Short Papers

Parametric Image Alignment Using Enhanced Correlation Coefficient Maximization

Georgios D. Evangelidis and
Emmanouil Z. Psarakis

Abstract—In this work, we propose the use of a modified version of the correlation coefficient as a performance criterion for the image alignment problem. The proposed modification has the desirable characteristic of being invariant with respect to photometric distortions. Since the resulting similarity measure is a nonlinear function of the warp parameters, we develop two iterative schemes for its maximization, one based on the forward additive approach and the second on the inverse compositional method. As is customary in iterative optimization, in each iteration, the nonlinear objective function is approximated by an alternative expression for which the corresponding optimization is simple. In our case, we propose an efficient approximation that leads to a closed-form solution (per iteration) which is of low computational complexity, the latter property being particularly strong in our inverse version. The proposed schemes are tested against the Forward Additive Lucas-Kanade and the Simultaneous Inverse Compositional (SIC) algorithm through simulations. Under noisy conditions and photometric distortions, our forward version achieves more accurate alignments and exhibits faster convergence, whereas our inverse version has similar performance as the SIC algorithm but at a lower computational complexity.

Index Terms—Image registration, motion estimation, gradient methods, parametric motion, correlation coefficient.

1 Introduction

The parametric image alignment problem consists of finding a transformation which aligns two image profiles. The profiles can either be entire images, as in the image registration problem [1], [2], or subimages, as in the region tracking [3], [4], [5], motion estimation [6], [7], [8], [9], and stereo correspondence [10], [11] problems. In image registration, the alignment problem needs to be solved only once, whereas, in region tracking, a template image has to be matched over a sequence of images. Finally, in motion estimation and stereo correspondences, the goal is to find the correspondence for all image points in a pair of images.

The alignment problem can be seen as a mapping between the coordinate systems of two images; therefore, the first step toward its solution is the suitable selection of a geometric transformation that adequately models this mapping. Existing models are basically parametric [12] and their exact form heavily depends on the specific application and the strategy selected to solve the alignment problem [3], [13]. The class of affine transformations and, in particular, several special cases (as pure translation) have been the center of attention in many applications [1], [2], [3], [4], [6], [10], [11], [13]. Alternative approaches rely on projective transformations (homography) and, more generally, on nonlinear transformations [5], [13], [14], [15].

• *The authors are with the Signal Processing and Communications Lab, Department of Computer Engineering and Informatics, University of Patras, 26504 Rio-Patras, Greece.*
E-mail: {evagelid, psarakis}@ceid.upatras.gr.

Manuscript received 17 Jan. 2007; revised 12 Feb. 2008; accepted 7 Apr. 2008; published online 2 May 2008.
Recommended for acceptance by F. Dellaert.
For information on obtaining reprints of this article, please send e-mail to: tpami@computer.org, and reference IEEECS Log Number TPAMI-0026-0107.
Digital Object Identifier no. 10.1109/TPAMI.2008.113.

Once the geometric parametric transformation has been defined, the alignment problem reduces itself to a parameter estimation problem. Therefore, the second step toward its solution consists of coming up with an appropriate performance measure, that is, an objective function. The latter, when optimized, will yield the optimum parameter estimates. Most existing approaches adopt measures that rely on l_p norms of the error between either the whole image profiles (*pixel-based techniques*) or a specific feature of the image profiles (*feature-based techniques*) [12]. Clearly, the l_2 norm is by far the most popular selection so far [1], [3], [6], [7], [9], [10], [13], [15], [16]. The l_2-based objective function is usually referred to as the Sum-Squared-Differences (SSD) measure and the corresponding optimization problem is known as the SSD technique [5], [9]. Variations on this approach have been proposed for the important problem of optical flow determination [5], [7], [17], and robust versions that can combat outliers were developed in [18].

For the *optimum* parameter estimation, all existing objective functions require nonlinear optimization techniques. Depending on the adopted solution strategy, the corresponding techniques can be broadly classified into two categories. The first includes gradient-based or differential approaches and the second includes direct search techniques [12]. Gradient-based schemes, because of their low computational cost, are regarded as more well fitted to CV applications [13], [19]. They are, however, characterized by noticeable convergence failure whenever homogeneous areas and/or single slanted edges (aperture problem [20]) are present. Meaningless estimates may also arise whenever we have strong displacement values. Direct search techniques, on the other hand, do not suffer the latter drawback. Indeed, these approaches can easily accommodate large motions since they rely on global image searches. Unfortunately, the latter require an exceedingly high computational cost, which becomes more intense in the cases of fine quantization needed in the case of accurate estimates [6]. Efforts to reduce complexity by adopting interpolation instead of fine quantization or hybrid techniques that combine the two classes can be found in [9], [15].

A common assumption encountered in most existing techniques is the *brightness constancy* of corresponding points or regions in the two profiles [20]. However, this assumption is valid only in specific cases and it is obviously violated under varying illumination conditions. There, it becomes clear that, in a practical situation, it is important that the alignment algorithm be able to take into account illumination changes. Alignment techniques that compensate for photometric distortions in contrast and brightness have been proposed in [1], [6], [8], [10], [16]. Alternative schemes make use of a set of basis images for handling arbitrary lighting conditions [3], [21] or use spatially dependent photometric models [7].

In this paper, we adopt a recently proposed similarity measure [11], the *enhanced correlation coefficient*, as our objective function for the alignment problem. Our measure is characterized by two very desirable properties. First, it is invariant to photometric distortions in contrast and brightness. Second, although it is a nonlinear function of the parameters, the iterative scheme we are going to develop for the optimization problem will turn out to be linear, thus requiring reduced computational complexity. Despite the resemblance of our final algorithm to well-known variants of the Lucas-Kanade alignment method which take lighting changes into account [10], [19], its performance, as we are going to see, is notably superior. We would like to mention that the enhanced correlation coefficient criterion was successfully applied to the problem of 1D translation estimation in stereo correspondence [11] and 2D translation estimation in registration [2].

The remainder of this paper is organized as follows: In Section 2, we formulate the parametric image alignment problem. Section 3 contains our main analytic results, namely, the definition of our objective function, the development of a forward and an inverse compositional iterative scheme for its optimization, and the relation of the proposed schemes to existing SSD techniques. In Section 4, our schemes are tested in a number of experiments against the currently most popular algorithms, namely, the Lucas-Kanade and Simultaneous Inverse Compositional (SIC) methods. Finally, Section 5 contains our conclusions.

2 Problem Formulation

Suppose we are given a pair of image profiles (intensities) $I_r(\mathbf{x})$, $I_w(\mathbf{y})$, where the first is the *reference* or *template* image and the second is the *warped* and $\mathbf{x} = [x_1, x_2]^t$, $\mathbf{y} = [y_1, y_2]^t$ denote coordinates. Suppose also that we are given a set of coordinates $\mathcal{T} = \{\mathbf{x}_k, k = 1, \dots, K\}$ in the reference image, which is called the *target area*. The alignment problem consists of finding the corresponding coordinate set in the warped image. Of course, we are not interested in arbitrary correspondences but, rather, in those that are structured and can be modeled with a well-defined vector mapping $\mathbf{y} = \boldsymbol{\phi}(\mathbf{x}; \mathbf{p})$, where $\mathbf{p} = [p_1, \cdots, p_N]^t$ is a vector of unknown parameters. Such correspondence problems often arise in practice, with the most common case being motion estimation in a sequence of images. In this application, due to the relative motion between scene and camera, whole (target) areas appear differently in time.

