A GENERAL PURPOSE GEOMETRIC DISTORTION MODEL FOR CENTRAL PROJECTION CAMERAS

By JAVIER A. MÚGICA
2005-2010

INTRODUCTION

THE idea of the forthcoming work stems from an article by K. Jacobsen: *Block Adjustment*. There the drawbacks of the traditional odd polynomial for the radial symmetric distortion are briefly depicted and a better model is provided, though it is not explained because the object of the paper is a different one. The parameters are however easy to interpret, and an analysis of them lead me to prefer a somewhat different set. I first developed the model for my diploma thesis back in 2005. Since that time the model has been subject to one important rectification and some minor improvements, and above all the opportunity for substantial testing and hence drawing conclusions arose two years ago, and here are the results presented.

The principle upon which the model was to be developed was the orthogonality among its components. The derivation of the orthogonal base was not so straightforward as it first appeared to be. Some distortion components are tightly bound to some orientation parameters. Once the different interdependencies are understood a polynomial base for the symmetric components arises naturally from its first, mandatory, components. The asymmetric components are thence derived in a natural way too.

The actual square or rectangular shape of photographs has to be taken into account for orthogonality —i.e. orthogonality is to be sought for a photograph-shaped domain.

THE MATHEMATICAL SETTING

The geometry

The taking of an image is idealised as a central projection. Such a projection is defined by the position in space, relative to the object being projected, of the projection centre O and the projection plane π, which account for six parameters. There is a distinguished point in the projection plane, viz. the point where the perpendicular, let fall from the point O onto the plane π, meets the plane π. It is called the principal point and will be denoted by the letter p.

If we further want plane coordinates to be assigned to the image points a coordinate system has to be placed onto the plane, adding three parameters and making a total of nine. The equations relating object coordinates to image coordinates are the well known collinearity equations.

I shall denote object points with capital Latin letters: *A*, *B*, *C*, etc., and image points with lower case ones: *a, b, c,* etc. I shall similarly denote coordinates in the object system by (X, Y, Z), thus $(X, Y, Z)_A$, X_A, etc., and real image coordinates by (x', y'). The image not being perfect, these coordinates do not correspond to any central projection.

The easiest definition of the 6 + 3 projection parameters is as follows: For the first six, the coordinates $(X, Y, Z)_O$ of the projection centre; the distance f from the centre O to the projection plane π, where f stands for focal length, but it should be better called projection distance; and two other parameters, $\varOmega$ and $\varPhi$, defining the orientation of the plane with respect to the object coordinate system. There is no easiest definition for these two parameters, and their geometrical meaning, if any, may be varied to better fit actual needs (in particular, to avoid singularities).

For the last three parameters there can be taken an angle κ defining the orientation of the image axes, to be measured from an origin the definition of which may be indirect, and the coordinates (x_p, y_p) in the image system of the principal point.

These parameters define the projection and hence, for any object point A, its theoretic image coordinates $(x, y)_a$, and for the entire space the whole theoretic image, where this should be thought as the whole set of (x, y) coordinates. The real image coordinates of the point a will be some different ones $(x', y')_a$, and the difference of these with respect to $(x, y)_a$ is the value of the distortion at the point a. In the like manner, the difference of the whole set (x', y') with respect to (x, y) is the distortion function. When I refer to the image, without any qualifier, I shall mean the real image.

The previous description divides the parameters according to their geometric nature:

$$X_O,\ Y_O,\ Z_O,\ \varOmega,\ \varPhi,\ f \qquad \kappa,\ x_p,\ y_p,$$

since from the geometric viewpoint the last three just don't exist. Here κ does mean a rotation on the x, y plane, while $\varOmega$ and $\varPhi$ are just symbolic names for two parameters defining the orientation of the plane π. I write a lower case κ to emphasize that its effect is a movement on the photograph, like x_p and y_p, in contrast with $\varOmega$ and $\varPhi$ that change the geometry of the image.

It is more common a classification according as to whether the parameters vary from one photograph to another or on the contrary remain fixed. These two sets are called exterior and interior parameters respectively:

$$X_O,\ Y_O,\ Z_O,\ \varOmega,\ \varPhi,\ K, \qquad f,\ x_p,\ y_p.$$

When developing a distortion model both these partitions are of little interest. The relevant division of the nine parameters together with the distortion function is as follows:

X_O, Y_O, Z_O,

$\varOmega$, $\varPhi$, x_p, y_p, κ, f, and the associated distortion components

The remaining distortion components.

This will be made apparent in the next sections.

The orientation plus distortion decomposition: the non-uniqueness problem

When calibrating an image there exists a set of (X, Y, Z) object coordinates, and there is also the set of corresponding (x', y') image coordinates. It is required to decompose the transformation from (X, Y, Z) to (x', y') in the form $D \circ T$, where T is a central projection and D a distortion, that is,

$$\underset{(X,Y,Z)}{\text{Obj.}} \xrightarrow{T} (x, y) \xrightarrow{D} \underset{(x',y')}{\text{Im.}}$$

The function D is very near unity, and it is therefore that its difference from unity, which we shall call $\mathcal{D}$, is more often used. This agrees with our previous definition of the distortion function as the difference between the coordinates (x', y') and (x, y).

It may seem from the definition of the distortion that any set of orientation parameters is possible, for once they are fixed the theoretic image is defined, and the distortion at any point would just be the difference between its real and theoretic coordinates. We will see that this is not true, but nonetheless the only restriction applies to the projection centre and it is indeed true that, in principle, any collection of values for the other six parameters may be selected at will.

Let two points from the object be aligned with a projection centre O_1. Their theoretic image will be a single point, and it will be as well their real image. If we now move the projection centre to a different position O_2, the two points will not in general be aligned with O_2, and their theoretic images will not be a single point any more, nor will their real images be. A set of points aligned with the projection centre, i.e. a straight line passing through the projection centre, is a projective ray. The set of projective rays is different for every projection centre, and so: Given an image $\{(x', y')\}$ there is only one possible projection centre O which, combined with some other values for the six other parameters, whatever they be, and composed with a distortion, whatever it be, can make the given image arise out from the given object.

The previous assertion has to be polished. We do not have the image of the whole space; we will in general not even have couples of points aligned with the projection centre. But the fact that some two point do not have the same image implies that the projection centre cannot lie on the straight line joining those two points. Furthermore, alignment is just one particular example, and the relative position on the photograph of the image of the measured object points will always provide information about the position of the projection centre. These considerations lead to the study of the correct placing of object points so that the projection centre can be well defined by the measured pairs $\{(X, Y, Z), (x', y')\}$. But this is not the aim of this paper and I shall not dwell on it.

The need of a perfect projection centre is not so much an idealisation as it may seem at first. The precision in the definition of the projection centre is of interest with respect to the object space, to the object size, while the actual manufacturing uncertainty is in the order of magnitude of image precision, a few microns.

The projection centre staying fixed, the effect of the variation of any other parameter is a function form the points in the plane to the points in the plain; that is, a distortion.

Fig. 1. Images differing in parameters other than the projection centre

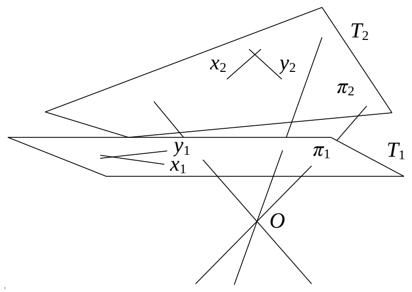


Fig. 1 pictures two distinct projections that differ in parameters other than the projection centre coordinates. If the parameters of T_1 and those of T_2 are known, in order to know the coordinates by the projection T_2 of a certain point it is not necessary to know the original position of the point in space, and the knowledge of its coordinates by the projection T_1 suffices. Hence the passage from T_1 to T_2 is a function from the plane into the plane, and this is a distortion; and if we have that $(T_1,\mathcal{D}_1)$ is a possible solution for the image, then $(T_2,\mathcal{D}_2)$ will also be a solution for some other $\mathcal{D}_2$. This $\mathcal{D}_2$ is such that $\mathcal{D}_2-\mathcal{D}_1$ is the opposite of the variation undergone by the theoretic image when passing from T_1 to T_2.

The set of possible $(T,\mathcal{D})$ is therefore infinite, and a decision has to be made to select one particular solution thereof.

Its consequence in designing the distortion model

The distortion function is usually expressed as a linear combination of some primitive functions. This set of functions is called base, and together with the selection criterion, whose need is derived from the previous discussion, defines the model. However, the base might be so designed that there is only one possible solution $(T,\mathcal{D})$, i.e. that linear combinations from the functions of the base can only express a restricted set of all possible distortions, in such a way that there is always one and just one solution $(T,\mathcal{D})$. In those cases there is no need for a subsequent selection criterion, and the base alone defines the model.

