# IMAGE-TO-IMAGE PERSPECTIVE RECTIFICATION

R. Roncella, G. Forlani

Dept. of Civil Engineering, 43100 Parma, Italy – gianfranco.forlani@unipr.it

**KEY WORDS:** Architecture, Cultural Heritage, Modelling, Rectification, Close Range

**ABSTRACT:**

The rectification of image is a good method for achieving fast representation of building facade and obtaining measures directly on the rectified frame; however the method becomes unusable when the facade it's not composed by a small number of different plane surfaces: the number of measures to take becomes rapidly too relevant and requires a too demanding topographical survey. We studied a method to project the different plane surfaces of a facade on a single reference plain using a limited number of direct measurement in order to eliminate the out of plane distortion, guaranteeing a fast surveying approach.

## 1. INTRODUCTION

A photogrammetric survey method relatively easy to use and conceptually simple is the rectification of planes. Its simplicity is mainly due to the fact that it asks for a frame only and normally doesn't require a careful evaluation and planning of the operation in takings the images and measuring ground control points. Nevertheless in the last years its relevance has diminished or at least has been superseded by the emergence of 3D models as the fashionable or high quality product. To the cost of a larger effort in topographic work, they provide a more intuitive and richer representation of the object. Besides, also the inherent limits in the method that bring to results that, although rapid and hasty, don't allow to get the accuracy level that the today's market requires. We are mainly speaking of the problem arising when we apply rectification to objects not stricly contained in a unique plane: the entity of the out-plane error, when not accounted for, easily bring to unacceptable results; one has to make a choice among two opposite compromises: to accept second quality results for the inexpensiveness (in terms of time and effort spent), or to separately rectify every single prominence (operation that implicates an exponential increase of times of realization of the project in reason for the increasing number of points to coincide in the topographical relief and on the frame) getting results comparable to the realization of orthophotoplanes. Thus a method that, for his simplicity and banality, could be ideal for rapid architectural surveys above all for not photogrammetrists (architects, art historian, archaeologists etc) is being sidelined by other approaches that, although more complexes, bring to better results. The purpose of our work has been to elaborate a new method that could be integrated to the traditional rectification, overcame in most cases its limits without resortin to orthoprojection, proposing tools that though simplifying on site survey the operations, guarantees at the same time a good precision on the final results, also widening in this way the fields of applicability of the method. Besides many of the treated aspects can subsequently be implemented to increase results' flexibility, and the range of applications, not only in of the photoplanes, but also in terms of three-dimensional computer-graphics, retouching and scene representation.

The concept pivoting our job it's found on a simple consideration: in many cases, in architecture, although the geometric model of what is wanted to represent is both articulated and therefore not approximated by just a plain surface, however it can often be assimilated, to a set of simple geometric primitives composed by plain faces; the architectural composition, obviously skipping the modern tendencies, has always had a preference for simmetry, perpendicularity and parallelism among the various component. The algorithm proposed is based on this simple assumption and it aims to rectify every plain surfaces, or better projecting them all on a single reference plane that can coincide either be chosen in the image space (and therefore is not yet rectified) or in object space, therefore yields the final rectified image, asking the user just for simple and necessary number of measures.

## 2. PROJECTING THE PLANES

### 2.1 The algorithm

Our purpose is to elaborate an algorithm that allows to project many parallel plains on a reference plane, limiting the number of measures to take on site; ideally we would be able to guarantee a rather expeditious methodology of relief, to free ourselves from the use of expensive and specific measure instrument such thodolites. The user should just get data that allow to describe the object's geometry, drawing measures with simple and economic tools, and measure the vertices of each frame the plain surface that compose the model in the image. If all the surfaces that compose the object are mutually orthogonal or parallel, we will show how it is possible, with only three measures of distance, to solve the whole problem.

