Matching Multiple Views by the Least Square Correlation

Georgy Gimel'farb and Jian Zhong

Centre for Image Technology and Robotics,
Department of Computer Science
Tamaki Campus, University of Auckland
Private Bag 92019, Auckland 1, New Zealand

Abstract. We consider the potentialities of matching multiple views of a 3D scene by the least square correlation provided that relative projective geometric distortions of the images are affinely approximated. The affine transformation yielding the (sub)optimal match is obtained by combining an exhaustive and directed search in the parameter space. The directed search is performed by a proposed modification of the Hooke–Jeeves unconstrained optimization. Experiments with the RADIUS multiple-view images of a model board show a feasibility of this approach.

1 Introduction

Generally, the uncalibrated multiple-view 3D scene reconstruction involves a set of images with significant relative geometric distortions (because of different exterior and interior parameters of cameras ised for image acquisition). This complicates the search for initial stereo correspondences for starting an iterative process of simultaneous cameras calibration and 3D surface recovery [5, 7, 8]. If the images form a sequence such that each neighbouring pair has rather small geometric deviations, then the search for correspondences is usually reduced to detection of identical points-of-interest (POI) such as corners [5, 8]. But generally due to significant geometric and photometric image distortions, the identical POIs may not be simultaneously detected in different images. Therefore it is more reliable to directly match large image areas by taking account of possible relative distortions.

We restrict our consideration to the simplified case when image distortions can be closely approximated by affine transformations [6, 10]. Then the least square correlation [2, 3, 4] can be used for finding a transformation that yields the largest cross-correlation of the images.

The least square correlation is widely used in computational binocular stereo if relative geometric distortions in a stereo pair are comparatively small [3, 4]. In this case, although the correlation function is generally multimodal, the gradient (steepest ascent) search is used to find the maximum correlation [2]. Such a search is based on normal equations obtained by linear approximation of the cross-correlation function in the vicinity of a starting point in the space of affine parameters.

The straightforward gradient search is not workable in the multiple–view case because of larger relative distortions of the images. The globally optimum match can be, in principle, found by exhausting all the values of affine parameters in a given range of possible distortions. But this is not computationally feasible.

In this paper we consider more practical (but only suboptimum) approach combining an exhaustion of some affine parameters over a sparse grid of their values with a directed search for all the parameters starting from every grid position. The directed search is based on a modified Hooke–Jeeves unconstrained optimization [1]. The proposed modification is intended to take account of the multi-modality of cross-correlation. Feasibility of the proposed approach is illustrated by experiments with the RADIUS multiple–view images of a 3D model board scene [9].

2 Basic notation

Let $\mathbf{R}_j$ be a finite arithmetic lattice supporting a greyscale image $g_j : \mathbf{R}_j \to \mathbf{G}$ where $\mathbf{G}$ is a finite set of grey values. Let $(x, y) \in \mathbf{R}_j$ denote a pixel with the column coordinate x and row coordinate y. For simplicity, the origin $(0, 0)$ of the (x, y)-coordinates is assumed to coincide with the lattice centre.

Let g_1 be a rectangular prototype matched in the image g_2 to a quadrangular area specified by an affine transformation $\mathbf{a} = [a_1, \ldots, a_6]$. The transformation relates each pixel (x, y) in the prototype g_1 to the point $(x_\mathbf{a}, y_\mathbf{a})$ in the image g_2:

$$\begin{aligned} x_{\bar{\mathbf{a}}} &= a_1 x + a_2 y + a_3; \\ y_{\bar{\mathbf{a}}} &= a_4 x + a_5 y + a_6. \end{aligned} \tag{1}$$

The affine parameters (a_1, a_5), (a_2, a_4), and (a_3, a_6) describe, respectively, the x- and y-scaling, shearing, and shifting of g_2 with respect to g_1.

Grey levels $g_2(x_\mathbf{a}, y_\mathbf{a})$ in the points with non-integer coordinates $(x_\mathbf{a}, y_\mathbf{a})$ are found by interpolating grey values in the neighbouring pixels of the lattice $\mathbf{R}_2$. If the transformed point $(x_\mathbf{a}, y_\mathbf{a})$ falls outside of the lattice, then the original pixel $(x, y) \in \mathbf{R}_1$ is assumed to be excluded from matching.

