Covering the Space of Tilts: Application to affine invariant image comparison
Mariano Rodríguez, Julie Delon, Jean-Michel Morel

To cite this version:

HAL Id: hal-01589522
https://hal.archives-ouvertes.fr/hal-01589522v2
Submitted on 19 Feb 2018

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L’archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d’enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.
Covering the Space of Tilts.
Application to affine invariant image comparison
Mariano Rodríguez, Julie Delon and Jean-Michel Morel
December 2017

Abstract
We propose a mathematical method to analyze the numerous algorithms performing Image Matching by Affine Simulation (IMAS). To become affine invariant they apply a discrete set of affine transforms to the images, previous to the comparison of all images by a Scale Invariant Image Matching (SIIM), like SIFT. Obviously this multiplication of images to be compared increases the image matching complexity. Three questions arise: a) what is the best set of affine transforms to apply to each image to gain full practical affine invariance? b) what is the lowest attainable complexity for the resulting method? c) how to choose the underlying SIIM method? We provide an explicit answer and a mathematical proof of quasi-optimality of the solution to the first question. As an answer to b) we find that the near-optimal complexity ratio between full affine matching and scale invariant matching is more than halved, compared to the current IMAS methods. This means that the number of key points necessary for affine matching can be halved, and that the matching complexity is divided by four for exactly the same performance. This also means that an affine invariant set of descriptors can be associated with any image. The price to pay for full affine invariance is that the cardinality of this set is around 6.4 times larger than for a SIIM.

1 Introduction
Image matching, which consists in detecting shapes common to two images, is a crucial issue for a large number of computer vision applications, such as scene recognition [60, 10, 51] and detection [15, 48], object tracking [65], robot localization [52, 59, 45], image stitching [2, 9], image registration [63, 32] and retrieval [18, 17], 3D modeling and reconstruction [14, 16, 61, 1], motion estimation [62], photo management [54], symmetry detection [34] or even image forgeries detection [13]. The problem has implementation variants depending on the set up. If for example the user knows that both compared images are related, the focus is on detecting the most reliable common set of shape descriptors. In the detection set up, an image is compared to a database of images and the question is to retrieve related images in the database. This is for example crucial for performing video search [55]. Local shape descriptors must be extracted for this purpose, and this description should be as invariant as possible to viewpoint changes and of course as sparse as possible. In our discussion we will most of the time refer to the simpler set up where two images are being compared. But the reduction of the number of descriptors is of course still more important for comparing an image to an image database as initially proposed in [53]. In this last reference, large sets of descriptors are sparsified by clustering techniques. This only indicates how important it is to reduce as much as possible the set of affine descriptors of each image.

Detectors, descriptors and affine invariance Given a query image of some physical object and a set of target images, the first goal of image matching is to decide if these target images contain a view of the same object. If the answer is positive, image matching aims at localizing this object in these target images. Deciding if the object is present is difficult and becomes especially tricky for large image databases, for which the control of false matches is crucial. Another difficulty of the matching problem comes from the change of camera viewpoints between images. In order to cope
with these viewpoint changes, the whole matching process should be as invariant as possible to the resulting image deformations. As we shall develop, this requires affine invariance for the recognition process.

The classical approach to image matching consists in three steps: detection, description and matching. First, keypoints are detected in the compared images. Second, regions around these points are described and encoded in local invariant descriptors. Finally, all these descriptors are compared and possibly matched. Using local descriptors yields robustness to context changes. Both the detection and description steps are usually designed to ensure some invariance to various geometrical or radiometric changes.

Local image point detectors are always translation invariant. While the venerable Harris point detector [19] is only invariant to translations and rotations, the Harris-Laplace [36], Hessian-Laplace [38] or DoG (Difference-of-Gaussian) region detectors [33] are invariant to similarity transformations, i.e. translations, rotations and scale changes. To ensure invariance to affine transforms, some authors have proposed moment-based region detectors [28, 6] including the Harris-Affine and Hessian-Affine region detectors [37, 38]. Locally affine invariant region detectors can also be based on edges [58, 57], intensity [56, 57], or entropy [21]. Finally, the detectors MSER (“maximally stable extremal region”) [35] and LLD (“level line descriptor”) [46, 47, 12] both rely on level lines. Yet the affine invariance of these detectors is limited by the fact that optical blur and affine transforms do not commute, as shown in [44]. Level line based detectors like MSER therefore are not fit to handle scale changes. Indeed, they do not take into account the effect of blur on the level line geometry [12].

In the last 15 years, numerous invariant image descriptors have been proposed in the literature, but the most well-known and the most widely used remains the scale-invariant feature transform (SIFT), introduced by Lowe in his landmark paper [33]. SIFT makes use of a DoG region detector. It is fully invariant to similarities (see [43] for a mathematical proof of this fact). Each SIFT descriptor is composed of histograms of gradient orientation around a key point, invariant to local radiometric changes and to geometrical image similarities. As a result, the SIFT method can be considered as partially invariant to illumination, fully invariant to geometrical similarities. But its success is certainly also due to its robustness to reasonable viewpoint changes.

The superiority of SIFT based descriptors has been demonstrated in several comparative studies [39, 42]. As a consequence, many variants of the SIFT descriptor have emerged, among which we can mention PCA-SIFT [23], GLOH (gradient location-orientation histogram) [39], SURF (speeded up robust features) [7] or RootSIFT [5]. The main claims of these variants are a lower complexity or a greater robustness to viewpoint changes. In the same vein, binary descriptors have also received much attention. Focusing on speed and efficiency, the BRIEF [11], BRISK [25] or LATCH [26] descriptors are compact and represented by sequences of bits, and can be compared more quickly than floating point descriptors like those used in SIFT. Descriptors based on nonlinear scale spaces, such as KAZE [3] or its accelerated version AKAZE [4], have also been proposed to locally adapt blur to the image data.

None of the previously mentioned state-of-the-art methods is fully affine invariant. The SIFT method does not cover the whole affine space and its performance drops under substantial viewpoint changes. SIFT and the other aforementioned descriptors cannot cope with viewpoint differences larger than 60° for planar objects [44, 40], and are still usable but much less efficient for angles larger than 45° [22]. We shall give and use here concrete measurements of their resilience to view angle changes.

To overcome this limitation, several simulation-based solutions have been recently proposed. The core idea of these algorithms, that we choose to call by the generic term IMAS (Image Matching by Affine Simulation), is to simulate a set of views from the initial images, by varying the camera orientation parameters. These simulations allow to capture far stronger viewpoint angles than standard matching approaches, up to 88°. Among those IMAS algorithms, we can mention ASIFT [64], FAIR-SURF [49] and MODS [40].

A first suggestion to simulate affine distortions before applying a SIIM (Scale Invariant Image Matching) appeared in [50] where the authors proposed to simulate two tilts and two shear deformations followed by SIFT in a cloth motion capture application. As argued in [64, 40, 49], if a physical object has a smooth or piecewise smooth boundary, its views obtained by cameras in different positions undergo smooth apparent deformations. These regular deformations are locally well approximated by affine transforms of the image plane. By focusing on local image descriptors, the changes of aspect of objects can therefore be modeled by affine image deformations.

The problem of constructing affine invariant image descriptors by using an affine Gaussian scale
space, that is equivalent to simulating affine distortions followed by the heat equation, has a long story starting with [20, 8, 27, 28]. The idea of affine shape adaptation underlying one of the methodologies for achieving affine invariance, was then in turn used as a base for the work on affine invariant interest points and affine invariant matching in [28, 6, 37, 38, 58, 57, 56]. The notion of an affine invariant reference frame was further developed in [30, 31]. Nevertheless, to the best of our knowledge, the direct constructions of affine invariant descriptors as fixed points for an iterative affine normalization process have never found a mathematical justification.

The first IMAS method provided with a mathematical proof of affine invariance is ASIFT [44, 64]. The authors of this paper proposed it as an affine invariant extension of SIFT and proved it to be fully affine invariant in a continuum model. The structure of ASIFT is generic in the sense that it can be implemented with any local descriptor, provided this descriptor has a robustness to viewpoint changes similar to SIFT descriptors. Unlike MSER, LLD, Harris-Affine and Hessian-Affine, which attempt at normalizing all of the six affine parameters, ASIFT simulates three parameters and normalizes the rest. More specifically, ASIFT simulates the two camera axis parameters, and then applies SIFT which simulates the scale and normalizes the rotation and the translation. Of the six parameters required for affine invariance, three are therefore simulated and three normalized.

Two recent successful methods follow the same affine simulation path. FAIR-SURF [49] combines the affine invariance of ASIFT and the efficiency of SURF. The MODS image comparison algorithm introduced in [40] also relies on this principle and affine simulations are generated on-demand if needed in the process of comparing two images. MODS employs a combination of different detectors when comparing images. It outperforms state-of-the-art image comparison approaches both in affine robustness and speed.

Other IMAS approaches without local descriptors have also been put up for template matching. FasT-Match [24] delivers affine invariance by assuming that the template (a patch in the query image) can be recovered inside the target image by a unique affine map. Meaning there is no subjacent projective map to identify. Contrary to IMAS with local descriptors, the six required parameters to attain affine invariance are simulated instead of three of the present paper.

In this paper, we are interested in generic IMAS algorithms based on local descriptors and in their geometric optimization. In order to measure the degree of viewpoint change between different views of the same scene, we draw on the concept of absolute and relative transition tilts, previously introduced in [44, 64], and we illustrate why simulating large tilts on both compared images is necessary to obtain a fully affine invariant recognition. Indeed, transition tilts can in practice be much larger than absolute tilts, since they may behave like the square of absolute tilts.

