

Image Matching with Spatially Variant Contrast and Offset: A Quadratic Programming Approach

Alexander Shorin, Georgy Gimel'farb, Patrice Delmas, and Jonh Morris

Department of Computer Science, University of Auckland
Private Box 92019, Auckland 1142, New Zealand

al@cs.auckland.ac.nz, {g.gimelfarb,p.delmas,j.morris}@auckland.ac.nz

Abstract. Template matching is widely used in machine vision, digital photogrammetry, and multimedia data mining to search for a target object by similarity between its prototype image (template) and a sensed image of a natural scene containing the target. In the real-world environment, similarity scores are frequently affected by contrast / offset deviations between the template and target signals. Most of the popular least-squares scores presume only simple smooth deviations that can be approximated with a low-order polynomial. This paper proposes an alternative and more general quadratic programming based matching score that extends the conventional least-squares framework onto both smooth and non-smooth signal deviations.

1 Introduction

Template matching is a crucial part of autonomous computer vision systems that detect, classify or retrieve natural objects under varying imaging conditions and operating environments. Inherent inflexibility of many matching algorithms is often exacerbated by poor expressiveness of their underlying models of noise (here 'noise' is an umbrella term embracing all sources that cause differences between corresponding target and template signals). Common models of image noise relate pixel-wise signal differences to interdependent spatially invariant or variant contrast factors and offsets and independent additive random differences with a known and typically central-symmetric probability distribution.

Most of the state-of-the-art matching methods support only smooth (i. e. mildly non-linear) contrast and offset deviations originating from varying illumination conditions [9,2]. Although such models are more or less appropriate for matching images of planar surfaces, they are inadequate as far as most three-dimensional objects are concerned. 3D geometry of those objects can create shadows following highly non-smooth contrast functions. To appreciate the sheer complexity of noise modelling, let us consider a rather simplistic case when additive noise for each individual pixel is a linear function of signals in the two adjacent neighbours. An exact (lossless) model of this complex noise function would require approximately $3N$ parameters where N is the total number of pixels. Comparing this model to the one in [9] where only five parameters are used clearly shows that the low-order polynomial model may not be satisfactory.

The model proposed in this paper deals with a much larger search space of noise parameters than any currently existing template matching algorithm. Specifically, it involves approximately $6N$ parameters that appear in a constrained least-squares matching problem with up to six constraints per pixel-wise signal to be solved with quadratic programming (QP). The resulting matching algorithm can successfully handle inhomogeneous contrast deviations over the target image that are introduced by a variety of noise sources including shadowing of complex non-planar objects.

The rest of the paper is structured as follows. Section 2 describes previous work done in this area and shows how our proposal fits into the existing body of knowledge. Section 3 presents the quadratic programming based matching algorithm. Experimental results and conclusions are covered in Sections 4 and 5.

2 Previous Work

In order to better understand main strengths and limitations of known contributions and novelty of the proposed approach, let us introduce a somewhat arbitrary dichotomy of contrast / offset deviations between the target and template images: *smooth* (or homogeneous) versus *non-smooth* (non-homogeneous). The former may be encountered due to inhomogeneities of an image sensor or because of illumination variations over planar objects. Three possible smooth functions are shown in Fig. 1. The existing matching methods [8,9] can handle this type of the deviations. The latter (non-homogeneous) signal deviations frequently arise when dealing with non-planar real-world objects because even a slight difference in the lighting angle can produce highly non-smooth contrast patterns like those shown in Fig. 2.

As a delineation criterion for these two broad categories of noise, a polynomial function of some low order p such that $p \ll N$ is adopted as the upper bound for the complexity which can be categorised as homogeneous. The choice is due to the fact that most of the known template matching methods have $p \leq 2$ as

