# Construction of a Spatio-Temporal Face Atlas: Experiments Using Down Syndrome Samples

I. Xavier[1], G. Giraldi[2], S. Gibson[3], G. J. F. Gattas[4], D. Rueckert[5] and C. Thomaz[1]
[1]Department of Eletrical Engineering, FEI, Sao Paulo, Brazil
[2]Department of Computer Science, LNCC, Rio de Janeiro, Brazil
[3]School of Physical Sciences, University of Kent, UK
[4]Department of Forensic Medicine, USP, Sao Paulo, Brazil
[5]Department of Computing, Imperial College London, UK

***Abstract*—This work describes a method for constructing a spatio-temporal face atlas based on pairwise non-rigid registration and kernel regression in order to explicitly consider the craniofacial differences that could exist across the time due to the subjects growth. The whole pipeline includes the following steps: (a) Facial Landmarks Positioning; (b) Sample Group Equalization and Kernel Interpolation; (c) Registration; (d) Atlas Construction. Our experiments on Down Syndrome face images show that the method generates realistic and unbiased face atlases from toddlers to teens.**

## I. Introduction

Face images have been used in a wide range of applications in the last years. We can list face recognition, facial expression analysis and facial composites as the most common ones. However, many approaches in these applications have not explicitly considered the craniofacial differences that could exist across the time due to the subjects growth [1], particularly in the first stages of human life.

Although faces are expected to have a global and common spatial layout with all its parts such eyes, nose and mouth arranged consistently, specific variations in these local features are fundamental to explain our perception of each individual singularity or samples of individuals [2]. These aspects of the face space have already been explored to analyze, for example, the gender and facial expression most discriminant and expressive features [3] of adults, as well as for diagnosing genetic diseases in children [4], but, to the best of our knowledge, have not explicitly explored the variation in these features across the time. In fact, a spatio-temporal face atlas can play the same role of the well-known human brain atlases [5] and can be explored to investigate shared grouping characteristics like race, gender, ageing, with potential application in clinical face phenotype and forensic imaging, for instance.

In this work, we have described and implemented a framework to construct spatio-temporal face atlases. Based on the approach proposed in the context of medical image [5], our method uses non-rigid point registration and face averaging to produce artificial and realistic face atlases taking into account temporal information. Differently from [4], which has used an unique face children template, we have produced, as main contribution, a different typical face image for each chosen time-interval.

The remainder of this paper is organised as follows. In section II, we describe the face database used to carried out the experiments and explain the method to build an spatio-temporal face atlas. All the experimental results have been shown in section III. Finally, in section IV, we evaluate the method and discuss further improvements.

## II. Materials and Methods

The spatio-temporal face atlas has been built by performing the following steps: face landmarking, sample group equalization, non-rigid registration (or equivalently non-linear spatial normalisation) and atlas construction by image kernel averaging. These techniques have been previously applied to facial composites [6], medical image registration [7] and age-dependent spatio-temporal brain atlas construction [5].

### A. Face Database

Frontal 2D face images, acquired as part of a research initiative to address the issue of missing children and adolescents with disabilities in Brazil [8], have been used to carry out the experiments. The subset used in this work is composed of 62 face images of Down Syndrome children (35 male subjects and 27 female subjects) with the age range from 5 months to 18 years old. The age histogram of this database can be seen in Figure 1. All images have been taken in an upright frontal position with some variation in scale. We have rigidly registered all these 62 face images previously, using the positions of the eyes as a measure of reference, so that the pixel-wise features extracted from the images correspond roughly to the same location across all subjects. All the images are encoded in gray-scale using 8-bits per pixel. Figure 2 illustrates some of these rigidly registered samples used in this work.

### B. Facial Landmarks Positioning

To spatially normalize all the frontal face images we need to identify a set of fiducial points or landmarks on each face. We have used the same approach applied in [3] to semi-automatically annotate these samples.

