

Universidad Autónoma de Baja California
Facultad de ciencias
Maestría y Doctorado en Ciencias e Ingeniería



Re-identificación facial basada en correlación y con predicción del campo de vista destino

TESIS

que para cubrir parcialmente los requisitos necesarios para obtener el
grado de

Doctor en Ciencias

Presenta

Everardo Santiago Ramírez

Ensenada, Baja California, Marzo del 2016.

Universidad Autónoma de Baja California

Facultad de Ciencias

Re-identificación facial basada en correlación
y con predicción del campo de vista destino

TESIS

que para cubrir parcialmente los requisitos necesarios para obtener el grado de

Doctor en Ciencias

Presenta

Everardo Santiago Ramírez

Aprobada por:



Dr. José Ángel González Fraga
Director de Tesis



Dr. Everardo Gutiérrez López
Miembro del Comité



Dr. Omar Álvarez Xochihua
Miembro del Comité



Dr. Juan Iván Nieto Hipólito
Miembro del Comité



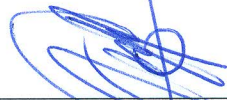
Dr. Juan José Tapia Armenta
Miembro del Comité

Ensenada, Baja California, Marzo del 2016.

Resumen de la tesis de **Everardo Santiago Ramírez**, presentada como requisito parcial para la obtención del grado de DOCTOR EN CIENCIAS del programa de Maestría y Doctorado en Ciencias e Ingeniería (MyDCI) de la UABC. Ensenada Baja California, México, Marzo del 2016.

Re-identificación facial basada en correlación y con predicción del campo de vista destino

Resumen aprobado por:



José Ángel González Fraga
Director de Tesis

El problema de la re-identificación facial consiste en el seguimiento e identificación de una cara a través de múltiples cámaras. Una solución a este problema tiene aplicaciones en vídeo vigilancia, casas inteligentes, entretenimiento, entre otras. En esta tesis se propone una metodología para la re-identificación facial basada en correlación y con predicción del campo de vista destino.

La metodología propuesta consiste de cuatro etapas principales. Primero, una cara es detectada sobre una imagen de un vídeo mediante un algoritmo de detección que usa un banco de filtros. Segundo, la identidad de la cara detectada se obtiene mediante un proceso de identificación facial. En este proceso se usan plantillas biométricas construidas con muestras faciales seleccionadas por una metodología basada en optimización numérica. Tercero, la cara detectada e identificada es seguida dentro del campo de vista usando un filtro que se adapta a los cambios de la apariencia facial y capaz de recuperarse de fallos. Durante el seguimiento se captura un conjunto de muestras faciales, el cual es usado para construir un filtro de correlación que es usado como descriptor facial. Cuarto, cuando la persona abandona el campo de vista, entonces, se predice el campo de vista destino. Estas cuatro etapas se repiten en cada campo de vista donde se detecta la cara objetivo. Finalmente, los descriptores faciales son comparados mediante correlación para determinar correspondencias entre las caras.


La metodología de re-identificación facial propuesto fue evaluada sobre un escenario real. Los resultados obtenidos confirman su desempeño en la práctica.

Palabras Clave: *Re-identificación facial, Seguimiento de personas Filtros de correlación, Predicción del campo de vista destino, Algoritmos de optimización*

Abstract of the thesis presented by **Everardo Santiago Ramírez** as partial requirement to obtain the Doctor of Science degree of the Maestría y Doctorado en Ciencias e Ingeniería (MyDCI) program at UABC. Ensenada, Baja California, México. March, 2016.

Facial re-identification based on correlation and with prediction of the destination field of view

Abstract approved by:



José Ángel González Fraga
Thesis Advisor

The facial re-identification problem consists of tracking and identification of a face across multiple cameras. A solution to this problem could be applied to video-surveillance, smart homes, entertainment, and so forth. In this thesis a correlation-based methodology for face re-identification with prediction of the destination field of view is proposed. The proposed methodology consists of four main phases. First, a face is detected in a frame by a detection algorithm using a filter bank. Second, the identity of the face detected is obtained by a facial identification process. Biometric templates constructed with facial samples, selected by a optimization-based methodology, are used in this process. Third, the detected and identified face is tracked within the field of view using a filter that adapts to facial appearance changes and is able to recover from failures. A set of facial samples are captured during the tracking, which is used to build a correlation filter that works as a facial descriptor. Fourth, when the subject leaves the field of view, then, his destination field of view is predicted. This four stages are repeated in each field of view where the target face is detected. Finally, the generated facial descriptors are compared by correlation to determinate correspondences between the faces. The proposed facial re-identification methodology was evaluated in a real scenario. The obtained results confirm its performance in practice.

Keywords: *Facial re-identification, Person tracking Correlation filters, Prediction of the destination field of view, Optimization algorithms*

Dedicatoria

A mis padres.

A mis hermanos y hermanas.

A mis familiares.

A todas aquellos que creyeron en mi y me brindaron su apoyo.

Agradecimientos

A Dios por darme la vida y proveer todo lo necesario para lograr mis anhelos. A todos aquellos que hicieron posible esta tesis. También quiero agradecer a todos los miembros de mi familia por apoyarme en mi educación y por darme ánimo cuando más lo necesité. Agradezco enormemente al Dr. José Ángel González Fraga por guiarme en mi educación y por todo su apoyo para iniciar esta aventura de la investigación. También agradezco a los doctores José Ángel González Fraga, Everardo Gutierrez López, Omar Álvarez Xochihua, Juan Iván Nieto Hipólito y Juan José Tapia Armenta por su disponibilidad para ser miembros de mi comité de tesis, por sus observaciones, comentarios, críticas, sugerencias y apoyo para la realización de esta tesis.

Al pueblo de México que a través del Consejo Nacional de Ciencia y Tecnología (CONACyT) me brindaron su apoyo económico al pagar sus correspondientes impuestos y otras obligaciones. A la Universidad Autónoma de Baja California (UABC) por las facilidades otorgadas para la realización de este trabajo.

Índice general

| | |
|--|-----------|
| 1. Introducción | 1 |
| 1.1. Objetivo de la investigación | 2 |
| 1.1.1. Objetivo general | 2 |
| 1.1.2. Objetivos específicos | 2 |
| 1.2. Contribuciones | 3 |
| 1.3. Organización de la tesis | 3 |
| 2. Re-identificación de personas | 5 |
| 2.1. Estado del arte | 8 |
| 2.2. Resumen | 10 |
| 3. Reconocimiento de patrones por correlación | 13 |
| 3.1. Acoplamiento de plantilla | 16 |
| 3.2. Filtros de correlación compuestos | 18 |
| 3.2.1. Funciones discriminantes sintéticas | 18 |
| 3.2.2. Filtrado no lineal | 19 |
| 3.2.3. Filtro del mínimo promedio de la energía de correlación | 19 |
| 3.2.4. Filtro del máximo promedio de la altura de correlación | 21 |
| 3.2.5. Función discriminante sintética de compromiso óptimo | 22 |
| 3.2.6. Función discriminante sintética de compromiso óptimo no restringido | 22 |
| 3.2.7. Filtro compuesto no lineal no restringido | 22 |
| 3.2.8. Filtro promedio de filtros exactos sintéticos | 23 |
| 3.2.9. Filtro de la mínima salida de la suma del error cuadrado | 23 |
| 3.2.10. Filtro de solo fase segmentado asimétrico | 24 |
| 3.3. Resumen | 25 |
| 4. Procesamiento facial | 26 |
| 4.1. Reconocimiento facial | 27 |
| 4.2. Detección facial | 34 |
| 4.3. Seguimiento facial | 36 |
| 4.4. Resumen | 37 |

