

# Enfoque basado en optimización heurística y estimación de homografía para el estudio de interfaces adaptativas

Approach based on heuristic optimization and homography estimation for the study of adaptive interfaces

Rafael Ángel Chacón-Fallas<sup>1</sup>

---

Chacón-Fallas, R.A. Enfoque basado en optimización heurística y estimación de homografía para el estudio de interfaces adaptativas. *Tecnología en Marcha*. Edición especial. Movilidad Estudiantil 6, 2019. Pág 47-57

 <https://doi.org/10.18845/tm.v32i8.4563>



<sup>1</sup> Estudiante del Área Académica de Ingeniería en Computadores. Instituto Tecnológico de Costa Rica. Cartago Costa Rica. Correo electrónico: rafangel1228@gmail.com



## Palabras clave

Optimización por enjambre de partículas; modelos de sustitución; homografía; interfaces adaptativas.

## Resumen

Como parte del estudio de los efectos de los sistemas con interfaces adaptativas en el desempeño de los usuarios, se propone el desarrollo de un sistema que permita determinar las alteraciones en la concentración y la atención, a la vez que se plantean soluciones a limitaciones técnicas presentadas en implementaciones anteriores del proyecto.

Concretamente, se propone la utilización de un algoritmo de optimización por enjambre de partículas en lugar de un algoritmo genético, así como la incorporación de un modelo de sustitución que permita aproximar la respuesta del usuario, logrando con ello disminuir la necesidad de evaluar directamente cada configuración de la interfaz y permitir así el uso del paralelismo en etapas futuras.

Para la detectar la concentración del usuario, se propone un uso novedoso a la técnica de estimación de homografía, para la detección del rostro y la orientación de la cabeza. Esto debido a la robustez de sus resultados preliminares en comparación a otras técnicas evaluadas, y a la necesidad de una solución de bajo costo, que evite el uso de equipo especializado para el rastreo de la mirada.

Se presentan los resultados de las diferentes pruebas realizadas al sistema y se plantean una serie de modificaciones o sugerencias para iteraciones futuras del proyecto.

## Keywords

Particle swarm optimization; surrogate models; homography estimation; adaptive interfaces.

## Abstract

A system is proposed to determine the alterations in concentration and attention, as part of the study of the effects on user performance when using systems with adaptive interfaces. At the same time, solutions are proposed to avoid technical limitations that were present in previous implementations of the project.

Specifically, it is proposed to use a particle swarm optimization algorithm instead of a genetic algorithm, as well as to incorporate a surrogate model that allows the user's response to be approximated, thereby reducing the need to directly evaluate each configuration of the interface with the user and allowing parallelism.

To detect the concentration of the user, a novel use to the homography estimation is proposed for the detection of the face and the orientation of the head, due to the robustness of its preliminary results compared to other techniques evaluated, as well as the need for a low-cost solution that prevented the use of specialized equipment in gaze tracking.

The results of the different system tests are presented and a series of modifications or suggestions are proposed for future iterations of the project.

## Introducción

Las interfaces de usuario adaptativas son un campo de estudio que ha tomado fuerza conforme los avances tecnológicos han ido minimizando sus desventajas respecto a capacidad de procesamiento y de almacenamiento necesarias. Si bien, parte de este campo incluye a funcionalidades simples, como el ajuste realizado al rotar las pantallas de los dispositivos móviles, se plantea que para poder considerar un sistema como adaptativo, este debe mejorar su interacción con el usuario, tomando en cuenta tanto las experiencias pasadas como las presentes del uso del sistema por un individuo específico [1].

La clave para el éxito de estos sistemas radica en cómo analizan los datos de los distintos dispositivos y sensores, para determinar qué componentes de la interfaz deben ser modificados en cada momento. Se han utilizado algoritmos genéticos para este fin [2], y se han encontrado resultados prometedores en cuanto a los beneficios de este tipo de interfaces en el desempeño de los usuarios al realizar una tarea específica en el sistema, así como un incremento en la activación psicofisiológica del individuo, comparado con la versión estándar o de interfaz estática. Sin embargo, también se encontraron limitaciones que impiden explotar al máximo las capacidades de los algoritmos genéticos, debido a la necesidad de evaluar individualmente cada posible solución directamente con el usuario.

