

Least-Squares Congealing for Large Numbers of Images

Mark Cox and Sridha Sridharan
Queensland University of Technology
Brisbane, QLD 4001, Australia
{md.cox, s.sridharan}@qut.edu.au

Simon Lucey and Jeffrey Cohn
The Robotics Institute
Carnegie Mellon University
Pittsburgh, PA 15213, USA
{slucey, jeffcjohn}@cs.cmu.edu

Abstract

In this paper we pursue the task of aligning an ensemble of images in an unsupervised manner. This task has been commonly referred to as “congealing” in literature. A form of congealing, using a least-squares criteria, has been recently demonstrated to have desirable properties over conventional congealing. Least-squares congealing can be viewed as an extension of the Lucas & Kanade (LK) image alignment algorithm. It is well understood that the alignment performance for the LK algorithm, when aligning a single image with another, is theoretically and empirically equivalent for additive and compositional warps. In this paper we: (i) demonstrate that this equivalence does not hold for the extended case of congealing, (ii) characterize the inherent drawbacks associated with least-squares congealing when dealing with large numbers of images, and (iii) propose a novel method for circumventing these limitations through the application of an inverse-compositional strategy that maintains the attractive properties of the original method while being able to handle very large numbers of images.

1. Introduction

The key assumption made in “congealing” is that the parametric form of the misalignment is known *a priori* (e.g. translation, similarity, affine, etc.) and that images in the ensemble have similar appearance when aligned. The final task in congealing is to automatically align these misaligned images in an unsupervised manner. Having the capability to align an ensemble of misaligned images stemming from the same object class has numerous applications in object recognition, detection and tracking. Canonically [5], congealing was performed through the minimization of an entropy-based cost function with respect to the parametric warp. Recently Cox *et al.* [3] demonstrated that this original approach to congealing, although performing well in some

circumstances, has some undesirable properties namely: (i) slow and poor convergence, (ii) the need to select a step-size/learning-rate, and (iii) sensitivity to how the warp is parametrized (see [3] for further details).

To remedy this problem, Cox *et al.* proposed replacing the entropy measure with a *least-squares* measure. Least-squares congealing has a number of inherent benefits over conventional congealing, namely: (i) the method exhibits fast convergence and requires no heuristic selection of a step-size or learning-rate, (ii) performance is invariant to how one linearly parametrizes the warp, (iii) the entire algorithm can be interpreted as an extended form of the well understood and studied Lucas & Kanade (LK) image alignment algorithm, and (iv) being based on the LK algorithm the method is able to perform well with a small number of images. These attractive characteristics stem from the employment of the least-squares cost function which permits the use of a Gauss-Newton optimization strategy for removing the misalignment. Utilizing the canonical entropy based cost function [5] does not lend itself to such a strategy and thus requires an inefficient optimization procedure prone to local minima. In its original formulation, the performance of least-squares congealing was shown to outperform entropy congealing on modestly sized image sets (< 100).

There are two major formulations of image alignment within the LK algorithm. The first estimates an additive increment of the warp parameters (the *additive* approach), the second a composed increment of the warp parameters (the *compositional* approach). In seminal work by Baker and Matthews [1] it has been shown that these two formulations are theoretically and empirically equivalent for the case of image-to-image matching. In this paper we will demonstrate that this case does *not* hold for the least-squares congealing algorithm. This result in itself is significant as least-squares congealing can be viewed as an extended form of the LK algorithm (see [3] for more details) where a “stack” of images is iteratively aligned, using a gradient-descent strategy, to a single image. Of more practical importance, however, is the realization that this in-

equivalence leads to an inherent benefit in the application of a compositional approach to congealing when dealing with large image sets. This result is contrary to the conventional wisdom in congealing literature, which has always employed an additive approach, irrespective of whether an entropy or least-squares cost function is being employed.

In this paper we make the following specific contributions:

- We show that least-squares congealing which uses an additive parameter update results in severe distortions for an ensemble with large initial misalignment. This is corrected by employing a compositional formulation which we show theoretically and empirically to be superior to the additive formulation.
- Argue that the object of interest within each image has a significantly higher risk of becoming irrecoverably lost when applying a single warp to a stack of images rather than the held out image. We show that this situation is not that severe for small numbers of images, but results in repeated catastrophic failure of the method when congealing larger image sets.
- To alleviate this second problem, we introduce an inverse composed formulation of least-squares congealing which is not susceptible to losing the object of interest. It achieves this while maintaining the important property of estimating the update from the highly textured stack rather than the held out image. The dramatic increase in alignment performance of the inverse-composed formulation over the composed formulation reinforces our theoretical claim that these two formulations are not equivalent.

