Image Processing Done Right

Jan J. Koenderink[1] and Andrea J. van Doorn[1]

Universiteit Utrecht, Buys Ballot Laboratory,
Princetonplein 5, Utrecht NL–3584CC, The Netherlands,
j.j.koenderink@phys.uu.nl,
http://www.phys.uu.nl/~wwwpm/HumPerc/koenderink.html

Abstract. A large part of "image processing" involves the computation of significant points, curves and areas ("features"). These can be defined as loci where absolute differential invariants of the image assume fiducial values, taking spatial scale and intensity (in a generic sense) scale into account. "Differential invariance" implies a group of "similarities" or "congruences". These "motions" define the geometrical structure of image space. Classical Euclidian invariants don't apply to images because image space is non–Euclidian. We analyze image structure from first principles and construct the fundamental group of image space motions. Image space is a Cayley–Klein geometry with one isotropic dimension. The analysis leads to a principled definition of "features" and the operators that define them.

Keywords. Image features, texture, image indexing, scale–space, image transformations, image space

1 Introduction

"Images" are often considered to be distributions of some "intensity" (a density of "stuff") over some spatial extent (the "picture plane" say). One thinks of a graph ("image surface") in the three dimensional product space defined by the picture plane and an intensity axis. Numerous conventional methods (often implicitly) involve the computation of differential invariants of this surface. One conventionally draws upon the large body of knowledge on the differential geometry of surfaces in $\mathbb{E}^3$.

There are grave problems with this approach though:
— the physical dimensions of the image plane are incommensurable with the dimension of the image intensity domain. The former are distances or angles, the latter radiances or irradiances;
— classical differential geometry deals with invariance under the Euclidian group of congruences. But to turn a region of "image space" about an axis parallel to the picture plane is impossible;
— the intensity domain is unlike the Euclidian line in that it is only a half–line and translations are undefined.
Thus the aforementioned methods lack a principled basis.

A. Heyden et al. (Eds.): ECCV 2002, LNCS 2350, pp. 158–172, 2002.

To accept that these methods are *ad hoc* doesn't necessarily mean that one should abandon them altogether. Rather, one should look for the proper group of congruences and adjust the differential geometry to be the study of invariants of that group[11]. One should investigate the proper geometry of the intensity domain. This is the quest undertaken in this paper: First we investigate the structure of the intensity domain, then we identify the proper group of congruences and similarities. We then proceed to develop the differential geometry of surfaces under the group actions in order to arrive at a principled discipline of "image processing".

2 Geometrical Structure of the "Intensity" Domain

"Intensity" is a generic name for a flux, the amount of stuff collected within a certain aperture centered at a certain location. In the case of a CCD array the aperture is set by the sensitive area and the flux is proportional with the number of absorbed photons collected in a given time window. One treats this as the continuous distribution of some "density", in the case of the CCD chip the number of absorbed photons per unit area per unit time, that is the irradiance. This goes beyond the observable and is often inadvisable because the "stuff" may be granular at the microscale. In the case of the CCD chip the grain is set by the photon shot noise. It is also inadvisable because natural images fail to be "nice" functions of time and place when one doesn't "tame" them via a finite collecting aperture or "inner scale". Only such tamed images are observable[4], this means that *any* image should come with an inner scale. This often is nilly willy the case. For instance, in the case of the CCD chip the inner scale is set by the size of its photosensitive elements. But no one stops you from changing the inner scale artificially. When possible this is a boon, because it rids one of the artificial pixelation. "Pixel fucking" is in a different ballpark from image processing proper, though it sometimes is a necessary evil due to real world constraints. Because of a number of technical reasons the preferred way to set the inner scale is to use Gaussian smoothing[4]. Here we assume that the "intensity" $z(x, y)$ is a smooth function of the Cartesian coordinates $\{x, y\}$ of the picture plane with a well defined inner scale. We assume that the intensity is positive definite throughout.

The photosensitive elements of the CCD chip can be used to illustrate another problem. Suppose we irradiate the chip with a constant, uniform beam. We count photons in a fixed time window. The photon counts are known to be Poisson distributed with parameter λ (say). Suppose a single measurement ("pixel value") yields n photons. What is the best estimate of the "intensity" λ on the basis of this sample? Let two observers A and B measure time in different units, *e.g.*, let $t_a = \mu t_B$, then their intensities must be related as $\lambda_A\, dt_A = \lambda_B\, dt_B$. Let A and B assign priors $f_A(\lambda_A)\, d\lambda_A$ and $f_B(\lambda_B)\, d\lambda_B$. We have to require that these be mutually consistent, that is to say, we require $f_A(\lambda_A)\, d\lambda_A = f_B(\lambda_B)\, d\lambda_B$. Both A and B are completely ignorant, and since their states of knowledge are equal, one arrives at the functional equation $f(\lambda) = \mu f(\mu\lambda)$, from which one concludes[9] that the prior that expresses complete ignorance and doesn't depend on the unit of time is $\lambda^{-1}\,d\lambda$, that is to say a *uniform prior on the log–intensity scale*.

