

Plan de Trabajo Final

Carrera Ingeniería de Sistemas

Facultad de Ciencias Exactas – UNICEN

Tema: Herramienta para un análisis robusto de la forma de objetos representados por matrices de landmarks en 3D

Alumno: Lucas Marquez

Director: Ing. Viviana Ferraggine

Codirector: Mg. Sebastián Torcida

1. Introducción

La forma puede definirse matemáticamente como aquella información de un objeto o estructura que es invariante por traslaciones, cambios de escala uniformes y rotaciones/reflexiones. Una de las maneras de registrar la forma de un objeto en 3D es utilizar las *coordenadas cartesianas* de una colección de puntos principales, denominados *landmarks*. La colección de landmarks que representa la forma de un objeto determinado se denomina *configuración*.

La *morfometría* es la disciplina que se ocupa del análisis cuantitativo de la forma de objetos, utilizando herramientas tanto geométricas como matemáticas. En particular, el formalismo matemático necesario para la descripción de una forma biológica se ha desarrollado básicamente en los últimos treinta años y de manera gradual. A partir de la llamada “revolución morfométrica de los 90” surgió como disciplina autónoma el análisis estadístico de la forma (en inglés, *statistical shape analysis*).

Para poder comparar la forma de dos o más configuraciones de landmarks, éstas deben ser inicialmente superpuestas. La denominada superposición Procrustes (SP) determina y aplica una traslación, un cambio de escala uniforme y una rotación/reflexión sobre cada una de las configuraciones para darles una posición, un tamaño y una orientación similar, respectivamente [3]. Así, luego de una SP las configuraciones quedarán perfectamente superpuestas si sus formas son idénticas, y en caso contrario las diferencias de forma existentes entre ellas quedarán a la vista.

2. Motivación

Para obtener la SP la mayoría de las herramientas de software utilizan el criterio de *cuadrados mínimos* (CM); la comparación de *dos* configuraciones se denomina *Análisis de Procrustes Ordinario* [3][6][7], y se distingue del *Análisis de Procrustes Generalizado* [3] que alude a la comparación de la forma de *tres o más* configuraciones utilizando una forma pivote o “consenso”.

Por su formulación matemática -su valor óptimo es un *promedio*- los métodos del tipo CM son altamente sensibles a la presencia de valores atípicos ó *outliers* en los datos, producidos por ej. a errores de medición o ruido. Por este motivo existe un enfoque alternativo conocido como la superposición de Procrustes Robusta [7][8][9][10], que utiliza la técnica de las *medianas repetidas* logrando dos ventajas fundamentales:

- limitar el efecto producido por la presencia de los valores atípicos;
- si dos configuraciones difieren en hasta la mitad menos uno de los landmarks, los landmarks en los que no hay diferencias son superpuestos perfectamente.

3. Objetivos

Para promover que los integrantes de la comunidad científica utilicen en sus diversos campos de aplicación los novedosos algoritmos [10] que permiten realizar un Análisis de Procrustes Robusto (Ordinario y Generalizado) en 3D, resulta conveniente contar con un software que los implemente de manera sencilla y flexible; este software estará disponible en forma libre a través de la web.

El software resultante podrá ser total o parcialmente original; en este último caso, la originalidad estará dada por la adaptación y mejora de alguna herramienta similar existente y de código abierto. La herramienta, autocontenida, deberá incluir además un par de capacidades gráficas también basadas en técnicas robustas [1][9], que permitan visualizar los resultados de las superposiciones.

En función de lo anterior, el objetivo principal de este proyecto es desarrollar o adaptar un software que permita realizar **un análisis de forma robusto en 3D a partir de datos de landmarks**, implementando:

- el algoritmo del Análisis de Procrustes Robusto Ordinario (para la comparación de 2 configuraciones) [10];
- el algoritmo del Análisis de Procrustes Robusto Generalizado (para la comparación de 3 ó más configuraciones) [10];

- algunas capacidades gráficas útiles para la visualización de resultados: robust MDS [1] y non-parametric MDS [9].

4. Cronograma de actividades

El trabajo de tesis se desarrollará en las siguientes etapas:

- **Análisis de la bibliografía y de los algoritmos**

En esta primera etapa se analizarán tanto la bibliografía que da el marco de referencia al proyecto como los algoritmos que se implementarán y las herramientas para la visualización de la información

- **Análisis de herramientas de software:**

En esta etapa se realizará un relevamiento y un análisis de las herramientas actualmente disponibles para realizar un Análisis Procrustes basado en cuadrados mínimos, focalizando especialmente las de código abierto.

- **Análisis y definición de las capacidades de la herramienta:**

En esta etapa se analizarán y se definirán los requerimientos funcionales y no funcionales de la herramienta, involucrando principalmente a las capacidades gráficas.

- **Diseño y desarrollo de la herramienta:**

En esta etapa se realizará el diseño y el desarrollo de los algoritmos y de las capacidades gráficas de visualización. Se contemplarán además alternativas para la captura y la exportación de datos y resultados.

- **Test de la herramienta:**

En esta etapa se realizará el test de la herramienta y se evaluarán sus resultados. Si bien a lo largo de todo el proceso de desarrollo se realizarán pruebas y test individuales, en esta última etapa se realizará la prueba de integración completa.

5. Bibliografía

1. Agarwal, A., Phillips, J. M., & Venkatasubramanian, S. (2010). *Universal multidimensional scaling*. In KDD'10: Proceedings of the 16th ACM SIGKDD international conference on knowledge discovery and data mining (pp. 1149–1158). New York: ACM.
2. Bookstein, F. L. (1991). *Morphometric tools for landmark data: Geometry and biology*. New York: Cambridge University Press.
3. Gower, J. C. (1975). *Generalized procrustes analysis*. *Psychometrika*, 40(1), 33–51.
4. Hampel, F. R., Ronchetti, E. M., Rousseeuw, P. J., & Stahel, W. A. (1986). *Robust statistics: The approach based on influence functions*. New York: Wiley.
5. Rohlf, F. J. (1990). *Rotational fit (Procrustes) methods*. In F.J. Rohlf et al., (Eds.) *Proceedings Michigan morphometrics workshop* (pp. 227–236). Special publication no. 2, Museum of Zoology, Michigan, University of Michigan.
6. Rohlf, F. J., & Slice, D. E. (1990). *Extensions of the Procrustes method for the optimal superimposition of landmarks*. *Systematic Zoology*, 39(1), 40–59.
7. Siegel, A. F., & Benson, R. H. (1982). *A robust comparison of biological shapes*. *Biometrics*, 38, 341–350.
8. Slice, D. E. (1996). *Three-dimensional generalized resistant fitting and the comparison of least-squares and resistant fit residuals*. In F. Marcus et al., (Eds.), *Advances in morphometrics* (pp.179–199). New York, Plenum Press.
9. Taguchi, Y. H., & Oono, Y. (2004). *Novel non-metric MDS algorithm with confidence level test*. <http://www.granular.com/MDS/src/paper.pdf>
10. Torcida, S., Perez, S. I. & Gonzalez, P. N. (2013). *An Integrated Approach for Landmark-Based Resistant Shape Analysis in 3D*. *Evol Biol*. DOI 10.1007/s11692-013-9264-1