| | |
|---|-----------|
| 5. Detección facial mediante filtros de correlación compuestos | 38 |
| 5.1. Método de detección facial basado en correlación | 39 |
| 5.1.1. Selección del conjunto de entrenamiento | 40 |
| 5.1.2. Síntesis del filtro de correlación compuesto no lineal | 42 |
| 5.1.3. Detección facial | 45 |
| 5.2. Resultados experimentales | 46 |
| 5.2.1. Configuración del conjunto de imágenes para la evaluación | 47 |
| 5.2.2. Medidas de desempeño para el algoritmo de detección facial | 50 |
| 5.3. Resultados numéricos | 51 |
| 5.4. Conclusiones | 54 |
| 6. Seguimiento facial mediante filtros adaptativos | 58 |
| 6.1. Seguimiento facial mediante filtros adaptativos | 58 |
| 6.2. Resultados | 64 |
| 6.2.1. Configuración de los datos de prueba | 64 |
| 6.2.2. Resultados numéricos | 64 |
| 6.3. Conclusión | 66 |
| 7. Selección automática del conjunto de entrenamiento | 68 |
| 7.1. Trabajo relacionado | 68 |
| 7.2. Fundamentos de los algoritmos de optimización | 70 |
| 7.2.1. Hill Climbing | 71 |
| 7.2.2. Simulated Annealing | 71 |
| 7.2.3. (1 + 1) Evolution Strategy | 72 |
| 7.2.4. Grey Wolf Optimizer | 73 |
| 7.3. Selección automática del conjunto de entrenamiento | 75 |
| 7.3.1. Metodología basada en optimización para seleccionar conjuntos de entrenamiento | 78 |
| 7.4. Resultados experimentales y discusión | 78 |
| 7.4.1. Configuración de los datos | 79 |
| 7.4.2. Resultados numéricos | 81 |
| 7.4.3. Discusión acerca de la importancia del conjunto de entrenamiento | 84 |
| 7.4.4. Comparación de la selección convencional contra la selección hecha por el método propuesto | 86 |
| 7.5. Conclusiones | 88 |
| 8. Re-identificación facial con predicción del campo de vista destino | 92 |
| 8.1. Método de re-identificación facial | 93 |
| 8.1.1. Registro facial | 93 |
| 8.1.2. Descriptor facial | 96 |
| 8.1.3. Establecimiento de correspondencia | 97 |
| 8.1.4. Predicción del campo de vista destino | 97 |
| 8.1.5. Algoritmo de re-identificación facial propuesto | 100 |

| | |
|---|------------|
| 8.2. Resultados experimentales | 102 |
| 8.2.1. Configuración de los datos de prueba | 102 |
| 8.2.2. Resultados numéricos y discusión | 105 |
| 8.3. Conclusiones | 110 |
| 9. Conclusiones | 114 |
| 9.1. Trabajo futuro | 115 |
| A. Predicción del campo de vista destino | 116 |
| A.1. Trabajos relacionados | 116 |
| A.2. Principios teóricos del algoritmo de hormigas | 118 |
| A.3. Modelo de predicción del CDV destino | 119 |
| A.3.1. Predicción del campo de vista destino | 120 |
| A.3.2. Resultados experimentales del modelo de predicción | 125 |
| A.4. Conclusiones | 135 |
| Referencias | 142 |

Índice de figuras

| | | |
|------|---|----|
| 2.1. | Representación gráfica de una red de cámaras sobre una zona de interés. | 6 |
| 2.2. | Representación gráfica del proceso de re-identificación. En cada CDV se sigue a la persona y se generan los descriptores los cuales son comparados para determinar coincidencias. | 7 |
| 3.1. | Proceso de correlación entre una imagen de entrada y un filtro de correlación en el dominio de Fourier. | 14 |
| 3.2. | Planos de salida de correlación para un objeto verdadero (izquierda) y para un objeto falso (derecha). | 15 |
| 3.3. | Plano de correlación en donde se delimitan el área del pico y su alrededor para el cálculo de PSR . | 15 |
| 3.4. | Área del plano de correlación sobre el cual se calcula la CD . | 16 |
| 3.5. | Resultado de correlacionar una imagen de entrada con una plantilla simple. | 17 |
| 3.6. | Resultado de correlacionar una imagen de entrada con un filtro SDF no lineal. | 20 |
| 4.1. | Esquema de un sistema de reconocimiento facial como un conjunto de procesos cooperativos. | 27 |
| 4.2. | Diferentes imágenes de una misma persona que forman una misma clase facial. | 28 |
| 4.3. | Esquema básico de la tarea de verificación facial. | 29 |
| 4.4. | Esquema básico de la tarea de identificación facial. | 30 |
| 5.1. | Principales etapas del método de detección facial propuesto. | 39 |
| 5.2. | Conjunto de imágenes disponible D , donde un círculo verde indica las imágenes seleccionadas por el algoritmo propuesto para el conjunto T . | 42 |
| 5.3. | Aplicación de las transformaciones al conjunto original para construir los clasificadores fuertes. | 44 |
| 5.4. | Proceso de correlación de la imagen de entrada con cada elemento del banco de filtros B . | 46 |
| 5.5. | Combinación de varias subdetecciones sobre una misma región facial para una sola detección final verdadera. | 47 |
| 5.6. | Muestras faciales de una persona en la Yale B recortado. | 48 |