Let $\mathcal{B}$ be a base; I shall denote by $\langle\mathcal{B}\rangle$ the set of all linear combinations of functions of the base. In the following sections a base $\mathcal{B}$ will be developed such that the only possible solution $(T,\mathcal{D})$ with $\mathcal{D}\in\langle\mathcal{B}\rangle$ is, amongst the infinite solution pairs $(T,\mathcal{D})$, the one in which $\mathcal{D}$ is smallest.

The measure which I adopted to quantify the size of a distortion function is the quadratic mean, viz. $\int_F \mathcal{D}^2\,ds$, where F denotes the photograph, that is, the domain of the integral is the photograph, and ds is the differential of surface. To actually obtain the mean we should further divide by the total surface of the photograph and extract the square root, but these can be omitted when comparing distortions one to another.

DISTORTIONS CORRESPONDING TO EACH PARAMETER

Radial and tangential decomposition

Because of the radial symmetry of objectives the distortion is better expressed in polar coordinates, and the values it takes are not of the form $(\Delta x, \Delta y)$, but rather $(\Delta r, \Delta t)$, where r is the radial coordinate and $t = r\theta$, where θ is the angular coordinate. Hence, $(\Delta r, \Delta t) = \mathcal{D}(r, \theta)$. Actually, Δt is applied along the perpendicular straight line to the radial direction, not along the arc.

The origin of the polar system is not the origin of the image (fiducial) coordinate system, which may be anywhere, but the principal point. For this reason (x, y) and (r, θ) will henceforth denote the theoretic coordinates with respect to the principal point, which do not depend upon the parameters x_p, y_p. The real coordinates are therefore

$$(x', y') = (x, y) + \mathcal{D}(x, y) + (x_p, y_p), \tag{1}$$

where the distortion is computed in polar coordinates as just explained.

Distortions corresponding to f and κ

In this and the next sections the distortions equivalent to the variation of each orientation parameter will be derived, with exception of (X_O, Y_O, Z_O) that are not equivalent to any distortion and must remain fixed. The easiest parameters to handle are the three ones defining the image coordinate system: a change in the parameters x_p, y_p will cause an equal change of the distortion in all points. It becomes less simple when expressed in polar coordinates, but in any case these parameters are not varied alone, and a change in them is always taken jointly with a change in the principal point itself, as will be explained in the next section.

A variation in κ will add an equal rotation to the distortion function, centred at the principal point (according to the definition of κ, x_p and y_p), and so it is $\Delta t = \Delta\kappa\, r$. Thus, if the κ parameter of a photograph is increased by an amount $\Delta\kappa$, the tangential distortion has to be modified by $-\Delta\kappa\, r$ so that the composition $D \circ T$ remains the same.

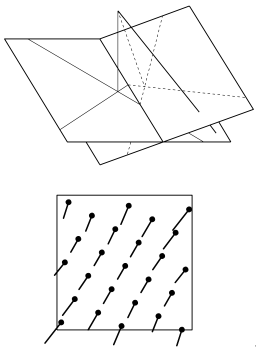
The variation of the projection distance f is easy to study as well. It will scale the image from the principal point as centre, namely $\Delta r = (\Delta f / f) r$. The f in the denominator is the original one, previous to the modification. Thus, the radial distortion will have to be modified by $-(\Delta f / f) r$, and vice versa.

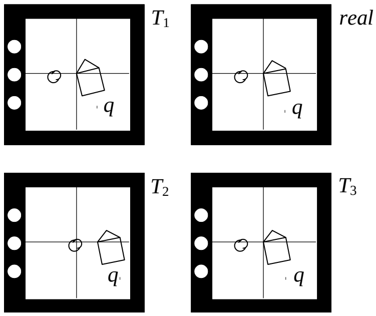
Distortions corresponding to Ω and Φ; qualitative analysis

Fig. 2 shows two projections where the only parameters that vary from one to the other are Ω and Φ. These parameters define the orientation of the projection plane, and hence that of the principal ray, the one corresponding to the principal point. Let us suppose that $x_p = 15\ \mu\text{m}$ and $y_p = 0\ \mu\text{m}$, as shown in Fig. 2 in π_1

(15 microns is very little and it is not visible). The principal ray is a definite line of the object space, that may intersect the object at, for instance, a building's corner. This definite point will have a real image, whose actual image coordinates need not be (x_p, y_p), in which case the distortion at the principal point will not be zero. The upper right of Fig. 3 represents the real image, and the upper left the theoretic image according to T_1.

Fig. 2. Images differing in Ω and Φ (left)
Fig. 3. A real image and three possible theoretic images for it (right)





If we now change the Ω and Φ parameters, the principal ray will change accordingly, and it may now come, for instance, from a point in a bush, which becomes the new principal point. We are supposing that no other parameter changes, and in particular x_p and y_p do not, so the principal point stays at fifteen microns form the theoretic centre of the image coordinate system, and the theoretic image is that shown in the lower left of Fig. 3. The bush is far away from the origin in the real image, so the distortion at the principal point is a huge one.

If we want the distortion at the principal point to be zero, the parameters x_p, y_p need to be changed so as to make them equal to the real image coordinates of the new principal point. The new projection (Fig. 3, lower right) is different from T_1, but not as much as T_2.

It is illuminating to imagine the theoretic image axes sliding through the plane as the parameters x_p, y_p and κ are varied. And not only the axes, but the photo frame bounded to the axes as well. And when a set of values for x_p, y_p and κ is fixed, the axes stop, and a definite window is selected form the infinite plane π, and this becomes the theoretic image. If different values were taken for those parameters, a different window would be drawn, and a different theoretic image would arise. The theoretic position of the principal point image with respect to the image coordinate centre, the pair (x_p, y_p), may or may not coincide with the real one.

This concept is largely misunderstood. The misconception stems from the identification of the (x_p, y_p) parameters with the principal point. These are not the principal point. *The Ω and Φ parameters* are *the principal point.* The parameters x_p and y_p are better understood as the position of the coordinate system centre with respect to the principal point, and not the other way round; and they are just the theoretic position, prior to distortion.

Therefore in formula (1) the values x_p, y_p are the theoretic coordinates, which will only coincide with the real ones if the distortion at the principal point is zero. The same applies when going back from the real coordinates (x', y') to the theoretic (x, y), which is just the inversion of formula (1).

To help comprehending this concept, a real image may be imagined perfect in all of it save at a small area around the principal point, as if a small bulb had been approached to the principal point, heating the film (these days the pixel matrix) up to deformation. The distortion at the principal point will not be zero and its image coordinates will not be (x_p, y_p). If on the contrary we insisted that the image coordinates of the principal point be (x_p, y_p), or which is the same that the distortion at that point be zero, the former distortion at that point will thereby be transferred to all the photograph.

It is however very inconvenient to have a non-zero distortion at the principal point, for even if when understood the concept may appear clear, it is not common to the photogrammetrist, as far as I know, to understand it in such a transparent manner, let alone to a majority of operators with little photogrammetric knowledge. For analogical cameras this is a mild restriction, because the nature of lenses makes the central part of the image the most perfect of it; but in digital cameras the camera distortion is the combination of both the objective and the pixel matrix distortions. The matrix is not a perfect grid, and irregularities may be present in any or the other area of it without any particular preference, and the imposition that the distortion be zero at the principal point is not a valid one from the theoretical point of view. However, for the above mentioned reason and lest confusion reigns, we had better restrict ourselves to distortion functions that are zero at the origin.

So when the base $\mathcal{B}$ shall be created, the only solution $(T, \mathcal{D})$ with $\mathcal{D} \in \langle \mathcal{B} \rangle$ will not exactly be the one for which $\mathcal{D}$ is smallest amongst all possible solutions, but the one in which it is smallest among those solutions where $\mathcal{D}(0,0) = (0,0)$.

In order that the distortion at the principal point is ever zero, when the Ω and Φ parameters are varied it is also necessary to vary x_p and y_p accordingly and vice versa. It is well known and not difficult to show that for any two central projections with a common projection centre there is a point from where image angles are equal, and so the transformation from one image to the other is a radial displacement of the points with respect to the isocentre, which is how that singular point is known. The isocentre's projective ray is the bisector of the two principal rays; therefore, the isocentre lies between the two principal points.

