Distance Metric between 3D Models and 2D Images for Recognition and Classification

Ronen Basri[*]
Dept. of Applied Math
The Weizmann Inst. of Science
Rehovot 76100, Israel
Internet: ronen@wisdom.weizmann.ac.il

Daphna Weinshall[†]
Institute of Computer Science
The Hebrew University of Jerusalem
91904 Jerusalem, Israel
Internet: daphna@cs.huji.ac.il

Abstract

Similarity measurements between $3D$ objects and $2D$ images are useful for the tasks of object recognition and classification. We distinguish between two types of similarity metrics: metrics computed in image-space (*image metrics*) and metrics computed in transformation-space (*transformation metrics*). Existing methods typically use image metrics; namely, metrics that measure the difference in the image between the observed image and the nearest view of the object. Example for such a measure is the Euclidean distance between feature points in the image and their corresponding points in the nearest view. (This measure can be computed by solving the *exterior orientation calibration problem.*) In this paper we introduce a different type of metrics: *transformation metrics.* These metrics penalize for the deformations applied to the object to produce the observed image.

In particular, we define a *transformation metric* that optimally penalizes for "affine deformations" under weak-perspective. A closed-form solution, together with the nearest view according to this metric, are derived. The metric is shown to be equivalent to the Euclidean image metric, in the sense that they bound each other from both above and below. It therefore provides an easy-to-use closed-form approximation for the commonly-used least-squares distance between models and images. We demonstrate an image understanding application, where the true dimensions of a photographed battery charger are estimated by minimizing the transformation metric.

1 Introduction

Object recognition is a process of selecting the object model that best matches the observed image. A common approach to recognition uses features (such as points or edges) to represent

[*]This research was conducted while affiliated with the McDonnell-Pew Center for Cognitive Neurocience and the Artificial Intelligence Lab. at MIT, Cambridge MA. Support for the laboratory's artificial intelligence research is provided in part by the Advanced Research Projects Agency of the Department of Defense under Office of Naval Research Contract N00014-91-J-4038.

[†]This research was conducted while affiliated with IBM T.J. Watson Research Center, Yorktown Heights, NY.

objects. An object is recognized in this approach if there exists a viewpoint from which the model features coincide with the corresponding image features, e.g. [R65, FB81, L85, HU87, BU93, TM87, UB91]. Since images often are noisy and models occasionally are imperfect, it is rarely the case that a model aligns perfectly with the image. Systems therefore look for a model that "reasonably" aligns with the image. Consequently, measures that assess the quality of a match become necessary.

Similarity measures between $3D$ objects and $2D$ images are needed for a range of applications:

- The recognition of specific objects in noisy images, as described above.
- The initial classification of novel objects. In this application a new object is associated with similar objects in the database. This way an image of, e.g., a Victorian chair is associated with models of (different) familiar chairs.
- The recognition of non-rigid objects whose geometry is not fully specified. An example is the recognition of $3D$ hand gestures. In this task only the generic shape of the gesture is known, and the particular instances differ according to the specific physiology of the hand.

Existing recognition methods are usually tailored to solve the first of these application, namely, the recognition of specific objects from noisy images. Many of these methods are sub-optimal (see Section 2 for a review), which may result in large number of either mis-recognition or false-positives. When these methods are extended to handle problems such as classification and recognition of non-rigid objects their performance may even be less predictable. The general problem of recognition therefore requires measures that provide a robust assessment of the similarity between objects and images. In this paper we describe two such measures, and develop a rigorous solution to the minimization problem that each measure entails.

A common measure for comparing $3D$ objects to $2D$ images is the Euclidean distance between feature points in the actual image and their corresponding points in the nearest view of the object. The assumption underlying this measure is that images are significantly less reliable than models, and so perturbations should be measured in the image plane. This assumption often suits recognition tasks. Other measures may better suit different assumptions. For example, when classifying objects, there is an inherent uncertainty in the structure of the classified object. One may therefore attempt to minimize the amount of deformations applied to the object to account for this uncertainty. Such a distance is measured in transformation space rather than in image space. A definition of these two types of measures is given in Section 3.

Measures to compare $3D$ models and $2D$ images generally are desired to have metrical properties; that is, they should monotonically increase with the difference between the measured entities. (A more exact definition is given in Appendix A.) The Euclidean distance between the image and the nearest view defines a metric. (We refer to this measure as the *image metric.*) The difficulty with employing this measure is that a closed-form solution to the problem has not yet been found, and therefore currently numerical methods must be employed to compute the measure. A common method to achieve a closed-form metric is to extend the set of transformations that objects are allowed to undergo from the rigid to the affine one. The problem with this measure is that it bounds the rigid measure from below, but not from above. Other methods either achieve only sub-optimal distances, or they do not define a metric. The existing approaches are reviewed in Section 2.

This paper presents a closed-form distance metric to compare $3D$ models and $2D$ images. The metric penalizes for the non-rigidity induced by the optimal affine transformation that aligns the model to the image under weak-perspective projection. Specifically, if A is the affine transformation that best aligns the model with the image, and $\mathcal{R}ig$ represents the set of all rigid transformations, then the metric is defined as

$$N_{tr} = \min_{R \in \mathcal{R}ig} \| A - R \|^2 \tag{1}$$

where the norm taken is the sum of squared elements. This metric is shown to bound the least-square distance between the model and the image both from above and below. We foresee three ways to use the metric developed in this paper:

1. Obtain a direct assessment of the similarity between $3D$ models and $2D$ images.
2. Obtain lower and upper bounds on the *image metric*. In many cases such bounds may suffice to unequivocally determine the identity of the observed object.
3. Provide an initial guess to be then used by a numerical procedure to solve the image distance.

The rest of this paper is organized as follows: In Section 2 we review related work. In Section 3 we define the concepts used in this paper. In Section 4 we summarize the main results of this paper. These results are discussed in detail and proved in section 5 for the *transformation metric*, and section 6 for the *image metric*. Sections 5 and 6 can be omitted on first reading. In Section 7 we describe possible applications of these metrics, and some comparisons to other methods; in one image understanding application we estimate the true dimensions of a photographed battery charger by minimizing the transformation metric.

2 Previous approaches

Previous approaches to the problem of model and image comparison using point features are divided into three major categories:

1. Least-square minimization in image space.
2. Sub-optimal methods using correspondence subsets.

3. Invariant functions.

The traditional photometric approach to the problem of model and image comparison involves retrieving a view of the object that minimizes the least-square distance to the image. This problem is referred to as the *exterior orientation calibration problem* (or the recovery of the *hand-eye transform*) and is defined as follows. Given a set of n $3D$ points (model points) and a corresponding set of n $2D$ points (image points), find the rigid transformation that minimizes the distance in the image plane between the transformed model points and the image points. An analytic solution to this problem has not yet been found. (Analytic solutions to the *absolute orientation problem*, the least-square distance between pairs of $3D$ objects, have been found, see [FH86, H87, H91]. An analytic solution to the least-square distance between pairs of $2D$ images has not yet been found.) Consequently, numerical methods are employed (see reviews in [T87, Y89]). Such solutions often suffer from stability problems, they are computationally intensive and require a good initial guess.

To avoid using numerical methods, frequently the object is allowed to undergo affine transformations instead of just rigid ones. Affine transformations are composed of general linear transformations (rather than rotations) and translations, and they include in addition to the rigid transformations also reflection, stretch, and shear. The solution in the affine case is simpler than that of the rigid case because the quadratic constraints imposed in the rigid case are not taken into account, enabling the construction of a closed-form solution. At least six points are required to find an affine solution under perspective projection [FB81], and four are required under orthographic projection [UB91].

The affine measure bounds the rigid measure from below. The rigid measure, however, is not bounded from above, and so the actual rigid measure may sometimes be significantly larger than the computed affine measure. This is demonstrated by the following example. Consider the case of matching four model points to four image points under weak-perspective. Since in this case there always exists a unique affine solution, the affine distance between the model and the image is zero. On the other hand, since three points uniquely determine the rigid transformation that aligns the model to the image, by perturbating one point we can increase the rigid distance unboundedly.

A second approach to comparing models to images, often called *alignment*, involves the selection of a small subset of correspondences (*alignment key*), solving for the transformation using this subset, and then transforming the other points and measuring their distance from the corresponding image points. Three [FB81, RBPD81, HLON91] or four [HCLL89] points are required under perspective projection, and three points under weak perspective [U89, HU87] . The obtained distance critically depends on the choice of alignment key. Different choices produce different distance measures between the model and the image. The results almost always are sub-optimal, since it is generally better to match all points with small errors than to exactly match a subset of points and project all the errors onto the others. However, by relying on small subsets of correspondences alignment can overcome occlusion and clutter.

A third approach involves the application of invariant functions. Such functions return a constant value when applied to any image of a particular model. Invariant functions were successfully used only with special kinds of models, such as planar objects (e.g., [LSW87, FMZCHR91]). More general objects can be recognized using model-based invariant functions [W93]. For noise-free data, model-based invariant functions return zero if the image is an exact instance of the object. To account for noise, the output of these functions usually is required to be below some fixed threshold. In general, very little research has been conducted to characterize the behavior of these functions when the model and the image do not perfectly align. The result of thresholding therefore becomes arbitrary.

3 Definitions and notation

In the following discussion, we assume weak-perspective projection. Namely, the object undergoes a $3D$ transformation that includes rotation, translation, and scaling, and is then orthographically projected onto the image. Perspective distortions are not accounted for and treated as noise. The weak-perspective projection model is particularly useful when objects are observed from a relatively long distance.

In order to define a similarity measure for comparing 3D objects to 2D images, as discussed in section 1, we first define the **best-view** of a $3D$ object given a $2D$ image:

Definition 1: *[best-view] Let ∂ denote a difference measure between two $2D$ images of n features. Given a $2D$ image of an object composed of n features, the* **best-view** *of a $3D$ object (model) composed of n corresponding features, is the view for which the smallest value of ∂ is obtained. The minimization is performed over all the possible views of the model; the views are obtained by applying a transformation T, taken from the set of permitted transformations Λ, and followed by a projection, Π.*

We compute ∂, the difference between two $2D$ images of n features in two ways:

image metric: *we measure position differences in the image, namely, it is the Euclidean distance between corresponding points in the two images, summed over all points.*

transformation metric: *the images are considered to be instances of a single $3D$ object. The metric measures the difference between the two transformations that align the object with the two images. This difference can be measured, for instance, by computing the Euclidean distance between the matrices that represent the two transformations (when the two transformations are linear).*

As is mentioned above, the measure ∂ is applied to the given image and to the views of the given model. These views are generated by applying a transformation from a set Λ of permitted transformations. The view that minimizes the distance ∂ to the image is considered as the **best view***, and the distance between the best view and the actual image is considered as the distance between the object and the image.*

We consider in this paper two families of transformations: rigid transformations[1] *and affine transformations, and we discuss the following metrics:*

N_{im}: *a metric that measures the image distance between the given image and the best* rigid *view of the object.*

N_{af}: *a metric that measures the image distance between the given image and the best* affine *view of the object.*

N_{tr}: a transformation metric. *We assume that the image is an affine view of the object. (When it is not, we substitute the image by the best affine view.) We look for the rigid view of the object so as to minimize the difference between the two transformations: the affine transformation (between the object and the image) and the rigid transformation (between the object and its possible rigid view.) In other words, we look for a view so as to minimize the amount of "affine deformations" applied to the object.*

To illustrate the difference between *image metrics* and *transformation metrics*, Fig. 1 shows an example of three $2D$ images, whose similarity relations reverse, depending on which kind of metric is used. Consider the planar object in Fig. 1(b) as a reference object, and assume Λ contains the set of rigid transformations in $2D$. The images in (a) and (c) are obtained by stretching the object horizontally (by 9/7) and vertically (by 3/2) respectively. (The image in (b) is obtained by applying a unit matrix to the object.)

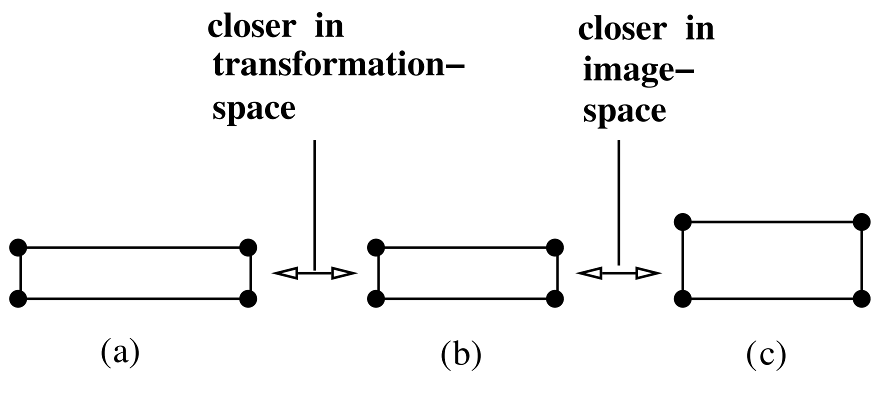


Figure 1: The $2D$ image shown in (b) is closer to the image in (a) when the difference is computed in transformation space, and closer to the image in (c) when the difference is the Euclidean difference between the two images.

[1] Note that a rigid transformation under weak perspective is equivalent to a similarity transformation followed by an orthographic projection

- The *image metric* between the images in (b) and (a) is 4, two pixel at each of the left corners of the rectangle.
  The *image metric* between the images in (b) and (c) is 2, one pixel at each of the upper corners of the rectangle.
  Therefore, according to the *image metric*, Fig. 1(c) is closer to (b) than (a) is.

