Covering the Space of Tilts: Application to affine invariant image comparison

Mariano Rodríguez, Julie Delon, Jean-Michel Morel

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

[*]CMLA, ENS Paris-Saclay
[†]MAP5, University Paris Descartes
[‡]CMLA, ENS Paris-Saclay

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 $u(\mathbf{x}) = u(x, y)$, 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 $u(\mathbf{x})$ by $Au(\mathbf{x}) = u(A\mathbf{x})$. We recall classic notation for three subsets of the general linear group $GL(2)$ of invertible linear maps of the plane,

$$\begin{aligned} GL^{+}(2) &= \{A \in GL(2) \mid \det(A) > 0\}, \\ GO^{+}(2) &= \left\{A \in GL^{+}(2) \mid A \text{ is a similarity}\right\}, \\ GL^{+}_{*}(2) &= GL^{+}(2) \setminus GO^{+}(2), \end{aligned}$$

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 *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*

$$A = \lambda R_1(\psi) T_t R_2(\phi) \tag{1}$$

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 λR_1.

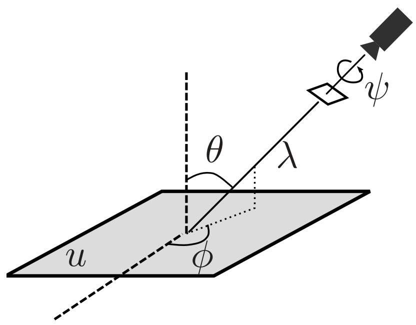


Figure 1: Geometric Interpretation of (1)

Figure 1 shows a camera viewpoint interpretation of this affine decomposition where the longitude ϕ and latitude $\theta = arccos\,\frac{1}{t}$ characterize the camera's viewpoint angles, ψ parameterizes the camera spin and λ 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 *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 *absolute tilt* of A the real number $\tau(A)$ defined by

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

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

$$A^{\star}A = R_2^t \begin{pmatrix} \gamma_1^2 & 0 \\ 0 & \gamma_2^2 \end{pmatrix} R_2$$

whose eigenvalues are

$$\lambda_1 = \gamma_1^2 \text{ and } \lambda_2 = \gamma_2^2$$

but $\gamma_1, \gamma_2 > 0$ imply

$$A = \sqrt{\lambda_2} R_1 \begin{pmatrix} \sqrt{\frac{\lambda_1}{\lambda_2}} & 0 \\ 0 & 1 \end{pmatrix} R_2$$

and finally $\tau(A) = \sqrt{\frac{\lambda_1}{\lambda_2}}$. In addition, it is well known that

$$|||A|||_2 = \sqrt{\rho(A^{\star}A)} = \sqrt{\lambda_1},$$

$$|||A^{-1}|||_2 = \sqrt{\rho\left((AA^{\star})^{-1}\right)} = \frac{1}{\sqrt{\lambda_2}}$$

where $\rho(A^{\star}A)$ is the largest eigenvalue of $A^{\star}A$, i.e, the largest singular value of A. □

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 $T_t R_2$ 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

$$A \sim B \quad \Leftrightarrow \quad AB^{-1} \in GO^{+}(2).$$

Remark 7. It is important to notice here that the right and left equivalence relations do differ because

$$AB^{-1} \in GO^{+} \not\Leftrightarrow B^{-1}A \in GO^{+}.$$

For example, take

$$A = T_2 R_{\frac{\pi}{4}} \text{ and } B^{-1} = R_{\frac{\pi}{4}} T_2,$$

then

$$AB^{-1} = 2R_{\frac{\pi}{2}} \in GO^{+}$$

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.

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

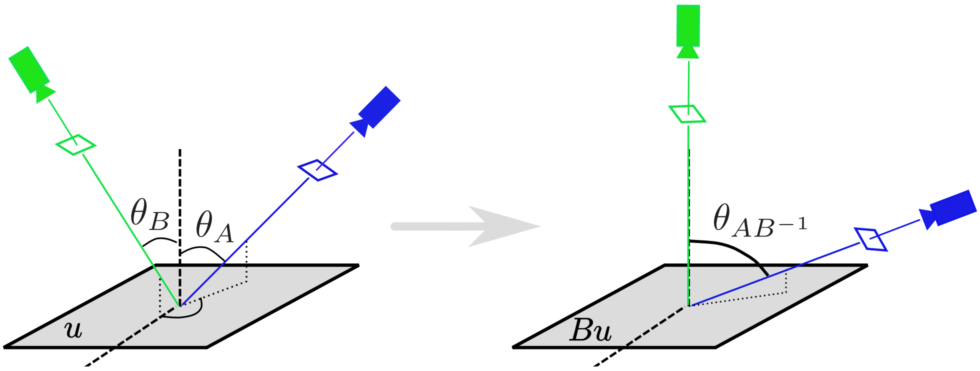


Figure 2: Passage from transition tilts (left side) to absolute tilts (right side).

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

$$\begin{aligned} \tau(A) &= |||A|||_2 \left|\left|\left|A^{-1}\right|\right|\right|_2 \\ &= \tau\left(A^{-1}\right) \end{aligned}$$

3) From 2) we have

$$\begin{aligned} \tau\left(AB^{-1}\right) &= \tau\left(\left(AB^{-1}\right)^{-1}\right) \\ &= \tau\left(BA^{-1}\right) \end{aligned}$$

4) By proposition 5

$$\begin{aligned} \tau\left(AB^{-1}\right) &= \left|\left|\left|AB^{-1}\right|\right|\right|_2 \left|\left|\left|\left(AB^{-1}\right)^{-1}\right|\right|\right|_2 \\ &\leq |||A|||_2 \left|\left|\left|B^{-1}\right|\right|\right|_2 |||B|||_2 \left|\left|\left|A^{-1}\right|\right|\right|_2 \\ &= \tau(A)\,\tau(B) \end{aligned}$$

5) From 4) we have

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

and the same relation for B. $\square$

Definition 10. We call *Space of Tilts, denoted by* Ω*,* 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.*

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

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

$$i : \left\{ \begin{array}{ccc} \Omega & \to & GL^{+}(2) \\ {[A]} & \mapsto & T_{\tau(A)} R_{\phi(A)} \end{array} \right. .$$

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

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

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*

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

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

$$\begin{array}{rcl} A & = & \lambda_A Q_A T_s R_A \\ B & = & \lambda_B Q_B T_t R_B \end{array}$$

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

$$\begin{aligned} AB^{-1} &= \lambda_{AB^{-1}} Q_{AB^{-1}} T_{\tau(AB^{-1})} R_{AB^{-1}} \\ T_s R_A R_B^{-1} T_t^{-1} &= \alpha Q_3 T_{t_3} R_3. \end{aligned} \tag{2}$$

Moreover

$$\begin{array}{rcl} AB^{-1} & = & \lambda_A Q_A T_s R_A \left(\lambda_B Q_B T_t R_B\right)^{-1} \\ & = & \dfrac{\alpha \lambda_A}{\lambda_B} \underbrace{(Q_A Q_3)}_{\text{rotation}} T_{t_3} \underbrace{\left(R_3 Q_B^{-1}\right)}_{\text{rotation}}. \end{array}$$

Then, by uniqueness of decomposition in equation (2) we have $T_{\tau(AB^{-1})} = T_{t_3}$, implying

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

Again, the same methodology applied to

$$\begin{array}{rcl} C & = & \lambda_C Q_C A \\ & = & \lambda_C \lambda_A Q_C Q_A T_s R_A \end{array}$$

and

$$\begin{aligned} D &= \lambda_D Q_D B \\ &= \lambda_D \lambda_B Q_D Q_B T_t R_B \end{aligned}$$

shows that

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

□

The next proposition follows directly from Proposition 9.

Proposition 14. *The function d*

$$d : \left\{ \begin{array}{ccc} \Omega \times \Omega & \rightarrow & \mathbb{R}_+ \\ ([A], [B]) & \mapsto & \log \tau\left(AB^{-1}\right) \end{array} \right.$$

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 $\tau\left(AB^{-1}\right) \geq 1$. This implies

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

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

$$\begin{aligned} d([A], [B]) = 0 &\Leftrightarrow \tau\left(AB^{-1}\right) = 1 \\ &\Leftrightarrow A \sim B \\ &\Leftrightarrow [A] = [B] \end{aligned}$$

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

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

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

$$\tau\left(BC^{-1}\left(AC^{-1}\right)^{-1}\right) \leq \tau\left(BC^{-1}\right)\tau\left(AC^{-1}\right).$$

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

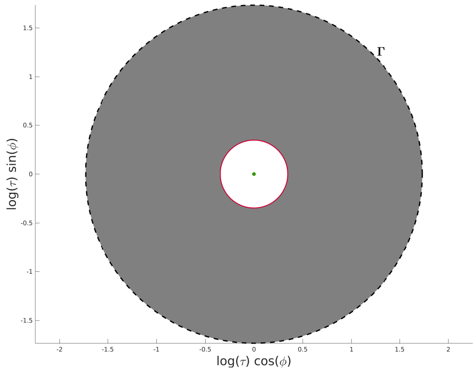
Now that we have introduced the space of tilts and the adequate metric on this space to measure image distortion, we wish to explore optimal coverings for this space. We start by establishing closed formulas for disks in this 2D space.

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

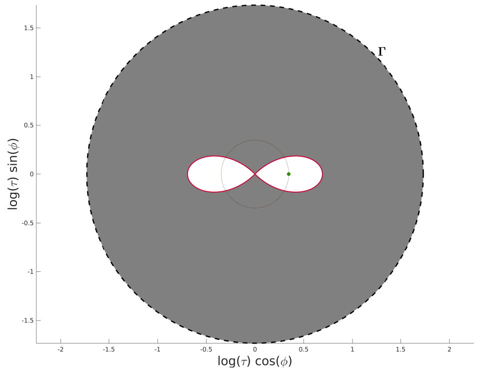
$$\left\{ [T_s R(\phi_2)] \mid G(t, s, \phi_1, \phi_2) \leq \frac{e^{2r} + 1}{2e^r} \right\}$$

where

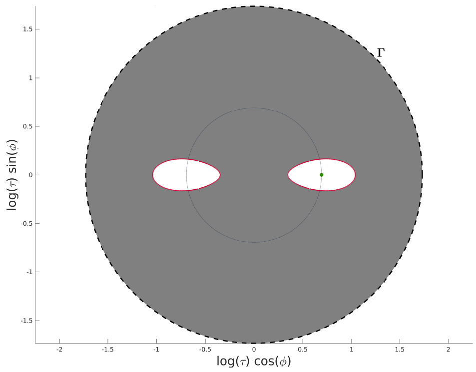
$$G(t, s, \phi_1, \phi_2) = \left(\frac{\frac{t}{s} + \frac{s}{t}}{2}\right)\cos^2(\phi_1 - \phi_2) + \left(\frac{\frac{1}{st} + st}{2}\right)\sin^2(\phi_1 - \phi_2).$$



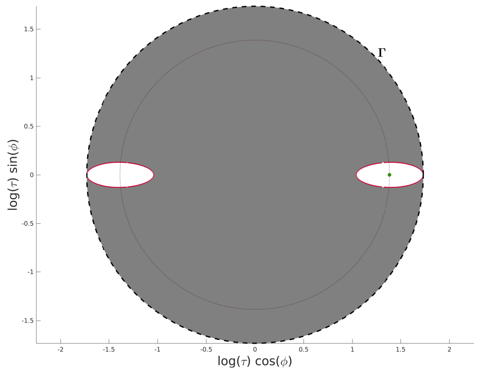
(a) $B\left([Id], \log\sqrt{2}\right)$



(b) $B\left(\left[T_{\sqrt{2}}R_0\right], \log\sqrt{2}\right)$



(c) $B\left([T_2R_0], \log\sqrt{2}\right)$



(d) $B\left([T_4R_0], \log\sqrt{2}\right)$

Figure 3: (Polar coordinates)

Green point - Affine transformation in question
Dashed line - $\partial B\left([Id], \log 4\sqrt{2}\right)$
Dotted line - Equal tilts
Red line - Disk's boundary

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 (τ, ϕ) 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 π longitude difference.

