

An Efficient Stochastic Approach to Groupwise Non-rigid Image Registration

Kirill A. Sidorov
School of Computer Science
Cardiff University, UK

K.Sidorov@cs.cardiff.ac.uk

Stephen Richmond
School of Dentistry
Cardiff University, UK

richmonds@cardiff.ac.uk

David Marshall
School of Computer Science
Cardiff University, UK

dave@cs.cardiff.ac.uk

Abstract

The groupwise approach to non-rigid image registration, solving the dense correspondence problem, has recently been shown to be a useful tool in many applications, including medical imaging, automatic construction of statistical models of appearance and analysis of facial dynamics. Such an approach overcomes limitations of traditional pairwise methods but at a cost of having to search for the solution (optimal registration) in a space of much higher dimensionality which grows rapidly with the number of examples (images) being registered. Techniques to overcome this dimensionality problem have not been addressed sufficiently in the groupwise registration literature.

In this paper, we propose a novel, fast and reliable, fully unsupervised stochastic algorithm to search for optimal groupwise dense correspondence in large sets of unmarked images. The efficiency of our approach stems from novel dimensionality reduction techniques specific to the problem of groupwise image registration and from comparative insensitivity of the adopted optimisation scheme (Simultaneous Perturbation Stochastic Approximation (SPSA)) to the high dimensionality of the search space. Additionally, our algorithm is formulated in way readily suited to implementation on graphics processing units (GPU).

In evaluation of our method we show a high robustness and success rate, fast convergence on various types of test data, including facial images featuring large degrees of both inter- and intra-person variation, and show considerable improvement in terms of accuracy of solution and speed compared to traditional methods.

1. Introduction

Image registration is a fundamental problem in computer vision with many applications ranging from character recognition [8], medical imaging [16] to modelling of facial dynamics [2]. Typically this involves the analysis of deformable structure in groups of images and the construction of some statistical model of appearance [3]. The fundamen-

tal challenge is to automatically find dense correspondences between images of deformable objects — a topic that has recently made significant progress. Furthermore, unsupervised non-rigid registration methods are important for large data sets for which manual annotation is too time consuming or impractical. Existence of such a method makes it possible to *automatically* construct statistical models of appearance in an entirely unsupervised fashion.

Conventional registration methods [17] work only on pairs of images. Repeated application of a pairwise registration algorithm will inevitably be biased to the choice of reference image which leads to errors in the final alignment [9]. In addition, an unfortunate choice of reference, for example missing features, will corrupt the algorithm further. To combat such issues, groupwise approaches that bring an entire group of images into registration simultaneously have been recently developed [1, 2, 9]. Here, the information from the *entire* data set is being utilised at each stage, rather than from only a pair. Only by considering *multiple examples* simultaneously can corresponding structures be reliably identified.

The groupwise approach to finding dense correspondence across a set of unlabelled examples (images or shapes) has been shown to be superior to pairwise methods [2]. However, unlike with pairwise methods, the dimensionality of the space in which the search for the optimal solution is performed, grows very rapidly with the number of samples in the set when groupwise approach is employed. In practical problems, very high dimensionality of the search space presents a significant obstacle to finding the optimal solution [3]. Suppose N images are to be registered, and the correspondences between images are controlled by K degrees of freedom per image, yielding $2NK$ degrees of freedom in total; even for a modest data set (hundreds of images) and a modest resolution of deformations (tens of degrees of freedom per image) the dimensionality of the space in which the solution is to be found is measured in thousands.

Groupwise registration may be decomposed into three subproblems [5]:

- A mechanism for representing and manipulating dense correspondence between images.
- An objective function F with a minimum at a point corresponding to the desired registration.
- A global minimisation algorithm which optimises F .

Global minimisation of the objective function, F , whose arguments are correspondences between all images and whose value measures the quality of registration, solves the problem.

The problem of efficient optimisation of the very high dimensional objective function in the context of groupwise image registration has not been extensively explored in the literature. Most traditional optimisation algorithms (used in *e.g.* [1, 2]) cannot deal with an optimisation problem of such magnitude and tend to converge to local minima. Some stochastic algorithms (*e.g.* simulated annealing and genetic algorithms), which attempt to avoid local minima, have impractical computation times even for small data sets. The main contribution of this paper is to describe an efficient optimisation framework for many-dimensional groupwise objective functions for non-rigid image registration that can quickly and reliably find very good (and in practise almost always the best) minima.

