Pattern Recognition Letters

Authorship Confirmation

Please save a copy of this file, complete and upload as the "Confirmation of Authorship" file.

As corresponding author I, Emmanouil Z. Psarakis, hereby confirm on behalf of all authors that:

1. This manuscript, or a large part of it, has not been published, was not, and is not being submitted to any other journal.
2. If presented at or submitted to or published at a conference(s), the conference(s) is (are) identified and substantial justification for re-publication is presented below. A copy of conference paper(s) is(are) uploaded with the manuscript.
3. If the manuscript appears as a preprint anywhere on the web, e.g. arXiv, etc., it is identified below. The preprint should include a statement that the paper is under consideration at Pattern Recognition Letters.
4. All text and graphics, except for those marked with sources, are original works of the authors, and all necessary permissions for publication were secured prior to submission of the manuscript.
5. All authors each made a significant contribution to the research reported and have read and approved the submitted manuscript.

Signature E. Z. Psarakis Date 1-10-2016

List any pre-prints:

Relevant Conference publication(s) (submitted, accepted, or published):

Justification for re-publication:

Graphical Abstract (Optional)

A New LS Based Congealing Technique

N. Nikolikos, E. Z. Psarakis and N. Lamprinou

ELSEVIER

The aim of this work is to improve upon the state-of-the-art pixel-level, Least-Squares (LS) based congealing methods. Specifically, we propose a new iterative algorithm, which outperforms in terms of speed, convergence rate and robustness the state-of-the-art inverse compositional LS based iterative scheme. Namely, by associating the geometric distortion of each image of the ensemble with the position of a particle of a multi particle system, we succeed to align the ensemble without having to align all the individual pairs resulting from it. Instead we align each image with the "mean", but unknown, image. To this end, by imposing the"centre of mass" of the particle system to be motionless during each iteration of the minimization process, a sequence of "centroid" images whose limit is the unknown "mean" image is defined, thus solving the congealing problem. The proposed congealing technique is invariant to the size of the image set and depends only on the image size, thus it can be used for the successful solution of the congealing problem on large image sets with low complexity.

Research Highlights (Required)

- A new fast and robust iterative LS congealing algorithm with high convergency rate
- Each image of the ensemble is aligned with the “mean”, but unknown, image
- The mean image is defined by imposing the“centre of system mass” to be motionless in each iteration of the minimization process.
- The method is invariant to the size of the image set thereby ensuring low complexity
- The proposed technique can be easily adapted to cope with feature based techniques

Pattern Recognition Letters
journal homepage: www.elsevier.com

A New LS Based Congealing Technique

Nikolaos Nikolikos[a], Emmanouil Z. Psarakis[a,**], Nefeli Lamprinou[a]

[a]*Department of Computer Engineering & Informatics, University of Patras, 26500 Patras, Greece*

ABSTRACT

The aim of this work is to improve upon the state-of-the-art pixel-level, Least-Squares (LS) based congealing methods. Specifically, we propose a new iterative algorithm, which outperforms in terms of speed, convergence rate and robustness the state-of-the-art inverse compositional LS based iterative scheme. Namely, by associating the geometric distortion of each image of the ensemble with the position of a particle of a multi particle system, we succeed to align the ensemble without having we succeed to align the ensemble without having to align all the individual pairs resulting from it. Instead we align each image with the "mean", but unknown, image. To this end, by imposing the"centre of mass" of the particle system to be motionless during each iteration of the minimization process, a sequence of "centroid" images whose limit is the unknown "mean" image is defined, thus solving the congealing problem. The proposed congealing technique is invariant to the size of the image set and depends only on the image size, thus it can be used for the successful solution of the congealing problem on large image sets with low complexity.

1. Introduction

The problem of image congealing (group-wise alignment/registration) is an important one within the computer vision community. A good congealing algorithm can be used as preprocessing to notably improve the performance of other vision tasks within different research areas. In recent literature, the existing image congealing techniques can be broadly classified into the following two categories Lankinen and Kamarainen (2011); Shokrollahi Yancheshmeh et al. (2015):

- Intensity/Pixel-level optimization and
- Visual-Object-Categoriation (VOC).

The proposed work improves upon the most widely recognised pixel-level LS based state-of-the-art approach, i.e. the inverse LS based algorithm with compositional warping. Thus, we will now describe the most relevant associated methods that fall under this term. All state-of-the-art algorithms can be considered as variations of a common base framework. The basic idea is to use one image at a time as the held out image and the rest of the ensemble as the stack. Having done that the goal is to minimize, in an iterative fashion, an error function defined over all of the ensemble, by estimating a warp update for the held out image that aligns it with the stack. Algorithms based on the aforementioned idea include congealing methods with entropy based cost functions, such as the algorithms proposed in Learned-Miller (2006); Zollei (2006); Vedaldi and Soatto (2006) as well as with LS based cost functions such as the methods proposed in Cox et al. (2008, 2009); Cox (2010); Xue and Liu (2012).