2. Congealing

Congealing can be defined as the minimization of a misalignment function $\mathcal{E}()$ which is calculated over a set of N warped images $\mathbf{Y}()$.

$$\arg \min_{\Phi} \mathcal{E}(\mathbf{Y}(\Phi)) \quad (1)$$

where $\Phi = \{\mathbf{q}_1, \mathbf{q}_2, \dots, \mathbf{q}_N\}$ is the set of N warp parameter vectors for the warping function $\mathcal{W}(\mathbf{x}; \mathbf{q})$ and the vector \mathbf{x} is a single pixel coordinate. It is the warping function $\mathcal{W}(\mathbf{x}; \mathbf{q})$ which models the parametric form of the misalignment to be removed. For images, $\mathcal{W}(\mathbf{x}; \mathbf{q})$ controls the image region used by congealing to determine if the images are misaligned or not. The warped image ensemble is defined as

$$\mathbf{Y}(\Phi) = \{I_1(\mathbf{q}_1), \dots, I(\mathbf{q}_N)\} \quad (2)$$

where the vector function $I_i(\mathbf{q}) = [I_i(\mathcal{W}(\mathbf{x}_1; \mathbf{q})), \dots, I_i(\mathcal{W}(\mathbf{x}_M; \mathbf{q}))]$ refers

to the warped image vector of M intensity pixels from each image $I_i(\mathbf{x})$.

Irrespective of the choice of $\mathcal{E}()$, minimizing Equation 1 is a highly non-linear and computationally costly operation. The work of Learned-Miller simplified Equation 1 into an iterative procedure where in each iteration only a single parameter vector \mathbf{q}_i from Φ is improved. The work of Cox [3] utilises a similar idea, but has a different procedure for updating the parameter vectors \mathbf{q}_i in Φ . In Figure 1 we further generalise the simplification introduced by Learned-Miller to account for the work of Cox and to support the novel work introduced in this paper. Equation 3 in Figure 1

```

repeat
  for  $i = 1$  to  $N$  do
     $\mathbf{p} = \arg \min_{\mathbf{p}} \mathcal{E}_i(\mathbf{Y}, \Phi, \mathbf{p})$  (3)
     $\Phi \leftarrow f(\Phi, \mathbf{p})$  (4)
  end for
until  $\mathcal{E}()$  has converged

```

Figure 1. The iterative congealing algorithm.

presents the simplification of obtaining a single parameter vector of identical length to \mathbf{q} given the ensemble of images \mathbf{Y} and the current estimation of parameters Φ .

Equation 4 updates the current estimation Φ using the update function $f()$ and the parameter vector \mathbf{p} .

3. Entropy Congealing

The measure of misalignment used in the congealing algorithm by Learned-Miller is a function based on entropy. The rationale for such a cost function is that entropy is able to capture the true pixel distribution present in the ensemble of images to be congealed. When the ensemble of images is aligned, and if they roughly stem from the same object appearance class, then the entropy of the pixel distributions should be minimized. Employing entropy for congealing, however, is problematic due to its poor characteristics with respect to optimization. This has resulted in a suboptimal optimization strategy being adopted which requires: (i) sequential improvement of parameters, (ii) redundant parameters which account for the common modes of variation, and (iii) a step size specification for each parameter [5, 3].

For entropy congealing, the simplified congealing Equation 3 is as follows

$$\mathbf{p} = \arg \min_{\mathbf{p}} \mathcal{E}_E(I_i(\mathbf{q}_i + \mathbf{p}), \mathbf{Y}_i(\Phi)) \quad (5)$$

where \mathcal{E}_E is the measure of misalignment using entropy, I_i is a held out image selected from the image ensemble \mathbf{Y} , and \mathbf{Y}_i is the *stack* which consists of the ensemble \mathbf{Y} without the held out image.

$$\mathbf{Y}_i(\Phi) = \{I_j(\mathbf{q}_j); \forall j \neq i\} \quad (6)$$

Since it is the held out image that is sequentially improved, the update function $f()$ is simply

$$\mathbf{q}_i \leftarrow \mathbf{q}_i + \mathbf{p} \quad (7)$$

4. Least-Squares Congealing

Least-squares congealing (LSC) is an extension of the Lucas & Kanade (LK) image-to-image alignment algorithm [7]. In fact, if just two images are given as input to LSC, the algorithm is the LK image alignment algorithm,

$$\sum_{\mathbf{x}} \varphi \left([I_1(\mathcal{W}(\mathbf{x}; \mathbf{q}_1 + \mathbf{p})) - I_2(\mathcal{W}(\mathbf{x}; \mathbf{q}_2))]^2 \right) \quad (8)$$

where the objective is to minimize \mathbf{p} within the sum of squared differences cost function that measures the misalignment between image $I_1()$ and $I_2()$. The robust error function φ is included to make the measure of misalignment less sensitive to outliers.