This means that we only obtain some degree of symmetry between the Cartesian dimensions $\{x, y\}$ (where no particular place is singled out, and no particular scale has been established) and the intensity dimension if we use $Z(x, y) = \log(z(x, y)/z_0)$ instead of the intensity itself. Here the constant z_0 is an (arbitrary) unit of intensity. The choice of unit cannot influence the differential geometry of image space because it represents a mere shift of origin along the affine Z–dimension. Then the Z–axis becomes the affine line. Of course the global $\{x, y, Z\}$ space still fails to be Euclidian because the (log–)intensity dimension is incommensurable with the $\{x, y\}$ dimensions, whereas the $\{x, y\}$ dimensions are mutually compatible. The latter are measured as length or optical angle, the former in some unrelated physical unit (for instance photon number flux per area in the case of the CCD chip). But the approach to Euclidian space is much closer than with the use of intensity as such: As shown below we arrive at one of the 27 (three–dimensional) Cayley–Klein geometries[10,2] (of which Euclidian space is another instance).

3 The Geometry of Image Space

Consider "image space" $\mathbb{I}^3$ with coordinates $\{x, y, Z\}$. It is a (trivial) fiber bundle with the "picture plane" $\mathbb{P}^2$ as base space and the "log–intensity domain $\mathbb{L}$" as fibers. An image $Z(x, y)$ is a cross section of the bundle. We will use the term "pixels" for the fibers. A point $\{x, y, Z\}$ is said to have log–intensity Z and "trace" $\{x, y\}$. Image space is an infinite three dimensional space with the Euclidian metric in planes parallel to the xy–plane and a—quite independent—metric on the Z–axis. $\mathbb{I}^3$ can hardly be expected to be anything like a Euclidian space $\mathbb{E}^3$. For instance, it is hard to conceive of a rotation of an object by a quarter turn about the x–axis, because that would align the y-directions in the object with the Z–axis. What sense can anyone make of a "coincidence" of a distance in the picture plane with the intensity domain? $\mathbb{P}^2$ and $\mathbb{L}$ are absolutely incommensurable. This entails that *such operations should be forbidden by the geometry of* $\mathbb{I}^3$. To be precise, any geometrical operation should leave the pixels invariant[13].

We refer to these invariant lines (pixels) as "normal lines" (for reasons to be explained later) and planes that contain normal lines "normal planes". Such entities clearly cannot occur as tangent planes or lines in images because that would entail that the image gradient wouldn't exist. When we say "line" or "plane" we exclude the normal ones. This means that any plane can be represented as $Z(u, v) = Z_0 + (g_x u + g_y v)$. Here Z_0 is the intercept with the normal line through the origin and $\mathbf{g} = \nabla Z = \{g_x, g_y\}$ is the (log intensity) gradient of the plane. The normal planes cannot thus be represented, they are *special.* A similar reasoning applies to lines.

3.1 The Group of Congruences

Thus the space we are after is a three–dimensional space, such that the intensity domain (one–dimensional) and the picture plane (two–dimensional) "don't mix". The only transformations of any relevance thus leave a family of parallel lines (the "pixels") invariant. What other constraints does one *a priori* have? One obvious candidate is to require the space to be "homogeneous", *i.e.*, to require that a group of "congruences" exists, such that "free mobility" of configurations is guaranteed all over space. This assumption merely expresses the fact that one piece of picture is as good as the next, it expresses our total prior ignorance and simply extends the assumption needed to arrive at the structure of the intensity domain. *But then the case is settled:* The space has to be of constant curvature (or free mobility is jeopardized) and is one of the 27 Cayley–Klein geometries. The invariance of a family of parallel lines fixes it to the "simple isotropic geometry". This geometry is obtained from projective geometry when one singles out two intersecting lines in the plane at infinity as the "absolute conic". One may take the lines $x = \pm iy$ in the plane $Z = \infty$ with intersection $\{x, y, Z\} = \{0, 0, \infty\}$ as the absolute conic. Then all lines parallel to the Z–axis meet in the "vanishing point" $\{0, 0, \infty\}$. The group G_8 (8 parameters) of "direct isotropic similarities" that is the analog of the (7 parameter) group of Euclidian similarities[15] contains the projective transforms that leave the absolute conic invariant. (In this paper we stick to the notation introduced by Sachs[15] for this group and its subgroups.) We write it in a form such that the identity is obtained when all parameters are set to zero and such that the factors e^h and e^δ are positive definite.

$$
\begin{aligned}
x' &= \mathrm{e}^h(x\cos\phi - y\sin\phi) + t_x \\
y' &= \mathrm{e}^h(x\sin\phi + y\cos\phi) + t_y \\
Z' &= \mathrm{e}^\delta Z(x,y) + \alpha_x x + \alpha_y y + \zeta
\end{aligned}
\tag{1}
$$

G_8 is indeed the only group of transformations that conserves normal lines and generic planes as families (see figure 1).

The "movements" are thus very simple, namely (a special kind of) linear transformations combined with a shift. It is not likely to be a coincidence that the familiar transformations that apparently leave the picture "invariant" are members of G_8. Examples include the use of various "gradations" ("soft", "normal", "hard") in photography, the introduction of lightness gradients when enlarging negatives[1], *etc.* The latter equalize "uneven density" in the negative, often introduced when an off–axis piece of the image is being enlarged. Similar transformations are common in many image processing packages. The use of paper gradations is aptly caught by the "gamma transformations" of the form $z' = (z/z_0)^\gamma$, whereas gradients are approximated with $z'(x, y) = z(x, y)\ \exp(\sigma_x x + \sigma_y y)$. Here the exponential serves to ensure the definite positiveness of the intensity, necessary because a "linear gradient" is a nonentity. Such gradients commonly occur in unevenly lit scenes. Notice that all these "image preserving" operations are *linear* in the log–intensity domain. The transformations affect only the log–intensity