In Storer and Urschler (2010) an error function based on mutual information that copes with possible variations in appearance beetween similar objects of the same class is defined. In addition, in Huang et al. (2007) and Huang et al. (2012) extended entropy based congealing for the usage on real world complex images is proposed. Such a framework can be incorporated within the rest of these techniques in order to deal with background variations. LS based congealing algorithms tend to perform better in terms of convergence rate and accuracy. In the LS case, there are two ways to align the held out image with the stack using gradient descend optimization techniques. The first one, which is known as the forward LS congealing approach, is to align the held out image with the mean image of the stack. This approach has poor alignment performance, especially for strong initial misalignments, but has a really low computational cost. The second one, which is known as the inverse LS congealing approach, computes a common warp update for all im-

**Corresponding author: Tel.: +30-261-099-6969; fax: +30-261-099-6969;
e-mail: psarakis@ceid.upatras.gr (Emmanouil Z. Psarakis)

ages of the stack by utilising the inverse approach presented in Cox et al. (2008), i.e. to align each image of the stack with the current held out image. This approach outperforms forward LS in both accuracy and robustibility, but has a high computational cost due to existing nested loops. This means that its cost becomes prohibitive for large image sets as the number of subproblems grows quadradically with respect to the image set's size. Another drawback lies in the additional robustification needed for its error function and warp computations in order to be able to handle outliers Cox (2010). This stems from the fact that the initial hypothesis behind the minimization strategy of the overall error function, which is the accumulation of all error functions per held out image, is flawed.

The proposed congealing method improves upon all the desirable characteristics of the state-of-the-art inverse method, while maintaining a linear to the set size computational cost, similar to that of the forward approach, plus the cost of the singular value decomposition on the centroid of the pseudo inverse of the Jacobian matrices over the entire ensemble of the images. Having mentioned the above, we still apply our warp updates compositionaly and not additively, since as shown in Cox et al. (2008) the compositional warping is more robust for large image sets.

The remainder of this paper is organized as follows: In Section 2, we formulate the image congealing problem and several issues related with it are examined. In Section 3 a particle system strongly related to the geometric transformtions of the image's set is introduced and the proposed solution is presented. In Section 4 the results of the experiments we have conducted are presented. Finally, Section 5 contains our conclusions.

2. Problem Formulation

2.1. Preliminaries

Let us consider a set containing N images:

$$\mathbb{S}_{\mathbf{i}} = \{\mathbf{i}_n\}_{n=1}^{N} \tag{1}$$

that belong to the same cluster, that is $\mathbb{S}_{\mathbf{i}}$ contains a group of similar in shape and aligned images, where $\mathbf{i}$ denotes the column-wise of length N_xN_y vectorized version of size $N_x \times N_y$ image I. It is well known that the "mean" image which is defined by:

$$\bar{\mathbf{i}}^{\star} = \frac{1}{N}\sum_{n=1}^{N}\mathbf{i}_n \tag{2}$$

constitutes the most representative image for the cluster and it can result from the solution of the following optimization problem:

$$\bar{\mathbf{i}}^{\star} = \arg\min_{\bar{\mathbf{i}} \in \mathbb{R}^{N_xN_y}} \{\sum_{n=1}^{N} \|\bar{\mathbf{i}} - \mathbf{i}_n\|_2^2\} \tag{3}$$

where $\|\mathbf{x}\|_2$ denotes the l_2 norm of vector $\mathbf{x}$.

Let us now consider, apart from the set $\mathbb{S}_{\mathbf{i}}$, the following set:

$$\mathbb{S}_{\mathbf{i}_w}(\mathbb{P}) = \{\mathbf{i}_w(\mathbf{p}_n)\}_{n=1}^{N} \tag{4}$$

containing the geometrically distorted vectorized images of set $\mathbb{S}_{\mathbf{i}}$ of (1) where

$$\mathbb{P} = \{\mathbf{p}_n\}_{n=1}^{N}, \tag{5}$$

is a set of N warp parameter vectors. Under the used warping transformation $w(\,.\,;\mathbf{p}_n)$[1], which is parameterized by the vector $\mathbf{p}_n \in \mathbb{R}^M$, each pixel $\mathbf{x}$ of the Region of Interest of image $\mathbf{i}_n$ of set (1) is mapped onto the pixel $\hat{\mathbf{x}}$ of the corresponding image $\mathbf{i}_w(\mathbf{p}_n)$ of set (4), i.e.:

$$I_w(\hat{\mathbf{x}};\ \mathbf{p}_n) = I_n(w(\mathbf{x};\ \mathbf{p}_n)). \tag{6}$$

Then, congealing can be defined as the minimization problem of a misalignment function, let us denote it by $\mathcal{E}(\mathbb{P})$, which is calculated over the set $\mathbb{S}_{\mathbf{i}_w}(\mathbb{P})$ of geometrically distorded images, for a warping function that models the parametric form of the misalignment to be removed. In general, solving the image congealing problem is not an easy task and its complexity heavily depends on several factors, such as the size of the ensemble and the strongness of the geometric distrortions, to name a few. However, in some cases the aforementioned problem can be easily solved. Such two characteristic cases follow:

1. *Image set $\mathbb{S}_{\mathbf{i}}$ defined in* (1) *is known*
   In this case, we can easily "align" the image sets by solving the following N optimization problems:

$$\mathbf{p}_n^{\star} = \arg\min_{\mathbf{p}_n \in \mathbb{R}^M} \|\mathbf{i}_n - \mathbf{i}_w(\mathbf{p}_n)\|_2^2,\ n = 1,\ 2,\ \cdots,\ N. \tag{7}$$