Assuming that a transformation model is given (and under the validity of the brightness constancy assumption), the alignment problem is simply reduced to the problem of *estimating* the parameters $\mathbf{p}$ such that

$$I_r(\mathbf{x}) = I_w(\boldsymbol{\phi}(\mathbf{x}; \mathbf{p})), \quad \forall\, \mathbf{x} \in \mathcal{T}. \tag{1}$$

In order to have a chance of obtaining a unique solution, it is necessary that the number N of unknown parameters does not exceed the number K of target coordinates. Of course, in practice, we usually have $N \ll K$, which suggests that (1) is an overdetermined system of (nonlinear) equations.

Most existing algorithms attempt to compute the parameter vector $\mathbf{p}$ by minimizing the *difference* or the *dissimilarity* of the two profiles. Dissimilarity is expressed through an objective function $E(\mathbf{p})$ which involves the l_p norm of the intensity difference of the two images. Since, in real applications, due to different viewing directions and/or different illumination conditions, the brightness constancy assumption is violated, it is necessary to include an additional photometric transformation $\Psi(I, \boldsymbol{\alpha})$ that accounts for the photometric changes and which is parameterized by a vector of unknown parameters $\boldsymbol{\alpha}$. A typical optimization problem has the following form:

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

We must mention that optimization problems of the form of (2) are often ill-posed and it is usually necessary to impose extra regularity (smoothness) conditions in order to obtain an acceptable solution [17].

Solving the optimization problem is clearly not a simple task because of the nonlinearity involved in the correspondence part. The computational complexity and estimation quality of the 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 our approach, which we detail in the next section.

3 Proposed Criterion and Main Results

Under the warping transformation $\boldsymbol{\phi}(\mathbf{x}; \mathbf{p})$, the coordinates $\mathbf{x}_k$, $k = 1, \dots, K$ of the target area $\mathcal{T}$ are mapped into the coordinates $\mathbf{y}_k(\mathbf{p}) = \boldsymbol{\phi}(\mathbf{x}_k; \mathbf{p})$, $k = 1, \dots, K$. Let us define the *reference vector* $\mathbf{i}_r$ and the corresponding *warped vector* $\mathbf{i}_w(\mathbf{p})$ as

$$\begin{aligned} \mathbf{i}_r &= [I_r(\mathbf{x}_1)\ I_r(\mathbf{x}_2) \cdots I_r(\mathbf{x}_K)]^t, \\ \mathbf{i}_w(\mathbf{p}) &= [I_w(\mathbf{y}_1(\mathbf{p}))\ I_w(\mathbf{y}_2(\mathbf{p})) \cdots I_w(\mathbf{y}_K(\mathbf{p}))]^t, \end{aligned} \tag{3}$$

and denote with $\bar{\mathbf{i}}_r$ and $\bar{\mathbf{i}}_w(\mathbf{p})$ their zero-mean versions, which are obtained by subtracting from each vector its corresponding arithmetic mean. We then propose the following criterion to quantify the performance of the warping transformation with parameters $\mathbf{p}$:

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

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

It is apparent from (4) that our criterion is invariant to bias and gain changes. This also suggests that our measure is going to be invariant to any photometric distortions in brightness and/or in contrast. Consequently, to a first approximation, we can completely disregard the photometric transformation and concentrate solely on the geometric. It is also interesting to mention that our measure exhibits statistical robustness against outliers, as is reported in [22]. All of these positive characteristics clearly support our expectation that the proposed criterion will turn out to be a suitable objective function for the parametric image alignment problem.

3.1 Performance Measure Optimization

Once the performance measure is specified, we then continue with its minimization in order to compute the optimum parameter values. It is straightforward to prove that minimizing $E_{\mathrm{ECC}}(\mathbf{p})$ is equivalent to *maximizing* the following *enhanced correlation coefficient* [11]:

$$\rho(\mathbf{p}) = \frac{\bar{\mathbf{i}}_r^t \bar{\mathbf{i}}_w(\mathbf{p})}{\|\bar{\mathbf{i}}_r\| \|\bar{\mathbf{i}}_w(\mathbf{p})\|} = \hat{\mathbf{i}}_r^t \frac{\bar{\mathbf{i}}_w(\mathbf{p})}{\|\bar{\mathbf{i}}_w(\mathbf{p})\|}, \tag{5}$$

where, for simplicity, we denote with $\hat{\mathbf{i}}_r = \bar{\mathbf{i}}_r / \|\bar{\mathbf{i}}_r\|$ the normalized version of the zero-mean reference vector, which is constant. 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.

As was mentioned in Section 1, maximizing $\rho(\mathbf{p})$ can be performed either by using direct search or by gradient-based approaches. Here, we are going to use the latter. As is customary in iterative techniques, we are going to replace the original optimization problem with a sequence of secondary optimizations. Each secondary optimization relies on the outcome of its predecessor, thus generating a chain of parameter estimates which hopefully converges to the desired optimizing vector. At each iteration, we do not have to optimize the objective function but an *approximation* to this function. Of course, the approximation must be selected so that the resulting optimizers are simple to compute. Next, let us introduce the approximation we are going to apply for our objective function and derive the solution that maximizes it.

Assume 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 perturbations. Let $\tilde{\mathbf{y}} = \boldsymbol{\phi}(\mathbf{x}; \tilde{\mathbf{p}})$ be the warped coordinates under the nominal parameter vector and $\mathbf{y} = \boldsymbol{\phi}(\mathbf{x}; \mathbf{p})$ under the perturbed ones. Considering the intensity of the warped image at coordinates $\mathbf{y}$ and applying a first-order Taylor expansion with respect to the parameters, then we can write

$$I_w(\mathbf{y}) \approx I_w(\tilde{\mathbf{y}}) + \left[\nabla_{\mathbf{y}} I_w(\tilde{\mathbf{y}})\right]^t \frac{\partial \boldsymbol{\phi}(\mathbf{x};\tilde{\mathbf{p}})}{\partial \mathbf{p}} \Delta\mathbf{p}, \tag{6}$$

where $\nabla_{\mathbf{y}} I_w(\tilde{\mathbf{y}})$ denotes the gradient vector of length 2 of the intensity function $I_w(\mathbf{y})$ of the warped image, evaluated at the nominal warped coordinates $\tilde{\mathbf{y}}$. Since $\boldsymbol{\phi}(\mathbf{x};\mathbf{p})$ is a vector transformation of length 2 (in order to yield the warped coordinates), then $\frac{\partial \boldsymbol{\phi}(\mathbf{x};\tilde{\mathbf{p}})}{\partial \mathbf{p}}$ denotes the size $2 \times N$ Jacobian matrix of the transform with respect to the parameters, evaluated at the nominal parameter values. Note that we have silently assumed that the intensity function I_w and the warping transformation $\boldsymbol{\phi}$ are of sufficient smoothness to allow for the existence of the required partial derivatives.

We can now apply (6) for all coordinates $\mathbf{x}_k$, $k = 1, \ldots, K$, of the target area $\mathcal{T}$. This will yield the following linearized version of the warped vector with parameters $\mathbf{p}$:

$$\mathbf{i}_w(\mathbf{p}) \approx \mathbf{i}_w(\tilde{\mathbf{p}}) + G(\tilde{\mathbf{p}})\Delta\mathbf{p}, \tag{7}$$

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 $\tilde{\mathbf{p}}$. In order to specify exactly this matrix, let us assume that the warping transformation is of the form

$$\boldsymbol{\phi}(\mathbf{x};\mathbf{p}) = [\phi_1(\mathbf{x};\mathbf{p}), \phi_2(\mathbf{x};\mathbf{p})]^t, \tag{8}$$

where ϕ_1, ϕ_2 are scalar functions. Then, the (k,n) element of the matrix G can be written as

$$G(\tilde{\mathbf{p}})_{k,n} = \sum_{i=1}^{2} \left(\left. \frac{\partial I_w(\mathbf{y})}{\partial y_i} \right|_{\mathbf{y}=\mathbf{y}_k(\tilde{\mathbf{p}})} \times \left. \frac{\partial \phi_i(\mathbf{x}_k;\mathbf{p})}{\partial p_n} \right|_{\mathbf{p}=\tilde{\mathbf{p}}} \right), \tag{9}$$

where $k = 1, \ldots, K$; $n = 1, \ldots, N$, and we recall that $\mathbf{y} = [y_1, y_2]^t$ are the coordinates in the warped image.

