ORTHORECTIFICATION OF HIGH RESOLUTION SATELLITE IMAGES

P. Boccardo [a] , E. Borgogno Mondino[a],*, F. Giulio Tonolo [a], A. Lingua [a]

[a] Politecnico di Torino, Dipartimento di Georisorse e Territorio, C.so Duca degli Abruzzi 24, 10129 Torino, ITALY
(piero.boccardo, enrico.borgogno, fabio.giuliotonolo, andrea.lingua)@polito.it

ABSTRACT:

The growing availability of high resolution satellite images leads to evaluations that are aimed at the definition of the mapping scale they can reasonably be defined for. The problems connected to the use of satellite images for the production of orthophotos in urban areas at a middle scale are dealt with in this work. An analysis of the residuals perspective errors connected to the presence of great altimetric discontinuites due to buildings and infrastructures is made. The effects of the use of a Dense DTM are evaluated and the problem of displacements on orthophotos induced by raised volumes are considered too.

A procedure has been developed which allows the orthoprojection of high resolution satellite images using a Dense DTM and non-parametric models. In particular non parametric algorithms based on Rational Function Models (RFM) and Neural Network (*Multy Layer Perceptron*) have been implemented. A multi-image (along-across track stereo images, multi-temporal images) and/or multi-sensor approach can be followed for the production of true orthophotos: this approach can prevent from the problem of data duplication next to elevation discontinuites.

1. INTRODUCTION

The introduction of high and very high resolution satellite images has made it necessary to revise the geometric correction techniques that are used in this field. There has been a transition, on the basis of evaluations aimed at the definition of map scale for which they can reasonably be defined, from simple 2D polynomial models to rigorous or non rigorous 3D models derived from digital aerial photogrammetry.

When the investigated zone is an urban area or when the territory is characterised by discontinuities, a classical type of orthoprojection could be insufficient for mapping purposes and might need to be substituted with a more rigorous approach (true orthophoto).

2. GEOMETRIC CORRECTION

The geometric correction of high-resolution satellite images can be carried out using two different approaches: rigorous modelling or non-parametric modelling.

Rigorous models are based on collinear equations (Toutin, 2004) that are adapted to *pushbroom* acquisition technique which is used by all high resolution satellites.

In this case, the orientation parameters are modelled as time dependent polynomials of a higher degree than the first: the estimation of the unknowns requires approximated initial values which are extracted from the metadata files usually supplied together with the images.

However the Companies that distribute images are not always willing to supply detailed technical information to the final users concerning the platform that is used or about the characteristics of the sensor that are necessary to implement rigorous models. It was for this reason that non-parametric models, or rather generalised models (independent of both the type of sensor and of the acquisition method) were introduced.

The most frequently used non-parametric methods are based on 3D rational polynomials, and which in literature are known as the *Rational Function Model, RFM – Rational Polynomial Coefficients, Rational Polynomial Camera, RPC – Rational Function Coefficients, RFC (*Dowman, Tao, 2002). Moreover a new prototype of a geometric correction procedure based on a Multy Layer Perceptron type (MLP) neural network, has been proposed in this paper. These latter methods have been analysed and verified in the application field.

2.1 Rational Function Model

The rational function is the most commonly used non-parametric model, which is implemented in almost all software packages for the processing of satellite images. This type of approach is used by image salesmen to allow the final user to obtain added value products, such as orthoprojection without the necessity of having a model of the sensor, but by only attaching the coefficients of the relation between the image coordinates and the ground coordinates.