The key question of IMAS methods is how to choose the list of affine transforms applied to the images before comparison. This list should be as short as possible to limit the computing time. But it should also sample the widest possible range of affine transforms. As we shall see, this question is closely related to the question of finding optimal coverings of the space of affine tilts. This question is formalized and solved in Section 2, where we find nearly optimal coverings. Section 3 applies this result to IMAS algorithms. It first presents a complete mathematical theory of IMAS algorithms, proving that they are fully affine invariant under the assumption that the underlying SIIM has a (quantifiable) limited affine invariance. Section 4 gives an experimental validation. It starts by measuring the exact extent of affine invariance for several SIIMs and deduces the corresponding complexity required to attain full affine invariance from each. Section 5 is a conclusion.

2 The space of affine tilts

In this section, we introduce the space of tilts for planar affine transforms, and we look for optimal coverings of this space. Optimal coverings will be used in the next section to define an optimal discrete set of affine transformations as the basis for IMAS algorithms. The rest of this section can be read as a sequence of purely geometric results. However, the reader might prefer to keep in mind that the affine transforms considered here can be interpreted as different viewpoints of a camera, or more generally as the transition from an image taken from a viewpoint to an image taken from another viewpoint. Indeed, given a frontal snapshot of a planar object \( u(x) = u(x, y) \), we can transition from any affine view \( Bu \) of the same object to any other affine view \( Au \) through the affine transformation \( AB^{-1} \). This requires some notation. For any linear invertible map \( A \in GL^+(2) \), we denote the affine transform \( A \) of a continuous image \( u(x) \) by \( Au(x) = u(Ax) \). We recall classic
notation for three subsets of the general linear group $GL(2)$ of invertible linear maps of the plane,

\[
GL^+(2) = \{ A \in GL(2) \mid \det(A) > 0 \},
\]

\[
GO^+ (2) = \{ A \in GL^+(2) \mid A \text{ is a similarity} \},
\]

\[
GL^+_*(2) = GL^+(2) \setminus GO^+ (2),
\]

where we call similarity any combination of a rotation and a zoom, and the symbol $\setminus$ denotes the set difference operator. Our central notion in the discussion is the \textit{tilt} of an affine transform, which we now define.

### 2.1 Absolute tilts

**Proposition 1** ([44]). Every $A \in GL^+_+(2)$ is uniquely decomposed as

\[
A = \lambda R_1 (\psi) T_t R_2 (\phi)
\]

where $R_1$, $R_2$ are rotations and $T_t = \begin{bmatrix} t & 0 \\ 0 & 1 \end{bmatrix}$ with $t > 1$, $\lambda > 0$, $\phi \in [0, \pi]$ and $\psi \in [0, 2\pi]$.

**Remark 2.** A similar decomposition to (1) was also presented in [29] for small deformations around the identity.

**Remark 3.** It follows from this proposition that any affine map $A \in GL^+_+(2)$ is either uniquely decomposed as in (1) or is directly expressed as a similarity $\lambda R_1$.

![Figure 1: Geometric Interpretation of (1)](image_url)

**Figure 1:** Geometric Interpretation of (1)

Figure 1 shows a camera viewpoint interpretation of this affine decomposition where the longitude $\phi$ and latitude $\theta = \arccos \frac{t}{1}$ characterize the camera’s viewpoint angles, $\psi$ parameterizes the camera spin and $\lambda$ corresponds to the zoom. In the ideal affine model, the camera is supposed to stand at infinite distance from a flat image $u$, so that the deformation of $u$ induced by the camera indeed is an affine map. But the above approximation is still valid provided the image’s size is small with respect to the camera distance. In other terms the affine model is locally valid for each small and approximately flat patch of a physical surface photographed by a camera at some distance. Yet, the affine deformation of the object’s aspect will be different for each of its patches. This explains why affine invariant recognition methods deal with local descriptors. The parameter $t$ defined above measures the so-called \textit{absolute tilt} between the frontal view and a slanted view. The uniqueness of the decomposition in (1) justifies the next definition.

**Definition 4.** We call \textit{absolute tilt} of $A$ the real number $\tau(A)$ defined by

\[
\begin{aligned}
GL^+(2) & \rightarrow [1, \infty] \\
A & \rightarrow \begin{cases}
1 & \text{if } A \in GO^+ (2) \\
t & \text{if } A \in GL^+_*(2)
\end{cases}
\end{aligned}
\]

where $t$ is the parameter found when applying Proposition 1 to $A$. 

Proposition 5. Let \( A \in GL^+(2) \). Then
\[
\tau(A) = \sqrt{\frac{\lambda_1}{\lambda_2}} = \|A\|_2 \|A^{-1}\|_2
\]
where \( \lambda_1 \geq \lambda_2 \) are the singular values of \( A \) and \( \| \cdot \|_2 \) is the usual Euclidean matrix norm.

Proof. The case of a similarity being straightforward, suppose that \( A \in GL^+(2) \). Then, using (1) we can re-write
\[
A = R_1 \begin{pmatrix} \gamma_1 & 0 \\ 0 & \gamma_2 \end{pmatrix} R_2
\]
where \( R_1, R_2 \) are two rotations and \( \gamma_1 \geq \gamma_2 > 0 \). So
\[
A^* A = R_2^t \begin{pmatrix} \gamma_1^2 & 0 \\ 0 & \gamma_2^2 \end{pmatrix} R_2
\]
whose eigenvalues are
\[
\lambda_1 = \gamma_1^2 \text{ and } \lambda_2 = \gamma_2^2
\]
but \( \gamma_1, \gamma_2 > 0 \) imply
\[
A = \sqrt{\frac{\lambda_1}{\lambda_2}} R_1 \begin{pmatrix} \sqrt{\frac{\lambda_1}{\lambda_2}} & 0 \\ 0 & 1 \end{pmatrix} R_2
\]
and finally \( \tau(A) = \sqrt{\frac{\lambda_1}{\lambda_2}} \). In addition, it is well known that
\[
\|A\|_2 = \sqrt{\rho(A^* A)} = \sqrt{\lambda_1},
\]
\[
\|A^{-1}\|_2 = \sqrt{\frac{\rho((AA^*)^{-1})}{\lambda_2}} = \frac{1}{\sqrt{\lambda_2}}
\]
where \( \rho(A^* A) \) is the largest eigenvalue of \( A^* A \), i.e., the largest singular value of \( A \). \qed

2.2 Transition Tilts

Image descriptors like those proposed in the SIFT method are invariant to translations, rotations and Gaussian zooms, which in terms of the camera position interpretation (see Figure 1) correspond to a fronto-parallel motion of the camera, a spin of the camera and to an optical zoom. We shall focus on the last part \( T_1 R_2 \) of the decomposition (1) because it is the one that is imperfectly dealt with by SIIMs. SIIMs are instead able to detect objects up to a similarity. This leads us to the next definition.

Definition 6. Let \( A, B \in GL^+(2) \). Then we define the right equivalence relation \( \sim \) as
\[
A \sim B \iff AB^{-1} \in GO^+(2).
\]

Remark 7. It is important to notice here that the right and left equivalence relations do differ because
\[
AB^{-1} \in GO^+ \Leftrightarrow B^{-1} A \in GO^+.
\]
For example, take
\[
A = T_2 R_\frac{\pi}{4} \text{ and } B^{-1} = R_\frac{\pi}{4} T_2,
\]
then
\[
AB^{-1} = 2R_\frac{\pi}{4} \in GO^+
\]
whereas
\[
B^{-1} A = R_\frac{\pi}{4} T_4 R_\frac{\pi}{4} \notin GO^+.
\]

Definition 8. Let \( A, B \in GL^+(2) \). We call transition tilt between \( A \) and \( B \) the absolute tilt of \( AB^{-1} \), i.e.
\[
\tau(AB^{-1}).
\]
Figure 2: Passage from transition tilts (left side) to absolute tilts (right side).

The transition tilt has an agreeable visual interpretation appearing in Figure 2. By Formula (1) applied to $AB^{-1}$, passing from an image $Bu$ to an image $Au$ comprises a single non-Euclidean transformation, namely the central tilt matrix $T_{\tau(AB^{-1})}$ which squeezes the image in the direction of $x$ after having rotated it. Thus the transition tilt measures the amount of image distortion caused by a change of view angle. We now state and give a brief proof of the formal properties of the transition tilt stated in [44].

**Proposition 9.** For $A, B \in GL^+(2)$ we have

1. $\tau\left(AB^{-1}\right) = 1 \Leftrightarrow A \sim B$;
2. $\tau(A) = \tau\left(A^{-1}\right)$;
3. $\tau\left(AB^{-1}\right) = \tau\left(BA^{-1}\right)$;
4. $\tau\left(AB^{-1}\right) \leq \tau(A)\tau(B)$;
5. $\max\left\{\frac{\tau(A)}{\tau(B)}, \frac{\tau(B)}{\tau(A)}\right\} \leq \tau\left(AB^{-1}\right)$.