We now need to compute the zero-mean version of the warped vector. With the help of (7), we obtain the following approximation of the objective function $\rho(\mathbf{p})$ defined in (5):

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

where $\bar{G}(\tilde{\mathbf{p}})$ and $\bar{\mathbf{i}}_w(\tilde{\mathbf{p}})$ are the column-zero-mean versions of $G(\tilde{\mathbf{p}})$ and $\mathbf{i}_w(\tilde{\mathbf{p}})$, respectively.

From now on, let us, for notational simplicity, drop the dependence of the warped vectors on $\mathbf{p}$; we can then write our previous approximation as follows:

$$\rho(\Delta\mathbf{p}|\tilde{\mathbf{p}}) = \frac{\hat{\mathbf{i}}_r^t \bar{\mathbf{i}}_w + \hat{\mathbf{i}}_r^t \bar{G}\Delta\mathbf{p}}{\sqrt{\|\bar{\mathbf{i}}_w\|^2 + 2\bar{\mathbf{i}}_w^t \bar{G}\Delta\mathbf{p} + \Delta\mathbf{p}^t \bar{G}^t \bar{G}\Delta\mathbf{p}}}. \tag{11}$$

Although $\rho(\Delta\mathbf{p}|\tilde{\mathbf{p}})$ is nonlinear in $\Delta\mathbf{p}$, its maximization is simple and results in a closed-form expression. This is a consequence of the next theorem, which provides the necessary result.

Theorem 1. *Consider the scalar function*

$$f(\mathbf{x}) = \frac{u + \mathbf{u}^t\mathbf{x}}{\sqrt{v + 2\mathbf{v}^t\mathbf{x} + \mathbf{x}^t Q\mathbf{x}}}, \tag{12}$$

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

$$v > \mathbf{v}^t Q^{-1} \mathbf{v}, \tag{13}$$

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

$$\max_{\mathbf{x}} f(\mathbf{x}) = \sqrt{\frac{(u - \mathbf{u}^t Q^{-1}\mathbf{v})^2}{v - \mathbf{v}^t Q^{-1}\mathbf{v}} + \mathbf{u}^t Q^{-1}\mathbf{u}}, \tag{14}$$

which is attainable for

$$\mathbf{x} = Q^{-1}\left\{\frac{v - \mathbf{v}^t Q^{-1}\mathbf{v}}{u - \mathbf{u}^t Q^{-1}\mathbf{v}}\mathbf{u} - \mathbf{v}\right\}. \tag{15}$$

Case $u \leq \mathbf{u}^t Q^{-1}\mathbf{v}$: Here, we have a supremum which is equal to

$$\sup_{\mathbf{x}} f(\mathbf{x}) = \sqrt{\mathbf{u}^t Q^{-1}\mathbf{u}} \tag{16}$$

and can be approached arbitrarily close by selecting

$$\mathbf{x} = Q^{-1}\{\lambda\mathbf{u} - \mathbf{v}\}, \tag{17}$$

with λ positive scalar and of sufficiently large value.[1]

Proof. The proof makes repeated use of the Schwartz inequality. All details are presented in the Appendix. □

Let us now examine whether we can apply Theorem 1 for the maximization of $\rho(\Delta\mathbf{p}|\tilde{\mathbf{p}})$ defined in (11). For this, we need to verify the validity of (13). For the problem of interest, this translates into the following inequality: $\|\bar{\mathbf{i}}_w\|^2 > \bar{\mathbf{i}}_w^t P_G \bar{\mathbf{i}}_w$, where $P_G = \bar{G}(\bar{G}^t\bar{G})^{-1}\bar{G}^t$. This relation is trivially satisfied because P_G is an orthogonal projection operator (i.e., $P_G^2 = P_G$ and $P_G^t = P_G$) and, therefore, we can write

$$\|\bar{\mathbf{i}}_w\|^2 = \|P_G\bar{\mathbf{i}}_w\|^2 + \|[I - P_G]\bar{\mathbf{i}}_w\|^2 \geq \|P_G\bar{\mathbf{i}}_w\|^2 = \bar{\mathbf{i}}_w^t P_G \bar{\mathbf{i}}_w, \tag{18}$$

where I denotes the identity matrix. We have equality if and only if $[I - P_G]\bar{\mathbf{i}}_w = 0$, which is true whenever $\bar{\mathbf{i}}_w$ is a linear combination of the columns of $\bar{G}$. Clearly, the probability of this happening is zero, especially under the presence of noise. Consequently, the desired inequality, for all practical purposes, is strict.

Since we can apply Theorem 1, according to (15), the optimizing perturbation is equal to

$$\Delta\mathbf{p} = (\bar{G}^t\bar{G})^{-1}\bar{G}^t\left\{\frac{\|\bar{\mathbf{i}}_w\|^2 - \bar{\mathbf{i}}_w^t P_G \bar{\mathbf{i}}_w}{\hat{\mathbf{i}}_r^t\bar{\mathbf{i}}_w - \hat{\mathbf{i}}_r^t P_G \bar{\mathbf{i}}_w}\hat{\mathbf{i}}_r - \bar{\mathbf{i}}_w\right\}, \tag{19}$$

when $\hat{\mathbf{i}}_r^t\bar{\mathbf{i}}_w > \hat{\mathbf{i}}_r^t P_G\bar{\mathbf{i}}_w$; or, according to (17),

$$\Delta\mathbf{p} = (\bar{G}^t\bar{G})^{-1}\bar{G}^t\{\lambda\hat{\mathbf{i}}_r - \bar{\mathbf{i}}_w\}, \tag{20}$$

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 $\rho(\Delta\mathbf{p}|\tilde{\mathbf{p}})$ satisfies $\rho(\Delta\mathbf{p}|\tilde{\mathbf{p}}) > \rho(0|\tilde{\mathbf{p}})$. In other words, we would like to select a perturbation that will increase the correlation and will make it nonnegative. The following lemma provides possible values for λ.

Lemma 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:*

$$\lambda_1 = \sqrt{\frac{\bar{\mathbf{i}}_w^t P_G \bar{\mathbf{i}}_w}{\hat{\mathbf{i}}_r^t P_G \hat{\mathbf{i}}_r}}, \quad \lambda_2 = \frac{\hat{\mathbf{i}}_r^t P_G\bar{\mathbf{i}}_w - \hat{\mathbf{i}}_r^t\bar{\mathbf{i}}_w}{\hat{\mathbf{i}}_r^t P_G \hat{\mathbf{i}}_r}. \tag{21}$$

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

1. More precisely, we mean that, for every $\epsilon > 0$, there exists a sufficiently large scalar λ_ϵ such that the resulting $f(\mathbf{x})$ is ϵ close to the upper bound.

