

Estimating the Fundamental Matrix Without Point Correspondences With Application to Transmission Imaging

Tobias Würfl¹, André Aichert², Nicole Maaß², Frank Dennerlein², Andreas Maier¹
¹Pattern Recognition Lab, Friedrich-Alexander-Universität Erlangen-Nürnberg (FAU)
²Siemens Healthcare GmbH

<https://www5.cs.fau.de>

Abstract

We present a general method to estimate the fundamental matrix from a pair of images under perspective projection without the need for image point correspondences. Our method is particularly well-suited for transmission imaging, where state-of-the-art feature detection and matching approaches generally do not perform well. Estimation of the fundamental matrix plays a central role in auto-calibration methods for reflection imaging. Such methods are currently not applicable to transmission imaging. Furthermore, our method extends an existing technique proposed for reflection imaging which potentially avoids the outlier-prone feature matching step from an orthographic projection model to a perspective model. Our method exploits the idea that under a linear attenuation model line integrals along corresponding epipolar lines are equal if we compute their derivatives in orthogonal direction to their common epipolar plane. We use the fundamental matrix to parametrize this equality. Our method estimates the matrix by formulating a non-convex optimization problem, minimizing an error in our measurement of this equality. We believe this technique will enable the application of the large body of work on image-based camera pose estimation to transmission imaging leading to more accurate and more general motion compensation and auto-calibration algorithms, particularly in medical X-ray and Computed Tomography imaging.

1. Introduction

Transmission imaging modalities are very popular in technical, scientific and medical applications. They measure attenuation of mechanical or electromagnetic waves instead of light reflected from surfaces. Examples include X-ray imaging, speed-of-sound ultrasound imaging [31, 32] and electron-microscopy. Fan *et al.* [10] illustrate the differences and similarities between transmission and reflection

imaging. Most importantly, they show that transmission images are continuous functions as opposed to the widely used discontinuous models for reflection images. This is a direct consequence of the image formation process which superimposes all objects of the three dimensional data along the projection rays. While feature detection and matching techniques suffer along occluding edges in reflection imaging, transmission images are still harder to analyze because objects are always superimposed. Two-view reconstruction is generally impossible in transmission imaging, however reconstruction of the 3D scene is routinely performed from a continuous camera trajectory to disentangle the objects. In X-ray imaging this is known as Computed Tomography (CT). In contrast to reflection imaging the resulting scene reconstructions are dense instead of mere surface models and require considerably more than two images at the minimum.

Most research in these modalities concentrates around solving the scene reconstruction problem while assuming the scene structure is calibrated offline [17, 27, 9]. However, in many scenarios this offline calibration falls short. E.g. in medical imaging movement of the patient often needs to be compensated [30]. Another difficult scenario are new generations of flexible X-ray systems on robotic arms which can perform arbitrary scan trajectories and therefore currently require a high calibration effort [34] and cannot utilize their full flexibility [22]. This estimation of the three-dimensional configuration of the imaging system has been studied extensively in computer vision. However, most current techniques naturally assume an estimation of corresponding points in images is generally possible. This works well in reflection imaging when the visual appearance of certain surface points is distinct and consistent across images. It is well-known that glossy and transparent surfaces pose problems to most algorithms. Matters are even more complicated in transmission imaging for three reasons. First, most systems do not distinguish between different energy spectra. Second, the resulting values only reflect the accumulated attenuation from passing through

the whole object or even multiple objects. A feature in a 2D projection image is usually not associated with a single point in the 3D object. Third, the appearance of those features varies completely between viewpoints as the view rays intersect entirely different parts of the object or objects. Therefore, both feature detection and feature matching techniques mostly fail in transmission imaging [18].

2. Related Work

We aim towards recovering the relative geometry of cameras based on transmission image data alone. We first present related work on reflection imaging.

In their seminal paper Luong and Faugeras introduced the fundamental matrix [24] \mathbf{F} and algorithms to estimate it. This matrix describes the relative geometry of a pair of cameras with a minimal number of parameters. At its core is the fundamental constraint of corresponding image points

$$\mathbf{x}_2^T \mathbf{F} \mathbf{x}_1, \quad (1)$$

where \mathbf{x}_1 and \mathbf{x}_2 are projections of the same 3D point. It has many applications in reconstruction of the 3D geometry of a scene like eliminating false positive corresponding points, speeding up the matching of points or reconstruction of the scene. Luong and Faugeras present multiple algorithms to estimate this matrix from point correspondences. However, feature matching, particularly in wide-baseline stereo, is an outlier-prone problem. Mostly this challenge is addressed by using robust probabilistic consensus-based matching methods like the RANSAC algorithm.

A different strategy is to bypass the computation of corresponding points and estimate \mathbf{F} directly without corresponding points. Recently, Omid *et al.* [28] presented such a method based on a deep neural network for reflection imaging. The network is presented with a pair of images and trained to output the corresponding fundamental Matrix \mathbf{F} . The downsides of such an implicit modelling approach are the absence of any guarantees for the performance of the algorithm on unseen data and the lack of interpretability of the learned algorithm. These problems make it challenging to deploy such an algorithm in safety critical applications.

An approach using classic vision methodology relies on using contours of objects [33], [35], [19]. A disadvantage of these methods is that they don't take all the intensity levels into account and rely on a binary segmentation instead.