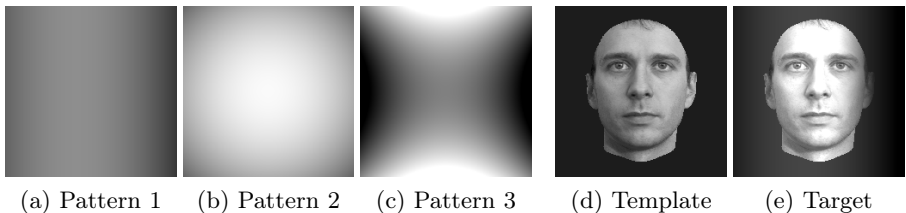


Fig. 1. Low-order polynomials as idealised contrast factors: three out of a host of possible 2D polynomial patterns of the 2nd degree (1a–1c), the template (1d), and the target image (1e) obtained by using the scaled pattern (1a) for the template (1d). The pattern (1e) was computed from (1d) by applying the transformation $f(x, y) = (0.035x - 0.00021x^2)t_{x,y} + 28$.

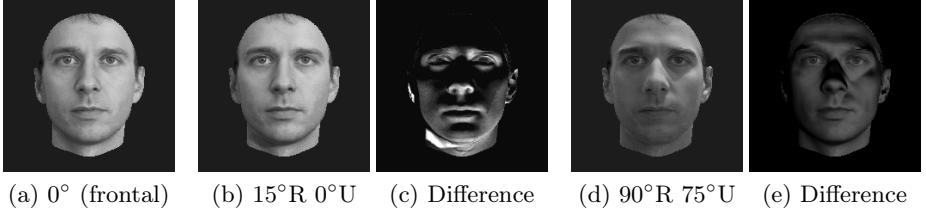


Fig. 2. Inhomogeneous contrast deviations: even a slight change in the light source orientation results in complex non-smooth contrast patterns. The contrast pattern (2c) for the target (2b) with respect to the template (2a) can hardly be approximated by a low-order polynomial. The same holds for more extreme illumination angles as shown in (2d) and (2e).

the upper bound and therefore would not be able to handle non-homogeneous contrast / offset deviations of higher order.

The interest in computationally efficient template matching has been around for quite some time. First attempts were done with various correlation-based methods [1] assuming the constant contrast and offset ($p = 0$). More recently, in response to the failure of traditional correlation techniques to handle the task [4], methods based on robust statistics and low-order polynomial models of contrast and offset such as proposed by Lai [8] made a step forward and branched out into a family of related algorithms [3,15,14]. In these algorithms, the traditional squared-error kernel is replaced by a statistical M-estimator being more robust in the presence of large signal differences and contrast and offset deviations are modelled with a low-order polynomial of pixel x - and y -coordinates.

Let $g : \mathcal{R} \rightarrow \mathcal{Q}$ denote a greyscale digital image \mathbf{g} on a finite arithmetic lattice $\mathcal{R} = [(x, y) : x = 0, \dots, m - 1; y = 0, \dots, n - 1]$ where $\mathcal{Q} = \{0, 1, \dots, Q - 1\}$ is a finite set of grey levels. Let $t : \mathcal{R} \rightarrow \mathcal{Q}$ be a template \mathbf{t} to be matched to \mathbf{g} .

Both the contrast factor $\alpha_{x,y}$ and offset $\beta_{x,y}$ for a pixel-wise signal $g_{x,y}$ in the target image are assumed to be polynomial functions of its co-ordinates (x, y) on the supporting pixel lattice so that the target signal depends on the respective signal $t_{x,y}$ in the template image as follows:

$$\begin{aligned} g_{x,y} &= \alpha_{x,y} t_{x,y} + \beta_{x,y} \\ &= (\alpha_0 + \alpha_1 x + \alpha_2 y + \dots) t_{x,y} + (\beta_0 + \beta_1 x + \beta_2 y + \dots) \end{aligned} \quad (1)$$

The matching algorithms use the basic Newton-Raphson descent method for iterative suboptimal minimisation of the Lorentzian or similar function chosen as the error kernel. Although the number of parameters could be expanded to a higher number (more than customary used five [8]), clearly there is a limit to what can be done under this formulation mainly due to high computational cost of its kernel and the kernel's influence function.

Other attempts to increase the number of parameters and thus improve the richness of the parameter space were done too. One approach is to avoid dealing with the entire N -dimensional space of images and resort to some selective

heuristics-based sampling of the pixel space [14]. Unfortunately, this method relies on manually crafting a heuristic requiring human expert intervention. Additionally, much of pixel space is completely ignored thus resulting in a rather limited applications of the approach.

