Similarity and Affine Distance Between Point Sets

Michael Werman and Daphna Weinshall *
Institute of Computer Science
The Hebrew University of Jerusalem
91904 Jerusalem, Israel
email: werman@cs.huji.ac.il

Abstract

We develop the expressions for measuring the distance between $2D$ point sets, which are invariant to either $2D$ affine transformations or $2D$ similarity transformations of the sets, and assuming a known correspondence between the point sets. We discuss the image normalization to be applied to the images before their comparison so that the computed distance is symmetric with respect to the two images. We then give a general (metric) definition of the distance between images, which leads to the same expressions for the similarity and affine cases. This definition avoids *ad-hoc* decisions about normalization. Moreover, it makes it possible to compute the distance between images under different conditions, including cases where the images are treated asymmetrically. We demonstrate these results with real and simulated images.

1 Background

When comparing images to other images or models, one would like to somehow cancel camera transformations. In general there is no way to normalize images of 3D objects so that all projections are equivalent (other than a normalization that makes *all* images equivalent). However, under the weak perspective (scaled orthographic) projection model assumed here, it is possible to remove the effects of certain camera transformations, such as rotations about the optical axis and translations.

More specifically, there exist standard methods of image normalization with respect to the following image transformations [3]:

Translations: the image is shifted so that its centroid is at the origin.

Rotations: the image is rotated so that its principal axis has some standard orientation.

Normalization with respect to rotations can be replaced by normalization with respect to linear transformations, or normalization by moments, where an image is transformed with a linear transformation so that its second moments have given values. This method is related to the Whitening transformation, a linear transformation of data which transforms its covariance matrix into the unit matrix [1, 4]. This transformation does not preserve Euclidean distances.

*This research was sponsored by the U.S. Office of Naval Research under Grant N00014-93-1-1202, R&T Project Code 4424341—01.

In Section 2 we give the expression for the distance between images up to $2D$ similarity transformations, which uses normalization with scale and rotations, and the expression for the distance up to $2D$ affine transformations, which uses normalization by moments. In Section 3 we show a different way to look at the diference betwen two point sets, which entails looking at all the possible transformations of a point set as a single object and measuring the distance between these objects. This approach defines distance measures between images that have metric properties. We show that both defintions give the same expressions in the similarity and affine cases. Finally, Section 4 contains examples with real and simulated images.

2 Image normalization and comparison

We assume here objects composed of n three dimensional fiducial points. An image of the object is obtained by a rigid transformation (of the object or the camera), followed by weak perspective (or scaled orthographic) projection[1] from three dimensional space to the two dimensional image.

An image is a set of n image points $\{(x_i, y_i)\}_{i=1}^n$. An equivalent representation of the image is the $2 \times n$ matrix $\mathbf{P}$, whose i-th column is the image coordinates of the i-th feature of the object. The use of matrix $\mathbf{P}$ to represent an image of an object implies a correspondence between the image features and the object features, where different correspondences lead to permutations of the matrix's columns.

Given two images, or the two matrices $\mathbf{P}$ and $\mathbf{Q}$, the question of comparing them is equivalent to matrix comparison. We are using the "usual" metric, which is the Frobenius norm of the difference matrix, and which is the same as the Euclidean distance between points in the images:

$$\|\mathbf{P} - \mathbf{Q}\|_F^2 = \sum \|\mathbf{P}[i,j] - \mathbf{Q}[i,j]\|^2 = tr[(\mathbf{P} - \mathbf{Q}) \cdot (\mathbf{P} - \mathbf{Q})^T] \tag{1}$$

Henceforth we will omit the subscript F, and an unsubscripted matrix norm will be the Frobenius norm.

Before taking the norm of the difference between the images, we want to remove differences which are due to irrelevant effects, such as the size of the image (which is arbitrary under scaled orthography) or the exact location of the object (e.g., due to an arbitrary translation and rotation of the object in the image). The following two operations, which remove irrelevant effects, are easy to do:

Normalization: the size of the images is normalized to some standard size.