Another approach was presented by Lehmann *et al.* [21]. In their work they propose to use a line-integral orthographic projection model for reflection imaging and use the projection-slice (Fourier-slice) theorem to formulate a cost-function for the parameters of the fundamental matrix. In their subsequent work [20] this approach was simplified by defining an intermediate function using the Radon transform \mathcal{R}

$$\phi_i(s, \theta) = \mathcal{R} \{q_i(u, v)\}, \quad (2)$$

where $q_i(u, v)$ denotes an image with index i . The Radon transform in 2D effectively represents the set of all possible line integrals over an image and is a continuous equivalent of the Hough-transform [7]. Because under the orthographic projection model all epipolar lines are parallel they find that

$$\phi_i(s, \theta_i) = \phi_j(s + d, \theta_j), \quad (3)$$

for a pair θ_i and θ_j which denote angles of corresponding epipolar lines and an offset d between these lines. This allows estimating the fundamental matrix by minimizing a cost-function of the difference between ρ_i and ρ_j with respect to θ_1, θ_2 and d . In their work they use maximization of the normalized cross-correlation as loss-function. They also show how to interpret the line-integral projection model as probabilistic model for reflection imaging. However, the line-integral image is much more natural for transmission imaging. The standard model for this is a linear attenuation law

$$I_i(u, v) = I_0(u, v) e^{-\int_0^\infty \mu(\beta_i + t\alpha(u, v)) dt} \quad (4)$$

where the line of integration is parametrized as $\beta_i + t\alpha$. Here β_i denotes the center of projection for an image indexed by i and $\alpha(u, v)$ is the ray direction depending on the projection coordinates u and v , while $\mu(\beta_i + t\alpha)$ denotes the spatial distribution of the attenuation values along a line defined by β_i and α . To densely reconstruct the linear attenuation coefficient $\mu(\beta_i + t\alpha)$ we rewrite Eq. 4 as:

$$q_i(u, v) = -\ln \frac{I_i(u, v)}{I_0(u, v)} = \int_0^\infty \mu(\beta_i + t\alpha(u, v)) dt. \quad (5)$$

This shows that transmission imaging naturally follows the line integral model proposed by Lehmann *et al.* without probabilistic interpretation. Still, the orthographic projection model is not applicable for most such imaging systems. The key problem for extending this method to perspective projection is the fact that Eq. 3 does not hold under this model. To extend the method a link between line integrals of a perspective projection in two views is necessary.

In the CT community such a link is known. By observing that line integrals over line integral images are equivalent to plane integrals we can see that Eq. 3 are just two measurements of the same plane-integral. Grangeat *et al.* [14] showed that in perspective geometry *derivatives* of line integrals over epipolar lines in orthogonal direction to their common epipolar plane are equivalent. This is known as a consistency condition in the CT community. Debbeler *et al.* [8] first used this to formulate such a consistency condition to estimate calibration parameters of a CT system. They also use a Radon transform similar to Eq. 3:

$$\rho_i(s, \theta) = \mathcal{R} \{\zeta_i(u, v) q_i(u, v)\}, \quad (6)$$

where $\zeta = \frac{\alpha(u, v)^T \bar{\alpha}_i}{\|\alpha(u, v)\| \|\bar{\alpha}_i\|}$ denotes the cosine between the central ray associated with this image: $\bar{\alpha}_i$ and the vec-

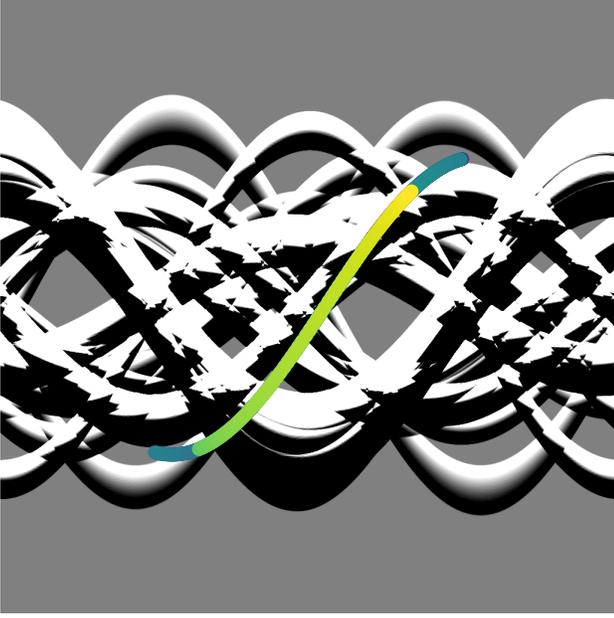


Figure 1: The intermediate function $\frac{\partial}{\partial s} \rho_i(s, \theta)$ for image $q_i(u, v)$. We also show the coordinates of the epipolar lines calculated by $\xi(\mathbf{F}, n)$ for different values of n .

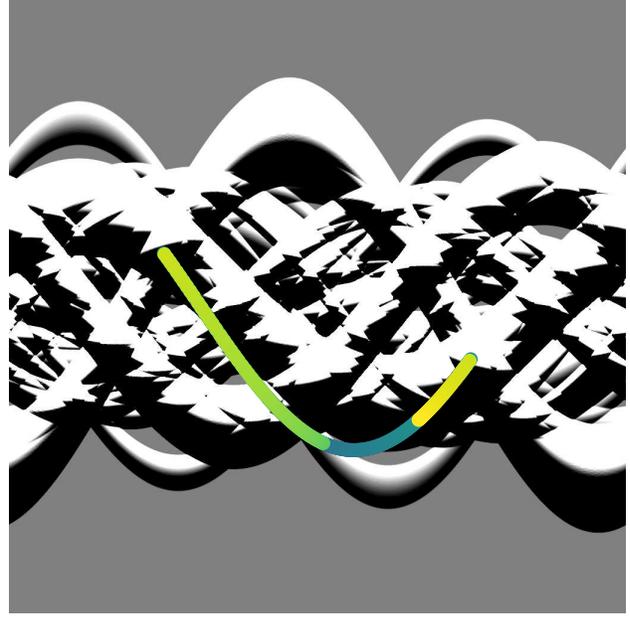


Figure 2: The intermediate function $\frac{\partial}{\partial s} \rho_j(s, \theta)$ for image $q_j(u, v)$. We also show the coordinates of the epipolar lines calculated by $\xi(\mathbf{F}, n)$ for different values of n .

