Pattern Recognition Letters 95 (2017) 58–64

Contents lists available at ScienceDirect

Pattern Recognition Letters

journal homepage: www.elsevier.com/locate/patrec

A new Least Squares based congealing technique

Nikolaos Nikolikos, Emmanouil Z. Psarakis*, Nefeli Lamprinou

Department of Computer Engineering & Informatics, University of Patras, Patras 26500, Greece

ARTICLE INFO

Article history:
Received 24 November 2016
Available online 2 June 2017

MSC:
41A05
41A10
65D05
65D17

Keywords:
Image alignment
Least Squares estimation
Congealing

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 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 "center 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 [7,11]:

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

Our work improves upon the most widely recognized pixel-level LS based state-of-the-art approach. All pixel level 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 [8,14,16] as well as with LS based cost functions such as the methods proposed in [2–4,15].

In [12] an error function based on mutual information that copes with possible variations in appearance between similar objects of the same class is defined. In addition, in [5,6] 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 images of the stack by utilizing the inverse approach presented in [3], 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 sub-problems grows quadratically 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 [2]. 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.

* Corresponding author.
E-mail address: psarakis@ceid.upatras.gr (E.Z. Psarakis).

http://dx.doi.org/10.1016/j.patrec.2017.06.001

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 compositionally and not additively, since as shown in [3] the compositional warping is more robust for large image sets. Finally, the proposed technique can be easily adapted to make use of feature descriptors, instead of intensity values, as the representation of each image, as the majority of the unsupervised intensity-based techniques during the joint alignment process do, in order to cope with background variations.

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 transformations 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. Then, 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:

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

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 $\mathbb{S}_{\mathbf{i}}$ is mapped onto the pixel $\hat{\mathbf{x}}$ of the corresponding image $\mathbf{i}_w(\mathbf{p}_n)$ of set $\mathbb{S}_{\mathbf{i}_w}(\mathbb{P})$, 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 $\mathbb{S}_{\mathbf{i}_w}(\mathbb{P})$, 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 distortions, 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 every 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. Let us now concentrate ourselves on one specific of the N optimization problems defined in (8). 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 kth 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 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 [1]:

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

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 $\tilde{\mathbf{p}}_n$ 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 choice of the misalignment function

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