicative inverse of the projective depth s^i .) Also note that the vectors $w_1 \dots w_4$ are scaled by $s^1 s^2$ while w_5 by $s^2 s^2$.

Therefore, the normal vector is independent of the translation between the two cameras since the last columns of the projection matrices are the product of the camera intrinsic parameters and the translation. For this reason, the following two cases must be distinguished:

1. Both projection matrixes P^1 and P^2 are known. (In other words, the cameras are calibrated.)
2. Only the upper-left 3×3 submatrices of the projections are known. In this case, the projective depth of the points is not known. However, the gradients can be computed up to a scale where this scale is the inverse of the projective depth.

Also remark that the normal vector cannot be estimated if $w_5 = 0$. This can only be true if $\nabla \Pi_x^1$ and $\nabla \Pi_y^1$ are parallel which is not a realistic case as it is only possible if the first and second rows of the 3×3 submatrix of projection matrix P_{persp} are parallel. If the camera calibration is valid it cannot be true. The problem itself is numerically stable if the angles between the vectors $w_1 \dots w_4$ are relatively large. To our experiments, this is true for realistic reconstruction problems.

3 NORMAL VECTOR ESTIMATION

This section shows different normal vector estimators. The first one is very fast and simple, later more sophisticated and accurate methods are introduced.

3.1 Fast Normal Estimation (FNE)

The base matrix equation (Eq. 3) consists of 4 elements. If two of those are selected and they are divided by each other, then an equation is obtained. If the same procedure is repeated for the rest of the matrix elements, then the second equation can be similarly yielded. For instance, two elements of the first and second rows give the equations

$$\frac{w_1^T n}{w_2^T n} = \frac{a_1}{a_2} \quad (4)$$

$$\frac{w_3^T n}{w_4^T n} = \frac{a_3}{a_4} \quad (5)$$

These equations can be trivially modified as

$$\begin{aligned} (a_2 w_1^T - a_1 w_2^T) n &= 0 \\ (a_4 w_3^T - a_3 w_4^T) n &= 0 \end{aligned}$$

The normal vector n is perpendicular to both $a_2 w_1^T - a_1 w_2^T$ and $a_4 w_3^T - a_3 w_4^T$. Therefore, the normal can be obtained as the cross product of these vectors:

$$n = (a_2 w_1^T - a_1 w_2^T) \times (a_4 w_3^T - a_3 w_4^T). \quad (6)$$

Of course, the obtained vector n should be normalized, its length must be 1. A very nice property of this normal estimation is that it is independent of the scales appearing in vectors $w_1 \dots w_4$. Therefore, the projective depths of the patch are not required to estimate the normal, because they influence only the length of n .

Remark, that the equation pairs (Eq. 4 & 5) can be selected in other two ways. In those cases, the normal vector is given by

$$n = (a_3 w_1^T - a_1 w_3^T) \times (a_4 w_2^T - a_2 w_4^T) \quad (7)$$

or

$$n = (a_4 w_1^T - a_1 w_4^T) \times (a_3 w_2^T - a_2 w_3^T) \quad (8)$$

3.2 Optimal Normal Estimation with Known Projective Depth (OPT)

The aim of the optimal method is to minimize the error in the matrix base equation (Eq. 3). Formally, the estimation itself can be written as the minimization of

Frobenius norm of Equation 3 with respect to normal n . This is equivalent to

$$\arg \min_n \sum_{k=1}^4 \left(\frac{n^T w_k}{n^T w_5} - a_k \right)^2 \quad (9)$$

It minimizes the normal vector in the least square sense assuming that the affine parameters are contaminated with noise. (This assumption is valid since the affine parameters are estimated as described later in Sec. 4.2 in short, and this estimation cannot be perfect since the images themselves contain noise.) The optimal solution is given in first part of the Appendix with $\alpha = 1$.

3.3 Normal Estimation with Unknown Projective Depths using Alternation (ALT)

If the projective depth is unknown then the base optimization equation (Eq. 9) cannot be applied since the parameter $\alpha = s_1/s_2$ is not known. The cost function defined in Eq. 9 has to be modified as

$$\arg \min_n \sum_{k=1}^4 \left(\frac{n^T w_k}{\alpha n^T w_5} - a_k \right)^2 \quad (10)$$

Unfortunately, this problem cannot be optimally solved to the best of our knowledge. We propose here an alternating-like approach which is overviewed in Alg. 1. The alternation has two steps:

1. EstimateAlpha: The cost function (Eq. 10) is a linear one with respect to $1/\alpha$ since it can be written as $A \frac{1}{\alpha} = b$ where $A = \left[\frac{n^T w_1}{n^T w_5}, \dots, \frac{n^T w_4}{n^T w_5} \right]^T$ and $b = [a_1, \dots, a_4]^T$. The optimal solution of an overdetermined linear system can be solved optimally. In this case, that is obtained as

$$\frac{1}{\alpha} = \frac{n^T w_5}{\sum_j (n^T w_j)^2} \sum_j n^T w_j a_j \quad (11)$$

2. EstimateNormal: The normal vector estimation is very similar to the optimal method described above, the only difference is that the parameter α appears in the denominators. However, the method described in Appendix can solve the sub-problem optimally.

