Photometric Invariant Projective Registration

Georgios D. Evangelidis *and* Emmanouil Z. Psarakis
Department of Computer Engineering and Informatics
University of Patras, 26500 Rio-Patras, Greece
email:{evagelid, psarakis}@ceid.upatras.gr *

The ability of an algorithm to accurately estimate the parameters of the geometric transformation which aligns two image profiles even in the presence of photometric distortions can be considered as a basic requirement in many computer vision applications. Projective transformations constitute a general class which includes as special cases the affine, as well as the metric subclasses of transformations. In this paper the applicability of a recently proposed iterative algorithm, which uses the Enhanced Correlation Coefficient as a performance criterion, in the projective image registration problem is investigated. The main theoretical results concerning the iterative algorithm and an efficient approximation that leads to an optimal closed form solution (per iteration) are presented. Furthermore, the performance of the iterative algorithm in the presence of nonlinear photometric distortions is compared against the leading Lucas-Kanade algorithm and its simultaneous inverse compositional variant by performing numerous simulations. In most cases the proposed algorithm outperforms the other algorithms in convergence speed and robustness against photometric distortions under ideal and noisy conditions.

1. Introduction

Many AI tools involve various computer vision application with the image registration being one of the most widely used. Image registration or alignment problem consists in finding a transformation which aligns two image profiles. Such correspondence problems arise often in practice with the most common cases being motion and optical flow estimation [1,2,5,7,9,10] tracking [8] and stereo correspondence problem [12,13] to name a few. The first critical step towards its solution is the choice of the appropriate class of geometric transformations (parametric models [16]) that can model this mapping. This choice is strictly related with the application at hands and may be restricted from the strategy that is followed to solve the alignment problem [4]. The class of projective transformations and in particular several special cases as affine, similariry transformations and pure translation have been in the center of attention in many applications [4,7,8].

Once the parametric transformation has been defined the alignment problem is reduced into a parameter estimation problem. Therefore, the second critical step in solving the alignment problem is the definition of an appropriate objective function and the solution of the induced optimization problem. Most of the used measures

*This work was supported by the General Secretariat for Research and Technology of Greek Government as part of the project "XROMA", PENED 01.

are l_p based norms of the error between the whole image profiles (*pixel-based techniques*) or between specific feature of image profiles (*feature-based techniques*) [16], with the l_2 norm being by far the most widely used [2,4,7,8,12,15]. This l_2 based objective function is usually referred as Sum-Squared-Differences (SSD) measure. Some variations of this technique have also been proposed, especially for the optical flow problem, adding spatial smoothness constraints and/or regularization parameters [9]. Robustified versions of the above mentioned measures that reduce their sensitivity in the presence of outliers have also been proposed [5].

Independently of the used measure, the parameter estimation problem results in a non-linear optimization problem and the alignment techniques, depending on the strategy they follow to solve it, can be broadly classified into two categories. Namely, gradient-based or differential and direct search techniques. Gradient based techniques because of their low computational cost can be considered as more well fitted to CV applications. However, homogeneous areas, single slanted edges (aperture problem) [10] as well as large motions constitute some pathological cases where these techniques fail to give meaningful results. On the other hand, direct search techniques can easily manage large motions, since the range of the search area is in our disposal. The drawbacks of these techniques are their heavy complexity and the fact that the precision of the optimum solution is affected from the quantization step [7]. Efforts to reduce complexity by adopting interpolation instead of fine quantization or hybrid techniques that combine the two classes can be found in [2,15].

A common assumption encountered in most of the existing techniques is the brightness constancy of corresponding points or regions in the two profiles [10]. However, this assumption is valid only in specific cases [10] and especially under varying illumination conditions it is violated. In a real world application it is vital an alignment algorithm to be capable in taking into account illumination changes. Several alignment techniques that compensate photometric distortions in contrast and brightness [7,10,12] and linear appearance variation [3] have been proposed in the literature. Moreover, alignment techniques which make use of a set of basis images for handling arbitrary lighting conditions [8] as well as more sophisticated spatially dependent photometric models [1] have also been proposed.

In [6] a modification of the correlation coefficient as a performance criterion for the image alignment problem is proposed. The proposed modification provides certain desirable invariance properties with respect to photometric distortions. Since the similarity measure is a highly nonlinear function of the warp parameters, an iterative maximization technique was developed. Based on an efficient approximation a closed form solution (per iteration) of the iterative algorithm is presented. Moreover, the class of affine transformations to model the warping process was used and the iterative scheme was compared against the Lucas-Kanade algorithm with the help of numerous simulations. In all cases the algorithm exhibited a noticeable performance improvement over the Lucas-Kanade scheme regarding convergence speed as well as in probability of convergence.

In this paper we present part of the theoretical results contained in [6] and perform a number of simulations in order to evaluate the performance of this new scheme modulated for projective registration in the presence of nonlinear photometric distortions. Moreover we compare it against the well known Lucas-Kanade alignment algorithm and the recently proposed Simultaneous Inverse Compositional algorithm [3].

The remainder of this paper is organized as follows. In Section 2, we briefly formulate the parametric image alignment problem. In Section 3, we present the l_2 based objective function and comment on its equivalency with the Enhanced Correlation Coefficient function as well as its relation to the two most popular approaches based on the SSD measure. In the same section the proposed gradient based iterative scheme and the basic non-linear optimization problem are defined. In addition, the optimal closed form solution of the basic optimization problem is given. In Section 4, we apply the proposed technique in a number of experiments and perform detailed comparisons with the Lucas-Kanade and the simultaneous inverse compositional alignment algorithm . Finally, Section 5 contains our conclusions.

2. Problem Formulation

In this section we briefly introduce the problem of alignment of two image profiles. To this end, let us assume that a *reference* image $I_r(\mathbf{x})$ and a *warped* image $I_w(\mathbf{x}')$ are given, where $\mathbf{x} = [x, y]$ and $\mathbf{x}' = [x', y']$ denote coordinates. Suppose also that we are given a set of coordinates $\mathcal{S} = \{\mathbf{x}_i |\ i = 1,\ \ldots,\ K\}$ in the reference image, which is called *target area*. Then, the alignment problem consists in finding the corresponding coordinate set in the warped image.

By considering that a transformation model $T(\mathbf{x}; \mathbf{p})$ where $\mathbf{p} = (p_1, p_2, \ldots, p_N)^t$ is a vector of unknown parameters is given, the alignment problem is reduced to the problem of estimating the parameter vector $\mathbf{p}$ such that

$$I_r(\mathbf{x}) = \Psi(I_w(T(\mathbf{x}; \mathbf{p})); \alpha),\ \mathbf{x} \in \mathcal{S}, \tag{1}$$

where transformation $\Psi(I, \alpha)$ which is parameterized by a vector α, accounts for possible photometric distortions that violate the brightness constancy assumption, a case which arises in real applications due to different viewing directions and/or different illumination conditions.