For the rest of this paper we will use the vector notation for compactness via the $\|\cdot\|_\varphi$ operator giving,

$$\arg \min_{\mathbf{p}} \|I_1(\mathbf{q}_1 + \mathbf{p}) - I_2(\mathbf{q}_2)\|_\varphi \quad (9)$$

as an equivalent definition of Equation 8. We now present the measure of misalignment function for LSC in Equation 10,

$$\mathcal{E}_{\text{LS}}(I_i(\mathbf{q}_i), \mathbf{Y}_i(\Phi + \mathbf{p})) = \sum_{j \neq i}^N \|I_j(\mathbf{q}_j + \mathbf{p}) - I_i(\mathbf{q}_i)\|_\varphi \quad (10)$$

which for $N = 2$ results in exactly the same measure of misalignment as originally defined in the LK algorithm by Equation 8. The parameter update function $f()$ for least-squares congealing applies the parameters \mathbf{p} to each image in the stack \mathbf{Y}_i and is therefore,

$$\mathbf{q}_j \leftarrow \mathbf{q}_j + \mathbf{p}; \forall j \neq i \quad (11)$$

Minimizing Equation 10 with respect to the warp parameters \mathbf{p} has been presented in depth in other work. We point the reader to [3, 1, 2] for a thorough presentation on minimizing Equation 10 using the Gauss-Newton optimization strategy that is used in this paper.

Of interest in Equation 10 is that the parameters \mathbf{p} are applied to the stack instead of the held out image. This is contrary to entropy congealing. Formulating the equation this way is crucial as it preserves important texture and gradient information from each image in the stack. The alternative version which has the parameters \mathbf{p} applied to the held out image $I_i()$ results in the minimization of the measure of misalignment collapsing to the alignment of the held out image with the *average* image of the stack. This situation is suboptimal, as the averaging process on misaligned images

destroys all the fine detail and texture information contained in each individual image, resulting in a single blurred image which is a poor candidate as a target alignment [3].

The central focus of this paper are the problems which arise from applying a single warp to multiple images and to provide solutions to these problems that preserve this important fine detail and texture information.

5. Compositional Updates

It has been demonstrated, in the seminal work of Baker and Matthews [1], that the LK algorithm has equivalent alignment performance theoretically and empirically to the *additive* approach in Equation 8 if we minimize,

$$\sum_{\mathbf{x}} \varphi \left([I_1(\mathcal{W}(\mathbf{x}; \mathbf{q}_1 \circ \mathbf{p})) - I_2(\mathcal{W}(\mathbf{x}; \mathbf{q}_2))]^2 \right) \quad (12)$$

with respect to \mathbf{p} which is often referred to as the *compositional* approach. Equation 13 presents the application of parameters \mathbf{p} to \mathbf{q} using composition [8].

$$\mathbf{q} \circ \mathbf{p} = \mathcal{W}(\mathcal{W}(\mathbf{x}; \mathbf{p}); \mathbf{q}) \quad (13)$$

Obviously for the case of LSC when the stack size is $N = 2$ this equivalence still holds. However, in this section we investigate the interesting case for when $N > 2$. Specifically, we shall demonstrate in this section that the additive approach for LSC when $N > 2$, unlike the compositional approach, is extremely sensitive to the initial warps of images in the stack. Figure 2 provides an illustration of this problem. Here, two images (shown in (a)) consisting of triangles are given as input to the congealing process. For argument's sake, a solution which perfectly aligns each triangle has been provided with the parameters \mathbf{q}_1 and \mathbf{q}_2 . The resulting aligned images are shown in (b). To illustrate the inconsistency of additive warps, we calculate the additive warp which rescales image $I_2(\mathbf{q}_2)$. The resulting triangle with the additive warp applied is shown in the bottom triangle of part (c). If we apply this same additive warp to $I_1(\mathbf{q}_1)$ the inconsistent behaviour is observed. As shown in the top triangle of (c), the triangle has been distorted to the point that it is no longer a triangle.

To rectify this situation so that the applied warp will operate consistently for all images in the stack regardless of initial scale, a compositional approach to image warps is required. Using the same strategy as the additive update in item (c) of Figure 2 we calculate a single warp \mathbf{p} which when composed with $I_2(\mathbf{q}_2)$ results in the bottom image shown in item (d). When applying the same warp to $I_1(\mathbf{q}_1)$ using composition we see that it too has been rescaled to exactly the same location. From this simple thought experiment we have demonstrated that the additive and compositional approaches to LSC are not equivalent for the case of $N > 2$.

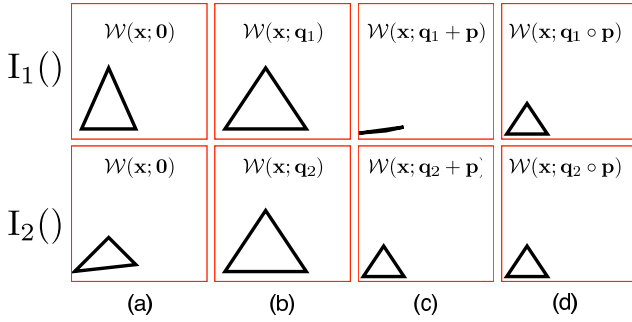


Figure 2. This figure illustrates the inconsistency of applying the same warp additively to multiple images. The triangles in (a) represent the original input images to the congealing algorithm; (b) represents the same triangles warped such that they are aligned; (c) shows the inconsistent behaviour of applying the same additive warp to each triangle (d) instead of applying an additive warp, a composed warp is applied resulting in consistent behaviour. This behavior is contrary to Lucas & Kanade (LK) image-to-image alignment where it has been shown that the additive and composed formulation are equivalent theoretically and empirically. In least-squares congealing however, the formulations are only equivalent when the number of images to be aligned is $N = 2$; which is when LSC is the LK alignment algorithm.