The alternation requires initial values for the parameters n and α to be optimized. We propose to use the linear methods described later in Sec 3.4.2 in order to compute the initial values. The alternation converges to the closest (local) minimum since it optimizes a non-negative cost function and each step decreases (or does not increase) the cost. Unfortunately, we

could not prove theoretically that the global optimum is reached in this way, however, to our practice, the method usually improves the initial normal n .

Algorithm 1: Alternation for Normal Estimation (ALT).

$n, \alpha \leftarrow$ Parameter Initialization by LNE-UPD
repeat
 $\alpha \leftarrow$ EstimateAlpha(n, w_1, \dots, w_5)
 $n \leftarrow$ EstimateNormal(α, w_1, \dots, w_5)
until convergence

3.4 Linear Normal Estimation (LNE)

The base matrix equation (Eq. 3) is a nonlinear one. The elements can be linearized if they are multiplied with their common denominator $\alpha w_5^T n$. Then a cost function can be formed for the elements as

$$\arg \min_n \sum_{k=1}^4 (n^T w_k - \alpha a_k n^T w_5)^2 \quad (12)$$

This is a usual trick, and the solution will not be optimal if this modification is carried out. However, the problem becomes linear, and it can be solved easily (Björck, 1996).

3.4.1 Linear Normal Estimation for Known Projective Depth (LNE-KPD)

If the projective depth is known, then $\alpha = 1$ and the problem can be rewritten as an overdetermined homogeneous linear equation system $An = 0$ subject to $n^T n = 1$, where

$$A = \begin{bmatrix} w_1 - a_1 w_5 \\ w_2 - a_2 w_5 \\ w_3 - a_3 w_5 \\ w_4 - a_4 w_5 \end{bmatrix} \quad (13)$$

The optimal solution of this system is the eigenvector of matrix $A^T A$ corresponding to the smallest eigenvalue (Björck, 1996).

3.4.2 Linear Normal Estimation for Unknown Projective Depth (LNE-UPD)

If the projective depth is unknown, then the function to be optimized in Eq. 12 gives an overdetermined homogeneous linear system $Bb = 0$, similarly to the previous case (Sec. 3.4.1), but the matrix of coefficients B and the vector b differ as follows.

$$B^T = \begin{bmatrix} w_1^T & -a_{11} \\ w_2^T & -a_{12} \\ w_3^T & -a_{21} \\ w_4^T & -a_{22} \end{bmatrix}$$

$$b = \begin{bmatrix} n \\ \alpha w_5^T n \end{bmatrix}$$

The solution is given from the eigenvector of matrix $B^T B$ corresponding to the smallest eigenvalue (Björck, 1996). If this vector is denoted by \hat{b} , then the estimation for the normal vector n is given by the first three coordinates of \hat{b} , but this vector should be normalized in order to fulfill the $n^T n = 1$ constraint. The parameter $\alpha = s_1/s_2$ can also be computed if n is known from the fourth coordinate of \hat{b} .

4 EXPERIMENTAL RESULTS

The proposed normal vector estimators have been tested both on synthesized data and real world images.

4.1 Test on Synthesized Data

During synthesized tests, our main goal was to generate different normal vectors with corresponding affine parameters. For this reason, a stereo image pair was first generated represented by their 3×4 projection matrices. Then a 3D sphere was generated as well sampled by spherical coordinates. The normal vector of the locations on the sphere can easily be calculated as it is the direction pointing from the sphere center to the current surface points. The synthetic sphere with the ground truth normals is visualized in Fig. 2. 72 different patches sampled by spherical coordinates are used in order to compare the methods.

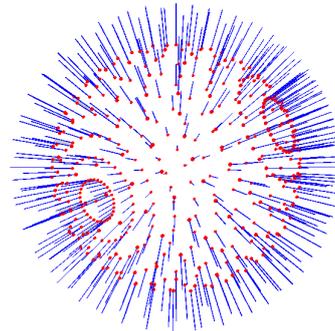


Figure 2: Sphere with normal vectors for synthesized test.

The affine parameters between the stereo images were calculated as follows: (i) The tangent plane of the sphere was determined first, (ii) then it was projected to the stereo images. (iii) The projections of

the plane determine two homographies with respect to the 3D tangent plane. (iv) The homography between the two images were given by concatenating the two 3D→2D homographies. (v) The affine transformation is the first order approximation of the 2D→2D homography at the given locations.