2. *Image set $\mathbb{S}_{\mathbf{i}}$ is unknown but the "mean" image $\bar{\mathbf{i}}^{\star}$ is known*
   We still can approximatelly solve the problem if we consider that $\bar{\mathbf{i}}^{\star}$ defined by (2) is the rank-one Singular Value Decomposition (SVD) of matrix:

$$S = [\mathbf{i}_1\ \mathbf{i}_2\ \cdots\ \mathbf{i}_N]$$

   whose each column is a member of the ensemble (1), by solving the following N optimization problems:

$$\mathbf{p}_n^{\star} = \arg\min_{\mathbf{p}_n \in \mathbb{R}^M} \|\bar{\mathbf{i}}^{\star} - \mathbf{i}_w(\mathbf{p}_n^{'})\|_2^2,\ n = 1,\ 2,\ \cdots,\ N. \tag{8}$$

Note that both the objective functions involved in the optimization problems (7) and (8) are nonlinear with respect to the parameter vector $\mathbf{p}_n$. This, of course, suggests that their minimization requires nonlinear optimization techniques either by using direct search or by following gradient-based approaches. As is customary in iterative techniques, the original optimization problem is replaced by 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. Assuming that at the k-th iteration of the iterative procedure $\mathbf{p}_n(k)$ is "close" to some nominal parameter vector $\tilde{\mathbf{p}}_n$, then we write $\mathbf{p}_n(k) = \tilde{\mathbf{p}}_n + \Delta\mathbf{p}_n(k)$, where $\Delta\mathbf{p}_n(k)$ denotes a vector of perturbations.

Let $w(\mathbf{x};\ \tilde{\mathbf{p}}_n)$ be the warped coordinates under the nominal parameter vector and $w(\mathbf{x};\ \mathbf{p}_n(k))$ under the perturbed one. Considering the intensity of the warped image at coordinates under

[1] In this paper, to model the warping process we are going to use the class of affine transformations with $\mathbf{p}_n \in \mathbb{R}^6$.

the nominal parameter vector and applying a first-order Taylor expansion with respect to the parameters, we can write:

$$\mathbf{i}_w(\mathbf{p}_n(k)) \approx \mathbf{i}_w(\tilde{\mathbf{p}}_n) + G_w(\tilde{\mathbf{p}}_n)\Delta\mathbf{p}_n(k) \tag{9}$$

where $G_w(\tilde{\mathbf{p}}_n)$ denotes the size $N_xN_y \times M$ Jacobian matrix of the warped intensity vector with respect to the parameters, evaluated at the nominal parameter values $\tilde{\mathbf{p}}_n$. Then, it is well known that the optimum vector of perturbations is defined by the following relation Baker and Matthews (2004):

$$\Delta\mathbf{p}_n(k) = A_w(\tilde{\mathbf{p}}_n)(\bar{\mathbf{i}}^\star - \mathbf{i}_w(\tilde{\mathbf{p}}_n)), \tag{10}$$

where:

$$A_w(\tilde{\mathbf{p}}_n) = (G_w(\tilde{\mathbf{p}}_n)^T G_w(\tilde{\mathbf{p}}_n))^{-1} G_w(\tilde{\mathbf{p}}_n)^T, \tag{11}$$

is the $M \times N_xN_y$ pseudo inverse of the Jacobian matrix $G_w(\tilde{\mathbf{p}}_n)$. Note that for the solution of the optimization problem by using an iterative procedure the "mean" image $\bar{\mathbf{i}}^\star$ as well as the nominal parameters $\tilde{\mathbf{p}}_n,\ n = 1,\ 2, \cdots,\ N$ are quantities that must be known. Note also that since $N_xN_y \gg M$, the column rank of the pseudo inverse of the matrix $G_w(\tilde{\mathbf{p}}_n)$ is upper bounded by M. Thus, for the definition of the optimum perturbations according to (10), the projection of the error image $\bar{\mathbf{i}}^\star - \mathbf{i}_w(\tilde{\mathbf{p}}_n)$ onto a subspace of $\mathbb{R}^{N_xN_y}$, of dimension at maximum M is needed. We are going to exploit this point in Section 3 in order to define a sequence of images whose limit will be the "mean" image.

Let us consider now that we would like to solve the above mentioned problem, but even the "mean" image is unknown. In that case, irrespectively of the choise of the misalignment function type, its minimization results in a highly nonlinear and computationally demanding procedure. This is basically because the goal of estimating the collection of the unknown parameters should be achieved by defining a misalignment function $\mathcal{E}(\mathbb{P})$ over the entire ensemble of images. Such a function, which is known as the Cummulative Squared Misalignment Error (CSME):

$$\mathcal{E}(\mathbb{P}) = \sum_{n=1}^{N} \epsilon(\mathbf{p}_n) \tag{12}$$

where

$$\epsilon(\mathbf{p}_n) = \sum_{m=1, m\neq n}^{N} \|\mathbf{i}_w(\mathbf{p}_n) - \mathbf{i}_w(\mathbf{p}_m)\|_2^2, \tag{13}$$

has all these characteristics and was proposed in Cox et al. (2008). However, since the above total cost function is difficult to be optimized directly Tong et al. (2009), the iterative minimization of the individual cost function $\epsilon(\mathbf{p}_n)$ for each geometrically distorted image $\mathbf{i}_w(\mathbf{p}_n)$, given an initial estimation of the warping parameter $\mathbf{p}_m,\ m = 1, 2, \cdots,\ N$, was proposed.