Most buildings have plenty of rectangular surfaces, bringing us to trying to exploit the geometry of the problem taking account the vanishing points of the straight lines parallel to the Cartesian axes (Van den Heuvel, 1998): in other words the algorithm, with a proper choice of the object's system, will not only use information from the points found on the surfaces to be projected but also from the border segments. To take into account also ideal points as the vanishing points we resort to use the rules of the projective geometry and to homogeneous coordinates (Hartley and Zisserman, 2000); as it is known there are no difficulties in the transformation between homogeneous coordinates and unhomogeneous coordinates and vice versa.

Be therefore $\mathbf{x}_i$ the generic image point with homogeneous coordinates $(x_{1i}, x_{2i}, x_{3i}) \in \mathbf{P}^2$ and $\mathbf{X}_i$ the correspondent point in object homogeneous coordinates $(X_{1i}, X_{2i}, X_{3i}, X_{4i}) \in \mathbf{P}^3$.

The relationship between the image point on the native surface $\mathbf{x}_i$ and its correspondent on the projected one $\mathbf{x}'_i$ over is:

$$x_i = H \cdot x_i' \qquad (1.1)$$

where H is a 3x3 matrix.
To determine H (that will depend on camera parameters and on the image geometry relative to the object) it is necessary to consider the more general relationship among object and frame coordinates:

$$x_i = P \cdot X_i \qquad (1.2)$$

where P (camera projection matrix) is a matrix 3x4(Hartley and Zisserman, 2000).
We want to find a relationship that allows to determine H as a function of P (that implicitly contains the image geometry) and of the entity of the projection *s*.
We start therefore to consider an image point $\mathbf{x}_i$ (with offset $s_1$)and its correspondent on the surface of projection $\mathbf{x}'_i$ (offset $s_2$). Attaching to the reference surface the ground system, with Z axis orthogonal to the various surfaces to be projected, the two homologous points will have identical value of **X** and **Y** object coordinates while **Z** will equal the entities of the prominence of the surface to project. We can write then:

$$x_i = P \cdot \begin{pmatrix} X_i \\ Y_i \\ s_1 \\ 1 \end{pmatrix} \qquad x_i' = P \cdot \begin{pmatrix} X_i \\ Y_i \\ s_2 \\ 1 \end{pmatrix} \qquad (1.3)$$

Accordingly:

$$x_i' = x_i - P \cdot \begin{pmatrix} 0 \\ 0 \\ s_2 - s_1 \\ 0 \end{pmatrix} = x_i - \Delta s \cdot \boldsymbol{p}_3 \qquad (1.4)$$

where $\mathbf{p}_3$ is the third column of the matrix P.
Nevertheless if we want to determine the matrix H of homologous transformation between the two surfaces (we must if we want to use the image coordinates taken from the frame) it becomes necessary to proceed in another way; we must consider, infact, that the homogeneous image coordinates in (1.4) are equal to the image coordinates taken from the frame up to an unkown scale factor.
We divide the equations (1.3) by the unknown scale factor $x_{3,i}$:

$$\begin{pmatrix} x_i \\ y_i \\ 1 \end{pmatrix} = \tilde{X}_{cam,1} \cdot \begin{pmatrix} \tilde{X} \\ \tilde{Y} \\ W \end{pmatrix}; \qquad \begin{pmatrix} x_{i1}' \\ x_{i2}' \\ x_{i3}' \end{pmatrix} = \tilde{X}_{cam,2} \cdot \begin{pmatrix} \tilde{X} \\ \tilde{Y} \\ W \end{pmatrix} \qquad (1.5)$$

where: $\tilde{X} = X / x_{i3}$

$\tilde{Y} = Y / x_{i3}$

$W = 1 / x_{i3}$

$\tilde{X}_{cam,1} = \left| p_1 \quad p_2 \quad (p_3 \cdot s_1 + p_4) \right|$

$\tilde{X}_{cam,2} = \left| p_1 \quad p_2 \quad (p_3 \cdot s_2 + p_4) \right|$

$p_i =$ column vectors of the matrix P

$s_i =$ offset of each plane from the reference plane

in this way the second hand member is known but for the unknown scale factor $x_{3,i}$ and the two ground coordinates X and Y, while the left-hand member coincides with the image coordinates; now we can invert (1.3) in order to obtain the unknown variables:

$$\begin{pmatrix} \tilde{X} \\ \tilde{Y} \\ W \end{pmatrix} = \tilde{X}_{cam,1}^{-1} \cdot \begin{pmatrix} x_i \\ y_i \\ 1 \end{pmatrix}$$

$$\begin{pmatrix} x_{i1}' \\ x_{i2}' \\ x_{i3}' \end{pmatrix} = \tilde{X}_{cam,2} \cdot \begin{pmatrix} \tilde{X} \\ \tilde{Y} \\ W \end{pmatrix} = \tilde{X}_{cam,2} \cdot \tilde{X}_{cam,1}^{-1} \cdot \begin{pmatrix} x_i \\ y_i \\ 1 \end{pmatrix} \qquad (1.6)$$

$$\Rightarrow H = \tilde{X}_{cam,2} \cdot \tilde{X}_{cam,1}^{-1}$$

Obviously, in order to obtain the image coordinates of the projected point, we have then to divide $\mathbf{x}'_i$ by $x'_{i3}$.
It is, therefore, clear that for the algorithm to provide accurate results it's of fundamental importance both to have a correct evaluation of the projection *s* and of the matrix P.
As far as *s* is concerned, it can be measure directly during the survey of the object, in which case the precision depends on the characteristics of the used instruments; it is also possible however to infer him in indirect way measuring on the frame any edge parallel to the Z axis that connects a surface of unknown prominence to one with *s* known. This, as shown in the following, entirely depends on the camera projection matrix (and obviously on the image coordinates of the end points of the edge) and therefore its evaluation is fundamental for the method to achieve good result.
Two image points on the edge $\mathbf{x}_s$ (on the unknown prominence surface) and $\mathbf{x}_r$ (on the known prominence) have indentical X and Y object coordinates (generally unknown). In homogeneous coordinates we can write:

$$x_s = P \cdot X_s \qquad x_r = P \cdot X_r \qquad (1.7)$$

$$x_s = \begin{pmatrix} x_{s,1} \\ x_{s,2} \\ x_{s,3} \end{pmatrix} = \begin{pmatrix} x_{s,imm} \\ y_{s,imm} \\ 1 \end{pmatrix} \cdot x_{s,3} \qquad x_s = \begin{pmatrix} x_{s,1} \\ x_{s,2} \\ x_{s,3} \end{pmatrix} = \begin{pmatrix} x_{s,imm} \\ y_{s,imm} \\ 1 \end{pmatrix} \cdot x_{s,3} \quad (1.8)$$

We can draw therefore from the preceding relationships up to six equations in the five unknown *s*, X, Y, $x_{s,3}$, $x_{r,3}$. From this we deduce that is enough for the evaluation of *s* that the two extreme of the segment that the user will go to specify on the frame are indifferently found on a same parallel surface to the plane XZ or YZ, in which case the solution is determined univocally, being constituted by five equations in five unknown; in the case, instead, in which the two points are found on a parallel straight line to the Z axis, the system is redundant.

It remains only to solve the problem best esteeming the camera projective matrix P: it is defined to less than a scale factor from the knowledge of its twelve parameters (they are had therefore eleven unknown) and it is therefore necessary the knowledge of at least six correspondences among image and object points; of these six correspondences at least three may be defined in an implicit way once the positions of the three vanishing points in the image are known. Of these only two are directly determinable, because we has departed from the assumption that the user goes to indentify on the frame only parallel segment (in terrestrial system) to the x axis or to the y axis. The third vanishing, correspondent to the straight lines with direction parallel to the terrestrial z axis, owes therefore to indirectly be drawn by the others two. To this respect the following property is of use:

> *Traced the triangle whose verteces are the three vanishing points correspondents to three directions mutually orthogonal, the principal point of the frame coincides with the orthocenter of the triangle.*

Knowing two vanishing points it is therefore always possible to determine the third, as the intersection of the line orthogonal to a side and passer-by for the principal point and the passing straight line for one of the vertexes and normal to the direction of the height exiting from the other vertex.