TABLE 1
Outline of the Proposed Forward Additive ECC (FA-ECC) Refinement Algorithm

Initialization	
Use reference image I_r to compute the zero-mean normalized vector $\hat{\mathbf{i}}_r$. Initialize $\mathbf{p}_0$ and set $j=1$.	
Iteration Steps	
S_1 :	Using $\boldsymbol{\phi}(\mathbf{x};\mathbf{p}_{j-1})$ warp I_w and compute its zero-mean counterpart vector $\bar{\mathbf{i}}_w(\mathbf{p}_{j-1})$.
S_2 :	Using $\boldsymbol{\phi}(\mathbf{x};\mathbf{p}_{j-1})$ compute the Jacobian $\bar{G}(\mathbf{p}_{j-1})$ using (9).
S_3 :	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 $\Delta\mathbf{p}_j$ either from (19) or using (20) and (21).
S_4 :	Update parameter vector $\mathbf{p}_j=\mathbf{p}_{j-1}+\Delta\mathbf{p}_j$. If $\|\Delta\mathbf{p}_j\|\geq T$ then, $j++$ and goto S_1; else stop.

Proof. By substituting the value of $\Delta\mathbf{p}$ from (20) in (11), the objective function becomes the following function of λ:

$$f(\lambda)=\frac{\left(\hat{\mathbf{i}}_r^t\bar{\mathbf{i}}_w-\hat{\mathbf{i}}_r^tP_G\bar{\mathbf{i}}_w\right)+\lambda\hat{\mathbf{i}}_r^tP_G\hat{\mathbf{i}}_r}{\sqrt{\left(\|\bar{\mathbf{i}}_w\|^2-\bar{\mathbf{i}}_w^tP_G\bar{\mathbf{i}}_w\right)+\lambda^2\hat{\mathbf{i}}_r^tP_G\hat{\mathbf{i}}_r}}. \tag{22}$$

It is easy to verify that the derivative of $f(\lambda)$ is nonnegative; therefore, $f(\lambda)$ is increasing in λ. This suggests that, for $\lambda\geq\lambda_2$, we have $f(\lambda)\geq 0$. Notice now that, for $\lambda=\lambda_1$, we can write

$$f(\lambda_1)=\frac{\hat{\mathbf{i}}_r^t\bar{\mathbf{i}}_w-\hat{\mathbf{i}}_r^tP_G\bar{\mathbf{i}}_w+\sqrt{\left(\bar{\mathbf{i}}_w^tP_G\bar{\mathbf{i}}_w\right)\left(\hat{\mathbf{i}}_r^tP_G\hat{\mathbf{i}}_r\right)}}{\|\bar{\mathbf{i}}_w\|}\geq\rho(0|\tilde{\mathbf{p}}), \tag{23}$$

with the last inequality being a consequence of applying the Schwartz inequality on $\hat{\mathbf{i}}_r^tP_G\bar{\mathbf{i}}_w$ and recalling that P_G is an orthogonal projection operator. □

Remarks. One should expect, as $\bar{\mathbf{i}}_w$ approaches $\bar{\mathbf{i}}_r$, to use mostly (19) since, for $\bar{\mathbf{i}}_w\approx\bar{\mathbf{i}}_r$, we have $\hat{\mathbf{i}}_r^t\bar{\mathbf{i}}_w\approx\hat{\mathbf{i}}_r^t\bar{\mathbf{i}}_r>\hat{\mathbf{i}}_r^tP_G\bar{\mathbf{i}}_r\approx\hat{\mathbf{i}}_r^tP_G\bar{\mathbf{i}}_w$. It is interesting, however, to note that, if one insists on using (19) at all times, then, whenever $\hat{\mathbf{i}}_r^t\bar{\mathbf{i}}_w\leq\hat{\mathbf{i}}_r^tP_G\bar{\mathbf{i}}_w$ holds, we end up with a *negative* correlation $\rho(\Delta\mathbf{p}|\tilde{\mathbf{p}})$ (this being true even if $\rho(0|\tilde{\mathbf{p}})>0$) which is always *smaller* than $\rho(0|\tilde{\mathbf{p}})$. In other words, instead of increasing the correlation coefficient (as is the desired goal), in this case, we *decrease* it. This clearly suggests that it is preferable to use (20) with a value of λ, as indicated in Lemma 1, (21).

3.2 Forward Additive ECC Iterative Algorithm

Let us now translate the above results into an *iterative* scheme in order to obtain the solution to the original nonlinear optimization problem. Assuming that estimate $\mathbf{p}_{j-1}$ of the parameter vector is available from iteration $j-1$, we can compute $\bar{\mathbf{i}}_w(\mathbf{p}_{j-1})$ and $\bar{G}(\mathbf{p}_{j-1})$; then, we can approximate $\rho(\mathbf{p})$ following (10) with the help of $\rho(\Delta\mathbf{p}_j|\mathbf{p}_{j-1})$ and optimize this approximation with respect to $\Delta\mathbf{p}_j$. This will lead to the parameter update rule $\mathbf{p}_j=\mathbf{p}_{j-1}+\Delta\mathbf{p}_j$. As is indicated in Step S_4, we stop iterating whenever the norm of the updating vector $\Delta\mathbf{p}_j$ becomes smaller than some predefined threshold value T. The iteration steps are summarized in Table 1 and we call the corresponding algorithm the Forward Additive ECC (FA-ECC).

Given the number K of pixels in the target area $\mathcal{T}$ and the parameter vector estimate $\mathbf{p}_{j-1}$ of length N, the complexity per iteration of the proposed scheme can be easily estimated. From Table 1 and taking into account that, usually, $K\gg N$, we realize that the most computationally demanding part is Step S_3, which involves the computation of $\Delta\mathbf{p}_j$ with the help of (19) or (20). As we can see, in this step, we need to form the matrix $\bar{G}^t\bar{G}$, which requires $O(KN^2)$ operations. This is the leading complexity in our algorithm since all other steps require at most $O(KN)$ per iteration.

3.3 Inverse Compositional ECC Iterative Algorithm

When the alignment problem is restricted to specific classes of parametric models, it is possible to devise more computationally efficient versions since certain parts of the algorithm can be computed offline [3], [13], [15]. If, for example, we adopt the methodology proposed in [19], we can come up with the Inverse Compositional ECC (IC-ECC) version of our algorithm which has the significantly reduced complexity $O(KN)$ per iteration. We briefly mention that the methodology found in [3], [13] relies on interchanging the role of $\mathbf{i}_w$ and $\mathbf{i}_r$. Consequently, matrix G becomes the Jacobian matrix of the reference intensity vector and since the warping function for this vector is the identity, matrix G is constant and $\bar{G}^t\bar{G}$ can be computed offline. The latter is the reason behind the one order of magnitude reduction in computational complexity. The outline of our alternative algorithmic version IC-ECC can be easily obtained from Table 1 by appropriately modifying our FA-ECC version.

Regarding inverse algorithms (additive and compositional) as well as the forward compositional algorithm [15], we should point out that they can be applied *only* to specific classes of warps. It is also known that inverse algorithms are more susceptible to noisy conditions than their forward counterparts [13]. These important weaknesses limit the usage of such algorithms in practice.