In general, aligning each image of the aforementioned set with a stack of images, results in a problem which is heavily depented on the initial conditions of the average image. This is obvious since independently of the blurriness of the basis images, the average image before the alignment of images is blured. Thus, using the average image to control the direction of the corrections of the parameters, with all the fine details of the image lost, the algorithm is at the mercy of the initial conditions Cox et al. (2008). In order to avoid this undesired effect, LS congealing avoids using the mean image and for improving the quality of the alignment an inverse compositional LS congealing algorithm was proposed.

However, in the next section we are going to explore ways of obtaining the alignment of the ensemble in (4) without having to align all the individual pairs resulting from it. Instead we are going to align each image with the "mean" image, which we are going to relate with the"centre of mass" of the geometric transformations set (5).

3. The proposed Solution

3.1. The particle System

To this end, let us consider that each member of set $\mathbb{S}_{\mathbf{i}_w}(\mathbb{P})$ denotes the starting position of a particle of mass $m = 1/N$ that can be moved into $\mathbb{R}^M$. Hence the set $\mathbb{S}_{\mathbf{i}_w}(\mathbb{P})$ could be considered as a set containing the starting positions of a system of N isobaric particles. Let $\mathbf{p}_n(k)$ be the position of the n-th patricle at the k-th instant as it moves into $\mathbb{R}^M$. Then, with the following position's set:

$$\mathbb{T}_n = \{\mathbf{p}_n(k),\ k = 0,\ 1,\ 2, \cdots\}, \tag{14}$$

we can define its trajectory, with its first and last element $\mathbf{p}_n(0)$, $\lim_{k\to\infty}\mathbf{p}_n(k)$ denoting its starting and ending position respectively.

As it is clear the successive positions of the n-th particle, or equivalently its motion, can be easily expressed by the following simple motion model:

$$\mathbf{p}_n(k) = \mathbf{p}_n(k-1) + \Delta\mathbf{p}_n(k),\ k = 1,\ 2,\ \cdots \tag{15}$$

with the vector $\Delta\mathbf{p}_n(k)$ denoting its differential movement. Note also that $\Delta\mathbf{p}_n(k)$ is proportional to the momentum $m\Delta\mathbf{p}_n(k)$ of the particle. The position of the centre of mass of the whole particle system, at the k-th instant, can be defined as follows:

$$\bar{\mathbf{p}}(k) = \frac{1}{N}\sum_{n=1}^{N}\mathbf{p}_n(k). \tag{16}$$

Having defined the centre of mass of the particle system, we can use it as the origin of the moving coordinate system. Specifically, the position of the n-th particle at the $k-$th instant with respect to the new origin $\bar{\mathbf{p}}(k)$ can be expressed as follows:

$$\mathbf{p}_n(k) = \bar{\mathbf{p}}(k) + \mathbf{q}_n(k),\ k = 1,\ 2,\ \cdots. \tag{17}$$

Then, we can easily see that the vectors $\mathbf{q}_n(k),\ n = 1,\ 2, \cdots,\ N$ denote the positions of the particles in the new coordinate system and are zero mean, that is:

$$\sum_{n=1}^{N}\mathbf{q}_n(k) = 0. \tag{18}$$

As time goes on, the position vectors of the particles change with time and thus, in general, the centre of mass moves too with its velocity given by:

$$\Delta\bar{\mathbf{p}}(k) = \bar{\mathbf{p}}(k) - \bar{\mathbf{p}}(k-1) = \frac{1}{N}\sum_{n=1}^{N}\Delta\mathbf{p}_n(k) \tag{19}$$

where $\bar{\mathbf{p}}(k),\ \bar{\mathbf{p}}(k-1),\ k=1,\ 2,\ \cdots$ are the centre of mass of the $N-$ sized particle system at two consecutive instances.

Our goal now is to express the motion equation in (15) into the new centre of mass based coordinate system and find out under what conditions the velocity, and equivalently the momentum, of each particle remains the same in both the aforementioned coordinate systems. The answer is given by the next proposition.

Proposition 3.1. *Let us consider a system of N isobaric particles of mass $m=\frac{1}{N}$ each. Then, the velocity of each particle remains the same in both the aforementioned coordinate systems, if its centre of mass is motionless or equivalently, its total momentum is zero, that is:*

$$\Delta\bar{\mathbf{p}}(k)=0. \tag{20}$$

Proof The proof is easy and is left to the reader. □

We must stress at this point that in terms of the image alignment problem we could safely say that the meaning of the condition expressed by (20), i.e:

$$\bar{\mathbf{p}}=\bar{\mathbf{p}}(k),\ k=0,\ 1,\cdots \tag{21}$$

is that the mean geometric transformation remains the same over the iteration process and totally specifies the geometric deformation of the "mean" image.
Concluding, instead of concentrating on the individual motion of each particle of the system we could concentrate on the motion of its centre of mass. Note that the mean image is unknown but all the warped ones that we would like to align with. Let us discriminate the following cases.