- To compute the *transformation metric* consider the planar object illustrated in (b). We compute the difference between the matrices that represent the affine transformation from (b) to both (a) and (c) and the matrix that represent the best rigid transformation (in this case it is the unit matrix): (a) is obtained from (b) by a horizontal stretch of 9/7. The *transformation metric* between (a) and (b) is therefore $2/7 = 9/7 - 1$.
  (c) is obtained from (b) by a vertical stretch of 3/2. The *transformation metric* in this case is $1/2 = 3/2 - 1$.
  Therefore, according to the *transformation metric*, Fig. 1(a) is closer to (b) than (c) is.

It is interesting to note that in this example the solution obtained by minimizing the *transformation metric* seems to better correlate with human perception than the solution obtained by minimizing the *image metric*.

3.1 Derivation of N_{im} and N_{af}

We now define the rigid and the affine *image metrics* precisely. Under weak-perspective projection, the position in the image, $\vec{q}_i = (x_i, y_i)$, of a model point $\vec{p}_i = (X_i, Y_i, Z_i)$ following a rigid transformation is given by

$$q_i = \Pi(R\vec{p}_i + \vec{t}) \tag{2}$$

where R is a scaled, 3×3 rotation matrix, $\vec{t}$ is a translation vector, and Π represents the orthographic projection operator. More explicitly, denote by $\vec{r}_1^T$ and $\vec{r}_2^T$ the top two row vectors of R, and denote $\vec{t} = (t_x, t_y, t_z)$; we have that

$$\begin{aligned} x_i &= \vec{r}_1^T \cdot \vec{p}_i + t_x \\ y_i &= \vec{r}_2^T \cdot \vec{p}_i + t_y \end{aligned} \tag{3}$$

where

$$\begin{aligned} \vec{r}_1^T \cdot \vec{r}_2 &= 0 \\ \vec{r}_1^T \cdot \vec{r}_1 &= \vec{r}_2^T \cdot \vec{r}_2 \end{aligned} \tag{4}$$

The rigid metric, N_{im}, minimizes (over all R and $\vec{t}$) the difference between the two sides of Eq. 3 subject to the constraints (4).

When the object is allowed to undergo affine transformations, the rotation matrix R is replaced by a general 3×3 linear matrix (denoted by A) and the constraints (4) are ignored. That is

$$q_i = \Pi(A\vec{p}_i + \vec{t}) \tag{5}$$

Denote by $\vec{a}_1^T$ and $\vec{a}_2^T$ the top two row vectors of A, we obtain

$$\begin{aligned} x_i &= \vec{a}_1^T \cdot \vec{p}_i + t_x \\ y_i &= \vec{a}_2^T \cdot \vec{p}_i + t_y \end{aligned} \tag{6}$$

The affine metric, N_{af}, minimizes (over all A and $\vec{t}$) the difference between the two sides of Eq. 6.

To define the rigid and the affine metrics, we first note that the translation component of both the best rigid and affine transformations can be ignored if the centroid of both model and image points are moved to the origin. In other words, we begin by translating the model and image points so that

$$\sum_{i=1}^{n} \vec{p}_i = \sum_{i=1}^{n} \vec{q}_i = 0 \tag{7}$$

We claim that now $\vec{t} = 0$ obtains the minimum. The proof is given in Appendix C.

Denote

$$P = \begin{pmatrix} X_1 & Y_1 & Z_1 \\ & \vdots & \\ X_n & Y_n & Z_n \end{pmatrix} \tag{8}$$

a matrix of model point coordinates, and denote

$$\vec{x} = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} \qquad \vec{y} = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} \tag{9}$$

the location vectors of the corresponding image points. A rigid metric that reflects the desired minimization is given by

$$\begin{aligned} N_{im} &= \min_{\vec{r}_1, \vec{r}_2 \in \mathcal{R}^3} \|\vec{x} - P\vec{r}_1\|^2 + \|\vec{y} - P\vec{r}_2\|^2 \\ & \quad \text{s.t. } \vec{r}_1^T \cdot \vec{r}_2 = 0, \quad \vec{r}_1^T \cdot \vec{r}_1 = \vec{r}_2^T \cdot \vec{r}_2 \end{aligned} \tag{10}$$

The corresponding affine metric is given by

$$N_{af} = \min_{\vec{a}_1, \vec{a}_2 \in \mathcal{R}^3} \|\vec{x} - P\vec{a}_1\|^2 + \|\vec{y} - P\vec{a}_2\|^2 \tag{11}$$

In the affine case the solution is simple. We assume that the rank of P is 3 (the case for general, not coplanar, $3D$ objects). Denote $P^+ = (P^T P)^{-1} P^T$, the pseudo-inverse of P; we obtain that

$$\begin{aligned} \vec{a}_1 &= P^+ \vec{x} \\ \vec{a}_2 &= P^+ \vec{y} \end{aligned} \tag{12}$$

And the affine distance is given by

$$N_{af} = \|(I - PP^+)\vec{x}\|^2 + \|(I - PP^+)\vec{y}\|^2 \tag{13}$$

Since the solution in the rigid case is significantly more difficult than the solution in the affine case, often the affine solution is considered, and the rigidity constraints are used only for verification (e.g. [UB91, W93, DD92]).

The constraints (4) (substituting $\vec{a}_i$ for $\vec{r}_i$, and using Eq. 12) can be rewritten as

$$\begin{array}{rcl} \vec{x}^T(P^+)^T P^+ \vec{y} & = & 0 \\ \vec{x}^T(P^+)^T P^+ \vec{x} & = & \vec{y}^T(P^+)^T P^+ \vec{y} \end{array} \tag{14}$$

Denote

$$B = (P^+)^T P^+ \tag{15}$$

we obtain that

$$\begin{array}{rcl} \vec{x}^T B \vec{y} & = & 0 \\ \vec{x}^T B \vec{x} & = & \vec{y}^T B \vec{y} \end{array} \tag{16}$$

where B is an $n \times n$ symmetric, positive-semidefinite matrix of rank 3. (The rank would be smaller if the object points are coplanar.)

We call B the **characteristic matrix** of the object. B is a natural extension to the 3×3 model-based invariant matrix defined in [W93]. A more general definition, and its efficient computation from images, is discussed in Appendix B.

3.2 Derivation of N_{tr}

We can now define a *transformation metric*. Consider the affine solution. The nearest "affine view" of the object is obtained by applying the model matrix, P, to a pair of vectors, $\vec{a}_1$ and $\vec{a}_2$, defined in Eq. 12. In general, this solution is not rigid, and so the rigid constraints (4) do not hold for these vectors. The metric described here is based on the following rule. We are looking for another pair of vectors, $\vec{r}_1$ and $\vec{r}_2$, which satisfy the rigid constraints, and minimize the Euclidean distance to the affine vectors $\vec{a}_1$, and $\vec{a}_2$. (This is equivalent to assuming that the transformation parameters vary normally around their true value.) $P\vec{r}_1$ and $P\vec{r}_2$ define the best rigid view of the object under the defined metric. The metric, N_{tr}, is defined by

$$N_{tr} = \min_{\vec{r}_1, \vec{r}_2 \in \mathcal{R}^3} \|\vec{a}_1 - \vec{r}_1\|^2 + \|\vec{a}_2 - \vec{r}_2\|^2 \qquad \text{s.t.} \;\; \vec{r}_1^T \cdot \vec{r}_2 = 0, \;\; \vec{r}_1^T \cdot \vec{r}_1 = \vec{r}_2^T \cdot \vec{r}_2 \tag{17}$$

where $\vec{a}_1$ and $\vec{a}_2$ constitutes the optimal affine solution, therefore

$$N_{tr} = \min_{\vec{r}_1, \vec{r}_2 \in \mathcal{R}^3} \|P^+\vec{x} - \vec{r}_1\|^2 + \|P^+\vec{y} - \vec{r}_2\|^2 \qquad \text{s.t.} \;\; \vec{r}_1^T \cdot \vec{r}_2 = 0, \;\; \vec{r}_1^T \cdot \vec{r}_1 = \vec{r}_2^T \cdot \vec{r}_2 \tag{18}$$

In Section 5 we present a closed-form solution for this metric, and in Section 6 we show how this metric can be used to bound the *image metric* from both above and below.

4 Summary of results

In the rest of the paper we prove the following results:

4.1 Transformation space:

The *transformation metric* defined in Eq. 18 has the following solution

$$N_{tr} = \frac{1}{2}\left(\vec{x}^T B\vec{x} + \vec{y}^T B\vec{y} - 2\sqrt{\vec{x}^T B\vec{x} \cdot \vec{y}^T B\vec{y} - (\vec{x}^T B\vec{y})^2}\right)$$

where B is defined in Eq. 15, and $\vec{x}, \vec{y}$ in Eq. 9. The **best view** according to this metric is given by

$$\begin{aligned} \vec{x}^* &= PP^+(\beta_1\vec{x} + \beta_2\vec{y}) \\ \vec{y}^* &= PP^+(\gamma_1\vec{x} + \gamma_2\vec{y}) \end{aligned}$$

where $\beta_1, \beta_2, \gamma_1, \gamma_2$ are defined in Appendix D.

4.2 Image space:

Using N_{tr} we can bound the *image metric* from both above and below. Denote

$$N_{af} = \|(I - PP^+)\vec{x}\|^2 + \|(I - PP^+)\vec{y}\|^2$$

we show that

$$N_{af} + \lambda_1 N_{tr} \le N_{im} \le N_{af} + \lambda_3 N_{tr} \tag{19}$$

where $\lambda_1 \le \lambda_2 \le \lambda_3$ are the eigenvalues of $P^T P$. A sub-optimal solution to N_{im} is given by

$$N_{af} + \frac{2\mu_1\mu_2}{\mu_1 + \mu_2} N_{tr}$$

where the computation of μ_1, μ_2 is described in Appendix E. A tighter upper bound is deduced from this sub-optimal solution

$$N_{im} \le N_{af} + h(\lambda_2, \lambda_3)N_{tr} \le N_{af} + 2\lambda_2 N_{tr}$$

where $h(\lambda_2, \lambda_3) = \frac{2}{\frac{1}{\lambda_2} + \frac{1}{\lambda_3}}$ is the Harmonic mean of λ_2, λ_3. The sub-optimal solution is proposed as an initial guess for an iterative algorithm to compute N_{im}.

5 Closed-form solution in transformation space

We now present a metric to compare between 3D models and 2D images under weak perspective projection. The metric is a closed-form solution to the *transformation metric*, N_{tr} defined in Eq. 18. We use the notation developed in Section 3. B is the $n \times n$ **characteristic matrix** of the object, $\vec{x}, \vec{y} \in \mathcal{R}^n$ contain the x- and y-coordinates of the image features. The metric is given by

$$N_{tr} = \frac{1}{2}\left(\vec{x}^T B\vec{x} + \vec{y}^T B\vec{y} - 2\sqrt{\vec{x}^T B\vec{x} \cdot \vec{y}^T B\vec{y} - (\vec{x}^T B\vec{y})^2}\right) \tag{20}$$

This metric penalizes for the nonrigidities of the optimal affine transformation. Note that $N_{tr} = 0$ if the two rigid constraints in Eq. 16 are satisfied. Otherwise, $N_{tr} > 0$ represents the optimal penalty for a deviation from satisfying the two constraints.

Derivation of the results:

In the rest of this section we prove that the expression for N_{tr}, given by Eq. 20, is indeed the solution to the *transformation metric* defined in Eq. 18. The proof proceeds as follows: Theorem 1 computes the minimal solution when $\vec{r}_1$ and $\vec{r}_2$ are restricted to the plane spanned by $\vec{a}_1$ and $\vec{a}_2$; Theorem 2 extends this result to three-space.

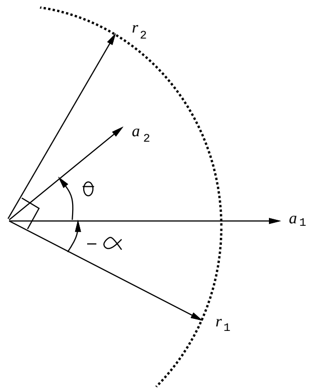


Figure 2: The vectors $\vec{a}_1$, $\vec{a}_2$, $\vec{r}_1$, and $\vec{r}_2$ in the coordinate system specified in Theorem 1. $\vec{a}_1$ and $\vec{a}_2$ represent the solution for the affine case. $\vec{r}_1$ and $\vec{r}_2$ are constrained to be in the same plane with $\vec{a}_1$ and $\vec{a}_2$, to be orthogonal, and to share the same norm.

Theorem 1: *When $\vec{r}_1$ and $\vec{r}_2$ are limited to span$\{\vec{a}_1, \vec{a}_2\}$, N_{tr} is given by Eq. 20.*

Proof: We first define a new coordinate system in which

$$\vec{a}_1 = w_1(1, 0)$$

$$\begin{aligned}\vec{a}_2 &= w_2(\cos\theta, \sin\theta)\\ \vec{r}_1 &= s(\cos\alpha, -\sin\alpha)\\ \vec{r}_2 &= s(\sin\alpha, \cos\alpha)\end{aligned}$$

(see Fig. 2). θ is the angle between $\vec{a}_1$ and $\vec{a}_2$, w_1 and w_2 are the lengths of $\vec{a}_1$ and $\vec{a}_2$ respectively. s is the common length of the two rotation vectors, $\vec{r}_1$ and $\vec{r}_2$, and $-\alpha$ is the angle between $\vec{a}_1$ and $\vec{r}_1$. Without loss of generality it is assumed below that $0° \le \theta \le 180°$ and $-90° \le \alpha \le 90°$. Notice that w_1, w_2, and θ are given and that s and α are unknown.