Son necesarias más investigaciones que permitan: respaldar los resultados sobre mejoras en el desempeño de los usuarios al utilizar interfaces adaptativas, realizar cambios que permitan aprovechar al máximo las capacidades de los algoritmos de optimización utilizados y concluir respecto a otras posibles implicaciones que las interfaces adaptativas tengan en el usuario, por ejemplo, en su concentración y atención.

Con respecto al presente trabajo, se presenta como objetivo principal: proponer un marco de trabajo basado en un algoritmo heurístico de optimización y que permita realizar una prueba de concepto, para la evaluación de las implicaciones en la concentración y la atención de los usuarios al utilizar interfaces adaptativas.

## Marco teórico

En esta sección se detallan brevemente las técnicas utilizadas, esto para una mejor comprensión del documento.

### Estimación de homografía

La homografía es un mapeo lineal de un punto a un punto con ciertos grados de libertad,

$$x' = Hx$$

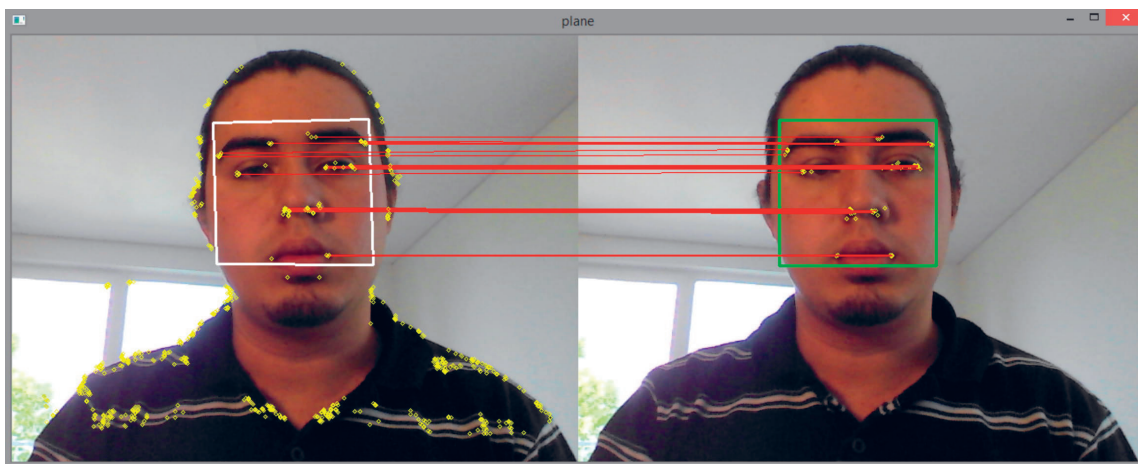
Donde H corresponde a la homografía. Por ejemplo, en imágenes se utiliza una matriz de 3x3, por lo que de la ecuación anterior se obtiene:

$$\begin{bmatrix} x'_1 \\ x'_2 \\ x'_3 \end{bmatrix} = \begin{bmatrix} h_{11} & h_{12} & h_{13} \\ h_{21} & h_{22} & h_{23} \\ h_{31} & h_{32} & h_{33} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ 1 \end{bmatrix}$$

Con  $(x_1, x_2)$  como la posición en la imagen de entrada y  $\begin{pmatrix} x'_1 & x'_2 \\ x'_3 & x'_3 \end{pmatrix}$  como la posición en la imagen de salida. La estimación de la homografía  $H$  entre dos imágenes es un proceso que involucra una cadena de 4 pasos [3]:

- I. Detectar puntos de interés en ambas imágenes.
- II. Calcular descriptores para cada punto de interés.
- III. Encontrar correspondencias entre puntos de interés basadas en los descriptores.
- IV. Estimar la homografía con algún método estadísticamente robusto, sobre las correspondencias detectadas.

En la figura 1 se muestra la premisa planteada para el uso de la estimación de homografía como técnica de reconocimiento facial. Para ello se toma una imagen de referencia (derecha) y se selecciona el rostro como zona de interés (rectángulo verde). Se plantea que la cantidad de puntos de interés (amarillos) entre las imágenes, solo decaerá cuando la persona gire su cabeza al mirar lejos de la pantalla, pero que se mantendrá alto al acercarse o alejarse de la misma. Esto debido a que el rostro se comportará como una imagen planar de 2 dimensiones para movimientos perpendiculares a la pantalla, pero no así para las otras rotaciones.