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

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

i.e, classes in Ω are stable by right multiplication. Moreover,

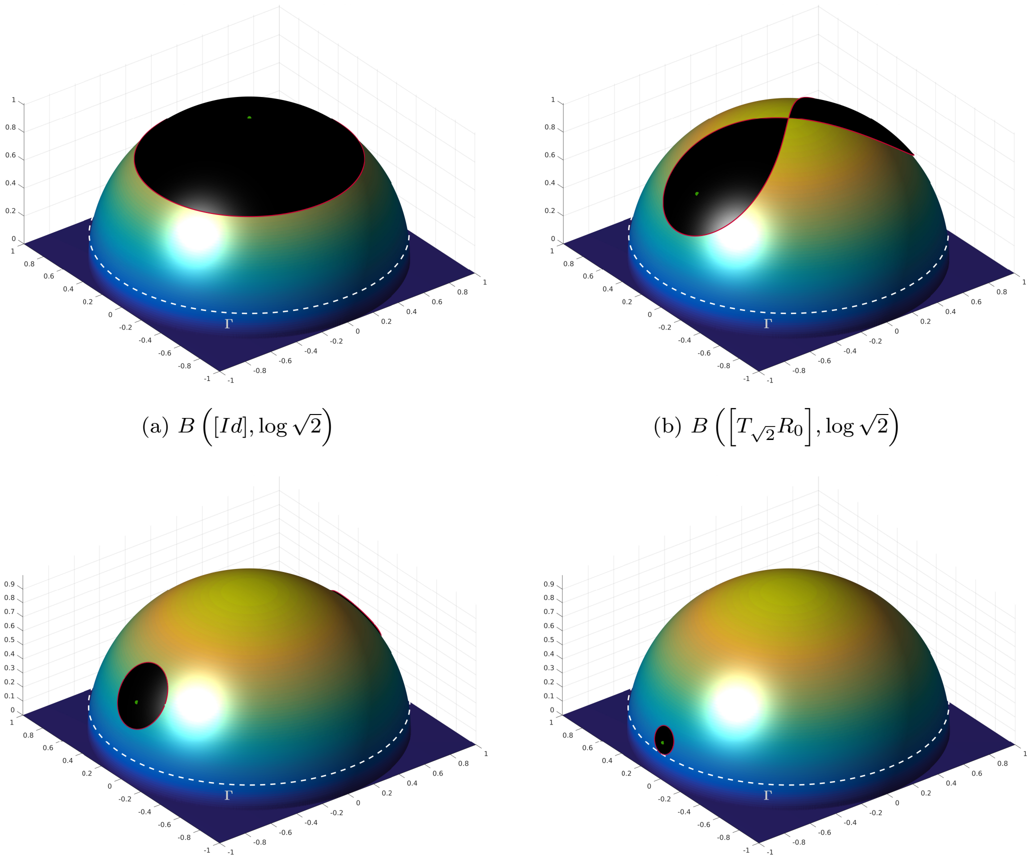
$$d\left([AC], [BC]\right) = d\left([A], [B]\right).$$

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

$$\begin{aligned} B \in [A] \quad &\Longrightarrow \quad B = \lambda RA \\ &\Longrightarrow \quad BC = \lambda RAC \\ &\Longrightarrow \quad BC \in [AC] \end{aligned}$$

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

$$\begin{aligned} D \in [AC] \quad &\Longrightarrow \quad D = \lambda RAC \\ &\Longrightarrow \quad D \in [A]\,C \end{aligned}$$



(a) $B\left([Id], \log\sqrt{2}\right)$

(b) $B\left(\left[T_{\sqrt{2}}R_0\right], \log\sqrt{2}\right)$

(c) $B\left([T_2R_0], \log\sqrt{2}\right)$

(d) $B\left([T_4R_0], \log\sqrt{2}\right)$

Figure 4: (Perspective views) Green point - Affine transformation in question
Dashed line - $\partial B\left([Id], \log 4\sqrt{2}\right)$
Black surface - Disk in question

3)

$$\begin{aligned} d\left([AC], [BC]\right) &= \log\tau\left(AC\left(BC\right)^{-1}\right) \\ &= \log\tau\left(AB^{-1}\right) \\ &= d\left(A, B\right) \end{aligned}$$

□

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

$$R_{\frac{\pi}{2}} = T_t\left(\frac{1}{t}R_{\frac{\pi}{2}}T_t\right) \notin [T_{t^2}].$$

Furthermore, for $C \in GL^{+}(2)$ one has

$$\begin{aligned}\tau\left(CAB^{-1}C^{-1}\right) &= c_2\left(CAB^{-1}C^{-1}\right)\\ &= \left|\!\left|\!\left|CAB^{-1}C^{-1}\right|\!\right|\!\right|_2 \left|\!\left|\!\left|C\left(AB^{-1}\right)^{-1}C^{-1}\right|\!\right|\!\right|_2\\ &\le \left|\!\left|\!\left|C\right|\!\right|\!\right|_2^2 \left|\!\left|\!\left|C^{-1}\right|\!\right|\!\right|_2^2 \left|\!\left|\!\left|AB^{-1}\right|\!\right|\!\right|_2 \left|\!\left|\!\left|\left(AB^{-1}\right)^{-1}\right|\!\right|\!\right|_2\\ &= \tau\left(C\right)^2\tau\left(AB^{-1}\right)\end{aligned}$$

so, in general

$$d\left(\left[CA\right],\left[CB\right]\right) \le 2d\left(\left[C\right],\left[Id\right]\right) + d\left(\left[A\right],\left[B\right]\right).$$

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

Theorem 19. *Let*

$$\begin{aligned}\Gamma_1 &= \mathcal{B}\left(\left[Id\right],\log\Lambda_1\right)\\ \Gamma_2 &= \mathcal{B}\left(\left[Id\right],\log\Lambda_2\right)\\ \Gamma' &= \mathcal{B}\left(\left[Id\right],\log\Lambda_2 r\right).\end{aligned}$$

be three neighborhoods of $[Id]$ in Ω where $\Lambda_1, \Lambda_2, r \in [1,\infty[$, and assume that $\mathbb{S}_1, \mathbb{S}_2 \subset \Omega$ are two $\log r$-coverings of Γ_1 and Γ', i.e

$$\begin{aligned}\Gamma_1 &\subset \bigcup_{S\in\mathbb{S}_1}\mathcal{B}\left(S,\log r\right)\\ \Gamma' &\subset \bigcup_{S\in\mathbb{S}_2}\mathcal{B}\left(S,\log r\right).\end{aligned}$$

Then, for every $[A] \in \Gamma_1$, $[B] \in \Gamma_2$, there exist $C \in GL^{+}(2)$ with $\tau(C) \le r$, $S_A \in \mathbb{S}_1$ and $S_B \in \mathbb{S}_2$ such that

$$\begin{aligned}d\left(S_A, \left[(AC)^{-1}\right]\right) &= 0\\ d\left(S_B, \left[(BC)^{-1}\right]\right) &\le \log r.\end{aligned}$$

A sketch of Theorem 19 appears in Figure 5.

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

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

Proposition 9-2) directly implies

$$d\left(\left[Id\right],\left[A\right]\right) = d\left(\left[Id\right],\left[A^{-1}\right]\right).$$

Then, as $\mathbb{S}_1$ is a $\log r$-covering of Γ_1, there exists $S_A \in \mathbb{S}_1$ such that

$$\left[A^{-1}\right] \in \mathcal{B}\left(S_A,\log r\right)$$

meaning that, the following inequality holds

$$\begin{aligned}d\left(\left[Id\right],\left[A^{-1}i\left(S_A\right)^{-1}\right]\right) &= \log\tau\left(A^{-1}i\left(S_A\right)^{-1}\right)\\ &= d\left(\left[A^{-1}\right], S_A\right)\\ &\le \log r.\end{aligned}$$

Finally, as d is a metric (by Proposition 14) we know

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

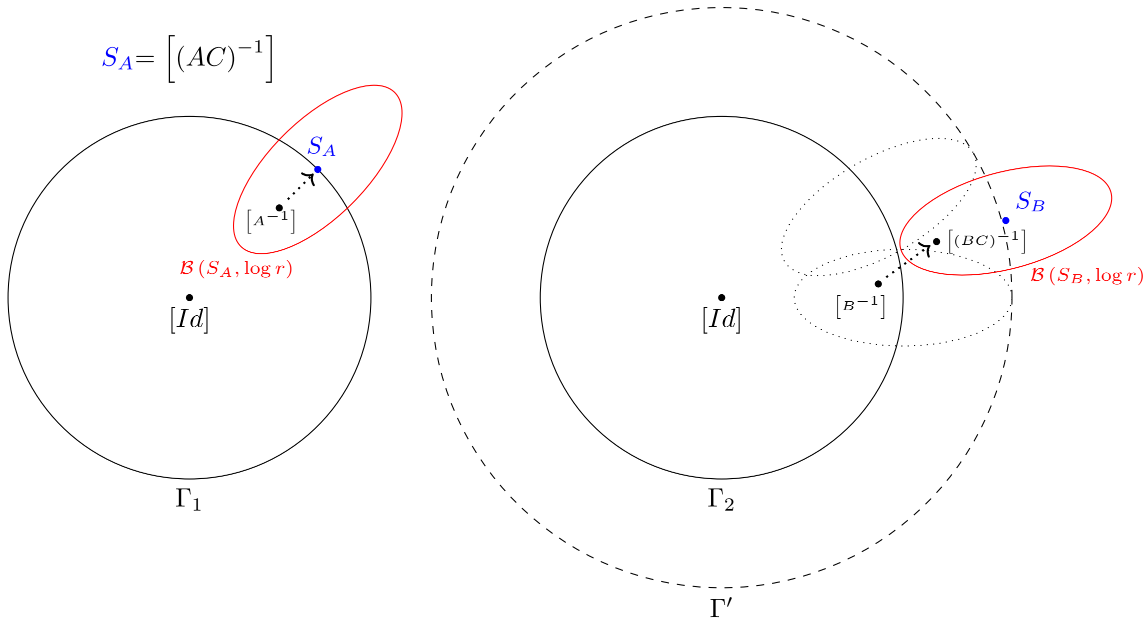


Figure 5: Sketch of Theorem 19.