Denote f the term to be minimized, that is

$$f(\alpha, s) = \|\vec{a}_1 - \vec{r}_1\|^2 + \|\vec{a}_2 - \vec{r}_2\|^2$$

then

$$\begin{aligned}f(\alpha, s) &= (w_1 - s\cos\alpha)^2 + s^2\sin^2\alpha + (s\sin\alpha - w_2\cos\theta)^2 + (s\cos\alpha - w_2\sin\theta)^2\\ &= w_1^2 + w_2^2 + 2s^2 - 2s([w_1 + w_2\sin\theta]\cos\alpha + w_2\cos\theta\sin\alpha)\end{aligned}$$

The partial derivatives of f are given by

$$\begin{aligned}f_\alpha &= 2s([w_1 + w_2\sin\theta]\sin\alpha - w_2\cos\theta\cos\alpha)\\ f_s &= 4s - 2([w_1 + w_2\sin\theta]\cos\alpha + w_2\cos\theta\sin\alpha)\end{aligned}$$

To find possible minima we equate these derivatives to zero

$$\begin{aligned}f_\alpha &= 0\\ f_s &= 0\end{aligned}$$

Solutions with $s = 0$ are not optimal. In this case $f(\alpha, 0) = w_1^2 + w_2^2$, and later we show that solutions with $s > 0$ always imply smaller values for f.

When $s \neq 0$, $f_\alpha = 0$ implies

$$\tan\alpha^{min} = \frac{w_2\cos\theta}{w_1 + w_2\sin\theta}$$

therefore

$$\cos\alpha^{min} = \frac{1}{\sqrt{1 + (\tan\alpha^{min})^2}} = \frac{w_1 + w_2\sin\theta}{\sqrt{w_1^2 + w_2^2 + 2w_1w_2\sin\theta}}$$

$f_s = 0$ implies

$$s^{min} = \frac{1}{2}([w_1 + w_2\sin\theta]\cos\alpha^{min} + w_2\cos\theta\sin\alpha^{min})$$

Notice the similarity of this expression to the expression for f. At the minimum point f can be rewritten as

$$f^{min} = w_1^2 + w_2^2 - 2(s^{min})^2 \tag{21}$$

(From which it is apparent that any solution for f with $s \neq 0$ would be smaller than the solution with $s = 0$.) Substituting for α^{min} we obtain

$$
\begin{aligned}
s^{min} &= \frac{1}{2}([w_1 + w_2 \sin\theta]\cos\alpha^{min} + w_2\cos\theta\sin\alpha^{min}) \\
&= \frac{1}{2}\cos\alpha^{min}(w_1 + w_2\sin\theta + w_2\cos\theta\tan\alpha^{min}) \\
&= \frac{w_1 + w_2\sin\theta}{2\sqrt{w_1^2 + w_2^2 + 2w_1w_2\sin\theta}}(w_1 + w_2\sin\theta + \frac{w_2^2\cos^2\theta}{w_1 + w_2\sin\theta}) \\
&= \frac{1}{2}\sqrt{w_1^2 + w_2^2 + 2w_1w_2\sin\theta}
\end{aligned}
$$

and therefore

$$f^{min} = w_1^2 + w_2^2 - 2(s^{min})^2 = w_1^2 + w_2^2 - \frac{1}{2}\left(w_1^2 + w_2^2 + 2w_1w_2\sin\theta\right)$$

or,

$$f^{min} = \frac{1}{2}(w_1^2 + w_2^2 - 2w_1w_2\sin\theta)$$

Recall that w_1 and w_2 are the lengths of $\vec{a}_1$ and $\vec{a}_2$, that is

$$
\begin{aligned}
w_1^2 &= \vec{a}_1^T \cdot \vec{a}_1 = \vec{x}^T B \vec{x} \\
w_2^2 &= \vec{a}_2^T \cdot \vec{a}_2 = \vec{y}^T B \vec{y}
\end{aligned}
$$

and θ is the angle between the two vectors, namely

$$w_1w_2\sin\theta = \sqrt{w_1^2w_2^2(1-\cos^2\theta)} = \sqrt{\vec{x}^T B\vec{x}\cdot\vec{y}^T B\vec{y} - (\vec{x}^T B\vec{y})^2}$$

We obtain that

$$f^{min} = \frac{1}{2}\left(\vec{x}^T B\vec{x} + \vec{y}^T B\vec{y} - 2\sqrt{\vec{x}^T B\vec{x}\cdot\vec{y}^T B\vec{y} - (\vec{x}^T B\vec{y})^2}\right)$$

□

In Theorem 1 we proved that if $\vec{r}_1$ and $\vec{r}_2$ are restricted to the plane spanned by $\vec{a}_1$ and $\vec{a}_2$, the metric N_{tr} is given by Eq. 20. In Theorem 2 below we prove that any other solution for $\vec{r}_1$ and $\vec{r}_2$ results in a larger value for f, and therefore the minimum for f is obtained inside the plane, implying that N_{tr} indeed is given by Eq. 20.

Theorem 2: *The optimal $\vec{r}_1$ and $\vec{r}_2$ lie in the plane spanned by $\vec{a}_1$ and $\vec{a}_2$.*

Proof: Assume, by way of contradiction, that $\vec{r}_1, \vec{r}_2 \notin span\{\vec{a}_1, \vec{a}_2\}$; we show that the corresponding value for f is not minimal.

Consider first the plane spanned by $\vec{r}_2$ and $\vec{a}_1$, and assume, by way of contradiction, that $\vec{r}_1 \notin span\{\vec{r}_2, \vec{a}_1\}$; we show that there exists a vector $\vec{r}_1'$ such that

$$\begin{aligned} \|\vec{r}_1'\| &= \|\vec{r}_2\| \\ \vec{r}_1' &\perp \vec{r}_2 \end{aligned}$$

and

$$\|\vec{r}_1' - \vec{a}_1\| < \|\vec{r}_1 - \vec{a}_1\|$$

contradicting the optimality of f.

Assume $\|\vec{r}_2\| = s$, and denote by $\vec{r}_1'$ a vector with length s in the direction $(\vec{r}_2 \times \vec{a}_1) \times \vec{r}_2$. This vector lies in $span\{\vec{r}_2, \vec{a}_1\}$ and satisfies

$$\begin{aligned} \|\vec{r}_1'\| &= \|\vec{r}_2\| \\ \vec{r}_1' &\perp \vec{r}_2 \end{aligned}$$

(There exist two such vectors, opposing in their direction. We consider the one nearest to $\vec{a}_1$.) We now show that

$$\|\vec{r}_1' - \vec{a}_1\| < \|\vec{r}_1 - \vec{a}_1\|$$

Denote the angle between $\vec{a}_1$ and $\vec{r}_1'$ by α, and denote the angle between $\vec{r}_1'$ and $\vec{r}_1$ by β. Also, denote $w_1 = \|\vec{a}_1\|$ and $s = \|\vec{r}_1\| = \|\vec{r}_2\| = \|\vec{r}_1'\|$. We can rotate the coordinate system so as to obtain

$$\begin{aligned} \vec{r}_1' &= s(1,0,0) \\ \vec{r}_2 &= s(0,1,0) \\ \vec{a}_1 &= w_1(\cos\alpha, \sin\alpha, 0) \\ \vec{r}_1 &= s(\cos\beta, 0, \sin\beta) \end{aligned}$$

Now,

$$\begin{aligned} \|\vec{r}_1' - \vec{a}_1\|^2 &= (s - w_1\cos\alpha)^2 + w_1^2 \sin^2\alpha \;=\; w_1^2 + s^2 - 2sw_1\cos\alpha \\ \|\vec{r}_1 - \vec{a}_1\|^2 &= (s\cos\beta - w_1\cos\alpha)^2 + w_1^2\sin^2\alpha + s^2\sin^2\beta \;=\; w_1^2 + s^2 - 2sw\cos\alpha\cos\beta \end{aligned}$$

and therefore, when $\alpha \neq 0°$ and $\beta \neq 0°$ (when $\beta = 0°$, $\vec{r}_1$ and $\vec{r}_1'$ coincide.)

$$\|\vec{r}_1' - \vec{a}_1\| < \|\vec{r}_1 - \vec{a}_1\|$$

contradicting the minimality property. Therefore, $\vec{r}_1 \in span\{\vec{r}_2, \vec{a}_1\}$. Similarly, it can be shown that $\vec{r}_2 \in span\{\vec{r}_1, \vec{a}_2\}$, therefore all four vectors $\vec{a}_1$, $\vec{a}_2$, $\vec{r}_1$, and $\vec{r}_2$ lie in a single plane.
$\square$

Corollary 3: *The* transformation metric *is given by*

$$N_{tr} = \frac{1}{2}\left(\vec{x}^T B\vec{x} + \vec{y}^T B\vec{y} - 2\sqrt{\vec{x}^T B\vec{x}\cdot\vec{y}^T B\vec{y} - (\vec{x}^T B\vec{y})^2}\right)$$

and the **best view** *for this metric is*

$$\begin{aligned}\vec{x}^* &= PP^+(\beta_1\vec{x} + \beta_2\vec{y})\\ \vec{y}^* &= PP^+(\gamma_1\vec{x} + \gamma_2\vec{y})\end{aligned}$$

where

$$\begin{aligned}\beta_1 &= \frac{1}{2}\left(1 + \frac{\vec{y}^T B\vec{y}}{\sqrt{\vec{x}^T B\vec{x}\cdot\vec{y}^T B\vec{y} - (\vec{x}^T B\vec{y})^2}}\right)\\ \beta_2 = \gamma_1 &= -\frac{\vec{x}^T B\vec{y}}{2\sqrt{\vec{x}^T B\vec{x}\cdot\vec{y}^T B\vec{y} - (\vec{x}^T B\vec{y})^2}}\\ \gamma_2 &= \frac{1}{2}\left(1 + \frac{\vec{x}^T B\vec{x}}{\sqrt{\vec{x}^T B\vec{x}\cdot\vec{y}^T B\vec{y} - (\vec{x}^T B\vec{y})^2}}\right)\end{aligned}$$

Proof: The expression for the metric immediately follows from Theorem 1 and 2. The expression for the **best view** is developed in the Appendix D.
□

6 Solution in image space

In order to compute the *image metric* as defined in section 3, we need to solve the constraint minimization problem defined in Eq. 10

$$N_{im} = \min_{\vec{r}_1,\vec{r}_2\in\mathcal{R}^3} \|\vec{x} - P\vec{r}_1\|^2 + \|\vec{y} - P\vec{r}_2\|^2 \qquad \text{s.t. } \vec{r}_1^T\cdot\vec{r}_2 = 0,\ \ \vec{r}_1^T\cdot\vec{r}_1 = \vec{r}_2^T\cdot\vec{r}_2$$

Section 6.1 shows that N_{tr}, computed in the previous section, can be used to bound N_{im} from both above and below. Section 6.2 describes a direct method to compute a sub-optimal approximation to N_{im} and outlines an iterative algorithm to improve this estimate to obtain the optimal N_{im}.

6.1 Bounding the image metric with the transformation metric

In this section we show that using the *transformation metric* defined in Section 5 N_{tr}, and the affine metric N_{af} (given in Eq. 13), we can bound the *image metric* N_{im} from both above and below. We prove the following theorem:

Theorem 4: *Let $0 \leq \lambda_1 \leq \lambda_2 \leq \lambda_3$ denote the three eigenvalues of $P^T P$, then*

$$N_{af} + \lambda_1 N_{tr} \;\leq\; N_{im} \;\leq\; N_{af} + \lambda_3 N_{tr} \tag{22}$$

Proof: Denote by $\vec{r}_1^*$ and $\vec{r}_2^*$ the vectors that minimize the term for the *image metric* given in Eq. 10, namely

$$N_{im} = \|\vec{x} - P\vec{r}_1^*\|^2 + \|\vec{y} - P\vec{r}_2^*\|^2$$

and denote by $\vec{r}_1$ and $\vec{r}_2$ the vectors that minimize the *transformation metric* given in Eq. 18, namely

$$N_{tr} = \|P^+\vec{x} - \vec{r}_1\|^2 + \|P^+\vec{y} - \vec{r}_2\|^2$$

We start by showing the upper bound. Since $\vec{r}_1^*$ and $\vec{r}_2^*$ minimize the term for N_{im}, we can write

$$\begin{aligned} N_{im} &= \|\vec{x} - P\vec{r}_1^*\|^2 + \|\vec{y} - P\vec{r}_2^*\|^2 \\ &\leq \|\vec{x} - P\vec{r}_1\|^2 + \|\vec{y} - P\vec{r}_2\|^2 \end{aligned}$$

We now break each term in this sum into two orthogonal components as follows

$$\vec{x} - P\vec{r}_1 = (\vec{x} - PP^+\vec{x}) + (PP^+\vec{x} - P\vec{r}_1)$$

for which it holds that

$$(\vec{x} - PP^+\vec{x})^T \cdot (PP^+\vec{x} - P\vec{r}_1) = 0$$

The orthogonality readily follows from the identity

$$(PP^+)^T P = (P^+)^T P^T P = P(P^T P)^{-1}(P^T P) = P$$

Since the two components are orthogonal it holds that

$$\|\vec{x} - P\vec{r}_1\|^2 = \|\vec{x} - PP^+\vec{x}\|^2 + \|PP^+\vec{x} - P\vec{r}_1\|^2$$

and, similarly,

$$\|\vec{y} - P\vec{r}_2\|^2 = \|\vec{y} - PP^+\vec{y}\|^2 + \|PP^+\vec{y} - P\vec{r}_2\|^2$$