It remains therefore only to express three of six correspondences: once the dimensions of one rectangular surface the problem, would apparently therefore to be resolved. Nevertheless the matrix P esteemed in this way excessively hears some lack, in his determination of proper points in direction z (along which the correspondent vanishing point is solely used). The matrix, in other words, it is able to correctly esteem the information of the points on the plan of the rectangle used for the determination of P, but it results entirely ineffective for how much it concerns pending surfaces on different prominences.

To resolve the problem therefore becomes essential to require to the user to point out a third measure on a leaning plan that allows the correct evaluation of P: this measure can be the precise statement of the coordinates terrestrial of a leaning point or more simply the length of a segment exiting from the used rectangle.

Once known a good respect of the matrix P the problem is geometrically resolved.

From how much dictate it appears evident as the algorithm try to set the bases for a mixed technique of rectification in which an image initially is resampled with the purpose to eliminate all the perspective deformations due to the presence of out-of-plane surfaces, which operation is possible only with a further specification of the geometric model of the represented object. The method therefore it is assimilable to the generation of an orthophoto whose is known the DTM or the DSM of the surface that is wanted to represent. Unlike these methods, however, the geometric model is described only through a parameter (the prominence s) for it is assumed, likely, that the various faces that compose the model are parallel to a same plane.

As we know, generally, for the rectification of a plane (representative of every single face) it is necessary to know the correspondence among the terrestrial and frame coordinates of at least four points belonging to the surface to be rectified; in the algorithm proposed a simplification is offered if we think of the possibility to consider this correspondence in terms of projective geometry and therefore considering not only belonging points to the surface to be rectified but also improper points that are found on the vanishing line (and which represent the direction of straight lines belonging to the plane to be rectified). Besides, as we will see better in the succession, almost all the operations of measure take place directly on the frame through collimation of points and individualization of segments, which it allows a great inexpensiveness of measure during the topographical relief.

## 2.2 The Program

For better understanding of th process and of the solving equations, we follow step by step the operations that the algorithm performs for resolving the problem highlighting the key points:

1. The user select a rectangular surface pointing out the dimensions and the distance of it from reference surface; already, at this point the program is able to complete a first estimate of the position of the points and of the image geometry: implicitly it puts the origin of the object reference system on the first point collimated in the image (which will have therefore homogeneous coordinates (0,0,0,1)) directing the Cartesian axes so that x and y are along the two sides of the rectangle, while the definition of a right-handed system is automatically determined the direction of the z axis; therefore, known the dimensions of the rectangle, it could be also calculated the ground coordinates of its vertexes; besides, having two couples of parallel segments to respectively the x and y axes we are able, for intersection, to draw the position of the vanishing points (in object coordinates (1,0,0,0) and (0,1,0,0)); besides, if the principal point coordinates are known, from the position of the two vanishing points it is possible to draw the position of the third. In this way we are able to write a first relationship among 6 (or 7) point found on the frame and the relative object points.
2. The user then selects other plain regions in the image; in all the cases he can decide whether the surfaces will be used for the evaluation of the vanishing points (they will have therefore sides parallel to the Cartesian axes), the dimensions of the sides (in the case in which the figure is still rectangular), or whether the distance from the reference surface. It’s clear that all these information are optional and they solely have the purpose to increase the precision of esteem of the estimation.
3. At this point it starts the procedures of evaluation: in first place the vanishing points (if has been specified rectangular surfaces). The frame orientation it’s calculated for following steps: a first evaluation is used for determining the matrix P thanks to only one surface through a DLT. This first evaluation is functional to allow the calculation of the terrestrial coordinates (held unknown) of the vertexes of the other surfaces. As already said, however, the matrix P is not able to get out correct coordinates out (in the direction of the zes) of the plane used for it evaluation. Then the user defines a segment in direction z or points out, in the same reference sistem, the coordinates of a terrestrial point not pending on the aforesaid surface. The matrix P is therefore re-esteemed and now it is able to allow the determination:
    a. Of the object coordinates of the vertexes of the surfaces if their positions are specified in direction Z (or the entity of the prominence on which they are found is given);

b. The entity of the prominence on which every surface is found if the user is able to specify a segment that connects a given surface to another with known Z.