| | | |
|-------|--|----|
| 5.7. | Imágenes de una persona para los subconjuntos Pose y Frontales. . . . | 49 |
| 5.8. | Imágenes de las personas con pose en Yale B extendido. | 49 |
| 5.9. | Conjunto de prueba P usado en la selección del conjunto de entrena- miento. | 52 |
| 5.10. | Valores PSR de las soluciones candidato en cada iteración (arriba) y disminución del tamaño del conjunto de entrenamiento debido a la opti- mización (abajo). | 53 |
| 5.11. | De izquierda a derecha: imagen original, plano de correlación del método propuesto y plano de correlación del filtros ASEF. Para cada plano de correlación, arriba se muestra la versión 3D y abajo la versión 2D. . . | 54 |
| 5.12. | Desempeño del método propuesto contra el filtro ASEF en términos de las métricas TLR (arriba) y TLDR (abajo). | 56 |
| 6.1. | Esquema general del método de seguimiento propuesto. | 59 |
| 6.2. | Ventana de seguimiento w con una dimensión constante sobre dos imáge- nes diferentes de un mismo vídeo S . A la izquierda la imagen $s_1(x, y)$ y a la derecha la imagen $s_{1066}(x, y)$ | 60 |
| 6.3. | En la parte inferior imagen facial promedio resultante de las diez imáge- nes faciales en la parte superior. | 61 |
| 6.4. | Diagrama de flujo del algoritmo propuesto para el seguimiento facial usando un filtro de correlación adaptable. | 62 |
| 6.5. | Salida de correlación de un filtro generado con una imagen facial y ver- siones generadas mediante la aplicación de transformaciones. | 63 |
| 6.6. | Imagen de muestra de los vídeos evaluados. | 66 |
| 7.1. | Diagrama de bloque de la metodología propuesta para la selección de conjuntos de entrenamiento para filtros compuestos. | 79 |
| 7.2. | Muestra del conjunto $T_{id,i}$ (izquierda) y P_i (derecha) para la i -ésima persona en CMU. | 80 |
| 7.3. | Muestra del conjunto $T_{id,i}$ (izquierda) y P_i (derecha) para la i -ésima persona en Yale B. | 80 |
| 7.4. | Imágenes resultantes del preprocesamiento. De izquierda a derecha: ima- gen original, filtrado de retina, diferencias de Gaussianas e imagen LBPI. | 81 |
| 7.5. | Gráficas de efectos principales para las métricas FERET con medias ajustadas. | 85 |
| 7.6. | Gráficas de intervalos del tamaño promedio del conjunto de entrenamien- to seleccionado por la metodología propuesta para CMU (izquierda) y Yale B (derecha). | 86 |
| 7.7. | Desempeño de filtros compuestos sintetizados con conjuntos de entrena- miento de diferentes tamaños. Arriba para conjuntos seleccionados con- vencionalmente y abajo para conjuntos seleccionados aleatoriamente. . | 87 |
| 7.8. | Desempeño de filtros compuestos (izquierda) y tamaño del conjunto de entrenamiento en cada iteración de la metodología propuesta (derecha). | 88 |

| | |
|---|-----|
| 7.9. Desempeño TRF del filtro ASPOF (arriba) y tamaño del conjunto de entrenamiento (abajo) usando la el conjunto CMU. Línea continua para el resultado obtenido por el método propuesto y línea segmentada para el resultado del método convencional. | 89 |
| 7.10. Desempeño TRF del filtro ASPOF (arriba) y tamaño del conjunto de entrenamiento (abajo) usando la el conjunto CMU. Línea continua para el resultado obtenido por el método propuesto y línea segmentada para el resultado del método convencional. | 90 |
| 8.1. Diagrama del método de re-identificación facial propuesto. | 94 |
| 8.2. Diagrama básico de bloques del proceso de registro facial para objetivos nuevos. | 95 |
| 8.3. Interfaz gráfica de usuario del módulo de registro facial. | 96 |
| 8.4. Interfaz gráfica de usuario del módulo de captura de registro de datos personales. | 97 |
| 8.5. Diferentes imágenes faciales de una misma persona capturadas por el módulo de <i>Registro facial</i> | 98 |
| 8.6. Conjuntos $T_{id,i}$ (izquierda) y P_i (derecha) correspondientes a la i -ésima persona. | 98 |
| 8.7. Conjunto de entrenamiento T_i seleccionado por el módulo de <i>Registro facial</i> | 99 |
| 8.8. Red de cámaras (izquierda) y su respectivo grafo G (derecha). | 102 |
| 8.9. Posición de la cámara con respecto a la cara de la persona (arriba) e imagen facial deseada (abajo). | 103 |
| 8.10. Vista aérea, croquis y grafo de campos de vista del área vigilada por cámaras denominado LPA2. | 104 |
| 8.11. Patrones de movimientos para la evaluación del algoritmo de re-identificación sobre el escenario LPA2. | 105 |
| 8.12. Desempeño RES de los patrones de movimiento $R'_1 = \{acdb\}$ y $R'_2 = \{acb\}$ sobre el grafo de la Figura 8.10. | 106 |
| 8.13. Desempeño RES del método de re-identificación facial propuesto sobre el escenario LPA2 de la Figura 8.10. | 110 |
| 8.14. Interfaz gráfica de usuario para el operador del sistema de re-identificación facial. | 110 |
| A.1. Red de cámaras representadas como un grafo conectado $G = (V, A)$ con una ruta de un nodo entrada a un nodo destino y del nodo destino al nodo salida. | 120 |
| A.2. Tabla de feromonas de la i -ésima persona. Esta estructura de datos contiene los niveles de feromonas del camino que une a los campos de vista i y j | 121 |

| | |
|---|-----|
| A.3. Aplicación del modelo de actualización sobre la arista (i, j) cuando es visitado varias veces. La línea continua indica el nivel de feromonas actualizado en cada visita, mientras que la línea segmentada indica la cantidad de feromonas agregado en la actualización. | 123 |
| A.4. Red de cámara y su correspondiente grafo. | 124 |
| A.5. Probabilidad de selección del CDV_i en 1000 predicciones para las vecindades de tamaño 2 y 3. | 125 |
| A.6. Probabilidad de selección del CDV_i en 1000 predicciones para las vecindades de tamaño 4 y 5. | 126 |
| A.7. Probabilidad de selección del CDV_i en 1000 predicciones para las vecindades de tamaño 6 y 10. | 127 |
| A.8. Probabilidad de selección del CDV_i en 1000 predicciones para las vecindades de tamaño 20 y 30. | 128 |
| A.9. Probabilidad de selección del CDV_i en 1000 predicciones para las vecindades de tamaño 100 y 200. | 129 |
| A.10. En la parte izquierda se muestra el grafo de campos de vista con entradas y salidas de color rojo, mientras que en el lado derecho se muestra un recorrido con tres ciclos. | 130 |
| A.11. Grafo una red de cinco cámaras. | 130 |
| A.12. Recorrido de una persona en una red de cámaras. Las flechas rojas indican la trayectoria tomada, mientras que las flechas verdes señalan el CDV en que se encuentra el agente. | 131 |
| A.13. Mapa de probabilidades para los vecinos del CDV c , los cuales son $V_c = \{a, b, e\}$ | 133 |
| A.14. Primer grupo de grafos usados en el experimento. | 134 |
| A.15. Segundo grupo de grafos usados en el experimento. | 135 |
| A.16. Tercer grupo de grafos usados en el experimento. | 136 |
| A.17. Cuarto grupo de grafos usados en el experimento. | 137 |
| A.18. Diagrama de cajas de la Razón de Éxito a Selecciones (RES) obtenido en cada grafo. | 137 |
| A.19. Diagrama de cajas de la cantidad de rutas tomadas por agente artificial sobre cada grafo. | 138 |