**Figura 1.** Estimación de homografía para el reconocimiento facial.

### Optimización por enjambre de partículas

Así como los algoritmos genéticos se basan en la teoría de la evolución de Darwin, la optimización por enjambre de partículas (OEP) se basa en la inteligencia colectiva y el comportamiento social de bandadas y escuelas de peces [4]. Cada partícula posee la información de su óptimo local y el enjambre posee un óptimo global que es comparado con los valores locales en cada iteración. El algoritmo básico de OEP consta de tres pasos: la generación de posiciones y velocidades de las partículas, la actualización de la velocidad y, finalmente, la actualización de la posición [5]. Para la generación de las posiciones y velocidades iniciales se generan valores aleatorios entre los rangos máximos y mínimos establecidos para cada variable.

Para la actualización de la velocidad se utiliza la fórmula [5]

$$v_{k+1}^i = wv_k^i + c_1 \text{rand} \left( \frac{(p^i - x_k^i)}{\Delta t} \right) + c_2 \text{rand} \left( \frac{(P_k^g - x_k^i)}{\Delta t} \right)$$

en la que se define la velocidad  $v$  de la partícula  $k+1$  en el tiempo como la suma de tres valores: movimiento actual (compuesto por el factor de inercia  $w$  y la velocidad actual de la partícula  $v_k^i$ ), influencia de la memoria de la partícula (compuesto por el factor de auto confianza  $c_1$  y un valor aleatorio dependiente del óptimo local  $p^i$  y de la posición actual  $x_k^i$ ) e influencia del enjambre (compuesto por el factor de confianza de enjambre  $c_2$  y un valor aleatorio dependiente del óptimo global  $P_k^g$  y de la posición actual  $x_k^i$ ).

Para la actualización de la posición se utiliza [5]

$$x_{k+1}^i = x_k^i + v_{k+1}^i \Delta t$$

en la que se define la posición  $x$  de la partícula  $i$  en el tiempo  $k+1$ , como la posición actual  $x_k^i$  sumado al producto de la velocidad en el siguiente instante de tiempo  $v_{k+1}^i$  y el diferencial de tiempo.

### Modelos de sustitución

Los modelos de sustitución permiten aproximar el resultado de experimentos o simulaciones que serían sumamente costosas de obtener de manera directa. El método consiste en tomar una serie de datos y, a partir de ellos, encontrar una función de aproximación, que permita predecir el comportamiento del problema de estudio sin la necesidad de evaluar directamente.

Las funciones de base radial (FBR) son aquellas cuyo valor solo depende de la distancia del punto al origen y se basan en técnicas para realizar interpolaciones exactas de un conjunto de datos o puntos en un espacio multi-dimensional [6]. Existen varios tipos de funciones de base radial, por ejemplo: de distancia, gaussiana, multi-cuadrática, entre otras, siendo la gaussiana

$$\phi(x) = \exp \left( -\frac{x^2}{2\sigma^2} \right)$$

la más común de todas [6]. El parámetro  $\sigma$  controla las propiedades de suavidad de la interpolación. La interpolación  $h(x)$  de una función cualquiera con  $n$  funciones de base radial se realiza con [6]

