On Similarity-Invariant Fairness Measures

Alexander Brook¹, Alfred M. Bruckstein², and Ron Kimmel²

¹ Dept. of Mathematics, Technion, Israel
abrook@technion.ac.il
² Dept. of Computer Science, Technion, Israel

Abstract. After introducing the basic principles behind the similarity-invariant smoothness measures for curves and surfaces, with references to the relevant literature, we discuss the ramifications of scale-invariance in various problems of image processing and analysis, and point out some unanswered questions.

1 Introduction

In this paper we consider variational measures of fairness which are invariant under similarity transformations and reparametrizations, that is integrals of the form

\[ J(\gamma) = \int_{\gamma} F(p, C, C', C'', \ldots) \, dp \]

(and analogous integrals for surfaces) invariant under

- change of parametrization \( p = p(\bar{p}) \), where \( p \) is a smooth monotonically increasing function,
- translation \( C = \bar{C} + v \), \( v \) being a constant vector,
- rotation \( C = U\bar{C} \), for any orthonormal matrix \( U \),
- scaling \( C = \alpha \bar{C} \), for any nonzero (usually positive) scalar \( \alpha \).

The combination of the first three properties is usually called Euclidean invariance. Most image processing theories and algorithms are Euclidean invariant, so we will focus on scale invariance.

In this paper we discuss the different (but often confused) meanings of invariance and review the previous work on fairness measures for curves and surfaces, trying to produce an exhaustive list of functionals that are such measures. We then consider the need for (and the possibility of) scale invariance in image processing and image analysis (in particular edge detection and integration).

2 Invariance of a Measure Versus Invariance of a Solution

An important distinction, which is often overlooked in works dealing with invariance of variational solutions, is the difference between the invariance of the
solution to a minimization problem and the invariance of the functional being minimized.

There are two possible uses for a fairness measure. One is the general question “Is this a nice curve?” or “Which of these two curves looks better?” (see [1–p. 30] for a discussion of definitions of fairness). Another is the specific question “What is the best curve satisfying the given conditions?” Most frequently, this is used in contour completion. The basic setting is of two short segments in the plane that are to be connected in the smoothest way possible. Such segments are usually treated as 1-jets, that is, the only information extracted from them is the position and the tangent.

The invariance of the functional is important if we use the functional as a universal measure of smoothness; we wish to give each curve a number that will be a measure of its smoothness, and this number should be scale-invariant.

On the other hand, if we are only interested in solving a specific problem of finding the most fair curve that connects two given segments, for example, we are only dealing with the minimizers of the functional. It is possible that the minimizers will be scale-invariant while the functional is not.

This phenomenon is remarked upon in [1–p. 72]:

Minimal Variation Curves are invariant under rigid body transformation and uniform scaling. The \textit{value} of the Minimal Variation Curve functional, however, changes with a change of scale.

Let us consider a very simple example. Say, we want to connect two points by the shortest possible curve. The corresponding functional is just the length of the curve, $\int_{\gamma} ds$, and the solution is, of course, the straight segment connecting the two points. Both the functional and the solution are invariant under rigid motions.

However, the solution is also scale invariant. Indeed, if we dilate the plane, the straight segment will remain the shortest possible line even after dilation. And yet the functional that defines this curve is not scale invariant, of course!

How does this happen? If we dilate the space by $\lambda$, the length functional is multiplied by $\lambda$: $\int_{\gamma} ds \mapsto \lambda \int_{\gamma} ds$, and so is the Euler-Lagrange equation:

$$C'' = 0 \mapsto \lambda C'' = 0,$$

but these two equations are equivalent. Thus, a functional which is not invariant under certain group (dilations) possesses an Euler-Lagrange equation which is invariant under this group.

This non-equivalence is mentioned in [2–p. 255] and [3–p. 236]. In both cases, after the theorem that says “every symmetry of the functional is a symmetry of the Euler-Lagrange equation”, it is mentioned that the converse is not true and that “the most common examples” are those of scaling transformations. Also, the above theorem (if the functional is invariant then the Euler-Lagrange
equation is invariant) is not a trivial one, due to the fact that generally, not all of the Euler-Lagrange equation’s solutions are minimizers of the functional.