Image alignment: a $2D$ alignment transformation, taken from a group of $2D$ transformations which includes $2D$ rotations, translations and scale, is applied to one image to obtain optimal alignment with the other image.

The **normalization** stage is intended to guarantee that the distance between two images, defined in Eq (1), is symmetric in the sense that we get the same distance when, in the alignment

[1]Weak perspective is an approximation to the pinhole perspective projection, which assumes that the field of view is not too large and therefore higher order terms in the size/distance ratio (the ratio between the size of the object and its distance to the camera) can be neglected. In this case, a world point whose coordinates in the camera coordinate system are (x, y, z) is projected to the image coordinates $s(x, y)$ for some scalar s, which is the same for all of the points.

stage, one image is aligned with another image or vice versa. The normalization enables us, therefore, to compare between different images of the *same* object and of *different* objects, since distances are always measured in a normalized frame.

For **alignment** we consider two groups of 2D transformations: the similarity group, which includes $2D$ rotations, translations, and scale, and the affine group, which includes $2D$ linear transformations and translations. An alignment with a similarity transformation is necessary, since under weak perspective projection, images that differ by image rotation or translation can be obtained from the same object, and should therefore be considered the *same image*.

Since the group of affine transformations includes the similarity group as a sub-group, the alignment with affine transformation is more general: it makes images that differ by $2D$ rotation and scale appear the same as needed, but it also makes images of different objects appear the same. It therefore leads to false identifications of different images as the same image. On the other hand, for planar objects, the affine alignment makes *all* the images of an object (from all viewpoints) appear the same (under the weak perspective assumption), which is advantageous when planar objects are expected. Both measures are therefore useful, and we discuss both possibilities here.

2.1 The similarity measure

Let us define the normalization, alignment, and comparison operations described above, where the alignment is done with a $2D$ similarity transformation, as applied to the matrix representations of two images $\mathbf{P}$ and $\mathbf{Q}$. We assume w.l.g. that the images are centered on their centroid, so that their first moments are 0 (this turns out to be the optimal translation when measuring distance by sum of square distances). To accomplish this normalization, we use a $2D$ translation. Therefore the remaining free components of the similarity transformation are a $2D$ rotation and scale.

Scale normalization:

$$\mathbf{P}' = \frac{\mathbf{P}}{||\mathbf{P}||}, \quad \mathbf{Q}' = \frac{\mathbf{Q}}{||\mathbf{Q}||} \tag{2}$$

Alignment: w.l.g. we align image $\mathbf{P}'$ with $\mathbf{Q}'$ using a scaled rotation

$$\mathbf{P}'' = sR \cdot \mathbf{P}'$$

where s is a scalar and R is a 2×2 rotation matrix

$$R = \begin{pmatrix} \cos(\mu) & \sin(\mu) \\ -\sin(\mu) & \cos(\mu) \end{pmatrix}$$

To accomplish optimal alignment, we choose μ and s which obtain

$$||sR\mathbf{P}' - \mathbf{Q}'||^2 = \min_{s',R'} ||s'R'\mathbf{P}' - \mathbf{Q}'||^2$$

This can be solved by differentiating the distance expression, $tr[(sR\mathbf{P}' - \mathbf{Q}') \cdot (sR\mathbf{P}' - \mathbf{Q}')^T]$, with respect to μ and s, and equating the partial derivatives to 0. Having done that, we get[2]:

$$s \;\; = \;\; \sqrt{rt^2[\mathbf{P}'(\mathbf{Q}')^T] + tr^2[\mathbf{P}'(\mathbf{Q}')^T]}$$

[2] $tr[]$ of a matrix returns the sum of its diagonal elements. In analogy, and for simplicity of presentation, we use the complementary operator $rt[]$; $rt[]$ of a 2×2 matrix returns the *difference* between its *non-diagonal* elements.