The error values are defined as the average of the angle error between the estimated and ground truth normal vectors. We have tested all the methods described in this study. In each test case, 72 patches of the sphere were generated, and the tests were repeated 50 times. Thus, the average error values come from $72 \cdot 50 = 3600$ run of the competitor methods. Two test cases were simulated: zero-mean Gaussian noise was added to the (i) affine parameters and (ii) to the elements of the projection matrices.

Remark that all the synthesized tests have been implemented on Octave¹.

4.1.1 Test with Contaminated Affine Parameters

We have compared the efficiency of the fast (FNE), alternation (ALT) and optimal (OPT) normal estimators. It is clear that the optimal estimator (OPT) outperforms the others as it is visualized in Fig. 3 since it optimally estimates the normal vector in the least square sense. It is also obvious that the fast method is the less accurate one as the other two methods (OPT and ALT) are significantly more sophisticated.

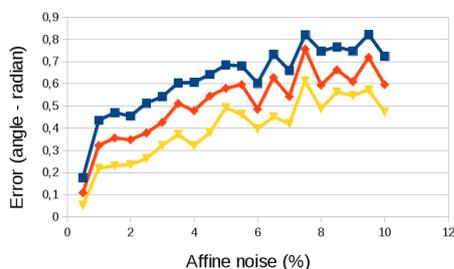


Figure 3: Comparison of methods when affine parameters contaminated.

We have also compared the linear methods to the corresponding non-linear ones. Namely, the linear method with unknown projective depth (LIN-UPD) algorithm is compared to the alternation (Fig. 5) and linear with known projective depth method (LIN-KPD) to the optimal one ((Fig. 6)). The differences are significant only if the optimal (OPT) method is used instead of its linear version.

We have also examined the expected values and the spreads of the five proposed methods. The expected value for the length of the difference between

¹www.octave.org.

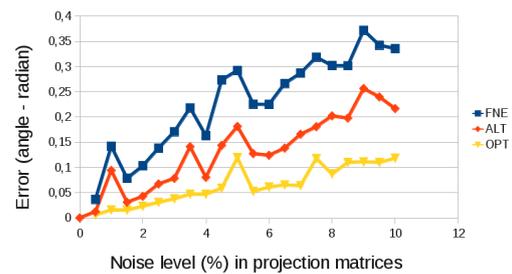


Figure 4: Comparison of methods when projective parameters contaminated.

the ground truth and estimated vectors are close to zero as it is expected. Therefore, the estimators are consistent. Their spreads are listed in Table 1. It shows that the optimal method has the lowest spread as it is expected, FNE is the highest one. It is interesting that the linear method with known projective depths gives significantly better result than the methods with unknown depths (LIN-UPD and ALT),

Table 1: Spread of error vector lengths.

FNE	LIN-UPD	ALT	LIN-KPD	OPT
0.55	0.449	0.433	0.352	0.2919

4.1.2 Test with Contaminated Projection Parameters

Another interesting experiment is when the elements of the projection matrices are contaminated with noise. We have examined the same test cases as in Sec.4.1.1.

When the base FNE, ALT and OPT methods are compared, the dominance of the optimal method (OPT) is more obvious. The performance of the fast (FNE) and alternation (ALT) methods are closer to each other than in the previous test case (Sec. 4.1.1). The accuracy of the result is also the best (highest) for the optimal method. It seems that the other two methods (FNE and ALT) are very sensitive to noise since the corresponding charts (Fig. 4) contain many peaks.

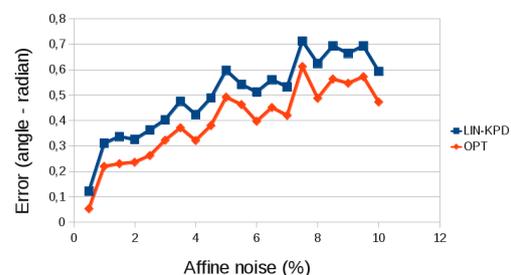


Figure 5: Comparison of linear and corresponding nonlinear methods.

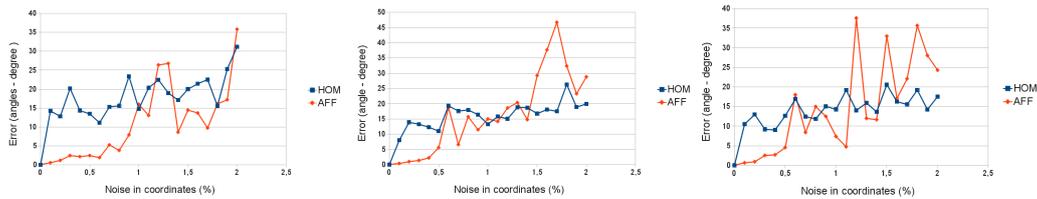


Figure 7: Comparison of normal vector estimators. HOM: normal from homography decomposition AFF: normal from affine parameters by proposed OPT method. Left: transformation estimated from 4 points, Center: 6 points, Right: 8 points.