All the suitable surfaces are considered therefore for a final compensation of the results: unknown of the system are also, besides the parameters of the matrix P, all the indirectly determined object points (they are initially esteemed with an approximation for the matrix P).

4. To summarize, the information that the user inserts can be:
   a. Plain regions (with sides of the polygonal that contain them not parallel to the coordinated axis) whose offset is known with respect to a reference surface: in this case the region is simply projected on the reference surface;
   b. Plain regions with sides of the polygonal that contain them parallel to the coordinated axis: in this case the information on the sides are used for determining with greater redundance the vanishing points;
   c. Rectangular plain regions with sides (at least one) of known length in the terrestrial system: in this case the information determined by the sides is used not only for the redundance of the vanishing points estimation, but also to compute more effectively the matrix P;
   d. Plain regions, rectangular or not, whose offset is not known but from which start a segment (parallel to the plain XZ or YZ) traceable on the frame that unites the region with a surface with known Z: in this case the prominence is esteemed indirectly and the region, with its characteristics, it comes subsequently used according to the previously criterions statements.
5. The program finally proceeds to re-sample the image, projecting all the surfaces on the reference one or directly performing the rectification. The three-dimensional known geometry of the object, and particularly the succession of the surfaces in direction of the Z axis, is also able to show the possible occlusions, both directly on the rectified image and on a separate image that can be used for instance as a mask for possible retouching operations.

## 3. CONCLUSION

### 3.1 Results

In terms of visual appearance of the rectified image, the tests with the simulated data (see last section) have shown that the elaborated algorithm broadly satisfy its goal: even with only three measures of distance the program generates a coorectly rectified image of satisfactory quality for objects with any number of planes projecting out of the reference plane; besides, if a greater precision and a greater control is required, the algorithm allows to give a description more faithful of the analyzed object's geometry. Also in this case the only measures the user should perform are the distance between two points, easily performed with simple instrumentation as hand-hold distance meters. The possibility to underline occlusions in the images and the generation of masks to facilitate a possible retouching were introduced to provide a flexible tool to users that in many cases cares more about the visual appearence of the results rather than metric accuracy or clarity of the procedure.

As far as the correctness of the results and the numerical stability is concerned, the matter is more delicate: although the computations have been arranged to make the method robust, in some tests it has emerged that inaccuracies in the collimation of the image points used to determine the vanishing points or the connection's segments among surfaces of different prominence could make solution unstable: particularly measurement on the first rectangles that the program uses to compute a first approximation of the camera matrix are essential (as may be expected) for the rest of the computation. In other terms the facilitation that is offered to the user, which has to take only few measures to get a result in many cases much satisfactory, has as counterpart the necessity to perform carefully the first data collimations.