The rational function model allows a relationship to be determined between the image coordinates (ξ, η) and the 3D coordinates of the object (X, Y and Z) through polynomial relations, as shown in (1)

$$\xi = \frac{P_a(X,Y,Z)}{P_b(X,Y,Z)}$$
$$\eta = \frac{P_c(X,Y,Z)}{P_d(X,Y,Z)} \quad (1)$$

where P_a, P_b, P_c, P_d, are usually maximum degree polynomials equal to 3, corresponding to 20 coefficients, which can be expressed through equation (2) or (3):

---

* Corresponding author.

$$P_a(X,Y,Z) = a_0 + a_1 X + a_2 Y + a_3 Z + a_4 X^2 +$$
$$+ a_5 XY + ... + a_{17} Y^2 Z + a_{18} YZ^2 + a_{19} Z^3 \qquad (2)$$

$$P_a(X,Y,Z) = \sum_{i=0}^{m_1} \sum_{j=0}^{m_2} \sum_{k=0}^{m_3} a_{ijk} X^i Y^j Z^k \qquad (3)$$

$$0 \le m_1 \le 3 \,;\, 0 \le m_2 \le 3 \,;\, 0 \le m_3 \le 3 \text{ e } m_1 + m_2 + m_3 \le 3$$

Equations (1) are known in literature as *Upward RFM* as they allow the image coordinates to be obtained starting from the 3D coordinates of a ground point. In order to proceed with the estimation of the transformation parameters a_i, b_i, c_i e d_i ($i = 0 \div 19$), it is necessary to trigger a least square iterative process after having linearised equations (1). This procedure has been implemented in IDL language (*Interactive Data Language)* so as to be able to adjust for the lack of transparency of commercial software packages.

Some critical elements of the algorithm are here underlined. First of all, in order to avoid numerical calculation problems (truncation or under/overflow errors), it is necessary to normalise both the image coordinates and those of the object in the interval (-1;+1) (OpenGIS-OCG, 1999). Furthermore, as the denominators of the polynomials assume very different values, in function of the distribution of the GCPs and of the altimetric range, it is probable that the coefficient matrix of the least square system results to be ill-conditioned. As a consequence, the normal matrix can result to be singular, in particular when polynomials of a high degree are used. The iterative process often does not converge in this case.

If we presume that the *camera model* is not available to the users, it is necessary to choose the ground control points in a conventional way, that is, through collimation of the homologous points on the cartography/DEM or through specific GPS survey campaigns. As it is not possible to obtain a regular distribution of the GCPs, it is therefore necessary to implement a numerical regularisation algorithm to make the iterative process converge.

Tikhonov algorithm is one of the most commonly used regularisation algorithms for the resolution of ill-conditioned systems: it consists in the adding of an arbitrarily small constant λ^2 to the diagonal of the normal matrix in order to improve the conditioning number. The resolution equation of the least square system (4) is then changed as shown in (5)

$$V = Q \cdot A \cdot X - Q \cdot L \qquad (4)$$
$$X = \left(A^T \cdot Q \cdot A + \lambda^2 \cdot I\right)^{-1} \cdot A^T \cdot Q \cdot L \qquad (5)$$

The choice of Tikhonov's parameter is not univocal and it is in fact obtained by empirically elaborating numerous solutions while varying the λ parameter, choosing the one that minimises the rejects on the control points.

The numerous tests that have been carried out have shown how, even though the iterative process converges, the subsequent orthoprojection step, in some cases, have some problems that are connected to distortion of the generated images which are probably due to the use of polynomials of too high a degree. To verify the correctness of this hypothesis and in order to avoid having to "a priori" choose the degree of the polynomials that has to be used (that is, the number of parameters for each polynomial), an adequacy analysis of the model was implemented. It is based on two different statistical tests in order to automatically determine whether a coefficient is necessary and which coefficients are not necessary.

The χ^2 test with allows to verify whether the model was overparametrized. In the case in which an overparametrization is shown, another test on the significance of the estimated unknowns is performed to determine how many and which polynomial coefficients are not necessary.