$$\tan\mu \quad = \quad \frac{rt[\mathbf{P}'(\mathbf{Q}')^T]}{tr[\mathbf{P}'(\mathbf{Q}')^T]}$$

Therefore the $2D$ alignment transformation is:

$$A = sR = \begin{pmatrix} tr[\mathbf{P}'(\mathbf{Q}')^T] & rt[\mathbf{P}'(\mathbf{Q}')^T] \\ -rt[\mathbf{P}'(\mathbf{Q}')^T] & tr[\mathbf{P}'(\mathbf{Q}')^T] \end{pmatrix} \tag{3}$$

Comparison: The similarity distance between images $\mathbf{P}$ and $\mathbf{Q}$ is

$$D^2_{sim}(\mathbf{P},\mathbf{Q}) = \|A\mathbf{P}' - \mathbf{Q}'\|^2$$

for $\mathbf{P}', \mathbf{Q}'$ defined in Eq (2) and A defined in Eq (3).

In order to express $D_{sim}(\mathbf{P},\mathbf{Q})$ directly in term of $\mathbf{P},\mathbf{Q}$ (rather than $\mathbf{P}',\mathbf{Q}'$), the similarity measure can be shown to be equal to

$$D^2_{sim}(\mathbf{P},\mathbf{Q}) = \frac{1}{h}\|sR\mathbf{P} - \mathbf{Q}\|^2 = \frac{1}{h}tr[(sR\mathbf{P} - \mathbf{Q}) \cdot (sR\mathbf{P} - \mathbf{Q})^T]$$

where

$$\begin{aligned} s &= \frac{\sqrt{rt^2[\mathbf{P}\mathbf{Q}^T] + tr^2[\mathbf{P}\mathbf{Q}^T]}}{tr[\mathbf{P}\mathbf{P}^T]} \\ \tan\mu &= \frac{rt[\mathbf{P}\mathbf{Q}^T]}{tr[\mathbf{P}\mathbf{Q}^T]} \\ h &= tr[\mathbf{Q}\mathbf{Q}^T] \end{aligned}$$

This expression can be simplified as follows:

$$D^2_{sim}(\mathbf{P},\mathbf{Q}) = \frac{1}{h}tr[(A\mathbf{P} - \mathbf{Q}) \cdot (A\mathbf{P} - \mathbf{Q})^T]$$

where

$$\begin{aligned} A &= \frac{1}{tr[\mathbf{P}\mathbf{P}^T]}[\mathbf{Q}\mathbf{P}^T + det(\mathbf{P}\mathbf{Q}^T)(\mathbf{P}\mathbf{Q}^T)^{-1}] \\ h &= tr[\mathbf{Q}\mathbf{Q}^T] = \|\mathbf{Q}\|^2 \end{aligned}$$

(Note that A is an orthogonal, but not orthonormal, 2×2 matrix.)

Additional simplifications give us the final form:

$$D^2_{sim}(\mathbf{P},\mathbf{Q}) \quad = \quad 1 - \frac{\|\mathbf{Q}\mathbf{P}^T\|^2 + 2det(\mathbf{Q}\mathbf{P}^T)}{\|\mathbf{P}\|^2\|\mathbf{Q}\|^2} \tag{4}$$

2.2 The affine measure

We now define the normalization, alignment, and comparison operations for an alignment done with a $2D$ affine transformation, as applied to the matrix representations of the two images $\mathbf{P}$ and $\mathbf{Q}$. Once again, we can assume w.l.g. that the images are centered on their centroid, so that their first moments are 0. To accomplish this normalization, we use a $2D$ translation. Therefore the remaining free component of the affine transformation is a $2D$ linear transformation.