Our approach alleviates the “curse of dimensionality” on two fronts:

- We propose a novel solution that implicitly reduces the dimensionality of the search space as the search progresses by incrementally learning optimal deformations.
- We propose a novel application of stochastic optimisation algorithms that do not significantly degrade in performance as the dimensionality grows.

Additionally, due to the efficient formulation of our approach, it is amenable for GPU implementation — apart from the control logic all steps can be performed on a GPU.

Due to the robustness of our approach we are also able to perform *inter-person* groupwise registration — we take a corpus of individual face images [11] and can successfully register them. This is the first time that the automatic non-rigid registration of data possessing such variety has been reported.

2. Problem statement: groupwise registration

The input to the registration algorithm is a (possibly unordered) set of N images $\{\mathcal{I}_i, i = 1 \dots N\}$ of different examples of a deformable object or a deformable structure in an object. We seek to automatically (without user intervention) derive a dense spatial correspondence between the examples. Deformation fields (one for each image) define spatial correspondences between the images, by specifying where each pixel on the underlying object structure is located on that image.

3. Our groupwise registration algorithm

As mentioned in Section 1, we regard the problem of groupwise registration as an optimisation problem. Before we explain the optimisation regime, we need to define the deformation mechanism and the objective function to minimise.

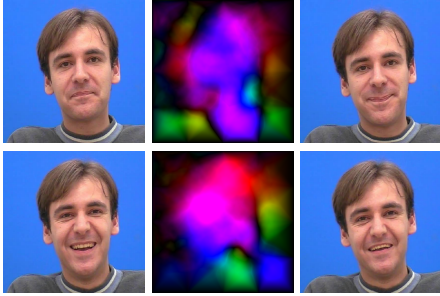
3.1. Deformation mechanism

We need to represent deformation fields that transform one image into another. The most common ways to represent them are: fields defined by a sparse set of control points which control either a set of splines [16] or nodes of a triangulated mesh [2], compositions of simple warps [1] or dense fields representing the displacement of each pixel [7]. In our approach we use both dense deformation fields (to store accumulated deformations) and fields controlled by sparse set of points (when incrementally improving accumulated deformations, see Section 3.2). In the latter case, for computational efficiency, we piece-wise linearly interpolate the deformation inside the triangles of a mesh obtained by triangulating the set of control points, given the deformations at vertices (control points), as in [2]. Although this representation has been criticised for being insufficiently smooth [1] and have limited flexibility and spatial resolution [12], as we will show below, this is no longer a problem with our methods and we are thus able to benefit from its computational efficiency without any sacrifices (see also Section 4).

Let \mathcal{D} denote a dense deformation field; we store a deformation field in a vector-valued matrix (deformation map, of the same size as the images) whose elements contain the pixel displacements. Thus, given an image \mathcal{I} and a dense deformation map \mathcal{D} , a warped image \mathcal{I}' is obtained thus: $\mathcal{I}'(\mathbf{x}) = \mathcal{I}(\mathbf{x} + \mathcal{D}(\mathbf{x}))$, $\forall \mathbf{x}$, which we will abbreviate to $\mathcal{I}' = W(\mathcal{I}, \mathcal{D})$, this is illustrated in Fig. 1. Given two (ordered) sets of control points C_1 and C_2 placed correspondingly in images \mathcal{I}_1 and \mathcal{I}_2 , a dense deformation map from \mathcal{I}_1 to \mathcal{I}_2 can be obtained by interpolation as described above; we will abbreviate it thusly: $\mathcal{D} = L(C_1, C_2)$.

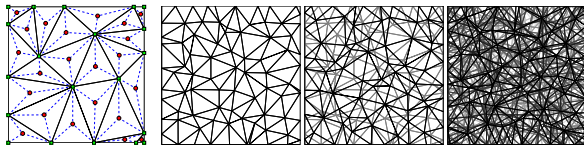
3.2. Incrementally learning optimal deformation

Let there exist a dense set of control points, C_D , and a sparse subset of C_D , $C_S \in C_D$. Suppose, for ease of explanation, that they are vertices of a triangular mesh describing the deformation of an image (in Fig. 2(a) C_D is all points, C_S is points on the solid lines). Let C_{D-S} denote the set of points in C_D not in C_S (in Fig. 2(a) solely on dashed lines). If we could express the optimal position of points in C_{D-S} as a function of optimal position of points in C_S then this would obviously yield a dimensionality reducing reparameterisation of the deformation — by means of which we would be able to control more complex deformations with the same number of control points (increased resolution) or



Sample Deformation Transformed

Figure 1. Samples from the data set [6] together with the computed deformation maps and shape-normalised images. In deformation maps, color indicates direction, and brightness — the magnitude of the displacement.