The standardised Z parameter is calculated according to the following relation:

$$W_i = \frac{\hat{x}_i}{\hat{\sigma}_{x_i}} \qquad (6)$$

where: $\hat{x}_i$ = the ith estimated coefficient and, $\hat{\sigma}_{x_i}$ = i[th] estimated r.m.s. In the case in which the following statistical test (7) based on the Student distribution, is verified the relative coefficient x_i is placed equal to zero:

$$|W_i| \le t_{\alpha/2} \qquad (7)$$

Where: $t_{\alpha/2}$ = the value of the Student distribution for the relative value of redundancy (n –r), n = number of equations, r = number of unknowns

The coefficient is annulled through the addition of a new condition equation to the initial system and by inserting an elevated weight in the corresponding position of the *weight* matrix..

2.2 Neural Network Model

The neural network approach to the orthoprojection of satellite/aerial images can be considered an innovative attempt to solve the problem of the correction of images through non-parametric methods.

The neural network consists of mathematical models whose operative philosophy is inspired by cerebral biological dynamics: the calculation process is schematised as a flow of distributed information whose elaboration occurs inside dedicated calculation units, which are known as "neurons" of the network. Some of these receive information from the external environment, others return answers to the environment and still others, if there are any, communicate with only the units inside the network: they are called input, output and hidden units, respectively, as shown in figure 1.
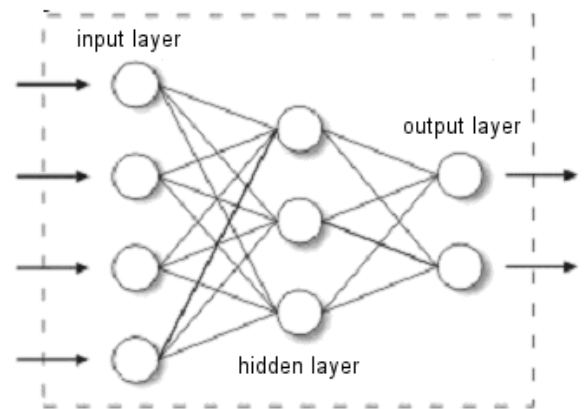

Figure 1 – Concept scheme of a two layer (hidden and output) computational MLP neural network

The neural panorama is extremely large and neural algorithms have been developed to resolve very different kinds of applications: it is the choice of application that determines the choice of algorithm.

Attention has been paid to the MLP algorithm (Multi Layer Perception) to obtain a geometric correction of satellite images. Function approximation and estimation properties of this algorithm (non-linear) have already been widely described in literature. The basic idea is that of substituting the *upward* projection model that relates the image coordinates (ξ, η) to the coordinates of the object (X, Y and Z) with an MLP neural network opportunely trained on the basis of the GCPs.

The reasons for this choice arose out of an analysis of the problems connected to the previously described RFM approach. Neural networks preserve from a forced linearisation of the equations around an approximated solution. They give a non-linear response to a non-linear problem, whose efficiency increases, just similarly to RFMs, with the growing of the number of GCPs and with the decrease of the original image deformations.

In the MLP network each neuron performs a very simple operation that consists in generating, through an opportune function, which is known as the transfer function, a response to the signals that converge on it through communication channels. These channels simulate the biological synapses and their duty consists in "weighting" the intensity of the transmitted signals: for this reason they are known as "synaptic weights" or simply "weights".

Formally the response signal (u_i) that is restituted by the generic neuron ith is equal to:

$$u_i = f(\sum_{j=1}^{N} w_{ij} p_{ij} + b_i) \qquad (8)$$

where f is the transfer function which normally takes the shape of a hyperbolic tangent (9) or of a logical sigmoid (10), w_{ij} are the weights of the ith neuron, p_{ij} are the input at the ith neuron (N) and b_i are scalar additives, called *bias,* that are considered as weights of unitary additional input (Figure 2).