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

$$\tau\left(BC\right) \leq \tau\left(B\right)\tau\left(C^{-1}\right) = \Lambda_2 r$$
$$\Downarrow$$
$$d\left(\left[Id\right], \left[(BC)^{-1}\right]\right) = \log \tau\left(BC\right) \leq \log \Lambda_2 r$$
$$\Downarrow$$
$$\left[(BC)^{-1}\right] \in \Gamma'.$$

Once more, as $\mathbb{S}_2$ is a $\log r$-covering of Γ', there exists $S_B \in \mathbb{S}_2$ such that

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

□

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 $u\left(\mathbf{x}\right)$ defined for every $\mathbf{x} = (x, y) \in \mathbb{R}^2$. 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 $A \in GL^+\left(2\right)$, we define the affine transform A of a continuous image u by

$$Au(\mathbf{x}) := u(A\mathbf{x}).$$

Homotheties and rotations acting on continuous images are similarly written as

$$H_\lambda u\left(\mathbf{x}\right) = u\left(\lambda\mathbf{x}\right);$$
$$R_\phi u\left(\mathbf{x}\right) = u\left(R_\phi\mathbf{x}\right).$$

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 $\mathbb{G}_\sigma$, $\mathbb{G}_\sigma^x$ and $\mathbb{G}_\sigma^y$ respectively the 2D and 1D convolution operators in the x and y directions with

$$\begin{aligned} G_{\mathbf{c}\sigma}(x,y) &:= \frac{1}{2\pi(\mathbf{c}\sigma)^2} e^{-\frac{x^2+y^2}{2(\mathbf{c}\sigma)^2}} \\ G^x_{\mathbf{c}\sigma}(x) &:= \frac{1}{\sqrt{2\pi}\mathbf{c}\sigma} e^{-\frac{x^2}{2(\mathbf{c}\sigma)^2}} \\ G^y_{\mathbf{c}\sigma}(y) &:= \frac{1}{\sqrt{2\pi}\mathbf{c}\sigma} e^{-\frac{y^2}{2(\mathbf{c}\sigma)^2}} \end{aligned}$$

namely

$$\begin{aligned} \mathbb{G}_\sigma u &:= G_{c\sigma} \star u \\ \mathbb{G}^x_\sigma u &:= G^x_{c\sigma} \star_x u \\ \mathbb{G}^y_\sigma u &:= G^y_{c\sigma} \star_y u. \end{aligned}$$

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 σ without causing significant aliasing.

Remark 21. $\mathbb{G}_\sigma$ satisfies the semigroup property

$$\mathbb{G}_\sigma \mathbb{G}_\beta = \mathbb{G}_{\sqrt{\sigma^2+\beta^2}}. \tag{3}$$

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

$$\mathbb{G}_\sigma H_\gamma u = H_\gamma \mathbb{G}_{\sigma\gamma} u. \tag{4}$$

In the classic Shannon-Nyquist framework, we shall denote the image sampling operator (on a unary grid) by $\mathbf{S}_1$. Thus $\mathbf{S}_1 u$ 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 $\mathbf{u} = \mathbf{S}_1 \mathbb{G}_1 A \mathcal{T} u_0$, where A is *any* linear map with positive singular values and $\mathcal{T}$ 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

$$\mathbf{u} =: \mathbf{S}_1 \mathbb{G}_1 A \mathcal{T} u_0 \tag{5}$$

where A is any linear map and $\mathcal{T}$ any plane translation. $\mathbb{G}_1$ denotes a Gaussian kernel broad enough to ensure no aliasing by 1-sampling, namely $I\mathbf{S}_1 \mathbb{G}_1 A \mathcal{T} u_0 = \mathbb{G}_1 A \mathcal{T} u_0$.

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

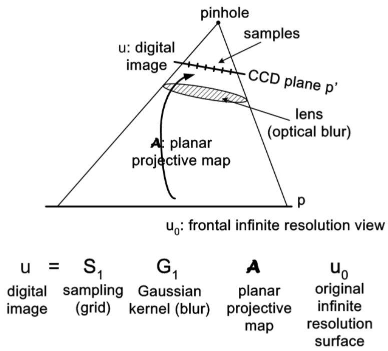


Figure 6: The projective camera model $u = \mathbf{S}_1\mathbb{G}_1\mathcal{A}u_0$. $\mathcal{A}$ is a planar projective transform (a homography). $\mathbb{G}_1$ is an anti-aliasing Gaussian filtering. $\mathbf{S}_1$ is the CCD sampling.

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

$$T_t^x u_0(x,y) =: u_0(tx,y);$$
$$T_t^y u_0(x,y) =: u_0(x,ty).$$

- Simulated tilts (taking into account camera blur)

$$\mathbb{T}_t^x v =: T_t^x \mathbb{G}^x_{\sqrt{t^2-1}} \star_x v;$$
$$\mathbb{T}_t^y v =: T_t^y \mathbb{G}^y_{\sqrt{t^2-1}} \star_y v.$$

- Digital tilts (transforming a digital image u into a digital image)

$$\mathbf{u} \to \mathbf{S}_1 \mathbb{T}_t^x I\mathbf{u};$$
$$\mathbf{u} \to \mathbf{S}_1 \mathbb{T}_t^y I\mathbf{u}.$$

Digital tilts are the ones used in practice. It all adds up because the simulated tilt yields a blur permitting $\mathbf{S}_1$-sampling.

If u_0 is an infinite resolution image observed with a camera tilt of t in the x direction, the observed image is $\mathbb{G}_1 T_t^x u_0$. 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 $\mathbb{T}_t^y$ is, up to a zoom out, a pseudo inverse to T_t^x.

The meaning of this result is that a tilted image $\mathbb{G}_1 T_t^x u$ 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 $\mathbf{S}_1$ yields an image that keeps that well sampling property.

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

$$\mathbb{T}_t^y \mathbb{G}_1 T_t^x = \mathbb{G}_1 H_t; \tag{6}$$
$$\mathbb{T}_t^y \mathbb{G}_1 = \mathbb{G}_1 T_t^y. \tag{7}$$

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

$$\mathbb{G}_1 = \mathbb{G}_1^x \mathbb{G}_1^y = \mathbb{G}_1^y \mathbb{G}_1^x \tag{8}$$

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

$$\mathbb{G}_1^x T_t^y = T_t^y \mathbb{G}_1^x \tag{9}$$

which is true because $\mathbb{G}_1^x$ and T_t^y act separably on the variables x and y. Using first (4) in the y dimension where T_t^y is a mere homothety, and then successively (9), (8), the semigroup property for the Gaussians, and Definition 23 we get

$$\begin{aligned}
T_t^y \mathbb{G}_t^y &= \mathbb{G}_1^y T_t^y \Rightarrow \\
\mathbb{G}_1^x T_t^y \mathbb{G}_t^y &= \mathbb{G}_1^x \mathbb{G}_1^y T_t^y \Rightarrow \\
T_t^y \mathbb{G}_t^y \mathbb{G}_1^x &= \mathbb{G}_1 T_t^y \Rightarrow \\
T_t^y \mathbb{G}_{\sqrt{t^2-1}}^y \mathbb{G}_1^y \mathbb{G}_1^x &= \mathbb{G}_1 T_t^y \Rightarrow \\
\mathbb{T}_t^y \mathbb{G}_1 &= \mathbb{G}_1 T_t^y,
\end{aligned}$$

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 $u = \mathbb{G}_1 A \mathcal{T}_1 w_0$ and $v = \mathbb{G}_1 B \mathcal{T}_2 w_0$ 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

$$\mathbb{S}_i = \left\{ \left[T^x_{t^i_k} R_{\phi^i_k} \right] \right\}_{k=1,\dots,n_i}.$$

Input:

Query and target images: u and v.

Start:

1: $\forall k = 1, ..., n_1$ do

$$u_k = \mathbb{T}^x_{t^1_k} R_{\phi^1_k} u.$$

2: $\forall k = 1, ..., n_2$ do

$$v_k = \mathbb{T}^x_{t^2_k} R_{\phi^2_k} v.$$

3: $\forall (k_1, k_2) \in \{1, ..., n_1\} \times \{1, ..., n_2\}$

$$M_{k_1,k_2} = \text{SIIM-Matches}(u_{k_1}, v_{k_2}).$$

Output:

$$M = \bigcup_{(k_1,k_2)\in\{1,\dots,n_1\}\times\{1,\dots,n_2\}} M_{k_1,k_2}.$$

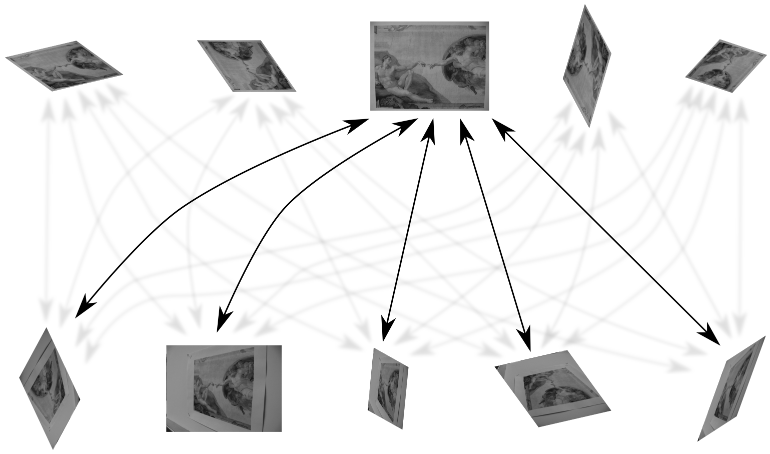


Figure 7: IMAS algorithms start by applying a finite set of optical affine simulations to u and v, followed by pairwise comparisons.