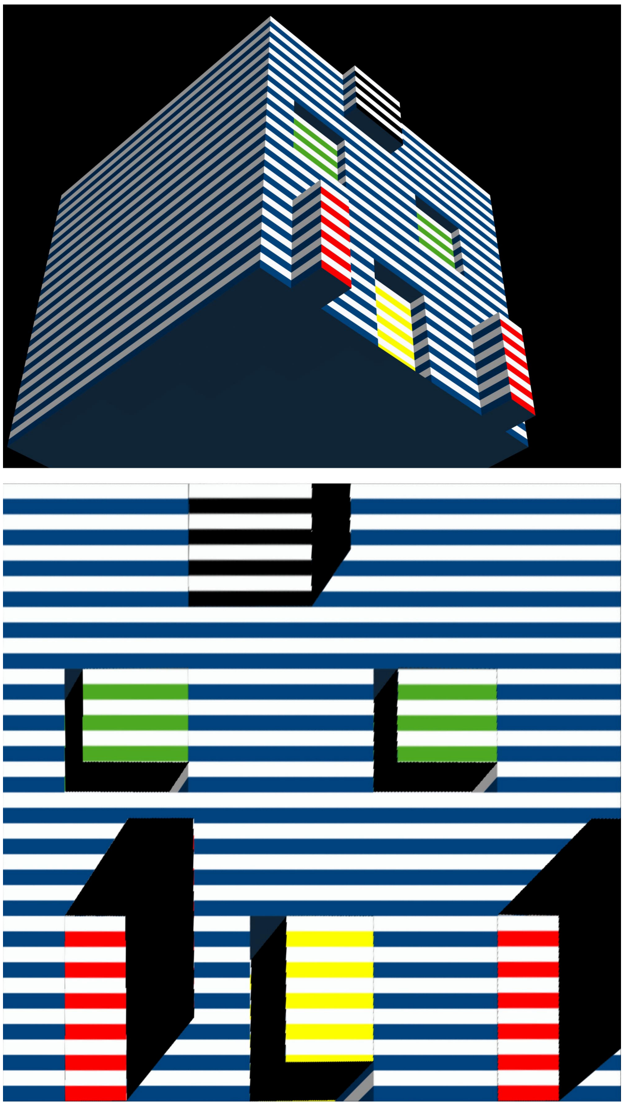

Figure 1. Top. *One of the artificial models used to test the result achived by the program.*
Bottom. *The final results: a rectification of the projected plane.*

With the artificial building models as those shown in figure 1 and in figure 2 (with maximum dimension 20x20 meters), the maximum error on evaluating the object position of the vertices of the projecting surfaces is of about 2÷3 cm: in other words the medium error is of about the 0.05÷0.1 % of the maximum dimension of the object. Obviously in reprojecting on the reference surface the repositioning error is much smaller for it's quite less sensitive in relation of the errors about the ground vertices' coordinates.

In the next section the results of the application of the method to artificially generated image are introduced: as the same images show the projection of the surfaces it results in all the cases demanded to the pixel (the original image is a 3000x2000 pixel).