$$f(x) = \frac{1 - e^{-x}}{1 + e^{-x}} \qquad \text{Hyperbolic tangent} \qquad (9)$$

$$f(x) = \frac{1}{1 + e^{-\alpha x}} \qquad \text{Logical sigmoid} \qquad (10)$$
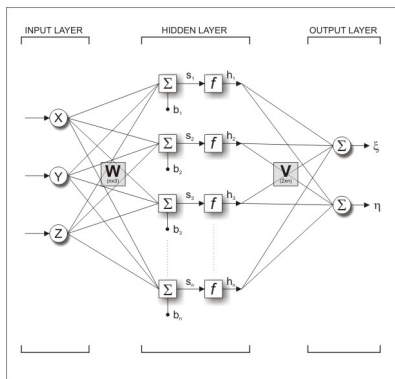

Figure 2 – Mathematical model of a two-layer computational MLP neural network (*hidden and output*)..

This type of algorithm belongs to the *feed-forward* family of neural networks, that is, networks in which the information travels in parallel and in a single direction.

The MLP network therefore constitutes a mathematical model in which the parameters are the *weights* and the *biases* of the *hidden* and of the *output layers.* The estimation of the values of these parameters on the basis of opportune samples (*patterns),* represents the training step of the network. In this application the training algorithm is the optimised (for a greater convergence speed) *Error Backpropogation* (EBP), known as Levenberg-Marquardt (LM).

In the EBP algorithm the network weights assume values that minimise (local minimums) the Performance Function (PF). This is defined, for a *batch* training, as:

$$PF(W(t)) = \sum_{p=1}^{P} \sum_{k=1}^{K} \left(d_{kp} - f_{kp} \right)^2 = E(t)^T E(t) \qquad (11)$$

where $W(t) = [w_1, w_2, \dots, w_N]^T$ is the weight vector of the network at epoch t, t counts the epochs of the training process and it is fixed by the operator, d_{kp} is the expected value (*target)* of the kth output relative to the pth training pattern, f_{kp} is the value of the kth output calculated by the network, $E(t) = [e_{11}, e_{21}, \dots e_{k1}, e_{12}, \dots e_{k2}, e_{1P}, \dots e_{KP}]^T$, in which $e_{kp} = (d_{kp} - f_{kp})$, $k = 1, \dots, K, p = 1, \dots, P,$ is the cumulative error of a batch training.

MATLAB 5.3 Neural Network Toolbox routines have been used. An *upward* "orthoprojection" approach has been adopted so that the coordinates of the object (X, Y, Z) constitute the network input and the coordinates of the image (ξ, η) constitute the output. Only one hidden layer and one output layer have been foreseen. Two network configurations have been verified and implemented with respects to the possible transfer functions that can be used. In the first case, the transfer function adopted for the hidden layer, which results more appropriate for the treatment of pushbroom images (to which the here presented results refer), is a hyperbolic tangent (9) while, for the output layer, it is a simple linear function (purely weighted sum).

In the second case, which is considered more suitable for the treatment of whiskbroom images, a logical sigmoid transfer function (10) has been adopted for the hidden layer while, also in this case, a simple linear function has been considered for the output layer.

The number of neurons (of the hidden layer) that drive to best performances has to be determined each time on the basis of repeated tests: in the MATLAB developed routine it is the calculator itself that does this automatically.

It should be recalled that an approximate estimation (as we are working in a non-linear ambit, these considerations are purely indicative) of the maximum number of admissible neurons could be obtained by comparing the training pattern number (the GCPs) with that of the parameters to be estimated (weights and bias). This results to be equal in number to:

$$N_{param} = (3_{(XYZ)} \cdot M_{PESIhidden}) + M_{BIAShidden} + 2_{(\xi\eta)} \cdot M + 2 \qquad (12)$$

where M is the number of neurons of the hidden layer.

More precise indications could be derived from a careful analysis of the results (residuals), verifying the possible appearance of overfitting phenomena, which, as a first approximation, can be identified in the progressive spread of

the differences between the residuals on the GCP and those on the Check Point.