Another alternative is to pre-process input data with edge detection and contouring [15]. Again as with earlier approaches, this direction managed to account for a very special low-order polynomial model of contrast / offset deviations but it fails when non-smooth variations such as shadows are introduced. Clearly, for all methods based on robust statistics additional computational costs associated with the robust estimators (error kernels) force these methods to tackle only a tiny subset of parameters preventing them from coping with more complex functions. As the result, adequate modeling of non-homogeneous noise becomes computationally intractable.

Alternative methods which do not use robust statistics have been also proposed. Unfortunately, they suffer from similar inability to handle rich parametric spaces and thus fail to capture non-homogeneous signal deviations. Examples include template matching handled in the frequency domain [6], modifications to correlation-based statistics [13,16], conquer-and-divide approaches to pixel space by breaking down the entire pixel space into smaller subspaces and subjecting them to localised matching [17], or even methods of matching stereo images where richer problem spaces are considered [2]. More recent alternative global or subglobal methods [11,12] rely on a number of restrictive assumptions and thus cannot handle more general non-homogeneous deviations. At present, to the best of our knowledge, no matching score for loosely constrained non-homogeneous contrast and offset deviations has been proposed¹.

3 Our Approach

Our approach is to capitalise on the computational convenience of the least-squares error function and formulate the matching problem as a global quadratic programming one (QP) with a simple set of linear constraints being sufficiently descriptive to handle arbitrary non-homogeneous contrast and offset deviations of target images. The re-formulation of the QP task in terms of the Karush-Kuhn-Tucker (KKT) conditions leads to an elegant and relatively fast implementation of the proposed new matching score using the Hildreth-D’Esopo algorithm [7,5].

The resulting non-parametric noise model combining spatially variant contrast and offset deviations with a central-symmetric independent noise and the derived on its basis matching score are far more general than its existing rivals and thus can successfully overcome the above limitations of the latter.

Let admissible contrast and offset deviations over the target image \mathbf{g} with respect to the template \mathbf{t} result in changes of the adjacent signals being constrained to a certain multiplicative range $E = [e_{\min}, e_{\max}]$ such that $0 < e_{\min} \leq e_{\max}$

¹ This paper deals only with photometric deviations, but the gradient-based search for geometric transformations of a template can be readily added much as in the other algorithms, e.g. in [8,11,12].

and with $e_{\min} < 1$ and $e_{\max} > 1$. Then contrast and offset of a transformed template can both decrease and increase. If $h : \mathcal{R} \rightarrow \mathcal{Q}$ is an image \mathbf{h} obtained by the admissible changes of the template \mathbf{t} , then the local constraints on \mathbf{h} are as follows:

$$\begin{array}{ccccc} \Delta_{\min;x,x-1;y} & \leq & h_{x;y} - h_{x-1;y} & \leq & \Delta_{\max;x,x-1;y} \\ \Delta_{\min;x;y,y-1} & \leq & h_{x;y} - h_{x;y-1} & \leq & \Delta_{\max;x;y,y-1} \\ 0 & \leq & h_{x;y} & \leq & Q - 1 \end{array} \quad (2)$$

for all the adjacent pairs $((x, y); (x - 1, y))$ and $((x, y); (x, y - 1))$ in \mathcal{R} where

$$\begin{aligned} \Delta_{\min;x,x-1;y} &= \min_{e \in E} \{e (t_{x,y} - t_{x-1,y})\}; \Delta_{\max;x,x-1;y} = \max_{e \in E} \{e (t_{x,y} - t_{x-1,y})\} \\ \Delta_{\min;x;y,y-1} &= \min_{e \in E} \{e (t_{x,y} - t_{x,y-1})\}; \Delta_{\max;x;y,y-1} = \max_{e \in E} \{e (t_{x,y} - t_{x,y-1})\} \end{aligned}$$

