Calibration of TV cameras through RBF networks

P. Cerveri[a,b], S. Ferrari[a,b], and N. A. Borghese[b]

[a] Centro di Bioingegneria
Fondazione ProJuventute Don Gnocchi
Politecnico di Milano, Italy
[b]Istituto Neuroscienze e Bioimmagini - C.N.R.
Via f.lli Cervi, 93 - 20090 Segrate (Milano) Italy
borghese@inb.mi.cnr.it

ABSTRACT

Distortions are introduced by standard TVcameras on the filmed images. These are due to optics and electronics and cause a systematic displacement of the image points with respect to their true position which is given by the geometrical perspective. In order to correct this error, in classical photogrammetrical approaches, the coordinates of the points are transformed through global polynomials whose coefficients are estimated from a set of reference points placed on regular grids. These approaches do not work well when, as in most of the cases, the distortions are highly irregular, and residual errors over the images can be relatively high. In this paper a solution based on RBF networks is proposed. It takes advantage of the quasi-local nature of the network elements to get a uniform small residual distortion error over all the TV image. The structural parameters of the network (namely their number and variance) are set according to criteria inspired to linear filtering theory and the weights are computed following a MAP criterion. Tests on simulated distortions and on real data have been carried. The results reported here show that the RBF networks achieve a better reduction of the distortions in all the tested conditions.

Keywords: RBF network, camera calibration, distortions.

1. INTRODUCTION

The reconstruction of the 3D geometrical information of a scene from 2D images acquired through CCD cameras requires camera calibration. This allows to establish an analytical relationships between a 3D point and its projection on the TVcamera image. To this scope, the camera is modelled as an ideal pinhole which is identified by two main sets of parameters: six extrinsic parameters define the pose of the camera with respect to an absolute reference system of coordinates and four intrinsic parameters take into account for the true position of principal point, the focal length and the image scale factor. Moreover, for very accurate 3D measurement, the systematic errors (distortions) introduced by the optical and electronic apparatus have to be compensated. This can be achieved by adding to the pinhole model a specific set of parameters which take into account distortions.

Classical approaches to camera calibration are based on the determination of a 3×4 transformation matrix whose coefficients are function of the extrinsic and intrinsic parameters. The entries of the matrix are determined starting from a set of 3D control points of known coordinates and their projection on the target of the TVcamera. Due to distortions these points are not measured in their true position. In metric photocameras this displacement is prevalently due to lenses and has been classified as radial and tangential distortions which produce typical pincushion or barrel distortions.[1] In standard CCD TVcameras, the offset of the principle point with respect to its true position and the misalignement of the lenses with the optical axes give rise to distortions which have very different shape from one TVcamera to the other.[2] The usual approach to correct these errors is to transform the measured coordinates through a polynomial whose coefficients can be determined from a set of control points usually positioned on a 2D grid positioned parallel to the TVcamera target.[3,4,5,6]

These appraoches are global in the sense that the effect of the polynomial extends over all the image plane. This is not efficient when very irregular distortions are produced by the lenses which can be seen in commercially available lenses, especially wide-angle. In this case, one would be able to correct locally this irregularities by a function which does not affect the neighbourhood. An algorithm of this kind has been proposed by Ferrigno et al.[7]. This algorithm is based on the partition of the image in a set of squared and to the application of a different polynomial to the points which lie in the different squares. This algorithm has proven to be very efficient in correcting whichever form of distortion but it has the drawback to introduce discontinuities at the border.

Here is proposed a new method for camera calibration which combines the use of local correcting functions with a fuzzy approach in the definition of the domain of influence of each function. This approach is based on a hierarchical radial basis function network approach recently proposed by Borghese et al.[7].

In Section 2 the distortion model used in the simulation is presented. In Section 3 the hierarchical RBF network model is briefly summarised and its use for distortion correction is justified. In Section 4 the method is described and the results obtained with simulated and real data are compared with those obtained with polynomial methods are reported.