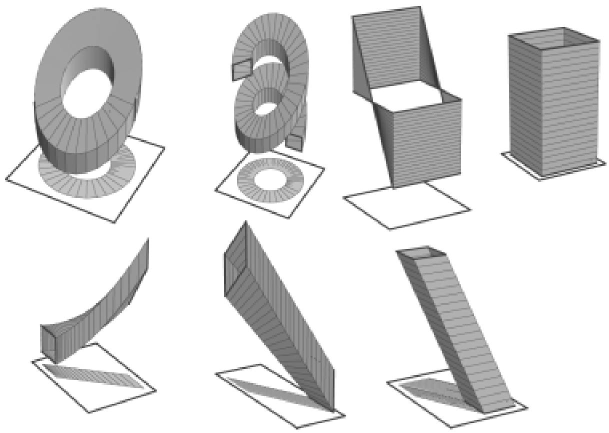

Fig. 1. Orbits of significant one–parameter subgroups. These groups appear either as identities, translations, or rotations in their traces on the picture plane. In image space the groups are far richer, for instance, the "rotations" may appear as screw motions with a normal line as axis, or as periodic motions that transform a paraboloid of rotation with normal axis in itself.

domain, not the traces. We suggest that the popularity of "gamma transformations" and "lightness gradients" derives from this fact. Image *structure* in the sense of curvatures of surfaces in $\mathbb{I}^3$ is indeed not in the least affected.

These transformations leave the "image" invariant (see figure 2) in the sense that photographers apply such transformations as they see fit, merely to "optimize" the image without in any way "changing" it[1]. Thus it is apt to say that *images are the invariants of* G_8. Hence we will consider G_8 as the *group of congruences and similitudes* of image space.

For any two points $\{x_{1,2}, y_{1,2}, Z_{1,2}\}$ we define the "reach" as the unsigned quantity $r_{12} = +\sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}$. When the reach vanishes but the points are distinct we call them "parallel". In that case (and in that case only!) we define the "span" $s_{12} = Z_2 - Z_1$ (a signed quantity). Both reach and span are relative invariants of G_8. Consider the subgroups $\mathcal{B}_7$ of reach preserving and S_7 of span preserving "isotropic similarities". Their intersection is the group $\mathcal{B}_6^{(1)}$ of "simple isotropic movements" (also known as "unimodular isotropic movements"). The group $\mathcal{B}_7$ is characterized by $h = 0$, S_7 by $\delta = 0$, and $\mathcal{B}_6^{(1)}$ by $h = \delta = 0$. We define the "distance" between points as their reach when the reach is different from zero and their span if not. The distance is an absolute invariant of the simple isotropic movements.

Planes $Z(x, y) = Z_0 + (g_x x + g_y y)$ have "plane coordinates" $\{u, v, w\} = \{g_x, g_y, Z_0\}$. Two planes $\{u_{1,2}, v_{1,2}, w_{1,2}\}$ subtend a "skew", that is the unsigned quantity $s_{12} = +\sqrt{(u_2 - u_1)^2 + (v_2 - v_1)^2}$. When the skew vanishes but the planes are distinct we call them "parallel". In that case (and in that case only!) we define the "gap" $g_{12} = w_2 - w_1$ (a signed quantity). The "angle" subtended

Fig. 2. An "image" is an invariant over variations such as these. Thus the figure suggests only a *single* image, not six! Infinitely other variations might be given of course.

by the planes is defined as the skew if the skew is not zero or the gap if it is. The skew is an absolute invariant of the subgroup $\mathcal{W}_7$ defined through $h = \delta$. The group of isotropic movements is the intersection of $\mathcal{W}_7$ and $\mathcal{B}_7$. Notice that there exist two types of similarities: Those of the "1^{st} kind" ($\mathcal{W}_7$) scale distances and preserve angles, whereas those of the "2^{nd} kind" ($\mathcal{B}_7$) scale angles and preserve distances (these are gamma transformations). There exists a full metric duality between points and planes, planes and points behave the same under the group G_8.

The metric $ds^2 = dx^2 + dy^2$ that is respected by $\mathcal{B}_6^{(1)}$ is of course degenerate. It is perhaps best understood as a degenerate Minkowski metric[7]. Then the pixels appear as degenerated "light cones" (of relativistic kinematics). All points on the normal line "above" a point (higher intensity) and "below" a point (lower intensity) are inside the light cone and thus *comparable* whereas a generic pair of points is outside the light cone ("elsewhere" in the relativistic kinematics) and *not comparable*. Such points only have a reach, namely their distance in the trace. Their log–intensities are not in a fixed relation but are typically changed by isometries of $\mathbb{I}^3$.

3.2 The Structure of Normal Planes

Much of the structure of image space can be understood from a study of the structure of normal planes[14]. In a way this is the study of one dimensional images and thus has frequent applications by itself. We use coordinates $\{x, y\}$, with the y–coordinate being interpreted as log–intensity, the x–coordinate as the distance in the image. This should yield no confusion since it will always be clear when we confine the discussion to normal planes.