Índice de tablas

| | |
|---|----|
| 2.1. Métodos de re-identificación de personas reportados en la literatura. Siglas en la tabla: campos de vista disjuntos (CVD), campos de vista superpuestos (CVS), vídeo (V) y cuerpo completo (CC). | 12 |
| 5.1. Transformaciones de imágenes para incrementar el tamaño de T | 43 |
| 5.2. Desempeño del método propuesto y del filtro ASEF en la detección facial. En verde las celdas con los mejores desempeños TLR y los peores para el resto de las métricas. | 55 |
| 6.1. Características de los vídeos usados en la evaluación. Notación: con cara (CC), sin cara (SC), iluminación intensa (II), sombra total (ST), sombra local (SL), oclusión (Oc) y expresiones (Ex). | 65 |
| 6.2. Desempeño del método propuesto en la tarea de seguimiento facial. En verde las celdas con los mejores desempeños TLR y los peores para el resto de las métricas. | 67 |
| 7.1. Desempeño de los filtros evaluados en términos de la Tasa de Identificación y Detección (TID) (%) con el mejor desempeño para cada función objetivo en negrita. | 82 |
| 7.2. Desempeño de los filtros evaluados en términos de la Tasa de Verificación (TV) (%) con el mejor desempeño para cada función objetivo en negrita. | 82 |
| 7.3. Desempeño de los filtros evaluados en términos de la Tasa de Falsa Alarma (TFA) (%) con el peor desempeño para cada función objetivo en negrita. | 83 |
| 7.4. Desempeño de los filtros evaluados en términos de la Tasa de Falsa Aceptación (TFAc) (%) con el peor desempeño para cada función objetivo en negrita. | 83 |
| 7.5. ANOVA para la métrica TID. | 84 |
| 7.6. ANOVA para la métrica TV | 84 |
| 7.7. ANOVA para la métrica TFA | 84 |
| 7.8. ANOVA para la métrica TFAc. | 84 |
| 7.9. Desempeño de conjuntos de entrenamiento seleccionados convencionalmente y mediante el método propuesto usando GWO y TRF. | 86 |

| | |
|---|-----|
| 8.1. Datos de los vídeos capturados para el patrón de movimiento 1. Notación: número de recorrido (NR), campo de vista (CDV) y total de imágenes (TI). | 105 |
| 8.2. Datos de los vídeos capturados para el patrón de movimiento 2. Notación: número de recorrido (NR), campo de vista (CDV) y total de imágenes (TI). | 105 |
| 8.3. Datos de los vídeos capturados para el patrón de movimiento 3. Notación: número de recorrido (NR), campo de vista (CDV) y total de imágenes (TI). | 106 |
| 8.4. Resultado de comparaciones de imágenes faciales promedio del patrón de movimientos 1. En verde las celdas de los valores donde no se pudo establecer correspondencia. | 108 |
| 8.5. Resultado de comparaciones de imágenes faciales promedio del patrón de movimientos 2. En verde las celdas de los valores donde no se pudo establecer correspondencia. | 109 |
| 8.6. Resultado de comparaciones de imágenes faciales promedio del patrón de movimientos 3. En verde las celdas de los valores donde no se pudo establecer correspondencia. | 112 |
| 8.7. Cantidad de caras capturadas en cada seguimiento. | 113 |
| A.1. Tabla de feromonas F_i inicial | 131 |
| A.2. Tabla de feromonas F_i actualizado después del un recorrido | 132 |
| A.3. Tabla de feromonas F_i después de la evaporación | 132 |
| A.4. Ruta Seleccionada (RS), Ruta Esperada (RE) y Frecuencia de la Ruta Seleccionada (FRS) de cada persona sobre los grafos 1 al 5. En verde las celdas con los recorridos donde RS y RE son diferente. | 138 |
| A.5. Ruta Seleccionada (RS), Ruta Esperada (RE) y Frecuencia de la Ruta Seleccionada (FRS) de cada persona de los grafos 6 al 10. En verde las celdas con los recorridos donde RS y RE son diferente. | 139 |
| A.6. Ruta Seleccionada (RS), Ruta Esperada (RE) y Frecuencia de la Ruta Seleccionada (FRS) de cada persona sobre los grafos 11 al 15. En verde las celdas con los recorridos donde RS y RE son diferente. | 139 |
| A.7. Ruta Seleccionada (RS), Ruta Esperada (RE) y Frecuencia de la Ruta Seleccionada (FRS) de cada persona sobre los grafos 16 al 20. En verde las celdas con los recorridos donde RS y RE son diferente. | 140 |
| A.8. Ruta Seleccionada (RS), Ruta Esperada (RE) y Frecuencia de la Ruta Seleccionada (FRS) de cada persona sobre los grafos 21 al 25. En verde las celdas con los recorridos donde RS y RE son diferente. | 141 |
| A.9. Ruta Seleccionada (RS), Ruta Esperada (RE) y Frecuencia de la Ruta Seleccionada (FRS) de cada persona sobre los grafos 26 al 31. En verde las celdas con los recorridos donde RS y RE son diferente. | 141 |

Capítulo 1

Introducción

La cara es la parte frontal de la cabeza que en los humanos se extiende de la frente a la barbilla e incluye la boca, la nariz, las mejillas y los ojos (Webster, 2015). Una imagen que exhibe la estructura física de la cara de una persona es la que se analiza para fines de reconocimiento. Sin embargo, el reconocimiento facial es una tarea que requiere de un conjunto de procesos cooperativos tales como la detección y el seguimiento. La detección facial tiene por objetivo encontrar y extraer caras en una imagen arbitraria, las cuales pudieran ser autenticadas mediante la tarea de reconocimiento. En la tarea de seguimiento facial, la cara detectada es seguida en una secuencia de vídeo. A las tareas de detección, reconocimiento y seguimiento se le denomina procesamiento facial en esta tesis.

Una de las disciplinas en donde se ha usado muy poco la cara es en la tarea de re-identificación definida de la siguiente manera. Sea un campo de vista (CDV) el espacio vigilado mediante una cámara. Entonces, dada la imagen de una persona que ha sido detectada en un CDV, el objetivo es determinar si esta imagen coincide con un conjunto candidato de imágenes de personas detectadas otros CDV's.

Uno de los principales retos en la re-identificación es la construcción de los *descriptores* que permitan una comparación robusta para determinar coincidencias. Para lograr esto, los *descriptores* deben contener información que caracterice eficientemente algún rasgo del objetivo, como por ejemplo el color de la ropa, la forma de caminar, la cara, entre otros. En (W. Li, Wu, Mukunoki, Kuang, y Minoh, 2015; Shi, Guo, Lai, Chen, y Hu, 2015; Nanni, Munaro, Ghidoni, Menegatti, y Brahnham, 2015) se presentan métodos recientes que emplean el cuerpo completo de la persona para la re-identificación. Estos enfoques principalmente usan el color de la ropa. Una desventaja de estos métodos, es que su desempeño disminuye cuando la persona se cambia o quita alguna de sus prendas. La forma de caminar es otra alternativa para la re-identificación de personas como se propone en (Roy, Sural, y Mukherjee, 2012), aunque al igual que el enfoque anterior requiere del cuerpo completo de la persona para funcionar. Sin embargo, no siempre es posible obtener una imagen del cuerpo completo de la persona. Por esta razón, se han propuesto en la literatura métodos que emplean solo alguna de las partes superiores del cuerpo, como por ejemplo, la cara (Ma, Sub, y Jurie, 2014), cabeza (Utsumi y

Tetsutani, 2004), entre otros.