The accuracy of the solution (which can be evaluated in terms of residuals on the GCP and on the CHK) varies to a great extent with different pseudo-casual initialisation of the network weights and different number of neurons, while the initial value of the training parameter μ of the LM algorithm results to be quite negligible even though values close to 10^{-3} are advisable.

The architectures of the network that the developed procedure is able to verify, depend on some parameters that the operator has to supply:

- the range of variability of the number of neurons: the maximum and minimum number of neurons to be tested has to be defined. These values are influenced by the number of GCPs that are supplied as training pattern. A too high number of neurons would decreases the generalisation capacity of the network while a too low number would not approximate the function in an adequate way.

- the number of successive initialisations for each node architecture: the number of times the training should be repeated for each architecture has to be indicated. The results that can be obtained, with the same number of neurons and same μ value can differ greatly according to how the weights are initialised. The risk, because of a bad initialisation, is that the algorithm stops inside local minimums of the performance function. The repetition of the training for a sufficiently high number of times (10 times for this work) prevents from this problem.

- level of accuracy required of the residuals on the GCPs and on the CHKs: the optimal architecture is identified on the basis of the simultaneous satisfaction of some conditions:

$$\begin{cases} RMSE_{GCP} < threshold \\ RMSE_{CHK} < N \cdot threshold \end{cases} \qquad (13)$$

where

$$RMSE = \sqrt{\frac{\sum_{i=1}^{P} RMS_i}{N-1}} \qquad RMS_i = \sqrt{\Delta \xi_i^2 + \Delta \eta_i^2}$$

The threshold value depends on the expected scale of the orthoimage. The simultaneous satisfaction of the two conditions helps, in a simple way, to prevent overfitting problems of the network (too many neurons) which can be detected, as a first approximation, as the difference between $RMSE_{GCP}$ and $RMSE_{CHK}$.

3. RESULTS

The RFM and MLP non-parametric methods were tested with images acquired from different satellite platforms. The accuracy of the planimetric positioning was evaluated through the evaluation of the residuals on both the GCPs that was used for the estimation of the model parameters and on the CHKs (which were different from the previous ones). The mean values of the residuals were also calculated to show any systematic error. A geometrically homogeneous distribution of the GCPs on the entire image was maintained during all the tests, as the validity of the non-parametric methods decreased with an increase in the distance from the support points.

Tables 3 and 4 report the results that were obtained using the polynomial relations and the neural networks.

Satellite	N° GCPs	N° CHKs	Δξ mean CHK	Δη mean CHK	RMSE CHK (pixel)	RMSE GCP (pixel)
Eros A1	51	6	0.00	0.00	3.19	0.83
QuickBird	60	30	-0.09	0.02	2.76	0.86
Spot5	50	5	-0.02	-0.09	2.09	1.01

Table 1 – Results obtained through the application of the RFM method

Satellite	N° GCPs	N° CHKs	ΔE Mean CHK	ΔN Mean CHK	RMSE CHK (pixel)	RMSE GCP (pixel)
Eros A1	51	6	-0.23	-1.10	2.46	1.08
QuickBird	60	30	-0.23	-0.14	2.37	1.40
Spot5	50	5	-2.04	0.02	2.96	1.38

Table 2 – Results obtained through the application of the MLP neural network method

4. TRUE ORTHOPHOTO

The orthoprojection of satellite images is a procedure that is used to correctly represent orthogonal projection, on a prefixed plane, of the area framed by the sensor during the acquisition. This product is obtained through the orthogonal projection of each pixel of the image of the territory onto a cartographic plane, in such a way that the original perspective representation (a deformed cylindrical perspective in the case of pushbroom acquisition) is transformed into an equivalent metrically correct image. It is in fact possible to measure angles and distances on the orthophoto, but also to read the cartographic coordinates of significant points exactly like on a map.