2. MODEL FOR DISTORTION ERROR

The projection $p(x,y)$ of a 3D point, $P(X,Y,Z)$, on the image plane of the TVcamera, is described by the collinearity equations[1]:

$$
\begin{aligned}
x - x_o + e_x + d_x(x,y) &= \lambda_x f \frac{r_{11}(X-X_o) + r_{12}(Y-Y_o) + r_{13}(Z-Z_o)}{r_{13}(X-X_o) + r_{32}(Y-Y_o) + r_{33}(Z-Z_o)} \\
y - y_o + e_y + d_y(x,y) &= \lambda_y f \frac{r_{21}(X-X_o) + r_{22}(Y-Y_o) + r_{23}(Z-Z_o)}{r_{13}(X-X_o) + r_{32}(Y-Y_o) + r_{33}(Z-Z_o)}
\end{aligned}
\tag{1}
$$

where $p_o(x_o,y_o)$ is the image centre (principal points), $T(X_o,Y_o,Z_o)$ and $R\{r_{ij}\}$ express the position and orientation of the TVcamera with respect to an external reference system. d_x(x,y) and d_y(x,y) represent the distortion on the image plane, and (e_x, e_y) account for the random errors in the measurement (e.g. quantisation error). A common representation of the distortions is in terms of a radial and a decentering (tangential) component which have the following analytical shape[2]:

$$
\begin{aligned}
d_x(x,y) &= (x-x_o)(k_1 r^2 + k_2 r^4 + k_3 r^6) + a_1(r^2 + 2x^2) + 2a_2(x-x_o)(y-y_o) \\
d_y(x,y) &= (y-y_o)(k_1 r^2 + k_2 r^4 + k_3 r^6) + a_2(r^2 + 2y^2) + 2a_1(x-x_o)(y-y_o)
\end{aligned}
\tag{2}
$$

where k_1, k_2 , k_3 are the coefficients of radial lens distortion and a_1 and a_2 are the coefficients of decentering distortion. High order terms can be introduced and may be useful for very particular lenses like fish-eye lens, but for usual applications with zoom lenses the coefficients in Equation (2) are considered adequate[1] and more parameters may lead easily to overfitting which introduces spurious oscillations in the function $d(x,y)$. To estimate these coefficients, a set of reference landmarks (x_k,y_k) placed on a regular grid, parallel to the TVcamera are surveyed. From their nominal position and the position measured by the TVcamera, a displacement (distortion) can be computed for each landmark as: $d_k(x_k,y_k)$. From this set of displacement, the parameters in Equation (2) can be estimated.

This approach is based on the exact knowledge of the position of the principal point (x_o,y_o) and this is often not the case. Small deviations from the true position of the principal point, produce large errors making this approach very unstable. For this reason, we adopt two polynomials which are a function of the image coordinates expressed in cartesian notation to make distortion correction independent on the position of the principal point. According to the above considerations, a fourth order polynomial has been chosen:

$$
\begin{aligned}
d_x(x,y) = {} & c_{0x} + c_{1x}x + c_{2x}y + c_{3x}x^2 + c_{4x}y^2 + c_{5x}xy + c_{6x}x^3 + c_{7x}x^2y + c_{8x}xy^2 + \\
& c_{9x}y^3 + c_{10x}x^4 + c_{11x}x^3y + c_{12x}x^2y^2 + c_{13x}xy^3 + c_{14x}y^4 \\
d_y(x,y) = {} & c_{0y} + c_{1y}x + c_{2y}y + c_{3y}x^2 + c_{4y}y^2 + c_{5y}xy + c_{6y}x^3 + c_{7y}x^2y + c_{8y}xy^2 + \\
& c_{9y}y^3 + c_{10y}x^4 + c_{11y}x^3y + c_{12y}x^2y^2 + c_{13y}xy^3 + c_{14y}y^4
\end{aligned}
\tag{3}
$$

for a total of 30 parameters, 15 for each direction.