**Proof.**

1) $\tau\left(AB^{-1}\right) = 1 \Leftrightarrow AB^{-1} = \lambda R \Leftrightarrow A = \lambda RB$

2) By proposition 5

$$\tau(A) = \|A\|_2 \|A^{-1}\|_2 = \tau(A^{-1})$$

3) From 2) we have

$$\tau\left(AB^{-1}\right) = \tau\left((AB^{-1})^{-1}\right) = \tau(BA^{-1})$$

4) By proposition 5

$$\tau\left(AB^{-1}\right) = \|AB^{-1}\|_2 \|BA^{-1}\|_2 \leq \|A\|_2 \|B^{-1}\|_2 \|B\|_2 \|A^{-1}\|_2 = \tau(A)\tau(B)$$

5) From 4) we have

$$\tau(A) = \tau(AB^{-1}B) \leq \tau\left(AB^{-1}\right)\tau(B)$$

and the same relation for $B$. □
Definition 10. We call Space of Tilts, denoted by $\Omega$, the quotient $GL^+(2)/\sim$ where the equivalence relation $\sim$ has been given in Definition 6.

This proposition completes Definition 6 and clarifies the geometrical interpretation of the space of tilts: an element in the space of tilts represents the set of all the camera spins and zooms associated with a certain tilt in a certain direction.

Notation 1. Let $A \in GL^+(2)$. We denote by $[A]$ the equivalence class in the space of tilts associated to $A$ i.e.

$$[A] = \{ B \in GL^+(2) \mid A \sim B \}.$$ 

Definition 11. We denote by $i$ the canonical injection from the space of tilts to $GL^+(2)$ defined by

$$i : \Omega \rightarrow GL^+(2) \quad [A] \mapsto T_{r(A)}R_{\phi(A)}.$$ 

This injection filters out the canonical representative from each class which is a mere tilt in the $x$ direction.

Remark 12. Clearly, the function $i$ satisfies

$$[A] = [i([A])]$$

and the space of tilts can be parameterized by picking these representative elements in each class as

$$\Omega = [Id] \bigcup \left\{ \bigcup_{(t,\phi)\in[1,\infty[\times[0,\pi]} [T_tR_{\phi}] \right\}.$$ 

The next proposition brings an additional justification to Definition 10. It means that the transition tilt does not depend on the choice of the class representative in the space of tilts.

Proposition 13. Let $A, B, C, D \in GL^+(2)$ satisfying $C \in [A]$ and $D \in [B]$. Then

$$\tau(AB^{-1}) = \tau(CD^{-1}).$$

Proof. Let $C \in [A], D \in [B]$. We first remark that if either $A \in GO^+(2)$ or $B \in GO^+(2)$ then the transition tilt operation is respectively the absolute tilt of $D$ or $C$, which does not depend on the class representative.

So without loss of generality suppose $A, B \in GL^+(2)$. Then, by proposition 1, they are re-written in a unique way as

$$A = \lambda_A Q_A T_s R_A$$

$$B = \lambda_B Q_B T_t R_B$$

and the same result can be applied to the following two matrices

$$AB^{-1} = \lambda_{AB^{-1}} Q_{AB^{-1}} T_{\tau(AB^{-1})} R_{AB^{-1}}$$

$$T_s R_A R_B^{-1} T_t^{-1} = \alpha Q_3 T_{t_3} R_3.$$ 

Moreover

$$AB^{-1} = \frac{\lambda_A Q_A T_s R_A (\lambda_B Q_B T_t R_B)^{-1}}{\lambda_B (Q_A Q_3) T_{t_3} (R_3 Q_B^{-1}).}$$

Then, by uniqueness of decomposition in equation (2) we have $T_{\tau(AB^{-1})} = T_{\alpha}$, implying

$$\tau(AB^{-1}) = \tau(T_s R_A R_B^{-1} T_t^{-1}).$$

Again, the same methodology applied to

$$C = \lambda_C Q_C A$$

$$= \lambda_C \lambda_A Q_A Q_A T_s R_A$$
and

\[
D = \lambda_D Q_D B = \lambda_D \lambda_B Q_D Q_B T_i R_B
\]

shows that

\[
\tau(CD^{-1}) = \tau(T_i R_A R_B^{-1} T_i^{-1}) = \tau(AB^{-1}).
\]

The next proposition follows directly from Proposition 9.

**Proposition 14.** The function \( d : \Omega \times \Omega \to \mathbb{R}_+ \frac{1}{\tau(AB^{-1})} \) is a metric acting on the space of tilts.

*Proof.* First, \( d \) is well defined thanks to Proposition 13 which ensures the independence from class representatives. Let us now prove the four metric axioms:

1) By definition of the absolute tilt \( \forall A, B \in GL^+(2) \) one has that \( \tau(AB^{-1}) \geq 1 \). This implies

\[
d([A], [B]) \geq 0.
\]

2) By Proposition 9-1) \( \forall A, B \in GL^+(2) \)

\[
d([A], [B]) = 0 \iff \tau(AB^{-1}) = 1 \iff A \sim B \iff [A] = [B]
\]

3) \( \forall A, B \in GL^+(2) \), Proposition 9-3) states that

\[
\tau(BA^{-1}) = \tau(AB^{-1})
\]

which implies

\[
d([A], [B]) = d([B], [A])
\]

4) \( \forall A, B, C \in GL^+(2) \), Proposition 9-4) assures that the following inequality holds

\[
\tau(BC^{-1}(AC^{-1})^{-1}) \leq \tau(BC^{-1}) \tau(AC^{-1}).
\]

As the logarithm is monotone in \([1, \infty]\), by simply applying it to both sides one obtains the triangular inequality for \( d \).

*2.3 Neighborhoods in the space of tilts*

Now that we have introduced the space of tilts and the adequate metric on this space to measure image distortion, we wish to explore optimal coverings for this space. We start by establishing closed formulas for disks in this 2D space.

**Theorem 15.** Given an element of the space of tilts in canonical form \([T_i R(\phi_1)]\), the disk \( [T_i R(\phi_1)], r \) in the space of tilts centered at this element and with radius \( r \) corresponds to the following set

\[
\left\{ [T_i R(\phi_2)] \mid G(t, s, \phi_1, \phi_2) \leq \frac{e^{2r} + 1}{2e^r} \right\}
\]

where

\[
G(t, s, \phi_1, \phi_2) = \left( \frac{\pi + \frac{r}{2}}{2} \right) \cos^2(\phi_1 - \phi_2) + \left( \frac{1 + st}{2} \right) \sin^2(\phi_1 - \phi_2).
\]
The proof of this theorem is given in the appendix. Figure 3 displays such disks in polar coordinates \((\log \tau \cos(\phi), \log \tau \sin(\phi))\). This representation will be convenient to visualize region coverings defined by disks in the space of tilts. Figure 4 is illustrating an observation hemisphere, which displays in a geometric environment the space of tilts, the class of affine transformations in question (green dots) and their neighborhoods (black surfaces). Notice that green dots represent camera viewpoints as depicted in Figure 1. In both representations, the pairs \((\tau, \phi)\) and \((\tau, \phi + \pi)\) are denoting the same element of the space of tilts. This is easily interpreted: Two identical images of a planar scene are indeed obtained by an affine camera positioned with a \(\pi\) longitude difference.

**Proposition 16.** Let \(A, B, C \in GL^+(2)\). Then

\[[A] C = [AC],\]

i.e, classes in \(\Omega\) are stable by right multiplication. Moreover,

\[d([AC], [BC]) = d([A], [B]).\]

**Proof.**

1) Proof of \([A] C \subset [AC].\)

\[B \in [A] \implies B = \lambda RA \implies BC = \lambda RAC \implies BC \in [AC]\]

2) Proof of \([AC] \subset [A] C.\)

\[D \in [AC] \implies D = \lambda RAC \implies D \in [A] C\]
Figure 4: (Perspective views)

Green point - Affine transformation in question
Dashed line - $\partial B \left( \left[ Id \right], \log \sqrt{2} \right)$
Black surface - Disk in question

3)

$$d([AC], [BC]) = \log \tau \left( AC (BC)^{-1} \right)$$

$$= \log \tau (AB^{-1})$$

$$= d(A, B)$$

Remark 17. Proposition 16 guarantees that transition tilts remain unchanged by right compositions. Furthermore, as argued in the proof of Proposition 25, the right composition with an element $C \in GL^+ (2)$ could be seen as a modification from a hypothetic frontal image $u$ to another hypothetic frontal image $C^{-1}u$. All this gives both motivation and meaning to the forthcoming Theorem 19.

Remark 18. One might also be interested in the way disks are transformed by left multiplication of elements belonging to $GL^+ (2)$. Unfortunately, in general

$$C [A] \neq [CA].$$

Take for example $C = A = T_t$ so

$$R_{2} = T_t \left( \frac{1}{t} R_{2} T_t \right) \notin [T_{t^2}].$$
Furthermore, for $C \in GL^+(2)$ one has
\[
\tau (CAB^{-1}C^{-1}) = c_2 (CAB^{-1}C^{-1}) = \|CAB^{-1}C^{-1}\|_2 \|C (AB^{-1})^{-1} C^{-1}\|_2 \\
\leq \|C\|_2^2 \|C^{-1}\|_2^2 \|AB^{-1}\|_2 \|(AB^{-1})^{-1}\|_2 \\
= \tau (C)^2 \tau (AB^{-1})
\]
so, in general
\[
d ([CA], [CB]) \leq 2d ([C], [Id]) + d ([A], [B]).
\]

The following theorem will be crucial in the next Section to explain why IMAS algorithms are truly affine invariant.