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*

$$\tau(A) \leq \Lambda_1 \text{ and } \tau(B) \leq \Lambda_2$$

such that

$$u = \mathbb{G}_1 A \mathcal{T}_1 w_0 \quad \text{and} \quad v = \mathbb{G}_1 B \mathcal{T}_2 w_0 \tag{10}$$

where $\mathcal{T}_1, \mathcal{T}_2$ 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 $S_A \in \mathbb{S}_1$, $S_B \in \mathbb{S}_2$ and $C \in GL^+(2)$ with $\tau(C) \leq r$ such that

$$\begin{aligned} d\left(S_A, \left[(AC)^{-1}\right]\right) &= 0 \\ d\left(S_B, \left[(BC)^{-1}\right]\right) &\leq \log r. \end{aligned}$$

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

$$u = \mathbb{G}_1 AC \mathcal{T}_1 w_1 \quad \text{and} \quad v = \mathbb{G}_1 BC \mathcal{T}_2 w_1.$$

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 $i(S_A) = T^x_{t_A} R_{\phi_A}$ and $i(S_B) = T^x_{t_B} R_{\phi_B}$ respectively on the query and target images. This is:

1. $\mathbb{T}^x_{t_A} R_{\phi_A}$ to u, which yields

$$\begin{aligned} \tilde{u} &= \mathbb{G}_1 i(S_A) AC \mathcal{T}_1 w_1 \\ &= \mathbb{G}_1 \lambda R \mathcal{T}_1 w_1. \end{aligned}$$

2. $\mathbb{T}^x_{t_B} R_{\phi_B}$ to v, which yields

$$\tilde{v} = \mathbb{G}_1 i(S_B) BC \mathcal{T}_2 w_1.$$

But

$$\begin{aligned} d([Id], [i(S_B) BC]) &= \log \tau(i(S_B) BC) \\ &= d\left(S_B, \left[(BC)^{-1}\right]\right) \\ &\leq \log r \end{aligned}$$

which proves that the affine relation between $\tilde{u}$ and $\tilde{v}$ involves a transition tilt under r. $\square$

Remark 26. Two $\log r$-coverings of the same region

$$\Gamma = \mathcal{B}\left([Id], \log \Lambda\right)$$

would then ensure that the formal IMAS Algorithm 1 manages to reduce transition tilts under $\frac{\Lambda^2}{r}$ 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.

Covered absolute tilts $(\tau(A) \leq \sqrt{r}\Lambda$ **and** $\tau(B) \leq \sqrt{r}\Lambda)$	**Attainable transition tilts** $\left(\tau\left(AB^{-1}\right) \leq \Lambda^2\right)$	**Viewpoint angle** $\left(arccos\frac{1}{\Lambda^2}\right)$
$\Lambda = 8$	64	89.1°
$\Lambda = 4\sqrt{2}$	32	88.2°
$\Lambda = 4$	16	86.4°
$\Lambda = 2\sqrt{2}$	8	82.8°
$\Lambda = 2$	4	75.5°
$\Lambda = \sqrt{2}$	2	60°

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 $\mathbb{S}_1, \mathbb{S}_2 \subset \Omega$ 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 those 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 $\mathbb{S}$ (a finite set of elements in Ω) the real number

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

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 $\mathbb{S}$ whose area ratio is close to the infimum among all $\log r$-coverings of Γ. 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 $\mathbb{S}$. Those constraints follow from simple requirements for an image matching method.

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

1. $[Id] \in \mathbb{S}$.

2. There exist $n \in \mathbb{N}^+$ and

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

such that

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

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

Remark 29. Definition 28-1) avoids an image resolution loss before comparison, an obvious requirement. Imposing groups of concentric equidistant tilts as in Definition 28-2) is a sound isotropy requirement.

Definition 30. Set $\Gamma = \mathcal{B}([Id], \log\Lambda)$. A feasible set $\mathbb{S} \in \Omega$ with parameters

$$(n, (t_1, ..., t_n, \phi_1, ..., \phi_n)) \in \mathbb{N}^+ \times [1, \infty[^n \times]0, \pi]^n$$

is said to be optimal among feasible sets if and only if it realizes the minimal area ratio. In other words, optimal feasible sets are solutions of:

$$\begin{aligned} &\underset{(n,(t_1,...t_n,\phi_1,...\phi_n))\in\mathbb{N}^+\times[1,\infty[^n\times]0,\pi]^n}{\arg\min} \quad 1 + \sum_{i=1}^{n} \frac{|J_{t_i,\phi_i}|}{t_i} \\ &\text{subject to:} \quad \Gamma \subset \mathcal{B}_{[Id]}^{\log r} \cup \left\{ \bigcup_{1\leq i\leq n} \bigcup_{S\in J_{t_i,\phi_i}} \mathcal{B}_{[S]}^{\log r} \right\} \end{aligned} \tag{11}$$

where J_{t_i,ϕ_i} is the set of transformations of the form

$$T_{t_i}R_{\phi_i},\ T_{t_i}R_{2\phi_i}\ ...,\ T_{t_i}R_{\left\lfloor \frac{\pi}{\phi_i} \right\rfloor \phi_i},$$

$|J_{t_i,\phi_i}|$ is the cardinal of J_{t_i,ϕ_i} and $\mathcal{B}_{[S]}^{\log r}$ is denoting $\mathcal{B}([S], \log r)$.

Fortunately for our problem with the realistic values $\Lambda = 6$ and $r = 1.8$, $n = 2$ can be fixed, as easy heuristics indicate that any covering with $n > 2$ has a far too large area ratio. Thus our optimization in a realistic setting ends up being performed in dimension 4 for sets $(t_1, t_2, \phi_1, \phi_2)$. With n thus fixed the optimization problem in (11) can be exhaustively optimized. In this minimization we deal with 4 dimensions and more specifically with 100^4 feasible sets by sampling each parameter. This yields an almost exact discrete exhaustive optimization by sampling densely the explored set $(t_1, t_2, \phi_1, \phi_2)$ with 100 different values for each parameter. The next proposition describes the result of this optimization and verifies that it is indeed feasible.

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* $\Gamma = \{[T_tR_\phi] \mid t \leq 6\}$*,* $n = 2$*. Therefore, the infimum area ratio among all* $\log 1.8$*-coverings of* $\{[T_tR_\phi] \mid t \leq 6\}$ *is lower than* 6.34*.*

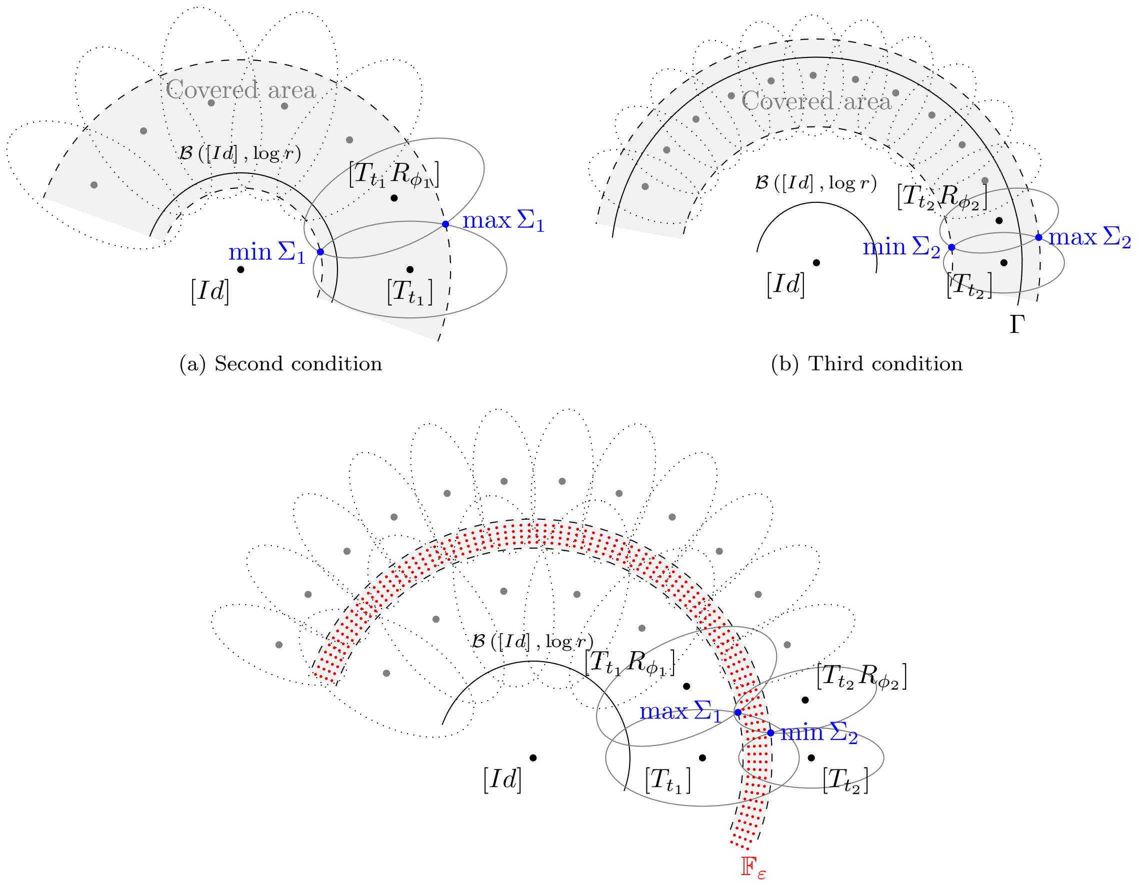
Proof. We are dealing with 4 dimensions to minimize and more specifically with 100^4 feasible sets. Computing area ratios for each feasible set is straightforward but validating the covering condition is a more involved computational issue. For the sake of clearness, the intersection of disks boundaries,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

$$\varSigma_1 = \partial\mathcal{B}_{[T_{t_1}]}^{\log 1.8} \cap \partial\mathcal{B}_{[T_{t_1}R_{\phi_1}]}^{\log 1.8} \qquad \varSigma_2 = \partial\mathcal{B}_{[T_{t_2}]}^{\log 1.8} \cap \partial\mathcal{B}_{[T_{t_2}R_{\phi_2}]}^{\log 1.8}$$

and their respective closest and farthest elements will be denoted by

$$\begin{aligned} \min \varSigma_1 &:= \arg\min_{S\in\varSigma_1} d(S, [Id]) \qquad & \max \varSigma_1 &:= \arg\max_{S\in\varSigma_1} d(S, [Id]), \\ \min \varSigma_2 &:= \arg\min_{S\in\varSigma_2} d(S, [Id]) \qquad & \max \varSigma_2 &:= \arg\max_{S\in\varSigma_2} d(S, [Id]). \end{aligned}$$

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:



(a) Second condition

(b) Third condition

(c) Fourth condition

Figure 8: Verifying covering conditions for feasible sets in Proposition 31.