To carry out an orthoprojection it is therefore necessary to have a geometric model of the sensor that is able to relate the 3D coordinates of the object to the coordinates of the image and a digital terrain model (DTM). It is very easy to perform an orthoprojection if the surface of the object is continuous (that is, smooth), but is not sufficiently accurate if the framed area is an urban area and if the geometric resolution of the image is high because of numerous discontinuities (breaklines) and frequent defilated areas.

A more complex method should therefore face the following problems:

- the lack of information in correspondence to the defilated zones (hidden areas): it is possible to eliminate or limit this inconvenience using a multi-image approach, if several images of the same area are available (along-track, across-track or multitemporal stereoscopy). In the case of a lack of images acquired by the same platform, it is possible to use the information acquired by other sensors using a multi-sensor type approach;

- the presence of discontinuities: this problem is resolved using more rigorous interpolation methods for the generation of a DTM whose grids are made up of a large number of points (DTM dense or Dense DTM). A dense DTM therefore allows a correct description of a surface to be obtained, a description that also takes into account the height of the buildings.

4.1 DDTM

The approach that was chosen for the generation of the dense DTM uses refined interpolation techniques applied to 3D digital map. The planimetric and altimetric information that is contained in them can in fact adequately describe the territory (height points, contour lines), the built entities (sections, vertices) and the buildings (centroids of known heights). The DDTM that were used to generate orthoprojections of satellite images were generated using the GeneDDTM software implemented in Visual Fortran at DIGET at the Politecnico di Torino (Dequal et al, 2002).
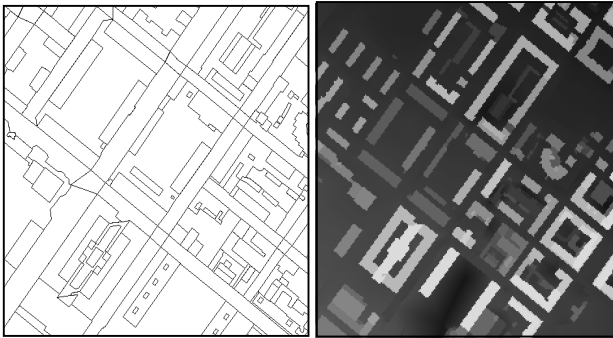
Figure 3 – Part of the DDTM used for the generation of the true orthophoto (on the right) and the relative digital map at 1:2000 scale (on the left).

4.2 Application field

The fields in which precision orthoprojection of high resolution satellite images could be more profitable than the usual procedure were investigated on the basis of the operative characteristics of the sensors. From this point of view, a test was first carried out of the conditions within which the displacements of the objects that can be put down to the presence of great discontinuities (buildings and infrastructures) results not to be negligible compared to the expected mapping scale to which the orthophoto should refer.

Displacements due to the presence of 3 different classes of buildings defined according to their height: $h_1=10m$, $h_2=30m$, $h_3=100m$ were considered. Then the negligibility limits were reported that were considered to be equal to the tolerance of the orthophoto (0.5 mm at the map scale) for the following scales: 1.2000, 1:5000 and 1:10000. These tolerances resulted to be 1 m, 2.5 m and 5 m, respectively.

In the graph of Figure 4 it is shown how with a variation of the mean view angle of the scene (γ) the entity of the displacement is increased due to the buildings. The three curves that are reported refer to the three different classes of buildings that were considered. The variability of the mean view angle was limited to an interval of (0-50) gon, considering the operative characteristics of the satellites, where the orientation capability of the sensor with respects to the nadir position never exceeds this value.