1. *The Ideal Case*: Let us consider that the set $\mathbb{S}_{\mathbf{i}_w}(\mathbb{P})$ is generated by the following process:

$$\bar{\mathbf{i}}^{\star}\xrightarrow{w(.;\bar{\mathbf{p}})}\mathbf{i}_w(\bar{\mathbf{p}})\xrightarrow{\{w(.;\mathbf{p}_n)\}_{n=1}^{N}}\{\mathbf{i}_w(\mathbf{p}_n)\}_{n=1}^{N}\equiv\mathbb{S}_{\mathbf{i}_w}(\mathbb{P}). \tag{22}$$

   Note that, in this case all the images of the set $\mathbb{S}_{\mathbf{i}}$ are identical and concequently each of them coincides with the mean image. Then, since each member of the ensemble of images in (4) is a geometrically distorded image of the warped by the mean of the geometric transformations $\bar{\mathbf{p}}$, "mean" image defined in (2), the following relation holds:

$$\lim_{k\to\infty}\mathbf{i}_w(\mathbf{p}_n(k))=\mathbf{i}_w(\bar{\mathbf{p}}). \tag{23}$$

   Hence, in the ideal case all the geometric transformations, or equivalently all the trajectories of the particles as they described by (17), converge to $\bar{\mathbf{p}}$, i.e:

$$\lim_{k\to\infty}\mathbf{p}_n(k)=\bar{\mathbf{p}}+\lim_{k\to\infty}\mathbf{q}_n(k)=\bar{\mathbf{p}} \tag{24}$$

   since $\lim_{k\to\infty}\mathbf{q}_n(k)=0,\ n=1,2,\cdots,N$. Finally, note that if $\bar{\mathbf{p}}$ is the identity transformation of the warping function $w(.;\mathbf{p})$, i.e.: $w(\mathbf{x};\ \bar{\mathbf{p}})=\mathbf{x}$ then, $\mathbf{i}_w(\bar{\mathbf{p}})\equiv\bar{\mathbf{i}}^{\star}$. For the sake of simplicity, from now on we consider that the parameter vector $\bar{\mathbf{p}}$ is the identity transformation of the warping function.

2. *The General Case*: In the general case the set $\mathbb{S}_{\mathbf{i}_w}(\mathbb{P})$ is generated by the following process:

$$\left.\begin{array}{ccc}\mathbf{i}_1 & \xrightarrow{w(.;\mathbf{p}_1)} & \mathbf{i}_w(\mathbf{p}_1)\\ \mathbf{i}_2 & \xrightarrow{w(.;\mathbf{p}_2)} & \mathbf{i}_w(\mathbf{p}_2)\\ \vdots & \vdots & \vdots\\ \mathbf{i}_N & \xrightarrow{w(.;\mathbf{p}_N)} & \mathbf{i}_w(\mathbf{p}_N)\end{array}\right\}\equiv\mathbb{S}_{\mathbf{i}_w}(\mathbb{P}) \tag{25}$$

   where, as it was already mentioned in Subsection 2.1, all images belong to the same cluster, are aligned and the "mean" image is defined by (2). In this case, even if (24) holds, we are expecting that relation (23) will hold approximatelly. This is because according to the definition (6), $\lim_{k\to\infty}\mathbf{i}_w(\mathbf{p}_n(k))=\mathbf{i}_n,\ n=1,2,\cdots,N$ and the average of these N images is the desired "mean" image $\bar{\mathbf{i}}^{\star}$.

3.2. The Proposed Misalignment Function

Let us consider the following total mean misalignment function:

$$\mathcal{E}_0(\mathbb{P})=\frac{1}{N}\sum_{n=1}^{N}\|\bar{\mathbf{i}}^{\star}-\mathbf{i}_w(\mathbf{p}_n)\|_2^2 \tag{26}$$

where $\bar{\mathbf{i}}^{\star}$ denotes the unknown "mean" image. As it is clear, the above defined misalignment function is separable, but is a non linear function of the warp parameters $\mathbf{p}_n,\ n=1,2,\cdots,N$. Thus, for each one of the cost functions involved into (26) the comments of Subsection 2.2 regarding their minimization, hold. We recall here that in order to be able at the $k-$th iteration of the minimization process to compute, using (10), the optimal perturbations $\Delta\mathbf{p}_n(k)$, the nominal parameter vector $\tilde{\mathbf{p}}_n$ as well as the "mean" image $\bar{\mathbf{i}}^{\star}$ must be known. By assigning to the nominal parameter vector $\tilde{\mathbf{p}}_n$ the position of the corresponding particle in the previous iteration, i.e.: $\mathbf{p}_n(k-1)\rightarrow\tilde{\mathbf{p}}_n$, (10) can be rewritten as follows:

$$\Delta\mathbf{p}_n(k)=A_w(\mathbf{p}_n(k-1))(\bar{\mathbf{i}}^{\star}-\mathbf{i}_w(\mathbf{p}_n(k-1))). \tag{27}$$

Note however, that since the "mean" image $\bar{\mathbf{i}}^{\star}$ is unknown the optimal values of the perturbations in (27) can not be computed. Moreover, the use of the following average of the warped vectorized images:

$$\bar{\mathbf{i}}_w(k)=\frac{1}{N}\sum_{n=1}^{N}\mathbf{i}_w(\mathbf{p}_n(k-1)) \tag{28}$$

leading to a similar approach with the forward one, will have a poor alignment performance Cox (2010). In order to overcome this obstacle, we are going to define in the next paragraph, the "centroid" vectorized image.