It is often convenient to identify the Euclidian plane with the complex number plane ($z = x + iy$ with $i^2 = -1$). The reason is that linear transformations induce the similarities: $z' = pz + q$ implies a scaling by $|p|$, rotation over $\arg p$ and translation by q. The distance between two points is $|z_1 - z_2|$. It is no less convenient to identify the normal planes with the dual number plane. Dual numbers[3,8] are written $z = x + \varepsilon y$ where the dual unit is nilpotent ($\varepsilon^2 = 0$). The distance $|z_1 - z_2|$ is $x_1 - x_2$ when we define the modulus $|z|$ as x (signed quantity!). A linear transformation $z' = pz + q$ with $p = p_1 + \varepsilon p_2$, $q = q_1 + \varepsilon q_2$ implies $x' = p_1 x + q_1$, $y' = p_2 x + p_1 y + q_2$, which is in G_8. We have a scaling of distances by $|p| = p_1$, a "rotation" over $p_2/p_1 = \arg p$, and a translation over q. Notice that we can indeed write $p_1 + \varepsilon p_2 = p_1(1 + \varepsilon p_2/p_1) = |p| \exp \arg p$. Almost all of the structure of the familiar complex plane can immediately be put to good use in the study of the normal planes. This study is much simplified by the nilpotency of the dual unit, for instance, the Taylor expansion truncates after the first order. Thus $\sin \varepsilon\psi = \varepsilon\psi$, $\cos \varepsilon\psi = 1$, $\exp \varepsilon\psi = 1 + \varepsilon\psi$ and so forth.

The unit circle is given by $x^2 = 1$ and consists of the two normal lines $x = \pm 1$. The normal line $x = 0$ contains points that may equally be considered "centers" of the unit circle (thus all real lines are "diameters"!). The group of pure shears $x' = x$, $y' = y + \phi x$ moves the unit circle in itself, leaving the centers invariant. It is called a "rotation over ϕ", for apparently the group is to be understood as the group of *rotations* about the origin. If we define the orientation of a line through the origin as the (special) arc length of the arc cut from the unit circle, then the line from the origin to $\{x, y\}$ has an orientation $\phi = y/x$ and is transformed into the x–axis by the rotation over $-\phi$. It is perhaps disconcerting at first that this orientation is not periodic. Angles in the normal plane take on values on $(-\infty, +\infty)$, you cannot make a full turn in the normal plane. This is of course exactly what is called for given the incommensurability of the log–intensity and picture plane dimensions. Notice that the rotations correspond to the application of gradients in image processing.

The normal plane differs from the Euclidian plane in that the angle metric (like the distance metric) is parabolic (the Euclidian plane has an elliptic angle metric). This has the virtue that one enjoys full duality between points and lines. In the $\mathbb{E}^2$ one has parallel lines, but no parallel points (points that cannot be connected by a line), thus duality fails in this respect. Due to this fact the structure of the normal plane is much simpler than that of the Euclidian plane.

The bilinear expression $y + Y = xX$ defines a *line* (as a set of collinear points in point coordinates $\{x, y\}$) when we consider constant $\{X, Y\}$ and a *point* (as a set of concurrent lines in line coordinates $\{X, Y\}$) when we consider constant $\{x, y\}$. We define the distance of two points as $d_{12} = x_2 - x_1$, except when the points are "parallel", then we substitute the "special" distance $\delta_{12} = y_2 - y_1$. Similarly, we define the distance of two lines as the angle subtended by them, thus $\delta_{12} = X_2 - X_1$, except when the lines are parallel, then we substitute the "special" distance $d_{12} = Y_2 - Y_1$. These definitions make sense because either distance is invariant under general movements. Consider the polarity π which interchanges the point p with (point–)coordinates $\{x, y\}$ and the line P with (line–)coordinates $\{X, Y\}$, such that (numerically) $X = x$, $Y = y$. Suppose that p is on Q, thus $y_p + Y_Q = x_p X_Q$: then $q = \pi(Q)$ must be on $P = \pi(p)$ because

$y_q + Y_P = x_q X_P$. If $\pi(A) = a$ and $\pi(B) = b$ then $\delta_{AB} = d_{ab}$ in the generic case, whereas when a and b are parallel $\delta_{ab} = d_{AB}$. We also define the distance between a point u and a line V (say) as the distance between u and the point $\overline{u}$ on V such that u and $\overline{u}$ are parallel (the point on V nearest to u, or the "projection of u on V). You have $|d_{uV}| = |x_u X_V - (y_u + Y_V)| = |d_{Uv}| = |X_U x_v - (Y_U + y_v)|$, which implies that when u is on V (thus $x_u X_V = y_u + Y_V$) then $d_{Uv} = d_{uV} = 0$. These properties indeed imply full metric duality of lines and points in the normal plane.

A circle like $x = \pm 1$ is more aptly called "circle of the 1^{st} kind" to distinguish it from more general "cycles" defined as entities that can be moved into themselves (a circle of the 1^{st} kind can also be rotated into itself and thus is a cycle too). In the normal plane the cycles include (apart from the circles of the 1^{st} kind), the so called "circles of the 2^{nd} kind" (which look like parabolas with normal axes), the generic lines and the normal lines. The circles are especially interesting. A circle $x + \varepsilon x^2/2\rho$ can be "rolled" over a line. One shows that distance traveled is ρ times the angle of rotation, which identifies ρ as "the radius" of the circle. The motion actually moves a whole family of "concentric" circles $x + \varepsilon(x^2/2\rho + \mu)$ in themselves. The "curvature" (reciprocal of the radius) equals y_{xx}. Notice that this expression is simpler than—and different from—the Euclidian $(y''/(1 + y'^2)^{3/2})$. The circles of the 2^{nd} kind hold many properties in common with the circles of the $\mathbb{E}^2$. Notice that the radius (or the curvature) may be negative though. The sign indicates whether the circle is a concave or convex one.