Moment normalization:

$$\mathbf{P}' = \mathbf{S}_p\mathbf{P}, \quad \mathbf{Q}' = \mathbf{S}_q\mathbf{Q} \tag{5}$$

where $\mathbf{S}_p$ and $\mathbf{S}_q$ are 2×2 invertible matrices such that $\mathbf{P}'(\mathbf{P}')^T = \mathbf{Q}'(\mathbf{Q}')^T = I$, and I denotes the 2×2 unity matrix.

Alignment: w.l.g. we align image $\mathbf{P}'$ with $\mathbf{Q}'$ with a linear transformation

$$\mathbf{P}'' = A\mathbf{P}'$$

where A is a 2×2 invertible matrix. To accomplish optimal alignment, we choose A which obtains

$$\|A\mathbf{P}' - \mathbf{Q}'\|^2 = \min_A \|A\mathbf{P}' - \mathbf{Q}'\|^2$$

The matrix A which obtains the above least square distance is the pseudo-inverse[3]:

$$A = \mathbf{Q}'(\mathbf{P}')^+ = \mathbf{Q}'(\mathbf{P}')^T(\mathbf{P}'(\mathbf{P}')^T)^{-1} = \mathbf{Q}'(\mathbf{P}')^T$$

Comparison: The affine distance between images $\mathbf{P}$ and $\mathbf{Q}$ is

$$D^2_{aff}(\mathbf{P},\mathbf{Q}) = \|A\mathbf{P}' - \mathbf{Q}'\|^2$$

for $\mathbf{P}', \mathbf{Q}'$ defined in Eq (5) and $A = \mathbf{Q}'(\mathbf{P}')^T$.

In order to express $D_{aff}(\mathbf{P},\mathbf{Q})$ directly in term of $\mathbf{P},\mathbf{Q}$ (rather than $\mathbf{P}',\mathbf{Q}'$), the affine measure can be shown to be equal to

$$D^2_{aff}(\mathbf{P},\mathbf{Q}) = tr[(\mathbf{Q}\mathbf{Q}^T)^{-1}(A\mathbf{P} - \mathbf{Q}) \cdot (A\mathbf{P} - \mathbf{Q})^T]$$

where

$$A = \mathbf{Q}\mathbf{P}^+$$

Additional simplifications give us the final form:

$$D^2_{aff}(\mathbf{P},\mathbf{Q}) \quad = \quad 2 - tr(\mathbf{P}^+\mathbf{P} \cdot \mathbf{Q}^+\mathbf{Q}) \tag{6}$$

[3]Given a $m \times n$ matrix Y where $m \leq n$, Y^+ denotes its pseudo-inverse where $Y^+ = Y^T(YY^T)^{-1}$.

3 A metric definition of image distance

In the previous section we showed how to normalize point sets in order to compare them in a "good" manner. In this section we give a different definition for the comparison between images, which gives a metric interpretation to the expressions developed above. The reason for bringing this new definition is that it gives another, maybe cleaner, way at looking at the comparison problem. It gives the same answers as the normalization procedure described above but in a simpler form. This lets us use known results to show that the distance defined is really a metric, and it allows us to compare images which undergo transformations from different classes.

We first identify an image, which is a set of k planar points $\{(x_1, y_1), (x_2, y_2), \ldots, (x_k, y_k)\}$, with a single point in R^{2k}: $\mathbf{p} = (x_1, x_2, \ldots, x_k, y_1, y_2 \ldots, y_k)$. $\mathbf{p}$ is a vector representation of the image, containing the concatenation of the rows of the matrix representation $\mathbf{P}$. In this space, all the similarity transformations of the image can be identified with a four-dimensional subspace of R^{2k} that includes $\mathbf{p}$, where the four parameters correspond to scale, rotation, and translation in the x and y directions. Similarly, all the affine transformations of the image can be identified with a six-dimensional subspace of R^{2k} that includes the origin and point $\mathbf{p}$. Thus a natural distance between images matched up to similarity is the distance between their corresponding four-dimensional subspaces, and the distance between images matched up to affine transformation is the distance between their corresponding six-dimensional subspaces.