Therefore (recall that $\vec{r}_1$ and $\vec{r}_2$ minimize N_{tr} and that λ_3 is the largest eigenvalue of $P^T P$)

$$\begin{aligned} N_{im} &\leq \|\vec{x} - P\vec{r}_1\|^2 + \|\vec{y} - P\vec{r}_2\|^2 \\ &= \|\vec{x} - PP^+\vec{x}\|^2 + \|PP^+\vec{x} - P\vec{r}_1\|^2 + \|\vec{y} - PP^+\vec{y}\|^2 + \|PP^+\vec{y} - P\vec{r}_2\|^2 \\ &= \|(I - PP^+)\vec{x}\|^2 + \|(I - PP^+)\vec{y}\|^2 + \|P(P^+\vec{x} - \vec{r}_1)\|^2 + \|P(P^+\vec{y} - \vec{r}_2)\|^2 \\ &= N_{af} + \|P(P^+\vec{x} - \vec{r}_1)\|^2 + \|P(P^+\vec{y} - \vec{r}_2)\|^2 \\ &\leq N_{af} + \lambda_3(\|(P^+\vec{x} - \vec{r}_1)\|^2 + \|(P^+\vec{y} - \vec{r}_2)\|^2) \\ &= N_{af} + \lambda_3 N_{tr} \end{aligned}$$

Next, we prove the lower bound. The proof is similar to the proof in the upper bound case, but this time we start by breaking up the terms into orthogonal components. Then we use the facts that $\vec{r}_1$ and $\vec{r}_2$ minimize N_{tr} and that λ_1 is the smallest eigenvalue of $P^T P$.

$$\begin{aligned}
N_{im} &= \|\vec{x} - P\vec{r}_1^*\|^2 + \|\vec{y} - P\vec{r}_2^*\|^2 \\
&= \|\vec{x} - PP^+\vec{x}\|^2 + \|PP^+\vec{x} - P\vec{r}_1^*\|^2 + \|\vec{y} - PP^+\vec{y}\|^2 + \|PP^+\vec{y} - P\vec{r}_2^*\|^2 \\
&= \|(I - PP^+)\vec{x}\|^2 + \|(I - PP^+)\vec{y}\|^2 + \|P(P^+\vec{x} - \vec{r}_1^*)\|^2 + \|P(P^+\vec{y} - \vec{r}_2^*)\|^2 \\
&= N_{af} + \|P(P^+\vec{x} - \vec{r}_1^*)\|^2 + \|P(P^+\vec{y} - \vec{r}_2^*)\|^2 \\
&\geq N_{af} + \lambda_1(\|(P^+\vec{x} - \vec{r}_1^*)\|^2 + \|(P^+\vec{y} - \vec{r}_2^*)\|^2) \\
&\geq N_{af} + \lambda_1 N_{tr}
\end{aligned}$$

Consequently

$$N_{af} + \lambda_1 N_{tr} \ \leq\ N_{im} \ \leq\ N_{af} + \lambda_3 N_{tr}$$

□

6.2 Direct solution for the image metric

In this section we develop tighter bounds on the *image metric* by direct methods, following the same steps we took in the derivation of the *transformation metric* in Section 5. Unlike for the *transformation metric*, we cannot obtain a closed-form solution for the *image metric*, but we can obtain a better estimator than we have previously described. This also enables us to develop an iterative method to compute the distance exactly.

In section 6.2.1 we describe a change of coordinate system, arriving at a minimization problem which is similar to the one we had to solve for the *transformation metric*. The difference is that the sought vectors are constrained to lie on an ellipsoid rather than a sphere, and the ellipsoid is defined by a 3×3 positive-definite version of the **characteristic matrix** B.

In section 6.2.2 we restrict the solution vectors, $\vec{u}, \vec{v}$, to lie in a plane with the data vectors, $\vec{x}, \vec{y}$ and we derive the optimal solution under this constraint. The solution, however, is only sub-optimal, since in contrast to the *transformation metric*, the optimal solution in this case does not have to lie in the plane. Using this solution we derive a tighter upper bound on the optimal solution.

In section 6.2.3 we describe the general problem that needs to be solved, and outline an iterative method. We propose the solution obtained in the plane as an initial guess for this method.

6.2.1 Reducing the dimensionality of the problem

In Section 6.1 we have shown that the *image metric* can be broken into two orthogonal terms, implying that

$$N_{im} = N_{af} + \|P(P^+\vec{x} - \vec{r}_1^*)\|^2 + \|P(P^+\vec{y} - \vec{r}_2^*)\|^2 \tag{23}$$

This property is useful for a direct computation of the *image metric*. The first term, N_{af}, does not depend on $\vec{r}_1, \vec{r}_2$. To compute N_{im}, therefore, only the second term needs to be minimized

$$\min_{\vec{r}_1,\vec{r}_1 \in R^3} \|PP^+\vec{x} - P\vec{r}_1\|^2 + \|PP^+\vec{y} - P\vec{r}_2\|^2 \qquad \text{s.t.} \;\; \vec{r}_1^T \cdot \vec{r}_2 = 0, \;\; \vec{r}_1^T \cdot \vec{r}_1 = \vec{r}_2^T \cdot \vec{r}_2 \tag{24}$$

Note first that $PP^+\vec{x}$ and $PP^+\vec{y}$, two vectors in R^n, both lie in a single linear subspace of dimension 3. (This follows from the fact, shown in [UB91], that every image of a $3D$ object can be written as a linear combination of three independent views.) Moreover, the three columns of P lie in the same subspace. It therefore follows that the vectors $\vec{u} = P\vec{r}_1$ and $\vec{v} = P\vec{r}_2$ must also lie in this subspace.

Denote $\vec{X} = PP^+\vec{x}$ and $\vec{Y} = PP^+\vec{y}$, the projection of $\vec{x}$ and $\vec{y}$ to the column space of P, and denote $\vec{u} = P\vec{r}_1$ and $\vec{v} = P\vec{r}_2$. (Note that $\vec{r}_1 = P^+\vec{u}$, $\vec{r}_2 = P^+\vec{v}$, and $B = (P^+)^T P^+$, the **characteristic matrix** of the object.) We rewrite the problem as follows

$$\min_{\vec{u},\vec{v} \in \mathcal{R}^n} \|\vec{X} - \vec{u}\|^2 + \|\vec{Y} - \vec{v}\|^2 \qquad \text{s.t.} \;\; \vec{u}^T B \vec{v} = 0, \;\; \vec{u}^T B \vec{u} = \vec{v}^T B \vec{v} \tag{25}$$

Since all the vectors, $\vec{X}$, $\vec{Y}$, $\vec{u}$, and $\vec{v}$, lie in a $3D$ subspace (the column space of P) we can perform the minimization in $\mathcal{R}^3$. To transform the system into $\mathcal{R}^3$, we rotate the vectors and the characteristic matrix B so as to get nontrivial (nonzero) values only in three of the coordinates. Recall that distances and quadratic forms are invariant under rotation. The rotation matrix Ω that should be applied to all terms is defined by the eigenvectors of B. Applying this matrix to B (in the form $\Omega^T B \Omega$) results in a diagonal matrix with the three positive eigenvalues of B.

6.2.2 Closed-form solution in the plane

Theorem 5: *When $\vec{u}$ and $\vec{v}$ are limited to span$\{\vec{X}, \vec{Y}\}$, the solution of Eq. 25 is given by*

$$\tilde{N}_{im} = \frac{\mu_1 \mu_2}{\mu_1 + \mu_2} \left(\vec{x}^T B \vec{x} + \vec{y}^T B \vec{y} - 2\sqrt{\vec{x}^T B \vec{x} \cdot \vec{y}^T B \vec{y} - (\vec{x}^T B \vec{y})^2} \right) \tag{26}$$

where $\sqrt{\mu_1} \leq \sqrt{\mu_2}$ are the principal axes of the ellipse, defined by the intersection of the ellipsoid B with the plane span$\{\vec{X}, \vec{Y}\}$.

Note the similarity between this solution and N_{tr} in Eq. 20. In fact,

$$\tilde{N}_{im} = \frac{2\mu_1 \mu_2}{\mu_1 + \mu_2} N_{tr} \tag{27}$$

The proof closely follows the proof for N_{tr} presented in Section 5 (Theorem 2). We therefore skip some of the details.

Proof: We first define a new coordinate system in which

$$
\begin{aligned}
\vec{X} &= w_1(\sqrt{\mu_1}\cos\eta, \sqrt{\mu_2}\sin\eta) \\
\vec{Y} &= w_2(\sqrt{\mu_1}\cos\theta, \sqrt{\mu_2}\sin\theta) \\
\vec{u} &= s(\sqrt{\mu_1}\cos\alpha, \sqrt{\mu_2}\sin\alpha) \\
\vec{v} &= s(-\sqrt{\mu_1}\sin\alpha, \sqrt{\mu_2}\cos\alpha) \\
B &= \begin{pmatrix} \frac{1}{\mu_1} & 0 & B_{13} \\ 0 & \frac{1}{\mu_2} & B_{23} \\ B_{13} & B_{23} & B_{33} \end{pmatrix}
\end{aligned}
$$

Without loss of generality it is assumed below that $-90° \le \eta \le 90°$, $\eta \le \theta \le \eta + 180°$, and $-90° \le \alpha \le 90°$. Notice that w_1, w_2, η and θ are given and that s and α are unknown.

Notice that this setting of coordinate system is similar to the one used in Theorem 1 with the exceptions that here $\vec{u}$ and $\vec{v}$ lie on an ellipse rather than on a circle, and that in general none of the points can be brought to lie on a principal axis.

Denote by f the term to be minimized, that is

$$f(\alpha, s) = \|\vec{X} - \vec{u}\|^2 + \|\vec{Y} - \vec{v}\|^2$$

then

$$
\begin{aligned}
f(\alpha, s) &= \mu_1(w_1\cos\eta - s\cos\alpha)^2 + \mu_2(w_1\sin\eta - s\sin\alpha)^2 + \mu_1(w_2\cos\theta + s\sin\alpha)^2 + \\
&\quad \mu_2(w_2\sin\theta - s\cos\alpha)^2 \\
&= w_1^2(\mu_1\cos^2\eta + \mu_2\sin^2\eta) + w_2^2(\mu_1\cos^2\theta + \mu_2\sin^2\theta) + s^2(\mu_1 + \mu_2) - \\
&\quad 2s(w_1\mu_1\cos\eta\cos\alpha + w_1\mu_2\sin\eta\sin\alpha - w_2\mu_1\cos\theta\sin\alpha + w_2\mu_2\sin\theta\cos\alpha)
\end{aligned}
$$

The partial derivatives of f are given by

$$
\begin{aligned}
f_\alpha &= 2s[(w_1\mu_1\cos\eta + w_2\mu_2\sin\theta)\sin\alpha - (w_1\mu_2\sin\eta - w_2\mu_1\cos\theta)\cos\alpha)] \\
f_s &= 2s(\mu_1 + \mu_2) - 2(w_1\mu_1\cos\eta\cos\alpha + w_1\mu_2\sin\eta\sin\alpha - w_2\mu_1\cos\theta\sin\alpha + w_2\mu_2\sin\theta\cos\alpha)
\end{aligned}
$$

To find possible minima we equate these derivatives to zero

$$
\begin{aligned}
f_\alpha &= 0 \\
f_s &= 0
\end{aligned}
$$

Again, solutions with $s = 0$ can be ignored since they do not correspond to the global minimum (for a similar reason as in the proof of Theorem 1).

When $s \ne 0$, $f_\alpha = 0$ implies

$$\tan\alpha^{min} = \frac{w_1\mu_2\sin\eta - w_2\mu_1\cos\theta}{w_1\mu_1\cos\eta + w_2\mu_2\sin\theta}$$

$f_s = 0$ implies

$$s^{min} = \frac{w_1 \mu_1 \cos \eta + w_2 \mu_2 \sin \theta}{\cos \alpha^{min} (\mu_1 + \mu_2)}$$

and, similarly to Eq. 21,

$$f^{min} = w_1^2 (\mu_1 \cos^2 \eta + \mu_2 \sin^2 \eta) + w_2^2 (\mu_1 \cos^2 \theta + \mu_2 \sin^2 \theta) - (\mu_1 + \mu_2)(s^{min})^2 \tag{28}$$

We substitute s^{min} and $\cos \alpha^{min}$, using the identity $\cos \alpha = \frac{1}{\sqrt{1+\tan^2 \alpha}}$, into Eq. 28. After some manipulations, we obtain

$$\tilde{N}_{im} = f^{min} = \frac{\mu_1 \mu_2}{\mu_1 + \mu_2} \left(w_1^2 + w_2^2 - 2 w_1 w_2 \sin(\theta - \eta) \right) \tag{29}$$

Note that

$$(PP^+)^T B (PP^+) = (P^+)^T P^T (P^+)^T P^+ P P^+ = (P^+)^T (P^+ P)^T (P^+ P) P^+ = (P^+)^T P^+ = B \tag{30}$$

from which it follows that

$$\begin{aligned} w_1^2 &= \vec{X}^T B \vec{X} = \vec{x}^T B \vec{x} \\ w_2^2 &= \vec{Y}^T B \vec{Y} = \vec{y}^T B \vec{y} \\ w_1 w_2 \cos(\theta - \eta) &= \vec{X}^T B \vec{Y} = \vec{x}^T B \vec{y} \end{aligned} \tag{31}$$

We substitute the identities from Eq. 31 into Eq. 29, obtaining the expression for $\tilde{N}_{im}$ in Eq. 26.
□

The derivation for μ_1 and μ_2 is given in Appendix E.

The sub-optimal solution in the plane can be used to improve the bounds on the *image metric*, which were previously discussed in Theorem 4.

Theorem 6: *Let $0 \leq \lambda_1 \leq \lambda_2 \leq \lambda_3$ be the three eigenvalues of $P^T P$, then*

$$N_{af} + \lambda_1 N_{tr} \leq N_{im} \leq N_{af} + h(\lambda_2, \lambda_3) N_{tr} \tag{32}$$

where $h(\lambda_2, \lambda_3) = \frac{2}{\frac{1}{\lambda_2} + \frac{1}{\lambda_3}}$, the Harmonic Mean of λ_2, λ_3.