3.4 Relation to Existing SSD-Based Measures

In this section, we are going to derive our performance measure in a different way. This will also help us in relating it to the two currently most popular SSD approaches in the literature. For our analysis, we are going to assume that photometric distortion is limited only to global brightness and contrast changes. Under this simple type of photometric changes, we can define the following performance measure for our parametric alignment problem:

$$E(\mathbf{p},\boldsymbol{\alpha})=\|\alpha_1\mathbf{i}_w(\mathbf{p})+\alpha_2-\mathbf{i}_r\|^2, \tag{24}$$

where $\boldsymbol{\alpha}=[\alpha_1\ \alpha_2]^t$ is the parameter vector for the photometric transformation. Our goal, of course, is to minimize the objective function with respect to all parameters. Regarding the first photometric parameter, we must point out that negative values of α_1 produce the *inversion* effect, where colors are reversed. Consequently, if there exists the a priori knowledge that such a color inversion cannot take place, then it is logical to limit α_1 only to *positive* values. Now, if we first minimize the objective function with respect to α_1,α_2, we obtain the following interesting result:

$$E(\mathbf{p})=\min_{\alpha_1\geq 0,\alpha_2}E(\mathbf{p},\boldsymbol{\alpha})=\|\bar{\mathbf{i}}_r\|^2\left\{1-[\max\{\rho(\mathbf{p}),0\}]^2\right\}, \tag{25}$$

where $\rho(\mathbf{p})$ is the correlation function defined in (5). Notice that, since the reference image is constant, so is the norm $\|\bar{\mathbf{i}}_r\|^2$ contained in the previous relation; therefore, further minimization with respect to $\mathbf{p}$ is equivalent to minimizing the term $(1-[\max\{\rho(\mathbf{p}),0\}]^2)$. But, this expression is decreasing in $\rho(\mathbf{p})$;

consequently, we can equivalently maximize the correlation function $\rho(\mathbf{p})$, thus recovering our criterion. The final optimization problem makes a lot of sense. Indeed, notice that, since $\rho(\mathbf{p})$ is free of photometric distortions (the simple type we consider here) and under the knowledge that there is no color inversion, it is quite plausible to look for the most *positive* correlation.

If we drop the constraint $\alpha_1 \geq 0$, then the minimization of the objective function in (25) is the optimization problem proposed by Fuh and Maragos [6]. By optimizing first with respect to α_1, α_2 yields

$$E_{\mathrm{FM}}(\mathbf{p}) = \min_{\alpha_1,\alpha_2} E(\mathbf{p},\boldsymbol{\alpha}) = \|\bar{\mathbf{i}}_r\|^2\{1-\rho^2(\mathbf{p})\}. \tag{26}$$

Notice that the resulting measure is now a decreasing function of $|\rho(\mathbf{p})|$; therefore, any further minimization with respect to $\mathbf{p}$ is equivalent to maximizing the absolute value $|\rho(\mathbf{p})|$ of the correlation function. It is clear that this optimization problem does not take into account the prior knowledge that there is no color inversion. In [6], maximization was achieved by adopting an exhaustive search approach in the N-D quantized parameter space. Clearly, in a noncolor-inversion situation, such a search will give rise to the correct maximum positive correlation (provided, of course, that the warped image does not contain parts that are the negative of the target area). However, as we mentioned in Section 1, exhaustive search approaches are characterized by high computational complexity, which becomes exceedingly demanding when we are interested in fine subpixel accuracy.

Although not proposed in [6], alternatively, we could adopt an iterative approach similar to the one suggested for our measure. If, however, we attempt to maximize $|\rho(\mathbf{p})|$ using the same approximation as in (10), then one can show that the optimum perturbation $\Delta\mathbf{p}$ is always given by (19). As was indicated in our remarks (after Lemma 1), adopting this strategy may result in negative correlations corresponding to local minima for $\rho(\mathbf{p})$ instead of the desired maxima. In other words, there are more chances for the iterative algorithm to be locked in erroneous local extrema than is the case with our approach.

An alternative measure arises if, in (24), we interchange the roles of $\mathbf{i}_w$ and $\mathbf{i}_r$, that is,

$$E(\mathbf{p},\boldsymbol{\alpha}) = \|\alpha_1\mathbf{i}_r + \alpha_2 - \mathbf{i}_w(\mathbf{p})\|^2. \tag{27}$$

This is the approach adopted by Lucas and Kanade [10] and it is known to generate, along with its variants, the most widely used algorithms in practice. Following similar steps as in the previous two cases, let us first minimize with respect to the two photometric parameters. This yields

$$E_{\mathrm{LK}}(\mathbf{p}) = \min_{\alpha_1,\alpha_2} E(\mathbf{p},\boldsymbol{\alpha}) = \left\|\bar{\mathbf{i}}_w(\mathbf{p})\right\|^2\{1-\rho^2(\mathbf{p})\}. \tag{28}$$

We observe in the current outcome that the resulting criterion has two terms that depend on the parameters $\mathbf{p}$, namely, the familiar part $\{1-\rho^2(\mathbf{p})\}$ and the magnitude of the warped image $\|\bar{\mathbf{i}}_w(\mathbf{p})\|^2$ (which is not constant). Therefore, minimizing $E_{\mathrm{LK}}(\mathbf{p})$ with respect to the parameters involves the minimization of the combination of the two terms. The first observation is that this criterion will not necessarily produce the same solution as our measure. Second, due to the term $\|\bar{\mathbf{i}}_w(\mathbf{p})\|^2$, it is clear that an iterative algorithm can lock in solutions which result in $\|\bar{\mathbf{i}}_w(\mathbf{p})\|^2 \approx 0$ (for example, areas with uniform intensity). And, third, because of the term $\rho^2(\mathbf{p})$, the algorithm can lock in negative correlations.

Despite the previous observations, the Lucas-Kanade performance measure gives rise to the most popular iterative algorithms for the image alignment problem. For this reason, we are going to use it as a point of reference and compare it against our scheme. Consequently, let us present its Forward Additive LK (FA-LK) updating version in more detail. Substituting the linear approximation of $\bar{\mathbf{i}}_w(\mathbf{p})$ in (28), then minimizing with respect to $\Delta\mathbf{p}$, we obtain the following optimum updating perturbation:

$$\Delta\mathbf{p}_{\mathrm{LK}} = (\bar{G}^t\bar{G})^{-1}\bar{G}^t\left\{\frac{\hat{\mathbf{i}}_r^t\bar{\mathbf{i}}_w - \hat{\mathbf{i}}_r^t P_G\bar{\mathbf{i}}_w}{1-\hat{\mathbf{i}}_r^t P_G\hat{\mathbf{i}}_r}\hat{\mathbf{i}}_r - \bar{\mathbf{i}}_w\right\}, \tag{29}$$

which is applicable at all times. Comparing (19) with (29), we realize that the difference is only in the scalar quantity that precedes the vector $\hat{\mathbf{i}}_r$. As we are going to see, this seemingly slight variation, in combination with (20), will result in significant performance improvements.

For the Lucas-Kanade approach, it is possible to define a special SSD-based measure that can handle arbitrary linear appearance variations. For its minimization, an iterative algorithm that makes use of the inverse additive update rule was proposed in [3] by Hager and Belhumeur. Based on the same SSD measure, Baker et al. [19], by adopting the inverse compositional approach, proposed several variants of the Hager-Belhumeur algorithm. Among these alternative algorithmic schemes, the SIC algorithm is reported to have the best performance [19]. Therefore, this algorithm will also be tested in the next section.

4 Simulation Results

In this section, we perform a number of simulations in order to evaluate our FA-ECC and IC-ECC algorithmic version. As we mentioned above, we will also simulate the FA-LK algorithmic version that copes with photometric distortions and the SIC algorithm, which is considered to be the most effective inverse LK scheme. For all aspects affecting the simulation experiments, we made an effort to stay exactly within the framework specified in [13], [19]. To model the warping process, we are going to use the class of affine transformations. We know that the 2D rigid body or similarity transformation are members of this class. Furthermore, the Jacobian of the affine model is a constant matrix, meaning that it can be computed offline. Before proceeding with the presentation of our simulation results, let us first briefly present the experimental setup and the figures of merit we are going to adopt.