Having illustrated the inconsistent behaviours of applying single warps additively to a stack of images, we modify LSC to utilize composable warps. These modifications are shown with the new measure of misalignment function \mathcal{E}_{LS} in Equation 14 and the parameter update function $f()$ in Equation 15.

$$\mathbf{p} = \arg \min_{\mathbf{p}} \sum_{j \neq i} \|I_j(\mathbf{q}_j \circ \mathbf{p}) - I_i(\mathbf{q}_i)\|_{\varphi} \quad (14)$$

$$\mathbf{q}_j \leftarrow \mathbf{q}_j \circ \mathbf{p}; \forall j \neq i \quad (15)$$

Minimizing Equation 14 is achieved the same way as the additive update equation with minor changes due to the composed warp.

6. Object Loss of Outlier Images

The previous section outlined the first major issue with LSC and showed that additive updates are not equivalent to composed updates when applying the same warp to multiple images. We now proceed to the second major issue which again stems from the fact that a single warp is applied to multiple images. Specifically, there is a risk that after applying a single warp to a collection of images, one or more of the objects of interest will be partly removed from view. In the extreme case, the object will be completely removed from view.

To illustrate, we point the reader to Figure 3 where we see this phenomenon occurring for the extreme case. Here, a stack of input images (a) are to be aligned to the currently

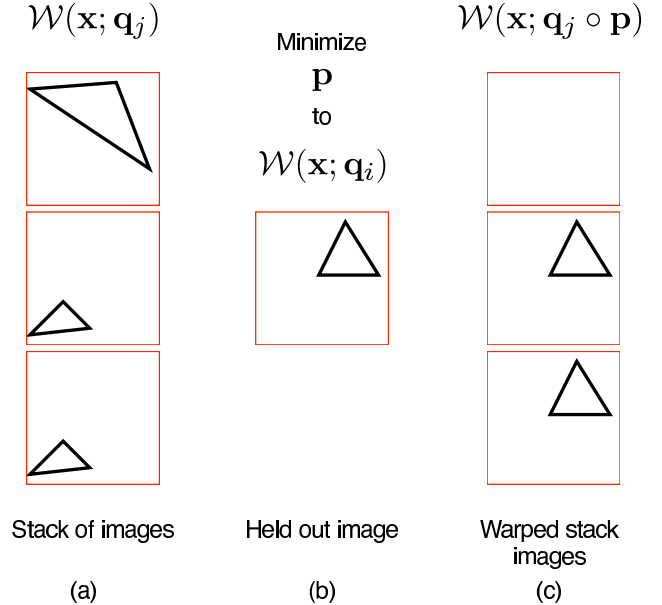


Figure 3. A stack of images shown in (a) is to be aligned to the held out image (b) using least squares congealing. We assume that the robust error function identifies the top triangle in (a) as an outlier image and de-emphasises that image's contribution to the calculation of the warp parameters \mathbf{p} . With the outlier essentially removed, the alignment can be thought of as calculating a warp which aligns the identical images with the held out image. In this case, \mathbf{p} has been calculated perfectly and applied to each image in the stack, including the outlier image. As is shown in (c), the two identical images have been aligned perfectly, but the outlier image has been warped such that the triangle has been completely removed from view.

selected held out image (b) utilizing a compositional LSC implementation. Within the stack of input images, two triangles are identical both in appearance and geometric location and the other triangle is substantially different both in terms of location and in appearance. Due to the substantial difference, we assume that any robust error function will significantly squash the outlier image's contribution to the alignment process. Thus obtaining the warp parameters \mathbf{p} which minimize the expression can be approximated as finding the warp which aligns the two identical images with the held out image. In Figure 3, the \mathbf{p} in this case has been calculated perfectly and applied to the two identical images resulting in the bottom two triangles of (c). Since the outlier image was part of the stack, it too was warped using \mathbf{p} and resulted in the top image of (c). Notice how the triangle has been completely removed from view. Once removed, the object is irrecoverably lost as the congealing process can no longer obtain any alignment information from the object.

One possible way of correcting this is to detect when the field of view is no longer intersecting with the bounding area of the original image and then reset the parameters for the offending image back to its initial estimate. This

may work for the case when there is only a small number of outliers with significant appearance and geometric variation. For instance, assume that the outlier image is selected to be the held out image. The stack consisting of mostly aligned images is then aligned with the outlier image. After completion of this process, the outlier image is now aligned with the stack, and is thus no longer an outlier image.