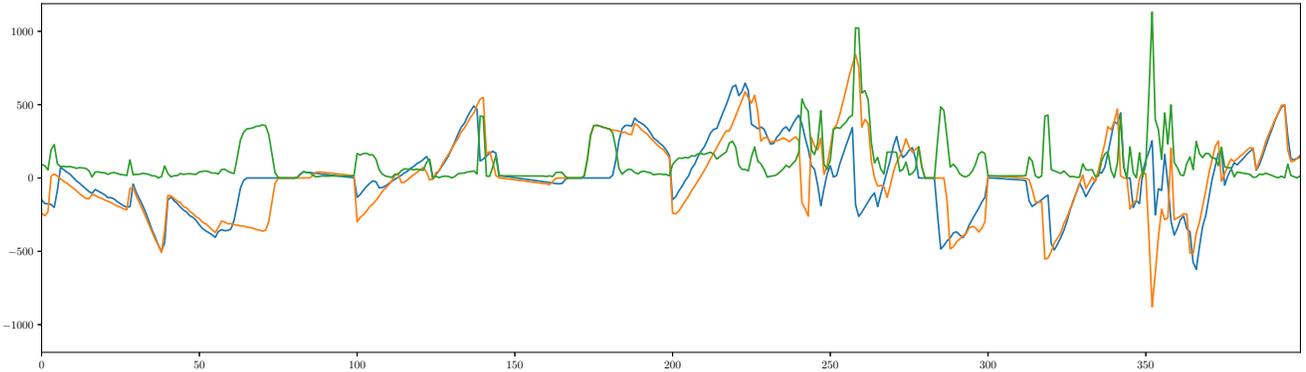


Figure 3: Values of the intermediate functions shown in Fig. 1 and Fig. 2 along the path of epipolar lines defined by $\xi(\mathbf{F}, n)$ denoted by the blue and orange lines. Element-wise difference between those two paths depicted in green. The proposed optimization problem minimizes a metric over the green line.

tor connecting the center of projection with the image coordinates u, v : $\alpha(u, v)$. Their intermediate function is now formed by applying a partial derivative in s direction: $\frac{\partial}{\partial s} \rho_i(s, \theta)$. Examples of such intermediate functions on two images are given in Fig. 1 and Fig. 2. However, corresponding epipolar lines are no longer parallel, so in order to extend Eq. 3, coordinates of corresponding epipolar lines need to be calculated. A visualization of those corresponding coordinates can also be found as the sinusoid curve in Fig. 1 and Fig. 2. Aichert et al. [2] proposed to use the projection matrices \mathbf{P}_i and \mathbf{P}_j of images $q_i(u, v)$ and $q_j(u, v)$ to

define a function:

$$\xi(\mathbf{P}_i, \mathbf{P}_j, \kappa) = (s_{(i, \kappa)}, \theta_{(i, \kappa)}, s_{(j, \kappa)}, \theta_{(j, \kappa)})^T, \quad (7)$$

which outputs coordinates $s_{(i, \kappa)}, \theta_{(i, \kappa)}$ and $s_{(j, \kappa)}, \theta_{(j, \kappa)}$ of corresponding epipolar lines in the intermediate function. The parameter κ can be interpreted as a rotation angle of the common epipolar plane around the baseline joining the two centers of projection. This allows to formulate the condition as:

$$\frac{\partial}{\partial s} \rho_i(s_{(i, \kappa)}, \theta_{(i, \kappa)}) = \frac{\partial}{\partial s} \rho_j(s_{(j, \kappa)}, \theta_{(j, \kappa)}), \quad (8)$$

where the coordinates are determined by the function

$\xi(\mathbf{P}_i, \mathbf{P}_j, \kappa)$. In conclusion Eq. 8 generalizes Eq. 3 and is already known for transmission imaging but new for reflection imaging.

Eq. 8 is now known as Grangeat consistency condition in the CT community and together with a newer consistency condition based on rectification they are known as epipolar consistency conditions. The relation between those types of epipolar consistency conditions have been shown by Lesaint *et al.* [23]. An application of this to image-based tracking of objects was presented by Aichert *et al.* [3]. The Grangeat consistency condition was also used for motion compensation by Fryschtal *et al.* [12] and image-based calibration [2, 8]. However, in those applications a model of an idealized trajectory was optimized instead of estimating the geometry generally.

The condition has also been used to transfer reflection imaging algorithms for estimation of symmetry planes to transmission imaging by Preuhs *et al.* [29] and to compensate for non-linear attenuation [36, 1].

3. Estimating the Fundamental Matrix without Point Correspondences

To formulate estimation of the fundamental matrix between two images using Eq. 8 we define a new parametrization $\xi(\mathbf{F}, n)$ instead of $\xi(\mathbf{P}_i, \mathbf{P}_j, \kappa)$. We propose to sample N points \mathbf{p}_n from the coordinates inside image $q_i(u, v)$. The points can form an arbitrary sampling pattern on the image. As no obvious sampling pattern can be derived without assumptions about the relative geometry we simply use a pattern spanning all edge pixels of the image. This will guarantee a certain amount of well spread epipolar lines regardless of the camera orientations. From the fundamental matrix we extract the epipole \mathbf{e}_i as the right null space of \mathbf{F} by solving $\mathbf{F}\mathbf{e}_i = \mathbf{0}$. A bundle of epipolar lines in image $q_i(u, v)$ is now given as

$$\mathbf{l}_{(i,\kappa)} = \mathbf{e}_i \times \mathbf{p}_n. \quad (9)$$

We transfer this bundle of epipolar lines to the other image $q_j(u, v)$ using the epipolar line transfer [16]

$$\mathbf{L} = \mathbf{F}[\mathbf{e}_i]_{\times} \quad (10)$$

where $[\mathbf{e}_i]_{\times}$ denotes forming a skew symmetric matrix. The epipolar line transfer allows us to form

$$\mathbf{l}_{(j,\kappa)} = \mathbf{L}\mathbf{l}_{(i,n)} \quad (11)$$

which is the corresponding line bundle in the second image. Using this formulation of the intermediate function the orientation of the corresponding lines matters. The function $\phi_i(s, \theta)$ is an even function denoting $\phi_i(s, \theta) = \phi_i(-s, -\theta)$ because of the Radon transform. However $\frac{\partial}{\partial s}\rho_i(s, \theta)$ is an

odd function denoting $-\frac{\partial}{\partial s}\rho_i(s, \theta) = \frac{\partial}{\partial s}\rho_i(-s, -\theta)$ because of the derivative operator. In order to respect this antisymmetry we calculate the orientation of all corresponding lines and enforce that they point towards the same half-space. The final step in this algorithm is the conversion from corresponding lines $\mathbf{l}_{(i,n)}, \mathbf{l}_{(j,n)}$ to coordinates $(s_{(i,n)}, \theta_{(i,n)}, s_{(j,n)}, \theta_{(j,n)})^T$ in the intermediate function. We can represent this algorithm to calculate corresponding epipolar lines as

$$\xi(\mathbf{F}, n) = (s_{(i,n)}, \theta_{(i,n)}, s_{(j,n)}, \theta_{(j,n)})^T \quad (12)$$

parametrized using only the fundamental matrix.