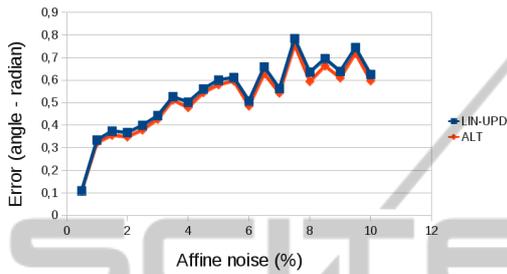


Figure 6: Comparison of linear and corresponding nonlinear methods.

To conclude this synthetic test, it can be declared that the optimal method is the best solution if the projective depth of the 3D point is known. If it is not, the alternation method serves the most efficient method, but its advantage over its linear version is very small. The alternation itself is an iterative algorithm, sometimes it can be very slow. Therefore, we propose the LIN-UPD method for time-critical application, ALT is the best selection for offline algorithms when the projective depths are unknown.

4.1.3 Normal Vector from Affine Parameters Versus Homography

The mainstream solution for computing the normal vector from two patches is to estimate the homography between the patches (Malis and Vargas, 2007). It has eight degrees of freedom, and it can be decomposed into the pose (3 DoF), the location (3 DoF), and the normal (2 DoF) of the plane.

We compare the accuracy of the homography-estimated normal vector with our optimal (OPT) estimator. The synthesized data is given by sampling the surface of a sphere similarly to the synthesized tests above. However, the homography and the affine transformation are both estimated from projected points: points are generated randomly on the 3D tangent plane (close to the location on sphere surface), and these points are projected to the image pair. Then noise is added to the projected coordinates in image space. The homography and the affine transformation are estimated using the corresponding points in im-

age pairs. The estimation of the affine transformation is easier since it is trivial that affine parameter estimation is a linear problem. We estimate the homography via numerical optimization method, the initial parameters are computed by DLT (Direct Linear Transformation) algorithm (Hartley and Zisserman, 2003). Note that at least 4 points are required to estimate the homography, while 3 points are sufficient to compute the affine transformation.

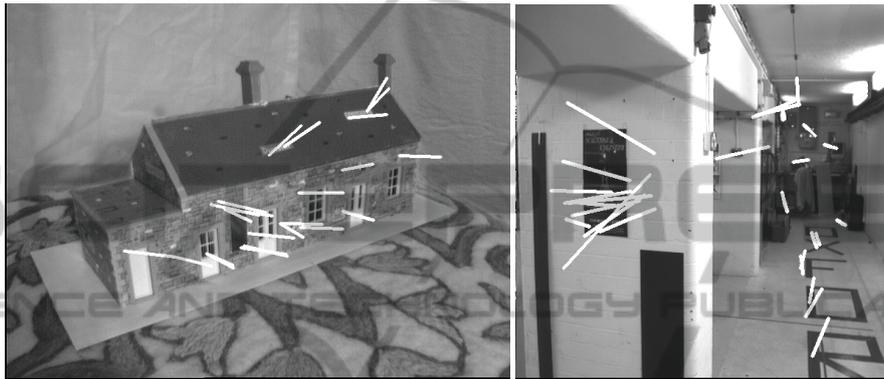
The results of the comparison is visualized in Fig. 7. The transformations are estimated from the same point correspondences. The number of corresponding points are 4, 6, and 8 as it is seen in Fig. 7. The methods are denoted by 'HOM' (normals from estimated homography) and 'AFF' (normals from affine transformation). It is interesting that AFF serves better results when the noise is low. It is true especially for the $P = 4$ case (left image in Fig. 7). It is because the homography is determined exactly by the given 4 points, while the affine transformation is overdetermined. When the number of points grows (center and right plots on Fig. 7), the normal from homography estimation becomes better than that from affine transformation since the projected coordinates are obtained via perspective projection, and homography represents the correct transformation between corresponding planes in image space.

4.2 Test on Real Image Pairs

Real Tests on Calibrated Images. The proposed optimal normal estimator has been also tested on real data. In order to use the normal estimator, the projection matrices have to be known. We have downloaded building images and reconstruction data with camera parameters from the web page of the Visual Geometry Group at Oxford University².