The goal of most existing algorithms is the minimization of the dissimilarity of the two image profiles, providing the optimum parameter values. Dissimilarity is usually expressed through an objective function $E(\mathbf{p}, \alpha)$ which involves the l_p norm of the intensity residual of the image profiles. A typical minimization problem has the following form

$$\min_{\mathbf{p},\alpha} E(\mathbf{p}, \alpha) = \min_{\mathbf{p},\alpha} \sum_{\mathbf{x}\in\mathcal{S}} |I_r(\mathbf{x}) - \Psi\left(I_w(T(\mathbf{x}; \mathbf{p})), \alpha\right)|^p. \tag{2}$$

Solving the above defined problem is not a simple task because of the nonlinearity involved in the correspondence part. Computational complexity and estimation

quality of existing schemes depends on the specific l_p norm and the models used for warping and photometric distortion. As far as the norm power p is concerned most methods use $p = 2$ (Euclidean norm). This will also be the case in the approach we briefly present in the next section.

3. The Alignment Algorithm

It is more convenient at this point to define the *reference vector* $\mathbf{i}_r$ and the *warped vector* $\mathbf{i}_w(\mathbf{p})$ as follows

$$\mathbf{i}_r = \begin{bmatrix} I_r(\mathbf{x}_1) \\ I_r(\mathbf{x}_2) \\ \vdots \\ I_r(\mathbf{x}_K) \end{bmatrix}, \ \mathbf{i}_w(\mathbf{p}) = \begin{bmatrix} I_w(T(\mathbf{x}_1;\mathbf{p})) \\ I_w(T(\mathbf{x}_2;\mathbf{p})) \\ \vdots \\ I_w(T(\mathbf{x}_K;\mathbf{p})) \end{bmatrix} \tag{3}$$

and denote by $\bar{\mathbf{i}}_r$ and $\bar{\mathbf{i}}_w(\mathbf{p})$ the zero-mean versions of the reference and warped vector respectively. We then propose the following l_2 based criterion to quantify the performance of the warping transformation with parameters $\mathbf{p}$

$$E_{ECC}(\mathbf{p}) = \left|\left| \frac{\bar{\mathbf{i}}_r}{||\bar{\mathbf{i}}_r||} - \frac{\bar{\mathbf{i}}_w(\mathbf{p})}{||\bar{\mathbf{i}}_w(\mathbf{p})||} \right|\right|^2, \tag{4}$$

where $||\cdot||$ denotes the usual Euclidean norm.

It is clear from (4) that this criterion is invariant to possibly existing contrast and/or brightness changes since involved vectors are zero-mean and normalized. So, to a first approximation, we can concentrate on the geometric transformation putting aside the photometric one. These characteristics clearly support the choice to be adopted this criterion for the image registration problem.

By assuming that photometric distortion is limited only to global brightness and contrast changes, we can derive the proposed criterion in a different way. This will help us in related it to the two, currently most popular SSD approaches in the literature. Under the above mentioned assumption on photometric changes, and by defining the following performance measure

$$E(\mathbf{p}, \alpha) = \|\alpha_1 \mathbf{i}_w(\mathbf{p}) + \alpha_2 - \mathbf{i}_r\|^2, \tag{5}$$

where $\alpha = [\alpha_1 \, \alpha_2]^t$ is the parameter vector for the photometric transformation, the proposed measure can easily be proved that is equivalent with the following constrained minimization problem

$$E_{ECC}(\mathbf{p}) = \min_{\alpha_1 \geq 0, \alpha_2} E(\mathbf{p}, \alpha). \tag{6}$$

If we drop the constraint $\alpha_1 \geq 0$ then the minimization of the objective function in (5) is the optimization problem proposed by Fuh and Maragos [7]. Notice that Fuh and Maragos followed a direct search method towards optimum solution.

An alternative measure arises if in (5) we interchange the roles of $\mathbf{i}_w$ and $\mathbf{i}_r$, that is,

$$E(\mathbf{p}, \alpha) = \|\alpha_1 \mathbf{i}_r + \alpha_2 - \mathbf{i}_w(\mathbf{p})\|^2. \tag{7}$$

This is the approach adopted by Lucas-Kanade [12] and is known to generate the most widely used algorithm in practice.

3.1. *Nonlinear Maximization*

Let us now concentrate ourselves on the minimization of the proposed measure. Since the residual in (4) is based on zero-mean and normalized vectors, it is straightforward to prove that minimizing $E_{ECC}(\mathbf{p})$ is equivalent to maximizing the *enhanced correlation coefficient* [13]

$$\rho(\mathbf{p}) = \hat{\mathbf{i}}_r^t \frac{\bar{\mathbf{i}}_w(\mathbf{p})}{||\bar{\mathbf{i}}_w(\mathbf{p})||} \tag{8}$$

where $\hat{\mathbf{i}}_r$ is the normalized reference vector. Notice that even if $\bar{\mathbf{i}}_w(\mathbf{p})$ depends linearly on the parameter vector $\mathbf{p}$, the resulting objective function is still nonlinear with respect to $\mathbf{p}$ due to the normalization of the warped vector. This of course suggests that its maximization requires nonlinear optimization techniques.

In order to maximize $\rho(\mathbf{p})$ we are going to use a gradient-based iterative approach. More specifically, we are going to replace the original optimization problem by a *sequence* of secondary optimizations. Each such optimization relies on the outcome of its predecessor thus generating a sequence of parameter estimates which hopefully converges to the desired optimizing vector of the original problem. Notice that, at each iteration we do not have to optimize the objective function, but an *approximation* to this function, such that the resulting optimizer are simple to compute. Let us therefore introduce next the approximation we intend to apply to our objective function and also derive with a theorem an analytic expression for the solution that maximizes it.