Using this parametrization of epipolar lines we form a cost function:

$$c(\mathbf{F}, \rho_i, \rho_j) = d(\mathbf{v}_i, \mathbf{v}_j) \quad (13)$$

where d is a distance measure between the vectors \mathbf{v}_i and \mathbf{v}_j sampled at the N corresponding coordinates from the intermediate function. Those corresponding vectors \mathbf{v} can be seen in Fig. 3 as pair of blue and orange lines. A valid choice for d is the squared euclidean norm of the difference $\|\mathbf{v}_i - \mathbf{v}_j\|_2^2$.

A potential issue with this cost function is of numerical nature. The costs from Eq. 13 scale with the numerical values of the image, i.e. epipolar lines through empty regions of the image have an associated cost of zero regardless of how well their geometry is represented. The problem is similar to missing corresponding points on a featureless surface. This can introduce a bias in this cost function to favour solutions which produce many epipolar lines in empty regions because those errors are smaller than the numerical errors on the path of the correct solutions. To reduce this problem we modify the cost function to incorporate a weight function $w(\mathbf{v}_i, \mathbf{v}_j)$ reducing the influence of empty regions:

$$\hat{d}(\mathbf{v}_i, \mathbf{v}_j) = \frac{d(\mathbf{v}_i, \mathbf{v}_j)}{w(\mathbf{v}_i, \mathbf{v}_j)}. \quad (14)$$

We define the weight function as

$$w(\mathbf{v}_i, \mathbf{v}_j) = \sum_n \frac{(v_{i,n} + v_{j,n})^2}{(v_{i,n} + v_{j,n})^2 + \epsilon}. \quad (15)$$

This function is inspired by the Geman-McClure cost function [13]. This function realizes a drop to zero for values with an absolute value smaller than ϵ while continuously increasing towards one for larger values. In effect this weight increases costs if mostly small values are sampled from the intermediate function remedying the numeric problem.

Another problem with this cost function is that it is in general not symmetric because the sampling of points on the different images might produce widely different samplings of epipolar lines. This problem also appears in traditional

methods to estimate fundamental matrices and a common way to solve it is by symmetrizing the cost function as follows:

$$\hat{c}(\mathbf{F}, \rho_i, \rho_j) = c(\mathbf{F}, \rho_i, \rho_j) + c(\mathbf{F}^T, \rho_j, \rho_i). \quad (16)$$

Estimation of the fundamental matrix can now be posed as an optimization problem:

$$\min_{\mathbf{F}} \{ \hat{c}(\mathbf{F}, \rho_i, \rho_j) \}. \quad (17)$$

This problem is in general a non-convex optimization problem because the intermediate functions ρ_i, ρ_j directly depend on the data and therefore local minima and ambiguities can exist. Because of the non-convex nature of the problem initialization of \mathbf{F} is important, which is a fundamental difference to algebraic methods based on matched points. A way to do so is to calculate it from projection matrices if an initial estimate of those is available as $F = [e_2]_{\times} \mathbf{P}_2 \mathbf{P}_1^+$, where \mathbf{P}_1^+ denotes the pseudo-inverse of \mathbf{P}_1 according to Hartley and Zisserman [16]. We found the Simplex method of Nelder and Mead [26] effective in solving our optimization problem.

Our optimization problem does not impose the known rank two condition on the fundamental matrix. Therefore, we project our result to the manifold of rank two matrices by factorizing it as $\mathbf{F} = \mathbf{U} \Sigma \mathbf{V}^T$ using the singular value decomposition and setting the smallest singular value contained in Σ to zero before restoring it.

4. Experiments

To validate our method we simulated an X-ray CT dataset consisting of 18 homogeneous beads of 4 different densities which were multiples of the density of water and varying radii from $1mm$ to $3mm$. The beads were distributed in a volume of $140mm^3$. From this volume we acquired 512×512 pixels resolved square line integral images with $2mm$ spacing from random centers of projection on a $200mm$ sphere with a $400mm$ focal length. Figure 4 depicts this acquisition geometry graphically. While the sphere is not a practical trajectory for CT imaging we intend to validate the generality of our method using this setup because this will imply it also works with more restricted continuous trajectories. We are interested in the estimation of relative geometry for each pair of images in any general position.

From this dataset we sample 100 pairs of images randomly and recover the fundamental matrix from this geometry after corrupting the groundtruth geometry. Geometric jitter from mechanical motion was simulated by adding a random offset of the principal point of up to 20 pixels in u and v direction separately to the principal point of a pair of projection matrices. We initialized the fundamental matrix from these corrupted projection matrices and recover