Proof: The eigenvalues of the **characteristic matrix** B are $\frac{1}{\lambda_1}$, $\frac{1}{\lambda_2}$, and $\frac{1}{\lambda_3}$. (This is shown in Appendix E.) Since $1/\mu_1$ and $1/\mu_2$ represent the eigenvalues of a section of B it holds that (see, e.g., [S76] p. 270)

$$\frac{1}{\lambda_1} \geq \frac{1}{\mu_1} \geq \frac{1}{\lambda_2} \geq \frac{1}{\mu_2} \geq \frac{1}{\lambda_3}$$

Using Eq. 27 we obtain that

$$\tilde{N}_{im} = \frac{2\mu_1\mu_2}{\mu_1+\mu_2}N_{tr} = \frac{2}{\frac{1}{\mu_1}+\frac{1}{\mu_2}}N_{tr} \;\leq\; \frac{2}{\frac{1}{\lambda_2}+\frac{1}{\lambda_3}}N_{tr} = h(\lambda_2,\lambda_3)N_{tr}$$

And, using Eq. 23 we obtain the upper bound

$$N_{im} \;\leq\; N_{af} + \tilde{N}_{im} \;\leq\; N_{af} + h(\lambda_2,\lambda_3)N_{tr}$$

□

Corollary 7:

$$N_{af} + \lambda_1 N_{tr} \;\leq\; N_{im} \;\leq\; N_{af} + \frac{2\mu_1\mu_2}{\mu_1+\mu_2}N_{tr} \tag{33}$$

Note that, since $h(a,b) \leq 2\min\{a,b\}$ for every a, b, we have the following corollary.

Corollary 8:

$$N_{af} + \lambda_1 N_{tr} \;\leq\; N_{im} \;\leq\; N_{af} + 2\lambda_2 N_{tr}$$

We cannot yet improve the lower bound in theorem 4; but we conjecture that

Conjecture 1: *Let $0 \leq \lambda_1 \leq \lambda_2 \leq \lambda_3$ be the three eigenvalues of $P^T P$, then*

$$N_{af} + h(\lambda_1,\lambda_2)N_{tr} \;\leq\; N_{im} \;\leq\; N_{af} + h(\lambda_2,\lambda_3)N_{tr} \tag{34}$$

Motivation: We know that if the two data points $\vec{X}$,$\vec{Y}$ lie on the ellipse whose principal axes are of length λ_1, λ_2 (the smallest cross-section of the ellipsoid B), then

$$N_{im} = N_{af} + h(\lambda_1,\lambda_2)N_{tr}$$

We can show that this solution is a local minimum, namely, it is not possible to improve the solution by applying small perturbations to the solution vectors.
□

6.2.3 An iterative optimal solution

The solution we obtained in Theorem 5 is sub-optimal; it is not the lowest distance. We now give the cost function, a function of four variables, which should be minimized to obtain the precise value of the *image metric*.

We first define a coordinate system such that

$$
\begin{aligned}
\vec{X} &= w_1(\sqrt{\lambda_1}\cos\theta\cos\nu, \sqrt{\lambda_2}\cos\theta\sin\nu, \sqrt{\lambda_3}\sin\theta) \\
\vec{Y} &= w_2(\sqrt{\lambda_1}\cos\zeta\cos\eta, \sqrt{\lambda_2}\cos\zeta\sin\eta, \sqrt{\lambda_3}\sin\zeta) \\
\vec{u} &= s(\sqrt{\lambda_1}\cos\alpha\cos\beta, \sqrt{\lambda_2}\cos\alpha\sin\beta, \sqrt{\lambda_3}\sin\alpha) \\
\vec{v} &= s(\sqrt{\lambda_1}(\sin\beta\cos\gamma + \sin\alpha\cos\beta\sin\gamma), \sqrt{\lambda_2}(-\cos\beta\cos\gamma + \sin\alpha\sin\beta\sin\gamma), \\
&\qquad -\sqrt{\lambda_3}\cos\alpha\sin\gamma) \\
B &= \begin{pmatrix} \frac{1}{\lambda_1} & 0 & 0 \\ 0 & \frac{1}{\lambda_2} & 0 \\ 0 & 0 & \frac{1}{\lambda_3} \end{pmatrix}
\end{aligned}
$$

where w_1, w_2, λ_1, λ_2, λ_3, ζ, η, θ, and ν are known, and s, α, β and γ are free.

Note that this setting of coordinate system is similar to the one used in Theorem 5, but now $\vec{u}$ and $\vec{v}$ lie on an ellipsoid rather than on an ellipse.

In this notation the free parameters are selected so as to satisfy the two rigid constraints, $\vec{u}^T B\vec{u} = \vec{v}^T B\vec{v}$ and $\vec{u}^T B\vec{v} = 0$. To compute the *image metric*, the following function should be minimized.

$$
\begin{aligned}
f(s,\alpha,\beta,\gamma) \;=\;& \lambda_1(s\cos\alpha\cos\beta - w_1\cos\theta\cos\nu)^2 + \lambda_2(s\cos\alpha\sin\beta - w_1\cos\theta\sin\nu)^2 + \\
& \lambda_3(s\sin\alpha - w_1\sin\theta)^2 + \\
& \lambda_1(s\sin\beta\cos\gamma + s\sin\alpha\cos\beta\sin\gamma - w_2\cos\zeta\cos\eta)^2 + \\
& \lambda_2(-s\cos\beta\cos\gamma + s\sin\alpha\sin\beta\sin\gamma - w_2\cos\zeta\sin\eta)^2 + \\
& \lambda_3(-s\cos\alpha\sin\gamma - w_2\sin\zeta)^2
\end{aligned} \tag{35}
$$

N_{im} is the global minimum of $f(s,\alpha,\beta,\gamma)$. Assuming that $f(s,\alpha,\beta,\gamma)$ is convex in the area that contains both the global minimum N_{im} and the sub-optimal solution $(N_{af} + \tilde{N}_{im})$, we can employ the following iterative method to compute N_{im}:

1. compute $\tilde{N}_{im}$;
2. improve the solution by any gradient-descent method until a local minimum is obtained.

If the convexity assumption is correct, this method returns the correct *image metric*, otherwise it may return a sub-optimal solution.

7 Applications

In this section we describe possible applications of the theory described above, and some comparisons to other methods: in 7.1 we illustrate the outcome of using the new transformation metric with a real example; in 7.2 we demonstrate an image understanding application, where the true dimensions of a photographed battery charger are estimated by minimizing the transformation metric; finally in 7.3 we compare the distances between $3D$ objects and $2D$ images, obtained by alignment, to our results.

7.1 Experiments with real images

We have applied the transformation metric to real images. In this experiment a $3D$ model of a chair, including twelve of its feature points, was given. Four images of the chair at different orientations, as well as two more images of two different chairs, were photographed (see Figs. 3-4). Twelve feature points corresponding to the model points were manually extracted from these images. The model was compared to the 6 pictures using the *transformation metric*. Figs. 3-4 also show the model points in the *best view* according to the *transformation metric*, overlaid on their corresponding points in the reference image. We can see that albeit the model chair is compared in Fig. 4 to different chairs, the matching obtained is relatively good. Note that in Fig. 3 the matching between the model and the images of the same chair is not perfect due to errors in the $3D$ measurements and the weak perspective approximation.

The distances between the model of the reference chair (condition number 5.25) and the six images of Figs. 3-4 are given in Table 1. It can be seen that the transformation metric values obtained for the images of the same chair (range between 0.04 and 0.06) are significantly smaller than those of the other chairs (range between 0.41 and 1.08). Similar results are obtained for the affine metric and the various bounds. As is expected, the affine metric always underestimates the image metric. The tightest upper bound is 10%-30% larger than the lower bound, and the worst upper bound for the same chair (2.73 for the top left image in Fig. 3) is still much lower than the lowest upper bound for the other chairs (5.587 for the left image in Fig. 4). Thus, the bounds suffice to discriminate between the images of the same chair from the images of the other chairs.

7.2 Using the metric in $3D$ reconstruction

Here we demonstrate the usefulness of the *transformation metric* by using it in an image understanding application. In this application we attempt to infer the dimensions of an object from a single view. We will use an image of a battery charger as an input (Fig. 5). Suppose that we can identify the object either by recognizing it as a box of some arbitrary dimensions or by identifying certain surface markings on the object. Our task now is to estimate the dimensions of the box from the image coordinates of the seven visible corners of the charger.

To find the actual dimensions of the battery charger, we search the parameter space $u \times v \times w$, where u is the depth of the charger (the width of the left face), v is its height, and w is the length of the front face. Since under the weak-perspective projection model we can infer the dimensions of objects up to a scale factor only, we may set one of these parameters to be

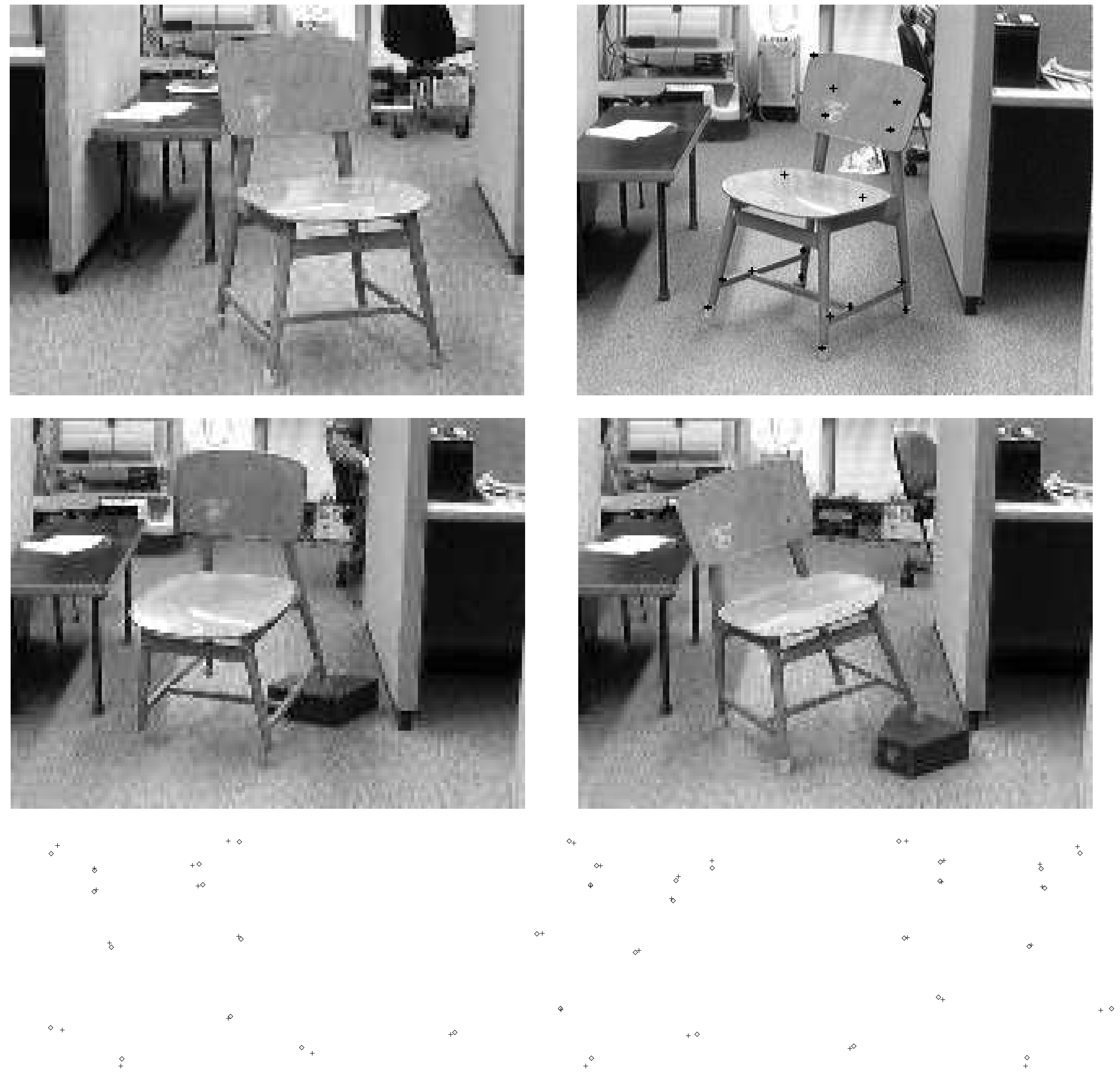


Figure 3: Top and middle rows: four images of a chair, with feature points marked on one of them for illustration (but not all points were used here), for which the $3D$ coordinates of the points are known (i.e., the model is given). Bottom row: for three of the images, the original feature points of the image are marked by $+$; for comparison, the feature points of the model, in the closest image according to N_{tr}, are marked by diamonds.