We use the following distance between subspaces [2]: a subspace of $\mathcal{R}^{2k}$ can be identified with the matrix that corresponds to the linear operator of orthogonal projection into that subspace. The distance between subspaces is the norm of the difference between their corresponding projection matrices. A projection matrix can be built from a set of basis vectors, which span the subspace, as follows:

1. the basis is orthonormalized;

2. as intermediate matrix C is constructed whose columns are the orthonormal basis vectors;

3. CC^T is the projection matrix.

The computation for the similarity case is as follows: if another image $\mathbf{S}$ is a similarity transformation of $\mathbf{P}$, then there exist four parameters a, b, e, f such that (using the matrix representation of the images)

$$\mathbf{S} = \begin{pmatrix} a & b \\ -b & a \end{pmatrix} \mathbf{P} + \begin{pmatrix} e \\ f \end{pmatrix} \tag{7}$$

We define four vectors in $\mathcal{R}^{2k}$

$$\begin{aligned} \mathbf{v}_1 &= (x_1, x_2, \ldots, x_k, y_1, y_2, \ldots, y_k) \\ \mathbf{v}_2 &= (y_1, y_2, \ldots, y_k, -x_1, -x_2, \ldots, -x_k) \\ \mathbf{v}_3 &= (1, 1, \ldots, 1, 0, 0, \ldots, 0) \\ \mathbf{v}_4 &= (0, 0, \ldots, 0, 1, 1, \ldots, 1) \end{aligned}$$

Changing Eq. (7) to vector representation, we get

$$\mathbf{s} = a\mathbf{v}_1 + b\mathbf{v}_2 + e\mathbf{v}_3 + f\mathbf{v}_4.$$

Therefore the four vectors $\{\mathbf{v}_i\}_{i=1}^4$ span the subspace of all images obtained from $\mathbf{p}$ with a similarity transformation. We use this basis to construction the projection matrix $\mathbf{V}$ as described above.

Let $\mathbf{V_p}$ denote the matrix corresponding to the orthogonal projection into the subspace of image $\mathbf{p}$, and similarly $\mathbf{V_q}$ denotes the projection matrix for image $\mathbf{q}$. The distance between $\mathbf{p}$ and $\mathbf{q}$ is

$$D_{sim}(\mathbf{P},\mathbf{Q}) = ||\mathbf{V_p} - \mathbf{V_q}||$$

(here we use again the Frobenius norm of the difference).

The computation for the affine case is the same as in the similarity case, except that the subspace corresponding to $\mathbf{p}$ is spanned by the six vectors:

$$\begin{aligned}
\mathbf{w}_1 &= (x_1, x_2, \ldots, x_k, 0, 0, \ldots, 0)\\
\mathbf{w}_2 &= (y_1, y_2, \ldots, y_k, 0, 0, \ldots, 0)\\
\mathbf{w}_3 &= (0, 0, \ldots, 0, x_1, x_2, \ldots, x_k)\\
\mathbf{w}_4 &= (0, 0, \ldots, 0, y_1, y_2, \ldots, y_k)\\
\mathbf{w}_5 &= (1, 1, \ldots, 1, 0, 0, \ldots, 0)\\
\mathbf{w}_6 &= (0, 0, \ldots, 0, 1, 1, \ldots, 1)
\end{aligned}$$

This is because an image $\mathbf{S}$ in the subspace corresponding to all the affine transformations of image $\mathbf{P}$ can be written as

$$\mathbf{S} = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \mathbf{P} + \begin{pmatrix} e \\ f \end{pmatrix}$$

for 6 parameters a, b, c, d, e, f. Changing to vector representation, we once again get

$$\mathbf{s} = a\mathbf{w}_1 + b\mathbf{w}_2 + c\mathbf{w}_3 + d\mathbf{w}_4 + e\mathbf{w}_5 + f\mathbf{w}_6$$