Assuming the central-symmetric random noise, the least-squares matching score minimises the Cartesian distance between the target image \mathbf{g} and the template \mathbf{t} under the constrained signal deviations of Eq. (2):

$$d(\mathbf{g}, \mathbf{t}|E) = \min_{\mathbf{h} \in \mathcal{H}(\mathbf{t}; E)} \left\{ \sum_{(x,y) \in \mathcal{R}} (h_{x;y} - g_{x;y})^2 \right\} \quad (3)$$

where $\mathcal{H}(\mathbf{t}; E)$ denotes a set of all images \mathbf{h} with the admissible signal deviations of Eq. (2) from the template \mathbf{t} . To solve this QP problem, it is converted to a dual problem that is solved efficiently with the Hildreth-d'Esopo QP algorithm [7,5] that is guaranteed to converge to a solution that is arbitrary close to the desired optimal solution. Since our constraint matrix is very sparse, this iterative implementation is efficient and converges quickly in spite of the large dimensionality $6N$ of the search space. There could be different stopping criteria, e.g. processing time (the number of iterations), convergence to the goal score (a threshold for its changes at successive iterations), or a magnitude of residual violations of the constraints. Since we found that suboptimal solutions can be adequate for image matching purposes as described in the next section, the latter criterion is not generally required.

Geometrically speaking, the algorithm starts the search for the solution \mathbf{h} from the target image \mathbf{g} , i.e. $\mathbf{h}^{[0]} = \mathbf{g}$ so that the matching score is zero at the very beginning and the initial location of $\mathbf{h}^{[0]}$ is most likely to be outside the feasible area specified by Eq. (2). Nevertheless the algorithm then quickly moves closer to the feasible area as the constraints are being negotiated. First several hundred iterations produce considerable improvement in meeting the constraints in the matched image (the score grows the fastest during the first iterations) and then the steps along the search space gradually level off.

4 Experimental Results

We have implemented the proposed QP-based matching algorithm and performed experiments with images having various non-homogeneous contrast /

offset deviations introduced by transforming templates to targets as follows (see Fig. 3 illustrating these experiments):

1. **Uniform contrast / offset:** A potentially lossy transformation for the high- and low-value signals because the transformed signals are truncated to fit the range $[0..255]$. This simple linear relationship corresponds to a frequently occurring situations such as variations due to software or hardware brightness and gain adjustments.
2. **Quadratic contrast:** Most likely it is even more lossy than the uniform one and may arise e.g. due to different angles of illumination of planar objects. See Fig. 1 for an example.
3. **Occlusions:** This scenario assumes that the original image signals are either partially or completely suppressed in a large area due to e.g. occlusions (a "step-function" contrast).
4. **Natural non-homogeneous deviations:** Shadowing of 3D object due to varying illumination angles is one of its examples. In practice it is one of the most relevant to computer vision types of noise because the exact image capturing conditions are almost never reproducible. See Fig. 2.
5. **Swapped uniform contrast / offset:** Finally, a transformation from less rich to more expressive images is modelled by swapping the target and template images; in our experiments, Case 1 has been used to swap the target and the template).

The main challenge to template matching of non-planar objects is posed by the natural non-homogeneity (Case 4) caused by shadows and varying illumination under different angles. It has never been addressed in the literature before. One of most suitable datasets for testing whether a matching algorithm can cope with this problem is the well-known MIT facial database [10], which we have used for the experiments below.

The dataset contains 360 images of 10 persons captured with the dominant light rotating such that its vertical orientation changes from 0° (direct) to 75° (top), and its horizontal orientation – from 0° (direct) to 90° (right). The additional ambient light source was used for making all facial features visible regardless of the dominant light position. All images are invariant to facial expression, geometry and backgrounds so that the experiments can be conducted in controlled conditions. The resulting complex shadows were the ultimate test for the proposed QP based approach (see Fig. 2c for an example of the deviations involved).

In order to benchmark the proposed algorithm, it was compared to the state-of-the-art alternative, namely, the gradient-descent based minimisation of the M-estimator of mismatches presented in [8] and capable to handle the contrast / offset deviations represented by simple polynomial models. We implemented a faster version [9] of this algorithm using the Newton-Raphson method.