Since there exist two types of circles there exist two distinct notions of "inversion in a circle". Since we are only interested in transformations that conserve the normal rays (pixels!) individually, only the inversions in more general cycles are of interest though. Since lines are degenerated circles, inversions in lines are included. For instance the inversion induced by the x–axis turns an image into its negative and vice versa. Since inversions are conformal, they preserve local image structure.

Because of the full duality there exist two distinct types of similitudes in the normal plane. One type conserves distances and scales all angles by a common factor. The other type conserves angles and scales all distances by a common factor. Because of the constraint that normal lines should be conserved individually only the first mentioned type is of interest to us. Of course these transformations are nothing but the familiar gamma transformations.

Curves that run fully in the normal planes ("normal curves" say) are important in the study of surfaces via normal cuts. A curve $y(x)$ is parameterized by arc length x, thus the unit tangent is $\mathbf{T}(x) = \{1, y'(x)\}$. The curve's normal is simply $\mathbf{N} = \{0, 1\}$. The derivative of the tangent is $\dot{\mathbf{T}}(x) = y''(x)\mathbf{N}$, whereas $\dot{\mathbf{N}} = 0$. Thus the curvature of the curve is $\kappa_n(x) = y''(x)$. Here we distinguish the "normal curvature κ_n" from the curvature κ which vanishes identically since the projection of the curve in $\mathbb{P}^2$ is a straight line. The osculating circle at $x = x_0$ is $x + \varepsilon[y(x_0) + \kappa_n(x_0)(x - x_0)^2/2]$. When you specify the curvature as a function of arc length (its "natural equation") the curve is determined up to a motion.

3.3 The Geometry of Image Space Proper

In image space $\mathbb{I}^3$ the new (as compared with the normal planes) linear entities are planes. It is obvious how to define the distance between planes or of a plane and a line. One takes a point on one entity and drops a plumb line (normal line) on the other. The length of the gap divided by the distance to the common point(s) is the angle between the planes or the plane and the line. It is invariant under arbitrary motions and can thus serve as the distance measure. In case the entities are parallel one uses the gap.

It is convenient to introduce an orthonormal basis $\{\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3\}$, $\mathbf{e}_1$ and $\mathbf{e}_2$ spanning the picture plane $\mathbb{P}^2$, and $\mathbf{e}_3$ the log–intensity dimension $\mathbb{L}$. One has $\mathbf{e}_i \cdot \mathbf{e}_j = 0$ for $i \neq j$, $\mathbf{e}_i \cdot \mathbf{e}_i = 1$ for $i = 1, 2$ and $\mathbf{e}_3 \cdot \mathbf{e}_3 = 0$. This introduces the degenerate metric

$$d_{12}^2 = (x_2 - x_1)^2 + (y_2 - y_1)^2. \tag{2}$$

For parallel points we again use the special distance.

The bivector $\pi = \mathbf{e}_1 \times \mathbf{e}_2$ represents the unit oriented area in the picture plane. The bivectors $\sigma_1 = \mathbf{e}_2 \times \mathbf{e}_3$ and $\sigma_2 = \mathbf{e}_3 \times \mathbf{e}_1$ represent oriented unit areas in normal planes and are of a different nature. The trivector $\tau = \mathbf{e}_1 \times \mathbf{e}_2 \times \mathbf{e}_3$ is the oriented unit volume of the space. Notice that you have $\pi^2 = -1$, a regular "pseudoscalar" in the picture plane, but $\sigma_i^2 = 0$ and also $\tau^2 = 0$. It is easy to set up the complete "geometric algebra" for image space. Many operations and geometrical relations can be handled elegantly in this framework. For instance, the bivector π generates the Euclidian rotations in the picture plane, whereas the bivectors σ_i generate shears in normal planes (also "rotations", but in the sense of image space).

Some properties of $\mathbb{I}^3$ may appear unfamiliar at first sight. For instance, two parallel lines in the picture plane are indeed parallel lines (in the sense of the metric) in $\mathbb{I}^3$, yet typically have different intensity gradients and thus are not necessarily coplanar. Any pair of points on a common perpendicular can be given equal intensities through a congruence, thus all such pairs of points are equivalent. This is similar to the phenomenon of "Clifford parallels" of elliptic space. There even exist surfaces ("Clifford planes") which are nonplanar, yet carry two mutually transverse families of parallel lines.

In the remainder of this paper we will be predominantly interested in differential properties of image space. Some of these are well known to photographers and universally used in the darkroom. An example is the practice of "burning" and "dodging" by which one introduces essentially arbitrary modulations of the type $z'(x, y) = z(x, y) \exp w(x, y)$, where $w(x, y)$ is quite arbitrary. Notice that $Z'(x, y) = Z(x, y) + w(x, y)$, thus locally (to first order) one has (at the point $\{x_0, y_0\}$ say) $Z'(x_0 + dx, y_0 + dy) = Z(x_0, y_0) + a + b\,dx + c\,dy$ (with $a = w(x_0, y_0)$, $\{b, c\} = \nabla(Z + w)(x_0, y_0)$), which is a *congruence* of image space. Thus dodging and burning represent *conformal transformations* of image space[16], which is most likely why they work as well as they do.

The topic of primary interest here is *differential geometry* of surfaces in image space[17,18,19,20,15]. We will only consider smooth surfaces (at least three times differentiable) with tangent planes that never contain a normal line. Notice that the degenerate metric allows us to point out the geodesics on arbitrary surfaces right away. They are the curves whose projections on the image plane are straight. The geodesic curvature of any curve on a surface is simply the Euclidian curvature of its projection on the picture plane.