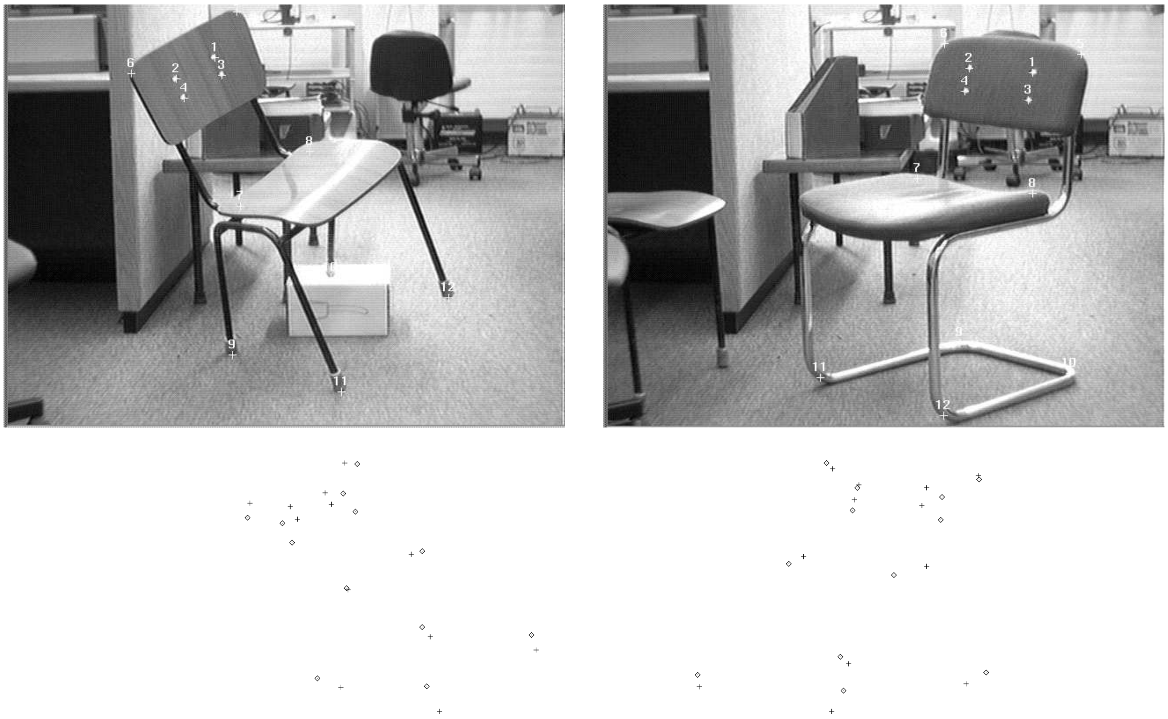


Figure 4: Top row: two images of different chairs, with feature points marked on one of them for illustration. Bottom row: the original feature points of the image are marked by +; for comparison, the feature points of the model chair (shown in Fig. 3), in the closest image according to N_{tr}, are marked by diamonds.

	Same chair (Fig. 3)				Other chairs (Fig. 4)	
	Top left	Top right	Bottom left	Bottom right	Left	Right
N_{tr}	0.060	0.048	0.053	0.040	1.080	0.410
N_{aff}	1.745	1.352	1.240	1.240	4.008	5.364
Lower bound (Eq. 19)	1.971	1.582	1.510	1.450	5.587	5.876
Upper bound (Eq. 19)	2.730	2.319	2.332	2.122	9.777	7.680
Tighter (Eq. 32)	2.336	1.941	1.916	1.778	7.719	6.731
Tightest (Eq. 33)	2.313	1.884	1.878	1.755	7.350	6.514

Table 1: The transformation metric values N_{tr}, the affine metric, and the various bounds computed for the four chairs in Fig. 3 and in Fig. 4. Except for the transformation metric, the values are normalized so they reflect the average distortion in pixels of a single feature point.

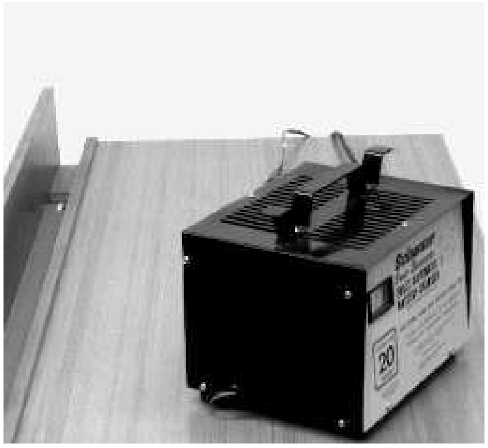


Figure 5: A picture of a battery charger, whose dimensions are: depth - 22.5cm, length - 28cm, and height - 19cm.

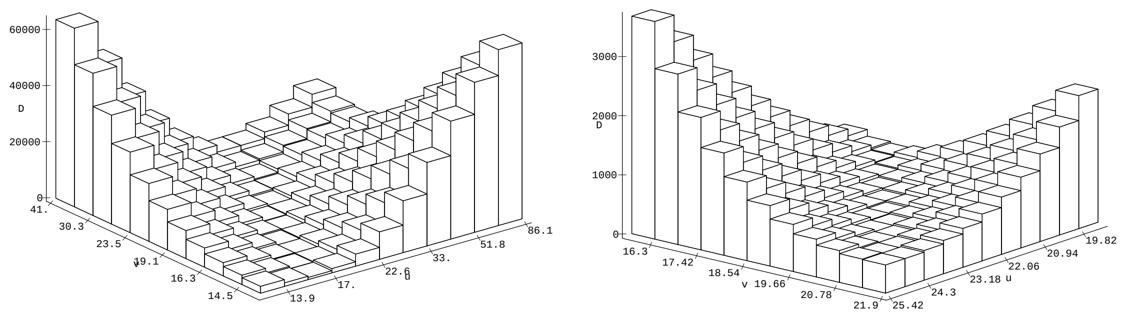
constant and search the space of the other two measurements. In our experiment we set w to its true value, 28cm, and searched the space of the other two parameters, u and v.

In Fig. 6 the upper limit on N_{im}, given in Eq. 32, is plotted for each pair of parameters. The first search was done on a coarse scale (Fig. 6a). The minimum of the error bound is obtained for $u = 22.6$, $v = 19.1$, which is the (correct) answer with certainty of ± 2 cm for u, and ± 1 cm for v. The second search was done on a finer scale (Fig. 6b). The minimum of the error bound is obtained for $u = 22.06$, $v = 19.1$, which is the answer with certainty of ± 0.28 cm in each dimension. This final result provides a reasonably good estimate of the dimensions of the battery charger, with an error of ≈ 0.5cm in one dimension (u).

7.3 Simulations

To test the presented metric we have compared it with the alignment method. As was mentioned in Section 2 the alignment method involves the selection of a small subset of correspondences (*alignment key*), solving for the transformation using this subset, and then transforming the rest of the points and measuring their distance from the corresponding image points. The obtained distance critically depends on the choice of alignment key. Different choices produce different distance measures between the model and the image. The results are almost always sub-optimal, since it is usually better to match all points with small errors than to exactly match a subset of points and project the errors entirely onto the others.

In our simulations, models composed of four points were projected to the image using weak perspective projection. Gaussian noise (with standard deviation 0.05 of the radius of the $3D$ object) was added to the obtained images. Using the expression for N_{tr} given in (20), we computed the upper and lower bounds on the *image metric* between the model to the noisy



(a) (b)

Figure 6: Plots of the upper bound on N_{im} (Eq. 32), when comparing the picture in Fig. 5 to a box model whose dimensions are $u \times v \times 28$cm. (a) Plot of the bound, for coarse sampling in log scale of $u \in [13.9 - 86.1\text{cm}]$ and $v \in [14.5 - 41\text{cm}]$. (b) Plot of the bound, for fine sampling in linear scale of $u \in [19.82 - 25.42\text{cm}]$ and $v \in [16.3 - 21.9\text{cm}]$.

images. In addition, we computed the corresponding *alignment distances*, each reflecting the distance between one model point and its predicted projection in the image after the alignment of the remaining three image points to the model. Note that in computing the alignment distances we considered only those distances that are obtained from correct matches (and so in the error free case these distances would vanish).

The figures below summarize our results. Fig. 7 shows the percentage of *alignment distances* which actually lie within the bounds on the *image metric* computed by our metric (given in Eq. 32). It can be seen that when the bounds are relatively tight (when the condition number on the **characteristic matrix** B is relatively low) most of the alignment solutions exceed the upper bound. This is surprising since the alignment distances are computed for correct matches. Only when the condition number gets larger do the alignment distances lie within the bounds. When a tighter upper bound is used (Eq. 33), a smaller portion of the alignment distances actually lie within the bounds.

Fig. 8 shows the maximal and minimal *alignment distances* obtained in different runs relative to the upper and lower bounds on the *image metric*, given in Eq. 32 and Eq. 33. It can be seen that in many cases even the best alignment solution (the one that minimizes the distance) still exceeds the upper bound.

The simulations demonstrate that the bounds on the image metric developed in this paper provide approximations to the metric that are often preferable to the distances obtained by the alignment method. These bounds are better for "symmetric" objects, objects whose convex-hull is close to a sphere, than for objects which are significantly stretched or contracted along one

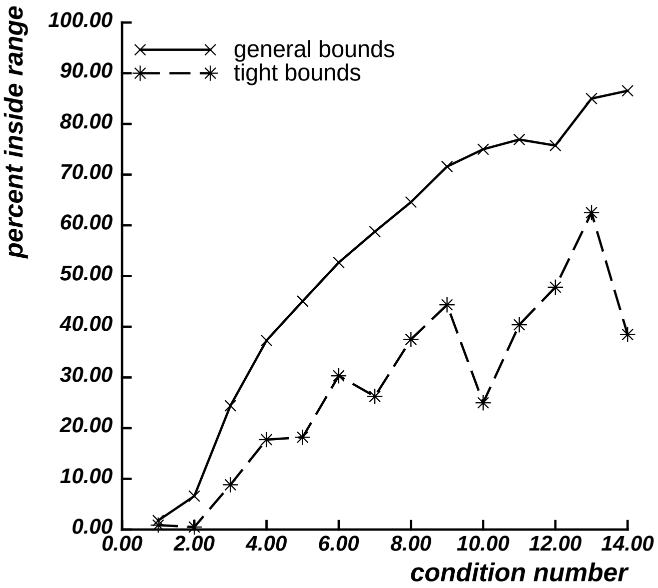


Figure 7: The percent of *alignment distances* which lie within the bounds on the *image metric* computed from our closed-form equations. The abscissa gives the condition number of the **characteristic matrix**, B, which determines how far apart the lower and upper bounds on the *image metric* are. The larger the condition number is, the further apart the bounds are. Solid graph: alignment distances relative to the wide bounds from Eq. 32. Dashed lines: alignment distances relative to the tight upper bound from Eq. 33.

spatial dimension.

It should be noted that measuring the difference between models and images is not the only objective of the alignment method. More importantly, by relying only on a small number of correspondences, alignment enables recognizing non-segmented objects in cluttered scenes in a worst-case polynomial time complexity. However such a strategy would often yield errors in estimating the transformation (see, [AG93, GHA92, GHJ92]), and so typical algorithms often try, after obtaining an initial alignment, to extend the match with more correspondences (e.g., [AG93, FB81]). Consequently, an accurate and fast estimation of the alignment transformation at this stage can reduce the amount of computation that is necessary to eliminate false hypotheses. The simulations above demonstrate the potential use of the transformation metric in this context, both by providing bounds to eliminate erroneous hypotheses and by providing an initial guess for an iterative pose estimation.

8 Conclusion

We have proposed a *transformation metric* to measure the similarity between $3D$ models and $2D$ images. The *transformation metric* measures the amount of affine deformation applied to

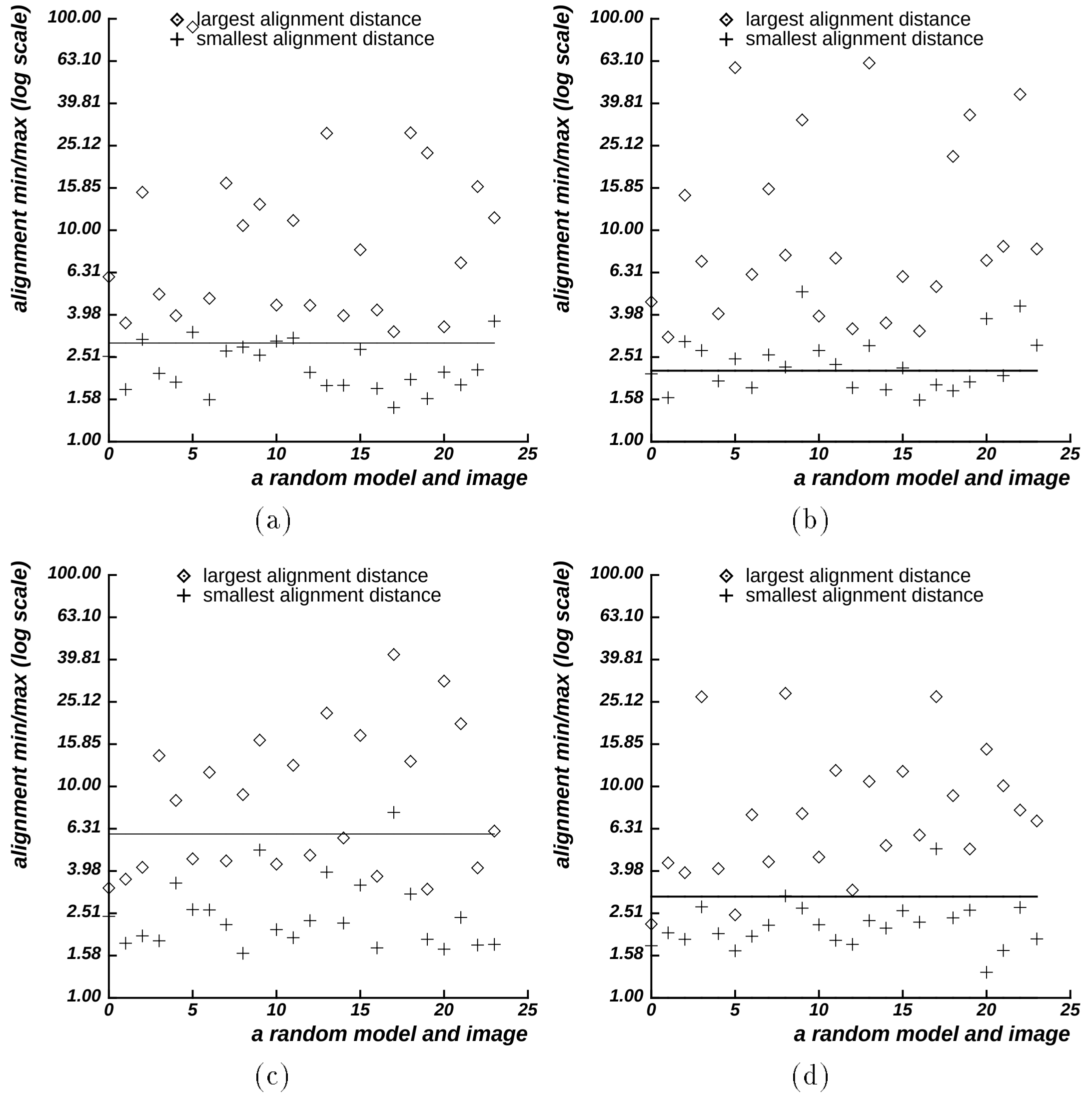


Figure 8: The maximal and minimal *alignment distances* are plotted for a number of models and objects, varying along the abscissa. The distances in these plots were normalized so as to obtain constant lower and upper bounds (the lower bound is set to 1; the upper bound is set to be the average ratio of the upper bound to the lower bound in each sequence of runs). Small (between 1.5 and 2.5) and large (between 4.5 and 5.5) condition numbers are used, and the results are compared to both the wide (Eq. 32) and the tight (Eq. 33) bounds. (a) Small condition number, wide bounds. (b) Small condition number, tight bounds. (c) Large condition number, wide bounds. (d) Large condition number, tight bounds.