(a) C_D and C_S (b) 1 mesh (c) 2 meshes (d) 8 meshes

Figure 2. (a) A dense mesh and its subset. (b)–(d) Superposition of random meshes.

control the deformations with a smaller number of control points keeping the same resolution (dimensionality reduction).

Unfortunately, such a function is not known in advance. Instead, we propose to incrementally learn and accumulate the optimal dense deformation everywhere *between* the sparse control points. The *power of our algorithm* comes from this fact. We only use control points when searching for an optimal improvement for an already established *dense* deformation map. Moreover, instead of using only one set of control points, as other approaches do, we randomly generate a completely new set of control points at each minimisation stage (see Alg. 2). This allows us to approximate more and more complex deformation fields as a superposition of simple deformation fields (parameterised by control points at each stage), as the algorithm progresses. Fig. 2(b)–(d) illustrates this idea: as we use a new random mesh each time we progressively sample points on the underlying accumulated deformation field with increasing resolution. It is critical to note that when a new set of control points is generated, the improvements obtained using the previous set are not lost but are accumulated in dense deformation maps. Also note, the number of control points (and thus the number of optimisation parameters) always remains the same. The key to the power stems therefore from the ability of our algorithm to search for more and more complex deformations whilst keeping the dimensionality of the search space constantly low.

3.3. Objective function

Let $G(\mathcal{I}_i, \mathcal{I}_j)$ be the pixelwise discrepancy function for images \mathcal{I}_i and \mathcal{I}_j (of size $r \times c$): $G(\mathcal{I}_i, \mathcal{I}_j) = \frac{1}{rc} \sum_{\forall \mathbf{x}} g(\mathcal{I}_i(\mathbf{x}), \mathcal{I}_j(\mathbf{x}))$, where $g(a, b)$ is a pixel discrepancy function which can be simply $g(a, b) = \|a - b\|$ or $g(x) = (a - b)^2 / (1 + (a - b)^2 / d^2)$ (Geman-McClure function). Also, let \mathcal{D}_i^a denote the corresponding deformation map for an image \mathcal{I}_i . Then we define the groupwise objective function, F , as follows:

$$F(\mathcal{D}_1^a, \dots, \mathcal{D}_N^a) = \frac{1}{N} \sum_{i=1}^N G \left(W(\mathcal{I}_i, \mathcal{D}_i^a), \frac{1}{N-1} \sum_{\forall j \neq i} W(\mathcal{I}_j, \mathcal{D}_j^a) \right) \quad (1)$$

This amounts to computing the average discrepancy between every image warped to reference space and the current estimate of the reference computed by averaging all other shape normalised images.

From the information-theoretical point of view, an attractive choice for the objective function is the so-called minimal description length (MDL) of the model (set of transformed images) [9, 16]. For implementation on a GPU, however, our discrepancy-based cost function is more suited. However, MDL could be substituted with no loss of generality.

3.4. Optimisation regime

A naïve attempt to minimise F in Eq. (1) would be to exhaustively search in the space of all possible deformation maps, \mathcal{D}_i^a , for all N images. As mentioned in Section 3.1, in practice some reparameterisation of \mathcal{D}_i^a is used, for example using control points defining triangular meshes [2] or splines [16]). Even after such reparameterisation, finding a minimum of our objective function is a difficult task since the number of parameters in the optimisation is still very large (see Section 1).

In [2, 3, 4] it has been shown that it is sufficient to optimise the deformations of one image at a time. This is equivalent to optimising a $2NK$ dimensional objective function along $2K$ dimensions at a time and is analogous to Powell's classic optimisation method. In our method we adopt a similar technique: as we show below, in an iterative fashion, for one image at a time, we improve it's corresponding deformation map which leads to a minimisation of Eq. 1.

We now give the complete description of the registration procedure which is summarised in Alg. 1. The deformation maps $\mathcal{D}_{i_0}^a$ for all images are initialised to identity transform (line 2). The algorithm then operates by incrementally improving the accumulated deformation maps, $\mathcal{D}_{i_k}^a$, in an iterative fashion. The iterative body (lines 4–11) is repeated until no further improvement is possible. In order to avoid biasing the algorithm, we randomise the order in



Figure 3. Evolution of the deformation map for one image as the algorithm progresses.

which the images in the set (and their corresponding deformation maps) are processed: at each iteration we randomly permute the set.