We do not know whether these are the only examples, or whether it is possible to give a complete classification of such counterexamples. We hope to answer these questions elsewhere.

3 Smoothness Measures for Curves

3.1 Basic Invariants

Classical Similarity Invariants. First of all, if we drop the requirement of invariance to scale, we are left with parametrization and Euclidean invariance. It is well known and easily established (see [4–p. 81] and [5–8.1.1]) that Euclidean and reparametrization invariant functionals must have the form

\[ \int F(\kappa, \kappa_s, \kappa_{ss}, \ldots) \, ds, \]

where \( s \) is the arclength parameter, and \( \kappa \) is the curvature. (For a curve in space we will have to add torsion to the list, but space curves are rarely, if ever, used in image processing.)

From these functionals we want to select those that are invariant to scaling. If we replace \( C \) with \( \alpha C \), \( \kappa^{(i)} \) becomes \( \kappa^{(i)}/\alpha^{i+1} \). Substituting these to \( F \), we see that we are looking for functions \( F \) that are homogeneous (of order \(-1\), so that \( F \, ds \) is invariant) with respect to \( \alpha \) (the same idea is used in [6]). Let us see what this leaves us. One possibility is just to write down the list of these functions:

\[ \kappa, \frac{\kappa_s}{\kappa}, \frac{\kappa_s^2}{\kappa^3}, \frac{\kappa_{ss}}{\kappa^2}, \frac{\kappa_{ss}}{\kappa_s} = \frac{d}{ds} (\ln |\kappa_s|) \]

and so on.

We can go on with this list, by considering higher order derivatives and higher powers, none of which is a good idea. High order derivatives are numerically challenging; PDEs of order higher than four are nearly impossible to implement. High powers are unreasonable in a smoothness measure because of their high sensitivity to noise (also known as sensitivity to outliers or non-robustness).

We can bring some order to the list (1) with the help of invariants theory. The invariant arclength and curvature for the similarity group are given in [3] as

\[ \frac{u_{xx}}{1 + u_x^2} \, dx \quad \text{and} \quad \frac{(1 + u_x^2)u_{xxx} - 3u_xu_{xx}^2}{u_{xx}^2}, \]

which upon inspection turn out to be \( d\theta = \kappa \, ds \) (notice that the turn angle \( \theta \) itself is not even Euclidean invariant!) and \( \mu = \kappa_s/\kappa^2 \). Appropriately, both quantities are homogeneous of degree 0 in the sense described above. Thus, the invariant measures are \( \int \Phi(\mu, \frac{d\mu}{d\theta}, \ldots) \, d\theta \), which generates the same list as above; in particular,

\[ d\mu = \left( \frac{\kappa_{ss}}{\kappa^2} - \frac{\kappa_s^2}{\kappa^3} \right) d\theta. \]
**Winding Number and Total Absolute Curvature.** The integral $\frac{1}{2\pi} \int_\gamma \kappa \, ds$ is called the winding number, for it equals $\frac{1}{2\pi} \int_\gamma d\theta$, which is the number of the turns the (closed) curve makes. The integral $\int \kappa \, ds$ on an open piece of a curve is the turn angle. This in particular means that $\kappa$ is a null Lagrangian, that is, an attempt to write down the Euler-Lagrange equation for the winding number will result in $0 = 0$. In any case, the winding number is obviously unsuitable as a smoothness measure.

We can consider another similarity invariant, the total absolute curvature $\int_{\gamma} |\kappa| \, ds$. The excess of this functional above $2\pi$ shows how much the curve wiggles before closing upon itself ([7] calls it “angular total variation”). This quantity was a subject of some research; see the review [8]. The relevant result is that $\int_{\gamma} |\kappa| \, ds \geq 2\pi$ for any closed curve, and the equality holds if and only if $\gamma$ is a convex curve. Thus, in principle, it is a fairness measure, and the corresponding flow should bring any closed curve to a convex one, which is a desirable behavior for a short time. However, $|\kappa|$ is also a null Lagrangian, except at the inflection points.