The variations in Ω and Φ which we would be dealing with will be very small, the subsequent displacement of the principal point being just a few pixels. Therefore the displacement of an image point towards the isocentre will not be distinguishable form a displacement towards one of the principal points. The displacement is

$$\Delta r = ar^2 \cos\theta + br^2 \sin\theta, \tag{2}$$

for two constants a and b depending upon the rotation that transforms one projection into the other. In the vicinity of the isocentre, where the difference between a displacement toward the isocentre and a displacement toward one of the principal points could be significant, the displacement function has a second order zero, so it is not the source of any problem. Hence: *A displacement of the principal point does not induce any tangential distortion.*

I have presented the previous result as a consequence of a combined change in the x_p, y_p, Ω and Φ parameters, but this is not exactly true. Formula (2) is the resulting distortion for the rotation that carries one projection plane onto the other. Since the orientation of the plane is defined by the Ω and Φ parameters, these parameters will do vary indeed. But, as we have already mentioned, the origin of the κ angles has an indirect definition, and a change in the projection plane orientation may affect it as well. This is not significant to our previous reasoning. The relevant fact is that a change in the x_p and y_p parameters can be combined with a change in the Ω, Φ and, possibly, κ parameters, so as to generate a distortion like that in (2); and similarly can Ω and Φ be combined with x_p, y_p and, possibly, κ. Moreover, the variation in the κ parameter is not necessary in order for the condition $\mathcal{D}(0,0) = (0,0)$ to hold. If it were not included, an extra κ rotation would appear in equation (2).

<u>*Distortions corresponding to Ω and Φ; Formulae*</u>

We want to relate the change in the principal point: $(\Delta x_p, \Delta y_p) = (\varepsilon_x, \varepsilon_y)$; the parameters a and b of the distortion of formula (2), and the variation of the parameters Ω, Φ and κ. Their variation is not independent, and if the parameters of any of the three groups are varied the others have to be changed accordingly so that i) the composition $D \circ T$ remains the same and ii) the condition $\mathcal{D}(0,0) = (0,0)$ continues to hold.

The relation of more interest now is that between $(\varepsilon_x, \varepsilon_y)$ and (a, b). By known geometric constructions it is found that if the principal ray is changed so that the position of the principal point varies an amount $(\varepsilon_x, \varepsilon_y)$, the theoretic image will change by an amount

$$\Delta r = -(\varepsilon_x / f^2) r^2 \cos\theta - (\varepsilon_y / f^2) r^2 \sin\theta.$$

Since the real image remains the same, the parameters a and b have to be varied by an amount

$$\Delta a = \varepsilon_x / f^2, \qquad \Delta b = \varepsilon_y / f^2 \tag{3}$$

and conversely, if the parameters a and b are increased by $(\Delta a, \Delta b)$ the principal point has to be moved by $(\Delta a f^2, \Delta b f^2)$.

The variation of the parameters Ω, Φ, κ is not needed for the analysis of the distortion and will be given for completeness. It can be found as follows. Let $\alpha_1 = \varepsilon_x / f$, $\alpha_2 = -\varepsilon_y / f$, and let $\mathbf{M}^x_{\alpha_2}$ be a rotation of angle α_2 centred at the projection centre O and around the direction of the object X axis, and $\mathbf{M}^y_{\alpha_1}$ an analogous rotation around the direction of the object Y axis. Whether the numerical values of the angles are α_1 and α_2 or $-\alpha_1$ and $-\alpha_2$ depends upon the choice of signs for the rotations. The matrices obviously remain the same, and have to be taken so that the displacement of the principal point due to the rotations is $(\varepsilon_x, \varepsilon_y)$. The new rotation matrix $\mathbf{M}'$ transforming from the object coordinate system to the image system as a function of the former matrix $\mathbf{M}$ is

$$\mathbf{M}' = \mathbf{M}^{y}_{\alpha_1}\mathbf{M}^{x}_{\alpha_2}\mathbf{M}$$

in the first order of approximation. The exact formula is somewhat more involved but given that the displacements $(\varepsilon_x, \varepsilon_y)$ are small it is not necessary. The new values for the parameters Ω, Φ, κ will be derived from $\mathbf{M}'$. Their relation to the original ones and the angles α_1 and α_2 depends on the expression relating the three parameters with the rotation matrix. If they are the usual primary, secondary and tertiary rotations the changes they undergo are

$$\Delta\Omega = \alpha_2' \sec\Omega, \qquad \Delta\Phi = \alpha_1', \qquad \Delta\kappa = \alpha_2' \tan\Omega,$$

where

$$\begin{aligned} \alpha_1' &= \alpha_1 \cos\kappa - \alpha_2 \sin\kappa \\ \alpha_2' &= \alpha_1 \sin\kappa + \alpha_2 \cos\kappa \end{aligned}$$

GENERAL DESIGNING OF THE BASE

Removal of orientation-equivalent components

If we collect up the main results from the previous section we have

$$\begin{aligned} f &\rightarrow \Delta r = ar, \\ \kappa &\rightarrow \Delta t = ar, \\ \Omega, \Phi, \kappa, x_p, y_p &\rightarrow \Delta r = ar^2 \cos\theta + br^2 \sin\theta, \end{aligned} \tag{4}$$

where the different a's are naturally not the same.

Expressions (4) mean that for any two possible decompositions $(T_1,\mathcal{D}_1)$ and $(T_1,\mathcal{D}_1)$ with $\mathcal{D}(0,0) = (0,0)$ the difference $\mathcal{D}_2 - \mathcal{D}_1$ will be expressible as a combination of those four components, i.e.

$$\begin{aligned} (\mathcal{D}_2 - \mathcal{D}_1)_r &= ar + br^2 \cos\theta + cr^2 \sin\theta, \\ (\mathcal{D}_2 - \mathcal{D}_1)_t &= dr. \end{aligned}$$

Let $\mathcal{B}$ be a base,

$$\mathcal{B} = \mathcal{B}_r \cup \mathcal{B}_t = \{g_{r1}, g_{r2}, \ldots, g_{t1}, g_{t2}, \ldots\}.$$

The expression of any distortion function $\mathcal{D}$ by means of the base $\mathcal{B}$ will be of the form

$$\begin{aligned} \mathcal{D}_r &= a_1 g_{r1} + a_2 g_{r2} + \ldots, \\ \mathcal{D}_t &= b_1 g_{t1} + b_2 g_{t2} + \ldots \end{aligned}$$

Let $\mathcal{D}_1$ and $\mathcal{D}_2$ be as above and write $\mathcal{D}_1 = \sum a_{i1} g_{ri} + \sum b_{i1} g_{ti}$ and $\mathcal{D}_2$ accordingly. If we let $g_{r1} = r$, $g_{r2} = r^2 \cos\theta$, $g_{r3} = r^2 \sin\theta$, and $g_{t1} = r$, then $a_{i1} = a_{i2}$ and $b_{i1} = b_{i2}$ will be satisfied for all coefficients with exception of a_1, a_2, a_3 and b_1; that is, if the projection T is varied within its possible values, the $\mathcal{D}$ from $(T,\mathcal{D})$ will vary in such a way that if we let $\mathcal{D} \in \langle\mathcal{B}\rangle$ all its coefficients will remain constant with exception of a_1, a_2, a_3 and b_1.

If T is varied within the space of possible solutions there will be a particular point at which a_1, a_2, a_3 and b_1 vanish. Let $\mathcal{B}_{rT} = \{r, r^2\cos\theta, r^2\sin\theta\}$, $\mathcal{B}_{tT} = \{r\}$ and $\mathcal{B}_T = \mathcal{B}_{rT} \cup \mathcal{B}_{tT}$, and let $\mathcal{B} = \mathcal{B}_T \cup \mathcal{B}^*$;that is, we write $\mathcal{B}$ as the union of orientation-equivalent components and other components. Now let $\mathcal{B}$ be designed so that every component of $\mathcal{B}^*$ is orthogonal in the photograph space to those of $\mathcal{B}_T$. Then, according to the properties of orthogonal functions, the previous particular point will have the least $\mathcal{D}$ of all possible distortions that can be assigned to the image, where the size is measured by the mean quadratic value, as explained at the end of the section “The mathematical setting”.

The restricted base $\mathcal{B}^*$ is the one to be used in calibration processes, providing simultaneously uniqueness and best possible solution.

Choice of the orthogonal bases

Even if orthogonality is only required among $\mathcal{B}_T$ and $\mathcal{B}^*$, it is still interesting that all functions be orthogonal among each other, due to many advantages of orthogonal bases that will be discussed later. Any radial distortion will be orthogonal to any tangential one, so orthogonality has only to be achieved within the radial and tangential components.

Distortion components can be classified as symmetric and asymmetric ones. Symmetric ones are those that do not depend on θ. Given that an r component is present in both $\mathcal{B}_{rT}$ and $\mathcal{B}_{tT}$, a polynomial base seems the natural solution for the symmetric distortions. Orthogonal polynomials may be derived recursively; in order to find the coefficients of the k-th polynomial a linear system with $k-1$ unknowns has to be solved, but the systems themselves may be solved recursively.