When optimising deformations for each image \mathcal{I}_i , first compute an estimate of the reference image \mathcal{M}_{i_k} by averaging all images \mathcal{I}_j (except \mathcal{I}_i) transformed to the reference space using the deformation learnt at the previous stages $\mathcal{D}_{j_{k-1}}^a$ and the improvement $\Delta\mathcal{D}_{j_k}$ learnt at the current step (line 7). The optimal improvement $\Delta\mathcal{D}_{i_k}$ is then computed using Alg. 2 by minimising $G(\cdot)$ over the space of all possible improvements $\Delta\mathcal{D}$ (line 8). After the deformations of all images have been improved, the improvements are added to the previously learnt deformations (line 10) and the process repeats. The evolution of the deformation map for an image is illustrated if Fig. 3.

We next address the outstanding problem of minimising $G(\mathcal{M}_{i_k}, W(\mathcal{I}_i, \mathcal{D}_{i_{k-1}}^a + \Delta\mathcal{D}))$ in line 8 as a function of $\Delta\mathcal{D}$ which is exacerbated by its non-linear nature and many local minima.

3.5. Stochastic optimisation

Most traditional optimisation algorithms, such as Downhill Simplex [1] and gradient descent [2], cannot efficiently deal with an optimisation problem of the above magnitude and tend to converge to local minima especially in the presence of measurement noise [3]. Some stochastic algorithms, such as simulated annealing and genetic algorithms, have been designed to avoid local minima. However, such techniques tend to be inefficient and have impractical computation times even for small data sets.

Thus, one of the most challenging aspects in the development of optimisation approaches to groupwise registration is designing an efficient minimisation routine [3]. Simultaneous Perturbation Stochastic Approximation (SPSA) is an attractive choice for the optimiser: it is capable of evading local minima due to its stochastic nature and, when adapted for our framework, is orders of magnitude more efficient [13] than the traditional stochastic algorithms. Moreover, while in traditional gradient-based methods the number of function evaluations required to estimate the gradient at a point grows linearly with the dimensionality of the space, SPSA offers independence of the number of function evaluations at each iteration on the dimensionality of the space. An overview of the SPSA algorithm can be found in [15], for completeness we summarise the idea of the al-

Algorithm 1 Register a batch of images

Require: Images $\mathcal{I}_i, i \in \{1 \dots N\}$

- 1: $k \leftarrow 1$
- 2: $\mathcal{D}_{i_0}^a \leftarrow \mathbf{0}, \forall i \in \{1 \dots N\}$
- 3: **while** not happy **do**
- 4: Randomly permute the order of images.
- 5: $\Delta\mathcal{D}_{i_k} \leftarrow \mathbf{0}, \forall i \in \{1 \dots N\}$
- 6: **for** $i = 1$ to N **do**
- 7: $\mathcal{M}_{i_k} \leftarrow \frac{1}{N-1} \sum_{\substack{j \in \{1 \dots N\} \\ j \neq i}} W(\mathcal{I}_j, \mathcal{D}_{j_{k-1}}^a + \Delta\mathcal{D}_{j_k})$
- 8: Using Algorithm 2, compute $\Delta\mathcal{D}_{i_k} \leftarrow \arg \min_{\Delta\mathcal{D}} G(\mathcal{M}_{i_k}, W(\mathcal{I}_i, \mathcal{D}_{i_{k-1}}^a + \Delta\mathcal{D}))$
- 9: **end for**
- 10: Learn improved deformation map: $\mathcal{D}_{i_k}^a \leftarrow \mathcal{D}_{i_{k-1}}^a + \Delta\mathcal{D}_{i_k}$
- 11: $k \leftarrow k + 1$
- 12: **end while**

Algorithm 2 Optimise improvement $\Delta\mathcal{D}$ to deformation $\mathcal{D}_{i_{k-1}}^a$ of \mathcal{I}_i into \mathcal{M}_{i_k} using SPSA [14, 15]

Require: $\mathcal{M}_{i_k}, \mathcal{I}_i, \mathcal{D}_{i_{k-1}}^a, m_{\max}, c_0, a_0, \alpha, \gamma, A$