The article [9] gives an algorithm for minimizing $\int_{\gamma} (a + b|\kappa|) \, ds$ by a direct construction of the minimum, which is a polygonal path. This approach will not work for $a = 0$.

**Other Invariants.** First of all, there are two functionals that are integrals of full differentials: the similarity arclength $\int d\theta$ (the turn angle) and $\int \mu \, d\theta = \int \frac{\kappa}{\kappa} \, ds = \int d\ln|\kappa|$. Since they result in 0 for any closed curve, they are unsuitable for our needs.

What is left is $\int \mu^2 \, d\theta = \int \frac{\kappa^2}{\kappa} \, ds$. It is unclear whether it has any meaning in familiar terms.

### 3.2 The Idea of Weiss and Others

So far, the situation is far from satisfying. What we want is something that is invariant, numerically feasible and has a clear meaning. Being unable to achieve this goal within the framework of standard variational problems, researchers turned to other possibilities, namely, to non-local functionals. The following idea was first published in [10], then in [11] and [12].

If we have a functional $J$ that is Euclidean invariant, parametrization invariant, and homogeneous with respect to scaling ($\alpha$ from Sect. 3.1), then $(\int_{\gamma} ds)^p J$ is similarity invariant for some value of $p$. The most popular example is the scale-invariant elastica (called Minimal Energy Curve in [12]) $\int_{\gamma} ds \int_{\gamma} \kappa^2 ds$.

### 3.3 Scale-Invariant Minimal Energy Curve (Elastica)

The functional $\int_{\gamma} \kappa^2 \, ds$ gives the bending energy of an elastic rod. The usual boundary conditions are derivatives at the endpoints, which is the common setting of the contour interpolation problem. Sometimes the constant length constraint $\int_{\gamma} ds = L$ is added.
The paper [11] introduces length-constrained elastica and scale-invariant elastica, and provides the Euler-Lagrange equations in terms of the turn angle. Another measure suggested there is the weighted functional \( L \int_0^L W(s) \kappa^2(s) \, ds \); the article does not mention the now obvious possibility to select an image dependent weight \( W \) and use this for segmentation.

The article [12] introduces all four functionals (Minimal Energy Curve and Minimal Variation Curve and their scale-invariant counterparts) in a quest for a fair and stable curve interpolation. The main rationale for the scale-invariant measures is that they are more stable. The minimization is by quintic Hermite splines (Hermite splines readily accommodate the constraints provided by the boundary segments). A different discretization approach to these functionals is offered in [13].

In [14] an area constraint is added to the usual length constraint on elastica:

\[
\alpha \int \kappa^2 \, dl + \mu \left( \int dl - L \right) + \sigma \left( \int \int_{\text{interior}} dx - A \right)
\]

(notice that the constraints are not squared). The authors are mainly interested in the dependence of the minimizers’ behavior on the Lagrange multipliers \( \mu \) and \( \sigma \). Their basic approach is through the phase plane of the Euler-Lagrange equations for \( \kappa \), that is the plane \( \kappa, \kappa' \); the basic ideas were present in [15], but here they are taken much further.

Non-closed elastica with fixed endpoints or fixed endpoints and end directions is the subject of [16]. Existence is shown and explicit formulas are given.

There is a standing conjecture of De Giorgi (mentioned in [17]) that \( \int \kappa^2 \, dl \) can be approximated in the sense of \( \Gamma \)-convergence by

\[
\frac{1}{\varepsilon} \int_{\Omega} \left( 2\varepsilon \Delta z - \frac{W'(z)}{4\varepsilon} \right)^2 \, dx \, dy,
\]

where \( W(z) = (1 - z^2)^2 \), and the curve is the zero level set of \( z(x, y) \). Some results on this can be found in [18].