The obtained coefficients will be different for different r_{max}, i.e., the maximum possible value of r. For this and other reasons it is much convenient to normalise the values of r as r/r_{max}, that is, formulae should be entered with the value $s = r/r_{max}$ rather than with r itself. Therefore we have $s \in [0,1]$.

The symmetric radial distortion has been traditionally expressed as an odd polynomial. This guarantees infinite derivability at the origin. However, I see no reason why it should be imposed that the distortion function be infinitely derivable at the origin. I find “smoothness” enough, and smoothness means continuity of the first derivative. I recall that there are cases when even a linear interpolation is performed between given points. So let smoothness be the condition to be satisfied by the functions of the base, in addition to $\mathcal{D}(0,0) = (0,0)$.

But odd polynomials have also some advantages. The model being developed will, in both cases, be capable of expressing any distortion that may exist; in particular, any distortion expressed by some other model shall be also expressible by this model by finding the values for the coefficients that yield the same total distortion than the one we are given. If the latter makes use of odd polynomials, let us suppose to fix our ideas that it includes r, r^3 and r^5 components. In order to express that distortion by a linear combination of the components of a base built up of complete polynomials five components will be necessary, while if the base is composed of odd polynomials three components will suffice. Odd polynomials are also endowed with the mathematical perfection that is so pleasant to see in whatever model.

For all the above it was finally decided to derive two bases: the complete model (CM) and the odd model (OM). The constant term will be zero for all the polynomials. The degree of the k-th polynomial will be equal to k in the CM and to $2k-1$ in the OM. The slower growth of the degree in the CM has numerical advantages, but these days they are negligible. On the other hand it turned out that the coefficients of the OM increase slower.

The asymmetric components

With regard to the non symmetric components, orthogonality will exist among any two components g_1, g_2 if, for instance, for any radius r the condition $\int_0^{2\pi} g_1 g_2 \,\mathrm{d}\theta = 0$ is satisfied. This is not exact for a rectangular shaped domain, for in that case θ cannot take any value for high values of r. We shall come to this later but for the moment we will take that condition as being true.

It is well known that in the interval $[0,2\pi]$ both $\int \cos m\theta \cos n\theta \,\mathrm{d}\theta$ and $\int \sin m\theta \sin n\theta \,\mathrm{d}\theta$ are zero whenever m and n are different, and $\int \cos m\theta \sin n\theta \,\mathrm{d}\theta = 0$ always. I will not repeat here the theory about orthogonal functions, and I will simply state that the simplest solution for the asymmetric part of the base is

$$\{p_k \cos c\theta, p_k \sin c\theta\},$$

where p_k are the polynomials from the symmetric base. In order to simplify the notation, I shall let S_c represent both $\cos c\theta$ and $\sin c\theta$, and if h is any function, by hS_c I shall denote the pair $\{h \cos c\theta, h \sin c\theta\}$.

This base does not conform to our requirement that its components be smooth. An analysis of these functions reveals that there cannot exist an r component if c is odd. For the tangential components this is made apparent by straight lines passing through the origin. These components would generate an angle at the origin. The p_k polynomials have to be replaced by some other polynomials q_k that do not include a first degree term; that is to say, a base has to be built starting from r^2. This does not break orthogonality between even c and odd c components, for if c and d are two such values, and let h_1, h_2 be two functions of r, for any fixed r we have

$$\int_0^{2\pi} h_1(r) S_c \, h_1(r) S_d \,\mathrm{d}\theta = h_1(r) h_2(r) \int_0^{2\pi} S_c S_d \,\mathrm{d}\theta,$$

and so h_1 and h_2 may be any functions.

We have $q_1 = r^2$, which implies that the base includes the components $r^2 \cos\theta$ and $r^2 \sin\theta$ that are needed for $\mathcal{B}_{rT}$. In the CM the q polynomials will include all the monomials starting from r^2, in the OM they will be even polynomials (when combined with the odd S_c term it results in an odd component).

The mean value of any asymmetric component for any fixed r is zero; it need be, for otherwise then the component would not be orthogonal to some p_k (and possibly to all). Let $\mathcal{D} \in \langle \mathcal{B} \rangle$, this property implies that the value of the symmetric part of $\mathcal{D}$ is at every r equal to the mean value of the distortion for that value of r all around the photograph, which makes sense.

THE FORMULAE

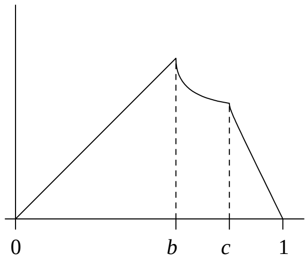
<u>*Derivation of the polynomial base*</u>

A base was first derived in the supposition that the domain of (s,θ) pairs be the rectangle $[0,1]\times[0,2\pi]$. The coefficient a_i of the k-th polynomial can be expressed for both the complete and odd models by a closed form involving factorials depending on k and i.

But these polynomials satisfy orthogonality for a faked domain which does not correspond to the shape of a photograph. In a real photograph there is “more” or “less” photograph for different values of s. This is represented by a weight function. If the photograph is a rectangle with its sides in a ratio b/c, such that $b \le c$, and the values of b and c are so normalised that $b^2 + c^2 = 1$, the weight function is that of Fig. 4 and its expression

$$w(s) = \begin{cases} \frac{\pi}{2} s, & 0 \le s < c \\ s \operatorname{asin} c/s, & c \le s < b \\ s(\operatorname{asin} c/s - \operatorname{acos} c/s), & b \le s < 1 \end{cases}$$

Fig. 4. Weight function for a rectangular photograph



The condition that must be satisfied by two polynomials h_1, h_2 from the base is

$$\int_0^1 w h_1 h_2 \,\mathrm{d}s = 0.$$

Several solutions may be adopted. One of them consists in dividing each p_k and q_k by $\sqrt{w}$. This however is not possible because the functions from $\mathcal{B}_T$ cannot be modified. Since the most important condition regarding orthogonality is that the functions from $\mathcal{B}^*$ be orthogonal to those from $\mathcal{B}_T$, we may divide each one from $\mathcal{B}^*$ by w, at the price of losing orthogonality within it. Apart from the loss of orthogonality, this approach has the disadvantage of implying $g(1,\theta) = 0$ for every g in $\mathcal{B}^*$. The drawbacks of this restriction will be discussed when analysing trigonometrical bases.

The above solutions being discarded, the one which remains and which is usually the best approach with respect to a weight function, is to derive the polynomials anew taking into account the weight function. In order to do so it is necessary to evaluate the integrals of the powers of s times the weight function. The details will be omitted here. The author will be pleased to provide them to whomever asks for them. Let b and c denote the sides of the rectangle, so normalised that $\sqrt{(b^2 + c^2)} = 1$. For $n = 0$, i.e. the 0'th power of s, the result is

$$\int w\,\mathrm{d}r = \int \mathrm{d}s = S = bc.$$

For even n the first values of $\int wr^n\,\mathrm{d}r / bc$ are

n	$\int / bc$
2	1/3
4	$1/5 - 4/5(bc)^2$
6	$1/7 - 8/35(bc)^2$

The expressions for odd n have an extra term which is not rational in b and c.

The values of these integrals an hence the coefficients of the polynomials arising thereof depend upon the ratio b/c. The author derived the polynomials for ratios ranging from 1/1 to 9/5. Since the polynomials vary little I decided to take a particular ratio for the definition of the polynomials. For the CM I chose the ratio 8/7 for the p_k series and 3/2 for the q_k ones, for these are the simple ratios that yield the values of the coefficients of the second polynomials 3, –2 and 4, –3 respectively, up to two decimal places (and in the first case up to three indeed). For the OM the chosen ratio was $\sqrt{3}/1$ for both series, because for this value the second polynomials are exactly $p_2 = 2s^3 - s$ and $q_2 = 2{\cdot}5s^4 - 1{\cdot}5s^2$.