The least square cross-correlation

$$C(\mathbf{a}^*) = \max_{\mathbf{a}} \{C(\mathbf{a})\} \tag{2}$$

maximizes by the affine parameters $\mathbf{a}$ the conventional cross-correlation

$$C(\mathbf{a}) = \sum_{(x,y) \in \mathbf{R}_{1,\mathbf{a}}} \frac{g_1(x, y) - m_1}{s_1} \cdot \frac{g_2(x_\mathbf{a}, y_\mathbf{a}) - m_{2,\mathbf{a}}}{s_{2,\mathbf{a}}} \tag{3}$$

between the prototype g_1 and affinely transformed image g_2. Here, m and s are the mean values and standard deviations, respectively:

$$m_1 = \frac{1}{|\mathbf{R}_{1,\mathbf{a}}|} \sum_{(x,y) \in \mathbf{R}_{1,\mathbf{a}}} g_1(x, y); \qquad s_1^2 = \frac{1}{|\mathbf{R}_{1,\mathbf{a}}|} \sum_{(x,y) \in \mathbf{R}_{1,\mathbf{a}}} (g_1(x, y) - m_1)^2;$$
$$m_{2,\mathbf{a}} = \frac{1}{|\mathbf{R}_{1,\mathbf{a}}|} \sum_{(x,y) \in \mathbf{R}_{1,\mathbf{a}}} g_2(x_\mathbf{a}, y_\mathbf{a}); \; s_{2,\mathbf{a}}^2 = \frac{1}{|\mathbf{R}_{1,\mathbf{a}}|} \sum_{(x,y) \in \mathbf{R}_{1,\mathbf{a}}} (g_2(x_\mathbf{a}, y_\mathbf{a}) - m_{2,\mathbf{a}})^2$$

and $\mathbf{R}_{1,\mathbf{a}} = \{(x,y) \ : \ ((x,y) \in \mathbf{R}_1) \wedge ((x_\mathbf{a}, y_\mathbf{a}) \in \mathbf{R}_2)\}$ denotes the sublattice which actually takes part in matching the prototype g_1 and the affinely transformed image g_2.

3 Combined search for suboptimal affine parameters

To approach the least square correlation in Eq. (2), we use the following combined exhaustive and directed search in the affine parameter space. For each given prototype g_1, a sparse grid of the relative shifts $a_3^{[0]}$ and $a_6^{[0]}$ of the matching area in the image g_2 is exhausted. Starting from each grid position, the modified Hooke–Jeeves directed optimization [1] is used to maximize the cross-correlation $C(\mathbf{a})$ by all six affine parameters. The largest correlation over the grid provides the desired affine parameters $\mathbf{a}^*$ of the (sub)optimal match.

The modified Hooke–Jeeves optimization consists of the following two successive stages which are repeated iteratively while the correlation value $C(\mathbf{a})$ continues to increase. Each parameter a_i, $i = 1,\ldots,6$, varies in a given range $[a_{i,\min}, a_{i,\max}]$, and the search starts with the initial parameter values $\mathbf{a}^{[0]} = [1, 0, a_3^{[0]}, 0, 1, a_6^{[0]}]$.

1. *Exploration stage.* At each step $t = 1,2,\ldots,T$, the locally best parameter, $j^{[t]}$, is chosen by changing each parameter $i \in \{1,\ldots,6\}$ under the fixed values, $[a_k^{[t-1]} : \ k \neq i; k \in \{1,\ldots,6\}]$, of other parameters. The choice yields the largest increase of the correlation $C(\mathbf{a}^{[t]})$ with respect to $C(\mathbf{a}^{[t-1]})$ providing the parameters $\mathbf{a}^{[t]}$ and $\mathbf{a}^{[t-1]}$ differ by only the value of the locally best parameter $j^{[t]}$. The exploration steps are repeated while the cross-correlation $C(\mathbf{a}^{[t]})$ increases further.
2. *Search stage.* The affine parameters $\mathbf{a}_\lambda = \mathbf{a}_T + \lambda\mathbf{d}$ are changed in the conjectured direction $\mathbf{d} = \mathbf{a}^{[T]} - \mathbf{a}^{[0]}$ of the steepest increase while the cross-correlation $C(\mathbf{a}_\lambda)$ increases further.