3.3. The "Centroid" Image

Let us define a sequence of vectorized images:

$$\mathbb{I}_c=\{\mathbf{i}(k)\}_{k=1}^{\infty}, \tag{29}$$

having the following two properties:

- $\mathcal{P}_1$: The $k-$th member of the sequence, is the "centroid" image of the corresponding iteration of the minimization process
- $\mathcal{P}_2$: The limit of the sequence is the unknown "mean" image, that is:

$$\lim_{k\to\infty} \mathbf{i}(k) = \bar{\mathbf{i}}^{\star}. \tag{30}$$

In order to achieve our goal in each iteration of the process, we decompose the "centroid" image into two components, i.e.:

$$\mathbf{i}(k) = \mathbf{i}_1(k) + \mathbf{i}_2(k) \tag{31}$$

with each one being associated with an appropriate subspace of the space $\mathbb{R}^{N_xN_y}$ of the vectorized images. To this end, let us define the following average quantities:

$$A(k) = \frac{1}{N}\sum_{n=1}^{N} A_w(\mathbf{p}_n(k-1)) \tag{32}$$

$$\mathbf{b}(k) = \frac{1}{N}\sum_{n=1}^{N} A_w(\mathbf{p}_n(k-1))\mathbf{i}_w(\mathbf{p}_n(k-1)) \tag{33}$$

$$\bar{\Delta\mathbf{p}}(k) = \frac{1}{N}\sum_{n=1}^{N} \Delta\mathbf{p}_n(k). \tag{34}$$

Let us also consider that:

$$A(k) = U(k)\Sigma(k)V(k)^T \tag{35}$$

is the singular value decomposition of the matrix $A(k)$. Note that the columns of the $M \times M$ matrix $U(k)$ and the $N_xN_y \times N_xN_y$ matrix $V(k)$ constitute orthonormal bases of the $\mathbb{R}^M$ and $\mathbb{R}^{N_xN_y}$ respectively, while matrix $\Sigma(k) = [\Sigma_0(k)\ 0_{M\times(N_xN_y-M)}]$ is of size $M \times N_xN_y$ with the diagonal matrix $\Sigma_0(k)$ containing the M singular values of matrix $A(k)$.
As it is clear from (34), $\bar{\Delta\mathbf{p}}(k)$ is the average of the optimal geometric perturbations at the $k-$th iteration of the minimization process. Note also that this average coincides with the velocity $\Delta\bar{\mathbf{p}}(k)$ of the centre of mass of the particle system defined in (19). By exploiting this fact, in the next lemma, we define the component $\mathbf{i}_1(k)$ of the "centroid" image (which differentiates the "centroid" image from the "mean" one) such that the condition (20) of Proposition 3.1 to be satisfied.

Lemma 3.2. *Let $V(k)$, $\Sigma_0(k)$, $U(k)$ be matrices resulting from the Singular Value Decomposition of matrix $A(k)$ defined in (32) and $[V_1(k)\ V_2(k)]$ be a partitioning of the matrix $V(k)$ with $V_1(k)$ and $V_2(k)$ containing the first M and $N_xN_y - M$ columns of matrix $V(k)$ defined in (35) respectively. Then, by defining the component $\mathbf{i}_1(k)$ of the "centroid" image $\mathbf{i}(k)$ as follows:*

$$\mathbf{i}_1(k) = V_1(k)\Sigma_0(k)^{-1}U(k)^T\mathbf{b}(k) \tag{36}$$

we ensure that the centre of mass of the $N-$sized particle system is motionless, i.e.:

$$\bar{\Delta\mathbf{p}}(k) \equiv \Delta\bar{\mathbf{p}}(k) = \mathbf{0}. \tag{37}$$

and thus the mean of the geometric transformations remains the same over the whole minimization process.

Proof The proof is easy and is left to the reader. □

We define its second component $\mathbf{i}_2(k)$ from the projection of the "mean" of the warped images defined in (28) onto the subspace of $\mathbb{R}^{N_xN_y}$ spanned by the column space of matrix $V_2(k)$, ie.:

$$\mathbf{i}_2(k) = V_2(k)V_2^T(k)\bar{\mathbf{i}}_w(k). \tag{38}$$

Having defined the "centroid" image, (27) can be rewriten as:

$$\Delta\mathbf{p}_n(k) = A_w(\mathbf{p}_n(k-1))(\mathbf{i}(k) - \mathbf{i}_w(\mathbf{p}_n(k-1))) \tag{39}$$

and it can be used at each iteration of the congealing algorithm for the computation of the optimal perturbations.
An outline of the proposed algorithm follows.

Input: Image set $\mathbb{S}_{\mathbf{i}_w}(\mathbb{P})$, the vector $\bar{\mathbf{p}}$ and the ROI
Set $k = 1$
repeat
  for $n = 1 : N$ **do**
    Compute the Jacobian of the image $\mathbf{i}_w(\mathbf{p}_n(k-1))$
  end
  Use (28-33) to update $\bar{\mathbf{i}}_w(k)$, $A(k)$ and $\mathbf{b}(k)$
  Compute the SVD of matrix $A(k)$ and form the matrices $V_i,\ i = 1, 2$
  Use (36) and (38) to compute the "centroid" $\mathbf{i}(k)$;
  for $n = 1 : N$ **do**
    Use (39) to compute the optimal perturbations $\Delta\mathbf{p}_n(k)$ that align image $\mathbf{i}_w(\mathbf{p}_n(k-1))$ to the "centroid" image $\mathbf{i}(k)$ and warp the image accordingly
  end
  k=k+1
until $\mathcal{E}_0(\mathbb{P})$ *has converged*;
Output: Image set $\mathbb{S}_{\mathbf{i}}$

Algorithm 1: Outline of the Proposed LS-Centroid Congealing Algorithm

4. Experimental Results

4.1. Experimental setup

In this section we are going to present our results. We conducted two experiments in order to demonstrate the superiority of the proposed scheme in terms of the computational cost, convergence rate as well as robustness. In order to control the strongness of the used warps applied to the initial images, the framework presented in Baker and Matthews (2004) was used, with the distortion parameter σ taking values in the interval $[1,\ 10]$ with the values 1 and 10 corresponding to the smallest and strongest geometric distortions respectively.