Here the polynomials are shown for both models. To the right of each of them its quadratic mean over the photograph is displayed. This mean varies with the image shape too, but it varies very little. The displayed values correspond to a ratio of 4/3 for the CM and $\sqrt{3}/1$ for the OM.

p_k polynomial series for the CM.

name	*polynomial*	$\lVert p_k \rVert$
p_1	s	0·58
p_2	$3s^2 - 2s$	0·28
p_3	$9s^3 - 11{\cdot}4s^2 + 3{\cdot}4s$	0·17
p_4	$29{\cdot}2s^4 - 53{\cdot}1s^3 + 30{\cdot}1s^2 - 5{\cdot}2s$	0·13
p_5	$95{\cdot}8s^5 - 225{\cdot}4s^4 + 187{\cdot}1s^3 - 63{\cdot}9s^2 + 7{\cdot}4s$	0·096
p_6	$320{\cdot}3s^6 - 922{\cdot}1s^5 + 1004{\cdot}9s^4 - 511{\cdot}4s^3 + 119{\cdot}2s^2 - 9{\cdot}9s$	0·077

q_k polynomial series for the CM.

name	*polynomial*	$\lVert q_k \rVert$
q_1	s^2	0·40
q_2	$4s^3 - 3s^2$	0·22
q_3	$14{\cdot}5s^4 - 20{\cdot}3s^3 + 6{\cdot}8s^2$	0·14
q_4	$53{\cdot}5s^5 - 107{\cdot}8s^4 + 69{\cdot}5s^3 - 14{\cdot}2s^2$	0·10
q_5	$197{\cdot}5s^6 - 511{\cdot}4s^5 + 476{\cdot}9s^4 - 188{\cdot}2s^3 + 26{\cdot}2s^2$	0·082

p_k polynomial series for the OM.

name	*polynomial*	$\|\| p_k \|\|$
p_1	s	0·58
p_2	$2s^2 - s$	0·26
p_3	$4{\cdot}8s^5 - 4{\cdot}7s^3 + 0{\cdot}9s$	0·14
p_4	$12{\cdot}8s^7 - 19{\cdot}1s^5 + 8{\cdot}2s^3 - 0{\cdot}9s$	0·093
p_5	$38{\cdot}4s^9 - 76{\cdot}2s^7 + 50{\cdot}5s^5 - 12{\cdot}6s^3 + 0{\cdot}9s$	0·069
p_6	$119{\cdot}5s^{11} - 296{\cdot}7s^9 + 268s^7 - 106{\cdot}5s^5 + 17{\cdot}6s^3 - 0{\cdot}9s$	0·055

q_k polynomial series for the OM.

name	*polynomial*	$\|\| q_k \|\|$
q_1	s^2	0·41
q_2	$2{\cdot}5s^4 - 1{\cdot}5s^2$	0·20
q_3	$6{\cdot}4s^6 - 7{\cdot}2s^4 + 1{\cdot}8s^2$	0·11
q_4	$19{\cdot}1s^8 - 31{\cdot}6s^6 + 15{\cdot}7s^4 - 2{\cdot}2s^2$	0·080
q_5	$60{\cdot}4s^{10} - 131s^8 + 97{\cdot}8s^6 - 28{\cdot}9s^4 + 2{\cdot}7s^2$	0·061

Figures 5–8 displays the first polynomials for both models.

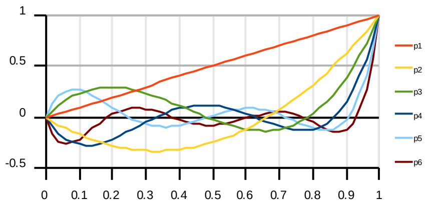


Fig. 5. p_k series for the CM

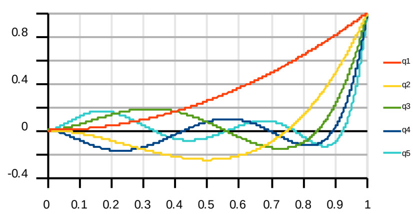


Fig. 6. q_k series for the CM

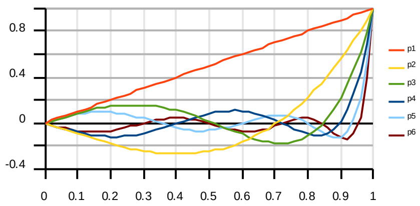


Fig. 7. p_k series for the OM

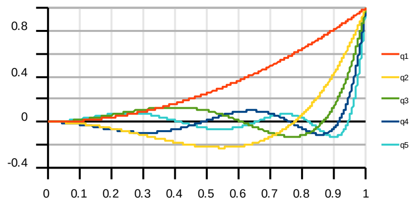


Fig. 8. q_k series for the OM

The distortion function

It is as follows:

Symmetric components:

$\mathcal{D}_r = a_2p_2 + a_3p_3 + a_4p_4 + \ldots$
$\mathcal{D}_t = b_2p_2 + b_3p_3 + b_4p_4 + \ldots$

Asymmetric components:

$$\mathcal{D}_r = c_3q_2\cos\theta + c_4q_2\sin\theta + c_5p_1\cos 2\theta + c_6p_1\sin 2\theta + c_7q_3\cos\theta + c_8q_3\sin\theta + c_9p_2\cos 2\theta + c_{10}p_2\sin 2\theta + c_{11}q_1\cos 3\theta + c_{12}q_1\sin 3\theta + \ldots$$

$$\mathcal{D}_t = d_1q_1\cos\theta + d_2q_1\sin\theta + d_3q_2\cos\theta + d_4q_2\sin\theta + d_5p_1\cos 2\theta + d_6p_1\sin 2\theta + \ldots$$

Thus, except for the components $ar^2\cos\theta$ and $r^2\sin\theta$ of $\mathcal{D}_{\mathrm{t}}$ the development is the same for the radial and tangential distortions. Those components in the radial distortion as well as a_1p_1 for the symmetric part of both may appear when a given distortion expressed by some other model is expressed by this model.

Alternative asymmetric decomposition

For each pair k, c each set of four parameters p_kS_c, if c is even, or q_kS_c if c is odd, may be replaced by four other parameters which, when taken together, are equivalent to the previous four. Let these components be written as

$$p_k\mathbf{u}_{-c}\,,\quad p_k\mathbf{v}_{-c}\,,\quad p_k\mathbf{u}_{+c}\,,\quad p_k\mathbf{v}_{+c}\,,$$

or q_k if c is odd.

The components $\mathbf{u}_{-c}$ ... $\mathbf{v}_{+c}$ are vectors of constant modulus equal to 1. The vector $\mathbf{u}_{-c}$ forms at a point with coordinates (r, θ) an angle with the radial direction equal to $-c\theta$, and the vector $\mathbf{u}_{+c}$ and angle equal to $c\theta$. The vectors $\mathbf{v}_{-c}$, $\mathbf{v}_{+c}$ form the same angles with respect to the tangential direction, and so they are perpendicular to $\mathbf{u}_{-c}$ and $\mathbf{u}_{+c}$ respectively. Therefore at a point on the positive side of the x axis the $\mathbf{u}$ vectors follow the axis direction while the $\mathbf{v}$ vectors are parallel to the y axis and directed toward the positive side. As we move along a circle centred at the origin the vectors rotate clockwise/counterclockwise ($-c$/$+c$), performing a total of c revolutions with respect to the radius when the circle is complete. I will hence call this the model of the rotating vector. With respect to a fixed direction in the x, y plane these vectors perform $1-c$ and $1+c$ rotations as they turn around the origin.

It should be noted that if $\alpha_{k,c}$, $\beta_{k,c}$ are the coefficients corresponding to $p_k\mathbf{u}_{-c}$ and $p_k\mathbf{v}_{-c}$ the composition $\alpha_{k,c}\, p_k\mathbf{u}_{-c} + \beta_{k,c}\, p_k\mathbf{v}_{-c}$ is a single vector with constant modulus that performs the c revolutions as it turns around the origin, the modulus of which as well as its direction and sense for $\theta = 0$ are determined by $\alpha_{k,c}$ and $\beta_{k,c}$, and the same is true for the $+c$ components. The coefficients corresponding to the $+c$ components I will call γ and δ.

The radial and tangential distortions generated by these components are, apart from the multiplier p_k or q_k and omitting the subindices,

$$\mathcal{D}_r = \alpha\cos(-c\theta) - \beta\sin(-c\theta) + \gamma\cos c\theta - \delta\sin c\theta,$$
$$\mathcal{D}_t = \alpha\sin(-c\theta) + \beta\cos(-c\theta) + \gamma\sin c\theta + \delta\cos c\theta;$$

and the components x and y,

$$\mathcal{D}_x = \alpha\cos(1-c)\theta - \beta\sin(1-c)\theta + \gamma\cos(1+c)\theta - \delta\sin(1+c)\theta,$$
$$\mathcal{D}_y = \alpha\sin(1-c)\theta + \beta\cos(1-c)\theta + \gamma\sin(1+c)\theta + \delta\cos(1+c)\theta.$$

If a set of four components of the radial/tangential model is

$$\mathcal{D}_r = ap_k\cos c'\theta + bp_k\sin c'\theta,$$
$$\mathcal{D}_t = cp_k\cos c'\theta + dp_k\sin c'\theta,$$

(I have written c' to distinguish it from the c coefficient) the relation between the two sets of four parameters is

$$a = \alpha + \gamma \qquad b = \beta - \delta \qquad\qquad \alpha = (a-d)/2 \qquad \beta = (b+c)/2$$
$$c = \beta + \delta \qquad d = -(\alpha - \gamma) \qquad\qquad \gamma = (a+d)/2 \qquad \delta = -(b-c)/2$$

If the model of the rotating vector is used an exception has to be observed at the $(k, c) = (1, 1)$ components, for we have seen that the components r^2S_c of the radial distortion need to be present in the base so that they can be omitted from the distortion function. This is equivalent to imposing $\alpha = -\gamma$, $\beta = \delta$ for those components, viz. $\alpha_{1,1} = -\gamma_{1,1}$, $\beta_{1,1} = \delta_{1,1}$.