In the event that there is no dominant group of aligned images in order to bring outliers into the dominant group, outliers will unfortunately remain outliers. Resetting the parameters in this case is treating a symptom of LSC, not its underlying cause. The cause in this case is, again, calculating a single warp that is applied to multiple images of differing geometric variations. If we look to the Learned-Miller’s entropy congealing algorithm[5], we see that this problem does not exist as the held out image is aligned with the stack. As stated previously, this formulation of the problem is sub-optimal for a least-squares cost function as the process collapses to aligning the held out image with the average image of the stack.

In the next section we exploit the geometric symmetry that exists with compositional warps to make LSC behave as if the held out image is aligned to the stack, but without the drawbacks of aligning the held out image with the average image.

7. Inverse-Composed LSC

In Section 5 we saw the move from additive warps to compositional warps. The move to a compositional warp has an interesting history in computer vision. The work of Baker and Matthews [1] exploited the compositional warp with great gains in computational efficiency with their inverse-composition approach to image-to-image alignment. However, for least-squares congealing, the computational gains are of lesser importance compared to the exploitation of the geometric symmetry the method exploits. LSC uses this symmetry in order to address the object loss problem described in the previous section.

The symmetry which inverse-composition takes advantage of stems from the assumption that two images of similar objects are related geometrically. A compositional warp which aligns one image with another can be inversely applied to the other image in order to achieve an equivalent result in alignment. i.e. in either case, the resulting objects are aligned, however their final alignment may have differing geometric appearance depending on which image was fixed as a template.

In this paper we use a compact notation to refer to the inverse composed application of warp parameters. This notation is as follows

$$\mathbf{q} \circ \mathbf{p}^{-1} = \mathcal{W}(\mathcal{W}^{-1}(\mathbf{x}; \mathbf{p}); \mathbf{q}) \quad (16)$$

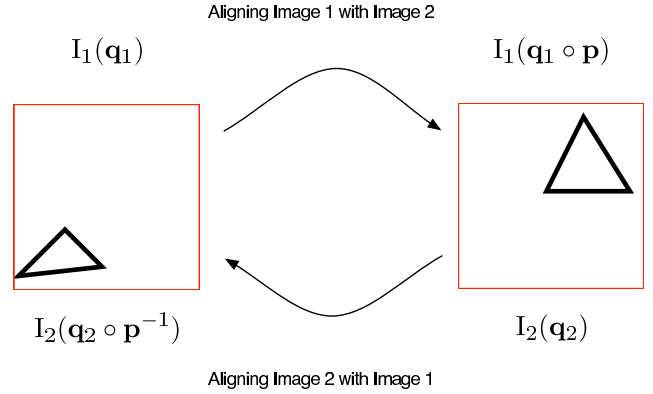


Figure 4. The illustration of the geometric symmetry that exists with composable warps. Here two images of triangles I_1 and I_2 are related by a composable warp $\mathcal{W}(\mathbf{x}; \mathbf{p})$ which when applied to I_1 results in an image like I_2 . Conversely, when the warp $\mathcal{W}(\mathbf{x}; \mathbf{p})$ is inversely applied to I_2 this results in an image like I_1 .

Figure 4 diagrammatically shows this symmetry of composable warps. Here two images of triangles $I_1(\mathbf{q}_1)$ and $I_2(\mathbf{q}_2)$ are related by a composable warp $\mathcal{W}(\mathbf{x}; \mathbf{p})$ which when applied to I_1 results in an image like I_2 . Conversely, when the warp $\mathcal{W}(\mathbf{x}; \mathbf{p})$ is inversely applied to I_2 this results in an image like I_1 .

How this applies to LSC should now be obvious. If we replace image I_1 with a stack \mathbf{Y} , the same geometric relationship between the stack and the held out image exists. Thus, we can calculate \mathbf{p} using exactly the same method as before, but instead of updating the parameters \mathbf{q}_j of the stack by \mathbf{p} , we apply the inverse to the held out image. This procedure is shown in Figure 5 and mathematically in Equations 17 and 18.

$$\mathbf{p} = \arg \min_{\mathbf{p}} \sum_{j \neq i} \|I_j(\mathbf{q}_j \circ \mathbf{p}) - I_i(\mathbf{q}_i)\|_{\varphi} \quad (17)$$

$$\mathbf{q}_i \leftarrow \mathbf{q}_i \circ \mathbf{p}^{-1} \quad (18)$$

One drawback of this formulation is that during the calculation of the warp parameters \mathbf{p} , object loss may still occur. Fortunately, this can be remedied by moving the inversion inside the Gauss-Newton strategy which calculates \mathbf{p} . This procedure is a direct extension of the inverse-composition method of image-to-image alignment proposed in [1].