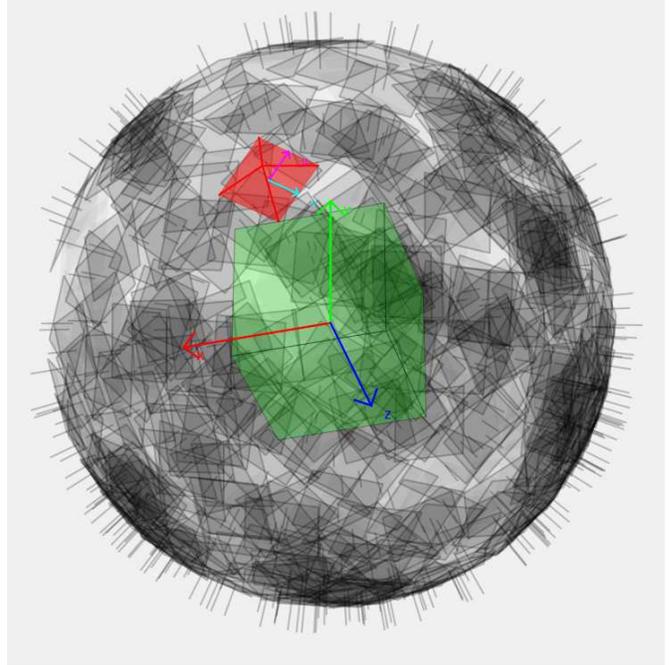


Figure 4: The acquisition geometry of the simulated dataset.

the groundtruth matrix using our proposed method and reference method using a classical algorithm. Note that while our corruption model does not cover all parameters we still verify that our algorithm returns all parameters of the fundamental matrix correctly.

We perform three experiments on this dataset. In the first experiment we compare the proposed method against a feature-matching-based reference method to show its advantages and verify its effectiveness. In a second experiment we validate the effectiveness of our proposed method to deal with the numeric effect favouring configurations where the epipolar planes pass through empty parts of the volume. We do this by performing an ablation study where we compare setting our parameter ϵ to 0 compared to the default choice of $1E-3$ which we use in all other experiments. Finally we compare the performance of different distance measures d on our results. We chose to evaluate different L_p norms with p equal to 3 and $3/2$ in comparison to the more standard choice of L_2^2 . The reason why we chose these norms is that $p = 3$ punishes outliers more compared to L_2^2 while $p = 3/2$ punishes them less and can therefore be more robust to outliers. We refrained from using the popular $p = 1$ norm to stay within the class of smooth optimization problems.

The reference method was implemented using the OpenCV library [5]. As feature descriptors we used KAZE features [11] and the fast approximate nearest neighbour matching algorithm of Muja and Lowe [25] as well as OpenCVs implementation of the fundamental matrix esti-

	Error L_2 -norm	Error epipolar-error
proposed	5.45E-03 ± 2.54E-03	2.62E-02 ± 5.57E-03
opencv	1.05E-01 ± 2.02E-02	6.48E-01 ± 2.53E-02
jittered	2.11E-02 ± 1.20E-02	4.21E-02 ± 6.23E-03

Table 1: Results for the reference method and the proposed method with the error of the initial geometry as reference. The proposed method is able to estimate the geometry accurately while the classical method performs very weak.

mation using the least median of squares algorithm. We chose KAZE features over the widely used SIFT features because there is no usage restriction on them.

For evaluation we use two quantitative measures in our experiments. First we report the L_2 -norm (Frobenius-norm) between the estimated fundamental matrix and the groundtruth matrix calculated from the uncorrupted projection matrices. Second we report a relative error on the estimated epipoles as used by Luong and Faugeras [24]. They define this epipolar-error as

$$\min \left\{ \frac{|x - x_0|}{\min(|x|, |x_0|)}, 1 \right\} \quad (18)$$

where x denotes a dimension of an estimated epipole and x_0 the corresponding value of the groundtruth epipole. In accordance to them, we calculate the mean of this error over each of the two dimensions and each of the pairs of epipoles. We report the mean and the standard error $SE = \frac{\sigma}{\sqrt{N}}$ of these measures with N being the number of random samples from the trajectory.

5. Results

We first present the results of the experiment comparing the classical method to our proposed method. For reference we also added the error measures which result from our simulation. The results are given in Tab. 1.

The proposed method performs best and is able to estimate the Fundamental matrix well. The results from the reference method are very weak in all cases recovering a completely wrong fundamental matrix. When inspecting a sample of the images from our results in Fig. 5 we can conclude that this is caused by the complete failure of the feature matching step. This is well in line with previous results from Klppel *et al.* [18]. The numeric results support this very well with a low error in the Frobenius distance of the estimated fundamental matrix to the groundtruth and with a significantly reduced epipolar error measure.

The results of the ablation study evaluating the effectiveness of weighting down very small contributions to the cost-function are given in Tab. 2. The Frobenius distance is improved drastically with the residual error roughly just a third of the error without the downweighting, while also the

	Error L_2 -norm	Error epipolar-error
proposed	5.45E-03 ± 2.54E-03	2.62E-02 ± 5.57E-03
$L_{2,\epsilon=0}^2$	1.53E-02 ± 1.11E-02	2.83E-02 ± 6.38E-03

Table 2: Results of our ablation study about the effectiveness of the weighting. The version with the active weighting shows improved numeric results.

	Error L_2 -norm	Error epipolar-error
L_3	1.39E-02 ± 9.95E-03	3.72E-02 ± 6.89E-03
$L_{3/2}$	1.25E-02 ± 8.88E-03	2.94E-02 ± 5.57E-03
proposed	5.45E-03 ± 2.54E-03	2.62E-02 ± 5.57E-03

Table 3: Results for using different distances measures d with our proposed method. The classical mean squared error metric performed better than the considered alternatives.

standard error is much lower. The error measure considering epipolar points shows a similar trend but a less drastic numeric improvement.

For our third experiment comparing different choices of distance measure d we present numeric results in Tab. 3. The result measures agree again and are best in the case of the simple mean squared error norm with a significant drop in the Frobenius distance almost halving the error, supported by a decrease in standard error.

6. Discussion

Our experiments show that the feature-based approach performs poorly numerically on our simulated dataset. The reason for this becomes apparent when considering Fig. 5. The matched points do not constitute true correspondences in an overwhelming majority. This means there is no opportunity for a consensus-based algorithm to find a correct subset. This result is general for arbitrary transmission images and any choice of classical feature descriptor because pixel values in transmission images convey different and much less distinctive information. Even if corners or similar features are available in the imaged objects, their appearance is e.g. a sharp point of high attenuation in one image and a line on a ramp of values under another configuration of the camera. This variation can not be captured by any classical descriptor. The feature-matching step of the reference method does not make use of the initial estimate of the fundamental matrix which is used in our proposed method. However, this is mostly for illustrative purposes because the feature matching will still fail but result in reporting no point correspondences at all if the initial guess is used to reject wrong matches.