**Theorem 19.** Let
\[
\Gamma_1 = B ([Id], \log \Lambda_1) \\
\Gamma_2 = B ([Id], \log \Lambda_2) \\
\Gamma' = B ([Id], \log \Lambda_2 r)
\]
be three neighborhoods of $[Id]$ in $\Omega$ where $\Lambda_1, \Lambda_2, r \in [1, \infty[$, and assume that $S_1, S_2 \subset \Omega$ are two log $r$-coverings of $\Gamma_1$ and $\Gamma'$, i.e
\[
\Gamma_1 \subset \bigcup_{S \in S_1} B (S, \log r) \\
\Gamma' \subset \bigcup_{S \in S_2} B (S, \log r).
\]

Then, for every $[A] \in \Gamma_1$, $[B] \in \Gamma_2$, there exist $C \in GL^+(2)$ with $\tau (C) \leq r$, $S_A \in S_1$ and $S_B \in S_2$ such that
\[
d (S_A, [(AC)^{-1}]) = 0 \\
d (S_B, [(BC)^{-1}]) \leq \log r.
\]

A sketch of Theorem 19 appears in Figure 5.

**Proof.** Let us set $C = A^{-1} i (S_A)^{-1}$ where $i$ appears in Definition 11.

1) Proof of $d \left( S_A, \left( (AC)^{-1} \right) \right) = 0$.

Proposition 9-2) directly implies
\[
d ([Id], [A]) = d ([Id], [A^{-1}]).
\]
Then, as $S_1$ is a log $r$-covering of $\Gamma_1$, there exists $S_A \in S_1$ such that
\[
[A^{-1}] \in B (S_A, \log r)
\]
meaning that, the following inequality holds
\[
d \left( [Id], \left[ A^{-1} i (S_A)^{-1} \right] \right) = \log \tau \left( A^{-1} i (S_A)^{-1} \right) \\
= d ([A^{-1}], S_A) \\
\leq \log r.
\]

Finally, as $d$ is a metric (by Proposition 14) we know
\[
d \left( S_A, \left[ (AC)^{-1} \right] \right) = d (S_A, [i (S_A)]) = 0.
\]
\[ S_A = [(AC)^{-1}] \]

\[ S_A \cup \mathcal{B}(S_A, \log r) \]

\[ \Gamma_1 \]

\[ [Id] \]

\[ S_B \]

\[ [B^{-1}] \]

\[ \mathcal{B}(S_B, \log r) \]

\[ \Gamma_2 \]

\[ \Gamma' \]

Figure 5: Sketch of Theorem 19.

2) Proof of \( d \left( S_B, \left[(BC)^{-1}\right] \right) \leq \log r \).

By first using Proposition 9 followed by Proposition 14 we have

\[ \tau(B) \leq \tau(C^{-1}) = \Lambda_2 r \]

\[ \downarrow \]

\[ d \left( [Id], \left[(BC)^{-1}\right] \right) = \log \tau(B) \leq \log \Lambda_2 r \]

\[ \downarrow \]

\[ \left[(BC)^{-1}\right] \in \Gamma'. \]

Once more, as \( S_2 \) is a \( \log r \)-covering of \( \Gamma' \), there exists \( S_B \in S_2 \) such that

\[ \left[(BC)^{-1}\right] \in \mathcal{B}(S_B, \log r). \]

\[ \square \]

3 Application: optimal affine invariant image matching algorithms

The theory and results presented above provide a well suited geometrical framework for image matching by affine simulation (IMAS). This section gives the mathematical formalism and a mathematical proof that IMAS based algorithms are fully affine invariant, up to sampling errors. While the former sections only dealt with affine geometry, we now must introduce in the formalism the camera blur, as we shall deal with digital image recognition. Our goal is to define rigorously affine invariant recognition for digital images.

Consider a continuous and bounded image \( u(x) \) defined for every \( x = (x, y) \in \mathbb{R}^2 \). All continuous image operators including the sampling will be written in capital letters \( A, B \) and their composition as a mere juxtaposition \( AB \).

**Definition 20.** For any \( A \in GL^+(2) \), we define the affine transform \( A \) of a continuous image \( u \) by

\[ Au(x) := u(Ax). \]

Homotheties and rotations acting on continuous images are similarly written as

\[ H_\lambda u(x) = u(\lambda x); \]

\[ R_\varphi u(x) = u(R_\varphi x). \]
We now introduce a compact notation for the various convolutions with Gaussians. We shall denote by \( \star_x \) the 1-D convolution operator in the \( x \)-direction, i.e.

\[
G \star_x u (x, y) = \int_{\mathbb{R}} G(z) \, u(x - z, y) \, dz.
\]

Similarly, we denote by \( \star_y \) the 1-D convolution operator in the \( y \)-direction. We denote by \( G_{\sigma}^{y} \), \( G_{\sigma}^{x} \) and \( G_{\sigma}^{y} \) respectively the 2D and 1D convolution operators in the \( x \) and \( y \) directions with

\[
G_{\sigma^{2}} (x, y) := \frac{1}{2\pi(\sigma^{2})} e^{-\frac{x^{2}+y^{2}}{2\sigma^{2}}},
\]

\[
G_{\sigma}^{y} (x) := \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{x^{2}}{2\sigma^{2}}},
\]

\[
G_{\sigma}^{y} (y) := \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{y^{2}}{2\sigma^{2}}},
\]

namely

\[
G_{\sigma} u := G_{\sigma}^{x} \star_{x} u, \\
G_{\sigma}^{y} u := G_{\sigma}^{y} \star_{y} u, \\
G_{\sigma}^{y} u := G_{\sigma}^{y} \star_{y} u.
\]

Here the constant \( c \geq 0.7 \) is large enough to ensure that all convolved images, initially sampled at 1 distance, can be sub-sampled at Nyquist distance \( \sigma \) without causing significant aliasing.

**Remark 21.** \( G_{\sigma} \) satisfies the semigroup property

\[
G_{\sigma} G_{\beta} = G_{\sqrt{\sigma^{2}+\beta^{2}}}. \tag{3}
\]

By a mere change of variables in the integral defining the convolution, the next formula holds and will be useful:

\[
G_{\sigma} H_{\gamma} u = H_{\gamma} G_{\sigma^{2}} u. \tag{4}
\]

In the classic Shannon-Nyquist framework, we shall denote the image sampling operator (on a unary grid) by \( S_{1} \). Thus \( S_{1} u \) is defined on the grid \( \mathbb{Z}^{2} \). The Shannon-Whittaker interpolator of a digital image on \( \mathbb{Z}^{2} \) will be denoted by \( I \).

As developed in [64], the whole image comparison process, based on local features, can proceed as though images where (locally) obtained by using digital cameras that stand far away, at infinity. The geometric deformations induced by the motion of such cameras are affine maps. A model is also needed for the two main camera parameters not deducible from its position, namely sampling and blur. The digital image is defined on the camera CCD plane. The pixel width can be taken as length unit, and the origin and axes chosen so that the camera pixels are indexed by \( \mathbb{Z}^{2} \). The digital initial image is always assumed well-sampled and obtained by a Gaussian blur with standard deviation around 0.8. In all that follows, \( u_{0} \) denotes the (theoretical) infinite resolution image that would be obtained by a frontal snapshot of a plane object with infinitely many pixels. The digital image obtained by any camera at infinity is therefore formalized as \( u = S_{1}G_{\sigma}^{y}ATu_{0} \), where \( A \) is any linear map with positive singular values and \( T \) any plane translation. Thus we can summarize the general image formation model with cameras at infinity as follows.

**Definition 22 (Image formation model).** Digital images of a planar object whose frontal infinite resolution image is \( u_{0} \), obtained by a digital camera far away from the object, satisfy

\[
u =: S_{1}G_{\sigma}^{y}ATu_{0} \tag{5}
\]

where \( A \) is any linear map and \( T \) any plane translation. \( G_{\sigma} \) denotes a Gaussian kernel broad enough to ensure no aliasing by 1-sampling, namely \( I S_{1}G_{\sigma}^{y}ATu_{0} = G_{\sigma}^{y}ATu_{0} \).

The image formation model in Definition 22 is illustrated in Figure 3.
3.1 Inverting tilts

We now formalize the notion of tilt. There are actually three different notions of tilt, that we must carefully distinguish.

Definition 23. Given $t > 1$, the tilt factor, define:

- Geometric tilts
  
  \[
  T^x_t u_0(x, y) := u_0(tx, y);
  \]
  
  \[
  T^y_t u_0(x, y) := u_0(x, ty).
  \]

- Simulated tilts (taking into account camera blur)
  
  \[
  T^x_t v := T^x_t G^x_1 \sqrt{t^2 - 1} \ast x v;
  \]
  
  \[
  T^y_t v := T^y_t G^y_1 \sqrt{t^2 - 1} \ast y v.
  \]

- Digital tilts (transforming a digital image $u$ into a digital image)
  
  \[
  u \rightarrow S_1 T^x_t I u;
  \]
  
  \[
  u \rightarrow S_1 T^y_t I u.
  \]

Digital tilts are the ones used in practice. It all adds up because the simulated tilt yields a blur permitting $S_1$-sampling.

If $u_0$ is an infinite resolution image observed with a camera tilt of $t$ in the $x$ direction, the observed image is $G_1 T^x_t u_0$. Our main problem is to reverse such tilts. This operation is in principle impossible, because geometric tilts do not commute with blur. However, the first formula of the next Theorem 24 shows that $T^y_t$ is, up to a zoom out, a pseudo inverse to $T^x_t$.

The meaning of this result is that a tilted image $G_1 T^x_t u$ can be tilted back by tilting in the orthogonal direction. The price to pay is a $t$ zoom out. The second relation in the theorem means that the application of the simulated tilt to an image that can be well sampled by $S_1$ yields an image that keeps that well sampling property.