1. $\varSigma_1 \neq \emptyset$ and $\varSigma_2 \neq \emptyset$.
2. $\min \varSigma_1$ must lie inside the ball $\mathcal{B}_{[Id]}^{\log 1.8}$, which ensures a covering of $\mathcal{B}_{[Id]}^{\log \tau(\max \varSigma_1)}$.
3. $\max \varSigma_2$ must lie outside the region Γ, which ensures a covering of the annulus defined by $\Gamma \setminus \mathcal{B}_{[Id]}^{\log \tau(\min \varSigma_2)}$.
4. For ε small, all elements $S \in \mathbb{F}_\varepsilon$ must lie inside some disks of radius $\log(1.8 - \varepsilon)$, i.e.

$$S \in \bigcup_{1 \leq i \leq 2} \bigcup_{S' \in J_{t_i, \phi_i}} \mathcal{B}_{[S']}^{\log(1.8-\varepsilon)},$$

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

$$\mathcal{B}_{[Id]}^{\log \tau(\min \varSigma_2)} \setminus \mathcal{B}_{[Id]}^{\log \tau(\max \varSigma_1)}.$$

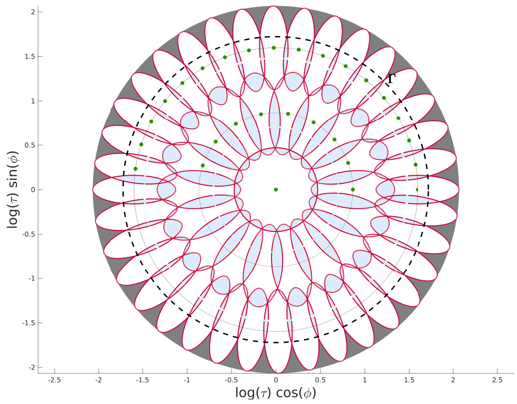
Notice that the fourth condition only ensures a $\log(1.8 - \varepsilon)$-covering up to an error

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

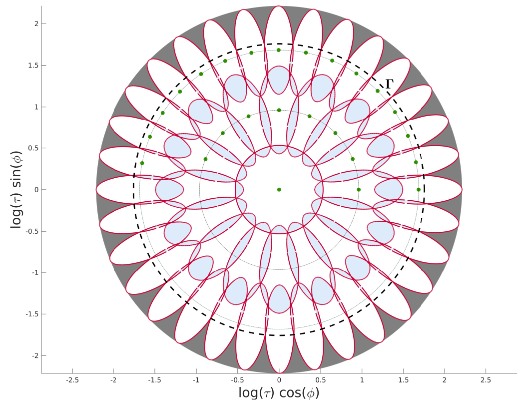
and so, by dilating back disks radius to 1.8 one ensures $\log 1.8$-coverings.

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. $\square$

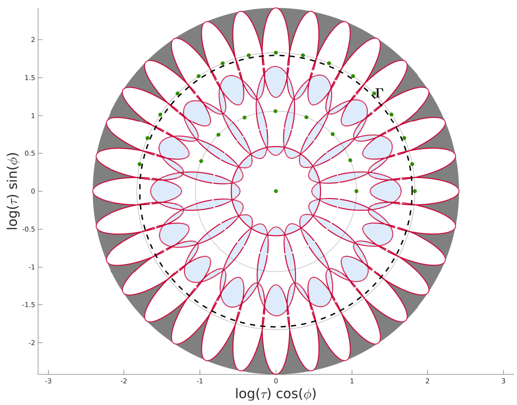
The procedure in the proof of Proposition 31 has also been applied to find more near optimal coverings appearing in Figure 9.



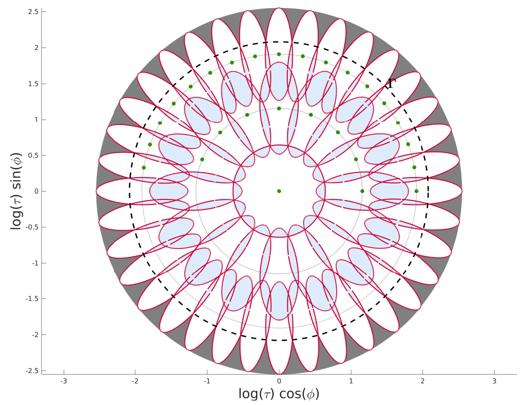
(a) Optimal $\log 1.6$-covering of $\left\{\left[T_t R_\phi\right] \mid t \leq 5.6\right\}$ with 28 affine simulations representing an area ratio of 8.42.



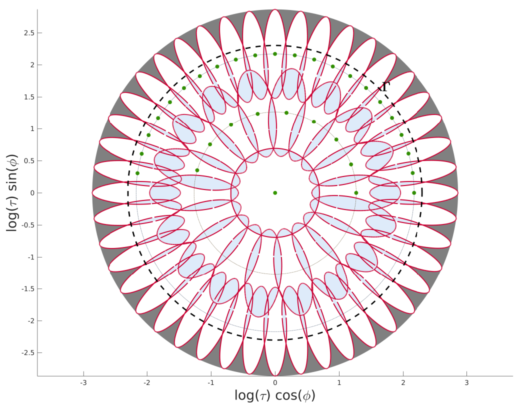
(b) Optimal $\log 1.7$-covering of $\left\{\left[T_t R_\phi\right] \mid t \leq 5.8\right\}$ with 25 affine simulations representing an area ratio of 7.06.



(c) Optimal $\log 1.8$-covering of $\left\{\left[T_t R_\phi\right] \mid t \leq 6\right\}$ with 25 affine simulations representing an area ratio of 6.34.



(d) Optimal $\log 1.9$-covering of $\left\{\left[T_t R_\phi\right] \mid t \leq 8\right\}$ with 27 affine simulations representing an area ratio of 6.18.



(e) Optimal $\log 2$-covering of $\left\{\left[T_t R_\phi\right] \mid t \leq 10\right\}$ with 32 affine simulations representing an area ratio of 6.02.

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.

Parameter	Value
t_1^{opt}	2.88447
ϕ_1^{opt}	0.394085
t_2^{opt}	6.2197
ϕ_2^{opt}	0.196389

Table 2: Approximated solution to the optimization problem in (11)

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 $t_{\max}^{s_1\times s_2}$ 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

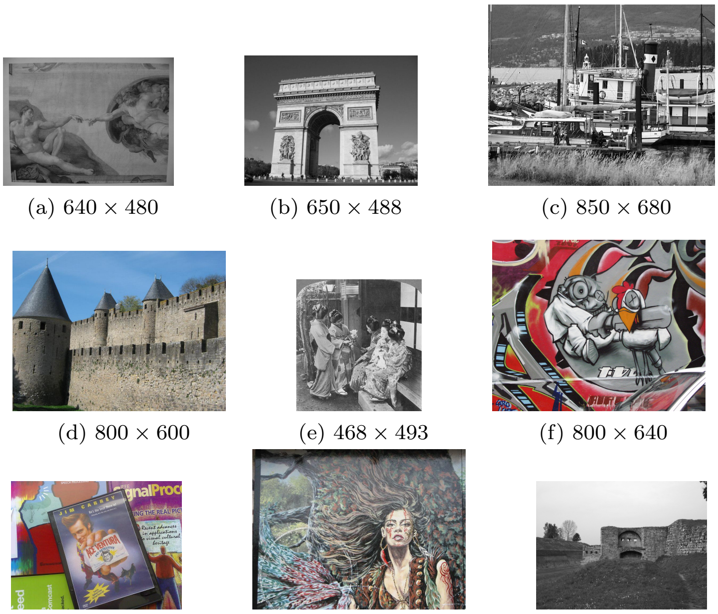
From the complexity viewpoint, the main quantitative parameter for extending a SIIM into an IMAS is its tilt tolerance. We do not question the invariance of descriptors with respect to zoom and rotations but rather how they perform against transition tilts changes incurred when matching, for example, $\mathbb{G}_1 Id\, u$ to $\mathbb{G}_1 T_t R_\phi u$ 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 $t_{\max}^{700\times 550}$, 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 $U_{\max}^{700\times 550}$ depicted in Figure 11 for nine of the best state-of-the-art SIIMs.

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

$$t_{\max}^{700\times 550} = \frac{U_{\max}^{700\times 550}}{\frac{1}{\left|\cos\left(\frac{\pi}{10}\right)\right|}} \approx \frac{U_{\max}^{700\times 550}}{1.05}$$

and constructed its affine invariant version with $\log t_{\max}^{700\times 550}$-coverings.



(a) 640×480 (b) 650×488 (c) 850×680

(d) 800×600 (e) 468×493 (f) 800×640

(g) 640×480 (h) 800×600 (i) 640×480

Figure 10: Tolerance image dataset.

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

$$\Gamma_1 = \Gamma' = \{[T_t R_\phi] \mid t \leq \Lambda\}$$

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 $\frac{\Lambda^2}{r}$ 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 $\left[(BC)^{-1}\right]$ 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), $\mathbb{G}_1 A_{t,\phi} w_0$ where $A_{t,\phi} = R_\phi T_t R_{\frac{\pi}{2}}$.
- Target image (fixed tilt), $\mathbb{G}_1 B_\phi w_0$ where $B_\phi = R_{\phi+\frac{\pi}{2}} T_\Lambda$.

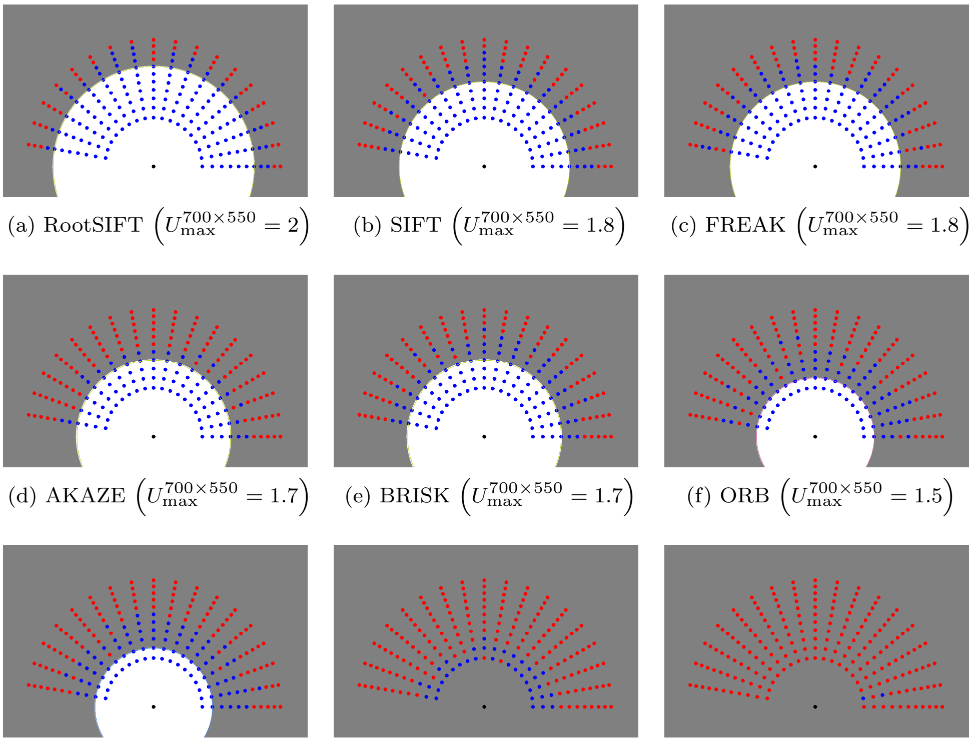
The veritable interest of these affine maps being the inverse maps they determine, namely,

$$\begin{aligned}\left[A_{t,\phi}^{-1}\right] &= \left[T_t R_{\frac{\pi}{2}-\phi}\right], \\ \left[B_\phi^{-1}\right] &= \left[T_\Lambda R_\phi\right],\end{aligned}$$

which according to Proposition 9-4, attain maximal transition tilts for fixed tilts such as t and Λ, 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 Γ' and Γ_2 are truly $\log r$-covered then



(a) RootSIFT $\left(U_{\max}^{700\times550}=2\right)$ (b) SIFT $\left(U_{\max}^{700\times550}=1.8\right)$ (c) FREAK $\left(U_{\max}^{700\times550}=1.8\right)$

(d) AKAZE $\left(U_{\max}^{700\times550}=1.7\right)$ (e) BRISK $\left(U_{\max}^{700\times550}=1.7\right)$ (f) ORB $\left(U_{\max}^{700\times550}=1.5\right)$

(g) SURF $\left(U_{\max}^{700\times550}=1.5\right)$ (h) LATCH $\left(U_{\max}^{700\times550}<1.4\right)$ (i) BRIEF $\left(U_{\max}^{700\times550}\ll 1.4\right)$

Figure 11: Represented in the space of tilts, the associated upper bounds ($U_{\max}^{700\times550}$) for maximum tilt tolerances.
Black dot - $[Id]$.
Coloured dots stand for tested tilts $[T_t R_\phi]$ where $t \in \{1.4, 1.5, \cdots, 2.4\}$ and $\phi \in \{0, 10, \cdots, 170\}$.
Blue dots - attainable tilts for all images in the dataset.
Red dots - unattainable tilts for at least one image in the dataset.
Gray areas - $\left\{[T_t R_\phi] \,|t \geq U_{\max}^{700\times550}\right\}$.
White areas - $\left\{[T_t R_\phi] \,|t \leq U_{\max}^{700\times550}\right\}$.