We will write a generic point on a surface or a curve as $\mathbf{R} = \mathbf{r} + Z(\mathbf{r})\mathbf{e}_3$, where $\mathbf{r}$ is the component in the picture plane. Because the tangent of a curve or tangent plane of a surface never contains a normal line, we may globally parameterize a curve as $x(s)\mathbf{e}_1 + y(s)\mathbf{e}_2 + Z(s)\mathbf{e}_3$ and a surface as $x\mathbf{e}_1 + y\mathbf{e}_2 + Z(x, y)\mathbf{e}_3$, a "Monge parameterization". This is typically the most convenient way to represent curves and surfaces.

3.4 Surfaces

Again, the most natural way to parameterize a surface in image space is as a "Monge parameterization" $x\mathbf{e}_1 + y\mathbf{e}_2 + Z(x, y)\mathbf{e}_3$. Because the metric is degenerate it is immediately obvious that the "First Fundamental Form" $d\mathbf{R} \cdot d\mathbf{R}$ (or metric) is simply

$$\mathrm{I}(dx, dy) = dx^2 + dy^2 \tag{3}$$

(thus $E = G = 1$, $F = 0$, that is what the Monge parameterization buys us). Thus all curves on the surface with straight traces are "geodesics".

A fruitful way to think of curvature is via the "spherical image". The tangent plane at $\{x_0, y_0\}$ is $\mathbf{R}(x_0 + dx, y_0 + dy) = \mathbf{R}(x_0, y_0) + [dx\mathbf{e}_1 + dy\mathbf{e}_2 + (Z_x\, dx + Z_y\, dy)\mathbf{e}_3]$. When we introduce the "Gaussian sphere" $\mathbf{G}(x, y) = x\mathbf{e}_1 + y\mathbf{e}_2 + (x^2 + y^2)/2\mathbf{e}_3$, we see that the tangent plane at $\mathbf{R}(x, y)$ is parallel to that at $\mathbf{G}(Z_x, Z_y)$. Thus the map that takes $\mathbf{R}(x, y)$ to $\mathbf{G}(Z_x, Z_y)$ is the analog of the "Gaussian image" of a surface[5]. When one applies a stereographical projection from the infinite focus of the Gaussian sphere, one obtains a mapping from a point $\mathbf{R}(x, y)$ of the surface to the point $\{Z_x, Z_y\}$ of what is generally known as "gradient space". Since the stereographic projection is isometric(!), gradient space is just as apt a representation as the Gaussian sphere itself. We hence refer to $\{Z_x, Z_y\} = \nabla Z(x, y)$ as the "attitude image" of the surface. It is indeed the analog of the spherical image, or Gauss map, of Euclidian geometry. Although the attitude image is in many respects simpler than the Gaussian image, it shares many of its properties. For instance, rotations of the surface in $\mathbb{I}^3$ lead to *translations* of the attitude image (of course translations of the surface don't affect the attitude image at all). Thus the shape of the attitude image is an invariant against arbitrary congruences of pictorial space. Similarities of $\mathbb{I}^3$ simply scale the attitude image (a possibility that cannot occur in Euclidian space which knows only similarities that conserve orientations.)

The area magnification of the attitude image is the intrinsic curvature $K(x, y)$. It is given by the determinant of the Hessian of the log–intensity

$$K(x, y) = Z_{xx}Z_{yy} - Z_{xy}^2. \tag{4}$$

The trace of the Hessian of the log–intensity $Z_{xx} + Z_{yy}$ is another important invariant. It can be interpreted as twice the "mean curvature" H, for because it is invariant against Euclidian rotations of the picture plane the average of normal curvatures in any pair of orthogonal directions equals the mean curvature.

The magnification of the surface attitude map is typically anisotropic. This is just another way of saying that the "sectional" curvature of the surface differs for different orientations of the section. Notice that we need to study the sectional curvature, as the equivalent of the normal curvature of Euclidian differential geometry. This is because the only reasonable definition of "surface normal" in $\mathbb{I}^3$ is to let them be normal directions. But then these "normals" are essentially useless to measure shape properties since they don't depend on the nature of the surface to begin with! However, we certainly have a right to speak of "normal curvature" as synonymous with "sectional curvature", just like in classical differential geometry. (This is also the origin of our term "normal plane" introduced above.)

In order to measure the curvature of a section we may find the rate of change of the image in attitude space, or—and this is completely equivalent—we may find the best fitting (highest "order of contact") of sectional circles. This latter definition is obviously the geometer's choice. As we rotate the planar section the radius of the best fitting normal circle changes (periodically of course). The sectional planes are indeed normal planes and the circles "normal circles" (parabolas with normal axis).

Remember that the sectional curvature is simply the second derivative of the depth in the direction of the section. There will be two directions at which the radius of the normal circle reaches an extremum, we may call these the "directions of principal curvature". Exceptions are points where the radius of the best fitting circle doesn't depend on the orientation of the section. Such points are rare (generically isolated) and may be called "umbilical points" of the surface.

The orientation of the directions of principal curvature are given by

$$Z_{xy}dx^2 - (Z_{xx} - Z_{yy})dx\,dy - Z_{xy}dy^2 = 0. \tag{5}$$

These directions are invariant under arbitrary congruences.