Theorem 24. Let $t \geq 1$. Then

\[
T^y_t G_1 T^x_t = G_1 H_t; \quad (6)
\]

\[
T^y_t G_1 = G_1 T^y_t. \quad (7)
\]

Proof. Since $H_t = T^y_t T^x_t$, (6) follows from (7) by composing both sides on the right by $T^x_t$. Let us now prove (7). We shall use the following obvious facts

\[
G_1 = G^x_1 G^y_1 = G^y_1 G^x_1
\]

which follows from the separability of the Gaussian and Fubini’s theorem and the commutation

\[
G^x_1 T^y_t = T^y_t G^x_1
\]
which is true because $G^y_1$ and $T^y_1$ act separably on the variables $x$ and $y$. Using first (4) in the $y$ dimension where $T^y_1$ is a mere homothety, and then successively (9), (8), the semigroup property for the Gaussians, and Definition 23 we get

\[
T^y_1 G^y_1 = G^y_1 T^y_1, \\
G^y_1 T^y_1 G^y_1 = G^y_1 G^y_1 T^y_1, \\
T^y_1 G^y_1 G^y_1 = G^y_1 T^y_1, \\
T^y_1 G^y_1 T^y_1 G^y_1 = G^y_1 T^y_1 G^y_1, \\
T^y_1 G^y_1 = G^y_1 T^y_1,
\]

which proves (7).

The meaning of Theorem 24 is that we can design an exact algorithm that simulates all inverse tilts for comparing two digital images. This algorithm handles two images $u = G_1 AT_1 w_0$ and $v = G_1 BT_2 w_0$ that are two snapshots from different view points of a flat object whose front infinite resolution image is denoted by $w_0$.

### 3.2 Proof that IMAS works

In this section, the formal IMAS algorithm is duly presented (Algorithm 1). Our goal is to prove that it works. This proof is a direct application of the results introduced of the previous section. The algorithm and its proof rely on the formal assumption that there exists an image comparison algorithm able to compare image pairs with tilts lower than $r$. The core idea of IMAS algorithms is illustrated in Figure 7.

---

**Algorithm 1** Formal IMAS (Image Matching by Affine Simulation)

**Enviroment:**

Parameters and assumptions from Theorem 19 with

\[S_i = \left\{ \left[ T^{x_k} R_{\phi_k^1} \right] \right\}_{k=1, \ldots, n_i} \]

**Input:**

Query and target images: $u$ and $v$.

**Start:**

1. $\forall k = 1, \ldots, n_1$ do
   
   \[u_k = T^{x_k} R_{\phi_k^1} u.\]

2. $\forall k = 1, \ldots, n_2$ do
   
   \[v_k = T^{x_k} R_{\phi_k^2} v.\]

3. $\forall (k_1, k_2) \in \{1, \ldots, n_1\} \times \{1, \ldots, n_2\}$
   
   \[M_{k_1, k_2} = \text{SHM-Matches}(u_{k_1}, v_{k_2}).\]

**Output:**

\[M = \bigcup_{(k_1, k_2) \in \{1, \ldots, n_1\} \times \{1, \ldots, n_2\}} M_{k_1, k_2}.\]
Figure 7: IMAS algorithms start by applying a finite set of optical affine simulations to $u$ and $v$, followed by pairwise comparisons.

**Proposition 25.** Let $u$ and $v$ be respectively query and target images which are related by a transition tilt under $\Lambda_1 \Lambda_2$, i.e. there exist a continuous image $w_0$ and $A, B \in GL^+$ (2) with

$$\tau(A) \leq \Lambda_1$$

$$\tau(B) \leq \Lambda_2$$

such that

$$u = G_1 AT_1 w_0 \quad \text{and} \quad v = G_1 BT_2 w_0 \quad (10)$$

where $T_1, T_2$ are planar translations. Then, under the assumptions of Theorem 19, the formal IMAS of Algorithm 1 generates two affine versions of the images $u$ and $v$ with a transition tilt lower than $r$.

**Proof.** By Theorem 19 there exist $S_A \in S_1$, $S_B \in S_2$ and $C \in GL^+$ (2) with $\tau(C) \leq r$ such that

$$d(S_A, \left[(AC)^{-1}\right]) = 0$$

$$d(S_B, \left[(BC)^{-1}\right]) \leq \log r.$$

Consider the slanted view of the frontal continuous image $w_0$ defined by $w_1 := C^{-1}w_0$. Then we can rewrite query and target images as

$$u = G_1 AC T_1 w_1 \quad \text{and} \quad v = G_1 BC T_2 w_1.$$

By Proposition 16, the above modification keeps transitions tilts stable, i.e.

$$d([AC], [BC]) = d([A], [B]),$$

so we can reason as if $w_1$ were the frontal image, instead of $w_0$.

Now, the formal IMAS Algorithm 1 will apply $i(S_A) = T_{iA} R_{\phi A}$ and $i(S_B) = T_{iB} R_{\phi B}$ respectively on the query and target images. This is:

1. $T_{iA} R_{\phi A}$ to $u$, which yields

$$\tilde{u} = G_1 i(S_A) AC T_1 w_1 = G_1 \lambda R T_1 w_1.$$

2. $T_{iB} R_{\phi B}$ to $v$, which yields

$$\tilde{v} = G_1 i(S_B) BC T_2 w_1.$$

But

$$d([Id], [i(S_B) BC]) = \log \tau(i(S_B) BC)$$

$$= d(S_B, \left[(BC)^{-1}\right])$$

$$\leq \log r$$

which proves that the affine relation between $\tilde{u}$ and $\tilde{v}$ involves a transition tilt under $r$.  \qed
Remark 26. Two log $r$-coverings of the same region

$$\Gamma = B([Id], \log \Lambda)$$

would then ensure that the formal IMAS Algorithm 1 manages to reduce transition tilts under $\Lambda^2$ between two images into transition tilts under $r$. A relation between covered absolute tilts, attainable transition tilts and maximal viewpoint angle can be found in Table 1.

<table>
<thead>
<tr>
<th>Covered absolute tilts ($\tau(A) \leq \sqrt{r}\Lambda$ and $\tau(B) \leq \sqrt{r}\Lambda$)</th>
<th>Attainable transition tilts ($\tau(AB^{-1}) \leq \Lambda^2$)</th>
<th>Viewpoint angle $\arccos \frac{1}{\Lambda^2}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\Lambda = 8$</td>
<td>64</td>
<td>89.1°</td>
</tr>
<tr>
<td>$\Lambda = 4\sqrt{2}$</td>
<td>32</td>
<td>88.2°</td>
</tr>
<tr>
<td>$\Lambda = 4$</td>
<td>16</td>
<td>86.4°</td>
</tr>
<tr>
<td>$\Lambda = 2\sqrt{2}$</td>
<td>8</td>
<td>82.8°</td>
</tr>
<tr>
<td>$\Lambda = 2$</td>
<td>4</td>
<td>75.5°</td>
</tr>
<tr>
<td>$\Lambda = \sqrt{2}$</td>
<td>2</td>
<td>60°</td>
</tr>
</tbody>
</table>

Table 1: Link between absolute tilts, transition tilts and viewpoint.

### 3.3 Optimal discrete coverings in the space of tilts

We now consider the problem of providing two optimal sets $S_1, S_2 \subset \Omega$ permitting the application of Theorem 19. These sets should ensure a minimal complexity for the IMAS algorithm. We thus need to define an optimality criterion. We observe that an IMAS algorithm simulates affine transformations on a digital image and then compares descriptors coming from these simulated versions. One would like to minimize the overall number of descriptor comparisons while maintaining the detection efficiency. This minimization is not equivalent to a minimization of the number of simulated versions being used. We shall base our efficiency criterion on two straightforward remarks. The first one is that if a digital image suffers a tilt $t$ in any direction, its area gets modified by a factor $\frac{1}{t}$. The second one is that the expected number of keypoints in a digital image is proportional to its area. Both remarks imply that the complexity of an IMAS algorithm will be given by the overall area of the simulated images being ultimately compared. This justifies the next definition.

**Definition 27.** We call area ratio of $S$ (a finite set of elements in $\Omega$) the real number

$$\sum_{S \in S} \frac{1}{\tau(S)}.$$

The area ratio fixes the factor (larger than 1) by which the image area is being multiplied when summing the areas of all of its tilted versions. Then, as the ultimate goal is to reduce the number of key points comparisons, it is natural to look for a set $S$ whose area ratio is close to the infimum among all log $r$-coverings of $\Gamma$. Unfortunately, even in $\mathbb{R}^2$, the mathematical problem of finding a covering of a certain set with a minimum amount of disks is well known to be NP-hard. It is therefore difficult to find an optimal solution for our problem, and unlikely that it will be proved to be optimal even if it is. Fortunately, our search space in the set of log $r$-coverings can be drastically reduced by imposing practical and theoretical constraints to $S$. Those constraints follow from simple requirements for an image matching method.

**Definition 28.** We shall say that a set $S \in \Omega$ is feasible if and only if:

1. $[Id] \in S$.
2. There exist $n \in \mathbb{N}^+$ and

$$(t_1, \ldots, t_n, \phi_1, \ldots, \phi_n) \in [1, \infty[^n \times [0, \pi]^n$$

such that

$$S \setminus \{[Id]\} = \bigcup_{i=1}^n \left\{[T_{t_i} R_{k\phi_i}] \in \Omega \mid k = 0, \ldots, \left\lfloor \frac{\pi}{\phi_i} \right\rfloor \right\}$$

where $\lfloor a \rfloor$ denotes the nearest integer less than or equal to a real number $a$. 