Proposition 25 implies that all tests for which $\left[A_{t,\phi}^{-1}\right] \in \Gamma_1$ 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 $\frac{\Lambda^2}{r}$.

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

Figure 12: Image w_0 (3264×1836) for the IMAS efficiency test

found in Table 3 and Table 4. In Table 3, Affine-ORB and Affine-BRIEF both fail because of too many false matches. The best scores found by ORSA to identify meaningful homographies were respectively 16 out of 905 and 6 out of 1409. Code optimization, smart tweaks and parallelism performance may vary from SIIM to SIIM and from IMAS to IMAS, which ultimately may lead to discrepant area ratio predictions on computation time. This is the case of SURF (and optimal Affine-SURF) whose implementation uses several fine and clever optimizations. Nonetheless, the optimal Affine-SIFT yields more matches for a lower computation time.

In Table 4 the reader will notice that Affine-ORB has less matches than ORB itself, which might seem contradictory. This happens when post-processing the matches, more specifically, when applying the second filter. The *multiple-to-one/one-to-multiple* filter, initially proposed in [44, 64], is meant to filter out undesired aberrant matches but, unfortunately, many good ones get also eliminated. In spite of this handicap, Affine-ORB is able to catch more matches with higher transition tilts.

5 Conclusion

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

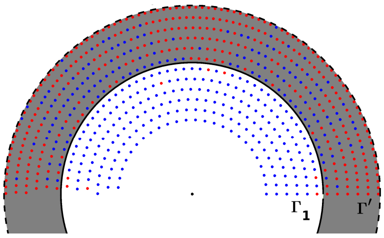
This work has been partially funded by:

	M	ar	ar^2	Keypoints (seconds)	Matching (seconds)	Filters (seconds)
SIFT	0	1	1	0.69	0.70	0.18
ASIFT	1013	13.7	189.6	12.46	138.59	3.05
(Optimal) Affine-SIFT	**795**	**7.06**	**49.8**	**6.04**	**29.61**	**1.39**
RootSIFT	0	1	1	0.72	0.71	0.18
Affine-RootSIFT	**658**	**6.9**	**47.6**	**5.05**	**20.70**	**1.44**
SURF	0	1	1	1.01	0.79	0.19
(Optimal) Affine-SURF	**471**	**14.82**	**219,6**	**12.53**	**35.24**	**1.40**
BRISK	0	1	1	1.75	0.27	0.18
Affine-BRISK	**421**	**8.42**	**70,89**	**18.95**	**8.68**	**2.06**
BRIEF	0	1	1	0.05	0.01	0.19
Affine-BRIEF	**0**	**14.82**	**219,6**	**4.20**	**2.18**	**6.08**
ORB	0	1	1	0.05	0.02	0.17
Affine-ORB	**0**	**14.82**	**219,6**	**4.34**	**5.13**	**3.25**
AKAZE	0	1	1	0.42	0.13	0.21
Affine-AKAZE	**194**	**8.42**	**70,89**	**5.00**	**6.23**	**3.74**
LATCH	0	1	1	0.11	0.02	0.00
Affine-LATCH	**37**	**14.82**	**219,6**	**4.52**	**2.16**	**0.17**
FREAK	0	1	1	0.34	0.15	0.18
Affine-FREAK	**145**	**7.06**	**49.8**	**4.37**	**2.38**	**1.94**

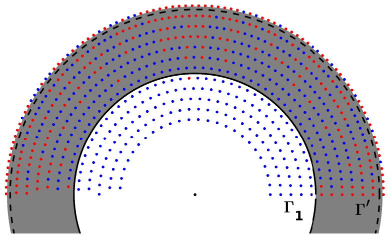
Table 3: Matching methods performance over query and target images from Figure 15. The proposed matching methods in this paper appear in bold. Computations were performed on an Intel(R) Core(TM) i5-4210U CPU 1.70GHz with 2 cores.
M - Matches.
ar - area ratio.

	M	ar	ar^2	Keypoints (seconds)	Matching (seconds)	Filters (seconds)
SIFT	102	1	1	0.23	0.01	0.09
ASIFT	317	13.7	189.6	5.43	1.68	0.47
(Optimal) Affine-SIFT	**292**	**7.06**	**49.8**	**2.71**	**0.38**	**0.30**
RootSIFT	110	1	1	0.25	0.01	0.09
Affine-RootSIFT	**219**	**6.9**	**47.6**	**2.23**	**0.28**	**0.24**
SURF	110	1	1	0.24	0.03	0.14
(Optimal) Affine-SURF	**663**	**14.82**	**219,6**	**3.68**	**1.19**	**0.73**
BRISK	29	1	1	1.57	0.00	0.04
Affine-BRISK	**49**	**8.42**	**70,89**	**17.57**	**0.06**	**0.08**
BRIEF	0	1	1	0.03	0.00	0.00
Affine-BRIEF	**7**	**14.82**	**219,6**	**2.06**	**0.09**	**0.03**
ORB	102	1	1	0.02	0.01	0.8
Affine-ORB	**90**	**14.82**	**219,6**	**2.12**	**0.31**	**0.40**
AKAZE	20	1	1	0.16	0.00	0.03
Affine-AKAZE	**51**	**8.42**	**70,89**	**2.31**	**0.06**	**0.09**
LATCH	54	1	1	0.07	0.01	0.04
Affine-LATCH	**101**	**14.82**	**219,6**	**1.72**	**0.12**	**0.10**
FREAK	124	1	1	0.14	0.01	0.10
Affine-FREAK	**182**	**7.06**	**49.8**	**2.54**	**0.11**	**0.31**

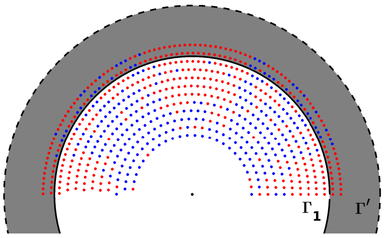
Table 4: Matching methods performance over query and target images from Figure 16. The proposed IMAS methods proposed here appear in bold. Computations were performed on an Intel(R) Core(TM) i5-4210U CPU 1.70GHz with 2 cores.
M - Matches.
ar - area ratio.



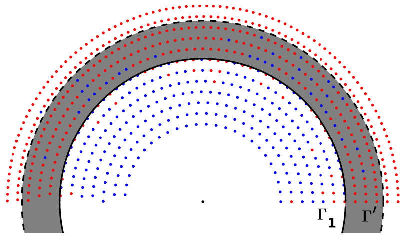
(a) Optimal Affine-SIFT ($r = 1.7$)
$\Gamma_1 = \left\{ [T_t R_\phi] \mid t \le 3.41 \right\}$
$\Gamma' = \left\{ [T_t R_\phi] \mid t \le 5.8 \right\}$



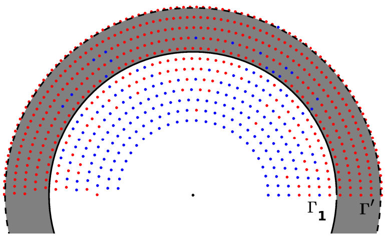
(b) ASIFT ($r = 1.8$)
$\Gamma_1 = \left\{ [T_t R_\phi] \mid t \le 3.05 \right\}$
$\Gamma' = \left\{ [T_t R_\phi] \mid t \le 5.5 \right\}$



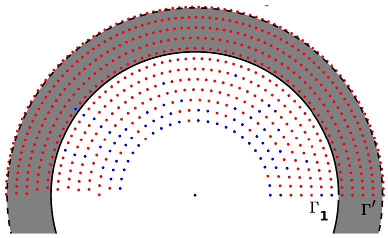
(c) MEDIUM configuration for DoG-SIFT ($r = 1.8$)
$\Gamma_1 = \left\{ [T_t R_\phi] \mid t \le 5 \right\}$
$\Gamma' = \left\{ [T_t R_\phi] \mid t \le 9 \right\}$



(d) Optimal Affine-SURF ($r = 1.4$)
$\Gamma_1 = \left\{ [T_t R_\phi] \mid t \le 3.57 \right\}$
$\Gamma' = \left\{ [T_t R_\phi] \mid t \le 5 \right\}$



(e) FAIR-SURF - simulated tilts ($r = 1.5$)
$\Gamma_1 = \left\{ [T_t R_\phi] \mid t \le 3.77 \right\}$
$\Gamma' = \left\{ [T_t R_\phi] \mid t \le 4\sqrt{2} \right\}$



(f) FAIR-SURF - fixed tilts ($r = 1.5$)
$\Gamma_1 = \left\{ [T_t R_\phi] \mid t \le 3.77 \right\}$
$\Gamma' = \left\{ [T_t R_\phi] \mid t \le 4\sqrt{2} \right\}$

Figure 13: Extreme test results.
Black dot - $[Id]$.
Coloured dots stand for $\left[A_{t,\phi}^{-1}\right]$ and belong to a fixed $\log 1.1$ uniform discretization of the annulus $\left\{ [T_t R_\phi] \mid 2 \le t \le 4\sqrt{2} \right\}$. The angle ϕ implicitly fixes $\left[B_\phi^{-1}\right] = [T_\Lambda R_\phi]$ where $\Lambda = \arg\max_t [T_t R_\phi] \in \Gamma'$.
Blue/Red dots - Success/Failure of ORSA Homography in identifying the underlying affine map.

- BPIFrance and Rgion 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.