3. HIERARCHICAL RADIAL BASIS FUNCTION NETWORKS

Distorsion correction can be seen as a more general problem of function approximation where the function to be approximated is the distortion on the image plane. When there is not a good model of the distortion, a possible solution is to resort to non-parametric estimate. In particular, due to their generalisation capabilities, artificial neural networks (ANN), and in particular radial basis functions (RBF) networks, have been proposed as general approximators[8]. In the following we will show that these architecture can be effectively implemented to correct for very complex distortions.

A RBF gaussian network have the following analytical shape:

$$s(\mathbf{x}) = \sum_{k=1}^{M} \mathbf{w}_k g(\mathbf{x}; \mathbf{c}_k, \boldsymbol{\sigma}_k) \qquad (4)$$

where M is the number of units forming the network, and the sets $\{c_k\}$, $\{\sigma_k\}$ and $\{w_k\}$ are respectively the positions, the covariance matrices and the weights assigned to each Gaussian unit.

These parameters are subdivided into two sets, that we will call *structural parameters* (M, $\{c_k\}$, $\{\sigma_k\}$) and *synaptic weights*$\{w_k\}$. To simplify the estimate of the parameters, the Gaussian units are positioned on a regular grid. In this situation, Equation (4) is equivalent to a digital linear filter whose cut-off frequency is regulated by the value of σ: the larger is σ the smaller is the cut-off frequency and viceversa. If normalised Gaussians are used, Equation (4) can be written as follows:

$$s(x) = \sum_{k=1}^{M} w_k \frac{e^{-\frac{(x-c_x)^2}{\sigma^2}}}{\sqrt{\pi}\sigma} = \sum_{k=1}^{M} s_k \Delta c^2 \frac{e^{-\frac{(x-c_x)^2}{\sigma^2}}}{\sqrt{\pi}\sigma} \qquad (5)$$

where Δx is the spacing between two Gaussians. In order to get $s(x)$ close enough to the real function, σ should be small enough to get a cut-off frequency larger than the maximum frequency of the function to be reconstructed. However σ should not be too large to avoid aliasing, and the following empirical relationship is obtained[9]:

$$2\Delta x \frac{\sqrt{-\ln \delta_2}}{\pi} \leq \sigma \leq \frac{\sqrt{-\ln \delta_1}}{\pi \nu_M} \qquad (6)$$

This equation sets the constraint on the choice of the value of σ.

This method can be applicable also when the control points are not equally spaced. In this case a local maximum a-posteriori estimate of the distortions at the grid crossing can be carried out using a local Maximum A-Posteriori estimate:

$$s(c_k) = \frac{\sum_{x_r \in R} s(x_r) e^{-\frac{(x_r - c_k)^2}{\sigma_w^2}}}{\sum_{x_r \in R} e^{-\frac{(x_r - c_k)^2}{\sigma_w^2}}} \qquad (7)$$

where the region R is taken somehow arbitrary as that included inside two grid meshes: R : $c_k \pm \Delta c$ ($\Delta c = c_{k+1} - c_k$). The value of σ_W has been taken equal to $\sigma/2$ to avoid filtering of the data. This choice is based on the observation that the closer is a control point to s_k, the closer is its distortion to that measurable in the point c_k. This suggest to weight the data points in the neighbourhood of c_k by a function decreasing with the distance of that point from s_k. Assuming the weighting function Gaussian, the MAP estimator is obtained.[8]

This approach is somehow rigid as it allows only one value for the variance (and for the cut-off frequency) over all the input space. This is undesirable when local distortions are present as these usually show a high frequency content. When a single variance is used for the entire image, σ should be chosen small enough to reconstruct the finest details, although these may occur in few regions of the input space. The consequence is a dense packing of the Gaussian units (Equation 6) with a loss of resources and possibly overfitting. A better solution is to distribute more Gaussian units in those regions where they are effectively required (highest frequency) and to use less Gaussians in the others. This has been achieved using a two-layer approach, each layer having its own grid spacing and a characteristic variance. The first layer will output an approximation of the distortions at a high scale (large value of σ); and the other layer will contribute with the correction of the most irregular distortions adding Gaussians units where the residual error in their neighbourhood (defined as in Equation 8) exceeds a certain threshold. It should be remarked that with this kind of architecture, the obtained residual distortion is lowered uniformly under a predefined threshold. To achieve a better fitting, as suggested by Girosi et al.[8] the HRBF network can be integrated with a low order polynomial function.