The Gauss-Newton strategy for minimizing \mathbf{p} is an iterative procedure in which the initial estimate of \mathbf{p} is incrementally improved by $\Delta \mathbf{p}$ in each iteration.

$$\arg \min_{\Delta \mathbf{p}} \sum_{j \neq i} \|I_j(\mathbf{q}_j \circ \mathbf{p} \circ \mathbf{0} + \Delta \mathbf{p}) - I_i(\mathbf{q}_i)\|_{\varphi} \quad (19)$$

$$\mathbf{p} \leftarrow \mathbf{p} \circ \Delta \mathbf{p} \quad (20)$$

As can be seen in Equation 19, the calculation of $\Delta \mathbf{p}$ is always from the stack to the held out image. Ensuring that the

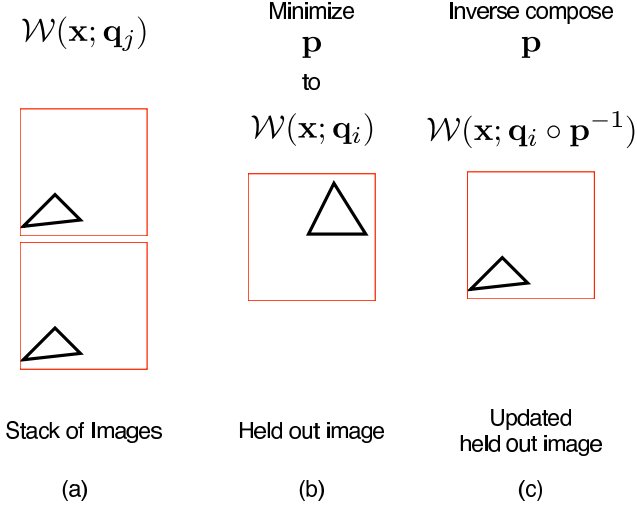


Figure 5. This figure demonstrates the inner workings of the inverse composed formulation of least squares congealing. A warp is calculated which minimizes the measure of misalignment between the stack of images (a) and the held out image (b). A forward composed formulation would then have the warp applied to the stack which could cause loss in outlier images. The inverse composed formulation avoids this situation by inverting the warp and having it applied to the held out image instead.

stack remains fixed is achieved by applying the incremental update to the held out image parameters \mathbf{q}_i or

$$\arg \min_{\Delta \mathbf{p}} \sum_{j \neq i} \|I_j(\mathbf{q}_j \circ \mathbf{0} + \Delta \mathbf{p}) - I_i(\mathbf{q}_i \circ \mathbf{p})\|_{\varphi} \quad (21)$$

$$\mathbf{p} \leftarrow \mathbf{p} \circ \Delta \mathbf{p}^{-1} \quad (22)$$

We see in Equation 21 that the incremental update $\Delta \mathbf{p}$ is minimized in a manner which still has the stack of images aligned to the held out image, but instead of applying the update to the stack of images, we use the geometric symmetry of composable warps to have its inverse applied to the held out image.

With this formulation of least squares congealing the parameter update function $f()$ is then simply,

$$\mathbf{q}_i \leftarrow \mathbf{q}_i \circ \mathbf{p} \quad (23)$$

We title this method *inverse-composed least-squares congealing*. For the remainder of this paper we distinguish the four different formulations of least squares congealing (LSC) as follows: (i) when using an additive update, we refer to this as additive LSC, (ii) when using a composed update applied to the stack of images, we refer to this as forward composed LSC, (iii) when using a composed update calculated on the stack of images which is then applied to the held out image via inverse composition, we refer to this as partial inverse-composed LSC and lastly (iv) is inverse-composed LSC which refers to the inverse composed in-

cremental update $\Delta \mathbf{p}$ that is applied to the held out image inside the Gauss-Newton optimization strategy.

8. Performance

In this section we present the alignment capability of each of the LSC formulations. We show (i) that a significant improvement in performance is obtained when moving from additive warps to composed warps and (ii) moving to an inverse-composed formulation has a dramatic effect on congealing performance for large numbers of images.

In all experiments presented in this section, images and ground truth labels of the eyes and nose were sourced from the MultiPIE data set [4]. In all cases, the images used were of front facing faces demonstrating no variations in expression and were taken under uniform lighting conditions.

To simulate initial misalignment we perturbed each of the images by a random warp. The magnitude of this warp corresponded to a displacement of the labelled landmarks equal to a percentage of the distance between the left eye and right eye. After applying the perturbation, we tracked the labeled landmarks throughout the congealing process until the process had converged or reached a maximum number of iterations.

In order to remove any global changes in scale caused by congealing, a transform which warped the average landmark position of the tracked labels to a landmark template was applied to the tracked landmarks. Once normalized, the distance between the normalized labels and the corresponding landmark template was collated. With the distribution of distances collected, it is now possible to determine the percentage of the tracked landmarks which are within a certain distance to the template. Obviously, the higher the percentage of points below this threshold the better.

For the purpose of comparing performance for the various factors which have an effect on congealing, we fixed a pixel distance threshold as the point to sample the above cumulative distribution. We term this percentage as the percentage of points converged. Other parameters that were fixed for all experiments were: the field of view which congealing uses to measure misalignment had an area of 30x30 pixels ($M = 900$) and the parametric function $\mathcal{W}(\mathbf{x}; \mathbf{p})$ which models the type of misalignment to be removed was set to be an affine warp.