The curvature of a normal section that subtends an angle Ψ with the first principal direction is

$$\kappa_n(\Psi) = \kappa_1 \cos^2 \Psi + \kappa_2 \sin^2 \Psi, \tag{6}$$

where κ_1, κ_2 are the principal curvatures. This is identical to Euler's formula from the classical differential geometry of surfaces.

At the umbilical points the curvilinear congruences of principal directions have a point singularity. Such points are key candidates for significant "features" of the surface.

The osculating paraboloid at a point is simply the Taylor expansion of log–intensity up to second order terms. For elliptic convex points these best fitting approximations are biaxial Gaussians in terms of intensity (not log–intensity).

Notice that the formalism is typically *much simpler* than that for Euclidian differential geometry, for instance, compare the simple expression for the mean curvature

$$2H = Z_{xx} + Z_{yy},$$

with the corresponding expression for Euclidian geometry:

$$\frac{(z_{xx} + z_{yy}) + z_{yy} z_x^2 + z_{xx} z_y^2 - 2 z_{xy} z_x z_y}{(1 + z_x^2 + z_y^2)^{3/2}}.$$

Apart from being simpler the new expression has the additional advantage of being correct for a change and is thus to be recommended, despite the fact that it runs counter to conventional practice (or wisdom?).

4 Features and Significant Regions

A "feature" is the geometrical locus where some differential invariant vanishes. The value of the invariant is obtained by running an image operator at some scale. In most cases the required scale is obvious, in others a "best" scale has to be established[12]. One needs yet another scale in order to define the vanishing of the invariant, essentially the "bin width" at which one wishes to sample the values[6]. A "level set" such as $I(x, y) = I_0$ (in this example generically a *curve*) is really an (hopefully narrow) *area* $(I(x, y) - I_0)^2 < \epsilon^2$, where ϵ is the resolution in the I–domain, essentially the "bin–width". This can easily be captured in a neat formalism when the image scale space is augmented with a histogram scale space[6]. Level sets of differential invariants define "fuzzy features" $\exp(-(I(x, y) - I_0)^2/2\epsilon^2)$, that is to say, they assume the (maximum) value unity at the location of the feature and are near zero far from the location of the feature. This is the robust and principled way to find (or *define*) "features".

Examples of significant features (see figure 3) are the parabolic curves ($K = 0$), minimal curves ($H = 0$), ridges and ruts, and umbilical points. Special points of interest are the inflection points of the parabolic curves, and the points of crossing of ridges of unlike type. Notice that "points" are detected as fuzzy spots and "curves" as fuzzy ribbons or even areas. This makes perfect sense when you think of the nature of level "curves" on shallow slopes and in the presence of noise. Whether a pixel belongs to the level "curve" is a statistical issue and the fuzzy membership function expresses this.

Significant *regions* are obtained through thresholding of differential invariants, usually at zero value. Again, one needs to define "fuzzy" characteristic functions. One defines the characteristic region via $(1 + \mathrm{erf}((I(x, y) - I_0)/\epsilon))/2$. It equals near unity inside and near zero outside the region.

Here is a simple application: It is tempting to use thresholding as a coarse way of segmentation. The problem is that it cannot yield an invariant result. A local extremum is meaningless because I can shift it with a congruence of image space. In the place of a maximum (or hill area) you may substitute an area where $K > 0$ (elliptic) and $H < 0$ (convex). Any point in such elliptic convex areas can be turned into a maximum through the application of a suitable movement whereas no other points can. Likewise, in place of the minima (or valley areas)

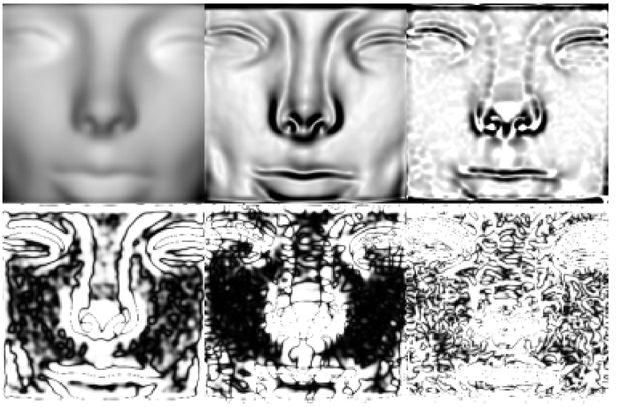

Fig. 3. Features for the case of a face image. Top from left to right: The image at the resolution of the operators (lighter is "more"), the gradient magnitude and the curvedness. These display scalar magnitudes. Bottom from left to right: Minimal ($H = 0$), parabolic ($K = 0$) loci and ridges (a 3^{rd}–order invariant). These are (fuzzy!) curves. For all the invariants darker is "more".

you may put areas where $K > 0$ and $H > 0$. Notice that at the boundary of such areas one principal curvature vanishes and $H \neq 0$. Indeed, $K = 0$ and $H = 0$ only occurs at planar points which are not present in the generic case. Thus the areas are bounded by the parabolic curves and the sign of the mean curvature merely decides upon the type. One may denote such areas "domes" and "bowls" which is what they look like. Unlike the results of raw thresholding these are significant regions because invariant against arbitrary image space movements.