La cara permanece estable por un periodo de tiempo mayor en comparación con la ropa. Además, que la identidad de la persona se puede obtener más fácilmente comparando la cara actual con plantillas previamente almacenadas. Por otra parte, los métodos de re-identificación en la literatura no conocen el CDV donde *buscar* a la persona. Por lo que la *búsqueda* es en todos los CDV vecinos disponibles, lo cual implica un costo computacional alto. Esto plantea la necesidad de una predicción del CDV destino para optimizar el uso de los recursos disponibles al enfocarlos solo en los CDV's donde hay una alta probabilidad de que aparezca la persona. En esta tesis se presenta una metodología de re-identificación facial basada en correlación y con capacidad de predecir el campo de vista destino.

1.1. Objetivo de la investigación

A partir de la problemática presentada arriba, la cual es fundamentada por los antecedentes en el estado del arte en el Capítulo 2, se ha definido el siguiente objetivo de la investigación.

1.1.1. Objetivo general

Desarrollar una metodología de re-identificación facial basada en filtros de correlación y con capacidad de predecir el campo de vista destino en una red de cámaras.

1.1.2. Objetivos específicos

Para lograr el objetivo general, se cuenta con los siguientes objetivo específicos:

- Desarrollar algoritmos basados en filtros de correlación para las tareas de detección, reconocimiento y seguimiento facial.
- Desarrollar un modelo para predecir el campo de vista destino de una persona.
- Desarrollar una estrategia para el establecimiento de correspondencias entre instancias de una misma persona en una red de cámaras.
- Integrar los resultados de los objetivos específicos anteriores en una metodología de re-identificación facial.
- Implementar y evaluar la metodología en un escenario real.

1.2. Contribuciones

En esta tesis se presenta un método de re-identificación facial. Primero, se desarrollaron algoritmos que usan eficientemente las propiedades de los filtros de correlación en las tareas de detección, seguimiento y reconocimiento facial. Segundo, se desarrolló una metodología basada en optimización numérica para la selección automática del conjunto de imágenes faciales de entrenamiento. Tercero, se desarrolló un modelo basado en los principios de los algoritmos de hormigas para predecir el campo de vista destino. Específicamente, las principales contribuciones de esta tesis son las siguientes:

1. Un método de re-identificación que usa la cara de la persona para su reconocimiento en el contexto biométrico. Este método integra un conjunto de algoritmos que usan eficientemente los filtros de correlación en las diferentes tareas de procesamiento facial.
2. Un modelo basado en los principios de los algoritmos de hormigas para la predicción del campo de vista destino. Este modelo aprende los patrones de movimientos de la persona para predecir probabilísticamente hacia qué CDV irá la persona. Cuando la persona cambia su patrón de movimiento, el modelo inmediatamente empieza a aprenderlo. La implementación de este modelo en la re-identificación permite optimizar los recursos tanto computacionales como humanos.
3. Una metodología basada en optimización numérica para la selección automática del conjunto de imágenes faciales entrenamiento. Esta metodología prueba diferentes combinaciones de imágenes disponibles para seleccionar solamente un subconjunto óptimo para el diseño de un filtro para el reconocimiento facial robusto.

1.3. Organización de la tesis

El resto de la tesis está estructurada de la siguiente manera. En el Capítulo 2 se describe en qué consiste la re-identificación de personas y se resumen los principales métodos existentes. En el Capítulo 3 se provee una descripción del reconocimiento de patrones por correlación y se revisan los filtros de correlación usados para evaluar los diferentes algoritmos propuesto. En el Capítulo 4 se describe en qué consiste cada tarea de procesamiento facial, así como los principales algoritmos reportado en la literatura. Los capítulos restantes contienen las principales propuestas de esta tesis. En el Capítulo 5 se presenta un algoritmo de detección facial que emplea un banco de filtros, cuyo desempeño fue evaluado en imágenes con fondos complejos, iluminación variable, expresiones y ligeras variaciones en rotación y pose. En el Capítulo 6 se presenta un algoritmo de seguimiento facial que combina las propiedades de los filtros y la detección para abordar principalmente la variabilidad en escala. En el Capítulo 7 se propone una metodología para la selección del conjunto de entrenamiento. También, se proponen cuatro funciones objetivo como criterios de optimización en la selección. El desempeño del modelo

fue evaluado mediante simulación, obteniendo resultados prometedores en la práctica. En el Capítulo 8 se presenta una metodología de re-identificación facial que integra los algoritmos de los Capítulos [5, 6, 7] y un modelo de predicción del CDV destino presentado en el Apéndice A. Finalmente, en el Capítulo 9 se resumen las principales conclusiones y algunos indicadores de trabajo futuro.

Capítulo 2

Re-identificación de personas

Una red de cámaras puede ser desplegada en diversas zonas de interés (ZI), tales como áreas públicas, bancos, universidades, estaciones de autobuses, edificios con oficinas, entre otros. Esa red permite vigilar puntos específicos a través de imágenes simples o vídeos. Estas imágenes o vídeos registran una gran cantidad de datos sobre personas y sucesos. Por lo tanto, la extracción y análisis de dicha información es útil en muchas actividades del ser humano, por ejemplo, para detectar intrusos, para identificar delincuentes en una escena, para seguir sospechosos, entre otros. Generalmente, esa extracción y análisis se realiza por un ser humano después del suceso, lo cual es factible a errores, requiere de mucho tiempo, es costoso en cuanto a recursos humanos requeridos y la efectividad de la vigilancia se reduce significativamente, especialmente cuando se tiene muchas imágenes o vídeos a procesar.

La extracción y análisis automático de la información contenida en dichos vídeos mejora la efectividad de la vídeo-vigilancia y reduce los costos tanto económico como materiales de los procesos. Además, esa automatización abre la posibilidad de implementar ambientes inteligentes personalizados e interfaces naturales de interacción humano-computadora.

La extracción y análisis de información contenida en los vídeos puede realizarse mediante la visión artificial. Sin embargo, es una tarea desafiante debido a que nos podemos encontrar con ambientes tanto restringidos como no restringidos. Los ambientes restringidos son aquellos que nos encontramos a diario, los cuales contienen personas no cooperativas, iluminación variable, ruido, oclusión, entre otros. Por otra parte, los ambientes no restringidos son aquellos donde las personas cooperan y existe control sobre otros factores importantes. Sin embargo, en la vídeo-vigilancia podemos encontrarnos con ambos tipos de ambientes que dificultan la aplicación directa de la visión artificial.

En la Figura 2.1 se muestra como ejemplo de ZI una imagen aérea de un campus universitario vigilado por múltiples cámaras con CDV's disjuntos. Sobre esta ZI se ubican varias cámara etiquetadas por las letras $\{a, b, c, d, e, f, g, h, i, j, k, l, m, n, o, p, q, r\}$. Como se puede observar, las cámaras están colocadas por pares en zonas que permiten obtener una imagen frontal de la persona.