The projection matrix $\mathbf{W_p}$ of the affine subspace of $\mathbf{p}$ can be built from the basis $\{\mathbf{w}_i\}_{i=1}^6$ as described above. The affine distance between points $\mathbf{P}$ and $\mathbf{Q}$ is:

$$D_{aff}(\mathbf{P},\mathbf{Q}) = ||\mathbf{W_p} - \mathbf{Wq}||$$

Note that if the vector $\mathbf{p}$ is normalized so that the sum of the x's is zero and the sum of the y's is zero, then there is no need to include the vectors $(1, 1, \ldots, 1, 0, 0, \ldots, 0)$ and $(0, 0, \ldots, 0, 1, 1, \ldots, 1)$ in the basis of the similarity and affine subspaces. This leaves a two-dimensional sub-space for the similarity case, and a four-dimensional subspace for the affine case.

It can be shown that the above expressions for D_{aff} and D_{sim} are exactly twice the expressions obtained in the previous section given in Eqs (4),(6). Note, however, that the present definition is a metric, and we can therefore conclude that the affine and similarity distances are in fact metrics, defining two metric spaces on the space of all images.

The general definition used here enables us to obtain more than just rederivations of the results of the previous section. For example, we match two images, one of which is allowed to go similarity transformations and the other is allowed affine transformations. This is done by again comparing their respective projection matrices. The squared distance between $\mathbf{P}$, where $\mathbf{P}$ can undergo affine transformations, and $\mathbf{Q}$, where $\mathbf{Q}$ can undergo similarity transformations, is

$$1 - \frac{tr(\mathbf{Q}\mathbf{Q}^T\mathbf{P}\mathbf{P}^+)}{||\mathbf{Q}||^2}$$

(where the points are centered on their centroid).

4 Examples:

Figure 1: An image of a toy tiger.

Fig. 2 shows simulated images of a cube, aligned with the optimal $2D$ similarity and affine transformations.

Fig. 1 shows a real reference image of a toy tiger. Three additional images of the tiger are shown in Fig. 3. We used ears, eyes, knees, tail and nose as features. Next to each of the three images in Fig. 3 we show a scaled rotated version of it, which is the same image rotated and scaled by the optimal similarity transformation that aligns it with the reference image given in Fig. 1.

similarity distance	affine distance
0.124	0.412
0.206	0.822
0.341	0.854

Table 1: The similarity and affine distances between three images of a tiger, shown in Fig. 3, to a reference image shown in Fig. 1.

Table 1 summarizes the similarity and affine distances between the images in Fig. 3 to the image in Fig. 1. The images in Fig. 3 were taken increasingly further away from the image in Fig. 1, as measured by the distance between them on the viewing sphere. The distances in Table 1 demonstrate that both the affine and similarity distances increase with the inter-view distance on the viewing sphere. Note that the similarity distances are larger than the affine distances, as expected.