It is easily seen that $\alpha_{1,2}$ and $\beta_{1,2}$ are the parameters of an affinity deformation: a difference in the x and y scales and a missorthogonality of the axes (i.e. the pixel grid). Thus, the use of the rotating vector may be of interest for the $(1, 2)$ components if we want to include just the affinity distortion or one of its two components.

Orthogonality of the asymmetric components

There remains a problem with respect to the asymmetric components. On a square photograph, in the interval $1/\sqrt{2} \le s \le 1$ the integral to be considered is not $\int_0^{2\pi}$ but $\int_{L(s)}$, where $L(s)$ is the union of four equal intervals centred at $\pi/4$, $3\pi/4$, $5\pi/4$ and $7\pi/4$, whose length diminishes as s grows. Several solutions were considered, but none of them resulted in simple expressions.

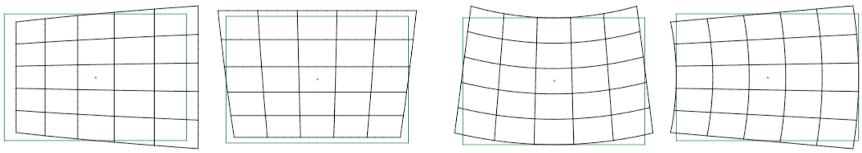
On the other hand, the integral of the product of any two functions in the circle $0 \le s \le 1/\sqrt{2}$ will always be zero, and that represents roughly ¾ of the photograph. As of the area $1/\sqrt{2} \le s \le 1$, due to the many symmetries of trigonometrical functions: (i) orthogonality is retained with respect to the symmetric components, (ii) orthogonality is retained between cos and sin components and (iii) orthogonality is retained between any two S_c and S_d if c and d are of different parity. This properties imply that the first loss of orthogonality does not appear till 3θ, and the first loss between a component of $\mathcal{B}_T$ and one of $\mathcal{B}^*$ does not till 4θ, where I, for instance, have never got to.

All of the above properties are still satisfied within a rectangular photograph, the only difference being that the circle where the circumferences are complete represents a lower proportion of the total image area.

Hence, for the sake of simplicity and for practical reasons, I judged we had better not perform any modification in this respect, and hereupon my development of the base finishes.

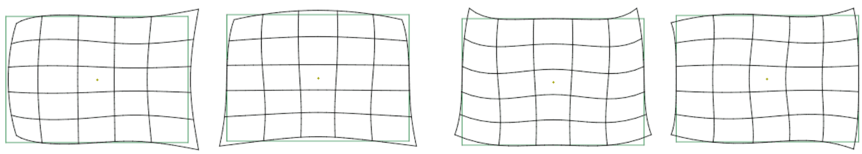
The graphs of the first asymmetric components

Fig. 9–12. c1,c2; d1,d2 components.

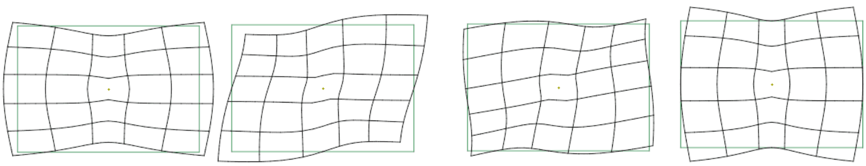


The components c1, c2 will normally not exist since they vanish if the principal point is calibrated.

Figs. 13–16. c3,c4; d3,d4 components.



Figs. 17–20. c5,c6; d5,d6 components.



About the coefficients

In this section, two decisions that were made in the design of the base will be discussed.

The first one deals with the number of figures of the coefficients. In order for the model to be usable the polynomials have to be unambiguously defined. This is why its coefficients were rounded to the first decimal place. But care should be taken when rounding polynomial coefficients. One could be tempted to suppose that the precision of any coefficient is to be regarded with respect to itself, that is, to pay attention to the number of supplied figures rather than to the position of the last one. But this is wrong. The coefficients of individual monomials within a polynomial may be very high and yet the polynomial take small values. Rounding errors must always be considered in absolute value, no matter how high the coefficient be, lest significant differences between the modified polynomial and the original one arise. Rounding to an integer or to half an integer is not acceptable. It is also absolutely necessary that the sum of the coefficients not be varied, viz. that it be kept equal to 1.

The other decision to be commented here is the one of designing the polynomials so as to make $\max|p_k| = p_k(1) = 1$, and the same is true for q_k. The alternative to be considered is that of multiplying the polynomials by a constant so that $(\int_F g)/bc$, i.e., the mean quadratic value, equals 1 for every $g \in \mathcal{B}$. If that condition is satisfied the functions are said to be normalised. Both possibilities have advantages and disadvantages, and what finally made me decide about the former was the fact that its coefficients are simpler.

ANALYSIS OF OTHER MODELS

Trigonometrical base for s

The trigonometrical base is possibly the easiest one to handle. There are two versions:

$$\{1, \sin 2\pi ns, \cos 2\pi ns\}$$
$$\{\sin \pi ns\}$$

The first one has terms that are different from zero at $s = 0$. If the infinite components could be taken this would not be a problem, but that is impossible. Those terms cannot be simply eliminated, since the resulting base would be restricted to very particular functions.

The second one has not this problem but it has the one that every component g has $g(1) = 0$. The infinite sum $\sum a_k g_k$ would be equal to the symmetric part of $\mathcal{D}_r$, which I shall call $\mathcal{D}_{rs}$, in [0,1) and equal to zero at s = 1 ($r = r_{\max}$), and similarly for $\mathcal{D}_{ts}$. Any finite sum will approximate very bad the last part of $\mathcal{D}_{rs}$, for it will always be a continuous function f satisfying $f(1) = 0$. Imposing $\mathcal{D}_s(1) = 0$ is not acceptable. Consider for instance the very common case where $\mathcal{D}_{rs}$ is always concave or always convex.

Before pointing a possible solution to this problem I will first state another one. This base is complete, so it can in particular be used to approximate s in [0,1). Hence, the more terms we take the more correlated the distortion function will be to the focal length. The most correlated of all terms is the first one. Now, the solution to the previous problem is to add the function s to the base, which worsens this problem.

A possible solution to both problems consists in modifying each $\sin \pi ns$ by adding a suitable multiple of s so that the resulting function be orthogonal to s. The first components thus created are

$$\begin{array}{ll} \sin \pi s - 1{\cdot}28s, & \sin 2\pi s + 0{\cdot}66s, \\ \sin 3\pi s - 0{\cdot}55s, & \sin 4\pi s - 0{\cdot}12s. \end{array} \tag{5}$$

Some of the properties of the trigonometrical base are lost, in particular orthogonality among its components, but this property had never hold due to the shape of the photograph. For it to hold it would be necessary that, in addition to the s term, $\sin m\pi s$ terms be added for all $m < n$. The formulae from (5) is not a bad solution at all, but it is not clear that it be simpler than the polynomial base.

The displayed components are the analogous of p_k, starting at k = 2; the ones replacing the q_k need also be obtained. On the one hand, they need to be orthogonal to s^2; on the other, they need to have zero derivative at the origin. These conditions can be satisfied if they are defined as $s(\sin n\pi s + c_n s)$, and again the simplicity of the original trigonometrical base is lost, as well as orthogonality among its components.

These problems with the trigonometrical base show that the requirement that the components p_k and q_k be orthogonal to s and s^2 respectively leads naturally to a polynomial base for both of them.

Odd polynomial for the symmetric radial distortion

The traditional model for the symmetric radial distortion is an odd polynomial whose coefficients are the parameters of the model. I have already pronounced myself about the restriction to odd terms (cf. section *Choice of the orthogonal bases*). It could be also argued against this restriction that parabolic like distortions will not be properly represented by this model. However, strange as it may be, odd polynomials can produce arbitrarily sharp approximations to an even polynomial within an interval; but surely a quadratic term would be better.

An important disadvantage arises if r is expressed in its original units, as is usually the case. This causes the powers of r to raise to very high values near the corners of the photograph or even at a mean distance, thereby the coefficients being very small (e.g. 10^{-14}), providing a faked appearance of being negligible. This is not actually a deficiency of the model itself, but of the way it is usually applied.