4.1 Experimental Setup and Figures of Merit

In order to create a reference and a warped image, we follow the procedure proposed in [13]. In brief, let $I(\mathbf{x})$ be a given image and $\mathbf{x}_i$, $i=1,2,3$, the coordinates of three points which define the boundaries of the desired target area. We perturb these points by adding Gaussian noise $\mathcal{N}(0,\sigma_p^2)$ (σ_p captures the strength of the geometric deformation), select a vector $\mathbf{x}_0$ such that the points $\mathbf{x}_0+\mathbf{x}_i$, $i=1,2,3$, lie in the interior of the support of the given image, and define the parameter vector $\mathbf{p}_r$ of the affine transformation that maps the original points to the translated noisy ones. We apply this transformation to all points of the target area to warp it. With the help of bilinear interpolation, we compute the new intensities. This process defines the reference profile $I_r(\mathbf{x})$. For the warped image, we use the given one.

All algorithms are initialized in the same way, namely, $\mathbf{p}_0 = [1\ 0\ 0\ 1\ \mathbf{x}_0^t]^t$. At iteration j, each algorithm provides the parameter estimates $\mathbf{p}_j$. In order to measure the quality of this estimate, we use the following quantity:

$$e(j) = \frac{1}{6}\sum_{i=1}^{3}\left\|\boldsymbol{\phi}(\mathbf{x_i};\mathbf{p}_r) - \boldsymbol{\phi}(\mathbf{x_i};\mathbf{p}_j)\right\|^2, \tag{30}$$

which quantifies the existing squared error between the exact warped version of the points $\mathbf{x}_i$, $i=1,2,3$, and their estimated counterparts.

By averaging this error over many realizations that differ in the point noise realization, we can compute the Mean Square Distance

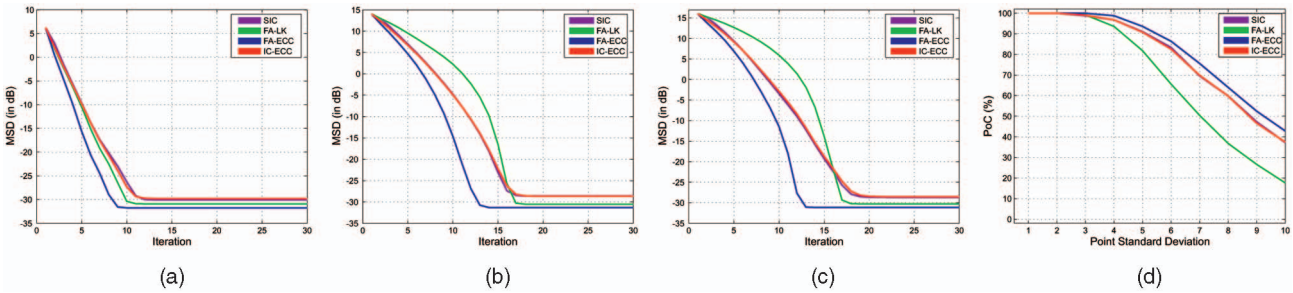

Fig. 1. MSD in decibels as a function of number of iterations under the presence of noise ($\sigma_i = 8$ gray levels). (a) $\sigma_p = 2$. (b) $\sigma_p = 6$. (c) $\sigma_p = 10$. In (d), PoC as a function of σ_p.

(MSD) value. Obviously, by computing this value in each iteration of an algorithm, we form a sequence that captures its *learning ability*. Of course, it is unrealistic to expect that any of the algorithms will 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 and have the right picture of this convergence characteristic, we adopt the idea followed in [13], namely, to define the MSD but *conditioned* on the event that *all* of the competing algorithms have converged. By "convergence," we mean that $e(j_{\max}) \leq T_{\text{MSD}}$. In other words, we consider that an algorithm has converged when its squared error $e(j)$ at a prescribed maximal iteration j_{max} is below a certain threshold level T_{MSD}.

The second quantity which is of importance is clearly the percentage of converging (PoC) runs. Therefore, we define this quantity as being the percentage of algorithms that converge up to a predefined maximal iteration 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 all algorithms.

Since it is only natural to prefer an algorithm that converges quickly *with high probability*, we propose a third figure of merit that captures exactly this aspect. Specifically, for characteristic values of σ_p and thresholds T_{MSD}, we apply the algorithms for a maximal number of iterations j_{MAX}. Then, we compute the *cumulative* PoC achieved by each algorithm as j_{max} increases from 0 to j_{MAX}. This third figure of merit is proposed here for the first time.

In all of the experiments, we use the "Takeo" image as the warped profile and generate a reference image as was previously described. We make 5,000 realizations of image pairs and we add independent and identically distributed, zero-mean Gaussian intensity noise of standard deviation σ_i before running the competing algorithms. Although in [13], [19] we find three different scenarios, here, due to lack of space, we only focus in the one where we add noise to both image profiles (since this is the most interesting from a practical viewpoint).

4.2 First Experiment

In this experiment, for the intensity noise, we use a standard deviation σ_i, which corresponds to eight gray levels, and compare the convergency characteristics of the competing algorithms for a maximum number of iterations[2] $j_{max} = 15$ and $T_{\text{MSD}} = 1$ pixel^2. Figs. 1a, 1b, and 1c depict the convergence profiles of the algorithms for different values of σ_p. We observe the appearance of an MSD floor value in each algorithm which is due to the presence of the intensity noise. Fig. 1d presents the corresponding PoC as a function of σ_p.

As we can see, each algorithm attains a different MSD floor value with our FA-ECC version converging to the lowest one and with a rate which can be significantly better. Specifically, for weak geometric deformations, all algorithms reach almost comparable floor values and have comparable convergence rates, with FA-ECC being slightly faster than its rivals. However, in the case of medium to strong deformations, FA-ECC reaches an MSD floor value which is 3 dB lower than the inverse versions and slightly lower than the FA-LK algorithm. On the other hand, the convergency rate of FA-ECC is significantly superior compared to all other algorithms. Regarding our IC-ECC version, as we can see, it has performance comparable to the SIC algorithm. The same characteristics also apply to PoC, where FA-ECC exhibits a larger percentage of successful convergences while IC-ECC matches the performance of SIC. Regarding the third figure of merit, we applied the algorithms for a maximal number of iterations $j_{MAX} = 100$. In order to test the accuracy of the alignment, we selected a threshold value $T_{\text{MSD}} = (1/18 \text{ pixel})^2$ (i.e., -25 dB), assuring that T_{MSD} is higher than the MSD floor value of all competing algorithms. Fig. 3a depicts the corresponding curves for three values of σ_p. As we can see, for weak deformations, all algorithms are almost completely successful after the 10th iteration. When, however, the geometric deformation becomes stronger, FA-ECC outperforms its competitors significantly. Again, IC-ECC is comparable to SIC.

4.3 Second Experiment

In this simulation, we consider the realistic case of photometrically distorted images under noisy conditions. We consider two different scenarios. We impose the photometric distortion 1) on the reference image and 2) on the warped one. Since all competing algorithms perfectly compensate for linear photometric distortions, we consider a nonlinear transformation of the form $I(\mathbf{x}) \leftarrow (I(\mathbf{x}) + 20)^{0.9}$, which is applied to the intensity of each image pixel. We repeat the same set of simulations as in the first experiment, only now we impose the photometric distortion before adding intensity noise.