Additive versus Compositional Warps: The results presented in Figure 6 show the alignment performance as a function of iteration when using a compositional warp over an additive warp. The stationary nature of the curve for the additive warp illustrates the inability of the additive formulation to cause any improvement in alignment due to the inconsistent changes in geometry that occur when applying the same warp additively to multiple images. The compositional approach however, does not suffer from this effect.

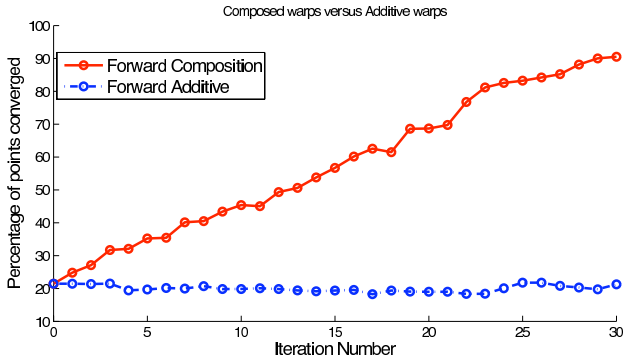


Figure 6. The difference in alignment capability between least-squares congealing using an additive update versus a forward composed update. Here we see that the additive approach performs terribly due to the inconsistent behaviour exhibited when applying a single warp additively to multiple images. This test was conducted utilizing 20 images of faces with an initial random error of magnitude equal to 40% of the distance between the left and right eye.

The curves shown correspond to the best performing robust error configurations for both the additive and composed least-squares congealing formulations. In this experiment, 20 images from MultiPIE were extracted and randomly misaligned to 40% of the distance between the left and right eye.

Object Loss and Inverse-Composed LSC: The next set of results shown in Figure 7 illustrate the effect that object loss has on alignment performance as a function of the numbers of images to align. Here, a forward composed least-squares congealing algorithm is compared against the partial and fully inverse composed formulation. As can be seen from the figure, a forward composed formulation is comparable to an inverse composed formulation when congealing no more than 30 images, but when aligning 40 or more images, the difference in alignment performance becomes more significant. Interestingly, retaining or resetting object loss appears to be the main factor to the degraded performance, as the difference in performance between partial and fully inverse-composed congealing is marginal. This is interesting, as object loss can still occur during the partial formulation as the parameter inversion occurs after the calculation of the warp parameters.

The significant difference in performance between the forward and inverse composed formulations indicate that the two methods are not equivalent. This is contrary to Lucas & Kanade (LK) image-to-image alignment where this is the case. This statement is only true when a stack of images is involved ($N > 2$). For the case when there are only two images to align using LSC ($N = 2$), the equivalence between the forward and inverse composed formulation still holds as LSC with $N = 2$ is the LK method.

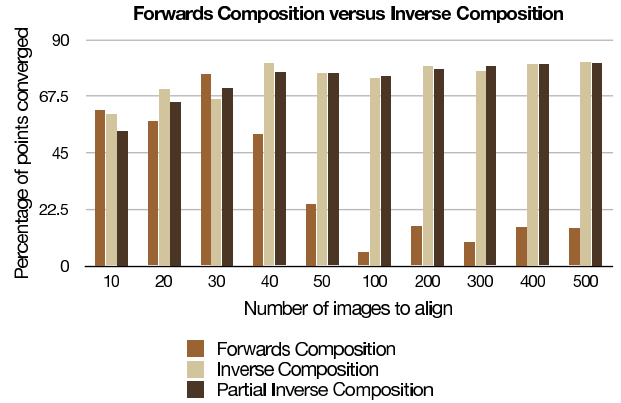


Figure 7. We illustrate the alignment performance difference between forward composition and inverse composition for differing numbers of images. The y-axis reflects the percentage of tracked points within a threshold of the same image aligned using the ground truth labels.

The experimental results presented in Figure 7 were conducted on collections of images which were initially misaligned to a distance of 50% of the distance between the left and right eye. In all cases, a number of robust error configurations were trialed and the best performing configuration was selected.

Sample output of applying inverse-composed congealing to a synthesized data set consisting of 500 misaligned images can be found in Figure 8. The run time of this application is approximately 16 hours on an Intel Xeon 2.66GHz CPU. Our (single threaded) implementation is based on the assumption that there exists a data set which can not be loaded entirely in to memory. Considerably faster implementations of inverse-composed congealing are possible when not making this assumption.

In Figure 9 we apply congealing to the MNIST handwritten digit database [6] to illustrate its ability to scale to data sets consisting of many thousands of images. We demonstrate its performance by showing the average image of the samples for each digit class before and after congealing.