4. RESULTS

Three sets of experiments have been carried out: two with simulated data and the third with real data. In the simulations an image of 256x256 pixels has been generated. Different type of distortions have been created by setting different values for the coefficients $\{k_i,a_i\}$ in Equation (2). The results obtained were very similar, therefore only the results obtained from one of the experiments are reported extensively. The parameters used

for this experiment are reported in the first column of Table I and their effect is evident from Figure 1a. Here a chessboard of 16x16 squares is deformed using the distortion coefficients defined above. As can be appreciated the distortions are quite large and somehow irregular.
To correct the effect of the distortions, the architecture described in Section 2 has been adopted. For each direction a gaussain network has been adopted. The first layer of the Hierarchical Network of Gaussians contained 13x11 Gaussian units, spaced by 36.00 with a $\sigma=52.75$. This first layer will allow to reconstruct spatial frequency up to 0.0035 pixel^{-1}. The second layer was constituted by Gaussian units spaced by 18.00 units with $\sigma=26.37$. This second layer will allow to reconstruct locally spatial frequencies up to 0.007 pixel^{-1}. These choices allow to correct for most of the distortions which can be found in real applications (Weng et al.[2]). At this stage, the number, the variance and the position of the Gaussians have been determined and only the value of the weights remains to be set in Equation (4). These weights are 143 for the first layer and a maximum of 323 for the second layer.