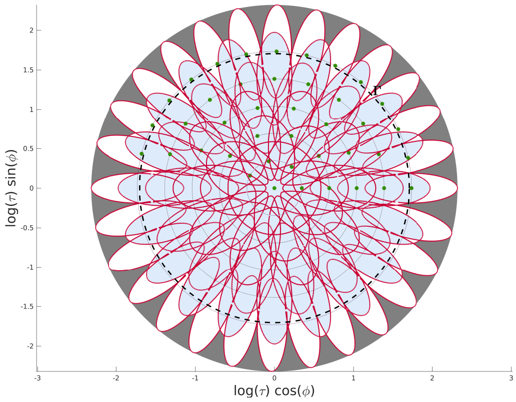
- The French Research Agency (ANR) under grant nro ANR-14-CE27-001 (MIRIAM)

6 Appendix

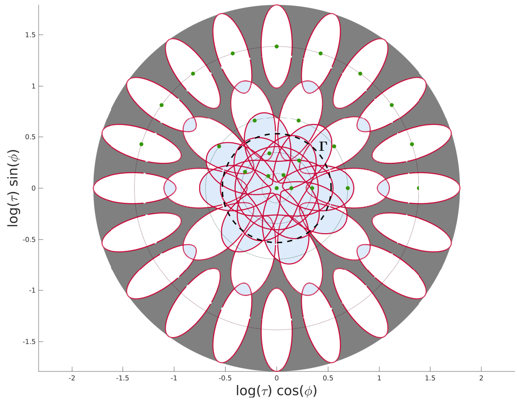
6.1 Proof of theorem 15

By proposition 13 we know that

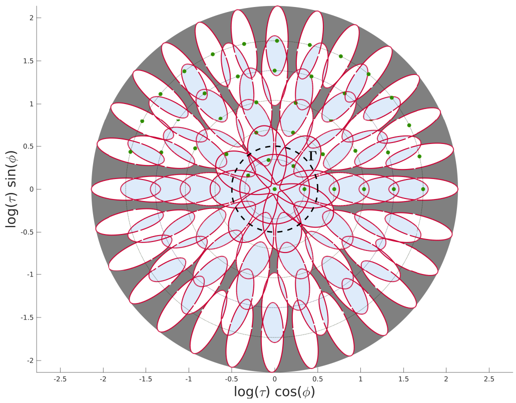
$$\tau\left(BA^{-1}\right) = \tau\left(i\left([B]\right) i\left([A]\right)^{-1}\right)$$



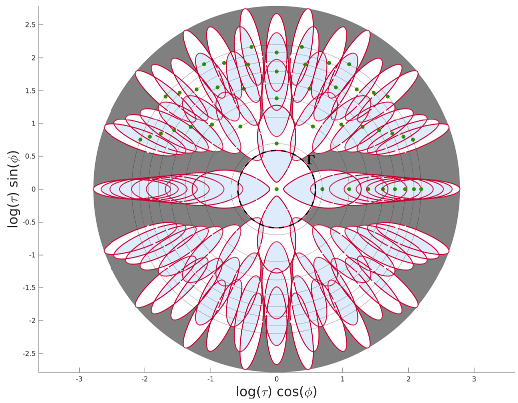
(a) Proposed covering for ASIFT in [44, 64]. This is a $\log 1.8$-covering of $\{[T_t R_\phi] \mid t \leq 5.5\}$ with 41 affine simulations representing an area ratio of 13.77.



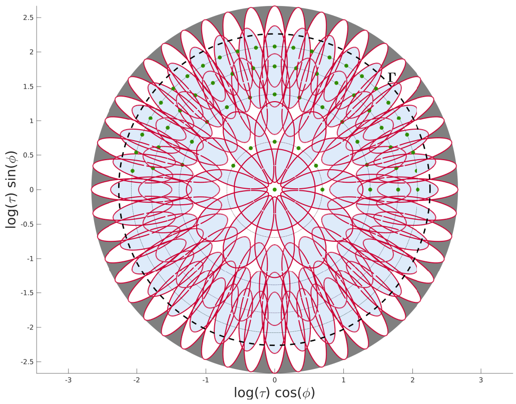
(b) Proposed covering for FAIR-SURF in [49], called fixed tilts. This is a $\log 1.5$-covering of $\{[T_t R_\phi] \mid t \leq 1.7\}$ with 23 affine simulations representing an area ratio of 11.42.



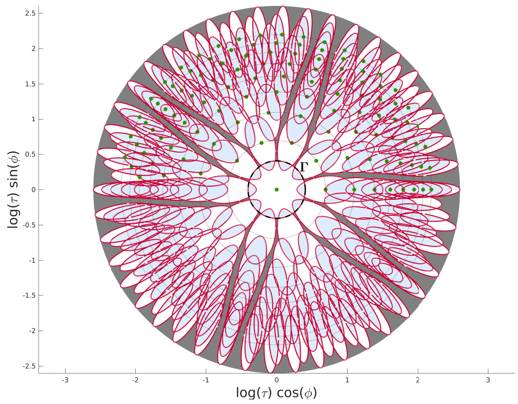
(c) Proposed covering for FAIR-SURF in [49], called simulated tilts. This is a $\log 1.5$-covering of $\{[T_t R_\phi] \mid t \leq 1.65\}$ with 41 affine simulations representing an area ratio of 13.77.



(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.

(a) 800×640 (b) 800×640

Figure 15: Graffiti. Both images generate a large number amount of keypoints for most methods.

(a) 600×450 (b) 600×450

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

$$\begin{aligned} C &= T_t R_2 Q_2^{-1} T_s^{-1} \\ &= T_t R(\phi) T_s^{-1} \end{aligned}$$

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

$$C^\star C = T_s^{-1} R(\phi)^{-1} T_t T_t R(\phi) T_s^{-1},$$

which are clearly

$$\det(C^\star C) = \frac{t^2}{s^2}$$

and

$$Tr(C^\star C) = \left(\frac{t^2}{s^2} + 1\right)\cos^2\phi + \left(\frac{1}{s^2} + t^2\right)\sin^2\phi.$$

6.1.2 The eigenvalues of $C^\star C$

Let $H = \begin{pmatrix} a & c \\ c & b \end{pmatrix} = C^\star C$ and λ_+, λ_- being the biggest and smallest eigenvalues of $C^\star C$ respectively. It is well known that

$$\begin{aligned} Tr(H) &= \lambda_+ + \lambda_- \\ \det(H) &= \lambda_+ \lambda_- \end{aligned}$$

and even more that both Tr and det also appear in the characteristic polynomial

$$\begin{aligned} |C^\star C - \lambda Id| &= \lambda^2 - \lambda(a+b) + \left(ab - c^2\right) \\ &= \lambda^2 - \lambda Tr\, H + \det H. \end{aligned}$$

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

$$\begin{aligned} \lambda_- &= \frac{Tr(H) - \sqrt{(Tr\, H)^2 - 4\det H}}{2}, \\ \lambda_+ &= \frac{Tr(H) + \sqrt{(Tr\, H)^2 - 4\det H}}{2}. \end{aligned}$$

Now, after some computations, the ratio between the biggest and smallest eigenvalues is

$$\begin{aligned} \frac{\lambda_+}{\lambda_-} &= \frac{\left(\frac{Tr\, H}{2} + \frac{\sqrt{(Tr\, H)^2 - 4\det H}}{2}\right)^2}{\det H} \\ &= \frac{s^2}{t^2}\left(\frac{g}{2} + \frac{\sqrt{g^2 - 4\frac{t^2}{s^2}}}{2}\right)^2 \end{aligned} \tag{12}$$

where g denotes the function

$$\begin{aligned} g(t, s, \phi) &:= Tr(C^\star C) \\ &= \left(\frac{t^2}{s^2} + 1\right)\cos^2\phi + \left(\frac{1}{s^2} + t^2\right)\sin^2\phi. \end{aligned}$$

6.1.3 Computing $\tau(C)$

Proposition 5 tells that the absolute tilt of C is

$$\begin{aligned}\tau(C) &= \sqrt{\frac{\lambda_+}{\lambda_-}} \\ &= \frac{s}{t}\left(\frac{g}{2}+\frac{\sqrt{g^2-4\frac{t^2}{s^2}}}{2}\right) \\ &= \frac{s}{t}\frac{g}{2}+\sqrt{\left(\frac{s}{t}\frac{g}{2}\right)^2-1} \\ &= G(s,t,\phi)+\sqrt{(G(s,t,\phi))^2-1}\end{aligned}$$

where

$$G(s,t,\phi) = \frac{s}{t}\frac{g(s,t,\phi)}{2}.$$

6.1.4 Disks in the space of tilts

Let $\boldsymbol{A} \coloneqq [T_t R_2] \in \Omega$ be fixed and let us find conditions on $\boldsymbol{B} \coloneqq [T_s Q_2] \in \Omega$ to satisfy

$$\boldsymbol{B} \in \mathcal{B}(\boldsymbol{A}, \log r)$$

which are clearly

$$\begin{gathered} d(\boldsymbol{A},\boldsymbol{B}) = \log\tau\left(i(\boldsymbol{A})\,i(\boldsymbol{B})^{-1}\right) \le \log r \\ \Updownarrow \\ \tau\left(i(\boldsymbol{A})\,i(\boldsymbol{B})^{-1}\right) \le r \end{gathered}$$

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

$$\begin{aligned} G(s,t,\phi)+\sqrt{(G(s,t,\phi))^2-1} &= \tau\left(AB^{-1}\right) \\ &\le r \end{aligned}$$

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

$$\begin{aligned} \sqrt{G^2-1} &\le r-G \\ &\Updownarrow \\ G^2-1 &\le r^2-2rG+G^2 \\ &\Updownarrow \\ G &\le \frac{r^2+1}{2r}. \end{aligned}$$

References

[1] Sameer Agarwal, Yasutaka Furukawa, Noah Snavely, Ian Simon, Brian Curless, Steven M Seitz, and Richard Szeliski. Building rome in a day. *Communications of the ACM*, 54(10):105–112, 2011.

[2] A Agarwala, M Agrawala, M Cohen, D Salesin, and R Szeliski. Photographing long scenes with multi-viewpoint panoramas. *International Conference on Computer Graphics and Interactive Techniques*, pages 853–861, 2006.

[3] Pablo Fernández Alcantarilla, Adrien Bartoli, and Andrew J. Davison. KAZE features. In *Lecture Notes in Computer Science*, volume 7577 LNCS, pages 214–227, 2012.

[4] Pablo Fernández Alcantarilla, Jesús Nuevo, and Adrien Bartoli. Fast Explicit Diffusion for Accelerated Features in Nonlinear Scale Spaces. *British Machine Vision Conference*, pages 13.1–13.11, 2013.

[5] Relja Arandjelovic and Andrew Zisserman. Three things everyone should know to improve object retrieval. In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 2911–2918, 2012.

[6] A Baumberg. Reliable feature matching across widely separated views. *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition*, 1:774–781, 2000.