### C. Sample Group Equalization

To handle the problem of distinct number of samples for each time-interval, we have implemented a slightly different version of the algorithm proposed in [5]. Such algorithm, summarized below, attempts to solve two problems at the same

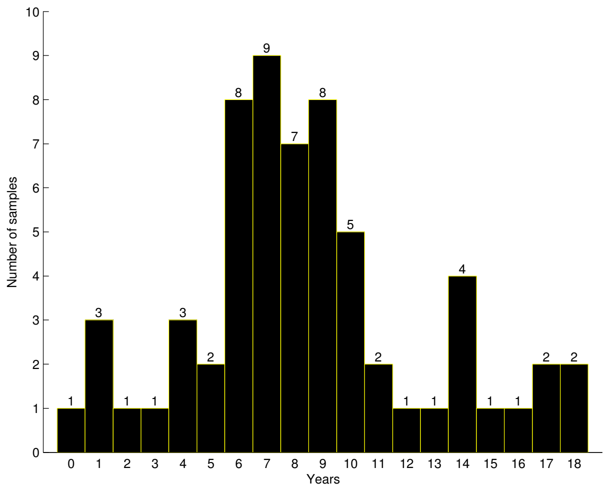


Fig. 1. Age histogram in years of the 62 subjects of the Down Syndrome children samples.

Fig. 2. A sample of the 2D frontal face images used in this work.

time, that is, equalizing the number of subjects among all time-intervals and estimating the kernel parameter $\sigma$ of each time-interval. These computed kernel parameters have been used to define the contribution of neighbors subjects in the atlas construction of each time-interval $t$.

Set $\bar{n}$ to the median of the number of samples of each time-interval $t$
Set $k \leftarrow 1$ as the tolerance value for the number of subjects.
Set $\hat{f}$ to the experimental increasing factor of $\sigma_t$.
**for all** time-interval $t$ **do**
  Set $n_t$ to represent the number of subjects at time-interval $t$
  Set $\sigma_t \leftarrow 1$ to represent the kernel size at time-interval $t$
  Set $\phi_t$ to represent the set of subjects at time-interval $t$
  **if** $n_t < \bar{n} - k$ **then**
    **while** $n_t < \bar{n} - k$ **do**
      Add to $\phi_t$ from the original set of $t-1$ time-interval
      Increment $n_t$
      Add to $\phi_t$ from the original set of $t+1$ time-interval
      Increment $n_t$
      $\sigma_t \leftarrow \sigma_t \cdot \hat{f}$
    **end while**
  **end if**
**end for**

Fig. 3. Sample group equalization algorithm.

### *D. Registration*

The registration or pre-processing of the data is an important step for any pattern recognition analysis. The purpose of the spatial normalisation stage is to remove any confounding effects from the data that are not relevant for the analysis. In the context of the face analysis this means that all images need to be mapped to a common coordinate system so that the pixel-wise features extracted from the images correspond ideally to the same anatomical location across all subjects [9].

We have used a 2D version of the free-form deformation (FFD) [10], [11] algorithm proposed by Rueckert et al. [7] to non-rigidly register all the face images. This algorithm was originally proposed to register 3D contrast-enhanced breast magnetic resonance images in order to minimize the non-rigid shape changes due to the patient breathing and motions, and, more recently, has been also applied to spatial normalisation of face images [12], [13], [3].

The estimation of the deformation fields can be briefly described as follows. Let

$$\Omega_{source} = \{(x, y) | 0 \leq x < h_1, 0 \leq y < h_2\} \qquad (1)$$

be a set of points representing the input image, where $h_1$ and $h_2$ are respectively the width and height in pixels of the image. Let $\Phi_{source}$ be an $n_x \times n_y$ lattice of $\phi_{xy}$ control points overlaid on $\Omega_{source}$, where $1 \leq n_x \leq h_1$ and $1 \leq n_y \leq h_2$.

The goal here is to minimize the misalignment of the 260 landmarks from the source image to the 260 landmarks of a target image. We have used the squared distance of the landmarks as the alignment criterion.