The most severe problem of this model is the extreme correlation among its components. A very small change in the distortion function may therefore cause a big change in the parameters. Another consequence is that the removal of the highest degree term requires all the other coefficients to be computed again. Another one is that the presence or absence of r term does not mean a presence or absence of linear tendency (and this in turn implies that the optimum solution, though not difficult to obtain, requires some more calculations). Yet another one is the fact that, even if we use the variable $s = r/r_{\max}$, the absolute value of the coefficients is by no means representative of the magnitude of the photograph distortions. For example, let $\mathcal{D}_{rs} = -24s + 97s^3 - 80s^5$, the maximum value of $\mathcal{D}_{rs}$ is not near 100 as it may seem, but equals 8; and that of $50s^3 - 130s^5 + 80s^7$ is 3. Finally, the high correlation may cause computation problems.

All these problems disappear with the designed base. Individual terms may be eliminated without the need to recompute the other ones, and the coefficients together with the given values of $\|p_k\|$ do represent the magnitude of the distortion. The first of the previous examples equals $\mathcal{D}_{rs} = 0{\cdot}3p_1 + 9{\cdot}3p_2 - 16{\cdot}7p_3$ (OM), whereby we find $\|\mathcal{D}_{rs}\| = \left((0{\cdot}3\|p_1\|)^2 + (9{\cdot}3\|p_2\|)^2 + (16{\cdot}7\|p_3\|)^2)\right)^{1/2} = 5{\cdot}1$ and $\max|\mathcal{D}_{rs}|=3{\cdot}4$.

Decentring distortion

It is a biparametric model:

$$\mathcal{D}_x = P_1(r^2 + 2x^2) + 2P_2xy,$$
$$\mathcal{D}_y = 2P_1xy + P_2(r^2 + 2x^2).$$

I looked for a justification of these components in some photogrammetry manuals, but I found none. It seems that manuals copy each other and all rely on an article of D. Brown.

The radial and tangential decomposition of this distortion is

$$\mathcal{D}_r = 3P_1\mathrm{r}^2 \cos\theta + 3P_2r^2 \sin\theta,$$
$$\mathcal{D}_t = -P_1\mathrm{r}^2 \sin\theta + P_2r^2 \cos\theta\,.$$

The radial components are the ones equivalent to an inclination of the plane π, that should not exist. Whenever an image is presented as having a decentring distortion, this may be split into radial and tangential components as shown, the radial part be simply dropped and the principal point be changed according to (3); and if the other asymmetric radial components of the model, if any, are orthogonal to the dropped ones, or if there are no more as is usually the case, and whatever the symmetric radial model be, the mean quadratic value of the image distortion will decrease, always.

The fact that a certain physical phenomenon causes a certain change in the image coordinates does not mean that such change has to be understood as a distortion. If for example a perfect camera be supposed with projection T_1; if certain rotation is applied to it a new perfect projection T_2 will result, that is not to be interpreted as projection T_1 plus a distortion, but as the perfect projection T_2. In the like manner, would we take a pair $(T_1,\mathcal{D}_1)$ as the solution if there exists another one $(T_2,\mathcal{D}_2)$ with smaller distortions, being the difference in distortions precisely the difference between T_1 and T_2? Certainly not. What should be done is to take as the reference projection T that which better fits the image.

The subject is often better understood if the camera physical reality is abstracted, as I have done throughout this work. If the coordinate variation due to some actual cause is derived, it should be studied whether the variation can be expressed as the composition of a change in the projection and an actual distortion. It is only necessary to write the coordinate variation as radial and tangential components and see if any of them belongs to $\mathcal{B}_T$.[a]

Affinity

A small affinity may be written as the union of radial and tangential distortions as

$$\mathcal{D}_r = \alpha\, r \cos 2\theta + \beta\, r \sin 2\theta,$$
$$\mathcal{D}_t = -\alpha\, r \sin 2\theta + \beta\, r \cos 2\theta.$$

[a] It may not be as easy as that, for terms may appear that are not equal to any from $\mathcal{B}_T$ and yet they are not orthogonal to them. However, it comes to pass that real terms can always be easily decomposed in a part included in $\langle\mathcal{B}_T\rangle$ and a part orthogonal to it.

It hence represents half of the p_1S_2 distortions. Its orthogonal complement within these distortions is

$$\mathcal{D}_r = \gamma\, r \cos 2\theta + \delta\, r \sin 2\theta,$$
$$\mathcal{D}_t = \gamma\, r \sin 2\theta - \delta\, r \cos 2\theta.$$

If we want the parameters α and β to explicitly appear in the model, the four p_1S_2 parameters should be replaced by $\{\alpha, \beta, \gamma, \delta\}$, which are those of the rotating vector model, as explained above.

Direct linear transformation (DLT)

The eleven parameters of a DLT are a mere recombination of the nine orientation ones and the two affinity ones, with the important drawback that if several images are taken, a different set of parameters will be computed for each one, including those that will not vary, namely: f, x_p, y_p, α and β.

It is not rare to read about the DLT that "it is suitable for any camera" and that "it is not necessary to know the focal length". These sentences are nonsense. Furthermore, the DLT is specially unsuited for large distortion cameras, for it only includes an affinity correction. With respect to the second one, I have already pointed that the eleven DLT parameters are *exactly* a central projection plus an affinity distortion, which certainly includes a focal length.

TESTING AND CONCLUSIONS

General

Since the development of the base in 2005 it has been applied to many different cameras, ranging from an amateur 3 Mpx camera to the self-calibration of aerial metric cameras. And the number of points measured for the calibration also varied greatly with a maximum of 16 000 for the calibration of a single photograph and 4 500 photographs for self-calibration. The computations were performed with a program written by the author on the occasion of the development of the model. The author also incorporated self-calibration to his aerotriangulation program, *Aerotri*, in the late 2009, thereby allowing the possibility of applying the model to self-calibration. Some of the drawn conclusions relate to the process of calibrating a camera rather than to the application of this model and will be passed over without mention.

Some points had been foreseen and the tests served to confirm them:

1. The model proved suitable for all the cameras to which it was applied. If few parameters can be determined, either because of few measured points or a low stability, few parameters from the model will be included; if there are many observations and the camera is stable more parameters can be determined. In particular, low-cost cameras with a fixed focal length have distortions which are stable enough to be modelled well below the pixel size.

2. If there is few data the computed distortion components fit the actual errors of the observations rather than the distortions. This will happen with any model, but is of importance in comparing this model versus the grid one, as will be seen shortly.

Some others had not been foreseen:

3. Even with thousands of measured points the distortions are remarkably well modelled with a few parameters. The next example will provide a remarkable exhibition of this.
4. Radial components are clearly greater than the tangential ones even for high degree components, when in principle the irregularities of the pixel grid were expected to nullify the radial preponderance. Therefore the main cause of distortion even well below the pixel size is still due to the lenses and the radial/tangential model is preferred over that of the rotating vector, for it will need less components in order to model the camera's distortion with equal accuracy. An affinity component due to the difference in the *x* and *y* scales is some times an exception to this.
5. The complete model was expected to behave better than the odd model. This is usually the case, for the most common situation is that both models yield the same quadratic mean of the residuals but the distortion according to the complete model is smaller, that is, the complete model fits better the camera distortions. There are cases however when the odd model fits better, and the difference may be substantial. This happens when the symmetric radial distortion (the greater by far, with the possible exception of the affinity component mentioned above), instead of having a parabolic-like shape increases abruptly at the edge and corners of the photograph. In this case the zero of the distortion function, and in general the detail in the function is better placer at a more extreme position, and the odd model is more suitable for this as can be seen at the graphs of their components.

   So instead of leaving just one model, both models were kept and for each camera the calibration is carried out with both of them and the solution which seems better is retained. It often happens that the difference is not significant.

A test with sixteen thousand points

The description of tests with few observations —up to a few hundred for calibrations of a single image in laboratory or a few thousand with self-calibration— would add little to the conclusions listed above. Regarding calibrations with a great amount of data, I think it is more interesting to give a detailed description of the most careful calibration carried out than to give a summarised account of the different cameras and settings to which this model was applied.

The camera is a Canon of 4752 × 3168 megapixels. It was calibrated for its use in creating a 3D model of the Discobolus when this statue was in Spain in the year 2009. The professors at the School of Topography of the Polytechnic

University of Madrid set up a calibration room by placing ten plotted sheets on a wall with 2106 marks on each making a total of 21060. Of these, 16100 fell within the photograph limits. In addition to this, some rows of small balls set on wires were placed in front of the wall for the correct determination of the focal length and the principal point. These have been omitted from the graphs of the residuals that follow in order not to disturb the appreciation of regular patterns that arise. The marks were correlated by a Matlab program written by the professors at the school. Incomplete rows or columns could not be correlated and were measured by hand. That break is clearly visible on the graphs of the residuals when the distortions are eliminated.