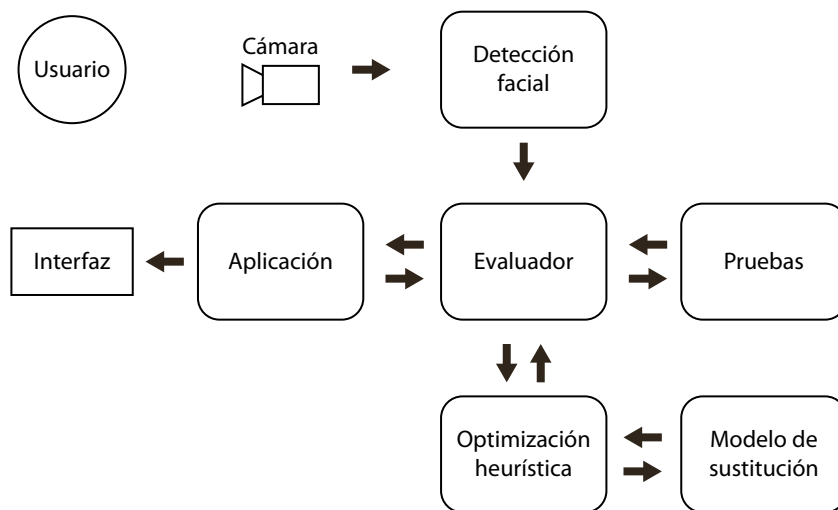
$$h(x) = \sum_n w_n \phi(\|x - x^n\|)$$

donde  $w_n$  son los pesos de cada función base centrada en  $x^n$ .  $\phi(x)$  corresponde a la FBR gaussiana y depende de la distancia  $\|x - x^n\|$  entre  $x$  y  $x^n$ , la cuál es usualmente euclídana.

Según estudios realizados [7], se concluyó que los modelos de sustitución que utilizan FBR funcionan mejor tanto para problemas de pequeña como de gran escala con alto orden de no linealidad.

## Descripción del sistema

En la figura 2 se muestra el diagrama de bloques del sistema implementado, el cual se divide en seis módulos o bloques: la aplicación seleccionada (sudoku), detección facial (estimación de homografía), el algoritmo de optimización heurística (OEP), el modelo de sustitución (utilizando FBR gaussianas), el evaluador (para la selección final de la configuración de interfaz mostrada al usuario) y el módulo de pruebas.



**Figura 2.** Diagrama de bloques de la solución con los algoritmos y técnicas utilizadas.

El flujo de ejecución del sistema inicia con la configuración de la cámara, para definir la región de interés (rostro) de la imagen de referencia y posteriormente se da inicio con la partida de la aplicación seleccionada (sudoku). Una vez finalizada dicha partida, se agrega la información de la interfaz utilizada y los resultados obtenidos a los datos conocidos, los cuales son enviados a través del sistema como entradas al modelo de sustitución, con el fin de recalculas las funciones de aptitud de cada uno de los 3 objetivos a minimizar (tiempo de la partida, tiempo sin mirar a la pantalla y cantidad de errores).

Con estas nuevas funciones se ejecuta una vez más la OEP y se genera una lista con las partículas en las posiciones más prometedoras, luego estas son enviadas al módulo evaluador para que este seleccione una única partícula óptima. La información de la posición de dicha partícula será utilizada para generar la interfaz de la próxima partida.

## Resultados

En esta sección se presentan los resultados de las evaluaciones a los módulos principales del sistema desarrollado.

### Optimización heurística y modelo de sustitución

En el cuadro 1 se observa la cantidad de evaluaciones de aptitud realizadas (por el modelo de sustitución) por objetivo, para 5 partidas de sudoku y 3 cantidades distintas de partículas en OEP.

**Cuadro 1.** Cantidad de valuaciones de funciones realizadas por el modelo de sustitución.

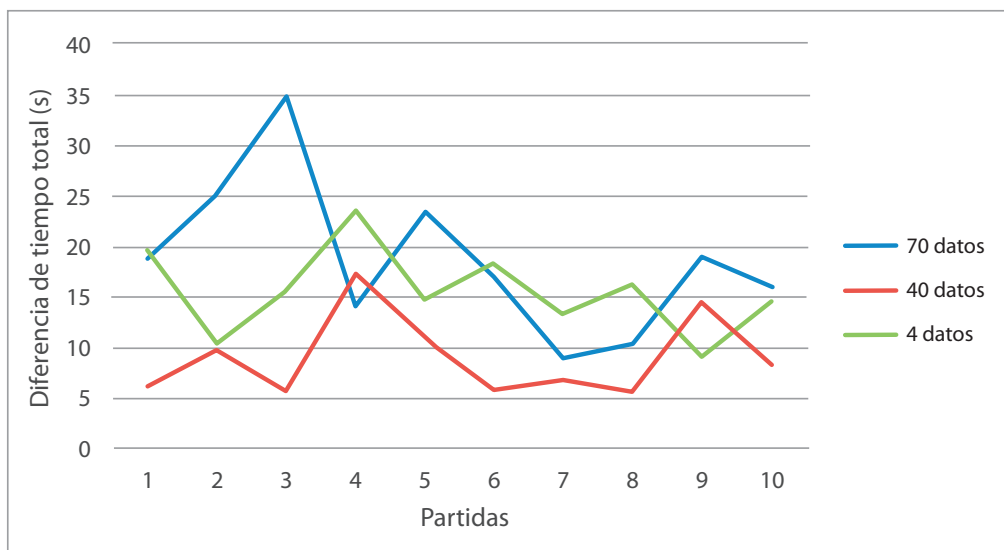
Partidas	10 partículas	25 partículas	50 partículas
1	200	275	900
2	90	550	450
3	230	425	500
4	180	600	400
5	90	475	350