- 1: $m \leftarrow 1$
- 2: Set gains $a_k \leftarrow \frac{a_0}{(A+m)^\alpha}$ and $c_m \leftarrow \frac{c_0}{m^\gamma}$
- 3: Select randomly C_R and initialise $\hat{\phi}_m \leftarrow C_R$
- 4: **while** not converged and $m < m_{\max}$ **do**
- 5: Generate $\delta_m, \delta_{m_i} \leftarrow \text{Bernoulli}(-1 \text{ or } 1)$
- 6: Using $G(\hat{\phi}_m) := G(W(\mathcal{I}_i, \mathcal{D}_{i_{k-1}}^a + L(\hat{\phi}_m, C_R)), \mathcal{M}_{i_k})$
Estimate $\mathbf{g}_m \leftarrow \mathbf{g}(G(\hat{\phi}_m), \hat{\phi}_m, c_m, \delta_m)$, see Eq. (2)
- 7: Update $\hat{\phi}_{m+1} \leftarrow \hat{\phi}_m - a_m \mathbf{g}_m$, see Eq. (3)
- 8: $m \leftarrow m + 1$
- 9: **end while**
- 10: **return** The optimal $\Delta\mathcal{D} \leftarrow L(\hat{\phi}_m, C_R)$

gorithm here. Let $f(\phi)$ be a real-valued function and ϕ be a p -dimensional vector of parameters. Assume that only the direct measurements of $f(\phi)$ are available, but not of its gradient. Measurements of $f(\phi)$ might also be noisy, and p might be very large. We aim to minimise $f(\phi)$ to find $\underline{\phi} = \arg \min f(\phi)$. Let $\delta = (\delta_1, \delta_2, \dots, \delta_p)^T$ be a vector of independent random variables with symmetric Bernoulli distribution: $\delta_i = \pm 1$ and $\Pr(\delta_i = 1) = \Pr(\delta_i = -1) = 1/2$. Let $\hat{\mathbf{g}}(\phi)$ denote the stochastic approximation of the gradient $\mathbf{g}(\phi)$:

$$\hat{\mathbf{g}}(f(\cdot), \phi, c_k, \delta_k) = \begin{pmatrix} \frac{(f(\hat{\phi}_k + c_k \delta_k) - f(\hat{\phi}_k - c_k \delta_k))}{2c_k \delta_{k_1}} \\ \dots \\ \frac{(f(\hat{\phi}_k + c_k \delta_k) - f(\hat{\phi}_k - c_k \delta_k))}{2c_k \delta_{k_p}} \end{pmatrix} \quad (2)$$

and $\hat{\phi}$ denote the ‘‘current’’ estimate for $\underline{\phi}$.

The SPSA algorithm incrementally updates $\hat{\phi}$ by the following process:

$$\hat{\phi}_{k+1} = \hat{\phi}_k - a_k \hat{g}(\hat{\phi}_k) \quad (3)$$

The gain sequences c_k and a_k are chosen as follows: $a_k = \frac{a_0}{(A+k)^\alpha}$ and $c_k = \frac{c_0}{k^\gamma}$. Note that at each iteration k only *two* $= O(1)$ evaluations of f are required, as opposed to $O(p)$ in traditional gradient-based methods. Extensive convergence theory [10] establishes performance guarantees for SPSA and shows that $\hat{\phi}_k \rightarrow \arg \min f(\phi)$ as $k \rightarrow \infty$.

Alg. 2 summarises our adaptation of SPSA for the task at hand. As stochastic algorithms are notoriously hard to tune, we now address issue of tuning SPSA. In Alg. 2, the tuning parameters c_0 and a_0 are measured in units of image size (say, pixels) and control the “greediness” of the algorithm — larger values correspond to less greedy search. We choose c_0 to be about 1–4% of the image size and experimentally choose a_0 to be of the same order of magnitude. The decay parameters α and γ are set to the theoretically optimal values $\alpha = 0.602$ and $\gamma = 0.101$ as discussed in [14] which also covers the choice of tuning parameters in various settings.

4. Experiments

Here, we describe the experiments we carried out to evaluate the efficiency of our optimisation framework.

In our experiments we have used a subset of 128 images from the publicly available FGNET “talking face” data set [6]. For inter-person registration experiment, we have used frontal images from the publicly available IMM face database [11]. We have also applied our algorithm to brain and face data sets acquired in our lab with similar results, space limitations preclude inclusion here.