The Region Of Interest (ROI) used in the experiments we have conducted is shown in Figure 1.(a) while in Figure 1.(b) sample images from the AR database Martinez and Benavente (1998) is shown.

Fig. 1: The ROI, shown in the initial average warped image, used in our experiments (a) and sample images from the AR database (b) (please see the text for the details).

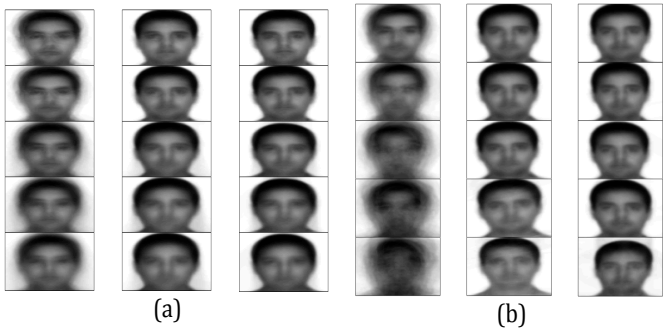

Fig. 2: (a): Initial (first column) and final average images obtained from the application of the LS-inverse algorithm (second column) and the proposed congealing algorithm (third column) on image sets of cardinality 50 (first-line), 100, 150, 200 and 300(fifth-line) respectively. (b): Initial (first column) and final average images obtained from the application of the LS-inverse algorithm (second column) and the proposed congealing algorithm (third column) on image sets of cardinality 50 with the strogness warp parameter σ taking per row the values 2, 4, 6, 8 and 10 respectively.

4.2. *Experiment I*

In this experiment we are going to apply the rivals on geometrically distorted images (resized to 96×128 pixels each) from AR database Martinez and Benavente (1998). Specifically, we used images randomly selected from the AR database. Such images initially are slightly geometrically distorted with respect to each other with σ to be bounded from above by 2. We tested the rivals on five sets of images of different size. Specifically, the cardinalities of the used image sets were 50, 100, 150, 200 and 300 respectively. The initial average of the warped images, of each image set and the optimal ones obtained by the rivals are shown in Figure 2.(a). As we can see the average images obtained by the proposed method (last column) seems to be slightly more sharp, especially for the image sets with large cardinalities. This is more evident in the "mean" image at the regions around the mouth. In addition, in Figure 3 the evolution of the missalignment function defined in (12) resulting from the application of the rivals for the first 50 iterations is shown. As we can see, for all ensemble sizes, each algorithm attains a different missalignment floor value with our congealing algorithm converging to the lowest one and with a rate which is significantly better.

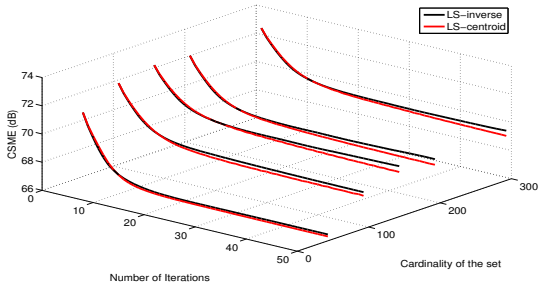

Fig. 3: Convergence curves obtained from the application of the LS-inverse (black-line) and the LS-centroid (red-line) congealing methods for the first 50 iterations, and for different sized image sets.

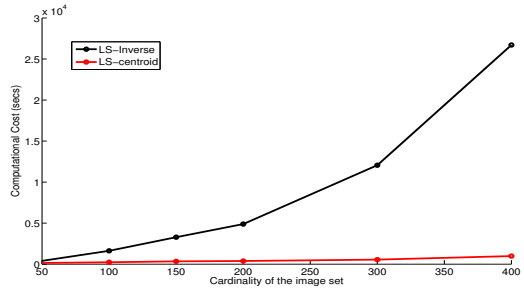

Fig. 4: The computational cost of the LS-inverse (black-line) and the LS-centroid (red-line) congealing techniques for different size of the image's sets.

Finally, in Figure 4 the computational burdens per iteration of the rivals in linear scale for the alignment of the aforementioned image sets are shown. As it is clear from this figure, the computational cost of the proposed congealing technique increases linearly with the cardinality of the image set, while the corresponding cost of its rival increases quadratically.

4.3. *Experiment II*

In this experiment we used the AR database again, applying further geometric distortions to the original images thus obtaing heavily misaligned image sets. Specifically, we apply the rivals on geometrically distorted images with the strogness of the distortions to be gradually increasing for each image set ($\sigma = 2, 4, 6, 8, 10$) while the cardinality of the set remains fixed ($N = 50$). The initial average images per set and the obtained ones from the application of the rivals, are shown in Figure 2.(b). Obviously both algorithms were affected more by the strong warps, considering that the ROI is the same we used in Experiment I. Still, our congealing technique seems to converge to sharper average images even for the strongest warps. In addition, Figure 5 shows how each method handles three specific images for each set. From this figure it is clear that alignment outliers, which remain misaligned after convergence, are treated more consistently by the proposed technique. Finally, Figure 6 holds the evolution of the misalignment function defined in (12) for each set. As it is clear the proposed technique again attains the lowest floor value while exhibiting a smooth, monotonic convergence in all cases.