Suppose that $\mathbf{p}$ is "close" to some nominal parameter vector $\tilde{\mathbf{p}}$ and write $\mathbf{p} = \tilde{\mathbf{p}} + \Delta\mathbf{p}$, where $\Delta\mathbf{p}$ denotes a vector of perturbation. Suppose also that the intensity function I_w and the warping transformation T are of sufficient smoothness to allow for the existence of the required partial derivatives. If we denote as $\tilde{\mathbf{x}}' = T(\mathbf{x}; \tilde{\mathbf{p}})$ the warped coordinates under the nominal parameter vector and $\mathbf{x}' = T(\mathbf{x}; \mathbf{p})$ under the perturbed, then, applying a first order Taylor expansion with respect to the parameters, we can write

$$I_w(\mathbf{x}') \approx I_w(\tilde{\mathbf{x}}') + \left[\nabla_{\mathbf{x}'} I_w(\tilde{\mathbf{x}}')\right]^t \frac{\partial T(\mathbf{x}; \tilde{\mathbf{p}})}{\partial \mathbf{p}} \Delta\mathbf{p}, \tag{9}$$

where $\nabla_{\mathbf{x}'} I_w(\tilde{\mathbf{x}}')$ denotes the gradient vector of length 2 of the intensity function $I_w(\mathbf{x}')$ of the warped image evaluated at the nominal coordinates $\tilde{\mathbf{x}}'$, and $\frac{\partial T(\mathbf{x}; \tilde{\mathbf{p}})}{\partial \mathbf{p}}$ denotes the size $2 \times N$ Jacobian matrix of the warp transform with respect to its parameters, evaluated at the nominal values $\tilde{\mathbf{p}}$.

By applying (9) to all points of target area $\mathcal{S}$, forming the linearized version of the warp vector $\mathbf{i}_w(\mathbf{p})$ and computing its zero mean counterpart we obtain the following approximation $\rho(\Delta\mathbf{p}|\tilde{\mathbf{p}})$ of the objective function $\rho(\mathbf{p})$ defined in (8):

$$\rho(\mathbf{p}) \approx \rho(\Delta\mathbf{p}|\tilde{\mathbf{p}}) = \frac{\hat{\bar{\mathbf{i}}}_r^t \bar{\mathbf{i}}_w + \hat{\bar{\mathbf{i}}}_r^t \bar{G} \Delta\mathbf{p}}{\sqrt{\|\bar{\mathbf{i}}_w\|^2 + 2\bar{\mathbf{i}}_w^t \bar{G} \Delta\mathbf{p} + \Delta\mathbf{p}^t \bar{G}^t \bar{G} \Delta\mathbf{p}}} \tag{10}$$

where $G(\tilde{\mathbf{p}})$ denotes the size $K \times N$ Jacobian matrix of the warped intensity vector with respect to the parameters, evaluated at the nominal parameter values $\tilde{\mathbf{p}}$, and $\bar{G}$ denotes its column-zero-mean counterpart. Notice that for notational simplicity, the dependence of the warped vectors on $\mathbf{p}$ has been dropped.

Although $\rho(\Delta\mathbf{p}|\tilde{\mathbf{p}})$ is a non-linear function of $\Delta\mathbf{p}$, its maximization results in a closed-form solution. This solution is given, without proof, by the next theorem.

Theorem 3.1. Consider the scalar function

$$f(\mathbf{x}) = \frac{u + \mathbf{u}^t \mathbf{x}}{\sqrt{v + 2\mathbf{v}^t \mathbf{x} + \mathbf{x}^t Q \mathbf{x}}} \tag{11}$$

where u, v are scalars; $\mathbf{u}, \mathbf{v}$ are vectors of length N; Q is a square, symmetric and positive definite matrix of size N and $v, \mathbf{v}, Q$ are such that

$$v > \mathbf{v}^t Q^{-1} \mathbf{v} \tag{12}$$

then, as far as the maximal value of $f(\mathbf{x})$ is concerned, we distinguish the following two cases:

Case $u > \mathbf{u}^t Q^{-1} \mathbf{v}$: here we have a maximum, specifically

$$\max_{\mathbf{x}} f(\mathbf{x}) = \sqrt{\frac{(u - \mathbf{u}^t Q^{-1} \mathbf{v})^2}{v - \mathbf{v}^t Q^{-1} \mathbf{v}} + \mathbf{u}^t Q^{-1} \mathbf{u}}, \tag{13}$$

which is attainable for

$$\mathbf{x} = Q^{-1} \left\{ \frac{v - \mathbf{v}^t Q^{-1} \mathbf{v}}{u - \mathbf{u}^t Q^{-1} \mathbf{v}} \mathbf{u} - \mathbf{v} \right\}. \tag{14}$$

Case $u \leq \mathbf{u}^t Q^{-1} \mathbf{v}$: here we have a supremum, specifically

$$\sup_{\mathbf{x}} f(\mathbf{x}) = \sqrt{\mathbf{u}^t Q^{-1} \mathbf{u}}, \tag{15}$$

which can be approached arbitrarily close by selecting

$$\mathbf{x} = Q^{-1} \left\{ \lambda \mathbf{u} - \mathbf{v} \right\}, \tag{16}$$

with λ positive scalar, of sufficiently large value.

Using the results of *Theorem I* and by defining the projection matrix $P_G = \bar{G}(\bar{G}^t \bar{G})^{-1} \bar{G}^t$, we have that the optimum perturbation is equal to

$$\Delta\mathbf{p}^o = (\bar{G}^t \bar{G})^{-1} \bar{G}^t \left\{ \frac{\|\bar{\mathbf{i}}_w\|^2 - \bar{\mathbf{i}}_w^t P_G \bar{\mathbf{i}}_w}{\hat{\bar{\mathbf{i}}}_r^t \bar{\mathbf{i}}_w - \hat{\bar{\mathbf{i}}}_r^t P_G \bar{\mathbf{i}}_w} \hat{\bar{\mathbf{i}}}_r - \bar{\mathbf{i}}_w \right\}, \tag{17}$$

when $\hat{\mathbf{i}}_r^t \bar{\mathbf{i}}_w > \hat{\mathbf{i}}_r^t P_G \bar{\mathbf{i}}_w$; or according to (16),

$$\Delta \mathbf{p}^o = (\bar{G}^t \bar{G})^{-1} \bar{G}^t \left\{ \lambda \hat{\mathbf{i}}_r - \bar{\mathbf{i}}_w \right\}, \tag{18}$$

when $\hat{\mathbf{i}}_r^t \bar{\mathbf{i}}_w \leq \hat{\mathbf{i}}_r^t P_G \bar{\mathbf{i}}_w$, where λ must be selected so that the resulting $\rho(\Delta \mathbf{p}^o | \tilde{\mathbf{p}})$ satisfies $\rho(\Delta \mathbf{p}^o | \tilde{\mathbf{p}}) > \rho(0 | \tilde{\mathbf{p}})$ and $\rho(\Delta \mathbf{p}^o | \tilde{\mathbf{p}}) \geq 0$. Possible values of λ provide the following lemma.