The Oxford data sets contain point correspondences, but we have used ASIFT method (Yu and Morel, 2011) of Yu et al. for this purpose instead of using the original data. The affine transformations for the pairs are computed as follows. (1) Two patches around the corresponding locations are cropped from

²<http://www.robots.ox.ac.uk/vgg/data/>

Figure 8: Stereo image pair with estimated normals (**Library** sequence).Figure 10: Estimated normals on sequence **House** (left) and **Corridor** (right).

the images. Their size are from 60×60 to 80×80 depending on test sequences. (2) Then the ASIFT method (Yu and Morel, 2011) is applied again for the patch pair obtaining point correspondences between patches. Estimating the affine transformation is an affine 2D registration problem based on point correspondences. It is easy to solve since the problem is linear w.r.t. affine parameters, the parameters can be obtained optimally (Björck, 1996) even if the problem is overdetermined. Remark that the affine estimator should be robust since ASIFT can give false correspondences. We used a RANSAC (Fischler and Bolles, 1981)-like algorithm to discard the outliers. The proposed optimal method is carried out on the computed affine transformations of the **Library**, **House**, and **Corridor** pairs as it is seen in Figs. 8. 10. The normals are visualized by white rods. It is evident that the quality of a normal depends on the baseline of the stereo images. The last stereo pair (**Corridor**) has shorter baseline than the other two, the accuracy of its estimated normals (on the floor and wall) are lower.

We were able to reconstruct the 3D surface of the estimated positions and corresponding normals using the Marching Cubes (APSS) filter of MeshLab³. It is visualized in Fig 9.

³www.meshlab.org



Figure 9: Reconstructed 3D model.

Normal from Real Planar Surfaces. The proposed normal vector estimator (OPT) is also tested on images of buildings as it is pictured in Fig 11. These objects mainly consist of planar walls and they can be matched by homography-based pairing methods especially when the images are rectified (Tanacs et al., 2014). Though the homography itself can be decomposed (Faugeras and Lustman, 1988) if the cameras are calibrated, and then the plane normal is obtained with the camera extrinsic parameters. However, the decomposition has ambiguity as it is discussed in (Liu, 2012) and two realistic normal vector