Fig. 5: Samples of images of the sets before and after the application of the rivals. First column contains the initial images warped by tranformations with σ increasing with step 2 from 2 (first row) to 10 (fifth row). Second and third columns demonstrate the removed misalignment plus some misses for the LS-inverse and LS-centroid respectively.

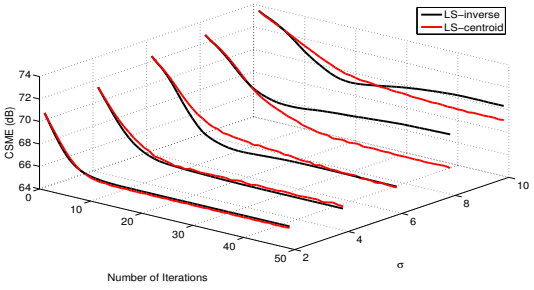

Fig. 6: Convergence curves obtained from the application of the LS-inverse (black-line) and the LS-centroid (red-line) congealing methods for the first 50 iterations, to image sets with initial misalignments of different strongness.

4.4. Experiment III

In this last experiment we are going to apply our method for the alignment of the digits contained in the MNIST database. The results we obtained are shown in Figure 7. Specifically, the average images per digit, before (first-line) and after (second-line) the proposed method is applied on the MNIST dataset LeCun et al. (1998) are shown. The increase in sharpness in the Average images shown in the second line indicates the improved alignment. We used both the training and test sets of the dataset numbering in a total of 70000 images, thus each digit holds approximately 7000 images. Finally, in Figure 8 the evolution of the missalignment function defined in (26) for each digit of the MNIST dataset resulting from the application of the proposed technique is shown. As we can see, the congealing algorithm converges for all the digits, with the corresponding obtained missalignment floor value indicating the toughness' of the digit subset to align.

5. Conclusions

This paper's goal was to further optimize the state of the art least squares based methods for the problem of joint alignment or congealing of an image ensemble. By changing the formulation of the problem, a new framework was proposed that exhibits improved characteristics when compared to the state of the art inverse composed least squares congealing technique. From a series of experiments we have conducted our algorithm, with its computational cost depending linearly on the size of the ensemble, exhibited a noticeably improvement in convergence rate and robustness.

Fig. 7: Average images per digit, before and after the proposed congealing method is applied on the MNIST dataset.

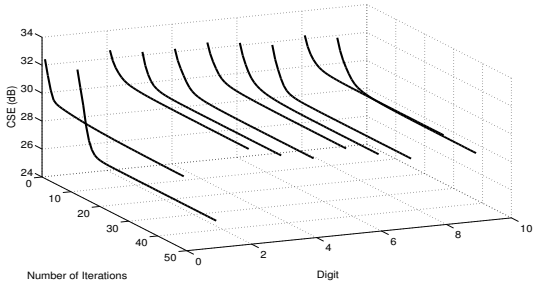

Fig. 8: Convergence curves obtained from the application of the proposed congealing method to each digit (0-9) of the MNIST dataset for the first 50 iterations.

References

Baker, S., Matthews, I., 2004. Lucas-kanade 20 years on: A unifying framework. International Journal of Computer Vision 56(3), 221255 .

Cox, M., 2010. Unsupervised alignment of thousands of images. Ph.D. thesis. Queensland University of Technology, Brisbane, Queensland.

Cox, M., Sridharan, S., Lucey, S., Cohn, J., 2008. Least squares congealign for unsupervised alignment of images. CVPR .

Cox, M., Sridharan, S., Lucey, S., Cohn, J., 2009. Least squares congealign for large number of images. CVPR .

Huang, G., Jain, V., Learned-Miller, E., 2007. Unsupervised joint alignment of complex images, in: ICCV-2007.

Huang, G.B., Mattar, M.and Lee, H., Learned-Miller, E., 2012. Learning to align from scratch. In Neural Information Processing Systems .

Lankinen, J., Kamarainen, J., 2011. Local feature based unsupervised alignment of object class images, in: BMVC, 2011.

Learned-Miller, E., 2006. Data driven image models through continuous joint alignment. IEEE T-PAMI, 28(2):236250 .

LeCun, Y., Cortes, C., Burges, C.J., 1998. The mnist database of handwritten digits. Technical Report .

Martinez, A., Benavente, R., 1998. The ar face database. Technical Report CVC .

Shokrollahi Yancheshmeh, F., Chen, K., Kamarainen, J.K., 2015. Unsupervised visual alignment with similarity graphs, in: CVPR, 2015.

Storer, M., Urschler, M., 2010. Intensity-based congealing for unsupervised joint image alignment, in: ICPR, 2010.

Tong, C., Liu, X., Willer, F., Tu, P., 2009. Automatic facial landmark labelin with minimal supervision, in: CVPR, 2009.

Vedaldi, A., Soatto, S., 2006. A complexity-distortion approach to joint pattern alignment, in: NIPS.

Xue, Y., Liu, X., 2012. Image congealign via efficient feature selection. Applications of Computer Vision (WACV) .

Zollei, L., 2006. A Unified Information Theoretic Framework for Pair- and Group-wise Registration of Medical Images. Ph.D. thesis. MIT Computer Science and Artificial Intelligence Laboratory.