Five tests were designed, corresponding to the above five scenarios of transforming templates to their target images. In our QP based matching the multiplicative range constants were set to $E = [0.25, 4]$ and the stopping rule was 7.500 iterations independently of whether other criteria were or not satisfied.

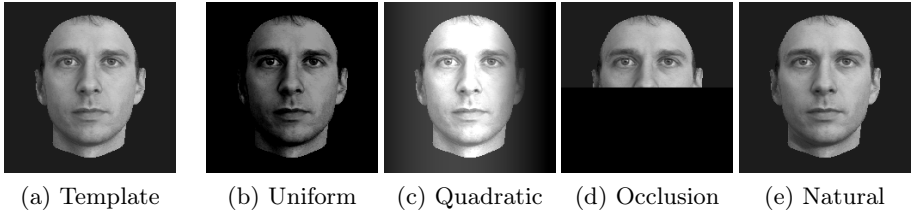


Fig. 3. Test cases used in our experiments: the template (a) and the target images for the above scenarios. In Case 5 the target (a) and template images and (b) are simply swapped. In the case of natural non-homogeneous deviations (e) due to illumination, the target may visually appear the same as (a), but in fact it does have mild shadowing variations e.g. under the subject’s left cheek bone.

For the benchmark counterpart, we use a 10-parameter polynomial model, i.e. 2nd-order polynomials without cross products for spatially variant contrast and offset. The stopping criterion was based on 20 iterations without improvement of the matching score.

For the uniform (or global) contrast / offset model, the target $f = \alpha t_{x,y} + \beta$ was obtained from its template with $\alpha = 1.3$, and $\beta = -40$ (see Fig. 3b), and the values $f(x, y)$ outside the range $[0..255]$ were truncated to their respective bounds. The contrast factor in Fig. 3c is a 2nd-degree polynomial along the horizontal axis only $f(x, y) = (0.035x - 0.00021x^2)t_{x,y} + 28$ (the value truncation was done after that transformation). For the step function, the lower half of the target image was set to $f_{x,y} = 0$ as in Fig. 3d. Targets with natural non-homogeneous deviations were chosen by randomly selecting another image of the same individual but with different lighting orientation (see Fig. 3e). Finally, for Case 5 the template and target images used in Case 1 were swapped. Results of experiments in comparison to the benchmark algorithm are shown in Fig. 4 and Table 1.

The QP-based matching algorithm displays clear superiority for the uniform contrast / offset, the natural shadows, and the swapped uniform case. It also is

Table 1. Scaled matching scores (total squared differences $d(\mathbf{g}, \mathbf{t}) \times 10^{-5}$ between the targets and the transformed templates) for the five test scenarios (mean, std, min, and max denote the mean score value, the standard deviation, and the minimum and maximum values of the matching scores in our experiments)

Test scenario	Benchmark				Our algorithm			
	Mean	Std	Min	Max	Mean	Std	Min	Max
1. Uniform contrast / offset	23	3.7	15	29	1.7	0.4	0.4	7.1
2. Quadratic contrast	6.1	7.6	0.8	28	64	1.6	49	85
3. Occlusion	210	59	110	350	87	4.5	46	134
4. Natural deviation	62	39	39	150	4.4	0.6	0.2	11
5. Swapped uniform case	74	27	37	150	6.6	0.8	1.5	17