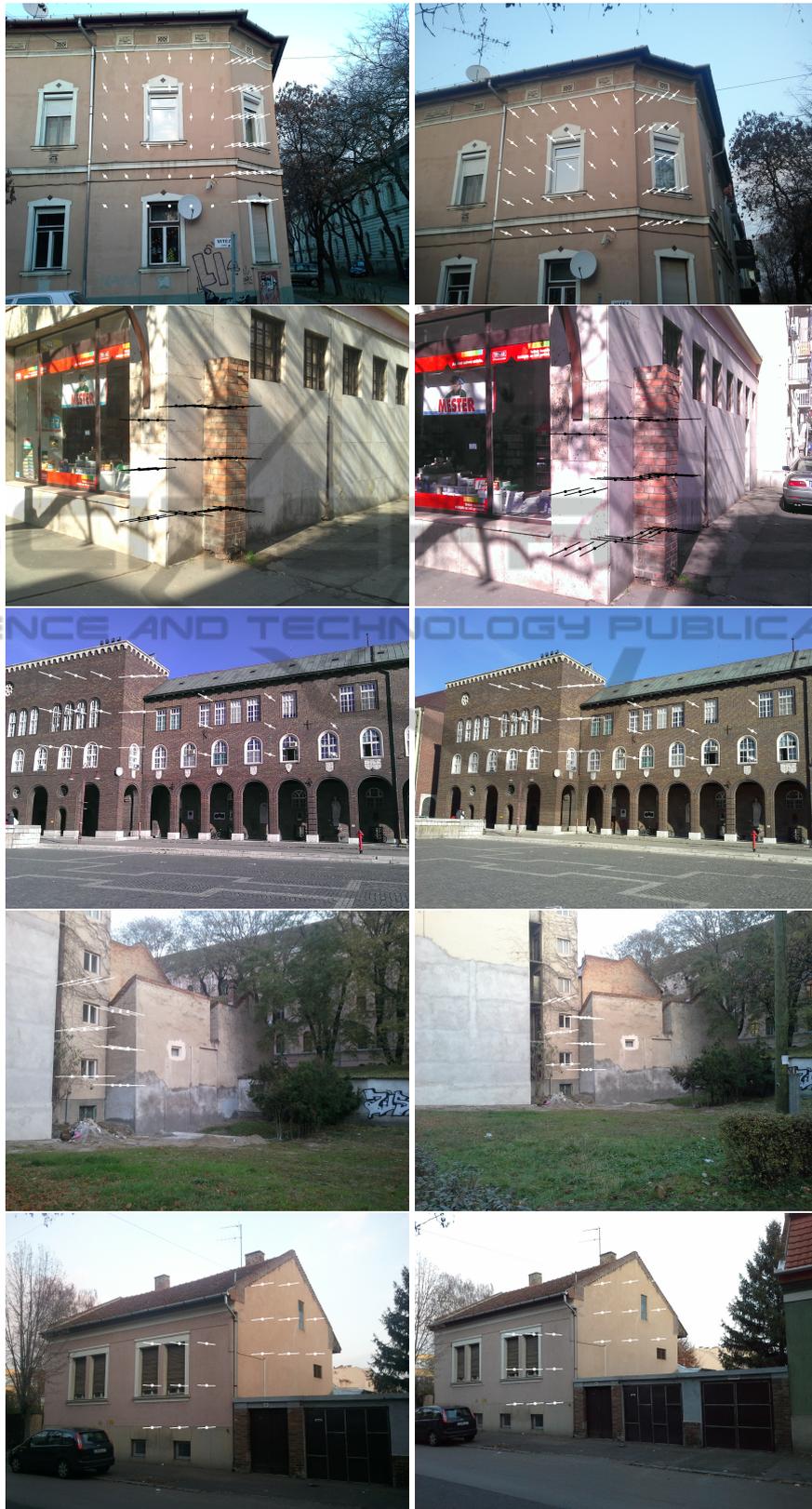


Figure 11: Estimated normals on planar surfaces.

can be achieved.

We reconstructed the plane normal via the affine transformation. The affine parameters can easily be calculated from homography as it is shown in (Molnár et al., 2014). The cameras are calibrated via point-based 3D reconstruction by bundle adjustment (B. Triggs and P. McLauchlan and R. I. Hartley and A. Fitzgibbon, 2000). Then the normals are computed by the proposed optimal method. We have tested the OPT algorithm on five different stereo pairs as it is visualized in Fig 11. They are short baseline stereo images. The yielded normal vectors and points of the planes are drawn on the input images. The corresponding points on the wall surfaces are denoted by small dots, the normals are drawn both inside and outside the plane. The proposed method is robust enough, it computes very accurately the surface normals.

5 CONCLUSION AND FUTURE WORK

Novel normal estimators have been proposed here that can estimate the normal of a surface patch if two perspective images of the patch are given and the affine transformations of the projected patches are known between the images. One of the proposed methods is optimal: if only the elements of the affine transformation are contaminated with noise, the proposed method (OPT) serves the optimal estimation in the least square sense. It can be applied if the perspective cameras are fully calibrated.

It is also obvious that normal estimation is very sensitive to the noise appearing in affine transformations. In the future, we plan to improve the affine transformation estimation in order to get more realistic results. We will also deal with developing novel reconstruction methods that use both point correspondences and estimated normals in order to obtain a more realistic 3D reconstruction of real-world 3D objects.

ACKNOWLEDGEMENT

This research was supported by the EU and the State of Hungary, co-financed by the European Social Fund through project FuturICT.hu (grant no.: TAMOP4.2.2.C11/1/KONV20120013)

REFERENCES

- B. Triggs and P. McLauchlan and R. I. Hartley and A. Fitzgibbon (2000). Bundle Adjustment – A Modern Synthesis. In Triggs, W., Zisserman, A., and Szeliski, R., editors, *Vision Algorithms: Theory and Practice*, LNCS, pages 298–375. Springer Verlag.
- Björck, Å. (1996). *Numerical Methods for Least Squares Problems*. Siam.
- Faugeras, O. and Lustman, F. (1988). Motion and structure from motion in a piecewise planar environment. Technical Report RR-0856, INRIA.
- Fischler, M. and Bolles, R. (1981). RANdom SAMpling Consensus: a paradigm for model fitting with application to image analysis and automated cartography. *Commun. Assoc. Comp. Mach.*, 24:358–367.
- Fodor, B., Kazó, C., Zsolt, J., and Hajder, L. (2014). Normal map recovery using bundle adjustment. *IET Computer Vision*, 8:66 – 75.
- Habbecke, M. and Kobbelt, L. (2006). Iterative multi-view plane fitting. In *In VMV06*, pages 73–80.
- Habbecke, M. and Kobbelt, L. (2007). A surface-growing approach to multi-view stereo reconstruction. In *CVPR*.
- Hartley, R. I. and Sturm, P. (1997). Triangulation. *Computer Vision and Image Understanding: CVIU*, 68(2):146–157.
- Hartley, R. I. and Zisserman, A. (2003). *Multiple View Geometry in Computer Vision*. Cambridge University Press.
- Kreyszig, E. (1991). *Differential geometry*. Dover Publications.
- Liu, H. (2012). *Deeper Understanding on Solution Ambiguity in Estimating 3D Motion Parameters by Homography Decomposition and its Improvement*. PhD thesis, University of Fukui.
- Malis, E. and Vargas, M. (2007). Deeper understanding of the homography decomposition for vision-based control. Technical Report RR-6303, INRIA.
- Molnár, J., Huang, R., and Kato, Z. (2014). 3d reconstruction of planar surface patches: A direct solution. *ACCV Big Data in 3D Vision Workshop*.
- Tanacs, A., Majdik, A., Molnar, J., Rai, A., and Kato, Z. (2014). Establishing correspondences between planar image patches. In *International Conference on Digital Image Computing: Techniques and Applications (DICTA)*.
- Woodham, R. J. (1978). Photometric stereo: A reflectance map technique for determining surface orientation from image intensity. In *Image Understanding Systems and Industrial Applications, Proc. SPIE*, volume 155, pages 136–143.
- Yu, G. and Morel, J.-M. (2011). ASIFT: An Algorithm for Fully Affine Invariant Comparison. *Image Processing On Line*, 2011.
- Z. Megyesi, G. and D.Chetverikov (2006). Dense 3d reconstruction from images by normal aided matching. *Machine Graphics and Vision*, 15:3–28.