Lemma 3.1. *Let $\hat{\mathbf{i}}_r^t \bar{\mathbf{i}}_w \leq \hat{\mathbf{i}}_r^t P_G \bar{\mathbf{i}}_w$ and define the following two values for λ*

$$\lambda_1 = \sqrt{\frac{\bar{\mathbf{i}}_w^t P_G \bar{\mathbf{i}}_w}{\hat{\mathbf{i}}_r^t P_G \hat{\mathbf{i}}_r}}, \quad \lambda_2 = \frac{\hat{\mathbf{i}}_r^t P_G \bar{\mathbf{i}}_w - \hat{\mathbf{i}}_r^t \bar{\mathbf{i}}_w}{\hat{\mathbf{i}}_r^t P_G \hat{\mathbf{i}}_r}. \tag{19}$$

Then for $\lambda \geq \lambda_1$ we have that $\rho(\Delta \mathbf{p}^o | \tilde{\mathbf{p}}) > \rho(0 | \tilde{\mathbf{p}})$; for $\lambda \geq \lambda_2$ that $\rho(\Delta \mathbf{p}^o | \tilde{\mathbf{p}}) \geq 0$; finally for $\lambda \geq \max\{\lambda_1, \lambda_2\}$ we have both inequalities valid.

Let us now translate the above results into an *iterative* scheme in order to obtain the solution to the original nonlinear optimization problem.

3.2. *Iterative Scheme*

To this end, let us assume that from iteration $j-1$ we have available the parameter estimate $\mathbf{p}_{j-1}$ and we adopt the following additive rule

$$\mathbf{p}_j = \mathbf{p}_{j-1} + \Delta \mathbf{p}_j. \tag{20}$$

Then, using $\mathbf{p}_{j-1}$ we can compute $\bar{\mathbf{i}}_w(\mathbf{p}_{j-1})$ and $\bar{G}(\mathbf{p}_{j-1})$ and optimize the approximation $\rho(\Delta \mathbf{p}_j | \mathbf{p}_{j-1})$ with respect to $\Delta \mathbf{p}_j$. The iterative algorithm is summarized below. Initialization

- *Use I_r to compute $\hat{i}_r$ defined in (3).*
- *Initialize $\mathbf{p}_0$ and set $j = 1$.*

Iteration Steps

- *Using $T(\mathbf{x}; \mathbf{p}_{j-1})$ warp I_w and compute $\bar{\mathbf{i}}_w(\mathbf{p}_{j-1})$*
- *Using $T(\mathbf{x}; \mathbf{p}_{j-1})$ warp the gradient ∇I_w of I_w and compute the Jacobian matrix $\bar{G}(\mathbf{p}_{j-1})$*
- *Compare $\hat{\mathbf{i}}_r^t \bar{\mathbf{i}}_w$ with $\hat{\mathbf{i}}_r^t P_G \bar{\mathbf{i}}_w$ and compute perturbations $\Delta \mathbf{p}_j^o$ either from (17) or using (18) and (19)*
- *Update parameter vector $\mathbf{p}_j = \mathbf{p}_{j-1} + \Delta \mathbf{p}_j^o$.*

If $\|\Delta \mathbf{p}_j^o\| \geq \epsilon_p$ then, $j++$ and repeat; else stop.

As it is indicated above, the algorithm is executed until the norm of the perturbation vector $\|\Delta \mathbf{p}_j^o\|$ becomes smaller than a predefined threshold ϵ_p.

We must stress here that the choice of the initial value of $\mathbf{p}_0$ is very critical for the convergence of the algorithm. For example, in the specific case of partially overlapped images or in a too strong geometric deformation case (i.e. flip up-down),

an appropriate value of $\mathbf{p}_0$ helps the algorithm to avoid local maxima. A reliable estimation for the initialization of the algorithm can be obtained by incorporating the above algorithm with a correlation style search method [15] or a landmark-based method [11]. However, in this paper we consider that the images have sufficient overlap.

3.3. *Complexity and computational issues*

The complexity per iteration of the above scheme is $O(KN^2)$ with the most expensive step being the third step. Specifically, this step requires the computation and the inversion of the Hessian matrix, task of complexity $O(KN^2 + N^3)$ ($K \gg N$), while all other steps have lower bounds. Hence, given the number M of iterations needed until convergence, the overall complexity becomes $O(MKN^2)$.

This bound is also the complexity of another well known scheme, the forwards additive scheme of the Lucas Kanade (LK) algorithm [12]. As we mentioned above, the aim of this scheme is the minimization of the squared error between the images. Taking into account changes in gain and bias the minimization of (7) results in a similar to (17) solution. What differentiates these algorithms is the scalar quantity that precedes the vector $\hat{\mathbf{i}}_r$ [6].

As far as the efficiency is concerned some variants of LK scheme have been proposed. To this end, a compositional warp update rule can be adopted and/or a backward (inverse) approach can be followed to achieve the alignment [8,15,4]. More recently, and based on the inverse compositional logic, a series of new variants to cope with the linear appearance variation of images have been proposed [3]. Among them, the simultaneous inverse compositional (SIC) algorithm seems to behave better in a wide sense, while the other versions are based on various approximation to achieve a lower computational cost. Because of the appearance parameters, the computational cost of SIC algorithm is $O(K(N + M)^2)$, where M is the size of the appearance parameter vector, cancelling thus the advantage of the backward approach. However, if the linear appearance variation is used to model gain and bias, the computational cost can be considered similar to $O(KN^2)$.

The compositional update rule as well as the backward logic in alignment can obviously be combined with the presented iterative scheme leading thus to more efficient approximations, but affecting the accuracy of alignment and restricting the applicability of the algorithms. Notice that these variants cannot be applied to any set of warps [4].

Since in this paper we concentrate ourselves in the alignment performance and efficiency is not a concern here, we are going to empirically compare in next section the three mentioned algorithms, though a series of tests using a known simulation framework [4].

4. Simulation Results

In this section we are going to evaluate our algorithm (ECC) and compared it against the forward additive version of the Lucas-Kanade algorithm (LK) as well as the simultaneous inverse compositional algorithm (SIC). LK and SIC algorithms are modulated to model gain and bias variation in images, while ECC algorithm is invariant to linear intensity deformation. However, in all experiments, a nonlinear photometric transformation is used for image distortion. As far as the geometric modelling is concerned, the class of projective transformation is used for the warping process. We must stress at this point that for all aspects affecting the simulation experiments, we made an effort to stay exactly within the frame specified in [4]. Before we present our results we give some details for the experimental setup as well as the figures of merit we are going to use.