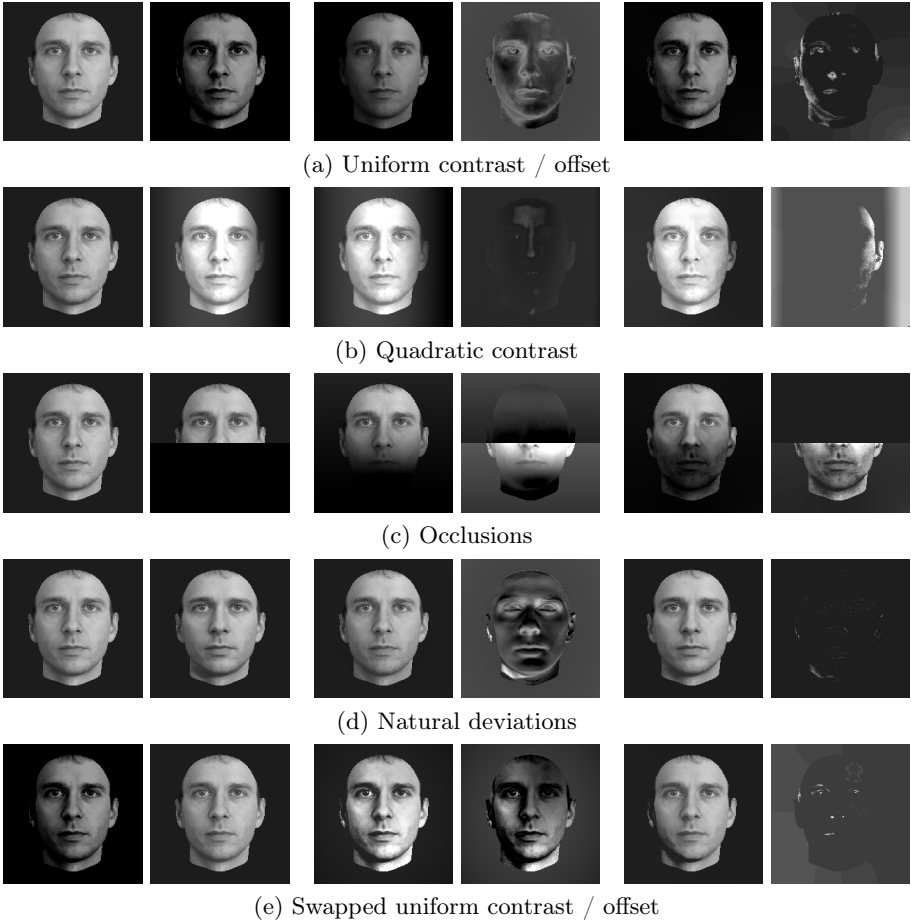


Fig. 4. Experimental results for the five test scenarios: the benchmark [8,9] vs. our QP-based matching: templates and targets are in columns 1 and 2, respectively; the matched image and its residual absolute difference from the target (scaled up for visualising purposes) for the benchmark is columns 3 and 4, respectively, and the like results for our matching are shown in columns 5 and 6. Note that the less the residual (i.e., the darker the residual), the closer (better) pixel-wise matching.

better in the case of occlusions, but is worse in the case of quadratic contrast function. The weaker performance of the benchmark algorithm [8,9] for the uniform contrast / offset and excellent performance for the quadratic contrast are worthy to note. The benchmark method presumes a second degree polynomial model so it is not surprising that it behaves nicely with the quadratic function. It is better than the more general QP based approach because of this perfect fit between the model and the function to be modeled. However, when handling simpler surfaces such as given in Case 1 the polynomial algorithm is just too numerically unstable due to its assumed higher (quadratic) polynomial factors.

In fact the best match in this case is achieved after just the first iteration, and then it begins to deteriorate.

The superior performance of our method has been demonstrated in all cases except the quadratic case when our approach behaves worse than the benchmark. It was quite expected because the benchmark polynomial model allows for different signs of local signal changes in adjacent pixels that correspond to one another in the transformed and original templates whereas in our model these signs have to remain the same. The other notable case relates to occlusions. Just as with the quadratic function, there were reasons to suspect that the benchmark could outperform our model since a single step function can be approximated by a 2nd-order polynomial. Nonetheless, our model gave better results in this case too. Statistical results for the 30 target–template pairs for each scenario are given in Table 1 and confirm the intuitive results visible on the output images.

5 Conclusions

This work focuses on template matching under spatially variant (inhomogeneous) contrast and offset deviations and presents the new theoretical framework and implementation of a general non-parametric QP-based algorithm. In a number of cases it shows superior results comparing to the current state-of-the-art polynomial approach. The latter benchmark algorithm had outperformed the proposed QP-matching only when the noise could be accurately approximated with the quadratic function assumed in the former method.