9. Conclusion

In this paper we showed that the original formulation of least-squares congealing utilizing an additive parameter update to each image in the stack results in inconsistent geometric transformations of images. We also showed that by using a compositional formulation, this inconsistent behavior is removed, resulting in a dramatic improvement in alignment performance. This result confirms that the additive and compositional formulation of least-squares congealing are not equivalent when aligning more than two images, unlike the case for image-to-image matching using the Lucas & Kanade algorithm.

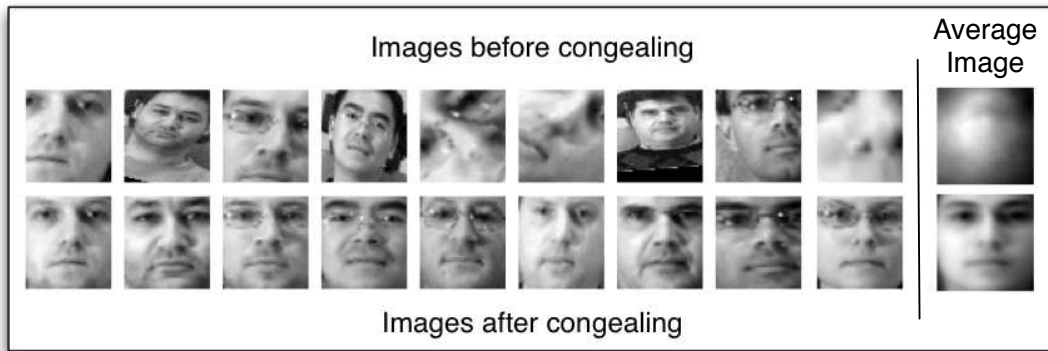


Figure 8. A sample of the output of the inverse-composed least squares congealing algorithm presented in this paper on 500 images. The initial misalignment for these images was set to a point displacement of 50% of the distance between the left and right eye. The average images on the right are on all 500 images. In this case 35 iterations of congealing were performed. The field of view used was fixed to 30x30 pixels. No coarse to fine strategy was used.

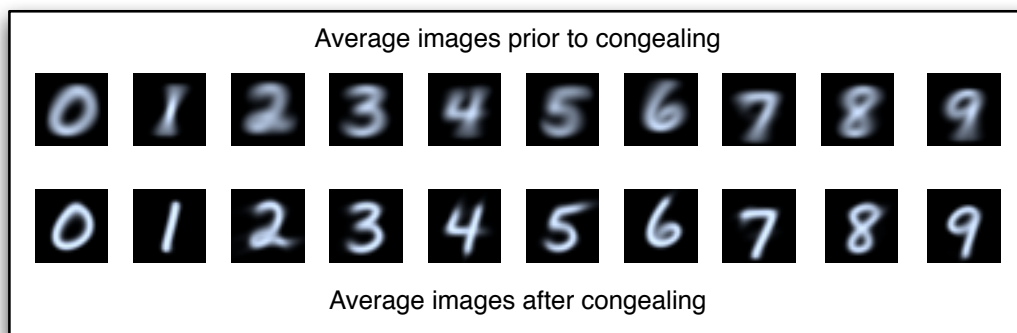


Figure 9. We apply congealing to approximately 6000 samples of each handwritten digit in order to demonstrate least-squares congealing's ability to scale to many thousands of images. The samples were obtained from the MNIST [6] digit database.

We made an additional contribution in this paper by identifying that outlier images in the image stack are at risk of having the object of interest being irrecoverably lost. To alleviate this problem, we introduced an inverse composed formulation of LSC which does not suffer from this problem. It achieves this whilst still maintaining the important property of aligning the stack of images with the held out image.

References

- [1] S. Baker and I. Matthews. Equivalence and efficiency of image alignment algorithms. In *Proceedings of the 2001 IEEE Conference on Computer Vision and Pattern Recognition*, volume 1, pages 1090 – 1097, December 2001.
- [2] S. Baker and I. Matthews. Lucas-Kanade 20 years on: A unifying framework. *International Journal of Computer Vision*, 56(3):221–255, Feb. 2004.
- [3] M. Cox, S. Lucey, S. Sridharan, and J. Cohn. Least squares congealing for unsupervised alignment of images. In *IEEE International Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2008.
- [4] R. Gross, I. M. S. Baker, and T. Kanade. The CMU Multiple pose, illumination and expression (MultiPIE) database. Technical Report CMU-RI-TR-07-08, Robotics Institute, Carnegie Mellon University, 2007.
- [5] E. Learned-Miller. Data driven image models through continuous joint alignment. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 28(2):236–250, 2006.
- [6] Y. LeCun and C. Cortes. The mnist database. Online, May 2007. <http://yann.lecun.com/exdb/mnist/>.
- [7] B. D. Lucas and T. Kanade. An iterative image registration technique with an application to stereo vision. In *Proceedings of the 1981 DARPA Image Understanding Workshop*, 1981.
- [8] H.-Y. Shum and R. Szeliski. Systems and experiment paper: Construction of panoramic image mosaics with global and local alignment. *International Journal of Computer Vision*, 36(2):101–130, 02 2000/02/01/.