4.1. *Parametric and Photometric Distortion Models*

In all experiments, to model the warping process we are going to use the projective transformation (homography) defined as follows

$$\mathbf{x}' = T(\mathbf{x};\mathbf{p}) = \frac{1}{P}\begin{bmatrix} p_1 & p_2 & p_3 \\ p_4 & p_5 & p_6 \end{bmatrix}\begin{bmatrix} \mathbf{x} \\ 1 \end{bmatrix} \tag{21}$$

where $P = p_7 x + p_8 y + 1$. This class of transformations is the most general class of the well known 2-D planar motion models that subsumes the most commonly used class of the affine transformations. For the Jacobian of the projective model we have

$$\frac{\partial T(\mathbf{x};\mathbf{p})}{\partial \mathbf{p}} = \frac{1}{P}\begin{bmatrix} x & y & 1 & 0 & 0 & 0 & -x'x & -x'y \\ 0 & 0 & 0 & x & y & 1 & -y'x & -y'y \end{bmatrix}, \tag{22}$$

where x', y' are the elements of $\mathbf{x}'$.

As it is clear from (21), projective transformation is a nonlinear function of its parameters and its stability as well as the continuity of its Jacobian depends on the values of the denominator P. To ensure its stability and the existence of its Jacobian, we restrict ourselves on admissible [14] estimations of the transform. Note also that in spite of the affine model which has a Jacobian that does not depend on the warping parameters, projective model, as it is obvious from (22), has a Jacobian that depends on the parameters $\mathbf{p}$ and thus it must be updated on each iteration of the iterative algorithms. These drawbacks can be overcome if we use approximated to projective transformation models as bilinear or polynomial models. However in this case, the projective deformation cannot be exactly adjusted and this approach may lead to meaningless alignment results in a real applications.

In order to investigate the performance of the algorithms under photometrically distorted images, we consider nonlinear transformations of the following form

$$I_r(\mathbf{x}) \leftarrow (I_r(\mathbf{x}) + \alpha_2)^{\gamma},\ \mathbf{x} \in \mathcal{S} \tag{23}$$

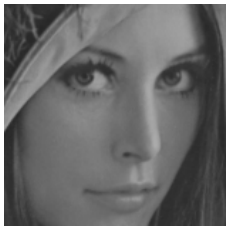
which are applied to the intensity of each pixel in the reference image. In the experiments presented in this work we used $\alpha_2 = 20$ and $\gamma = 0.9$ respectively. Notice that in order not to favor the competing algorithms which are invariant to linear photometric distortions, we adopt a nonlinear photometric distortion model.

4.2. *Experimental Setup*

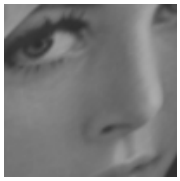
In order to create a reference and a target (warped) image we follow the procedure proposed in [4]. Specifically, let $I(\mathbf{y})$ be a given image. Consider an orthogonal target area which, without loss of generality, we assume to have a left-bottom corner with coordinates (0,0) and the top-right corner with coordinates $(L-1, J-1)$ (i.e a target area of size $L \times J$). Since the affine transformation has six parameters, in order to be completely specified it is sufficient to define the correspondence between three points in the plane. For this we select $\mathbf{x}_{lb} = [0,0]^t, \mathbf{x}_{rb} = [0, L-1]^t, \mathbf{x}_{mt} = [\frac{L-1}{2}, J-1]^t$. We now need to define where these three points correspond to. If with $\mathbf{x}_{ij}$ we denote any of the three vectors $\{\mathbf{x}_{lb}, \mathbf{x}_{rb}, \mathbf{x}_{mt}\}$ then we first select a vector $\mathbf{x}_0$ such that the three vectors $\mathbf{x}_0 + \mathbf{x}_{ij}$ correspond to points that lie in the interior of the support of the given image. In order to change this simple translation model into a more general warping, we simply perturb the three points by adding Gaussian noise. In other words we generate three points using the following formula $\mathbf{y}_{ij} = \mathbf{x}_0 + \mathbf{x}_{ij} + \delta_{ij}$, where δ_{ij} are vectors of length 2 with elements constituting independent realizations of zero-mean Gaussian random variables with the same standard deviation σ_p. The three initial points $\mathbf{x}_{ij}$ and their corresponding warped versions $\mathbf{y}_{ij}$, when substituted in (**??**), generate a system of six *linear equations* in six unknowns which, when solved, defines the parameter vector of the affine transformation. We denote this parameter vector as $\mathbf{p}_r$.

With the parameter vector $\mathbf{p}_r$ (and therefore the affine warping transformation) being specified, we now need to compute the intensity of the reference image $I_r(\mathbf{x})$. For this we consider all coordinates $\mathbf{x} = [x_1, x_2]^t$ with $0 \leq x_1 \leq L-1, 0 \leq x_2 \leq J-1$ and warp them into coordinates $\mathbf{y} = [y_1, y_2]^t$ of the given image, using the warping transformation $\mathbf{y} = \phi(\mathbf{x}; \mathbf{p}_r)$. Of course the resulting points $\mathbf{y}$ will not necessarily correspond to existing pixels in the given image, we thus use bilinear interpolation to define their intensities. This process, as we said, defines the reference image $I_r(\mathbf{x})$. For the warped image we can select the given one, that is, $I_w(\mathbf{y}) = I(\mathbf{y})$. To make things even more difficult, in a set of experiments we also add noise to the two images. Specifically we perturb each *pixel intensity* by adding independent and identically distributed, zero-mean Gaussian noise of standard deviation σ_i. As an example, in Fig. 1 we present the outcome of this procedure when applied to the "Lenna" image with $\sigma_p = 10, \sigma_i = 0$ and $L = J = 100$. In all experiments we are going to perform, we make use of the "Takeo", "Simon", "Knee" and "Car" images but not of "Lena", this is because we attempt to stay consistent with the experiments in [4].

In our simulations we compare our algorithm against the LK. Therefore for all



(a)



(b)

Fig. 1. (a) The original "Lena" image and (b) the resulting reference profile with $L = J = 100$ and $\sigma_p = 10$.

aspects affecting the simulation experiments, we made an effort to stay exactly within the frame specified in [4]. In particular as far as initialization of the parameter vector estimates is concerned for both schemes we use $p_1 = p_2 = p_3 = p_4 = 0$ and $[p_5, p_6]^t = \mathbf{x}_0$. At iteration j the corresponding algorithm will provide estimates $\mathbf{p}_j$. In order to measure the quality of this estimate we propose the use of the MSE between the exact warped version of the points $\{\mathbf{x}_{lb}, \mathbf{x}_{rb}, \mathbf{x}_{mt}\}$ and their estimated counterparts. More formally, we are using the following quantity

$$e(j) = \frac{1}{6} \sum_{\mathbf{x}=\mathbf{x}_{lb},\mathbf{x}_{rb},\mathbf{x}_{mt}} \mathbb{E}\left\{\|\phi(\mathbf{x};\mathbf{p}_r) - \phi(\mathbf{x};\mathbf{p}_j)\|^2\right\}. \tag{24}$$

In the previous definition $\mathbb{E}\{\cdot\}$ denotes expectation with respect to all types of noises included in the two image profiles.