Se realiza una única evaluación directa por partida, por lo que para cada columna de la tabla 3.1 (o ejecución del sistema con 5 partidas de longitud) se realizaron 5 evaluaciones con el usuario y más de 1900 evaluaciones, en promedio, con el modelo de sustitución. Esto evidencia cómo el marco de trabajo desarrollado permite estudiar problemas que, por sus características y dimensiones, eran imposibles de evaluar con el sistema anterior, debido al tiempo que le tomaría a un humano realizar tal cantidad de evaluaciones.

Por otro lado, se realizó un experimento que permite evaluar la exactitud del modelo y cómo se ve afectado por distintas cantidades de datos conocidos inicialmente. Con el fin de minimizar las variables de la prueba, se seleccionó un único usuario experto en la aplicación de sudoku y con el cuál el algoritmo de estimación de homografía presentaba resultados estables acorde a la premisa planteada.

Dicho usuario realizó, en varias sesiones, 3 rondas de 10 partidas de sudoku para 3 cantidades de datos iniciales (4, 40 y 70 datos), completando un total de 90 partidas de sudoku. Cada uno de estos datos iniciales consta de la configuración de la interfaz y los resultados reales de su evaluación con otro usuario.

Para presentar los resultados de dicho experimento, se calculó la diferencia entre los valores promedio reales y los estimados por el modelo de sustitución para cada uno de los tres objetivos que se desean minimizar: tiempo total de partida, tiempo sin mirar a la pantalla y cantidad de errores. En la figura 3 se muestran los resultados para el tiempo total de partida.

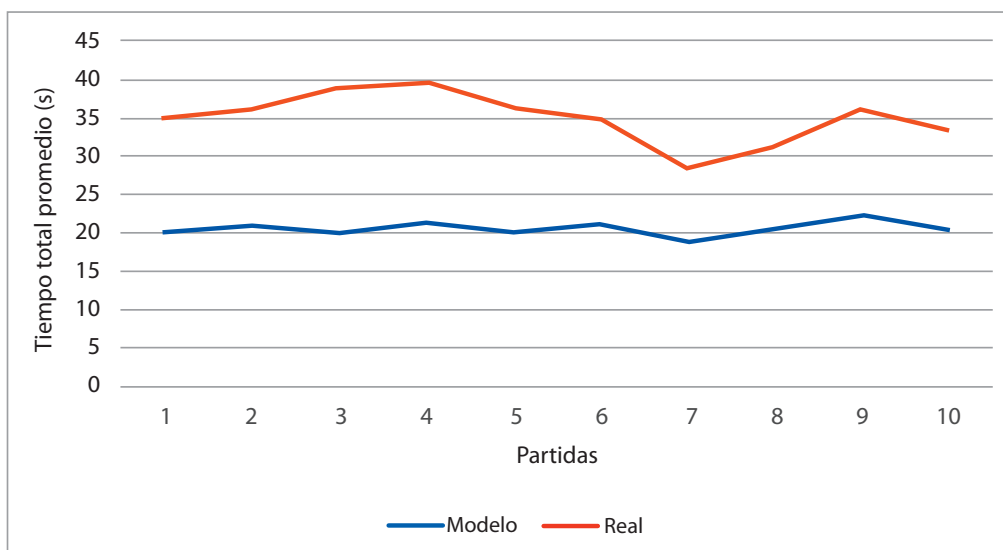


**Figura 3.** Diferencia de tiempos totales reales y según el modelo de sustitución, para 3 cantidades de datos iniciales.