The experiments have also established that our up-weighting method for small contributions to the cost function based on the Geman-McClure cost function is benefi-

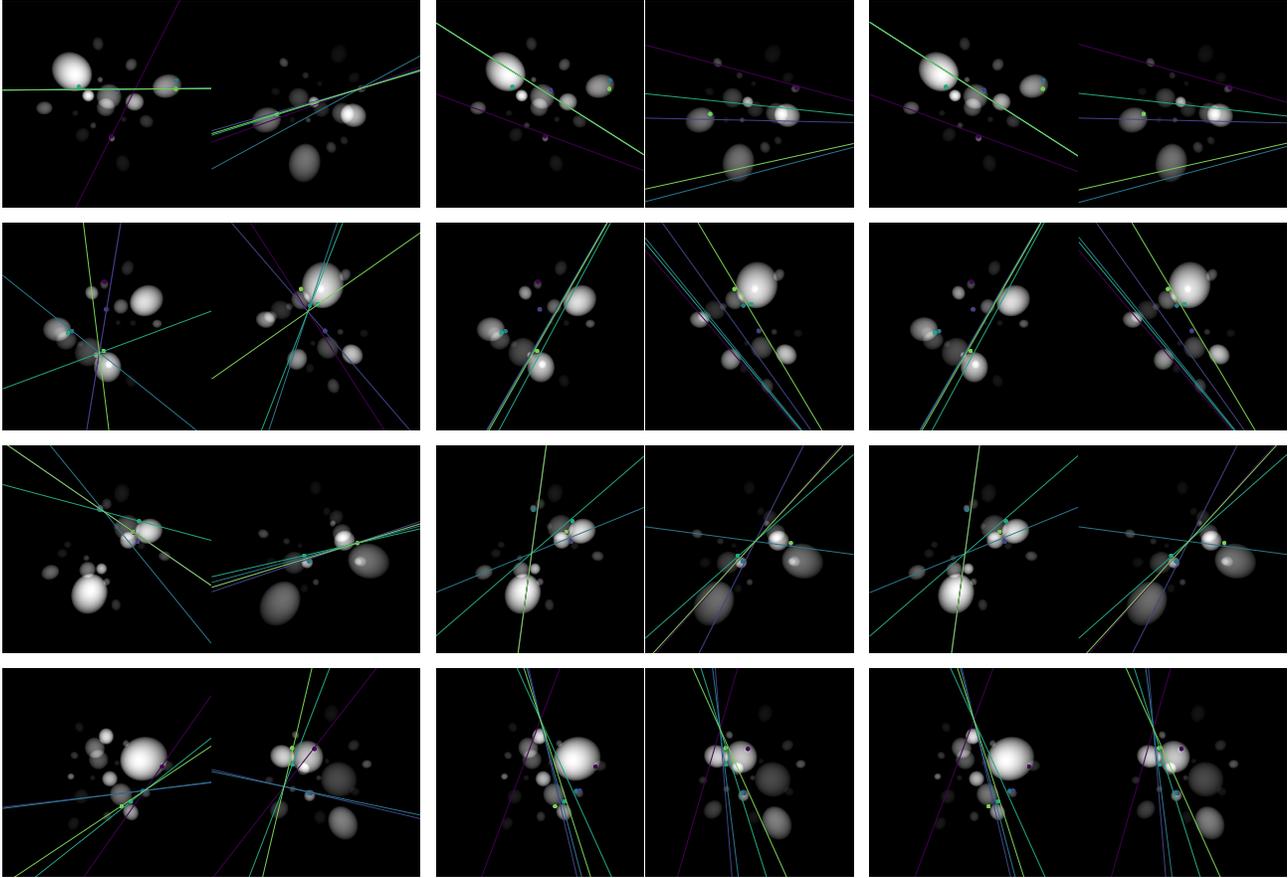


Figure 5: Extracted points using a descriptor and superimposed epipolar lines from different fundamental matrices. From left to right source of fundamental matrix: Reference method, proposed method, groundtruth. The reference method is clearly unable to deal with the featureless transmission imaging data. The proposed method matches the groundtruth well visually. Note that because the points are no valid correspondences they do not lie on the epipolar lines produced by their matched points in the other image for correct fundamental matrices.

cial for the considered object. For arbitrary objects we believe its advantage is even higher because larger areas without attenuating material often exist in the images. The parameter ϵ should than be chosen based on a validation set of independent images for applications. The experiments on different distances yielded surprising results because we expected the $L_{3/2}$ norm to improve our results by weighting down the contribution of outliers. Instead the mean squared error norm performed best. We believe for certain applications to transmission imaging modalities other norms could still outperform it.

The proposed method involves a non-convex optimization problem because the objective depends on the object data which is not restricted in any meaningful way. This seems like a major downside at first sight when compared to the classical algorithm. But despite this fact we found that optimization is well posed if the imaged scene provides information in terms of varying line integrals. This situation

is similar to rigid registration algorithms which also involve non-convex optimization but are known to be solvable stable and repeatable for most objects. In contrast, the convexity of the optimization problem of the classical algorithm is only achieved because the matching points are determined in advance and therefore the problem no longer depends on the data. However, if one considers the complete algorithm, the case where not enough corresponding points can be established unambiguously is similar to an ambiguous non-convex objective for our proposed algorithm. To sum this up we first conjecture that the non-convex optimization problem will mostly be well posed for objects providing enough information. Secone we found that the nature of that information is different from the nature of information required for good performance in the classical algorithm. Our experiment can serve as a showcase for this.