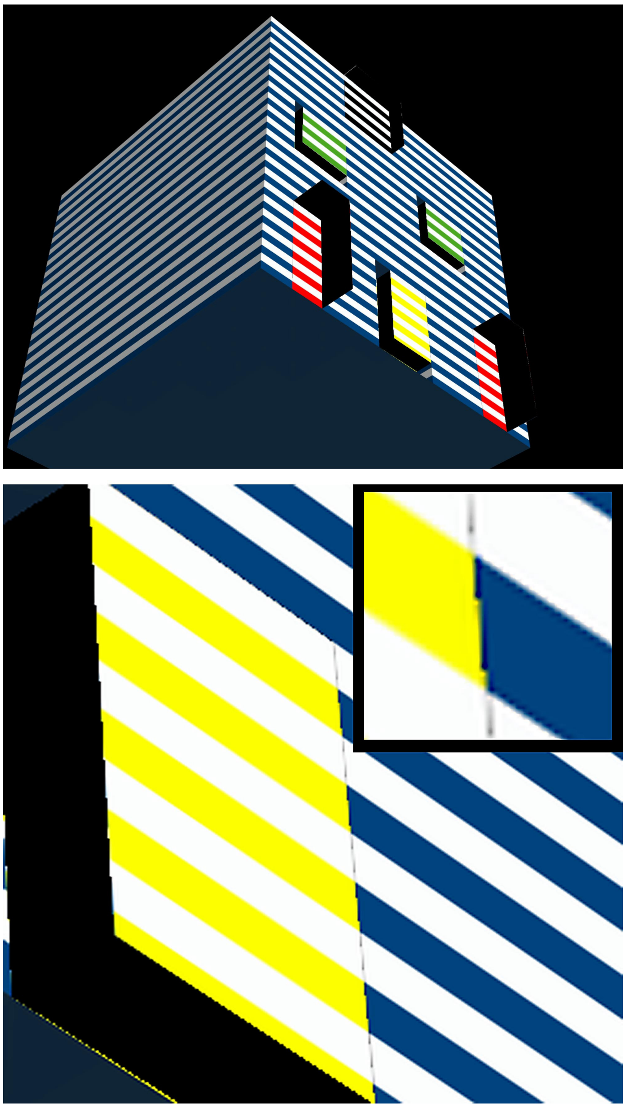

Figure 1. *Zooming the projected area we can see how the errors on reprojecting the plane is less than one pixel.*

### 3.2 Perspectives

It is too early to make statements about the method's performance in practice: a trial period using real images will tell more about. The simulation so far show the correctness of the method under rather ideal condition. The algorithm proposed does not pretend to be the answer to the quest for simple and universal tool for photogrammetric surveys: undoubtedly it just give a mean to solve with simplicity a certain ranges of cases. Although the concept at the base of our job has been enough deepened in its implications, we care thinking of (and it will be clarified in the next months) how effective they are some developments that will make the tool more flexible and performant.
First of all it will be sought to increase the number of geometric primitives acceptable to the method to allow the description of a larger number of architectural objects: this always trying to maintain enough precision of restitution by keeping at a minimum the information that the user must provide with the program.
Another improvement of the current program version, of simple solution, will be the introduction of more images for the final restitution or the ability to realize mosaiking.
Another interesting aspect will be trying to export the results to a 3d modeler: currently the program allows only to get a resampled image of the projected object, correcting all regions parallel to the XY plain; nevertheless it is simple to produce a group of images, each corresponding to a region different specified, to use as texture for a model. Of simple implementation it's also to produce images that contain the faces of connection among different regions that is the faces (to the moment neglected) contained in the plane xz and yz.
Besides to maintain accuracy while opening up to non photogrammetrist it is essential to find a rapid and possibly automated procedure of camera calibration: At present the program allows to insert the camera calibration parameters so that to eliminate the optic distorsions; user doesn't have knowledges neither the tools to calibrate his own camera; if the user is not a professional surveyor, the images to be elaborated are likely to have been taken with middle-low level film or digital cameras, with variable principal distance and with lens of not particular quality. To the moment we are just studying the possibility to integrate the program with the plumbline calibration technique (Fraser et al.): the straight lines the nethod relies on for the evaluation of the vanishing points and for the determination of the surfaces to project make the plumbline method the simplest and most efficient tool to get a calibration work sufficiently accurate.

## 4. REFERENCES


Fryer, J.G., Brown D.C., 1986. Lens distortion for Close-Range Photogrammetry. *Photogrammetric Engineering & Remote Sensing*, 52(4), pp. 51-58.

Gruen, A., Huang, T.S., 2001. *Calibration and Orientation of Cameras in Computer Vision*. Springer, Berlin, pp. 116-119.

Hartley, R., Zisserman, A., 2000. *Multiple View Geometry in computer vision*. Cambridge University Press, Cambridge, pp. 3-44, 139-153, 166-213.

Parker, J.R., 1997. *Algorithms for Image Processing and Computer Vision*. Springer, Berlin, pp. 1-19.

Press, W., Flannery, B., Teukolsky, S., Vetterling, W., 1988. *Numerical recipes in C*. Cambridge University Press, Cambridge.

Sutherland, I.E., 1980. *Sketchpad: A man-machine graphical comunication system*. Garland Publishing, New York.

Van den Heuvel, F., 1998. Vanishing point detection for architectural Photogrammetry. In: *The International Archives of the Photogrammetry and Remote*, Hakodate, Japan, Vol. XXXII, Part 5, pp. 652-659.