The results we obtained are shown in Fig. 2. As we can see, the performance of our forward algorithm seems to be almost unaffected, achieving, under both scenarios, almost the same and the lowest MSD floor value. On the other hand, the performance of both inverse algorithms and FA-LK scheme seems to be vitally affected. Comparing Fig. 2 with Fig. 1, we observe that, under the first scenario, FA-ECC performs even better than before. In fact, the MSD floor value is now 3 and 5 dB lower than the value attained by the FA-LK algorithm and the inverse algorithms, respectively. We should note here that the MSD floor is due not only to the intensity noise but also to the photometric *model mismatch*. Under the second scenario, all algorithms achieve the same MSD floor value. As far as PoC is concerned, we observe a rather steady and robust behavior for the forward algorithms under both scenarios while inverse schemes, under the first scenario, exhibit a significant performance reduction as compared to the second one.

2. In order to make the different MSD floor values achieved by the competing algorithms in Figs. 1a, 1b, and 1c and Figs. 2a, 2b, and 2c visible, 30 iterations are shown.

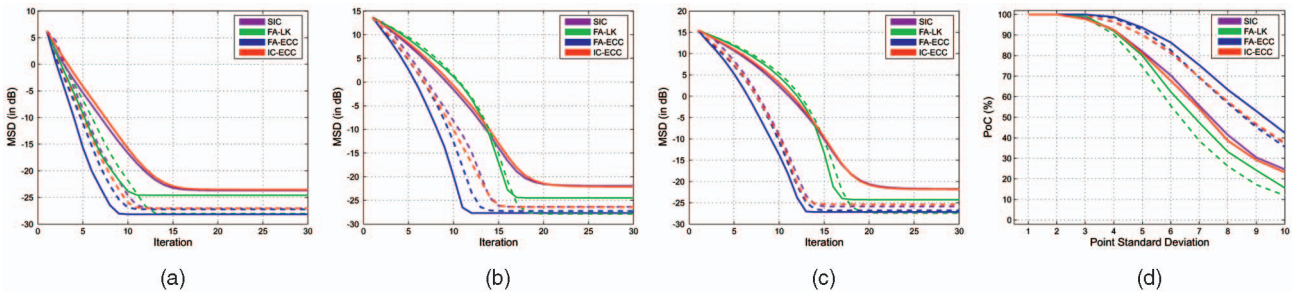

Fig. 2. MSD in decibels as a function of number of iterations for photometrically distorted reference (solid lines) and warped (dashed lines) image under the presence of noise ($\sigma_i = 8$ gray levels). (a) $\sigma_p = 2$. (b) $\sigma_p = 6$. (c) $\sigma_p = 10$. In (d), PoC as a function of σ_p.

Finally, we present the corresponding curves of the third figure of merit in Fig. 3b under the first scenario since, under the second one, both the inverse and the FA-ECC algorithm exhibited a similar performance. As in the previous experiment, we permit a maximal number of 100 iterations with a threshold $T_{\text{MSD}} = (1/10 \text{ pixel})^2$ (i.e., -20 dB), since now we have higher MSD floor values. Again, FA-ECC outperforms the other algorithms. Comparing Fig. 3a with Fig. 3b, we can also notice a robust and consistent behavior of FA-ECC with respect to intensity noise and photometric distortion model mismatch.

In summary, we can safely conclude that our proposed schemes are preferable to the corresponding variants of the LK algorithm. Clearly, our forward version is more effective than the forward LK scheme regarding both speed and percentage of convergence. On the other hand, our inverse version has performance which is comparable to the performance of SIC, which is the best inverse version of the LK algorithm. However, the point that makes our IC-ECC version preferable to SIC is the reduced computational complexity, which is $O(KN)$ as compared to SIC, which requires $O(K(N+2)^2)$ operations.

We should also mention that we evaluated the algorithms under diverse uncertainty conditions. Only in the case of zero intensity noise (in other words, when the warped image follows the warping model exactly), we observed the performance of both inverse algorithms and the FA-ECC to be similar and to outperform the FA-LK algorithm in all figures of merit. This performance difference can, in fact, become quite significant if the geometric deformations are strong (i.e., $\sigma_p \geq 6$). However, due to lack of space, we cannot present these results in more detail.

5 Conclusions

In this paper, we have proposed a new l_2-based iterative algorithm tailored to the parametric image alignment problem. The new scheme is aimed at maximizing the Enhanced Correlation Coefficient function, which constitutes a measure that is robust against geometric and photometric distortions. The optimal parameters were obtained by iteratively solving a sequence of approximate nonlinear optimization problems which enjoy a simple closed-form solution with low computational cost. In addition, based on the inverse compositional update rule, we developed an efficient modification of the forward algorithm. Our iterative schemes were compared against two variants of the LK algorithm through numerous simulations. Under ideal conditions, the proposed algorithms and the SIC algorithm exhibited similar performance, outperforming the forward LK algorithm. However, in the more realistic case of noisy conditions and photometric distortions, our forward algorithm exhibited a noticeably superior performance in convergence speed, accuracy, and percentage of convergence.

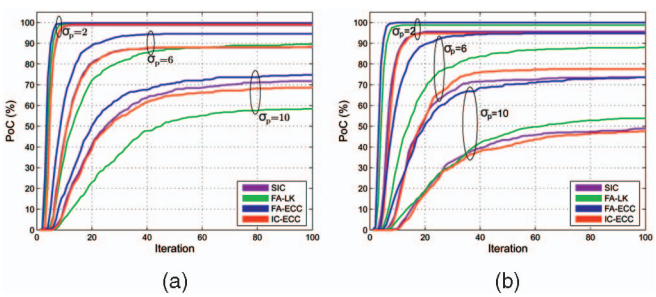

Fig. 3. PoC as a function of iterations: (a) noisy images ($\sigma_i = 8$ gray levels) and (b) noisy ($\sigma_i = 8$ gray levels) and photometrically distorted images.

Appendix

Proof of Theorem 1

The proof of Theorem 1 relies on the application of Schwartz inequality. In order to simplify our presentation, let us impose the following change of variables:

$$\mathbf{z} = Q^{1/2}\mathbf{x} + Q^{-1/2}\mathbf{v};\ \tilde{\mathbf{u}} = Q^{-1/2}\mathbf{u};\ \tilde{\mathbf{v}} = Q^{-1/2}\mathbf{v}; \tag{31}$$

then the function we want to optimize becomes a function of $\mathbf{z}$ and has the form

$$f(\mathbf{z}) = \frac{(u - \tilde{\mathbf{u}}^t\tilde{\mathbf{v}}) + \tilde{\mathbf{u}}^t\mathbf{z}}{\sqrt{\left(v - \|\tilde{\mathbf{v}}\|^2\right) + \|\mathbf{z}\|^2}}. \tag{32}$$

Note that condition $v > \|\tilde{\mathbf{v}}\|^2$ guarantees that the quantity under the square root in the denominator is positive.