Each exploration step exhausts a given number L of the equispaced parameter values in their range to approach the local correlation maximum along a parameter axis, given the fixed previous values of all other parameters. In the experiments below L = 3,...,15. The quadratic approximation of these L correlations provides another possible position of the local maximum. The maximum of the $L+1$ values found is then locally refined using small increments $\pm\delta_i$ of the parameter.

The exploration steps converge to a final local maximum value $C(\mathbf{a}^{[T]})$, and the parameters $\mathbf{a}^{[T]}$ allow for inferring the possible steepest ascent direction in the parameter space. The search along that direction refines further the obtained least square correlation.

The proposed algorithm replaces the coordinate-wise local search for the closest correlation maximum of the original Hooke–Jeeves exploration stage with the combined exhaustive and directed search. This allows to roughly take into account the multimodal character of the cross-correlation function in the parameter space and escape some non-characteristic minor modes. Figure 1 shows

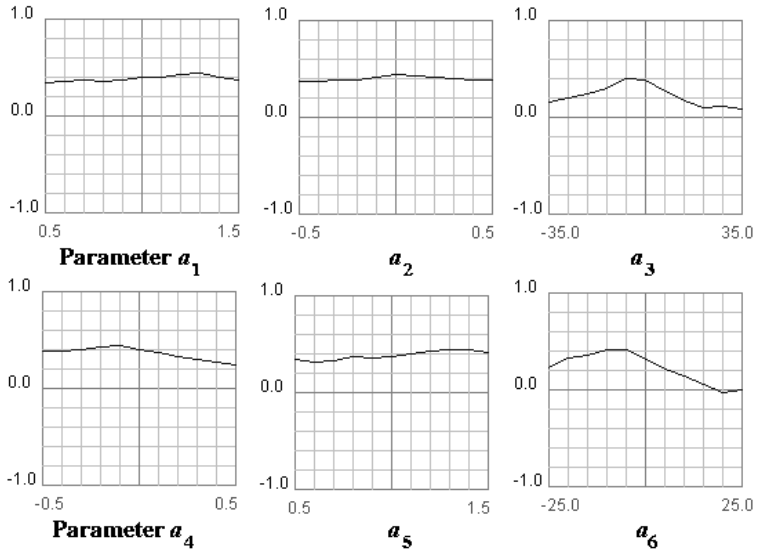

Fig. 1. Typical cross-correlations at the exploration steps.

the typical multi-modal dependence of the cross-correlation from a single affine parameter, given the fixed values of other affine parameters.

Comparing to the conventional least square matching [2], our algorithm does not linearize the correlation function in the parameter space and hence does not build and use the normal equation matrix. This latter is usually ill-conditioned because it depends on the image derivatives with respect to affine parameters.

4 Experiments with the RADIUS images

Image pairs M15–M28, M24–M25, and M29–M30 selected for experiments from the RADIUS-M set [9] are shown in Figure 2. The images of size 122×96 and 244×192 represent, respectively, the top and the next–to–top levels of image pyramids. Each pyramid is built by reducing the original image 1350×1035 to 976×768 at the first level and then by the twofold demagnification at each next level of the pyramid.

Some results of matching the top–level image pairs in Figure 2 using the rectangular prototype windows of size 49×81 are shown in Table 1 and Figure 3. The prototype g_1 is placed to the central position $(61, 48)$ in the initial image. In these experiments, the search grid 5×5 of step 5 in both directions is centered to the same position $(61, 48)$ in the other image (that is, the shift parameters for the central grid point are $a_3^{[0]} = 0$ and $a_6^{[0]} = 0$), and the parameter $L = 11$.