### 3.4 Scale-Invariant Minimal Variation Curve

The functional \( \left( \int_{\gamma} ds \right)^3 \int_{\gamma} \kappa_s^2 \, ds \) is proposed in [12] as a scale-invariant version of the minimal variation curve \( \int_{\gamma} \kappa_s^2 \, ds \) and later studied numerically in [13]. The generated curves are usually more smooth than elasticae, of course, and the scale-invariant version is more stable than the usual Minimal Variation Curve [12].

A recent article [19] deals with the problem of curve completion. The authors propose a scale-invariant completion model which they think is scale-invariant Minimal Variation Curve; unfortunately, they are wrong. Their solution, namely Euler spiral, satisfies \( \kappa_{ss} = 0 \); the authors erroneously claim that this is the Euler-Lagrange equation of the Minimal Variation Curve functional \( \int_{\gamma} \kappa_s^2 \, ds \).

The correct equation \( \kappa'' + \kappa^2 \kappa - \frac{1}{2} \kappa \kappa' = 0 \) is given in [20], and is also contained in [21], derived by the methods from [15]. Note that since the equation \( \kappa_{ss} = 0 \)
is homogeneous it is indeed, obviously, scale-invariant. This is similar to the situation we have described in Sect. 2.

In [22] the energy \( \int (\alpha \kappa_2^2 + \beta) \, ds \) arises as the mode of a certain type of random walk, that attempts to model curve completion.

### 3.5 Other Possibilities

One possibility to go in a different (though not necessarily more promising) direction is to use the area inside the curve instead of the length, for example \( [\text{area}(\gamma)]^{1/2} \int \kappa^2 \, ds \). The most obvious invariant using the area is the isoperimetric ratio \( \frac{\text{length}^2}{\text{area}} \). Of course, it is only suited for closed curves [23].

Also, if we want a solution to a specific problem, we might as well consider flows (PDEs) that do not come from an Euler-Lagrange equation of a functional. For some examples see [19, 24, 25].

### 4 Smoothness Measures for Surfaces

#### 4.1 Integrals of Curvatures

Just as with curves, we are led to the study of integrals of powers of curvatures, namely of the Gaussian curvature \( K \) and the mean curvature \( H \).

\( K \) is a null Lagrangian; the Gauss-Bonnet theorem states that on a closed surface \( \int_S K = 2\pi \chi(S) \), where \( \chi \) the Euler characteristic of the surface.

The integrals \( \int_S K^2 \) and \( \int_S (K^2 + \lambda) \) are treated in [25]. The integral of the mean curvature vector is zero on a closed surface: \( \int_S H\hat{n} = \mathbf{0} \) [26]. The integral of \( H \) itself is widely studied in integral geometry (e.g. [27]) and gives, in a sense, the mean width of the body enclosed by a surface. It is unlikely that it will make a good smoothness measure.

The most widely studied is the Willmore functional \( \int_S H^2 \), which is similarity invariant. The book [28] gives some basic results on this functional, including the Euler-Lagrange equation, and provides an introduction to the literature. In particular, there are works on numerical minimization of the Willmore functional.

#### 4.2 Other Invariant Measures

Besides Willmore functional \( \int_S H^2 \) there are other similarity invariant smoothness measures for surfaces.

The paper [12] proposes the following “minimal variation surface” scale-invariant functional:

\[
\int dS \int \left[ \left( \frac{\partial \kappa_1}{\partial \hat{e}_1} \right)^2 + \left( \frac{\partial \kappa_2}{\partial \hat{e}_2} \right)^2 \right] \, dS.
\]

Here \( \frac{\partial \kappa_1}{\partial \hat{e}_1} \) is a directional derivative of a principal curvature in the corresponding principal direction (see [1] for details).