As shown in Fig. 4, as the algorithm progresses and incrementally establishes the correct correspondences, the shape-normalised average of the images converges to a true, crisp picture of the underlying structure. This means that all images have been well aligned by the groupwise registration. In Fig. 5, on inter-person data, the average of non-shape normalised (no registration) images does not exhibit much structure but after applying our algorithm we get a good approximation of the average face. There is clearly much variation in the inter-person data set but our algorithm admirably copes with this.

As discussed in [10], SPSA proved to be very suitable for global optimisation. Fig. 6 (left) shows the comparison of our approach with incrementally learning the deformations to the traditional approaches that do not learn. To evaluate the performance we evaluate the objective function at each iteration. SPSA was used as the optimiser in each case here. The FGNET data set was used. It is clear from Fig. 6 that our approach yields an improvement in the speed of convergence and converges to better minima. It is also

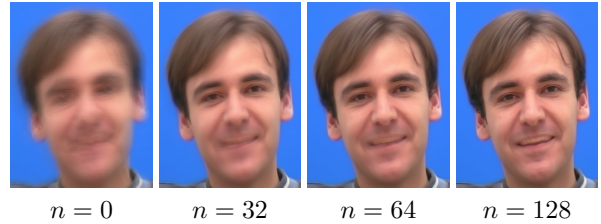


Figure 4. Evolution of the mean shape normalised face. Shown after n iterations of the Alg. 1.

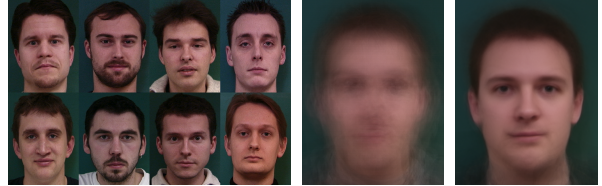


Figure 5. Inter-personal registration. *Left*: Samples data. *Middle*: Averaged images (no registration). *Right*: Averaged shape-normalised images after registration with our algorithm.

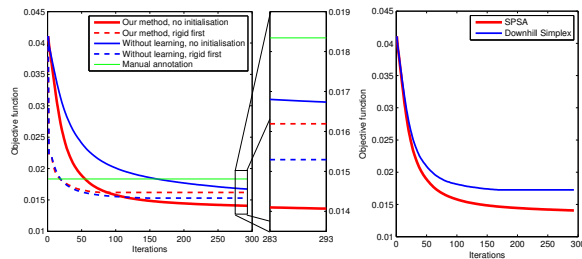


Figure 6. Comparison of algorithm performance. *Left*: With and without incrementally learning the deformations. *Rightmost part* magnified. *Right*: SPSA vs. Downhill Simplex.

important to note that our method requires no initialisation (although conventional approaches benefit from initialisation [2]). For example, the FGNET data set contains a lot of affine deformations (head movement); it helps to initialise the algorithm by performing rigid registration first (dashed lines). In our experiments, after such initialisation the algorithm quicker arrives to the solution but at a cost of minor loss in accuracy: this initialisation is essentially too greedy a search and tends to throw the algorithm into a strong but not best local minima. All approaches beat a hand labelled annotation of the data (Fig. 6 (left) horizontal line). In all experiments our algorithm numerically demonstrates linear convergence. Note, however, that the critical performance cost is the number of objective function evaluation, not the number of iterations, as shown in Section 3.5 — this is an important property and compares favourably with other optimisers. In Fig. 6 (right) we compare SPSA with the commonly used Downhill Simplex method and obtain a better optimal solution with a better convergence rate.

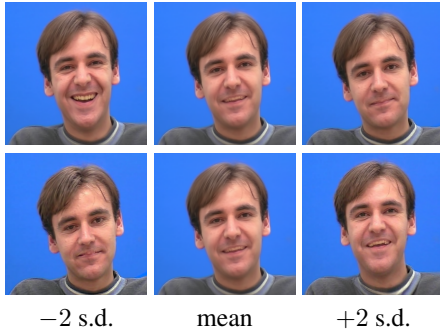


Figure 7. An Active Appearance Model of a talking head.

Having registered the images, we can construct statistical appearance models, using deformation maps directly to build high resolution shape model. If memory is a concern, however, one might obtain traditional control point-based representation of shape in the end by sampling the deformation maps. The first two modes of variation of combined model of the FGNET data set are shown in Fig. 7.