In all our experiments the ranges of the affine parameters a_1, a_4 and a_2,a_3

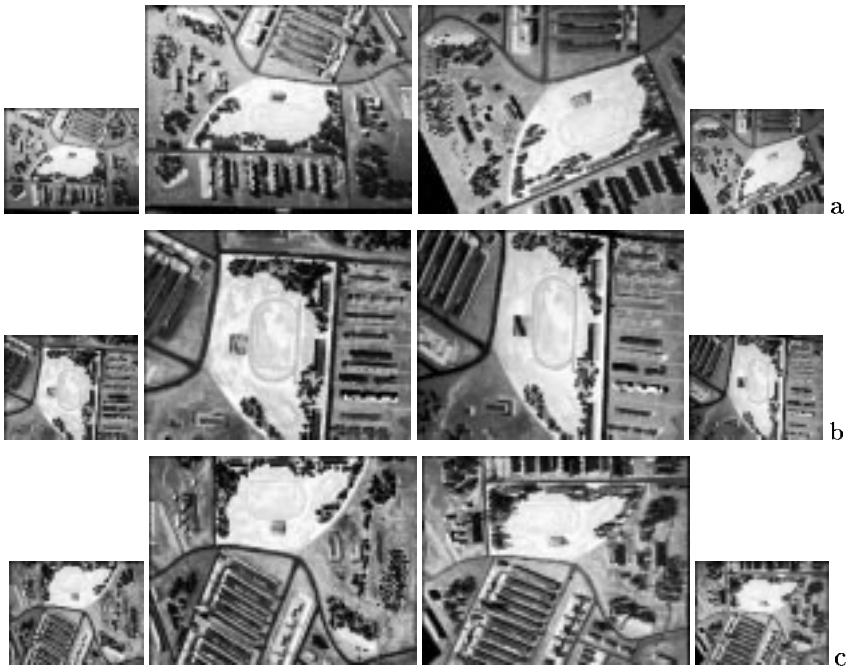

Fig. 2. RADIUS images M15–M28 (a), M24–M25 (b), and M29–M30 (c) at the top and next–to–top pyramid level.

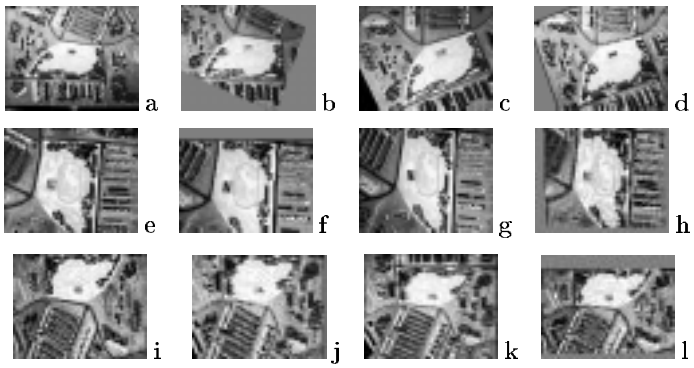

Fig. 3. Initial top–level RADIUS images M15 (a), M28 (c), M24 (e), M25 (g), M29 (i), M30 (k) and the affinely transformed images M28 (b), M15 (d), M25 (f), M24 (h), M30 (j), M29 (l) adjusted to M15, M28, M24, M25, M29, M30, respectively, with the parameters presented in Table 1.

Table 1. The least square correlation values and the corresponding affine parameters obtained by matching the top–level RADIUS images M15–M28, M24–M25, and M29–M30 in Figure 3.

Transformation	$C(\mathbf{a}^*)$	Affine parameters $\mathbf{a}^*$					
		a_1^*	a_2^*	a_3^*	a_4^*	a_5^*	a_6^*
M28 to M15	0.62	1.03	0.35	18.4	-0.50	1.05	5.6
M15 to M28	0.52	0.90	-0.30	-11.0	0.30	0.70	-7.0
M25 to M24	0.66	0.90	0.00	2.0	0.00	1.00	-10.8
M24 to M25	0.79	1.12	-0.02	-3.2	-0.02	0.90	7.5
M30 to M29	0.69	1.00	0.00	2.0	-0.01	0.77	3.0
M29 to M30	0.64	0.97	0.00	-2.0	0.00	1.23	-5.0

Table 2. Characteristics of the prototype windows to be matched, the least square correlation values, and the affine parameters found by matching.