17
Proposition 31. There exists a feasible log 1.8-covering, depicted in Figure 9c, with area ratio equal to 6.34. It is an approximated solution of the optimization problem in (11) for $\Gamma = \{T_i R_\phi \mid t \leq 6\}$, $n = 2$. Therefore, the infimum area ratio among all log 1.8-coverings of $\{T_i R_\phi \mid t \leq 6\}$ is lower than 6.34.

Proof. We are dealing with 4 dimensions to minimize and more specifically with $100^4$ feasible sets. Computing area ratios for each feasible set is straightforward but validating the covering condition is a more involved computational issue. For the sake of clearness, the intersection of disks boundaries, which are composed at most of two elements for non identical disks, shall be denoted by

$$\Sigma_1 = \partial B^{\text{log } 1.8}_{[T_1]} \cap \partial B^{\text{log } 1.8}_{[T_1 R_{\phi_1}]} \quad \text{and} \quad \Sigma_2 = \partial B^{\text{log } 1.8}_{[T_2]} \cap \partial B^{\text{log } 1.8}_{[T_2 R_{\phi_2}]}$$

and their respective closest and farthest elements will be denoted by

$$\min \Sigma_1 := \arg \min_{S \in \Sigma_1} d (S, [Id]) \quad \max \Sigma_1 := \arg \max_{S \in \Sigma_1} d (S, [Id]) \quad \min \Sigma_2 := \arg \min_{S \in \Sigma_2} d (S, [Id]) \quad \max \Sigma_2 := \arg \max_{S \in \Sigma_2} d (S, [Id]).$$

In order to check if a feasible set does cover the specified region we propose to verify the following four conditions depicted in Figure 8:
1. $\Sigma_1 \neq \emptyset$ and $\Sigma_2 \neq \emptyset$.

2. $\min \Sigma_1$ must lie inside the ball $B_{\log 1.8}$, which ensures a covering of $B_{\log \tau(\max \Sigma_1)}$.

3. $\max \Sigma_2$ must lie outside the region $\Gamma$, which ensures a covering of the annulus defined by $\Gamma \setminus B_{\log \tau(\min \Sigma_2)}$.

4. For $\epsilon$ small, all elements $S \in \mathbb{F}_\epsilon$ must lie inside some disks of radius $\log (1.8 - \epsilon)$, i.e.

$$S \in \bigcup_{1 \leq i \leq 2} \bigcup_{S' \in J_{ti, \phi_i}} B_{\log(1.8 - \epsilon)}^{\log \tau(\min \Sigma_2)} \setminus B_{\log \tau(\max \Sigma_1)}^{\log \tau(\min \Sigma_2)}.$$

where $\mathbb{F}_\epsilon$ is a finite $\epsilon$-dense set of the annulus defined by

$$\mathbb{F}_\epsilon = B_{\log \tau(\min \Sigma_2)}^{\log \tau(\min \Sigma_2)} \setminus B_{\log \tau(\max \Sigma_1)}^{\log \tau(\min \Sigma_2)}.$$

Notice that the fourth condition only ensures a $\log (1.8 - \epsilon)$-covering up to an error

$$\epsilon = \max_{S' \in \Gamma} \min_{S \in \mathbb{F}_\epsilon} d(S, S')$$

and so, by dilating back disks radius to 1.8 one ensures log 1.8-coverings.

By using the procedure described above, an approximated solution to the optimization problem in (11) has been obtained. Its parameters can be found in Table 2. Its corresponding representation in the space of tilts appears in Figure 9c.

The procedure in the proof of Proposition 31 has also been applied to find more near optimal coverings appearing in Figure 9.
Figure 9: Near-optimal coverings in the space of tilts.
Gray areas - Uncovered.
Blue areas - Covered by at least two disks.
White areas - Covered by only one disk.
### Table 2: Approximated solution to the optimization problem in (11)

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>$t^{opt}_1$</td>
<td>2.8847</td>
</tr>
<tr>
<td>$\phi^{opt}_1$</td>
<td>0.394085</td>
</tr>
<tr>
<td>$t^{opt}_2$</td>
<td>6.2197</td>
</tr>
<tr>
<td>$\phi^{opt}_2$</td>
<td>0.196389</td>
</tr>
</tbody>
</table>

4 Experimental Validation

We are now able to propose and evaluate for each SIIM method its IMAS, namely its affine-invariant extension. This affine invariant version relies on two facts. First, each SIIM identifies viewpoint changes, under a certain transition tilt threshold (that we shall estimate in this section). Second, any smooth map is locally approximable by an affine map. Hence, under the assumption that the surface of photographed objects is locally smooth, all viewpoint changes can be understood as local transition tilts changes (see Figure 1). Third, once provided with a log $r$-covering of $\Gamma = \Gamma'$, where $r$ is less than the transition tilt threshold of the SIIM, Proposition 25 states that Algorithm 1 offers an affine-invariant version of the considered SIIM. Indeed, there is at least one pair of simulated images whose transition tilt is less than $r$, and on these two images the SIIM can succeed. The affine invariance property is ensured for transition tilts changes up to $\Lambda_1 \Lambda_2$, i.e. for viewpoint angle changes of about $\arccos \left( \frac{1}{\Lambda_1 \Lambda_2} \right)$. We shall denote by $t^{s_1 \times s_2}_{max}$ the associated maximum tilt tolerance with respect to a matching method for images with size larger than $s_1 \times s_2$.

In our experiments, all SIIM methods were immersed in the same affine extension set-up. The simulation of optical tilts, matching and filtering were handled in the very same way. This set-up received as a parameter the name of the base detector+extractor method to perform, then a brute force matcher was performed with the second-closest neighbor acceptance criterion proposed by D. Lowe in [33]. Finally, as presented in [44, 64], three main filters were applied: first, only unique matches were taken into account; second, groups of multiple-to-one and one-to-multiple matches were removed; finally, only matches coming from the most significant geometric model (if it existed!) were kept. In our case, as all tests were based on planar transformations, the ORSA homography detector [41] (a parameterless variant of RANSAC) was applied to filter out matches not compatible with the dominant homography.

All detectors, all extractors and the matcher were taken from the Open Source Computer Vision (OPENCV) Library, version 3.2.0.

4.1 Maximal tilt tolerance computation for each SIIM

From the complexity viewpoint, the main quantitative parameter for extending a SIIM into an IMAS is its tilt tolerance. We do not question the invariance of descriptors with respect to zoom and rotations but rather how they perform against transition tilts changes incurred when matching, for example, $G_1 I d u$ to $G_1 T_i R_\phi u$ where $t \in [1, \infty]$ and $\phi \in [0, \pi]$.

We used the tolerance image dataset displayed in Figure 10 to evaluate the maximal tilt tolerance of each SIIM with respect to images of similar size. Images in this dataset have a fixed size and were selected to obtain a diversity of challenging scenarios. In order to approximate $t^{700 \times 550}_{max}$, we simulated optical tilts on the tolerance image dataset and then tested whether this affine simulation was identified by ORSA Homography with a precision of 3 pixels. This test determined upper bounds $U^{700 \times 550}_{max}$ depicted in Figure 11 for nine of the best state-of-the-art SIIMs.

This test yielded upper bounds for $r^{700 \times 550}_{max}$, based on its application to nine images whose sizes are close to 700 $\times$ 550. Supposing a maximal angle error computation of $\frac{\pi}{10}$, we assumed that for each SIIM

$$ t^{700 \times 550}_{max} = \frac{U^{700 \times 550}_{max}}{1 - \cos \left( \frac{\pi}{10} \right)} \approx \frac{U^{700 \times 550}_{max}}{1.05} $$

and constructed its affine invariant version with log $t^{700 \times 550}_{max}$-coverings.
4.2 Affine-invariant methods

The matching process is as symmetric as possible. No significant changes should come along by interchanging the roles of the query and target images. In the case of IMAS algorithms this symmetry implies a unique set of optical tilts to simulate on both query and target images. Thus, if this unique set of optical tilts represents a log \( r \)-covering of

\[ \Gamma_1 = \Gamma' = \{ [T_t R_{\phi}] \mid \phi \leq \Lambda \} \]

then Proposition 25 ensures that any IMAS based on a SIIM whose maximum tilt tolerance is greater than \( r \) is able to identify all tilts under \( \Lambda \) by simulating all affine maps in the log \( r \)-covering.

Several coverings in the space of tilts have been proposed in [44, 64, 49, 40] for SIFT and SURF. Figure 14 displays these coverings. They are clearly not optimal. Indeed, most of these coverings do not really cover the region they were meant to, except for ASIFT [44, 64] (which instead is visually redundant) and for the affine DoG-SIFT version in [40].

In order to compare the efficiency of those coverings, query and target images were generated in a way so as to test Algorithm 1 to the limit, i.e., forcing the worst case scenario in which \( (BC)^{-1} \) lies in \( \Gamma' \setminus \Gamma_2 \). We simulated the optical tilts on query and target images coming from one single image. This image, denoted by \( w_0 \) and appearing in Figure 12, was then used to compute the inputs of Algorithm 1 as follows:

- Query image (non-fixed tilt), \( G_1 A_{t,\phi} w_0 \) where \( A_{t,\phi} = R_{\phi} T_t R_{\pi/2} \).
- Target image (fixed tilt), \( G_1 B_{\phi} w_0 \) where \( B_{\phi} = R_{\phi + \pi/2} T_{\Lambda} \).