En la re-identificación, la persona es seguida en cada CDV y cada seguimiento es

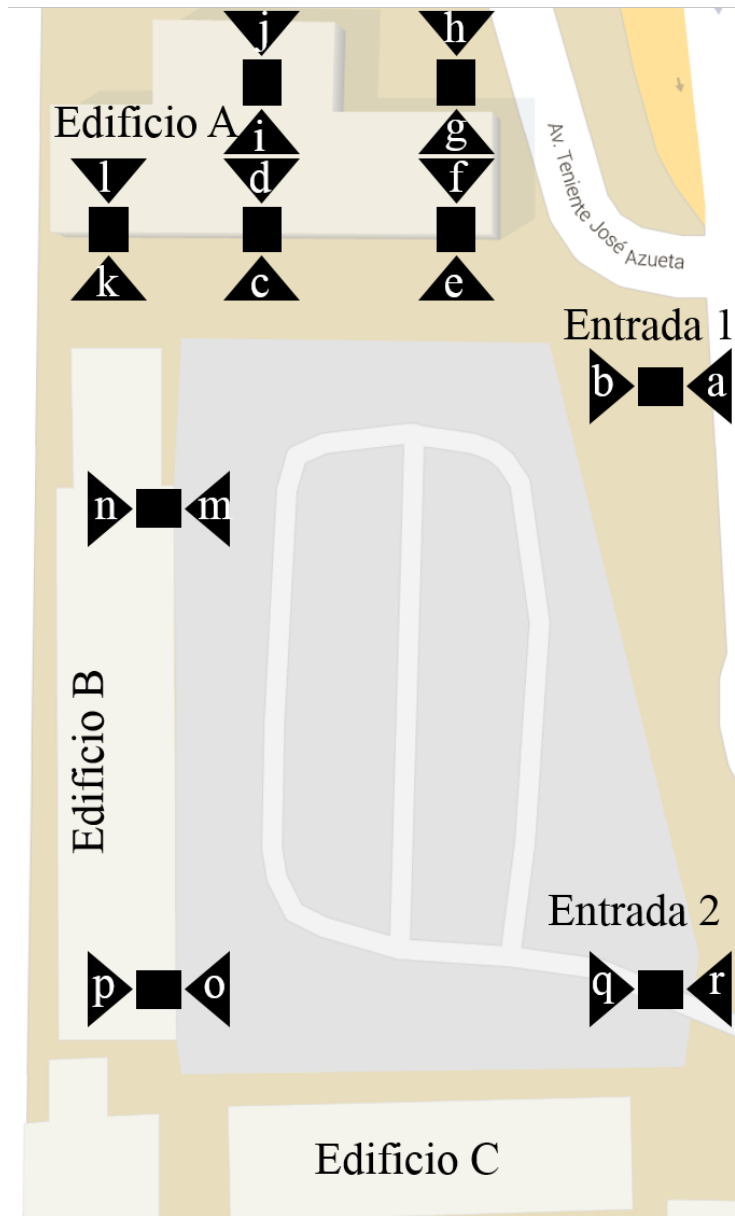


Figura 2.1: Representación gráfica de una red de cámaras sobre una zona de interés.

distinto e independiente. En estos seguimientos se obtienen muestras de las personas que son usadas para generar descriptores, los cuales son comparados para determinar si existe correspondencia entre las imágenes faciales capturadas en las diferentes campos de vista. En la Figura 2.2 se presenta un ejemplo de tres CDV's etiquetadas con $\{a, b, c\}$. En cada una se realiza un seguimiento de la persona para extraer sus principales características y con ellos generar un descriptor. Luego, los descriptores generados son comparados mediante diferentes técnicas, como por ejemplo, cálculo de distan-

cia o correlación para determinar si existe correspondencia. Los descriptores pueden ser contruidos usando características del color de la ropa, partes del cuerpo y rasgos biométricos (ojos, cara, huella dactilar, oreja, forma de caminar, entre otros). De esta forma, los descriptores deben sintetizar información general y específica de esas características para establecer la correspondencia entre diferentes instancias de la persona bajo condiciones complejas.

El seguimiento de una persona en varios campos de vista permite caracterizar su patrón de movimiento, el cual puede ser usado para predecir el CDV destino. Un patrón de movimiento es un conjunto de campos de vista que una persona frecuentemente recorre de forma sistemática. A continuación se describen algunos enfoques de re-identificación de personas que se han reportado en la literatura.

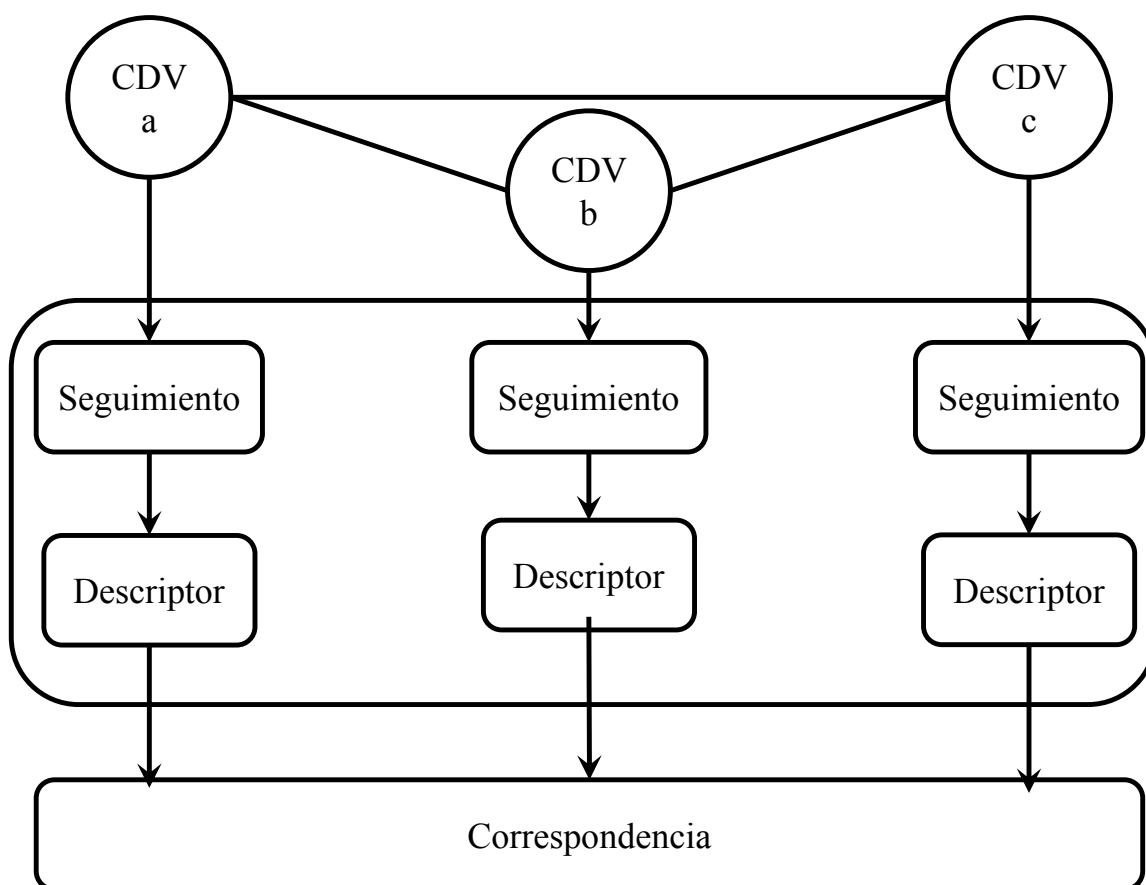


Figura 2.2: Representación gráfica del proceso de re-identificación. En cada CDV se sigue a la persona y se generan los descriptores los cuales son comparados para determinar coincidencias.