To calculate the new position of a point at the location $(x, y)$, we have applied the 2D tensor product of B-splines considering the sixteen control points in its neighbourhood [11], that is

$$\mathbf{w}(x, y) = \sum_{k=0}^{3} \sum_{l=0}^{3} B_k(s) B_l(t) \phi_{(i+k)(j+l)}, \qquad (2)$$

where $i = \lfloor x/n_x \rfloor - 1$, $j = \lfloor y/n_y \rfloor - 1$, $s = x - \lfloor x/n_x \rfloor$ e $t = y - \lfloor y/n_y \rfloor$. $B_k(s)$ and $B_l(t)$ correspond to the uniform cubic B-spline basis functions evaluated at $s$ and $t$ [11], respectively, and defined as:

$$\begin{aligned}
B_0(u) &= (-u^3 + 3u^2 - 3u + 1)/6 \\
B_1(u) &= (3u^3 - 6u^2 + 4)/6 \\
B_2(u) &= (-u^3 + 3u^2 + 3u + 1)/6 \\
B_3(u) &= u^3/6,
\end{aligned}$$

where $0 \leq u < 1$.

Analogously to [7], the algorithm initially performs an affine transformation to align the landmarks of the $\Omega_{source}$ to the same landmarks of the target image. Then the procedure runs an iterative coarse-to-fine deformation, displacing the control points of the lattice $\Phi_{source}$ until it achieves the alignment criterion. This has been repeated until the finest level of the FFD resolution is computed, estimating the final deformation fields for each source face image.

### E. Atlas Construction

The method used in this study to build an spatio-temporal face atlas is analogous to [5]. It performs pairwise registrations among all the subjects within a time-interval and, in the end, averages the transformed face images to compose the spatio-temporal face atlas. For each time-interval, we carry out this pairwise registration by, in turn, selecting one image as source and the rest of the images as targets, generating a set of transformations. These transformations are averaged and applied to the current source image. Finally, we compute the estimated atlas by averaging the transformed source images. This technique attempts to eliminate the bias toward any original subject in the atlas composition. Figure 4 pictures the idea of this spatio-temporal atlas construction.

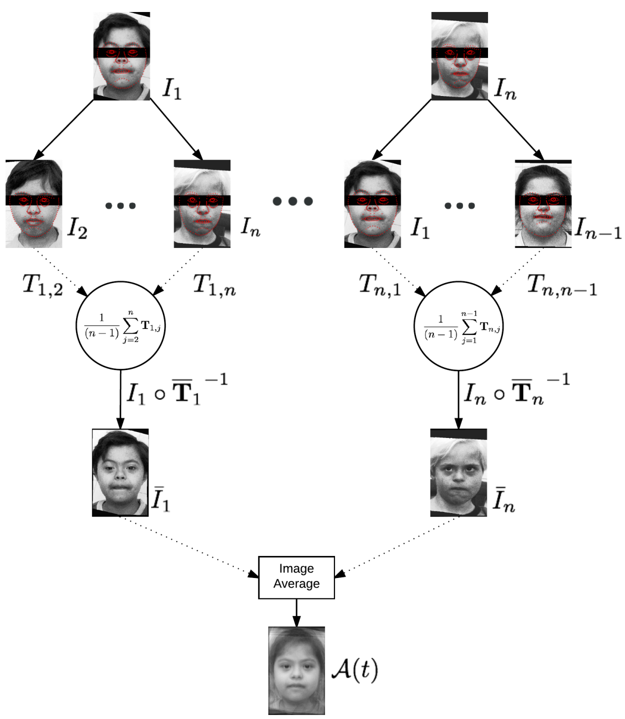


Fig. 4. The pairwise point based registration process for construction of artificial face atlas. All the original images have been masked (with black strips) to preserve the identity of the subjects.

An extended approach in the atlas construction is to calibrate the contribution of the neighbours included in the time-interval, using an adaptive kernel regression [5]. The kernel not only serves to estimate the weighted support given by neighbours, but, it could be used to interpolate between the subjects when there is no subjects at the exact age of interest. Let the weight assigned to the $i^{\text{th}}$ subject at the time $t$ be given by a Gaussian kernel:

$$g(t_i, t) = \frac{1}{\sigma_t\sqrt{2\pi}} e^{\frac{-(t_1-t)^2}{2\sigma_t{}^2}}, \tag{3}$$

where $t_i$ denotes the age of the subject $i$ in years.