APPENDIX

Algorithm for Optimal Normal Estimation. The task is to minimize the cost function defined in Eq. 10 with respect to normal vector n . The scale of the vector is arbitrary, only the direction of the normal is required. Such kind of problems are typically solved using Lagrange-multipliers, however, it cannot be applied here since the derivatives become very difficult. For this reason, we utilize another constraint for the normal: let the sum of the coordinates be 1. Thus, n is written as $n = [n_x, n_y, 1 - n_x - n_y]^T$. Eq. 10 can be reformulated as,

$$\arg \min_m \sum_{k=1}^4 \left(\frac{m^T q_k + w_{k,z}}{\alpha m^T q_5 + \alpha w_{5,z}} - a_k \right)^2,$$

where $m = [n_x, n_y]$, $q_i = [w_{i,x} - w_{i,z}, w_{i,y} - w_{i,z}]^T$. (Indices x , y , and z denote the first, second, and third coordinates of vectors, respectively.)

The minima/maxima can be obtained by taking the derivative with respect to vector m :

$$2 \sum_{k=1}^4 \beta_k \gamma_k = 0$$

where

$$\beta_k = \left(\frac{m^T q_k + w_{k,z}}{\alpha m^T q_5 + \alpha w_{5,z}} - a_k \right)$$

$$\gamma_k = \left(\alpha \frac{(m^T q_5 + w_{5,z})q_k - (m^T q_k + w_{k,z})q_5}{(\alpha m^T q_5 + \alpha w_{5,z})^2} \right)$$

After taking the lowest common multiple of the fractions, the left side should be equal to zero as $\sum_{k=1}^4 \delta_k \kappa_k = 0$, where

$$\delta_k = (m^T q_k + w_{k,z} - a_k \alpha m^T q_5 - a_k \alpha w_{5,z})$$

$$\kappa_k = ((m^T q_5 + w_{5,z})q_k - (m^T q_k + w_{k,z})q_5)$$

It can be simplified as $\sum_{k=1}^4 e_k^1 e_k^2 = 0$, where

$$e_k^1 = (m^T (q_k - a_k \alpha q_5) + (w_{k,z} - a_k \alpha w_{5,z}))$$

$$e_k^2 = ((m^T q_5)q_k - (m^T q_k)q_5 + w_{5,z}q_k - w_{k,z}q_5)$$

This is an equation with a 2D-vector:

$$\sum_{k=1}^4 r \begin{pmatrix} m^T (q_5 q_{k,x} - q_i q_{5,x}) + w_{5,z} q_{k,x} - w_{k,z} q_{5,x} \\ m^T (q_5 q_{k,y} - q_i q_{5,y}) + w_{5,z} q_{k,y} - w_{k,z} q_{5,y} \end{pmatrix} = 0$$

where $r = (m^T (q_k - a_k \alpha q_5) + (w_{k,z} - a_k \alpha w_{5,z}))$

By introducing the $m = [x, y]^T$ notation, the vector equation is modified as follows

$$\sum_{k=1}^4 (\Omega_k x + \Psi_k y + \Gamma_k) \begin{pmatrix} \Omega_k^1 x + \Psi_k^1 y + \Gamma_k^1 \\ \Omega_k^2 x + \Psi_k^2 y + \Gamma_k^2 \end{pmatrix} = 0$$

where

$$\begin{aligned} \Omega_k &= q_{k,x} - \alpha q_{5,x} a_k & \Psi_k &= q_{k,y} - \alpha q_{5,y} a_k \\ \Gamma_k &= w_{k,z} - a_k \alpha w_{5,z} & \Omega_k^1 &= 0 \\ \Psi_k^1 &= q_{5,y} q_{k,x} - q_{k,y} q_{5,x} & \Gamma_k^1 &= w_{5,z} q_{k,x} - w_{k,z} q_{5,x} \\ \Omega_k^2 &= q_{5,x} q_{k,y} - q_{k,x} q_{5,y} & \Psi_k^2 &= 0 \\ \Gamma_k^2 &= w_{5,z} q_{i,y} - w_{i,z} q_{5,y} \end{aligned}$$