2.1. Estado del arte

En la literatura se han propuesto varios enfoques de re-identificación de personas. En (Doretto, Sebastian, Tu, y Rittscher, 2011; Gong, Cristani, Loy, y Hospedales, 2014; Bedagkar-Gala y Shah, 2014) se presentan los principales retos y requerimientos de la re-identificación. En la Tabla 2.1 se resumen los principales métodos encontrados en el estado del arte de la re-identificación de personas. La mayoría de los enfoques procesan vídeo y usan el cuerpo completo para la re-identificación. Aunque la cara no ha sido muy explotado en la re-identificación, ofrece muchas ventajas sobre otras características físicas descriptivas de la persona como por ejemplo que permanece estable por mayor tiempo, es universal y es única para cada persona. A continuación se describe de forma breve algunos de los enfoques en la Tabla 2.1.

En (hsun Chang y Gong, 2001) se presentó un sistema multi-cámara basado en el método de fusión modalidad Bayesiana (Toyama y Horvitz, 2000) para seguir múltiples personas en un ambiente interior. En dicho trabajo se usan redes bayesianas para el seguimiento de la persona. Cuando el sistema detecta una nueva persona le asigna una identidad y lo sigue dentro del campo de vista. Si la persona ya ha aparecido en otros campos de vista o se ha perdido la identidad durante el seguimiento, el sistema pasa la identidad al CDV actual y lo re-asigna al nuevo sujeto mediante una comparación con previas re-identificaciones. Este algoritmo usa la información del cuerpo completo mediante la detección de *blobs* que se mueven. Este método fue evaluado con 20 secuencias de vídeo obteniendo una exactitud de 99.1 % con una desviación de 2.4 % bajo condiciones de iluminación y oclusión. En (Javed y cols., 2003) se explota la redundancia en las rutas que las personas tienden a seguir, por ejemplo, caminos y corredores. La correspondencia se establece mediante los datos generados por el movimiento y la apariencia de la persona. Este enfoque requiere de un proceso de entrenamiento, y después la correspondencia se establece mediante parámetros aprendidos y actualizados con los cambios de patrones de trayectorias. En (Mittal y Davis, 2003) se presenta un enfoque que segmenta, detecta y sigue múltiples personas en una red de cámaras. Para la re-identificación se usa la evidencia recolectada de muchos pares de cámaras. Un aspecto importante de este enfoque es que es capaz de encontrar puntos 3D en una persona conocida solo por una proyección de la misma en dos vistas, luego el algoritmo de segmentación usa un clasificador bayesiano y realiza un análisis de oclusión para combinar evidencias de diferentes pares de cámaras.

En (Gandhi y Trivedi, 2007) cada persona es seguida en múltiples cámaras y su posición en el piso se calcula mediante triangulación. Un mapa panorámico se crea usando la geometría de la cámara y la ubicación de la persona. Cada píxel en el mapa obtiene la información del color de la cámara que lo observa. La correspondencia entre diferentes seguimiento se obtiene mediante la comparación de estos mapas. En (Hamdoun y cols., 2008) se propone un esquema que usa el acoplamiento de firmas basados en descriptores de puntos de interés recolectados sobre secuencias de vídeos cortos. Los principales pasos de este esquema son: 1) construcción del modelo, donde se genera una firma para cada persona detectada; 2) construcción de consultas; 3) comparación de descriptores;

4) acoplamiento rápido y robusto; y 5) identificación de la persona mediante la comparación de los puntos de interés. En (Berdugo y cols., 2010) se presenta un enfoque de dos etapas. Primero, la creación de la base de datos del modelo de apariencia. La entrada en esta etapa es un vídeo, de la cual se segmenta y sigue a la persona para extraer sus características (color y textura) para almacenarlas en la base de datos. En la segunda etapa, la persona es detectada y seguida para extraer sus características las cuales se comparan con las previamente almacenadas. En el proceso de comparación se determina si existe o no una coincidencia. En (Mazzon y cols., 2012) se propone un método que considera la apariencia de la gente, la localización espacial de las cámaras y las rutas potenciales que una persona puede seguir.

En muchos escenarios no es posible obtener la apariencia del cuerpo completo de la persona, por lo tanto se han propuesto enfoques que solo usan partes del cuerpo humano para la re-identificación. En (Utsumi y Tetsutani, 2004) se propone un enfoque que emplea la apariencia de la cabeza de la persona para el seguimiento en múltiples cámaras. La apariencia de la cabeza se modela dinámicamente usando observaciones de varios CDV's. Se usa principalmente la información de color los cuales se almacenan como un conjunto de parches de color distribuidos en un espacio 3D. Para el proceso de acoplamiento, primero se calcula la posición de la cabeza a través del modelo 3D. Luego, el modelo 3D es re-escalado para acoplarse con el modelo 2D. Entonces, el modelo 3D es examinado para encontrar coincidencias. Cuando se encuentra una coincidencia, el modelo 3D se actualiza usando la imagen detectada. La cara humana se caracteriza por ser única para cada persona, universal y es cuantificable. Por esta razón, puede ser usada para la re-identificación robusta de personas en periodos de tiempos largos, por ejemplo, días, semanas o incluso meses. Esto debido a su poca variabilidad en cuando a su forma, aunque su apariencia cambia con el tiempo debido a la edad. En (Bauml y cols., 2010) se usan las características de la apariencia facial para la re-identificación de la persona. Este enfoque es capaz de tratar con caras tan pequeñas como 18×18 píxeles y con rotación fuera de plano de hasta 60° . Primero se detecta la cara, la cual es dividida en bloques de 8×8 . Sobre cada bloque se aplica la transformada de Coseno discreto (TCD) y luego se seleccionan los bloques más discriminatorios para construir el vector de características de la apariencia facial (descriptor facial). Se usa la máquina de vector de soporte (MVS) como clasificador. En (Fischer y cols., 2010) se presenta un enfoque de re-identificación facial en series de televisión dentro del contexto de recuperación de vídeo. La cara es seguida en cada vídeo para recolectar muestras que mejoren el modelo facial. Luego, se usa un algoritmo de reconocimiento para encontrar caras que coincidan con el modelo. Este enfoque ofrece un buen desempeño debido a que procesa vídeos con buena calidad en cuanto a iluminación y nitidez de las imágenes. Otro trabajo basado en la imagen facial y en el contexto de indexado de vídeo es el que se presenta en (Marsico y cols., 2012). Primero se hace un seguimiento de la cara para la recolección de muestras faciales, a los cuales se les extraen las principales características para la clasificación de la persona. En (Satta y cols., 2012) se usan múltiples componentes y un modelo de subdivisión del cuerpo dentro de un vector de disimilitudes. Otros rasgos biométricos comunes han sido usados en la re-identificación tales como la oreja (Chan y Kumar,

2011; Prakash y Gupta, 2012), esclerótica (Crihalmeanu y Ross, 2011), patrones de las venas de la mano (Kang, 2012) y la dinámica de la forma de caminar (Roy y cols., 2012).