Let $I_1, \ldots, I_n$ represent the images for all subjects within a time-interval. For each image $I_i$, selected as a source image, the procedure of Figure 4 generates transformations $\mathbf{T}_{i,j}$ for $j = 1, \ldots, n, j \neq i$, which are averaged out to produce $\overline{\mathbf{T}}_i$:

$$\overline{\mathbf{T}}_i = \frac{1}{(n-1)} \sum_{j=1}^{n} \mathbf{T}_{i,j} \tag{4}$$

for each image $i$ at a given time-interval. The mean image $\bar{I}_i$ is defined as the image obtained from $I_i$ spatially transformed by $\overline{\mathbf{T}}_i$:

$$\bar{I}_i = I_i \circ \overline{\mathbf{T}}_i^{-1} \tag{5}$$

The spatio-temporal face atlas can be estimated as:

$$\mathcal{A}(t) = \frac{\sum_{j=1}^{n} g(t_j, t)\bar{I}_j}{\sum_{j=1}^{n} g(t_j, t)} \tag{6}$$

## III. RESULTS

We have applied the aforementioned algorithm on the image database described on section II-A in order to equalize the number of samples for each time-interval. Figure 5 shows the result of the sample equalisation. Most time-interval groups of images have changed by adding samples from their neighbours, while Gaussian kernels have been estimated. These kernels have been plotted over each corresponding bar. The parameter $\sigma$ depends on how many samples have been taken from the neighbours. Figure 5 shows these generated Gaussian curves as well.

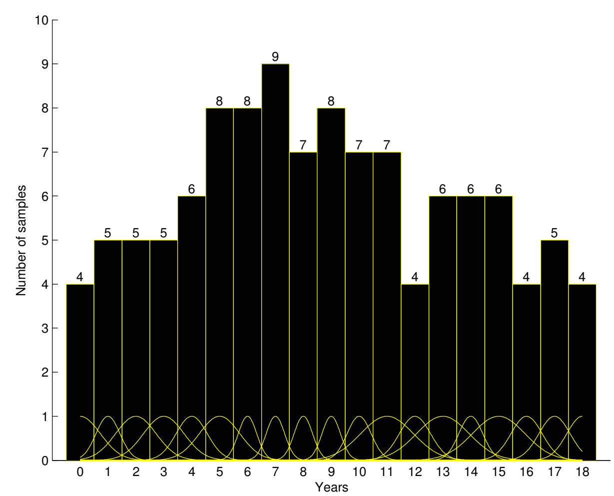


Fig. 5. The equalized time-interval group histogram after processing the original samples.

We have evaluated the method by building three types of spatio-temporal face atlases. Firstly, we just have taken the original database and for each time-interval group of images with more than two samples we have applied the registration procedure described in Figure 4. The result can be seen in the top row of Figure 6. Then, we have applied the latter approach for the equalized database generated. The results can be visualized in the middle row of Figure 6. Finally, we have computed the atlas by applying the equation (6) to average out the images in each time-interval ($\hat{f} = 1.3$). The atlas

Fig. 6. The spatio-temporal face atlases artificially constructed. From top to bottom: original face database, equalized version of the database, and equalized version with the neighbors contribution weighted by an estimated Gaussian kernel. From left to right, average images from 0 to 18 years old of the registered faces representing the atlas of each age.

generated by the weighted kernel approach can be visualized in the bottom row of Figure 6.

We can see that the equalization procedure has filled the lack of original database by increasing the number of samples. More interestingly, comparing the middle row with the bottom row of Figure 6, we can notice less artifacts in the faces region, which can be explained by the reduced contribution from the neighbors to the atlas composition.

## IV. Conclusion

In this work, we presented a method for constructing a spatio-temporal face atlas. Despite the fact that we have used Down Syndrome face images only to carry out the experiments, the method is not restricted to any particular set of samples and can be applied to construct a spatio-temporal face atlas for any face samples of interest. In this study, we have used very few samples indeed. Even so, the method has produced realistic artificial face images, each of them corresponding to an unbiased face atlas of the specific age under investigation. The experiments have shown that the kernel based approach produced a more detailed and less confounding effect version of the atlas and the equalization algorithm has overcome the lack of face images in some time intervals. Further work would be necessary to fully automate this framework. To do so, an automatic landmarking processing step will be evaluated. Additionally, the use of larger sample groups would definitely enhance shared sample group characteristics like age, gender or race from ordinary photos.