An analysis of the graph allows us to see how, for a building height equal to 30 m, taken as an example (central curve), it is easy to establish the following mean view angle values so that the displacements are kept within the permitted tolerance for the orthophotos at different scales:

- lower than ~ 2.5 gon for an orthophoto in a 1:2000 scale;
- lower than ~ 6 gon for an orthophoto in a 1:5000 scale;
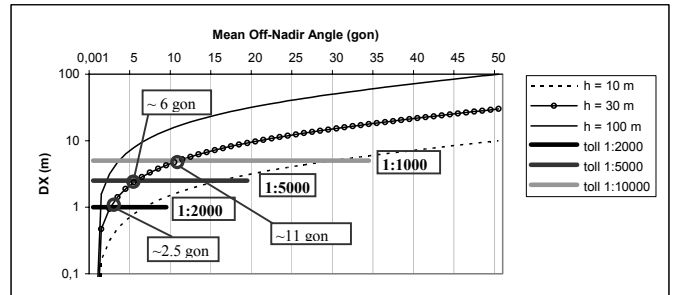- lower than ~ 11 gon for an orthophoto in a 1:10000 scale.

Figure 4 – Entity of the displacements (due to the presence of buildings) in function of the mean view angle.

It has been shown that orthoprojection of non-nadir satellite images requires in most cases a more accurate approach and that altimetric height data of the buildings cannot be neglected, above all in urban areas.

4.3 Tests

In order to verify the efficiency and real incidence on an application case when adopting a "precision" approach for the generation of orthophotos, a non-nadir EROS A1 image (GSD = 1.9 m) was processed first using a traditional DEM (50 m steps) and then a DDEM previously generated from 3D digital map using the previously described software. The scene refers to the city of Cuneo (Piedmont).

It can be seen in figure 5 how the viaduct is reported in its correct position thanks to the altimetric information that is derived from the DDTM, but the radiometric values that identify it are also repeated for the portions of scenes that were hidden from the sensor and which are "uncovered" after orthoprojection.

This lack of information, in a "rigorous" approach, is resolved by removing the radiometric value from images acquired from other points of view and whose orientations are known. These can be images from the same sensor or from another sensor (multi-sensor approach).

If no other data are available, it is possible to resolve this lack of information using a masking of the hidden areas with a background value that is easy to identify, as shown in figure 6 and figure 7 (test on a QuickBird image). In this way, the interpretation of orthoprojected satellite images is easier for unskilled users that are not deceived from erroneous data.
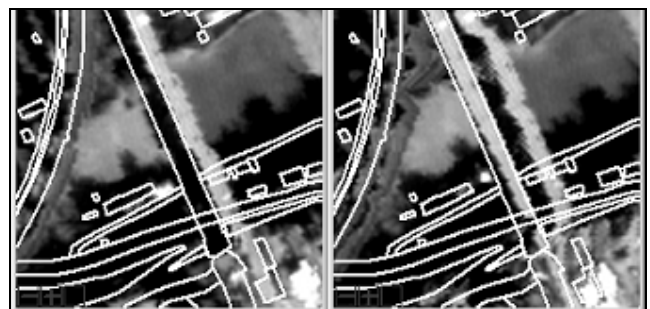