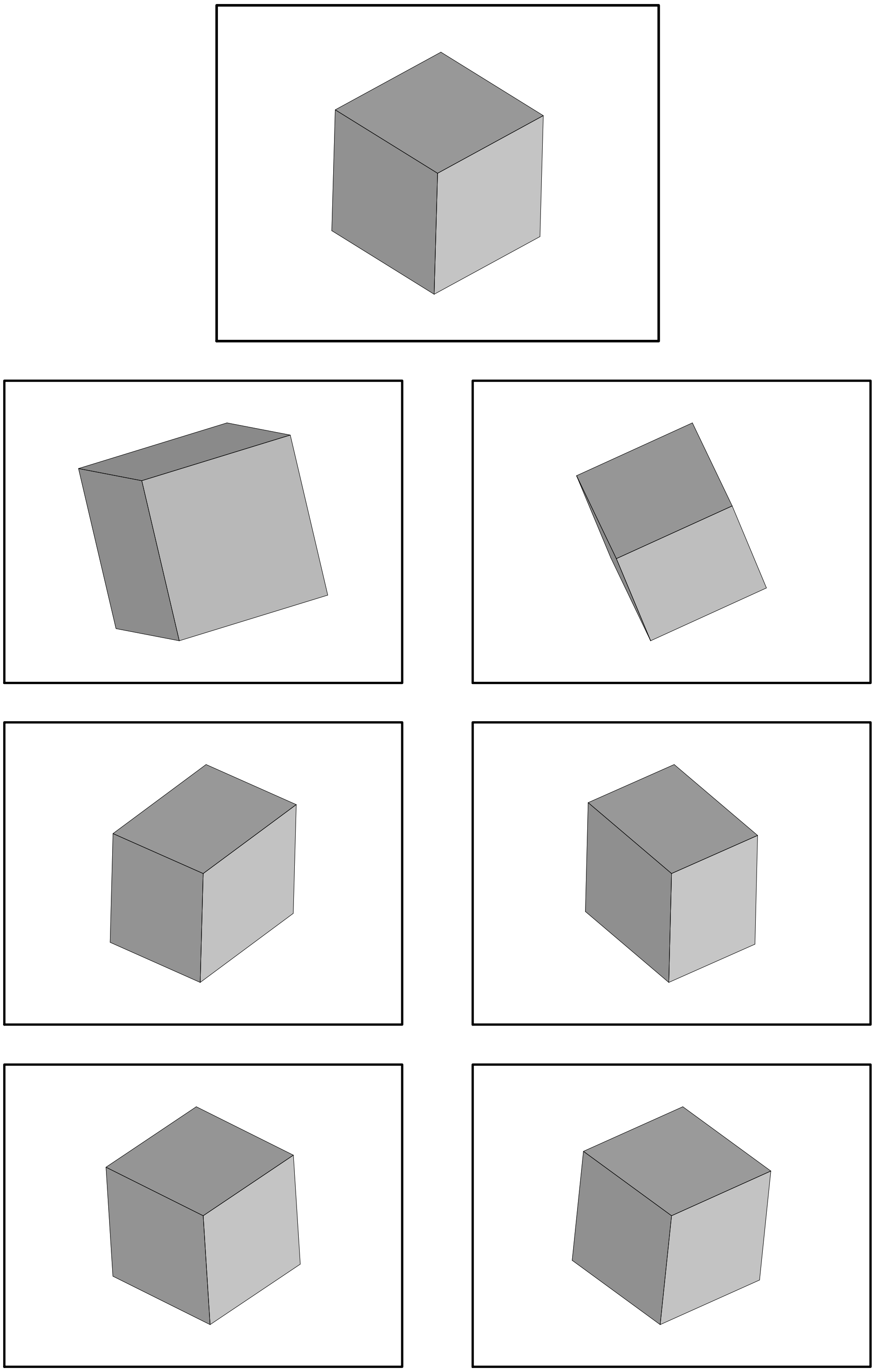

Figure 2: 1st row - a reference image of a cube, to be aligned with other images; 2nd row - two other images of a cube; 3rd row - the reference image aligned with the images in the second row using the optimal affine transformation; 4th row - the reference image aligned with the images in the second row using the optimal similarity transformation. The left and right columns depict two different examples.