The isoperimetric ratio can be generalized to closed surfaces as \( \frac{\text{area}^3}{\text{volume}^2} \) [23].
5 Smoothness Measures for Images and Scale-Invariant Image Processing

5.1 Scale-Invariant Smoothness Measures for Images

If we accept the “morphological” point of view, and regard an image as a collection of isophotes, we can generate a smoothness measure on images from any smoothness measure on curves, by integrating over the intensity range: \( \int_R f(I^{-1}(u)) \, du \), where \( I : \Omega \rightarrow \mathbb{R} \) is the image. Using the coarea formula we obtain for \( f(\gamma) = \int_\gamma ds \) and for \( f(\gamma) = \int_\gamma |\kappa|^p \, ds \)

\[
\int_R \int_{I^{-1}(u)} du = \int_\Omega |\nabla u| \, dx \, dy,
\]

\[
\int_R \int_{I^{-1}(u)} |\kappa|^p \, du = \int_\Omega |\kappa|^p |\nabla u| \, dx \, dy = \int_\Omega \left| \nabla \cdot \frac{\nabla u}{|\nabla u|} \right|^p |\nabla u| \, dx \, dy.
\]

These formulas appear in [9] and [29] as a justification of a functional

\[
\int_\Omega (a + b|\kappa|^p)|\nabla u| \, dx \, dy,
\]

used to solve the problem of disocclusion. To obtain a scale-invariant functional, we have to multiply the functionals instead of adding them, arriving at

\[
\int_\Omega |\nabla u| \, dx \, dy \cdot \int_\Omega \kappa^2 |\nabla u| \, dx \, dy
\]

(2)

and

\[
\left( \int_\Omega |\nabla u| \, dx \, dy \right)^3 \cdot \int_\Omega \kappa_s^2 |\nabla u| \, dx \, dy
\]

(note that \( \kappa_s = \nabla \left( \nabla \cdot \frac{\nabla u}{|\nabla u|} \right) \cdot \frac{(u_y, -u_x)}{|\nabla u|} \)).

5.2 Do we Want Image Processing to be Scale Invariant?

It seems difficult to really justify the need for scale invariance in image processing. Our main objection to the idea from [11] that

a small circle should be considered as smooth as a large one

is that in an image, the ultimately small one-pixel circle is most probably due to noise, and most image restoration techniques rely on this distinction in scale between noise and signal. Thus, scale-invariant image processing will be very noise-sensitive, if not unstable.

We did not check this objection numerically. It might ultimately turn out that the regularizing effect of a discretization will overcome these problems somehow.
Also, it is possible to achieve partial scale invariance by building an image processing algorithm on a pyramid, that is, on a range of scales; in this case, we can decide explicitly on the minimal scale at which we “stop the invariance”. But complete scale invariance means inability to detect noise.

The only kind of image processing that probably needs scale invariance is inpainting (disocclusion) [30, 31]. As noted (in a somewhat different context) in [31):

In application, an average image often contains objects of a large dynamic range of scales. Hence in most inpainting problems, it is commonly found that “slim” objects are broken by the inpainting domains even though the domains themselves are small (to human observers). A good inpainting scheme should encourage the connection of these broken slim objects.

Still, so far there are no inpainting algorithms that are completely invariant.

5.3 What is Similarity-Invariant Image Processing?

What condition should an image processing algorithm satisfy to be called “scale-invariant”? There are at least two answers to this question, both unsatisfactory in our opinion.

The first possibility is to demand that the processing results on a scaled version of an image will be the same as the results on the image itself, up to scaling. This is the usual commutativity demand for invariance. The only problem with this criterion is that it is too easily satisfied—we just scale all images to a fixed size before processing. Usually this is done even without actually scaling the image, which will introduce errors, but by selecting the image size as the unit of measurement [32, 33].

Another variant on basically the same principle is to ask that an object will be processed the same, regardless of its size in a given picture. This, again, is customarily done by selecting some characteristic dimension of the object as the unit length for the algorithm [34, 35].