Let us first consider the case $u - \tilde{\mathbf{u}}^t\tilde{\mathbf{v}} > 0$, then we can define

$$\tilde{\mathbf{z}} = \left[\mathbf{z}^t\sqrt{v - \|\tilde{\mathbf{v}}\|^2}\right]^t;\quad \tilde{\mathbf{w}} = \left[\tilde{\mathbf{u}}^t\frac{u - \tilde{\mathbf{u}}^t\tilde{\mathbf{v}}}{\sqrt{v - \|\tilde{\mathbf{v}}\|^2}}\right]^t \tag{33}$$

and our objective function becomes

$$f(\mathbf{z}) = \frac{\tilde{\mathbf{w}}^t\tilde{\mathbf{z}}}{\|\tilde{\mathbf{z}}\|} \leq \frac{|\tilde{\mathbf{w}}^t\tilde{\mathbf{z}}|}{\|\tilde{\mathbf{z}}\|} \leq \|\tilde{\mathbf{w}}\|, \tag{34}$$

with the last inequality being the result of applying the Schwartz inequality. Now notice that $\|\tilde{\mathbf{w}}\|$ is constant, constituting an upper bound to our objective function. This bound is attainable when both inequalities become equalities. From the Schwartz inequality, we know that we have equality whenever we select $\tilde{\mathbf{z}} = \lambda\tilde{\mathbf{w}}$, where λ is some scalar quantity. Under this selection, in order for the first inequality to become equality, we need $\lambda > 0$. From $\tilde{\mathbf{z}} = \lambda\tilde{\mathbf{w}}$, by equating the last vector elements, we conclude that $\lambda = (v - \|\tilde{\mathbf{v}}\|^2)/(u - \tilde{\mathbf{u}}^t\tilde{\mathbf{v}})$, which is positive only when $u - \tilde{\mathbf{u}}^t\tilde{\mathbf{v}} > 0$, yielding $\mathbf{z} = (v - \|\tilde{\mathbf{v}}\|^2)\tilde{\mathbf{u}}/(u - \tilde{\mathbf{u}}^t\tilde{\mathbf{v}})$. It is

interesting to note that, when $u - \tilde{\mathbf{u}}^t\tilde{\mathbf{v}} \leq 0$, the upper bound $\|\tilde{\mathbf{w}}\|$ *is not tight* (not attainable) and, therefore, this case needs the separate treatment that follows.

When $u - \tilde{\mathbf{u}}^t\tilde{\mathbf{v}} \leq 0$, in order to find the supremum, we apply the following inequalities:

$$\begin{aligned} f(\mathbf{z}) &= \frac{(u - \tilde{\mathbf{u}}^t\tilde{\mathbf{v}}) + \tilde{\mathbf{u}}^t\mathbf{z}}{\sqrt{\left(v - \|\tilde{\mathbf{v}}\|^2\right) + \|\mathbf{z}\|^2}} \\ &\leq \frac{\tilde{\mathbf{u}}^t\mathbf{z}}{\sqrt{\left(v - \|\tilde{\mathbf{v}}\|^2\right) + \|\mathbf{z}\|^2}} \\ &\leq \|\tilde{\mathbf{u}}\| \frac{\|\mathbf{z}\|}{\sqrt{\left(v - \|\tilde{\mathbf{v}}\|^2\right) + \|\mathbf{z}\|^2}} \\ &\leq \|\tilde{\mathbf{u}}\|. \end{aligned} \tag{35}$$

The first inequality is true because of the nonpositivity of $u - \tilde{\mathbf{u}}^t\tilde{\mathbf{v}}$ (from our assumption); for the second, we applied the Schwartz inequality in the numerator; finally, for the last, we used the fact that the ratio is smaller than 1. We observe that, in this case, we end up with a different (smaller) upper bound. In order to verify its tightness (i.e., whether it constitutes a supremum), we use the selection prescribed by the Schwartz inequality, that is, $\mathbf{z} = \lambda\tilde{\mathbf{u}}$ again with $\lambda > 0$ and compute the corresponding value of the objective function. By letting $\lambda \to \infty$, we realize that we converge to $\|\tilde{\mathbf{u}}\|$. This suggests that, for sufficiently large λ, we can approach the desired upper bound arbitrarily close (but there is no finite $\mathbf{z}$ for which we can attain it exactly!). This concludes the proof. □

ACKNOWLEDGMENTS

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

REFERENCES

[1] S. Periaswamy and H. Farid, "Elastic Registration in the Presence of Intensity Variation," *IEEE Trans. Medical Imaging*, vol. 22, no. 7, pp. 865-874, 2003.
[2] I. Karybali, E.Z. Psarakis, K. Berberidis, and G.D. Evangelidis, "Efficient Image Registration with Subpixel Accuracy," *Proc. 14th European Signal Processing Conf.*, 2006.
[3] G.D. Hager and P.N. Belhumeur, "Efficient Region Tracking with Parametric Models of Geometry and Illumination," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 20, no. 10, pp. 1025-1039, Oct. 1998.
[4] J. Shi and C. Tomasi, "Good Features to Track," *Proc. IEEE Int'l Conf. Computer Vision and Pattern Recognition*, 1994.
[5] M. Gleicher, "Projective Registration with Difference Decomposition," *Proc. IEEE Int'l Conf. Computer Vision and Pattern Recognition*, 1997.
[6] C. Fuh and P. Maragos, "Motion Displacement Estimation Using an Affine Model for Image Matching," *Optical Eng.*, vol. 30, no. 7, pp. 881-887, 1991.
[7] Y. Altunbasak, R.M. Mersereau, and A.J. Patti, "A Fast Parametric Motion Estimation Algorithm with Illumination and Lens Distortion Correction," *IEEE Trans. Image Processing*, vol. 12, no. 4, pp. 395-408, 2003.
[8] B.K.P. Horn and E.J. Weldon, "Direct Methods for Recovering Motion," *Int'l J. Computer Vision*, vol. 2, no. 1, pp. 51-76, 1988.
[9] P. Anandan, "A Computational Framework and an Algorithm for the Measurement of Visual Motion," *Int'l J. Computer Vision*, vol. 2, no. 3, pp. 283-310, 1989.
[10] B.D. Lucas and T. Kanade, "An Iterative Image Registration Technique with an Application to Stereo Vision," *Proc. Seventh Int'l Joint Conf. Artificial Intelligence*, 1981.
[11] E.Z. Psarakis and G.D. Evangelidis, "An Enhanced Correlation-Based Method for Stereo Correspondence with Sub-Pixel Accuracy," *Proc. 10th IEEE Int'l Conf. Computer Vision*, 2005.
[12] R. Szeliski, *Handbook of Mathematical Models of Computer Vision*, N. Paragios, Y. Chen, and O. Faugeras, eds., chapter 17. Springer, 2005.
[13] 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," *Int'l J. Computer Vision*, vol. 56, no. 3, pp. 221-255, 2004.
[14] M.J. Black and Y. Yacoob, "Tracking and Recognizing Rigid and Non-Rigid Facial Motions Using Local Parametric Models of Image Motion," *Proc. Fifth IEEE Int'l Conf. Computer Vision*, 1995.
[15] H. Shum and R. Szeliski, "Construction of Panoramic Image Mosaics with Global and Local Alignment," *Int'l J. Computer Vision*, vol. 36, no. 2, pp. 101-130, 2000.
[16] S. Nagahdaripour and C.H. Yu, "A Generalized Brightness Change Model for Computing Optical Flow," *Proc. Fourth IEEE Int'l Conf. Computer Vision*, 1993.
[17] B.K.P. Horn and B.G. Schunk, "Determining Optical Flow," *Artificial Intelligence*, vol. 17, pp. 185-203, 1981.
[18] M.J. Black and P. Anandan, "A Framework for the Robust Estimation of Optical Flow," *Proc. Fourth IEEE Int'l Conf. Computer Vision*, 1993.
[19] S. Baker, R. Gross, and I. Matthews, "Lucas-Kanade 20 Years On: A Unifying Framework: Part 3," CMU-RI-TR-03-35, Robotics Inst., Carnegie Mellon Univ., 2004.
[20] B.K.P. Horn, *Robot Vision*. MIT Press, McGraw-Hill, 1986.
[21] P. Hallinan, "A Low-Dimensional Representation of Human Faces for Arbitrary Lighting Conditions," *Proc. IEEE Int'l Conf. Computer Vision and Pattern Recognition*, 1994.
[22] V. Barnett and T. Lewis, *Outliers in Statistical Data*. John Wiley & Sons, 1978.