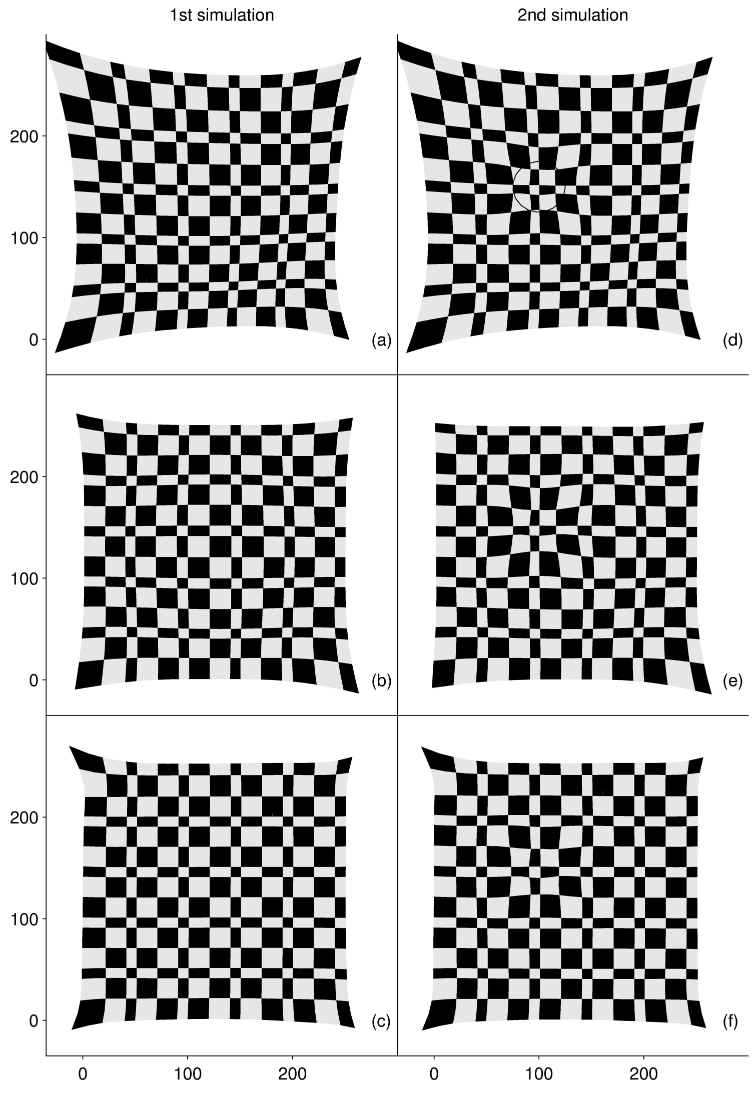

Figure 1. The results of the simulations are reported. In Figure a the deformation of a 16x16 grid through global distortion parameters is reported. It is represented as a chessboard. In figures b and c the grid after the distortion correction with the polynomial and the HRBF architecture is reported. The same results are reported in the second column in Figures 1d-1f for the simulations with Global and Local distortions.

To estimate the value of the weights, we suppose to acquire the coordinates of a set of 182 landmarks distributed over a regular grid (13x14), surveyed parallel to the image plane. For this set of landmarks, which will be called training set, the displacement due to distortions is computed according to Equation (2) with the parameters in Table I. From the displacement of this set of 182 markers the 143 weights are determined for the first layer through the MAP in Equation (7). A first approximation of the distortion function is obtained. In general, this approximation is sufficient in some regions of the input space, where the residual distortion on the training point is of the order of magnitude of the quantisation error, and higher in other regions. To reduce the error here, the Gaussian units of the second layer which lie inside these regions are activated. Their weights are computed through a MAP estimation carried out on the residual distortion error in that region and not on the original distortion value. This is because the first layer had already corrected for most of the distortion and the actual output value of the network is the sum of the output of the first and of the second layer.

The Hierarchical RBF method is compared with the polynomial function in Equation (3) whose coefficients are determined through a Least Mean Square estimate from the nominal coordinates at the grid crossing and their distorted coodinates (through Equation (2)).

To determine the accuracy which can be reached with the two methods, a set of different landmarks, not used for the estimation of the parameters, is considered. It will be called test set and it is constituted of 256 landmarks which are placed on the crossings of a 16x16 regular grid. For the sake of clarity, this test grid is plotted as black and white squares in Figures 1a-f. The transformation in Equation (2) is applied with the parameters in Table I also to this test set such that the landmarks will be affected by the same distortions of the one in the training set. After correction the residual distortion is taken as a measure of the efficacy of both methods. This is measured through the mean and standard deviation of the difference between the nominal position of all the test landmarks and their position after distortion correction.

Results are reported at the bottom of the first column of Table I and Figures 1b and 2c. These show that HRBF network is consistently better than polynomial correction. The result did not change if polynomial of higher order were considered. On the contrary, the accuracy achieved though polynomial correction started to degrade after fourth order in most of the cases.

The same results were true also when local distortions were added to the global ones and are reported in the second column of Table I. Local distortions were generated by choosing a point on the target p(x,y) around which local distortions are generated, and by computing the displacement through Equation (2) where the point p(x,y) is substituted to po(xo,yo). To keep this error local, the resulting displacement is weighted with a gaussian function centered in p(100.02,150.50) with standard deviation equal to 25 target units. These local distortions are added to the global ones and their effect is evident from 1d where they have been added to the distortions reported in Figure 1a. Local distortions are added in the region enclosed in the circle marked with a thick line (Figure 1d). The parameters used for the local distortions are reported in the second column of Table I and the results in Figures 1e and 1f.

Table 1.Setup and results of the simulation.