[7] H Bay, T Tuytelaars, and L Van Gool. Surf: Speeded up robust features. *European Conference on Computer Vision*, 1:404–417, 2006.

[8] J. Blom. Topological and Geometrical Aspects of Image Structure. *University of Utrecht*, 1992.

[9] M Brown and D Lowe. Recognising panoramas. In *Proc. the 9th Int. Conf. Computer Vision, October*, pages 1218–1225, 2003.

[10] Matthew Brown and Sabine Süsstrunk. Multi-spectral SIFT for scene category recognition. In *Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on*, pages 177–184. IEEE, 2011.

[11] Michael Calonder, Vincent Lepetit, Christoph Strecha, and Pascal Fua. BRIEF: Binary robust independent elementary features. In *Lecture Notes in Computer Science*, volume 6314 LNCS, pages 778–792, 2010.

[12] F Cao, J.-L. Lisani, J.-M. Morel, P Musé, and F Sur. *A Theory of Shape Identification*. Springer Verlag, 2008.

[13] Davide Cozzolino, Giovanni Poggi, and Luisa Verdoliva. Efficient dense-field copy–move forgery detection. *IEEE Transactions on Information Forensics and Security*, 10(11):2284–2297, 2015.

[14] O Faugeras. *Three-Dimensional Computer Vision: A Geometric Viewpoint*. MIT Press, 1993.

[15] G Fritz, C Seifert, M Kumar, and L Paletta. Building detection from mobile imagery using informative SIFT descriptors. *Lecture Notes in Computer Science*, pages 629–638.

[16] Andreas Geiger, Julius Ziegler, and Christoph Stiller. Stereoscan: Dense 3d reconstruction in real-time. In *Intelligent Vehicles Symposium (IV), 2011 IEEE*, pages 963–968. Ieee, 2011.

[17] Yunchao Gong, Svetlana Lazebnik, Albert Gordo, and Florent Perronnin. Iterative quantization: A procrustean approach to learning binary codes for large-scale image retrieval. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 35(12):2916–2929, 2013.

[18] J S Hare and P H Lewis. Salient regions for query by image content. *Image and Video Retrieval: Third International Conference, CIVR*, pages 317–325, 2004.

[19] C Harris and M Stephens. A combined corner and edge detector. *Alvey Vision Conference*, 15:50, 1988.

[20] T. Iijima. Basic equation of figure and and observational transformation. *Systems, Computers, Controls*, 2(4):70–77, 1971.

[21] T Kadir, A Zisserman, and M Brady. An Affine Invariant Salient Region Detector. In *European Conference on Computer Vision*, pages 228–241, 2004.

[22] Maxim Karpushin. *Local features for RGBD image matching under viewpoint changes*. PhD thesis, 2016.

[23] Y Ke and R Sukthankar. PCA-SIFT: A more distinctive representation for local image descriptors. *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition*, 2:506–513, 2004.

[24] Simon Korman, Daniel Reichman, Gilad Tsur, and Shai Avidan. Fast-Match: Fast Affine Template Matching. In *Computer Vision and Pattern Recognition (CVPR), 2013 IEEE Conference on*, pages 1940–1947. IEEE, 2013.

[25] Stefan Leutenegger, Margarita Chli, and Roland Y. Siegwart. BRISK: Binary Robust invariant scalable keypoints. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 2548–2555, 2011.

[26] Gil Levi and Tal Hassner. LATCH: Learned arrangements of three patch codes. In *2016 IEEE Winter Conference on Applications of Computer Vision, WACV 2016*, 2016.

[27] T. Lindeberg. *Scale-Space Theory in Computer Vision.* Royal Institute of Technology, Stockholm, Sweden, 1993.

[28] T Lindeberg and J Garding. Shape-adapted smoothing in estimation of 3-D depth cues from affine distortions of local 2-D brightness structure. *Proc. ECCV*, pages 389–400, 1994.

[29] Tony Lindeberg. Direct estimation of affine image deformations using visual front-end operations with automatic scale selection. In *Computer Vision, 1995. Proceedings., Fifth International Conference on*, pages 134–141. IEEE, 1995.

[30] Tony Lindeberg. Generalized gaussian scale-space axiomatics comprising linear scale-space, affine scale-space and spatio-temporal scale-space. *Journal of Mathematical Imaging and Vision*, 40(1):36–81, 2011.

[31] Tony Lindeberg. Invariance of visual operations at the level of receptive fields. *BMC Neuroscience*, 14(1):P242, 2013.

[32] Ce Liu, Jenny Yuen, and Antonio Torralba. Sift flow: Dense correspondence across scenes and its applications. *IEEE transactions on pattern analysis and machine intelligence*, 33(5):978–994, 2011.

[33] D G Lowe. Distinctive image features from scale-invariant key points. *International Journal of Computer Vision*, 60(2):91–110, 2004.

[34] G Loy and J O Eklundh. Detecting symmetry and symmetric constellations of features. *Proceedings of ECCV*, 2:508–521, 2006.

[35] J Matas, O Chum, M Urban, and T Pajdla. Robust wide-baseline stereo from maximally stable extremal regions. *Image and Vision Computing*, 22(10):761–767, 2004.

[36] K Mikolajczyk and C Schmid. Indexing based on scale invariant interest points. *Proc. ICCV*, 1:525–531, 2001.

[37] K Mikolajczyk and C Schmid. An affine invariant interest point detector. *Proc. ECCV*, 1:128–142, 2002.

[38] K Mikolajczyk and C Schmid. Scale and Affine Invariant Interest Point Detectors. *International Journal of Computer Vision*, 60(1):63–86, 2004.

[39] K Mikolajczyk and C Schmid. A Performance Evaluation of Local Descriptors. *IEEE Trans. PAMI*, pages 1615–1630, 2005.

[40] Dmytro Mishkin, Jiri Matas, and Michal Perdoch. MODS: Fast and robust method for two-view matching. *Computer Vision and Image Understanding*, 141:81–93, 2015.

[41] Lionel Moisan, Pierre Moulon, and Pascal Monasse. Automatic Homographic Registration of a Pair of Images, with A Contrario Elimination of Outliers. *Image Processing On Line*, 2:56–73, 2012.

[42] P Moreels and P Perona. Evaluation of Features Detectors and Descriptors based on 3D Objects. *International Journal of Computer Vision*, 73(3):263–284, 2007.

[43] J M Morel and G Yu. On the consistency of the SIFT Method. Technical Report Prepublication, to appear in Inverse Problems and Imaging (IPI), CMLA, ENS Cachan, 2008.

[44] Jean-Michel Morel and Guoshen Yu. ASIFT: A new framework for fully affine invariant image comparison. *SIAM Journal on Imaging Sciences*, 2(2):438–469, 2009.

[45] A Murarka, J Modayil, and B Kuipers. Building Local Safety Maps for a Wheelchair Robot using Vision and Lasers. In *Proceedings of the The 3rd Canadian Conference on Computer and Robot Vision*. IEEE Computer Society Washington, DC, USA, 2006.

[46] P Musé, F Sur, F Cao, and Y Gousseau. Unsupervised thresholds for shape matching. *Proc. of the International Conference on Image Processing*, 2:647–650.

[47] P Musé, F Sur, F Cao, Y Gousseau, and J M Morel. An A Contrario Decision Method for Shape Element Recognition. *International Journal of Computer Vision*, 69(3):295–315, 2006.

[48] A Negre, H Tran, N Gourier, D Hall, A Lux, and J L Crowley. Comparative study of People Detection in Surveillance Scenes. *Structural, Syntactic and Statistical Pattern Recognition, Proceedings Lecture Notes in Computer Science*, 4109:100–108, 2006.

[49] Yanwei Pang, Wei Li, Yuan Yuan, and Jing Pan. Fully affine invariant SURF for image matching. *Neurocomputing*, 85:6–10, 2012.

[50] D Pritchard and W Heidrich. Cloth Motion Capture. *Computer Graphics Forum*, 22(3):263–271, 2003.

[51] Paul Scovanner, Saad Ali, and Mubarak Shah. A 3-dimensional SIFT descriptor and its application to action recognition. In *MULTIMEDIA '07: Proceedings of the 15th international conference on Multimedia*, pages 357–360, New York, NY, USA, 2007. ACM.

[52] S Se, D Lowe, and J Little. Vision-based mobile robot localization and mapping using scale-invariant features. *Robotics and Automation, 2001. Proceedings 2001 ICRA. IEEE International Conference on*, 2, 2001.

[53] Josef Sivic, Andrew Zisserman, et al. Video google: A text retrieval approach to object matching in videos. In *iccv*, volume 2, pages 1470–1477, 2003.

[54] N Snavely, S M Seitz, and R Szeliski. Photo tourism: exploring photo collections in 3D. *ACM Transactions on Graphics (TOG)*, 25(3):835–846, 2006.

[55] Cees Snoek, Kvd Sande, OD Rooij, Bouke Huurnink, J Uijlings, M van Liempt, M Bugalhoy, I Trancosoy, F Yan, M Tahir, et al. The MediaMill TRECVID 2009 semantic video search engine. In *TRECVID workshop*, 2009.

[56] T Tuytelaars and L Van Gool. Wide baseline stereo matching based on local, affinely invariant regions. *British Machine Vision Conference*, pages 412–425, 2000.

[57] T Tuytelaars and L Van Gool. Matching Widely Separated Views Based on Affine Invariant Regions. *International Journal of Computer Vision*, 59(1):61–85, 2004.

[58] T Tuytelaars, L Van Gool, and Others. Content-based image retrieval based on local affinely invariant regions. *Int. Conf. on Visual Information Systems*, pages 493–500, 1999.

[59] Christoffer Valgren and Achim J Lilienthal. SIFT, SURF & seasons: Appearance-based long-term localization in outdoor environments. *Robotics and Autonomous Systems*, 58(2):149–156, 2010.

[60] Koen Van De Sande, Theo Gevers, and Cees Snoek. Evaluating color descriptors for object and scene recognition. *IEEE transactions on pattern analysis and machine intelligence*, 32(9):1582–1596, 2010.

[61] M Vergauwen and L Van Gool. Web-based 3D Reconstruction Service. *Machine Vision and Applications*, 17(6):411–426, 2005.

[62] Philippe Weinzaepfel, Jerome Revaud, Zaid Harchaoui, and Cordelia Schmid. Deepflow: Large displacement optical flow with deep matching. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 1385–1392, 2013.

[63] G Yang, C V Stewart, M Sofka, and C L Tsai. Alignment of challenging image pairs: Refinement and region growing starting from a single keypoint correspondence. *IEEE Trans. Pattern Anal. Machine Intell.*, 2007.

[64] Guoshen Yu and Jean-Michel Morel. ASIFT: An Algorithm for Fully Affine Invariant Comparison. *Image Processing On Line*, 1:1–28, 2011.

[65] Huiyu Zhou, Yuan Yuan, and Chunmei Shi. Object tracking using SIFT features and mean shift. *Computer vision and image understanding*, 113(3):345–352, 2009.