Figure 5 – Details that show the overlapping in correspondence to the viaduct using the DDEM (on the right) and the duplication of radiometric tones over the hidden area.
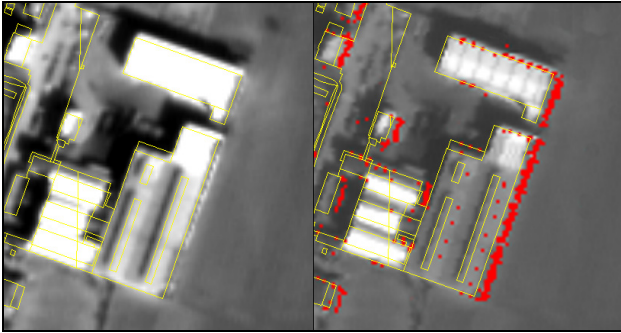
Figure 6 – An example of Eros A1 (1.9 m) orthoprojection. Duplication of the radiometric information (on the left) and masking of the same hidden area (on the right).
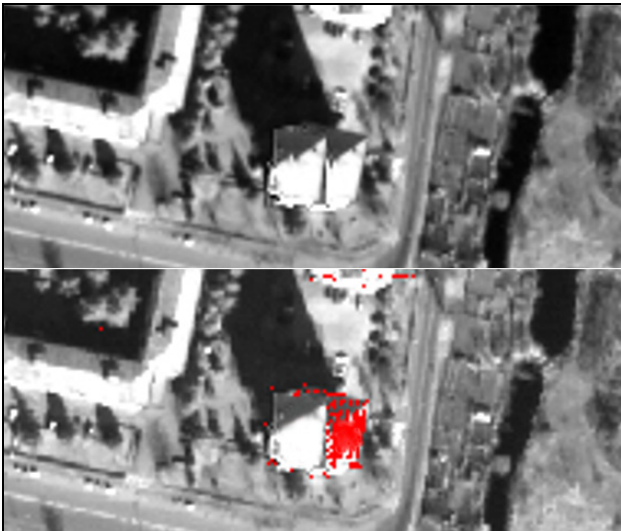

Figure 7 – An example of QuickBird (0.61m) orthoprojection. Duplication of the radiometric information (upper) and masking of the same hidden area (lower).

5. CONCLUSIONS

The here proposed case clearly shows the problems connected to the use of a precise approach for the orthoprojection of satellite images and confirms that altimetric height data of the buildings cannot be neglected, above all in urban areas. An analysis of the application fields of such approach is basically confirmed by the metadata supplied with the image. The mean view angle for the test reported in the previous section is about 10°. The entity of the displacements due to the mean height of the buildings, on the basis of the proposed graphs, is therefore not negligible for any orthophoto scale. Only the use of a DDEM can allow a correct resolution of the problem to be made. The suitability of the orthophoto at a certain scale in this way remains conditioned by only the quality of the projection model. The masking of the hidden areas makes the interpretation of orthoprojected satellite images easier, especially for unskilled users.

References:

Bello, M. G. , 1992, Enhanced training algorithms, and integrated training/architecture selection for multilayer perceptron networks, *IEEE Trans. on Neural Networks*, 3,864-875.

Boccardo, P., Borgogno Mondino, E., Giulio Tonolo, F., 2003, High resolution satellite images position accuracy tests, *IGARSS 2003*, Toulouse (proceedings on CD).

Dequal S., Lingua A., 2002, L'ortofoto di precisione del Comune di Torino, *Atti VI conferenza nazionale ASITA 2002*

Dowman, I., Tao, V., 2002, An update on the use of rational functions for photogrammetric restituition, *ISPRS, Vol. 7, N°3*, September 2002, pp. 26-29

Giulio Tonolo, F., Poli, D., 2003, Georeferencing of EROS-A1 high resolution images with rigorous and rational function model*, ISPRS , October 2003*, Hannover, Germany.

Howard Demuth, Mark Beale, MATLAB Neural Network Toolbox User's Guide, The Mathworks, Version 4, pp.5-8,5-30.

Lingua, A., Borgogno Mondino, E., 2002, *High Resolution Satellite Images Orthoprojection Using Dense DEM*, *Spie 2002 proceedings*, Creete

Open GIS, 1999, Abstract specification. Topic 7: The earth imagery case, Open GIS Consortium, March 1999

Rumelhart, D. E., Hinton, G. E. and Williams, R. J, Learning internal representations by error propagation, *In Parallel Distributed Processing*, vol 1, pp. 318-362. Cambridge, MA: MIT Press.

Tao, C., Hu, Y., 2001, A comprensive study of the Rational Function Model for Photogrammetric processing, Photogrammetric engineering & Remote Sensing, December 2001, pp.1347-1357

Toutin, T., 2004, Geometric processing of remote sensing images: models, algorithms and methods, International Journal of Remote Sensing, 24, 2003 (to be published)