The veritable interest of these affine maps being the inverse maps they determine, namely,

\[
\begin{align*}
A_{t,\phi}^{-1} & = [T_t R_{\pi/2-\phi}], \\
B_{\phi}^{-1} & = [T_{\Lambda} R_{\phi}].
\end{align*}
\]

which according to Proposition 9-4, attain maximal transition tilts for fixed tilts such as \( t \) and \( \Lambda \), i.e.

\[ \tau \left( A_{t,\phi}^{-1} B_{\phi} \right) = t \Lambda. \]

When ORSA Homography was able to identify the affine map that relates query and target images, we counted the event as a success. Clearly, if \( \Gamma' \) and \( \Gamma_2 \) are truly log \( r \)-covered then
Figure 11: Represented in the space of tilts, the associated upper bounds \( (U_{700 \times 550}^{\text{max}}) \) for maximum tilt tolerances.

Black dot - \([id]\).

Coloured dots stand for tested tilts \([T_tR_\phi]\) where \(t \in \{1.4, 1.5, \ldots, 2.4\}\) and \(\phi \in \{0, 10, \ldots, 170\}\).

Blue dots - attainable tilts for all images in the dataset.

Red dots - unattainable tilts for at least one image in the dataset.

Gray areas - \(\{[T_tR_\phi] | t \geq U_{700 \times 550}^{\text{max}}\}\).  

White areas - \(\{[T_tR_\phi] | t \leq U_{700 \times 550}^{\text{max}}\}\).

Proposition 25 implies that all tests for which \(A^{-1}_{t,\phi} \in \Gamma_1\) should be counted as a success. Results in Figure 13 were as expected and highlight the importance of using the right coverings for extreme cases. Both ASIFT and Optimal Affine-SIFT were able to capt most of all transition tilts that Proposition 25 predicted, namely those under \(\Lambda_2\).

We must keep in mind that these log \(r\)-coverings depend on tilt tolerances found over images in Figure 10. Maximal tilt tolerances are linked to the size of images being compared and as a consequence the disks radius might grow or shrink proportionally to the minimum size of all simulated images. Moreover, Proposition 25 does not take into account discretization errors and relies on two main hypotheses:

1. The considered SIIM is truly rotation and zoom invariant.

2. For images similar to the input image, the SIIM under consideration has a maximal tilt tolerance not smaller than \(r\).

As anticipated, the area ratio associated to a covering reliably evaluates the difference of performance between affine versions of the same matching method. Being proportionally linked to the total amount of keypoints, the area ratio of Definition 27 predicts the order of growth in computation time. For example, the SIFT keypoint computation part induced by the optimal covering in Figure 9b is twice faster than the one induced by the ASIFT covering. The same goes for the matching part, only this time the optimal version is four times faster. Since both coverings cover about the same region, our Optimal Affine-SIFT supplants ASIFT with no qualitative matching loss.

Two examples of performance over query and target images from Figure 15 and 16 are respectively
Image matching by affine simulations (IMAS) is acknowledged as the best methodology to match images of the same scene regardless of the viewpoint change. Its time complexity is one of the main drawbacks that has been widely criticized in the literature. The mathematical derivations in this paper imply that IMAS based methods really are affine-invariant provided the base SIIM satisfies: scale+rotation invariance, sufficient distinctiveness, and an acceptable viewpoint tolerance measured as its transition tilt. We have proved that, as summarized in Figure 14, all former IMAS methods are over-simulating optical tilts. We therefore have developed a method finding for each SIIM an optimal IMAS method which only depends on the tilt tolerance of the SIIM. This led us to measure the tilt tolerance of a number of classic SIIMs. We found for example that the optimal IMAS extension of SIFT needs twice less descriptors and therefore is four times faster than ASIFT. This improvement applies to all state of the art IMAS, that can be accelerated by a factor of four. Another consequence is that the set of affine descriptors associated with an image can be halved.

Acknowledgement

This work has been partially funded by:
<table>
<thead>
<tr>
<th>Method</th>
<th>M</th>
<th>ar</th>
<th>ar^2</th>
<th>Keypoints (seconds)</th>
<th>Matching (seconds)</th>
<th>Filters (seconds)</th>
</tr>
</thead>
<tbody>
<tr>
<td>SIFT</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0.69</td>
<td>0.70</td>
<td>0.18</td>
</tr>
<tr>
<td>ASIFT</td>
<td>102</td>
<td>13.7</td>
<td>189.6</td>
<td>12.46</td>
<td>138.59</td>
<td>3.05</td>
</tr>
<tr>
<td>(Optimal) Affine-SIFT</td>
<td>795</td>
<td>7.06</td>
<td>49.8</td>
<td>6.04</td>
<td>29.61</td>
<td>1.39</td>
</tr>
<tr>
<td>RootSIFT</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0.72</td>
<td>0.71</td>
<td>0.18</td>
</tr>
<tr>
<td>Affine-RootSIFT</td>
<td>658</td>
<td>6.9</td>
<td>47.6</td>
<td>5.05</td>
<td>20.70</td>
<td>1.44</td>
</tr>
<tr>
<td>SURF</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1.01</td>
<td>0.79</td>
<td>0.19</td>
</tr>
<tr>
<td>(Optimal) Affine-SURF</td>
<td>471</td>
<td>14.82</td>
<td>219.6</td>
<td>12.53</td>
<td>35.24</td>
<td>1.40</td>
</tr>
<tr>
<td>BRISK</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1.75</td>
<td>0.27</td>
<td>0.18</td>
</tr>
<tr>
<td>Affine-BRISK</td>
<td>421</td>
<td>8.42</td>
<td>70.89</td>
<td>18.95</td>
<td>8.68</td>
<td>2.06</td>
</tr>
<tr>
<td>BRIEF</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0.05</td>
<td>0.01</td>
<td>0.19</td>
</tr>
<tr>
<td>Affine-BRIEF</td>
<td>0</td>
<td>14.82</td>
<td>219.6</td>
<td>4.20</td>
<td>2.18</td>
<td>6.08</td>
</tr>
<tr>
<td>ORB</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0.05</td>
<td>0.02</td>
<td>0.17</td>
</tr>
<tr>
<td>Affine-ORB</td>
<td>0</td>
<td>14.82</td>
<td>219.6</td>
<td>4.34</td>
<td>5.13</td>
<td>3.25</td>
</tr>
<tr>
<td>AKAZE</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0.42</td>
<td>0.13</td>
<td>0.21</td>
</tr>
<tr>
<td>Affine-AKAZE</td>
<td>194</td>
<td>8.42</td>
<td>70.89</td>
<td>5.00</td>
<td>6.23</td>
<td>3.74</td>
</tr>
<tr>
<td>LATCH</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0.11</td>
<td>0.02</td>
<td>0.00</td>
</tr>
<tr>
<td>Affine-LATCH</td>
<td>37</td>
<td>14.82</td>
<td>219.6</td>
<td>4.52</td>
<td>2.16</td>
<td>0.17</td>
</tr>
<tr>
<td>FREAK</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0.34</td>
<td>0.15</td>
<td>0.18</td>
</tr>
<tr>
<td>Affine-FREAK</td>
<td>145</td>
<td>7.06</td>
<td>49.8</td>
<td>4.37</td>
<td>2.38</td>
<td>1.94</td>
</tr>
</tbody>
</table>

Table 3: Matching methods performance over query and target images from Figure 15. The proposed matching methods in this paper appear in bold. Computations were performed on an Intel(R) Core(TM) i5-4210U CPU 1.70GHz with 2 cores.

<table>
<thead>
<tr>
<th>Method</th>
<th>M</th>
<th>ar</th>
<th>ar^2</th>
<th>Keypoints (seconds)</th>
<th>Matching (seconds)</th>
<th>Filters (seconds)</th>
</tr>
</thead>
<tbody>
<tr>
<td>SIFT</td>
<td>102</td>
<td>1</td>
<td>1</td>
<td>0.23</td>
<td>0.01</td>
<td>0.09</td>
</tr>
<tr>
<td>ASIFT</td>
<td>317</td>
<td>13.7</td>
<td>189.6</td>
<td>5.43</td>
<td>1.68</td>
<td>0.47</td>
</tr>
<tr>
<td>(Optimal) Affine-SIFT</td>
<td>292</td>
<td>7.06</td>
<td>49.8</td>
<td>2.71</td>
<td>0.38</td>
<td>0.30</td>
</tr>
<tr>
<td>RootSIFT</td>
<td>110</td>
<td>1</td>
<td>1</td>
<td>0.25</td>
<td>0.01</td>
<td>0.09</td>
</tr>
<tr>
<td>Affine-RootSIFT</td>
<td>219</td>
<td>6.9</td>
<td>47.6</td>
<td>2.23</td>
<td>0.28</td>
<td>0.24</td>
</tr>
<tr>
<td>SURF</td>
<td>110</td>
<td>1</td>
<td>1</td>
<td>0.24</td>
<td>0.03</td>
<td>0.14</td>
</tr>
<tr>
<td>(Optimal) Affine-SURF</td>
<td>663</td>
<td>14.82</td>
<td>219.6</td>
<td>3.68</td>
<td>1.19</td>
<td>0.73</td>
</tr>
<tr>
<td>BRISK</td>
<td>29</td>
<td>1</td>
<td>1</td>
<td>1.57</td>
<td>0.00</td>
<td>0.04</td>
</tr>
<tr>
<td>Affine-BRISK</td>
<td>49</td>
<td>8.42</td>
<td>70.89</td>
<td>17.57</td>
<td>0.06</td>
<td>0.08</td>
</tr>
<tr>
<td>BRIEF</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0.03</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>Affine-BRIEF</td>
<td>7</td>
<td>14.82</td>
<td>219.6</td>
<td>2.06</td>
<td>0.09</td>
<td>0.03</td>
</tr>
<tr>
<td>ORB</td>
<td>102</td>
<td>1</td>
<td>1</td>
<td>0.02</td>
<td>0.01</td>
<td>0.8</td>
</tr>
<tr>
<td>Affine-ORB</td>
<td>90</td>
<td>14.82</td>
<td>219.6</td>
<td>2.12</td>
<td>0.31</td>
<td>0.40</td>
</tr>
<tr>
<td>AKAZE</td>
<td>20</td>
<td>1</td>
<td>1</td>
<td>0.16</td>
<td>0.00</td>
<td>0.03</td>
</tr>
<tr>
<td>Affine-AKAZE</td>
<td>51</td>
<td>8.42</td>
<td>70.89</td>
<td>2.31</td>
<td>0.06</td>
<td>0.09</td>
</tr>
<tr>
<td>LATCH</td>
<td>54</td>
<td>1</td>
<td>1</td>
<td>0.07</td>
<td>0.01</td>
<td>0.04</td>
</tr>
<tr>
<td>Affine-LATCH</td>
<td>101</td>
<td>14.82</td>
<td>219.6</td>
<td>1.72</td>
<td>0.12</td>
<td>0.10</td>
</tr>
<tr>
<td>FREAK</td>
<td>124</td>
<td>1</td>
<td>1</td>
<td>0.14</td>
<td>0.01</td>
<td>0.10</td>
</tr>
<tr>
<td>Affine-FREAK</td>
<td>182</td>
<td>7.06</td>
<td>49.8</td>
<td>2.54</td>
<td>0.11</td>
<td>0.31</td>
</tr>
</tbody>
</table>