Fig. 4	Window	Position		Search grid		L	$C(\mathbf{a}^*)$	Affine parameters $\mathbf{a}^*$					
	size	in g_1	in g_2	size	step			a_1^*	a_2^*	a_3^*	a_4^*	a_5^*	a_6^*
a	50×35	85,55	90,50	5×5	10	3	0.67	1.11	0.50	11.0	-0.50	1.23	14.0
b	50×35	94,53	97,57	5×5	10	3	0.69	1.10	0.50	10.0	-0.50	1.24	9.0
c	50×35	90,50	95,55	5×5	10	3	0.69	1.09	0.50	8.3	-0.50	1.24	10.0
d	50×35	80,65	100,65	5×5	10	3	0.69	1.09	0.50	15.2	-0.50	1.20	19.0
e	50×35	94,53	105,65	5×5	10	3	0.70	1.09	0.50	10.0	-0.50	1.24	8.9
f	50×35	97,56	109,69	5×5	10	3	0.71	1.08	0.50	11.8	-0.50	1.25	8.0
g	50×35	100,59	110,70	5×5	10	15	0.72	1.10	0.50	14.0	-0.50	1.21	8.0
h	50×35	99,58	111,71	5×5	10	3	0.72	1.10	0.50	13.0	-0.50	1.23	7.8
i	50×35	100,59	110,70	3×3	10	7	0.73	1.13	0.50	14.0	-0.54	1.28	7.7
j	50×50	100,59	110,70	1×1	10	11	0.73	1.10	0.50	14.0	-0.50	1.24	7.0
k	50×35	100,59	110,70	5×5	10	3	0.73	1.10	0.50	14.0	-0.50	1.26	7.4
l	50×35	99,58	110,70	5×5	10	3	0.73	1.11	0.50	13.4	-0.50	1.22	7.9

are $[0.5, 1.5]$ and $[-0.5, 0.5]$, respectively. The ranges of the parameters a_3 and a_6 are given by the width and height of the chosen prototype window.

In these cases, the least square correlation matching allows, at least as a first approximation, to relatively orient all the three image pairs. For comparison, Table 4 presents results of matching the top–level images M15 and M28 using the two larger prototype windows of size 71×51 and 81×61 placed to the central position $(60, 50)$ in M15. Here, the search grid 5×5 of step 10 is sequentially centered to the nine neighbouring positions $(60 \pm 1, 50 \pm 1)$ in M28. Although the photometric distortions of the images are non-uniform, the median values of the obtained affine parameters $[a_1^*, \ldots, a_6^*]$ for the confident matches with $C(\mathbf{a}^*) \geq 0.55$ are quite similar to the like parameters in Table 1.

Tables 2 – 3 and Figures 4 – 5 show results of matching the images M15 and

Table 3. Characteristics of the prototypes to be matched, the least square correlation values, and the affine parameters found by matching.

Fig. 4	Window	Position		Search grid		L	$C(\mathbf{a}^*)$	Affine parameters $\mathbf{a}^*$					
	size	in g_1	in g_2	size	step			a_1^*	a_2^*	a_3^*	a_4^*	a_5^*	a_6^*
a	50×35	100,60	100,60	5×5	15	5	0.49	0.54	0.06	31.0	0.21	0.86	-2.5
b	50×50	110,70	100,60	5×5	10	3	0.52	0.75	-0.28	-12.0	0.12	1.00	-4.0
c	50×50	100,65	100,65	5×5	10	15	0.53	0.83	-0.29	-7.0	0.14	0.97	-6.3
d	75×50	100,65	100,65	5×5	10	15	0.53	0.83	-0.29	-7.0	0.14	0.97	-6.3
e	50×35	100,60	100,60	5×5	10	15	0.68	0.75	-0.36	-10.0	0.32	0.66	-10.0
f	50×50	100,60	100,60	5×5	10	15	0.69	0.74	-0.32	-9.0	0.29	0.68	-10.0
g	50×50	110,70	100,60	5×5	10	11	0.71	0.77	-0.30	-14.0	0.30	0.70	-10.0
h	50×50	110,70	100,60	5×5	10	5	0.71	0.75	-0.30	-14.0	0.31	0.69	-10.0
i	50×35	100,65	100,65	5×5	10	15	0.71	0.75	-0.29	-10.0	0.32	0.64	-11.0
j	50×50	100,59	110,70	1×1	10	11	0.72	0.76	-0.28	-14.0	0.33	0.67	-10.3
k	50×35	110,70	110,70	5×5	10	3	0.72	0.77	-0.31	-14.0	0.32	0.68	-10.0
l	50×35	110,70	100,60	5×5	15	5	0.72	0.76	-0.31	-14.0	0.32	0.68	-10.0

Table 4. Central positions of the search grid in M28, the least square correlation values, and the affine parameters found by matching M28 to M15.