Below the residuals of the marks after different adjustments are shown. The first one is the result of computing the camera orientation without the inclusion of any distortion parameter. Actually, the camera exhibited a great difference in the *x* and *y* scales. I corrected this affinity component as part of the transformation from raw pixel coordinates to coordinates centred at the principal point so that the radial distortion could be appreciated on the graphs of the residuals, that would otherwise be outweighed by the affinity component. This first graph shows clearly the prevailing symmetric radial distortion. Note the position of the zero at a high distance. This is one of the cases mentioned in conclusion 5 above. Therefore the OM models this camera's distortions better than the CM, and the next graphs correspond to the OM.

Fig. 21. Residuals: no distortion corrected.

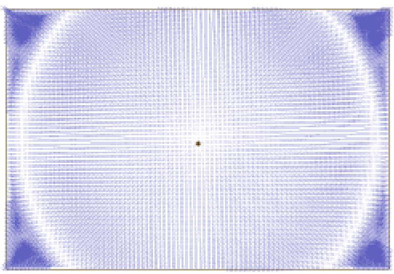


Fig. 22. Residuals: a2 corrected.

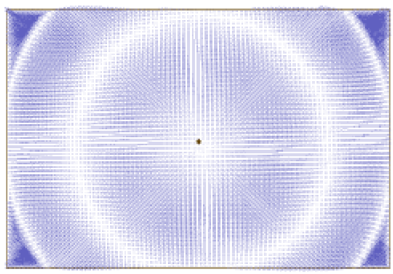


Fig. 23. Residuals: a_3 corrected.

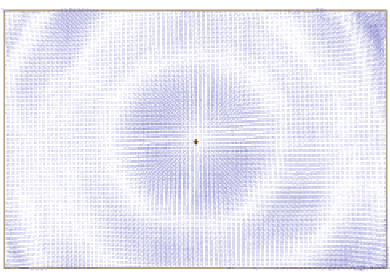
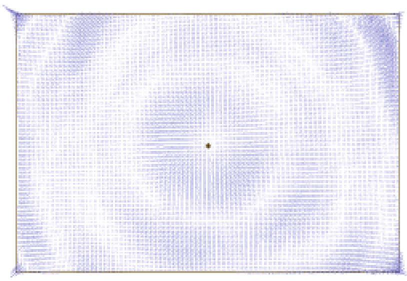


Fig. 24. Residuals: a_4 corrected.



The second image includes just the a_2 parameter, and the a_3 component with its two zeros arises strikingly sharp. In this image the residuals are displayed multiplied by a factor of 10. In the first image the factor is 5, and in the subsequent images the factor is 20, except for the last one where it is 30.
The third image has the component a_3 applied. Note that in spite of the increase in the factor the residuals appear smaller. In this camera as in many others either the first two components of the $\mathcal{D}_{rs}$ or more often the first component alone carry the bulk of the distortion. In this image the three rings of zeros of the a_4 component can be seen, but asymmetries begin to show up. In the fourth image the five zeros of the a_5 component can still be seen.

The fifth image has the component a_5 applied. The zeros of the next symmetric radial component are no longer discernible. Following this the $\mathcal{D}_{rs}$ components b_2–b_5 were applied. The resulting image is almost identical with this one and is not shown.

Next, the asymmetric components c_3–c_6 and d_1–d_6 where computed, and the result is shown in the last but one image. Finally the components c_7–c_{12} and d_7–d_{12} were included. This last image features the residuals multiplied by 30 instead of the factor 20 of previous images, for even these higher order component are significant and the residual are reduced.

Fig. 25. Residuals: a5 corrected.

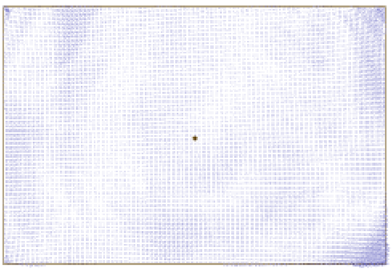


Fig. 26. Residuals: up to c6, d6 corrected.

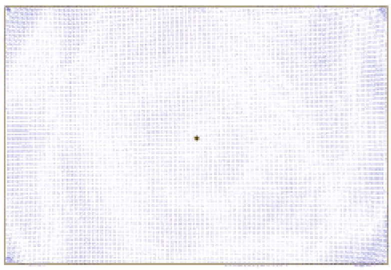
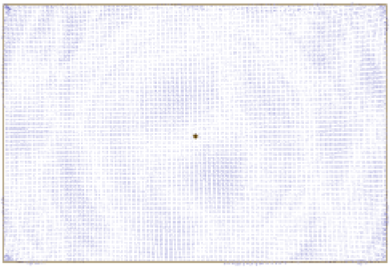


Fig. 27. Residuals: up to c_{12}, d_{12} corrected.



The results of this calibration are typical, except for the fact that the CM usually fits the distortion better than the OM. All the 16 tangential components are below the pixel size, the greatest being the asymmetric d_2 and d_5 with values of $-0{\cdot}77$ and $-0{\cdot}68$ and mean square value over the photograph of $0{\cdot}22$ and $0{\cdot}28$ respectively. I recall that the components are so designed that the value of each of them equals the greatest value of the distortion due to that component, which is reached at the corners of the photograph. The greatest distortions are the Drs, and within this the components a2 and a3 account for 99% of the whole distortion.

The four components are above the pixel size (this is not always the case), the least one being a_4 with a value of $-6{\cdot}86$ and mean square of $1{\cdot}23$. The asymmetric radial ones are some of them above, some below the pixel size, and all below that limit in their mean square, the overall mean square being exactly one pixel.

Finally the mean square of the residuals after the inclusion of the components up to c_6, d_6 is $0{\cdot}34$ px and after the inclusion of all the components is $0{\cdot}24$ px. This same value is achieved if the whole $\mathcal{D}_{ts}$ and the components d_7–d_{12} are omitted, and the graph of the residuals is almost the same as with all the components included. Such a good fitting of the model to the camera distortion with just 20 components was not expected, and this proved to be always the case.

The grid model

In view of the previous discussion, in case there is few data there seems to be no room for the division of the image in a rectangular grid and the computation of several parameters for each of them, making a total of 40 or more. As an example of conclusion 2 above, a flight with three strips, 45 photographs in all, GPS and INS data and a total of 561 measures over the photographs, when self-calibrated including the $\mathcal{D}_s$ and asymmetric components up to c_6 and d_6, appears to have very clear asymmetric components, like d_2 with a value of $25{\cdot}4$ (microns) and a precision for this computed value of $1{\cdot}1$, and similarly for d_1, d_4 and c_4, all of them with values above 10 microns. The pixel size was 12 microns. The observation of this same pattern in other flights, the smaller the flight the greater the distortions, and the fact that they never appear when calibrating a single image with hundreds of observations or self-calibrating a block with thousands of photographs, means that these are not actual distortions but the modelling of the actual measuring errors by the components of distortion. The only reliable values in these small flights are the ones of the symmetric components (if the flight is very small, only a_2).

A grid model will exaggerate this effect. It will have no consequences in the computed values of the image orientations, but the fact that what would otherwise appear as residuals is absorbed by the distortion parameters will result in an artificially low value of the computed a posteriori standard deviation. If the image has an actual a_3 (say) component, and this is not removed prior to the application of the grid, statistical tests performed on the parameters will conclude that they are significant, and that is the truth since they are describing an actual distortion, but all the components will be describing the behaviour of the single a_3 component within its rectangle in addition to local measuring errors.

In case there are thousands of measured points the parameters of the grid model will very likely be right, but the same distortion which this model describes with, let us suppose, 50 parameters, can be described with fewer overall parameters, very likely with just 20 as the previous test of 16 000 observations and similar ones have shown. Note that in the previous test after the computation of 20 components the standard deviation of the residuals was already at 0·24 px.

The problem behind the grid model is not that the concept be wrong, but that it cannot be applied to the geometric calibration of one-piece images. The irregularities of the distortion are very small with respect to the pixel size to allow a meaningful division of the same in several areas. We may look at other cases when a function over a surface is described with increasing precision by the progressive addition of components. When compressing an image by means of the Fourier inversion, the image is divided in small squares, but still the irregularity of the function, i.e. the image, within each square is usually much greater than that of a camera's distortion function over the whole image. And when the Geoid is described, local models are developed indeed, but scientist have computed thousands of global parameters of the orthogonal basis and continue to do so.

REFERENCES

JACOBSEN, K., 1998. Block Adjustment. *Institute for Photogrammetry and Surveying Engineering*. University of Hanover. Accessible at http://www.ipi.uni-hannover.de/uploads/tx_tkpublikationen/block_adjustment.pdf. 25 pages.

BROWN, D., 1966. Decentring distortions of lenses. *Journal of Photogrammetric Engineering and Remote Sensing*, 32(3): 444–462.