Simulation 1 (only Global distortion)		Simulation 2 (Global + Local distortion)	
Global distortion	Local distortion	Global distortion	Local distortion
k1 2.0161e-07	k1 0	k1 2.0161e-07	k1 -4.032e-05
k2 -6.0483e-11	k2 0	k2 -6.0483e-11	k2 -6.048e-07
k3 3.0219e-15	k3 0	k3 3.0219e-15	k3 2.0161e-11
a1 8.4677e-05	a1 0	a1 8.4677e-05	a1 -4.435e-03
a2 7.6612e-05	a2 0	a2 7.6612e-05	a2 0
Polynomial correction 0.96±0.63		Polynomial correction 1.30±1.20	
HBRF correction 0.20±0.17		HBRF correction 0.47±0.0.56	

The better accuracy reached through HRBF correction is confirmed by the experiment of real data. For this test the Elite system has been adopted.[7] This is an automatic motion analyser which recognises landmarks constituted of hemispheric plastic support covered with retroreflective material. With the Elite system a grid of 13x14 landmarks is filmed parallel to the TVcamera and the parameters of the two-layer HRBF network and the coefficients of the polynomial are estimated. The accuracy was carried out on the same grid used to estimate theparameters and the results are the following: residual error with HRBF network: (0.085±0.0484) and residual error with polynomial corrections: (0.1±0.0542).

5. DISCUSSION AND CONCLUSION

From simulated and real data, HRBF network has shown to be a very flexible method to correct distortions. Thanks to their quasi-local nature, they can take into account not only very different type of distortions both local and global. The residual error which can be achieved is overall smaller than that obtained with polynomial correction. The advantage of this approach is emphasized when huge set of data points are used. In fact solving Equation (4) with SVD is feasible only in case of low cardinality of the data point set. HRBF network can be extended from point-based distortion correction to line-based approach in which a great number of points have to be matched.

REFERENCES

1. P.R. Wolf, *Elements of photogrammetry*. New York, McGraw-Hill, 1983.
2. J. Weng, P. Cohen, and M. Henriou, Camera calibration with distortion models and accuracy evaluation *IEEE Trans. Pattern. Analysis and Machine Intellingence.*, vol. 14, no. 10, pp. 965-979, 1992.
3. M. Li and J.M. Lavest, *Some aspects of zoom lens* camera *calibration, IEEE Trans. on Pattern Analysis and Machine Intelligence,* vol. 18, no. 11, pp. 1105-1110, 1996.
4. Y. Nomura, M. Sagara, H. Naruse, and A. Ide, *Simple calibration algorithm for high-distortion -lens camera, IEEE Trans. on Pattern Analysis and Machine Intelligence,* vol. 14, no. 11, pp. 1095-1099, 1992.
5. G-Q Wei and S. D. Ma, *Implicit and explicit camera calibration: theory and experiment, IEEE Trans. on Pattern Analysis and Machine Intelligence,* vol. 16, no. 5, pp. 469-480, 1994.
6. B. Prescott and G.F McLean, *Line-based correction of radial lens distortion, Grafical Models and Image Processing,* vol. 59, no. 1, pp. 39-47, 1997.
7. G. Ferrigno, N.A. Borghese and A. Pedotti. *Pattern recognition in 3D automatic human motion analysis, ISPRS J. of Photogr. and Rem. Sens.*, vol. 45, pp. 227-246 1990.
8. F. Girosi, M. Jones and T. Poggio, Regularization theory and Neural Networks Architectures, *Neural Computation* 7. pp. 219-269, 1995
9. N.A. Borghese and S. Ferrari, *Hierarchical RBF networks and local parameter estimate,* submitted to Neural Computing.
10. S. Shah and J. K. Aggarwal, Intrinsic parameter calibration procedure for a (high distortion) fish-eye lens camera with distortion model and accuracy estimation. *Pattern Reconition* vol. 29, no. 11, pp. 1775-1788, 1996.
11. R. K. Lenz and R. Y. Tsai, Tecniques for calibration of the scale factor and image center for high accuracy 3D machine vision metrolog, *IEEE Trans. Pattern. Analysis and Machine Intelligence* vol. 10, no.5, pp. 713-720, 1988.
12. R.C. Gonzales and R.E. Woods, *Digital Image Processing* (Addison-Wesley, 1993).