We must point out, that we do not expect any of the two algorithms to converge at all times. This is particularly apparent for high values of σ_p. For this reason, in order to quantify the algorithmic performance in a meaningful way, in the next subsection we introduce several interesting figures of merit.

4.3. *Figures of Merit*

In iterative schemes it is clearly the speed of convergence that determines the quality of the corresponding algorithm. In this sense the sequence $\{e(j)\}$ defined in (24) can be used as our first figure of merit. Indeed $\{e(j)\}$ captures the *learning ability* of the algorithm. Unfortunately this selection turns out to be problematic basically because, as we said previously, the two algorithms do not have a guaranteed convergence. In fact in the cases where the fail the corresponding MSE can be quite important, therefore altering significantly the average MSE value $e(j)$. Consequently, in order to have the right picture of the convergence characteristic, we could adopt the idea followed in [4], namely to define the MSE but *conditioned* on

the event that the algorithm has converged, that is

$$e(j) = \frac{1}{6} \sum_{\mathbf{x}=\mathbf{x}_{lb},\mathbf{x}_{rb},\mathbf{x}_{mt}} \mathbb{E}\left\{ \|\phi(\mathbf{x};\mathbf{p}_r) - \phi(\mathbf{x};\mathbf{p}_j)\|^2 | \text{ the algorithm has converged} \right\}. \quad (25)$$

We can clearly apply (25) for each algorithm separately. However since the two algorithms may fail convergence for different realization, we decided to compute the MSE only for those where *both* algorithms converge, namely

$$e(j) = \frac{1}{6} \sum_{\mathbf{x}=\mathbf{x}_{lb},\mathbf{x}_{rb},\mathbf{x}_{mt}} \mathbb{E}\left\{ \|\phi(\mathbf{x};\mathbf{p}_r) - \phi(\mathbf{x};\mathbf{p}_j)\|^2 | \text{ both algorithms have converged} \right\}. \quad (26)$$

A practical computation of (26) utilizes the Law of Large Numbers and approximates expectation with average over many realizations. In other words we run our algorithms over a large number of image pairs that differ in the noise realization. For each pair and for each iteration step j we compute the quantity

$$\bar{e}(j) = \frac{1}{6} \sum_{\mathbf{x}=\mathbf{x}_{lb},\mathbf{x}_{rb},\mathbf{x}_{mt}} \|\phi(\mathbf{x};\mathbf{p}_r) - \phi(\mathbf{x};\mathbf{p}_j)\|^2 \quad (27)$$

and for each j we take the average only of those $\bar{e}(j)$ where both algorithms have converged. By "convergence" we mean that $\bar{e}(j_{\max}) \leq T_{\text{MSE}}$. In other words we consider that an algorithm has converged when its squared error $\bar{e}(j)$ at a prescribed maximal iteration j_{max} is below a certain threshold level T_{MSE}.

Although the sequence $\{e(j)\}$, with $e(j)$ as defined in (26), gives a concise picture of the learning (convergence) capabilities of an algorithm, this is nonetheless an *average* index (it is like representing a random variable simply by its mean), not to mention of course that it hides the frequency failures. Since it is only natural to prefer an algorithm that converges quickly *with high probability*, we propose a second figure of merit that captures exactly this point. In other words we propose the generation of a histogram depicting the probability of successful convergence at each iteration. Specifically a run of an algorithm on an image pair realization will be considered as having converged at iteration n when the squared error $\bar{e}(j)$ defined in (27) goes below the threshold T_{MSE} for *the first time* at iteration $j = n$. It is clear that we prefer a histogram to be concentrated (or equivalently the probability of successful convergence to be distributed) over mostly small iteration-numbers.

A final quantity which is of importance is clearly the percentage of converging (PoC) runs. This index is not directly captured with the previous two figures of merit. We therefore define this last quantity as being the percentage of algorithms that converge up to a predefined maximal iteration j_{max}. Where "convergence" as defined above. In fact this index is simply the sum of all bin values in the previous histogram that correspond to iteration-numbers no greater than j_{max}. PoC will be depicted as a function of the point standard deviation σ_p which is the most important factor that affects the performance of both algorithms.

In all experiments that follow we use $T_{\text{MSE}} = 1\,\text{pixel}^2$. In order to evaluate the behavior of the two algorithms under diverse uncertainty conditions imposed by the

presence of noise we consider different combinations standard deviation values for σ_p and σ_i.

4.3.1. *Experiment I*

In the first experiment, the alignment algorithms try to compensate both the geometric and photometric distortions that have been applied to images. Specifically, we use the "Takeo" image [4] as the warped profile image and by using the procedure described in Subsection 4.2 we generate a reference image which in the sequel is photometrically distorted as described in (23). We make 500 realizations of image pairs, run the three algorithms for all integer values of σ_p in the range $[1, 10]$, and compare their convergence characteristics for a maximum number of iterations $j_{max} = 15$ and for a threshold value $T_{MSD} = 1\text{pixel}^2$.

Figure 2 depicts the relative performance of the algorithms. Three cases are investigated; (a) $\sigma_p = 2$, (b) $\sigma_p = 6$ and (c) $\sigma_p = 10$. In most cases the proposed algorithm exhibits a smaller MSD which, especially compared to LK scheme, is order(s) of magnitude better. Specifically, for weak geometric deformation all algorithms reach almost the same floor value, with ECC being slightly faster than the two others. In the case of medium or strong deformations, ECC and SIC seems to behave equivalently at first iterations, while in the sequel ECC makes quite ($\sigma_p = 6$) or a little ($\sigma_p = 10$) bigger steps. Furthermore concerning the PoC, as we can see from Figure 2.(d), ECC and SIC algorithms achieve near PoC scores for almost all values of σ_p, while LK algorithm is vitally affected, especially from strong deformation.