Because different concrete transmission imaging modalities will provide slightly different real world challenges

our experiment was conducted on simulated data. Our object was designed to highlight the benefits of our approach. When applying our method to measured data we expect it will be beneficial to take deviations of the linear attenuation model into account. These non-linear effects will vary with the transmission imaging modality at hand. E.g. speed-of-sound ultrasound will show different effects than X-ray imaging. However, while certainly affecting our method we believe that it can handle those deviations well without modifications, as the linear attenuation model is well tested in practice by forming the basis of dense 3D reconstruction methods in these applications.

While the fundamental matrix itself encodes the relative geometry of two images it does not suffice to recover the two projection matrices associated with the scene if the internal camera parameters are unknown. Therefore this work represents the presentation of a technical method which enables many concrete applications in the future. An immediate application is the improvement on patient-model-based anatomical landmark estimation techniques in medical X-ray images as presented by Bier *et al.* [4]. Our method could be used to verify matching detections of anatomical landmarks from their algorithm with a complementary source of information using the optimal triangulation algorithm by Hartley and Sturm [15]. Another immediate application is the method of Carrasco and Mery [6] which uses epipolar geometry to verify defects in aluminium cast wheels by matching detections from multiple frames. Our method could improve this because it draws information from the whole geometry of the aluminium cast wheel instead of being based on the detected landmarks itself. Another application is the estimation of the geometry of computed tomography scans with known intrinsic parameters like stiff C-arm CT systems. In these cases essential matrices can be computed from the fundamental matrices and decomposed into the relative geometry.

7. Conclusions and Outlook

We have presented a new algorithm to estimate the fundamental matrix between two images without using point correspondences. It extends a method previously proposed for reflection imaging from an orthographic to a perspective projection model. However, we show that applying our method to transmission imaging is more natural since it can be interpreted as being based on a solid physical model of the imaging process. In addition we showed that the advantage of our method of not relying on corresponding points is even more important in transmission imaging.

In a simulated experiment we verified the effectiveness of all components of our algorithm and its advantage over a classical reference method. We found our algorithm is accurate and reliable based on a reasonable initial guess of the geometry.

In future work it would be interesting to explore different applications of using fundamental matrices in transmission imaging. Additionally it will be interesting to test our method on different concrete modalities. We expect our algorithm to work on all those modalities which can be modelled by the linear attenuation law. Still they may permit additional physical constraints to improve over our general method. Additionally the method could be applied to reflection imaging as proposed by Lehmann *et al.* [21] in scenarios where feature matching is complicated.

We believe the most promising kind of future work based on our method is the development of general autocalibration methods for transmission imaging from a minimal number of images. Right now most work in that direction operates in very constrained settings like assuming a certain trajectory of the imaging system. In the future we think more flexible acquisition techniques for transmission imaging like robotic CT scanners would benefit greatly from fully image-based algorithms to recover the 3D geometry for subsequent scene reconstruction similar to the available algorithms for reflection imaging. This would enable them to acquire datasets on flexible trajectories which are defined on the fly to ideally adapt to the objects at hand. Currently this is very challenging because it requires extremely accurate odometry from the robot system. For this we want to explore in future work how to adapt autocalibration and structure from motion methods for reflection imaging to transmission imaging building on this work.

Disclaimer

The concepts and information presented in this paper are based on research and are not commercially available.

References

- [1] Shiras Abdurahman, Robert Frysch, Richard Bismark, Stefan Melnik, Oliver Beuing, and Georg Rose. Beam hardening correction using cone beam consistency conditions. *IEEE Transactions on Medical Imaging*, 2018. 4
- [2] André Aichert, Martin Berger, Jian Wang, Nicole Maaß, Arnd Doerfler, Joachim Hornegger, and Andreas K. Maier. Epipolar consistency in transmission imaging. *IEEE Transactions on Medical Imaging*, 34(11):2205–2219, 2015. 3, 4
- [3] André Aichert, Jian Wang, Roman Schaffert, Arnd Dörfler, Joachim Hornegger, and Andreas K. Maier. Epipolar consistency in fluoroscopy for image-based tracking. In *BMVC*, pages 82–1, 2015. 4
- [4] Bastian Bier, Mathias Unberath, Jan-Nico Zaech, Javad Fotouhi, Mehran Armand, Greg Osgood, Nassir Navab, and Andreas K. Maier. X-ray-transform Invariant Anatomical Landmark Detection for Pelvic Trauma Surgery. In Springer, editor, *Medical Image Computing and Computer-Assisted Intervention*, pages 55–63, 2018. 8