the object to produce the given image. A simple, closed-form solution for this metric has been presented. This solution is optimal in transformation space, and it is used to bound the *image metric* from both above and below. The *transformation metric* presented in this paper can be used to obtain a direct assessment of the similarity between models and images or as a mean to evaluate the image metric. The proposed metric can be used in several different ways in the recognition and classification tasks. We conclude the paper with a brief discussion of possible applications of the metric.

The *transformation metric* provides a sub-optimal closed-form estimate for the *image metric*. A scheme which uses this measure will prefer "symmetric" objects, objects whose convex-hull is close to a sphere, over other objects which are significantly stretched or contracted along one spatial dimension. This solution can also be used as an initial guess in an iterative process that computes the optimal value of the image metric numerically. The sub-optimal solution derived using the image metric provides a better estimate for the image metric than the affine solution, which has been used for example in [DD92] as the initial guess for computing the perspective image metric numerically.

Another potential application of the metric is in evaluating hypothesized correspondences in an alignment algorithm. Alignment is a method for evaluating the similarity between models and images based on a small number of correspondences. While the use of few correspondences is advantageous for recognizing objects in polynomial time complexity while overcoming partial occlusion, it may often yield errors in estimating the distance between models and images (see, e.g., [GHA92]). Therefore, typical algorithms often try, after obtaining an initial alignment, to extend the match with additional correspondences (e.g., [FB81]). The bounds derived on the image metric may be used at this stage to evaluate potential correspondences. Our simulations show that these bounds often provide better estimates than those provided by using alignment.

Finally, our transformation metric can be used in schemes that attempt to classify objects. A scheme for classification was recently proposed [B93], in which classes contain objects that share the same basic features in distorted positions. Our metric can be used under such a scheme to evaluate the amount of affine distortion applied to the object relative to a prototype object in order to determine its class identity.

Appendices

A Metric properties

The measures described in this paper compare entities of different dimensionalities: $3D$ objects and $2D$ images. We define a *metric* for comparing such entities as follows. Let P be a set of n model points, and let q be a set of n corresponding image points. A distance function, $N(P, q)$, defined using a difference function $\partial(q, q')$ between two views (see Section 3), is called a *metric* if

1. $N(P,q) \geq 0$ for every model P and image q.
2. $N(P,q) = 0$ if, and only if, q is a rigid view of P.
3. $\forall q', \quad N(P,q) \leq N(P,q') + \partial(q - q')$

For the *image metric*, N_{im}, ∂ is simply the Euclidean distance between corresponding points in the compared images. It is straightforward to see that the conditions hold for this case. In the rest of this appendix we prove that these conditions also hold for the *transformation metric*, N_{tr}.

Transformation metric

The *transformation metric*, N_{tr}, measures the amount of "affine deformation" applied to the object in the image. The metric conditions for N_{tr} are defined as follows.

1. $N(P,q) \geq 0$ for every model P and image q.
2. $N(P,q) = 0$ if, and only if, there exists a rigid view which coincides with PP^+q. (In other words, the best affine view of the object is a rigid view and there is no "affine deformation".)
3. $\forall q', \quad N(P,q) \leq N(P,q') + \|P^+(q - q')\|$

Theorem 9: *N_{tr} is a metric.*

Proof:

1. $N_{tr} \geq 0$. N_{tr} minimizes a non-negative distance function. It is therefore always non-negative.
2. $N_{tr} = 0$ if, and only if, the best affine view is rigid. Denote $\vec{x}$ and $\vec{y}$ the x and y coordinates of the points in q, according to Eq. 20

$$\begin{array}{ll} & N_{tr} = 0 \\ \Longleftrightarrow & (\vec{x}^T B\vec{x} + \vec{y}^T B\vec{y})^2 = 4(\vec{x}^T B\vec{x} \cdot \vec{y}^T B\vec{y} - (\vec{x}^T B\vec{y})^2) \\ \Longleftrightarrow & (\vec{x}^T B\vec{x})^2 + 2(\vec{x}^T B\vec{x} \cdot \vec{y}^T B\vec{y}) + (\vec{y}^T B\vec{y})^2 = 4(\vec{x}^T B\vec{x} \cdot \vec{y}^T B\vec{y}) - 4(\vec{x}^T B\vec{y})^2 \\ \Longleftrightarrow & (\vec{x}^T B\vec{x} - \vec{y}^T B\vec{y})^2 = -4(\vec{x}^T B\vec{y})^2 \end{array}$$

   This equation holds if, and only if, both sides are zero implying that

$$\begin{array}{rcl} \vec{x}^T B\vec{x} & = & \vec{y}^T B\vec{y} \\ \vec{x}^T B\vec{y} & = & 0 \end{array}$$

   The best affine view of the object is given by PP^+x, PP^+y. Following Eq. 16, the best affine view also satisfies the rigidity constraints above, and therefore it forms a rigid view.

3. The metric N_{tr} is defined in Eq. 18 as:

$$N_{tr}(P,q) = \min_{\vec{r}_1,\vec{r}_2 \in \mathcal{R}^3} \|P^+\vec{x} - \vec{r}_1\|^2 + \|P^+\vec{y} - \vec{r}_2\|^2 \qquad \text{s.t. } \vec{r}_1^{\,T}\cdot\vec{r}_2 = 0, \;\; \vec{r}_1^{\,T}\cdot\vec{r}_1 = \vec{r}_2^{\,T}\cdot\vec{r}_2$$

Let $\vec{w}_1$ and $\vec{w}_2$ be the optimal vectors for q', that is

$$N_{tr}(P,q') = \|P^+\vec{x}' - \vec{w}_1\|^2 + \|P^+\vec{y}' - \vec{w}_2\|^2$$

And we obtain

$$\begin{aligned} & N_{tr}(P,q') + \|P^+(q-q')\| \\ = \;& \|P^+\vec{x}' - \vec{w}_1\|^2 + \|P^+\vec{y}' - \vec{w}_2\|^2 + \|P^+\vec{x} - P^+\vec{x}'\|^2 + \|P^+\vec{y} - P^+\vec{y}'\|^2 \\ \geq \;& \|P^+\vec{x} - \vec{w}_1\|^2 + \|P^+\vec{y} - \vec{w}_2\|^2 \\ \geq \;& \min_{\vec{r}_1,\vec{r}_2 \in \mathcal{R}^3} \|P^+\vec{x} - \vec{r}_1\|^2 + \|P^+\vec{y} - \vec{r}_2\|^2 \; = \; N_{tr}(P,q) \end{aligned}$$

B The computation of the characteristic matrix

In Eq (15) the characteristic matrix B was defined using the matrix of *Euclidean* model point coordinates P. We now give a more general (though equivalent) definition of B using a matrix of *affine* model point coordinates Q. Namely, the point coordinates in Q are given in a coordinate system whose axes are not necessarily orthonormal. This definition makes it possible to compute B directly from three or more images with a completely *linear* algorithm, which requires no more than pseudo-inverse.

We select an affine coordinate system whose independent axes are defined by three of the object points, to be called the basis points. Let P_{bas} denote the submatrix of P corresponding to the coordinates of the basis points, and let Q denote the affine coordinates of all the object points in this basis. It immediately follows that:

$$P = Q \cdot P_{bas}$$

Let B_{bas} denote the *characteristic matrix* of the three basis points. From Eq (15) it follows that

$$B_{bas} = (P_{bas}^{-1})^T P_{bas}^{-1} \tag{36}$$

Finally, from the definition of pseudo-inverse it can be readily verified that

$$P^+ = (Q \cdot P_{bas})^+ = P_{bas}^{-1} Q^+ \tag{37}$$

We now describe B in terms of Q and B_{bas}. Substituting Eq (37) into the definition of B in Eq (15), and using Eq (36), we obtain

$$B = (P^+)^T P^+ = (Q^+)^T \cdot B_{bas} \cdot Q^+$$

The linear and incremental computation of the matrices Q and B_{bas} from at least three images of the object points is described in [WT95].

C Eliminating translation

In this appendix we show that translation can be ignored if we set the centroids of both model and image points to be the origin. To show this, we prove that the best rigid and affine transformations maps the model centroid to the image centroid. We begin by showing that, given two sets of n $2D$ points (images), the best translation that relates the two images maps the centroid of the first image to that of the second.

Lemma 10: *Let $p_1, ..., p_n \in \mathcal{R}^2$ and $q_1, ..., q_n \in \mathcal{R}^2$ be two sets of corresponding points. Denote by $\bar{p} = \frac{1}{n}\sum_{i=1}^{n} p_i$ and $\bar{q} = \frac{1}{n}\sum_{i=1}^{n} q_i$ the centroids of $p_1, ..., p_n$ and $q_1, ..., q_n$ respectively. The translation $t^* \in \mathcal{R}^2$ that minimizes the term*

$$D^* = \min_{t \in \mathcal{R}^2} \sum_{i=1}^{n} \|p_i + t - q_i\|^2$$

is given by

$$t^* = \bar{q} - \bar{p}$$

Proof: Assume, by way of contradiction, that the best translation is given by

$$t' = t^* + \delta$$

for some nonzero $\delta \in \mathcal{R}^2$. Denote the new term by D'

$$\begin{aligned} D' &= \sum_{i=1}^{n} \|p_i + t' - q_i\|^2 \\ &= \sum_{i=1}^{n} \|p_i + t^* + \delta - q_i\|^2 \\ &= \sum_{i=1}^{n} \|p_i + t^* - q_i\|^2 + 2\sum_{i=1}^{n}(p_i + t^* - q_i) \cdot \delta + \sum_{i=1}^{n} \|\delta\|^2 \\ &= D^* + 2n(\bar{p} + t^* - \bar{q}) \cdot \delta + n\|\delta\|^2 \end{aligned}$$

Since $t^* = \bar{q} - \bar{p}$, we obtain that

$$\bar{p} + t^* - \bar{q} = 0$$

and, therefore,

$$D' = D^* + n\|\delta\|^2$$

which implies that

$$D^* < D'$$

contradicting the initial assumption.
□

Using Lemma 10 we prove that the best rigid and affine transformations map the model centroid to the image centroid.