Table 4: Matching methods performance over query and target images from Figure 16. The proposed IMAS methods proposed here appear in bold. Computations were performed on an Intel(R) Core(TM) i5-4210U CPU 1.70GHz with 2 cores.

M - Matches.
ar - area ratio.
Figure 13: Extreme test results.

Black dot - $[Id]$. Coloured dots stand for $[A_{-1}^{1}]$ and belong to a fixed log 1.1 uniform discretization of the annulus $\{[T_tR_\phi] \ | \ 2 \leq t \leq 4\sqrt{2} \}$. The angle $\phi$ implicitly fixes $[B_{-1}^{1}] = [T_\Lambda R_\phi]$ where $\Lambda = \arg\max_{t} [T_tR_\phi] \in \Gamma'$.

Blue/Red dots - Success/Failure of ORSA Homography in identifying the underlying affine map.

- BPIFrance and Region Ile de France, in the framework of the FUI 18 Plein Phare project, the Office of Naval research by grant N00014-17-1-2552, ANR-DGA project ANR-12-ASTR-0035.
- The French Research Agency (ANR) under grant nro ANR-14-CE27-001 (MIRIAM)

6 Appendix

6.1 Proof of theorem 15

By proposition 13 we know that

$$\tau(BA^{-1}) = \tau(i([B])i([A])^{-1})$$
(a) Proposed covering for ASIFT in [44, 64]. This is a log 1.8-covering of \( \{ T_t R_\phi \mid t \leq 5.5 \} \) with 41 affine simulations representing an area ratio of 13.77.

(b) Proposed covering for FAIR-SURF in [49], called fixed tilts. This is a log 1.5-covering of \( \{ T_t R_\phi \mid t \leq 1.7 \} \) with 23 affine simulations representing an area ratio of 11.42.

(c) Proposed covering for FAIR-SURF in [49], called simulated tilts. This is a log 1.5-covering of \( \{ T_t R_\phi \mid t \leq 1.65 \} \) with 41 affine simulations representing an area ratio of 13.77.

(d) Proposed covering in [40], called MEDIUM configuration for DoG-SIFT. This is a log 1.8-covering of \( \{ T_t R_\phi \mid t \leq 1.8 \} \) with 45 affine simulations representing an area ratio of 9.

(e) Proposed covering in [40], called HARD configuration for DoG-SIFT. This is a log 1.8-covering of \( \{ T_t R_\phi \mid t \leq 9.6 \} \) with 61 affine simulations representing an area ratio of 13.

(f) Proposed covering in [40], called HARD configuration for SURF-SURF. This is a log 1.5-covering of \( \{ T_t R_\phi \mid t \leq 1.5 \} \) with 112 affine simulations representing an area ratio of 21.28.

Figure 14: Examples of coverings found in the literature for maximum tilt tolerances as in Figure 11.

- Gray areas - Uncovered.
- Blue areas - Covered by at least two disks.
- White areas - Covered by only one disk.
Figure 15: Graffiti. Both images generate a large number amount of keypoints for most methods.

Figure 16: Adam. Both images generate a small number of keypoints for most methods.
where \( i \) is the injection in Definition 11. Thus, without loss of generality, we focus in computing the absolute tilt of
\[
C = T_i R_2 Q_2^{-1} T_s^{-1}
\]
where \( R(\phi) = R_2 Q_2^{-1} \). Proposition 5 states that the ratio between the singular values of \( C \) can be used to compute its absolute tilt.

6.1.1 Trace and determinant
First, we start by computing the trace and determinant of
\[
C^* C = T_s^{-1} R(\phi)^{-1} T_i T_i R(\phi) T_s^{-1},
\]
which are clearly
\[
\det (C^* C) = \frac{t^2}{s^2}
\]
and
\[
Tr (C^* C) = \left( \frac{t^2}{s^2} + 1 \right) \cos^2 \phi + \left( \frac{1}{s^2} + t^2 \right) \sin^2 \phi.
\]

6.1.2 The eigenvalues of \( C^* C \)
Let \( H = \begin{pmatrix} a & c \\ c & b \end{pmatrix} = C^* C \) and \( \lambda_+, \lambda_- \) being the biggest and smallest eigenvalues of \( C^* C \) respectively. It is well known that
\[
Tr (H) = \lambda_+ + \lambda_-
\]
\[
\det (H) = \lambda_+ \lambda_- 
\]
and even more that both \( Tr \) and \( \det \) also appear in the characteristic polynomial
\[
|C^* C - \lambda Id| = \lambda^2 - \lambda (a + b) + (ab - c^2) = \lambda^2 - 4 Tr H + \det H.
\]
On the other hand, the eigenvalues of a symmetric positive definite matrix are in \( \mathbb{R} \), which implies that \( \sqrt{(Tr H)^2 - 4 \det H} \geq 0 \), and so one can write
\[
\lambda_- = \frac{Tr (H) - \sqrt{(Tr H)^2 - 4 \det H}}{2},
\]
\[
\lambda_+ = \frac{Tr (H) + \sqrt{(Tr H)^2 - 4 \det H}}{2}.
\]
Now, after some computations, the ratio between the biggest and smallest eigenvalues is
\[
\frac{\lambda_+}{\lambda_-} = \frac{\left( \frac{Tr H}{2} + \sqrt{(Tr H)^2 - 4 \det H} \right)^2}{\det H}
\]
\[
= \frac{s^2}{t^2} \left( \frac{g}{2} + \sqrt{\frac{g^2 - 4 t^2}{2}} \right)^2
\]
where \( g \) denotes the function
\[
g(t, s, \phi) := Tr (C^* C)
\]
\[
= \left( \frac{t^2}{s^2} + 1 \right) \cos^2 \phi + \left( \frac{1}{s^2} + t^2 \right) \sin^2 \phi.
\]
6.1.3 Computing \( \tau(C) \)

Proposition 5 tells that the absolute tilt of \( C \) is

\[
\tau(C) = \sqrt{\frac{\lambda_+}{\lambda_-}}
\]

\[
= \frac{s}{t} \left( \frac{g}{2} + \frac{\sqrt{g^2 - 4t^2}}{2} \right)
\]

\[
= \frac{s \cdot g}{t \cdot 2} + \sqrt{\left( \frac{s \cdot g}{t \cdot 2} \right)^2 - 1}
\]

\[
= G(s,t,\phi) + \sqrt{(G(s,t,\phi))^2 - 1}
\]

where

\[
G(s,t,\phi) = \frac{s \cdot g(s,t,\phi)}{t}.
\]

6.1.4 Disks in the space of tilts

Let \( A := [T_1 R_2] \in \Omega \) be fixed and let us find conditions on \( B := [T_3 Q_2] \in \Omega \) to satisfy

\[
B \in B(A, \log r)
\]

which are clearly

\[
d(A, B) = \log \tau \left( i(A) i(B)^{-1} \right) \leq \log r
\]

\[
\Downarrow
\]

\[
\tau \left( i(A) i(B)^{-1} \right) \leq r
\]

where \( i \) is the injection in Definition 11. Thus, just by applying the above to \( C := i(A) i(B)^{-1} \) we obtained

\[
G(s,t,\phi) + \sqrt{(G(s,t,\phi))^2 - 1} = \tau(AB^{-1}) \leq r
\]

where \( R(\phi) = R_2 Q_2^{-1} \). So

\[
\sqrt{G^2 - 1} \leq r - G
\]

\[
\Downarrow
\]

\[
G^2 - 1 \leq r^2 - 2rG + G^2
\]

\[
\Downarrow
\]

\[
G \leq \frac{r^2 + 1}{2r}.
\]

References