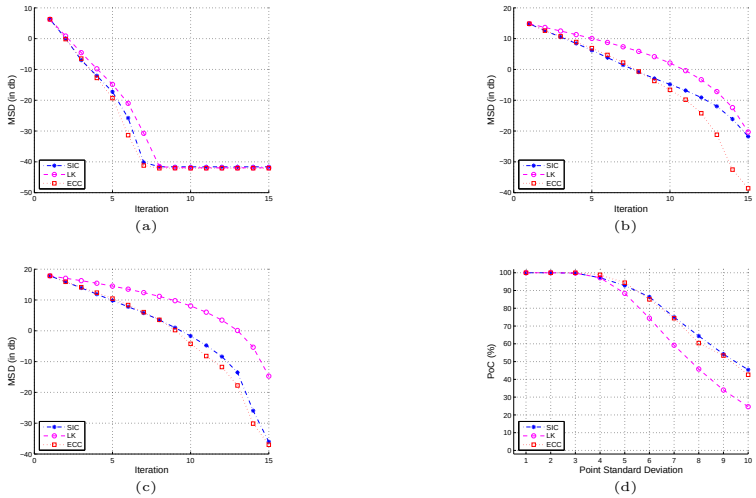


Fig. 2. MSD in dB as a function of number of iterations; (a) $\sigma_p = 2$, (b) $\sigma_p = 6$, (c) $\sigma_p = 10$. In (d), PoC as a function of σ_p for $j_{\max} = 15$.

Regarding the second figure of merit, that is the histogram of successful convergence, the resulting graphs are shown in Figure 3. The left graph of Figure 3 has the histogram for the case of $\sigma_p = 6$ and the right one for $\sigma_p = 10$. We can see that LK behaves worse than ECC and SIC algorithm, with the SIC being a little better than ECC. For example, 13% of SIC runs and 10% of ECC runs passed the threshold in the fourth iteration for $\sigma_p = 6$ while $6,5\%$ of SIC runs and 5% of ECC runs passed the threshold in the sixth iteration for the case of $\sigma_p = 10$.

Fig. 3. Intensity noise-free case. Histogram of successful convergence as a function of iteration number; left graph correspond to $\sigma_p = 6$ and right to $\sigma_p = 10$.

4.3.2. *Experiment II*

In this experiment, we repeat the simulation of the previous case, but now we add intensity noise to both images before their alignment. Specifically, the standard deviation of the noise we add into the images is equal to 8 gray levels.

In Figure 4 the simulation results we obtained are shown. For the case of $\sigma_p = 2$ we observe that the algorithms reach a different MSD floor value, with the proposed algorithm achieving a floor value which is almost $3 - dB$ and $5 - dB$ lower than the floor value achieved by LK and SIC algorithm respectively. Besides, ECC algorithms converges sensibly faster than the others algorithms. In the case of medium and strong geometric deformation, LK and SIC algorithm present a very low rate of convergence, while ECC seems to be relatively more robust. Regarding the PoC, ECC algorithm exhibits a larger PoC score for all values of σ_p confirming thus its superiority. We must stress here that Figure 4.(c) is meaningless since the common set of realizations where the algorithms have converged is just 7% of all realizations.

The superiority of ECC algorithm against the others can also be validated from the histograms of Figure 5. As we can clearly see, a sufficient part of ECC runs passed the threshold of $0dB$ in small iteration numbers. On the other hand, the support area of SIC and LK distributions contains bigger iteration indices. So, ECC algorithm seems to converge faster and with higher probability than the other two algorithms.

From a comparison of noise and noise-free case results, we conclude that noise

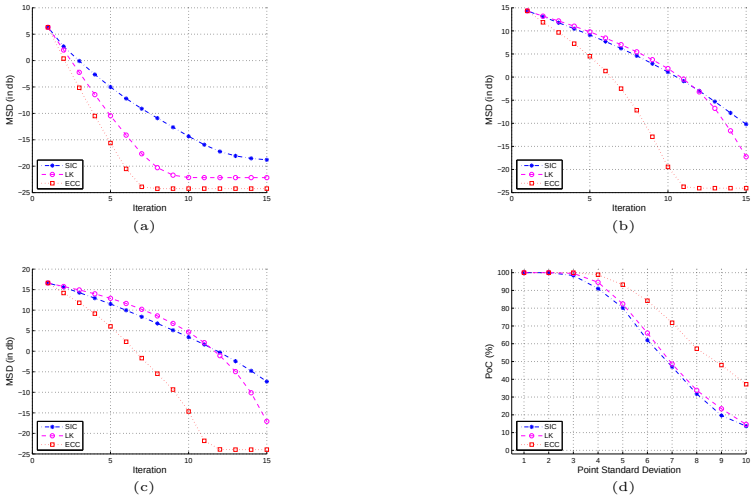


Fig. 4. MSD in dB as a function of number of iterations for the noisy (8 gray lelels) "Takeo" image; (a) $\sigma_p = 2$, (b) $\sigma_p = 6$, (c) $\sigma_p = 10$. In (d), PoC as a function of σ_p for $j_{\max} = 15$.

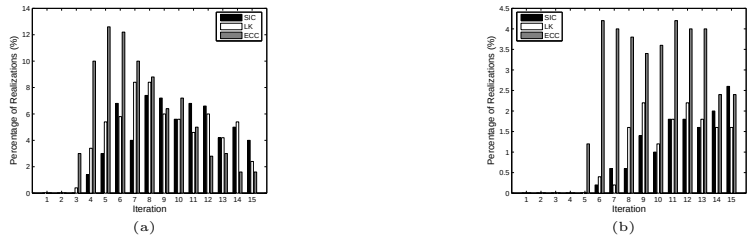


Fig. 5. Intensity noise case. Histogram of successful convergence as a function of iteration number; left graph correspond to $\sigma_p = 6$ and right to $\sigma_p = 10$.

affects relatively more SIC algorithm than ECC or LK scheme. This is in accordance with [4], where it is pointed out that forwards variants of algorithms are more robust to intensity noise compared with the backward ones.

4.3.3. *Experiment III*

In this experiment we examine the behavior of the algorithms under the influence of over-modelling. Specifically, we create the reference profiles by using an affine transform, but we still model the warping process by using a nonlinear projective transformation. As long as six parameters are required for the affine transform, the values of two more parameters $(p7; p8)$ must be estimated by the alignment algorithms. Ideally these values must be equals to zero. Because of the over-modelling case, a slight reduction in the performance of the algorithms is expected. Since we

also use here noisy and intensity distorted images, the evaluation of the algorithms for strong geometric transformation becomes meaningless. To this end, we restrict ourselves on weak or medium deformation; so σ_p takes values in the range $[1,\ 5]$.

In Figure 6 two cases are presented; $\sigma_p = 2$ and $\sigma_p = 4$. In the case of $\sigma_p = 2$, ECC algorithm reach $-25dB$ floor in 9 iterations, while LK scheme needs 12 iteration steps to achieve an MSD value near to $-23dB$ and SIC algorithm requires more than 15 iterations to reach a steady state (approximately $-20dB$). The superiority of ECC algorithm is also confirmed from the learning curves for the case of $\sigma_p = 4$, since the other two algorithms have sensibly lower rate of convergence.