Theorem 11: *Let $P_1, ..., P_n \in \mathcal{R}^3$ be a set of n model points, and let $q_1, ..., q_n \in \mathcal{R}^2$ be the corresponding n image points. The rigid transformation $\{s^*, R^*, t^*\}$ that minimizes the term*

$$D^* = \min_{\{s,R,t\}} \sum_{i=1}^{n} \|s\Pi R P_i + t - q_i\|^2$$

where Π denotes the orthographic projection, satisfies

$$\bar{q} = s^* \Pi R^* \bar{P} + t^*$$

Proof: Denote by $p_i = s^* \Pi R^* P_i$; according to Lemma 10

$$t^* = \bar{q} - \bar{p}$$

Since

$$\bar{p} = \frac{1}{n} \sum_{i=1}^{n} p_i = \frac{1}{n} \sum_{i=1}^{n} s^* \Pi R^* P_i = s^* \Pi R^* \bar{P}$$

we obtain that

$$\bar{q} = \bar{p} + t^* = s^* \Pi R^* \bar{P} + t^*$$

The theorem holds also if we consider affine transformations rather then only the rigid ones. The rotation matrix R is replaced in this case by a general linear transformation A.
□

Theorem 11 shows that the best rigid and affine transformations map the model centroid to the image centroid. Consequently, if the two centroids are moved to the origin, the translation component vanishes. This follows immediately from Theorem 11, since

$$\bar{q} = s^* \Pi R^* \bar{P} + t^*$$

then

$$\bar{P} = \bar{q} = 0$$

implies

$$t^* = 0$$

D Best View

In this appendix we develop an expression for the **best view** of the *transformation metric*, N_{tr}. The derivations here follow the notations used in the proof of Theorem 1, from which we have that

$$\begin{aligned} s &= \frac{1}{2}\sqrt{w_1^2 + w_2^2 + 2w_1 w_2 \sin\theta} \\ s\cos\alpha &= \frac{1}{2}(w_1 + w_2 \sin\theta) \\ s\sin\alpha &= \frac{1}{2} w_2 \cos\theta \end{aligned}$$

According to Theorem 2, $\vec{r}_1, \vec{r}_2 \in span\{\vec{a}_1, \vec{a}_2\}$. We can therefore express $\vec{r}_1$ and $\vec{r}_2$ by

$$\begin{aligned} \vec{r}_1 &= \beta_1 \vec{a}_1 + \beta_2 \vec{a}_2 \\ \vec{r}_2 &= \gamma_1 \vec{a}_1 + \gamma_2 \vec{a}_2 \end{aligned}$$

where β_1, β_2, γ_1, and γ_2 are scalars. Substituting the definitions of the vectors $\vec{r}_1$, $\vec{r}_2$, $\vec{a}_1$, and $\vec{a}_2$ we obtain

$$\begin{aligned} s\cos\alpha &= \beta_1 w_1 + \beta_2 w_2 \cos\theta \\ -s\sin\alpha &= \beta_2 w_2 \sin\theta \end{aligned}$$

and

$$\begin{aligned} s\sin\alpha &= \gamma_1 w_1 + \gamma_2 w_2 \cos\theta \\ s\cos\alpha &= \gamma_2 w_2 \sin\theta \end{aligned}$$

Therefore

$$\begin{aligned} \beta_1 &= \frac{s\sin\alpha\cos\theta + s\cos\alpha\sin\theta}{w_1 \sin\theta} \\ \beta_2 &= -\frac{s\sin\alpha}{w_2\sin\theta} \\ \gamma_1 &= \frac{s\sin\alpha\sin\theta - s\cos\alpha\cos\theta}{w_1\sin\theta} \\ \gamma_2 &= \frac{s\cos\alpha}{w_2\sin\theta} \end{aligned}$$

Substituting for s and α we obtain

$$\begin{aligned} \beta_1 &= \frac{1}{2}(1 + \frac{w_2}{w_1\sin\theta}) \\ \beta_2 = \gamma_1 &= -\frac{\cos\theta}{2\sin\theta} \\ \gamma_2 &= \frac{1}{2}(1 + \frac{w_1}{w_2\sin\theta}) \end{aligned}$$

And substituting for w_1, w_2, and θ

$$\begin{aligned} \beta_1 &= \frac{1}{2}\left(1 + \frac{\vec{y}^T B\vec{y}}{\sqrt{\vec{x}^T B\vec{x} \cdot \vec{y}^T B\vec{y} - (\vec{x}^T B\vec{y})^2}}\right) \\ \beta_2 = \gamma_1 &= -\frac{\vec{x}^T B\vec{y}}{2\sqrt{\vec{x}^T B\vec{x} \cdot \vec{y}^T B\vec{y} - (\vec{x}^T B\vec{y})^2}} \\ \gamma_2 &= \frac{1}{2}\left(1 + \frac{\vec{x}^T B\vec{x}}{\sqrt{\vec{x}^T B\vec{x} \cdot \vec{y}^T B\vec{y} - (\vec{x}^T B\vec{y})^2}}\right) \end{aligned}$$

Now, to obtain the **best view** we use the following identities

$$\begin{array}{lllllllll} \vec{x}^* & = & P\vec{r}_1 & \vec{r}_1 & = & \beta_1\vec{a}_1 + \beta_2\vec{a}_2 & \vec{a}_1 & = & P^+\vec{x} \\ \vec{y}^* & = & P\vec{r}_2 & \vec{r}_2 & = & \gamma_1\vec{a}_1 + \gamma_2\vec{a}_2 & \vec{a}_2 & = & P^+\vec{y} \end{array}$$

Therefore

$$\begin{aligned} \vec{x}^* &= PP^+(\beta_1\vec{x} + \beta_2\vec{y}) \\ \vec{y}^* &= PP^+(\gamma_1\vec{x} + \gamma_2\vec{y}) \end{aligned}$$

E Computing the eigenvalues of an ellipse

In this appendix we compute the eigenvalues of the ellipsoid B and the eigenvalue of an elliptic section of this ellipsoid.

We first show that the eigenvalues of the **characteristic matrix**, B, are $\frac{1}{\lambda_1}$, $\frac{1}{\lambda_2}$, and $\frac{1}{\lambda_3}$, where λ_1, λ_2, and λ_3 are the three positive eigenvalues of P^TP. This is derived as follows.

$$B\vec{a} = \frac{1}{\lambda}\vec{a} \Longleftrightarrow P(P^TP)^{-1}(P^TP)^{-1}P^T\vec{a} = \frac{1}{\lambda}\vec{a}$$

Multiplying both sides by P^T we obtain that

$$(P^TP)^{-1}P^T\vec{a} = \frac{1}{\lambda}P^T\vec{a}$$

Denote $\vec{b} = P^T\vec{a}$

$$(P^TP)^{-1}\vec{b} = \frac{1}{\lambda}\vec{b}$$

which implies that

$$(P^TP)\vec{b} = \lambda\vec{b}$$

Given $\vec{X} = PP^+\vec{x}$ and $\vec{Y} = PP^+\vec{y}$ in R^3, and a positive definite 3×3 matrix B, let B' denote the ellipse defined by the intersection of the ellipsoid B with the plane $span\{\vec{X}, \vec{Y}\}$. We need to find the eigenvalues of B', $\frac{1}{\mu_1}$ and $\frac{1}{\mu_2}$.

Without loss of generality we assume that $\vec{X}$ and $\vec{Y}$ lie on the ellipsoid defined by B (namely, we normalize the vectors so that $\vec{X}^TB\vec{X} = \vec{x}^TB\vec{x} = 1$ and $\vec{Y}^TB\vec{Y} = \vec{y}^TB\vec{y} = 1$). Let θ denote the angle between $\vec{X}$ and $\vec{Y}$. We define two orthonormal vectors $\vec{x}'$ and $\vec{y}'$, which span the plane $span\{\vec{X}, \vec{Y}\}$, as follows:

$$\begin{aligned} \vec{x}' &= \frac{\vec{X}}{|\vec{X}|} \\ \vec{y}' &= \frac{\vec{Y} - \frac{\vec{X}\cdot\vec{Y}}{|X|^2}\vec{X}}{|Y|\sin\theta} \end{aligned}$$

Every vector $\vec{v} \in span\{\vec{X}, \vec{Y}\}$ can be written as

$$\vec{v} = \alpha\vec{x}' + \beta\vec{y}'$$

and the intersection ellipse B' is given by

$$\vec{v}B'\vec{v} = 1 \qquad \Longleftrightarrow \qquad (\,\alpha \quad \beta\,)A^T BA\begin{pmatrix}\alpha\\ \beta\end{pmatrix} = 1$$

for A the 3×2 matrix whose columns are $\vec{x}'$ and $\vec{y}'$. We therefore have that

$$B' = A^T BA = \begin{pmatrix}(\vec{x}')^T B\vec{x}' & (\vec{x}')^T B\vec{y}'\\ (\vec{x}')^T B\vec{y}' & (\vec{y}')^T B\vec{y}'\end{pmatrix}$$

Substituting the expressions for $\vec{x}'$ and $\vec{y}'$, we get

$$\begin{aligned}(\vec{x}')^T B\vec{x}' &= \frac{1}{|X|^2}\\ (\vec{y}')^T B\vec{y}' &= \frac{|X|^2 - 2|X||Y|\cos\theta(\vec{X}^T B\vec{Y}) + |Y|^2\cos^2\theta}{|X|^2|Y|^2\sin^2\theta}\\ (\vec{x}')^T B\vec{y}' &= \frac{(\vec{X}^T B\vec{Y})|X| - |Y|\cos\theta}{|X|^2|Y|\sin\theta}\end{aligned}$$

To obtain the two eigenvalues of B' $\frac{1}{\mu_1}$ and $\frac{1}{\mu_2}$, we solve the characteristic equation of B', whose roots are

$$\frac{|X|^2 + |Y|^2 - 2|X||Y|\cos\theta\cdot\kappa \pm \sqrt{(|X|^2 + |Y|^2 - 2|X||Y|\cos\theta\cdot\kappa)^2 - 4|X|^2|Y|^2\sin^2\theta(1-\kappa^2)}}{2|X|^2|Y|^2\sin^2\theta}$$

for $\kappa = \vec{X}^T B\vec{Y} = \vec{x}^T B\vec{y}$, $|X| = |PP^+\vec{x}|$, $|Y| = |PP^+\vec{y}|$, and $\cos\theta = \frac{\vec{X}\cdot\vec{Y}}{|X||Y|}$.

Acknowledgments

We thank Larry Maloney for valuable discussions and technical help, and Eric Grimson, Yael Moses, and Tomasso Poggio for their comments to earlier drafts of this paper.

References

[AG93] Alter, T. D. and Grimson W. E. L., (1993). Fast and Robust 3D Recognition by Alignment, in *Proc. Fourth Inter. Conf. Computer Vision*, Berlin.

[B93] Basri, R. (1993). Recognition by prototypes. *Computer Vision and Pattern Recognition (CVPR-93), New York City, NY.*

[BU93] Basri, R. and Ullman, S. (1993). The alignment of objects with smooth surfaces. *Computer Vision, Graphics, and Image Processing: Image Understanding*, 57(3):331–345.

[DD92] DeMenthon, D. F. and Davis, L. S. (1992). Model-based object pose in 25 lines of code. In *Proceedings of the 2nd European Conference on Computer Vision*, Santa Margherita Ligure, Italy. Springer-Verlag.

[FH86] Faugeras, O. D. and Hebert, M. (1986). The representation, recognition, and locating of 3-D objects. *Int. J. of Robotics Research*, 5(3):27–52.

[FB81] Fischler, M. A. and Bolles, R. C. (1981). Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. *Communications of the ACM*, 24:381–395.

[FMZCHR91] Forsyth, D., Mundy, J. L., Zisserman, A., Coelho, C., Heller, A., and Rothwell, C. (1991). Invariant descriptors for 3-D object recognition and pose. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 13:971–991.

[GHA92] Grimson, W. E. L., Huttenlocher, D. P., and Alter, T. D., (1992). Recognizing 3D Objects from 2D Images: An Error Analysis. in *Proc. IEEE Conf. Computer Vision Pat. Rec.*, Urbana.

[GHJ92] Grimson, W. E. L., Huttenlocher, D. P., and Jacobs, D. W., (1992). A Study of Affine Matching with Bounded Sensor Error. in *Proc. Second European Conf. Computer Vision*, 291–306.

[HLON91] Haralick, R. M., Lee, C., Ottenberg, K., and Nolle, M. (19). Analysis and solutions of the three point perspective pose estimation problem. In *Proceedings IEEE Conf. on Computer Vision and Pattern Recognition*, pages 592–598, Urbana-Champaign, IL.

[HCLL89] Horaud, R., Conio, B., Leboulleux, O., and Lacolle, B. (1989). An analytic solution for the perspective 4-point problem. *Computer Vision, Graphics, and Image Processing*, 47:33–44.

[H87] Horn, B. K. P. (1987). Closed-form solution of absolute orientation using unit quaternions. *Journal of the Optical Society of America*, 4:629–642.

[H91] Horn, B. K. P. (1991). Relative orientation revisited. *Journal of the Optical Society of America*, 8:1630–1638.

[HU87] Huttenlocher, D. P. and Ullman, S. (1987). Object recognition using alignment. In *Proceedings of the 1st International Conference on Computer Vision*, pages 102–111, London, England. IEEE, Washington, DC.

[LSW87] Lamdan, Y., Schwartz, J. T., and Wolfson, H. (1987). On recognition of 3-d objects from 2-d images. Robotics research Technical Report 122, Courant Institute of Math. Sciences, N.Y. University.

[L85] Lowe, D. G. (1985). Three-dimensional object recognition from single two-dimensional images. Robotics research Technical Report 202, Courant Institute of Math. Sciences, N.Y. University.

[RBPD81] Rives, P., Bouthemy, B., Prasada, B., and Dubois, E. (1981). Recovering the orientation and the position of a rigid body in space from a single view. Technical report, INRS-Telecommunications, Quebec, Canada.

[R65] Roberts, L. G. (1965). Machine perception of three-dimensional solids. In et al, T., editor, *Optical and Electro-Optical Information Processing*. M.I.T. Press, Cambridge, MA.

[S76] Strang, G. (1976). *Linear algebra and its applications.* Harcourt Brace Jovanovich, Orlando, Florida.

[TM87] Thompson, D. W. and Mundy, J. L. (1987). Three-dimensional model matching from an unconstrained viewpoint. In *Proceedings of IEEE Conference on Robotics and Automation*, pages 208–220, Raleigh, NC.

[T87] Tsai, R. (1987). A versatile camera calibration technique for high-accuracy 3d machine vision metrology using off-the-shelf tv cameras and lenses. *IEEE J. of Robotics and Automation*, RA-3(4):323–344.

[U89] Ullman, S. (1989). Aligning pictorial descriptions: an approach to object recognition. *Cognition*, 32:193–254.

[UB91] Ullman, S. and Basri, R. (1991). Recognition by linear combinations of models. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 13:992–1006.

[W93] Weinshall, D. (1993). Model-based invariants for 3D vision. *International Journal on Computer Vision*, 10(1):27–42.

[WT95] Weinshall, D. and Tomasi, C. (1995). Linear and incremental acquisition of invariant shape models from image sequences. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 17(5):512–517.

[Y89] Yuan, J. S. C. (1989). A general photogrammetric method for determining object position and orientation. *IEEE Trans. on Robotics and Automation*, 5(2):129–142.