Window size	Position in M28	$C(\mathbf{a}^*)$	Affine parameters $\mathbf{a}^*$					
			a_1^*	a_2^*	a_3^*	a_4^*	a_5^*	a_6^*
71×51	59,49	0.66	1.10	0.50	18.0	-0.50	1.21	3.0
	60,49	0.59	1.10	0.40	18.0	-0.50	1.10	5.0
	61,49	0.59	1.10	0.40	18.0	-0.50	1.10	5.0
	59,50	0.67	1.09	0.48	18.0	-0.50	1.20	3.0
	60,50	0.67	1.09	0.48	18.0	-0.50	1.20	3.0
	61,50	0.67	1.09	0.48	18.0	-0.50	1.20	3.0
	59,51	0.54	1.10	0.30	19.0	-0.50	1.00	7.0
	60,51	0.52	1.00	0.30	20.0	-0.56	1.00	7.0
	61,51	0.53	1.00	0.40	20.0	-0.56	1.28	4.0
Median parameter values for the confident matches			1.09	0.48	18.0	-0.50	1.20	4.0
81×61	59,49	0.55	1.02	0.50	16.0	-0.40	1.10	6.0
	60,49	0.58	1.10	0.44	18.0	-0.50	1.36	5.0
	61,49	0.65	1.10	0.44	18.0	-0.50	1.20	3.0
	59,50	0.55	1.02	0.50	16.0	-0.40	1.10	6.0
	60,50	0.58	1.10	0.44	18.0	-0.50	1.36	5.0
	61,50	0.65	1.10	0.44	18.0	-0.50	1.20	3.0
	59,51	0.55	1.02	0.50	16.0	-0.40	1.10	6.0
	60,51	0.57	1.00	0.50	15.4	-0.50	1.20	4.0
	61,51	0.60	1.05	0.50	16.0	-0.50	1.20	4.0
Median parameter values for the confident matches			1.10	0.50	16.0	-0.50	1.20	5.0

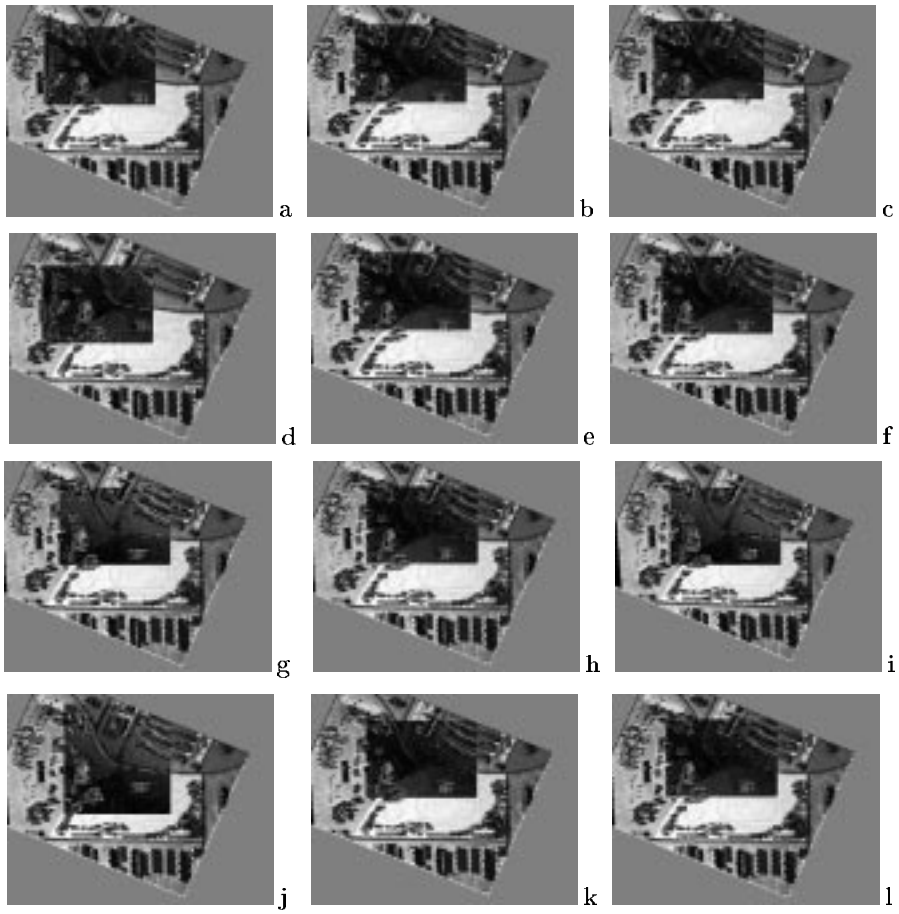

Fig. 4. Next–to–top image M28 affinely adjusted to M15 using the affine parameters in Table 2. The dark rectangles show positions of the prototype windows and the grey-coded values of the residual pixel-wise errors for the least square correlation matching.