No se observa ninguna mejora considerable en el desempeño del algoritmo para las diferentes cantidades de datos iniciales. Por otro lado, sí es apreciable una tendencia a la baja, lo que indica que las aproximaciones del modelo se acercan gradualmente a las mediciones reales con el usuario.

Para apreciar de mejor manera el comportamiento del algoritmo, en la figura 4 se presentan los valores promedio para el tiempo total de la partida. Para este caso, se tomaron en cuenta todas las partidas jugadas, esto debido a lo expuesto sobre no encontrarse diferencias considerables, ni cambios de tendencia, entre distintas cantidades de datos iniciales.

Se observa cómo los valores aproximados por el modelo de sustitución, en la figura 4, se mantienen cercanos al valor óptimo definido de 20 segundos (tiempo mínimo necesario para concluir la partida sin cometer errores). A partir de la partida 6, se aprecia cómo el modelo aproxima de manera más acertada el comportamiento del usuario, esto debido a la información adicional generada luego de cada partida.



**Figura 4.** Comparación entre el tiempo total promedio según el modelo de sustitución y el real.

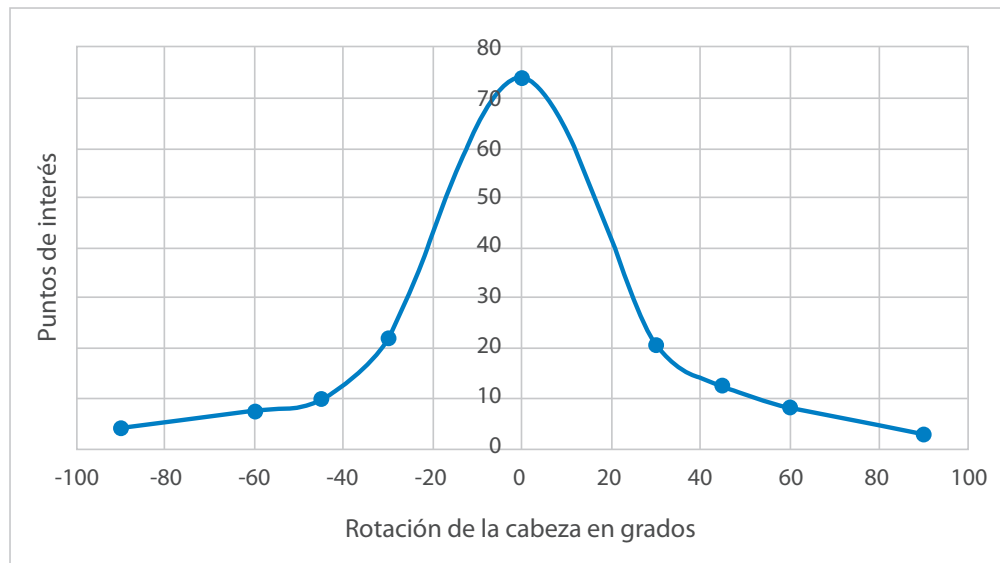
### Detección facial

En el gráfico de la figura 5 se observan las mediciones de puntos de interés promedio detectados para tres usuarios distintos, de acuerdo al ángulo de rotación de la cabeza en grados.

Para la obtención de los datos se colocaron objetos frente al usuario, de manera que al verlos directamente se obtuvieran los ángulos de rotación deseados. La tendencia de la gráfica concuerda con la premisa planteada de que al mirar directamente a la pantalla se presenta un comportamiento de imagen en 2 dimensiones, pero que al girar la cabeza la cantidad de puntos de interés en común con la referencia decae drásticamente conforme el ángulo de rotación se aleja de 0°.

Si bien estos resultados son prometedores y validan la premisa, se deben complementar con lo observado durante la prueba de concepto, en un ambiente menos controlado y con usuarios que desconozcan completamente el funcionamiento del método planteado.