- [5] Gary Bradski. The OpenCV Library. *Dr. Dobb's Journal of Software Tools*, 2000. 5
- [6] Miguel Carrasco and Domingo Mery. Automatic multiple view inspection using geometrical tracking and feature analysis in aluminum wheels. *Mach. Vis. Appl.*, 22:157–170, 01 2011. 8
- [7] Stanley R. Deans. Hough transform from the radon transform. *IEEE transactions on pattern analysis and machine intelligence*, 3(2):185–188, 1981. 2
- [8] Christina Debbeler, Nicole Maaß, Matthias Elter, Frank Dennerlein, and Thorsten M. Buzug. A new CT rawdata redundancy measure applied to automated misalignment correction. In *Fully Three-Dimensional Image Reconstruction in Radiology and Nuclear Medicine*, pages 264–267, 2013. 2, 4
- [9] Michel Defrise and Rolf Clackdoyle. A cone-beam reconstruction algorithm using shift-variant filtering and cone-beam backprojection. *Medical Imaging, IEEE Transactions on*, 13:186 – 195, 04 1994. 1
- [10] Zhitao Fan, Feng Guan, Chunlin Wu, and Ming Yan. The continuity of images by transmission imaging revisited. *arXiv preprint arXiv:1401.1558*, 2014. 1
- [11] Pablo Fernández Alcantarilla, Adrien Bartoli, and Andrew J. Davison. Kaze features. In *Computer Vision – ECCV 2012*, pages 214–227, 10 2012. 5
- [12] Robert Frysch and Georg Rose. Rigid motion compensation in interventional c-arm CT using consistency measure on projection data. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 298–306. Springer, 2015. 4
- [13] Stuart Geman and Donald E. McClure. Bayesian images analysis: An application to single photon emission tomography. *Proc Stat Comp Sec Am Stat Assoc*, pages 12–18, 01 1985. 4
- [14] Pierre Grangeat. Mathematical framework of cone beam 3d reconstruction via the first derivative of the radon transform. *Lecture Notes in Mathematics*, 1497:66–97, 1991. 2
- [15] Richard I. Hartley and Peter Sturm. Triangulation. *Computer vision and image understanding*, 68(2):146–157, 1997. 8
- [16] Richard I. Hartley and Andrew Zisserman. *Multiple View Geometry in Computer Vision*. Cambridge University Press, ISBN: 0521540518, second edition, 2004. 4, 5
- [17] Alexander Katsevich, Katsuyuki Taguchi, and Alexander A. Zamyatin. Formulation of four katsevich algorithms in native geometry. *IEEE transactions on medical imaging*, 25(7):855–868, 2006. 1
- [18] Moritz Klüppel, Jian Wang, David Bernecker, Peter Fischer, and Joachim Hornegger. On Feature Tracking in X-Ray Images. In T.M. Deserno, H. Handels, H.-P. Meinzer, and T. Tolxdorff, editors, *Bildverarbeitung für die Medizin 2014*, Informatik aktuell, pages 132–137, Berlin Heidelberg, 2014. 2, 6
- [19] Svetlana Lazebnik, Yasutaka Furukawa, and Jean Ponce. Projective visual hulls. *International Journal of Computer Vision*, 74(2):137–165, 2007. 2
- [20] Stefan Lehmann, Andrew P. Bradley, and I. Vaughan L. Clarkson. Estimation of epipolar geometry via the radon transform. In *2006 IEEE International Conference on Acoustics Speech and Signal Processing Proceedings*, volume 2, pages II–II. IEEE, 2006. 2
- [21] Stefan Lehmann, Andrew P. Bradley, I. Vaughan L. Clarkson, John Williams, and Peter J. Kootsookos. Correspondence-free determination of the affine fundamental matrix. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 29:82–97, 02 2007. 2, 8
- [22] Jerome Lesaint. *Conditions de rang en tomographie de rayons X et leur application au problème d'auto-étalonnage*. PhD thesis, Grenoble Alpes, 2018. 1
- [23] Jérôme Lesaint, Simon Rit, Rolf Clackdoyle, and Laurent Desbat. GCC and FBCC for linear tomosynthesis. In F. Noo, editor, *Proceedings of the Fifth International Conference on Image Formation in X-Ray Computed Tomography (CT-Meeting)*, pages 114–118, 2018. 4
- [24] Quan-Tuan Luong and Olivier Faugeras. The fundamental matrix: Theory, algorithms, and stability analysis. *International Journal of Computer Vision*, 17:43–75, 01 1996. 2, 6
- [25] Marius Muja and David G. Lowe. Fast approximate nearest neighbors with automatic algorithm configuration. *VISAPP (I)*, 2(331-340):2, 2009. 5
- [26] John A. Nelder and Roger Mead. A simplex method for function minimization. *The computer journal*, 7(4):308–313, 1965. 5
- [27] Frédéric Noo, Rolf Clackdoyle, and Jed D. Pack. A two-step hilbert transform method for 2d image reconstruction. *Physics in Medicine and Biology*, 49(17):3903–3923, aug 2004. 1
- [28] Omid Poursaeed, Guandao Yang, Aditya Prakash, Qiuren Fang, Hanqing Jiang, Bharath Hariharan, and Serge Belongie. Deep fundamental matrix estimation without correspondences. In Laura Leal-Taixé and Stefan Roth, editors, *Computer Vision – ECCV 2018 Workshops*, pages 485–497, Cham, 2019. Springer International Publishing. 2
- [29] Alexander Preuhs, Andreas K. Maier, Michael Manhart, Javad Fotouhi, Nassir Navab, and Mathias Unberath. Double Your Views - Exploiting Symmetry in Transmission Imaging. In Schnabel J. Frangi A., editor, *Medical Image Computing and Computer Assisted Intervention*, volume 1, pages 356–364, 2018. 4
- [30] Marcus Prümmer, Joachim Hornegger, Günter Lauritsch, Lars Wigström, Erin Girard-Hughes, and Rebecca Fahrig. Cardiac C-arm CT: a unified framework for motion estimation and dynamic CT. *IEEE Trans Med Imaging*, 28(11):1836–49, 2009. 1
- [31] Sergio J. Sanabria and Orcun Goksel. Hand-held sound-speed imaging based on ultrasound reflector delineation. In *MICCAI*, 2016. 1
- [32] Sergio J. Sanabria, Katharina Martini, Gregor Freystaetter, Colleen A. Kraft, Orcun Goksel, Thomas Frauenfelder, and Marga B. Rominger. Speed of sound ultrasound: a pilot study on a novel technique to identify sarcopenia in seniors. *European Radiology*, 29:3–12, 2018. 1
- [33] Sudipta N. Sinha, Marc Pollefeys, and Leonard McMillan. Camera network calibration from dynamic silhouettes. In

Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2004. CVPR 2004., volume 1, pages I–I. IEEE, 2004. 2

- [34] Sabine Thürauf, Oliver Hornung, Mario Körner, Florian Vogt, Alois Knoll, and M. Ali Nasser. Model-based calibration of a robotic c-arm system using x-ray imaging. *Journal of Medical Robotics Research*, 3(03n04):1841002, 2018. 1
- [35] Kwan-Yee K. Wong and Roberto Cipolla. Reconstruction of sculpture from its profiles with unknown camera positions. *IEEE Transactions on Image Processing*, 13(3):381–389, 2004. 2
- [36] Tobias Würfl, Nicole Maaß, Frank Dennerlein, Xiaolin Huang, and Andreas K. Maier. Epipolar Consistency Guided Beam Hardening Reduction - ECC². In Ge Wang and Xuanqin Mou, editors, *Proceedings of the 14th International Meeting on Fully Three-Dimensional Image Reconstruction in Radiology and Nuclear Medicine*, pages 181–185, 2017. 4