M28 on the next–to–top level of the pyramids. Here, different prototype windows and various search grids and approximation orders are compared. The position of the prototype window with respect to the image is shown by a dark rectangle giving the grey–coded residual pixel–wise errors of matching (the darker the pixel, the smaller the error).

The matching results are mostly similar although in the general case they depend on the search characteristics, in particular, on the chosen search grid and the parameter L (e.g., Figures 5,a,b,e,g, and the corresponding data in Table 3). Also, the larger prototype windows may affect the precision of the affine

Fig. 5. Next–to–top image M15 affinely adjusted to M28 using the affine parameters in Table 2. The dark rectangles show positions of the prototypes and the grey-coded values of the residual pixel-wise errors for the least square correlation matching.

approximation of actually projective image distortions (Figures 5,c,d,i).

The median values of the obtained affine parameters $[a_1^*, \ldots, a_6^*]$ for the seven best matches are as follows:

$g_1 - g_2$	a_1^*	a_2^*	a_3^*	a_4^*	a_5^*	a_6^*
M15–M28:	1.10	0.50	14.0	-0.50	1.24	7.9
M28–M15:	0.75	-0.30	-14.0	0.32	0.68	-10.0

These values are close to the parameters found by matching the top–level images so that the fast top–level matching can provide a first approximation of the relative geometric distortions of these images to be refined at the next levels

of the image pyramids. Similar results are also obtained for the image pairs M24–M25 and M29–M30 as well as for other RADIUS images (e.g., M24–M10, M10–M11, M11–M19, M19–M20, M8–M9, M9–M23, M23–M29, M30–M36, etc).

5 Concluding remarks

These and other experiments show that the proposed modified Hooke–Jeeves optimisation algorithm permits us to successfully match large-size areas in the multiple-view images of a 3D scene by the least square correlation, provided the relative image distortions can be affinely approximated. This approach has a moderate computational complexity, hence in principle it can be used at the initial stage of the uncalibrated multiple-view terrain reconstruction.

The approach exploits almost no prior information about a 3D scene, except for the ranges of the affine parameters for matching. Also, the final cross-correlation value provides a confidence measure for the obtained results: if the correlation is less than or equal to $0.5 - 0.55$, one may conclude that the matching fails, otherwise the larger the correlation, the higher the confidence and, in the most cases, the better the affine approximation of the relative image distortions.

References

1. Dennis J.E., Schnabel R.B.: *Numerical Methods for Unconstrained Optimization and Nonlinear Equations.* Prentice-Hall, Englewood Cliffs (1983)
2. Förstner, W.: Image matching. In Haralick, R.M., Shapiro, L.G.: *Computer and Robot Vision*, vol.2, ch.16, p.p.289–378. Addison-Wesley (1993).
3. Hannah, M. J.: Digital stereo image matching techniques. *International Archives on Photogrammetry and Remote Sensing*, 27 (1988) 280–293.
4. Helava, U. V.: Object space least square correlation. *Photogrammetric Engineering and Remote Sensing*, 54 (1988) 711–714.
5. Koch, R. and van Gool, L., eds.: *3D Structure from Multiple Images of Large-Scale Environments (European Workshop, SMILE'98, Freiburg, Germany, June 6-7, 1998, Proc.)*, Lecture Notes in Computer Science 1506. Springer (1998).
6. Koenderink, J., van Doorn, A.: Affine structure from motion. *Journal of the Optical Society of America*, 8:2 (1991) 377–382.
7. Maybank, S.J., Faugeras, O.: A theory of self-calibration of a moving camera. *International Journal of Computer Vision*, 8 (1992) 123–152.
8. Pollefeys, M., Koch, R., van Gool, L.: Self-calibration and metric reconstruction in spite of varying and unknown internal camera parameters. *Int. Journal of Computer Vision*, 32:1 (1999) 7–25.
9. *RADIUS model board imagery and ground truth* (CD-ROM). Intelligent Machines Lab., University of Washington. Seattle (1996).
10. Shapiro, L.S.: *Affine Analysis of Image Sequences.* Cambridge University Press (1995).

This article was processed using the LaTeX macro package with LLNCS style