## Acknowledgements

The authors would like to thank Dr. Andreas Shuh from Imperial College London, UK, who has provided valuable support to adapt the MIRTK tools[1] to 2D face image registration. Also, we would like to thank the financial support provided by FAPESP (2009/53556-0 and 2012/22377-6) and Royal Society (NA140147).

## References

[1] N. Ramanathan and R. Chellappa, “Modeling age progression in young faces,” in *2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR’06)*, vol. 1. IEEE, 2006, pp. 387–394.

[2] C. E. Thomaz, V. Amaral, D. F. Gillies, and D. Rueckert, “Priori-driven dimensions of face-space: experiments incorporating eye-tracking information,” in *Proceedings of the Ninth Biennial ACM Symposium on Eye Tracking Research & Applications*. ACM, 2016, pp. 279–282.

[3] I. Xavier, M. Pereira, G. Giraldi, S. Gibson, C. Solomon, D. Rueckert, D. Gillies, and C. Thomaz, “A photo-realistic generator of most expressive and discriminant changes in 2d face images,” in *2015 Sixth International Conference on Emerging Security Technologies (EST)*. IEEE, 2015, pp. 80–85.

[4] Q. Ferry, J. Steinberg, C. Webber, D. R. FitzPatrick, C. P. Ponting, A. Zisserman, and C. Nellåker, “Diagnostically relevant facial gestalt information from ordinary photos,” *Elife*, vol. 3, p. e02020, 2014.

[5] A. Serag, P. Aljabar, G. Ball, S. J. Counsell, J. P. Boardman, M. A. Rutherford, A. D. Edwards, J. V. Hajnal, and D. Rueckert, “Construction of a consistent high-definition spatio-temporal atlas of the developing brain using adaptive kernel regression,” *NeuroImage*, vol. 59, no. 3, pp. 2255–2265, 2012.

[6] S. J. Gibson, “Eigenfit: A statistical learning approach to facial composites,” Ph.D. dissertation, University of Kent, 2006.

[7] D. Rueckert, L. I. Sonoda, C. Hayes, D. L. Hill, M. O. Leach, and D. J. Hawkes, “Nonrigid registration using free-form deformations: application to breast mr images,” *IEEE Transactions on Medical Imaging*, vol. 18(8), pp. 712–721, 1999.

[8] G. J. F. Gattas, C. Figaro-Garcia, C. E. Thomaz, R. Cardinali, V. Amaral, M. Pereira, F. T. Goncalves, V. L. R. Alves, A. C. M. Slimene, C. Fridman, and L. R. Battistella, “The risk of missing children and adolescents with disabilities: a brazilian experience,” in *International Conference on Forensic Research & Technology*, vol. 3. Chicago. J.Forensic Res, 2012, pp. 71–71.

[9] C. E. Thomaz, J. P. Boardman, S. Counsell, D. L. Hill, J. V. Hajnal, A. D. Edwards, M. A. Rutherford, D. F. Gillies, and D. Rueckert, “A multivariate statistical analysis of the developing human brain in preterm infants,” *Image and Vision Computing*, vol. 25, no. 6, pp. 981–994, 2007.

[10] T. W. Sederberg and S. R. Parry, “Free-form deformation of solid geometric models,” in *Proceedings of SIGGRAPH’86*. Computer Graphics, 1986, pp. 151–160.

[11] S. Lee, G. Wolberg, K. Chwa, and S. Y. Shin, “Image metamorphosis with scattered feature constraints,” *IEEE Transactions on Visualization and Computer Graphics*, vol. 2(4), pp. 337–354, 1996.

[12] S. Koelstra and M. Pantic, “Non-rigid registration using free-form deformations for recognition of facial actions and their temporal dynamics,” in *Automatic Face & Gesture Recognition, 2008. FG’08. 8th IEEE International Conference on*. IEEE, 2008, pp. 1–8.

[13] S. Koelstra, M. Pantic, and I. Patras, “A dynamic texture-based approach to recognition of facial actions and their temporal models,” *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 32, no. 11, pp. 1940–1954, 2010.

[1] https://mirtk.github.io