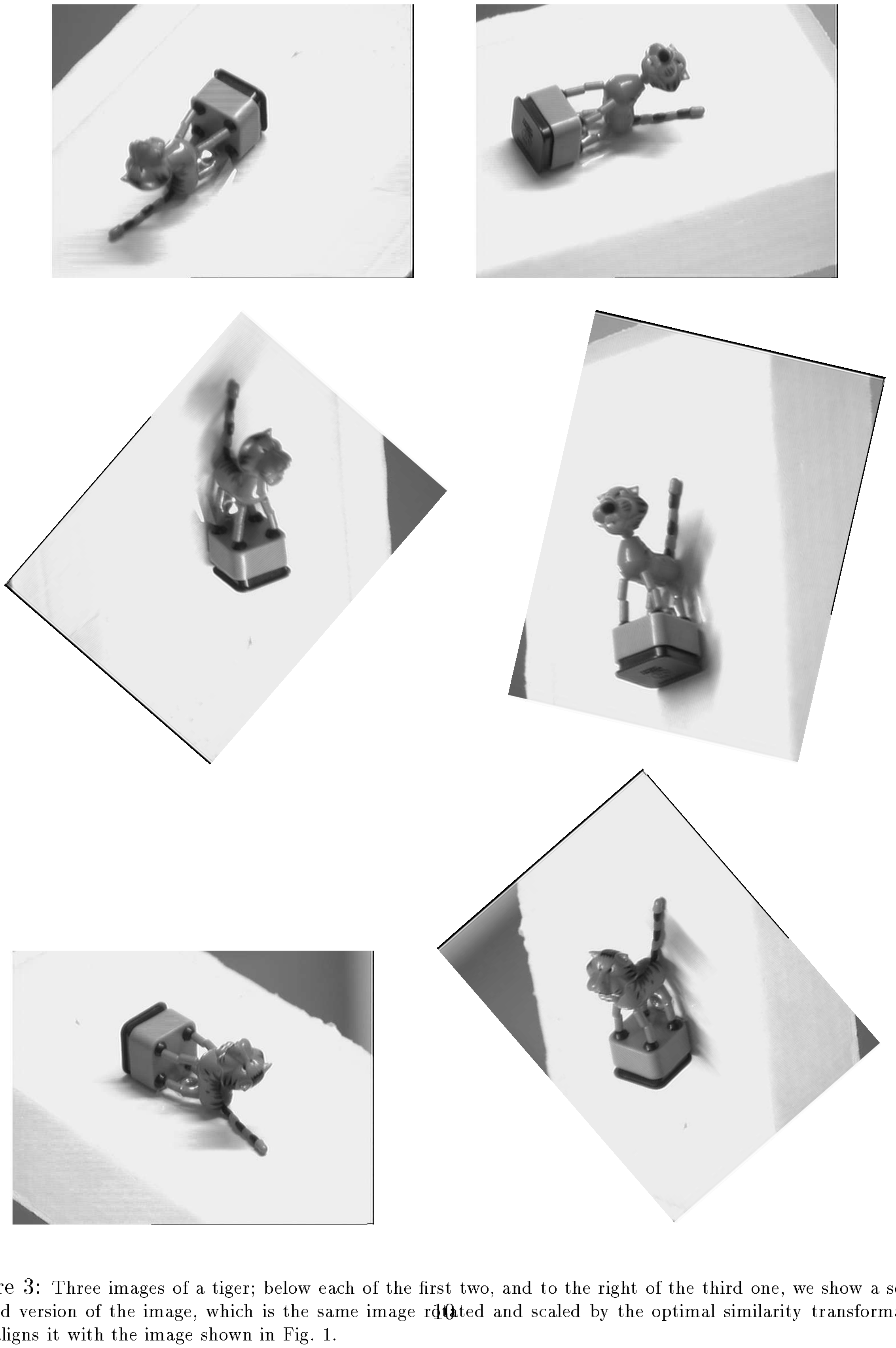

Figure 3: Three images of a tiger; below each of the first two, and to the right of the third one, we show a scaled rotated version of the image, which is the same image rotated and scaled by the optimal similarity transformation that aligns it with the image shown in Fig. 1.

References

[1] K. Fukunaga. *Introduction to statistical pattern recognition.* Academic Press, Boston, 1990.

[2] G. H. Golub and C. F. Van Loan. *Matrix Computations.* Johns Hopkins University Press, Baltimore, 1989.

[3] A. Rosenfeld and A. C. Kak. *Digital picture processing.* Academic Press, Boston, 1982.

[4] J. Sprinzak and M. Werman. Affine point matching. *Pattern Recognition Letters*, 1993.