The rows of the vector equation give two special quadratic curves. They are written by their implicit equations as $\sum_{k=1}^4 A_k^l x^2 + \sum_{k=1}^4 B_k^l y^2 + \sum_{k=1}^4 C_k^l xy + \sum_{k=1}^4 D_k^l x + \sum_{k=1}^4 E_k^l y + \sum_{k=1}^4 F_k^l = 0$, where $A_k^l = \Omega_k \Omega_k^l$, $B_k^l = \Psi_k \Psi_k^l$, $C_k^l = \Omega_k^l \Psi_k + \Psi_k^l \Omega_k$, $D_k^l = \Omega_k^l \Gamma_k + \Gamma_k^l \Omega_k$, $E_k^l = \Psi_k^l \Gamma_k + \Gamma_k^l \Psi_k$ and $F_k^l = \Gamma_k^l \Gamma_k$, $l \in 1, 2$. They are special because $A_k^1 = 0$ and $B_k^2 = 0$.

The solution of the optimal method described in the study (within appendix) is given by the intersection of two quadratic equations.

$$B_1 y^2 + C_1 xy + D_1 x + E_1 y + F_1 = 0$$

$$A_2 x^2 + C_2 xy + D_2 x + E_2 y + F_2 = 0$$

Parameter y can be obtained from the latter equation as

$$y = \frac{A_2 x^2 + D_2 x + F_2}{C_2 x + E_2}$$

Substituting y into the first equation the following expression is obtained

$$B_1 \left(\frac{A_2 x^2 + D_2 x + F_2}{C_2 x + E_2} \right)^2 -$$

$$C_1 x \frac{A_2 x^2 + D_2 x + F_2}{C_2 x + E_2} + D_1 x -$$

$$E_1 \frac{A_2 x^2 + D_2 x + F_2}{C_2 x + E_2} + F_1 = 0$$

If both sides are multiplied with $(C_2 x + E_2)^2$, then the equation modifies as follows

$$B_1 (A_2 x^2 + D_2 x + F_2)^2 -$$

$$C_1 x (A_2 x^2 + D_2 x + F_2) (C_2 x + E_2) +$$

$$D_1 x (C_2 x + E_2)^2 - E_1 (A_2 x^2 + D_2 x + F_2) (C_2 x + E_2)$$

$$+ F_1 (C_2 x + E_2)^2 = 0$$

This is a fourth-order polynomial where the coefficients are as follows

$$\begin{aligned} x^4 : & B_1 A_2^2 - C_1 A_2 C_2 \\ x^3 : & 2B_1 A_2 D_2 - C_1 A_2 E_2 - \\ & C_1 D_2 C_2 + D_1 C_2^2 - E_1 A_2 C_2 \\ & B_1 D_2^2 + 2B_1 A_2 F_2 - \\ x^2 : & C_1 D_2 E_2 - C_1 F_2 C_2 + 2D_1 C_2 E_2 - \\ & E_1 A_2 E_2 - E_1 D_2 C_2 + F_1 C_2^2 \\ & 2B_1 D_2 F_2 - C_1 F_2 E_2 + \\ x^1 : & D_1 E_2^2 - E_1 D_2 E_2 - E_1 F_2 C_2 + \\ & 2F_1 C_2 E_2 \\ x^0 : & B_1 F_2^2 - E_1 F_2 E_2 + F_1 E_2^2 \end{aligned}$$

Remark that the equation $C_2x + E_2 = 0$ can also be considered. (In this case the first equation is independent from y .)

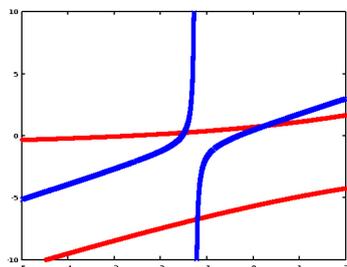


Figure 12: Quadratic curves.

An example for two quadratic curves with three real intersections is visualized in Fig. 12. (The parameters of curves are $B1 = -1.9055$, $C1 = 2.2632$, $D1 = 2.8577$, $E1 = -9.4392$, $F1 = 7.7081$, and $A2 = -2.2632$, $C2 = 1.9055$, $D2 = -4.2074$, $E2 = 2.3903$, $F2 = -1.1190$.)