5. Conclusion

We have proposed a novel approach to groupwise non-rigid image registration which requires no initialisation. We have developed methods that implicitly reduce the dimensionality of the search space by representing increasingly complex deformations as a superposition of simpler deformations. Due to this formulation we are able to take advantage of the simplicity and speed of piece-wise linear interpolation to model deformations and overcome previous limitations of this approach due to limited smoothness, flexibility and spatial resolution. We also use a novel efficient and reliable, fully unsupervised stochastic optimiser — an adaptation of SPSA — which is independent of the number of function evaluations at each iteration on the dimensionality of the space.

In evaluation of our method we have demonstrated a high robustness and success rate, fast (linear) convergence on various types of test data which shows considerable improvement in terms of accuracy of solution and speed compared to existing methods. Due to the robustness of approach we are also able to perform *inter-person* registration. This is the first time that the groupwise registration of data possessing such variety has been reported.

Due to the efficient formulation of our approach, it is amenable for GPU implementation — in our experiments, apart from the control logic, all steps are performed on a GPU.

References

- [1] T. F. Cootes, S. Marsland, C. J. Twining, K. Smith, and C. J. Taylor. Groupwise diffeomorphic non-rigid registration for automatic model building. In *Proc. ECCV*, pages 316–327. Springer, 2004.
- [2] T. F. Cootes, C. J. Twining, V. Petrovic, R. Schestowitz, and C. J. Taylor. Groupwise construction of appearance models using piece-wise affine deformations. In *Proceedings of BMVC05*, pages 879–888, 2005.
- [3] R. Davies, C. Twining, and C. Taylor. *Statistical Models of Shape: Optimisation and Evaluation*. Springer Publishing Company, Incorporated, 2008.
- [4] R. H. Davies, C. J. Twining, T. F. Cootes, J. C. Waterton, and C. J. Taylor. A minimum description length approach to statistical shape modeling. *IEEE Transactions on Medical Imaging*, 21:525–537, 2002.
- [5] R. H. Davies, C. J. Twining, and C. Taylor. Groupwise surface correspondence by optimization: Representation and regularization. *Medical Image Analysis*, 12(6):787–796, 2008.
- [6] FGNET. Face and gesture network: Talking face database. http://www-prima.inrialpes.fr/FGnet/data/01-TalkingFace/talking_face.html.
- [7] M. J. Jones and T. Poggio. Multidimensional morphable models. In *6th International Conference on Computer Vision*, pages 683–688, 1998.
- [8] E. G. Learned-Miller. Data driven image models through continuous joint alignment. *PAMI*, 28:236–250, 2006.
- [9] S. Marsland, C. J. Twining, and C. J. Taylor. A minimum description length objective function for groupwise non-rigid image registration. *Image and Vision Computing*, 26(3):333–346, 2008.
- [10] J. L. Maryak and D. C. Chin. Global random optimization by simultaneous perturbation stochastic approximation. *IEEE Transactions on Automatic Control*, 53(3):780–783, 2008.
- [11] M. M. Nordstrøm, M. Larsen, J. Sierakowski, and M. B. Stegmann. The IMM face database — an annotated dataset of 240 face images. Technical report, Informatics and Mathematical Modelling, Technical University of Denmark, DTU, may 2004.
- [12] V. Petrovic, T. Cootes, A. Mills, C. Twining, and C. Taylor. Automated analysis of deformable structure in groups of images. In *BMVC07*, pages 302–311, 2007.
- [13] J. Spall, S. Hill, and D. Stark. Theoretical framework for comparing several stochastic optimization approaches. In *Probabilistic and Randomized Methods for Design under Uncertainty*, pages 99–117. Springer, 2006.
- [14] J. C. Spall. Implementation of the simultaneous perturbation algorithm for stochastic optimization. *IEEE Transactions on Aerospace Electronic Systems*, 34:817–823, July 1998.
- [15] J. C. Spall. An overview of the simultaneous perturbation method for efficient optimization. *Johns Hopkins APL Technical Digest*, 19(4), 1998.
- [16] C. Twining, S. Marsland, and C. Taylor. Groupwise non-rigid registration of medical images: The minimum description length approach. In *Medical Image Analysis and Understanding*, pages 81–84, 2004.
- [17] B. Zitova and J. Flusser. Image registration methods: a survey. *Image and Vision Computing*, 21(11):977–1000, October 2003.