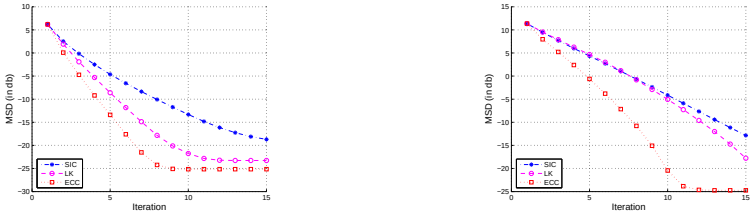


Fig. 6. MSD in dB as a function of number of iterations for the over-modelling case; (a) $\sigma_p = 2$, (b) $\sigma_p = 4$.

As long as in this case we investigate minor geometric distortions, instead of plotting PoC curves, we prefer to present PoC scores of the algorithms for different convergence threshold values. Table 1 shows PoC scores for $0dB$, $-10dB$ and $-20dB$ when σ_p takes values in $[1,\ 5]$. Table results confirm that ECC algorithm converges with high probability, even in cases where a strict threshold is required. Notice that in the case of $\sigma_p = 5$ and when the acceptance threshold is $-20dB$, ECC algorithm achieves $80,6\%$ PoC score, while LK and SIC algorithm achieve 56% and $18,2\%$ respectively.

	SIC			LK			ECC		
Threshold:	$0dB$	$-10dB$	$-20dB$	$0dB$	$-10dB$	$-20dB$	$0dB$	$-10dB$	$-20dB$
$\sigma_p = 1$	100%	100%	$47,4\%$	100%	100%	$92,8\%$	100%	100%	$98,8\%$
$\sigma_p = 2$	$99,8\%$	$99,4\%$	$49,2\%$	100%	100%	$92,8\%$	100%	100%	$98,0\%$
$\sigma_p = 3$	$96,8\%$	$92,0\%$	$39,6\%$	$98,8\%$	$97,6\%$	$88,6\%$	$99,8\%$	$99,6\%$	$96,8\%$
$\sigma_p = 4$	$89,8\%$	$80,0\%$	$27,6\%$	$90,2\%$	$86,6\%$	$75,2\%$	$96,6\%$	$95,0\%$	$92,2\%$
$\sigma_p = 5$	$75,8\%$	$63,4\%$	$18,2\%$	$74,2\%$	$69,2\%$	$56,0\%$	$86,2\%$	$84,4\%$	$80,6\%$

Finally similar conclusion with that of the previous experiment can be drawn from Figure 7 where the obtained histograms with the percentages of successful convergence are depicted.

Fig. 7. Over-modelling case. Histogram of successful convergence as a function of iteration number; left graph correspond to $\sigma_p = 2$ and right to $\sigma_p = 4$.

5. Conclusions

In this work a recently proposed alignment algorithm was used in the projective registration problem. This algorithm aims at maximizing the Enhanced Correlation Coefficient function which is a robust similarity measure against both geometric and photometric distortions. The optimal parameters are obtained by solving, iteratively, a sequence of approximate nonlinear optimization problems, which enjoy a simple closed-form solution with low computational cost. Regarding the applicability of the algorithm, three interesting experiments with photometrically distorted images were investigated. The iterative algorithm was compared against the original Lucas-Kanade algorithm and the recently proposed Simultaneous Inverse Compositional algorithm, through numerous simulation examples involving ideal and noisy conditions, strong or weak geometric deformations and modelling or over-modelling case. In most cases the proposed algorithm exhibited a better performance with an improvement in speed, as well as in probability of convergence as compared to both the other algorithms. *However, in noise-free case, SIC algorithm exhibits a comparable behavior to the proposed algorithm.*

References

1. Y. Altunbasak, R. M. Mersereau, and A. J. Patti. A fast parametric motion estimation algorithm with illumination and lens distortion correction. *IEEE Transactions on Image Processing*, 12(4):395–408, 2003.
2. P. Anandan. A computational framework and an algorithm for the measurement of visual motion. *International Journal of Computer Vision*, 2(3):283–310, 1989.
3. S. Baker, R. Gross, and I. Matthews. *Lucas-Kanade 20 Years On: A Unifying Framework: Part 3. Technical Report CMU-RI-TR-03-35*. Robotics Institute, Carnegie Mellon University, 2003.
4. S. Baker and I. Matthews. Lucas-kanade 20 years on: A unifying framework: Part 1: The quantity approximated, the warp update rule, and the gradient descent approximation. *International Journal on Computer Vision*, 56(3):221–255, 2004.
5. M. J. Black and P. Anandan. A framework for the robust estimation of optical flow. In *Proc. of 4th IEEE International Conference on Computer Vision (ICCV'93)*, 1993. Berlin, Germany.
6. G. D. Evangelidis and E. Z. Psarakis. Parametric image alignment using enhanced correlation coeeficient maximization. *submitted to IEEE Trans. on PAMI, submission TPAMI-0026-0107*.

7. C. Fuh and P. Maragos. Motion dislpacement estimation using an affine model for image matching. *Optical Engineering*, 30(7):881–887, 1991.
8. G. D. Hager and P. N. Belhumeur. Efficient region tracking with parametric models of geometry and illumination. *IEEE Trans. on PAMI*, 20(10):1025–1039, 1998.
9. B. K. P. Horn and B. G. Schunk. Determining optical flow. *Artificial Intelligence*, 17:185–203, 1981.
10. B. K. P. Horn and E. J. Weldon. Direct methods for recovering motion. *International Journal of Computer Vision*, 2:51–76, 1988.
11. H. Johnson and G. Christensen. Consistent landmark and intensity-based image registration. *IEEE Transactions on Medical Imaging*, 21(5):450–461, 2002.
12. B. D. Lucas and T. Kanade. An iterative image registration technique with an application to stereo vision. In *Proc. of 7th International Joint Conf on Artificial Intelligence (IJCAI)*, 1981. Vancouver, British Columbia.
13. E. Z. Psarakis and G. D. Evangelidis. An enhanced correlation-based method for stereo correspondence with sub-pixel accuracy. In *Proc. of 10th IEEE International Conference on Computer Vision (ICCV 2005)*, 2005. Beijing, China.
14. R. Radke, P. Ramadge, T. Echigo, and S. Iisaku. Efficiently estimating projective transformations. In *Proc. of IEEE International Conference on Image Processing (ICIP'00)*, pages 232–235, 2000.
15. H. Shum and R. Szeliski. Construction of panoramic image mosaics with global and local alignment. *International Journal on Computer Vision*, 36(2):101–130, 2000.
16. R. Szeliski. *Handbook of Mathematical Models of Computer Vision*. Springer, 2005. ch. 17.