5 Conclusion

We have introduced a geometrical framework that allows one to handle image structures in a principled manner. The basic structure depends upon two major considerations. The first is a careful examination of the physical nature of the intensity domain. It turns out to be the case that only the log–intensity representation can be considered "natural" (in the geometrical sense) because it does not commit one to any particular choice of unit or fiducial intensity. The second is a careful examination of the group of transformations that leave image structure invariant (the "similitudes" and proper motions). We identify this group with the transformations that conserve the spatial structure ("pixels") and conserve lines and planes in image space as families. That this choice is particularly apt is evident from the fact that many of the common transformations used for more by a century by photographers and commonly available in image processing packages (of the Photoshop type) are easily identified as subgroups. Even nonlinear transformations as "dodging" and "burning" are easily identified as *conformal transformations* of the space[16].

Notice that all this derives from a single, very simple assumption, namely *Complete ignorance as to location in image space*. From this it follows that no intensity range is singled out and that the space is homogeneous (that "free mobility" is guaranteed). That the movements should conserve a family of parallel lines can hardly be counted as an "assumption": It is the only way to be consistent.

The geometry we obtain is one of the 27 Cayley–Klein geometries. It is the product of the Euclidian plane (with parabolic distance metric and elliptic angle metric) and the isotropic line[13] (with parabolic distance metric). This makes that the geometry of the isotropic planes (called "normal planes" in this paper) is governed by a degenerate, parabolic distance metric and a parabolic angle metric. This is exactly what makes this geometry a natural representation of image space. Since the slant of planes is not periodic but may take values on the full real axis, one cannot "turn around" in image space. This corrects the irritating oddity implicit in the conventional Euclidian choice where one may turn (in principle) the intensity domain so as to lie in the picture plane. Although we have seen no author explicitly admit this, it is implicit in the (very common) use of Euclidian expressions (for the curvature of image surfaces for instance) in image processing. Such Euclidian expressions are invariants under the group of *Euclidian* isometries, including rotations about *any* axis.

"Image processing done right" indeed implies that one uses the *isotropic* geometry. This may appear exotic at first blush, but some exposure soon leads to an intuitive understanding. The fact that the formalism becomes generally much simpler is perhaps an incentive to move over. The inadvertent use of differential invariants of G_8 is quite widespread in image processing applications. Such practices are justified through the fact that they work. The present theory puts such practices upon a principled foundation and—more importantly—allows a disciplined analysis and extension.

References

1. Adams, A.: *The Print: Contact Printing and Enlarging*. Basic Photo **3**. Morgan and Lester, New York, 1950
2. Cayley, A.: Sixth Memoir upon Quantics. Phil.Trans.Roy.Soc.Lond. **149** (1859) 61–70
3. Clifford, W. K.: Preliminary sketch of biquaternions. Proc.Lond.Math.Soc. (1873) 381–395
4. Florack, L.: *Image Structure*. Kluwer, Dordrecht, 1997
5. Gauss, C. F.: *Algemeine Flächentheorie*. (German translation of *Disquisitiones generales circa Superficies Curvas*), Hrsg. A. Wangerin, Ostwalds Klassiker der exakten Wissenschaften **5**. Engelmann, Leipzig, 1889 (orig. 1827)
6. Ginneken, B. van, Haar Romeny, B. M. ter: Applications of locally orderless Images. In: *Scale–Space Theories in Computer Vision*. Eds. M. Nielsen, P. Johansen, O. F. Olsen and J. Weickert. Second Int.Conf. Scale–Space'99, Lect.Notes in Computer Science **1682**. Springer, Berlin (1999) 10–21

7. Jaglom, I. M.: *A simple non–Euclidian geometry and its physical basis: an elementary account of Galilean geometry and the Galilean principle of relativity*. Transl. A. Shenitzer, ed.ass. B. Gordon. Springer, New York, 1979
8. Jaglom, I. M.: *Complex numbers in geometry*. Transl. E. J. F. Primrose. Academic Paperbacks, New York, 1968
9. Jaynes,E. T.: Prior Probabilities. IEEE Trans. on Systems Science and Cybernetics **4**(3) (1968) 227–241
10. Klein, F.: Über die sogenannte nichteuklidische Geometrie. Mathematische Annalen Bd.**6** (1871) 112–145
11. Klein, F.: *Vergleichende Betrachtungen über neue geometrische Forschungen* (the "Erlangen Program"). Programm zu Entritt in die philosophische Fakultät und den Senat der Universität zu Erlangen. Deichert, Erlangen, 1872
12. Lindeberg, T.: *Scale–Space Theory in Computer Vision*. Kluwer, Dordrecht, 1994
13. Pottmann, H., Opitz, K.: Curvature analysis and visualization for functions defined on Euclidian spaces or surfaces. Computer Aided Geometric Design **11** (1994) 655–674
14. Sachs, H.: *Ebene Isotrope Geometrie*. Vieweg, Braunschweig/Wiesbaden, 1987
15. Sachs, H.: *Isotrope Geometrie des Raumes*. Vieweg, Braunschweig/Wiesbaden, 1990
16. Scheffer, G.: Verallgemeinerung der Grundlagen der gewöhnlichen komplexen Funktionen. Sitz.ber.Sächs.Ges.Wiss., Math.–phys.Klasse, Bnd. **45** (1893) 828–842
17. Strubecker, K.: Differentialgeometrie des isotropen Raumes I. Sitzungsberichte der Akademie der Wissenschaften Wien **150** (1941) 1-43
18. Strubecker, K.: Differentialgeometrie des isotropen Raumes II. Math.Z. **47** (1942) 743–777
19. Strubecker, K.: Differentialgeometrie des isotropen Raumes III. Math.Z. **48** (1943) 369–427
20. Strubecker, K.: Differentialgeometrie des isotropen Raumes IV. Math.Z. **50** (1945) 1–92