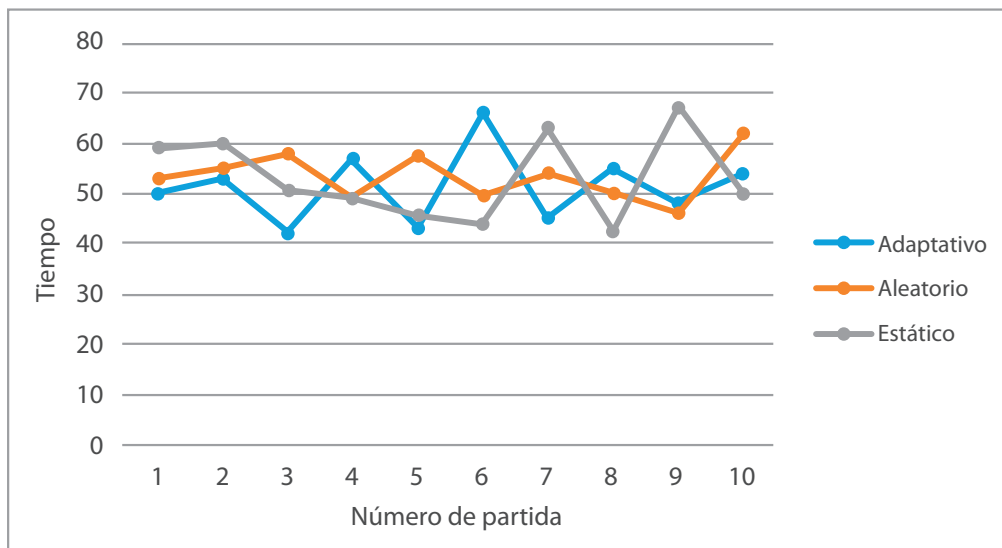
**Figura 5.** Promedio de puntos de interés detectados para distintos ángulos de rotación de la cabeza.

### Prueba de concepto

Se ejecutaron un total de 21 pruebas con distintos usuarios, 12 mujeres y 9 hombres, con edades comprendidas entre los 20 y los 58 años. En adición a esta variedad generacional, el grupo de participantes incluye tanto a personas que juegan sudoku diariamente, como a personas que desconocían el juego y se les explicaron las reglas momentos antes de iniciar la primera partida.

Cada usuario fue asignado aleatoriamente a uno de los tres tipos de prueba (adaptativo, aleatorio o estático) y completó un total de 10 partidas de sudoku con 8 casillas en blanco cada una. De cada partida se almacenó el tiempo total, el tiempo sin mirar a la pantalla, la cantidad de errores cometidos, los valores para replicar la configuración de la interfaz (16 variables) y las respuestas a cada una de las cinco preguntas.

En el gráfico de la figura 6 se muestran los valores promedio de los tiempos totales por partida para los tres tipos de pruebas realizadas. No se aprecia ninguna tendencia para ninguno de los tres tipos de pruebas, lo que parece indicar un comportamiento del juego de sudoku que no se ve afectado por variaciones en la interfaz utilizada.



**Figura 6.** Promedio de los tiempos totales por partida para cada tipo de prueba.

### Observaciones adicionales

Aparte de los resultados cuantificables, las pruebas y evaluaciones dejaron en evidencia una serie de fenómenos que no se consideraron inicialmente y que es relevante mencionarlos a manera de observaciones o lecciones aprendidas.

### Estimación de homografía

En primer lugar, esta técnica demostró no ser efectiva para todo tipo de rostros, específicamente, usuarios con anteojos y vello facial abundante, no son reconocidos adecuadamente por el algoritmo. Esto debido a que los puntos de interés dentro de estas secciones cambian con gran facilidad debido a la iluminación, disminuyendo considerablemente la cantidad de puntos de coincidencia con la imagen de referencia a lo largo de la prueba, incumpliendo la premisa planteada.

El algoritmo también presenta problemas con ciertos comportamientos presentados en la prueba de concepto. Por ejemplo, algunos usuarios tienden a mover los labios para realizar conteos o incluso los realizan en voz alta, este movimiento de los labios también modifica los puntos de interés de dicha zona, disminuyendo la coincidencia con la referencia. Otros usuarios, tienden a cambiar de posición por cansancio alterando el ángulo de rotación de su cabeza significativamente y en ocasiones cubren parcialmente su rostro con una o ambas manos, afectando, de igual manera, la precisión de la técnica utilizada.