On the other hand, we can ask for an algorithm that will act invariantly on each object in the image, regardless of the scale, this without trying to actually measure the object as part of the algorithm. This is what we see in most Euclidean invariant algorithms, so it seems to be a reasonable requirement. For example, if two identical chairs are seen in an image from different distances, we would like to have in the processed image two corresponding parts, differing only by scale, and that without any attempt on the part of the algorithm to identify the objects and their dimensions.

If we want to have a variational model to have this property, the functional must be local. In particular, we can not allow products of non-trivial integrals, like in (2). We are not aware of any image processing or image analysis algorithm, which is scale-invariant in this sense.
Scale Invariance in Edge Detection and Contour Completion

If we turn from image processing to image analysis, the situation changes. Edge detection, contour completion and object recognition algorithms will probably benefit from scale invariance, as noted e.g. in [19] with regard to curve completion:

Scale invariance is necessary since these curves are used to model, say gap completion, in a world where the distance from the observer to the imaged object varies constantly, yet the gaps must be completed consistently.

We should also mention the work [36], pointing out the need for scale invariance in salient curve detection, and the lack of such invariance in the well-known model of [37]. It seems that currently there is no scale invariant contour completion or saliency detection algorithm.

Still, we must be careful not to introduce scale invariance at all costs. For example, [13] motivates introduction of the scale-invariant elastica by mentioning

...a counterintuitive result of Horn asserting that the least energy curve (minimizing $\int \kappa^2 ds$) which starts vertically up at $(0,0)$ and arrives vertically down at $(1,0)$ is not a semicircle.

The hidden claim here is that if we have two co-circular segments (small pieces of a curve), then a good curve interpolation algorithm must connect them by a circle. However, we should distinguish between purely “geometric” co-circularity and “perceptional” co-circularity. Two parallel segments, while being indeed geometrically co-circular, are not perceived as such. When modeling the perceptual co-circularity we must take into account the relative direction of the segments, the distance between them, their “strength”, and possibly other parameters. Some ideas on modeling co-circularity perception appear in the work on tensor voting [38] and other works on contour saliency. Some examples that use the elastic energy are [39, 22].

Usually, edge detection that is done by zero-crossings is scale-invariant. In particular, the Laplacian zero crossing (Marr-Hildreth) and the zeros of the second directional derivative in the gradient’s direction (Haralik-Canny) are similarity-invariant.

As an illustrative example, we may consider the edge integration functional (snake) proposed in [40]. It is a variant of the Haralik-Canny scheme, being the difference of two double integrals, both similarity-invariant: the integral of the Laplacian $\int \int \Delta I \, dx \, dy$ (which by itself will detect the Laplacian zero crossings) and the integral of the second directional derivative in the direction normal to the gradient $\int \int I_{\eta} \, dx \, dy$.

Figure 1 shows the results of this edge integration algorithm and nicely illustrates several of our points.

Since the functional is scale-invariant, there is no incentive for the edges to be short. This means that the detected edge will follow the zero-crossing as close as
possible within the limits of the discretization, which in most cases in much more wiggling than desirable. What prevents endless squirming in this specific case is probably the implicit regularization provided by the level set implementation. Still, it can be seen that the edge is very rugged.

In the same figure we can also see the problem predicted in Sect. 5.2: the detection of very small “objects”, that are obviously a result of a few noisy pixels (or even one pixel). The solution proposed in [40] is an additional energy term for regularization; unsurprisingly, this term is not scale invariant.

Another scale-invariant active contour formulation was proposed in [41], with better stability. However, it seems that this method demands a very close initialization. Also, since the paper does not go into any implementation details, it is difficult to say whether the stability is inherent to the method, or (more likely) to the numerical implementation.

7 Conclusions

We think that the ramifications of the scale invariance demand are not as clear as sometimes thought. In particular,

- it is possible for a non-invariant minimization problem to provide invariant solutions;
- scale invariance is probably undesirable in image processing, since it will make noise suppression impossible;
- in edge detection, scale invariance is more reasonable but still has to be very carefully handled.

References