In the practical sense, the most important improvement of the QP approach is its ability to handle natural non-homogeneous variations resulting from changing lighting source orientation. This requirement is crucial when image recognition of non-planar objects is involved since it is almost impossible to reproduce the exact lighting conditions under which the template image was taken. Shadows are quite non-homogeneous constructs, and they can hardly have accurate low-order polynomial approximations.

The proposed noise model is more general and has greater expressive power. Its adaptability to signal deviations lies in the boundaries of the multiplicative range which were used to define the constraints. The greater the range, the more variations the model can successfully capture. This is true since the magnitude of the constraints represents the size of the admissible deviations the algorithm is aiming at. The selection of the multiplicative range should be done carefully which is a goal for future work. In this paper the authors set up a conservative range which nevertheless resulted in impressive improvement over the benchmark in four out of five test cases.

Finally, the authors are currently working on extending the current approach to include a mechanism for handling geometric transformations and masks for outliers. This will improve flexibility and power of template matching in a range of tasks where up to now matching has not been successful.

References

1. Aschwanden, P., Guggenbuhl, W.: Experimental results from a comparative study on correlation-type registration algorithms. In: Förstner, W., Ruwiedel, S. (eds.) *Robust Computer Vision*, pp. 268–289. Karlsruhe, Wichmann (1992)
2. Basri, R., Jacobs, D., Kemelmacher, I.: Photometric stereo with general, unknown lighting. *International Journal of Computer Vision* 72(3), 239–257 (2007)
3. Chen, J., Chen, C., Chen, Y.: Fast algorithm for robust template matching with m-estimators. *IEEE Transactions on Signal Processing* 51(1), 230–243 (2003)
4. Crowley, J., Martin, J.: Comparison of correlation techniques. In: *Proc. International Conference on Intelligent Autonomous Systems (IAS-4)*, Karlsruhe, Germany, March 27–30, pp. 86–93. IOS Press, Amsterdam (1995)
5. D’Esopo, D.: A convex programming procedure. *Naval Research Logistics Quarterly* 6, 33–42 (1959)
6. Fitch, A., Kadyrov, A., Christmas, W., Kittler, J.: Fast robust correlation. *IEEE Transactions on Image Processing* 14(8), 1063–1073 (2005)
7. Hildreth, C.: A quadratic programming procedure. *Naval Research Logistics Quarterly* 4, 79–85 (1957)
8. Lai, S.: Robust image matching under partial occlusion and spatially varying illumination change. *Computer Vision and Image Understanding* 78(1), 84–98 (2000)
9. Lai, S., Fang, M.: Method for matching images using spatially-varying illumination change models. US patent 6, 621, 929 (2003)
10. MIT face database (accessed August 24, 2006), <http://vismod.media.mit.edu/pub/images/>
11. Pizarro, D., Peyras, J., Bartoli, A.: Light-invariant fitting of active appearance models. In: *Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVIP 2008)*, Anchorage, Alaska, USA, pp. 1–6 (June 2008)
12. Silveira, G., Malis, E.: Real-time visual tracking under arbitrary illumination changes. In: *Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVIP 2007)*, Minneapolis, USA, pp. 1–6 (June 2007)
13. Tombari, F., Di Stefano, L., Mattoccia, S.: A robust measure for visual correspondence. In: *Proc. 14th Int. Conf. on Image Analysis and Processing (ICIAP)*, Modena, Italy, 2007, pp. 376–381 (September 2007)
14. Wei, S., Lai, S.: Robust and efficient image alignment based on relative gradient matching. *IEEE Trans. on Image Processing* 15(10), 2936–2943 (2006)
15. Yang, C., Lai, S., Chang, L.: Robust face image matching under illumination variations. *EURASIP Journal on Applied Signal Processing* 2004(16), 2533–2543 (2004)
16. Zhu, G., Zhang, S., Chen, X., Wang, C.: Efficient illumination insensitive object tracking by normalized gradient matching. *IEEE Signal Processing Letters* 14(12), 944–947 (2007)
17. Zou, J., Ji, Q., Nagy, G.: A comparative study of local matching approach for face recognition. *IEEE Trans. on Image Processing* 16(10), 2617–2628 (2007)