### Concentración y atención

Todos los usuarios se mantuvieron enfocados en completar el juego de sudoku en todo momento. Los errores fueron escasos e independientes del tipo de interfaz y todo el tiempo sin mirar a la pantalla se debió a problemas de detección y a voltear la cabeza hacia el teclado para presionar la tecla correcta. Tal como los resultados lo respaldan, no se encontró ninguna alteración en la concentración ni en la atención al utilizar interfaces adaptativas. Sin embargo, esto puede ser una característica de la prueba de concepto realizada y no una generalidad, por lo que se insta a utilizar el marco de trabajo con aplicaciones distintas.

## Conclusiones

Se desarrolló un marco de trabajo de un sistema adaptativo utilizando un algoritmo de optimización por enjambre de partículas y un modelo de sustitución para eliminar la dependencia a la evaluación de cada interfaz directamente con el usuario.

Para problemas que involucren evaluaciones directas con el usuario a manera de función de aptitud en algoritmos heurísticos de optimización, la utilización de modelos de sustitución para la aproximación de dichas funciones, permite variar el tamaño de las poblaciones y mantener constante la cantidad necesaria de evaluaciones directas con el usuario. Esto abre la puerta al paralelismo y a un incremento en el tamaño de los espacios de búsqueda y objetivo. Se demostró, además, que el modelo de sustitución implementado mejora su aproximación del comportamiento real del usuario al incorporar la información adicional generada con el paso del tiempo.

Se comprobó la premisa planteada para la utilización de la estimación de homografía como detección de rotación facial. Sin embargo, también se encontraron debilidades del método (problemas con ciertas conductas, con el uso de anteojos y la presencia de vello facial) que deben ser consideradas para su correcta utilización.

Se plantea, como trabajo futuro, sustituir el módulo de optimización por uno que utilice un algoritmo genético, para compararlo con la optimización por enjambre de partículas para este problema en específico. Por otro lado, se sugiere realizar diferentes pruebas de concepto, que permitan comparar el comportamiento del sistema para distintas aplicaciones.

## Agradecimientos

Agradezco al Dr. Pablo Alvarado, al Dr. Guido Kempter y al equipo de trabajo del departamento User Centred Technologies (UCT) de la universidad FH Vorarlberg en Austria, ya que este trabajo no hubiese sido posible sin el apoyo y consejos brindados.

## Referencias

- [1] Pat Langley. «User modeling in adaptive interface». En: *UM99 User Modeling*. Springer, 1999, págs. 357-370.
- [2] Guido Kempter y Walter Ritter. «A Genetic Algorithm Based Approach to Psychophysiological Human-Computer Interaction». En: *Universal access in the information society* (oct. de 2005).
- [3] Richard Szeliski. *Computer Vision: Algorithms and Applications*. Springer Science & Business Media, 30 de sep. de 2010. 824 págs. isbn: 978-1-84882-935-0.
- [4] Andrea E. Olsson. *Particle Swarm Optimization: Theory, Techniques and Applications*. Commack, NY, USA: Nova Science Publishers, Inc., 2010.
- [5] Rania Hassan y col. «A comparison of particle swarm optimization and the genetic algorithm». En: *46th AIAA/ASME/ASCE/AHS/ASC structures, structural dynamics and materials conference*. 2005.
- [6] Chris Bishop, Christopher M. Bishop y col. *Neural Networks for Pattern Recognition*. Oxford university press, 23 de nov. de 1995. 501 págs. isbn: 978-0-19-853864-6.
- [7] R. Jin, W. Chen y T. W. Simpson. «Comparative studies of metamodelling techniques under multiple modelling criteria». En: *Structural and Multidisciplinary Optimization* 23. 1 (dic. de 2001), págs. 1-13. issn: 1615-1488. doi: 10.1007/s00158-001-0160-4. url: <https://doi.org/10.1007/s00158-001-0160-4>.
- [8] Guido Kempter, Karl-Heinz Weidmann y Pascale Roux. «What are the benefits of analogous communication in human computer interaction». En: *Universal access in hci: inclusive design in the information society* (2003), págs. 1427-1431.