Como se mencionó previamente, uno de los principales aspectos a considerar es la construcción de un descriptor invariante a los cambios de apariencia que sufre la persona en un ambiente no restringido. En (Ayedi y cols., 2011) se propone que el descriptor de cualquier objeto se construya mediante una imagen de covarianza y una estructura de datos *quadtree*. Ésta es una estructura usada para representación de imágenes multi-escala, que contiene información de posición de las características de color y forma. Los trabajos en (Prosser y cols., 2010) y (Chellappa y cols., 2011) abordan también el problema de la escala, el primero propone el *Ensemble Rank* basado en la MVS, mientras que el segundo resalta los aspectos principales en el reconocimiento de caras a distancia (10-250 m. de distancia entre la cámara y la persona). Los investigadores encontraron que las imágenes adquiridas remotamente son afectadas por las variaciones en iluminación, desenfoque, oclusión y variaciones en pose. En (Bedagkar-Gala y Shah, 2011) se propone el diseño de un descriptor usando el color de la ropa y características faciales. Este descriptor se construye sobre una secuencia de imágenes de un individuo, capturando de esta forma las características de las apariencias así como su variación sobre el tiempo. Uno de los principales problemas que se enfrenta en el reconocimiento facial es el pose. Por tal motivo, (Gupta y cols., 2012) proponen un método para sintetizar un pose de una imagen. Esto con el fin de predecir la imagen facial con un error mínimo en un pose deseado de un pose dado. Esto es frecuentemente requerido en aplicaciones como películas animadas, ciencias forenses y generación 3D de la geometría facial.

En (W. Li y cols., 2015) se propone un modelo de acoplamiento basado en un conjunto de muestras. Por otra parte, en (Shi y cols., 2015) se propone un modelo de apariencia multi-nivel, correspondencia adaptativa y fusión. En (Nanni y cols., 2015) primero se extrae la información de textura y color de la imagen de la persona. Entonces, diferentes descriptores son comparados usando medidas de distancia.

Como se puede observar en la Tabla 2.1, la cara y el cuerpo completo son los rasgos biométricos más usados en la re-identificación de personas. En (Nappi y Wechsler, 2012) se presenta una revisión del campo de la re-identificación y plantea rutas prometedoras para la investigación y el desarrollo. Además, proponen una arquitectura que consiste en siete módulos usando aprendizaje semi-supervisado y transducción. La transducción es método que analiza la estabilidad del aprendizaje usando tanto entrenamiento y datos de prueba.

2.2. Resumen

En este capítulo se ha visto que la re-identificación de personas requiere de al menos tres tareas: 1) detección, 2) identificación y 3) seguimiento. También, para que una persona pueda ser reconocida se requiere la construcción de un descriptor que sintetice las características generales y específicas de la persona. Y por último, una técnica que

acople los descriptores que se generan en cada CDV para determinar la correspondencia. La mayoría de los métodos reportados en la literatura usan la imagen completa del cuerpo humano. Sin embargo, no siempre es posible obtener una imagen completa por lo que se han propuesto enfoques que usan solo alguna de las partes superiores como la cabeza, cara, ojos, entre otros.

En este trabajo de investigación se emplea la cara para la re-identificación debido a que permanece estable por más tiempo y que la identidad de la persona es obtenible a través de un análisis facial. También, se explota la topología de la red de cámaras y el patrón de movimiento de la persona para predecir el campo de vista destino de la persona. Como descriptores se emplean los filtros de correlación debido a su robustez al usar tanto la forma como el contenido del objeto.

Tabla 2.1: Métodos de re-identificación de personas reportados en la literatura. Siglas en la tabla: campos de vista disjuntos (CVD), campos de vista superpuestos (CVS), vídeo (V) y cuerpo completo (CC).

| Referencia | CVD | CVS | V | Img. | Cara | CC |
|-----------------------------------|-----|-----|---|------|------|----|
| hsun Chang y Gong (2001) | × | | × | | | × |
| Javed y cols. (2003) | × | | × | | | × |
| Mittal y Davis (2003) | × | | × | | | × |
| Utsumi y Tetsutani (2004) | × | | × | | | |
| Gandhi y Trivedi (2007) | × | | × | | | × |
| Hamdoun y cols. (2008) | × | | × | | | × |
| Bauml y cols. (2010) | × | | × | | × | |
| Berdugo y cols. (2010) | × | | × | | | × |
| Fischer y cols. (2010) | × | | × | | × | |
| Prosser y cols. (2010) | × | | × | | | × |
| Ayedi y cols. (2011) | | | × | | | × |
| Bedagkar-Gala y Shah (2011) | × | | × | | × | × |
| Chan y Kumar (2011) | | × | | × | | |
| Chellappa y cols. (2011) | × | | × | | × | |
| Crihalmeanu y Ross (2011) | | × | | × | | |
| Doretto y cols. (2011) | × | | × | | | × |
| Gupta y cols. (2012) | | × | | × | × | |
| Kang (2012) | | × | | × | | |
| Marsico y cols. (2012) | × | | × | | × | |
| Mazzon y cols. (2012) | × | | × | | | × |
| Nappi y Wechsler (2012) | × | | × | | × | |
| Prakash y Gupta (2012) | | × | | × | | |
| Roy y cols. (2012) | | × | × | | | |
| Satta y cols. (2012) | × | | × | | | × |
| An, Kafai, Member, y Bhanu (2013) | × | | × | | × | |
| An, Kafai, Yang, y Bhanu (2013) | × | | × | | | × |
| X. Chen y cols. (2013) | × | | × | | | × |
| Soori y cols. (2013) | × | | × | | | × |
| C. Liu y cols. (2013) | × | | × | | | × |
| Ma y cols. (2014) | × | | × | | × | × |
| Kenk y cols. (2014) | × | | × | | | × |
| Gong y cols. (2014) | × | | × | | | × |
| Bedagkar-Gala y Shah (2014) | × | | × | | | × |
| Karaman y cols. (2014) | × | | × | | | × |
| Zhou y cols. (2014) | × | | × | | | × |
| W. Li y cols. (2015) | × | | × | | | × |
| Shi y cols. (2015) | × | | × | | | × |
| Nanni y cols. (2015) | × | | × | | | × |

Capítulo 3

Reconocimiento de patrones por correlación

La correlación es una técnica de reconocimiento de patrones usada en aplicaciones tales como reconocimiento automático de objetivos, reconocimiento biométrico, reconocimiento de caracteres ópticos, detección y seguimiento de objetos. Esta técnica ha causado interés en los investigadores debido a las siguientes propiedades: invariante a desplazamiento (Vijaya-Kumar, Mahalanobis, y Juday, 2005), tolerante a ruido (Savvides, Bhagavatula, hui Li, y Abiantun, 2007), degradación suavizada de la respuesta a las oclusiones (Naresh-Boddeti, 2012), habilidad para simultáneamente localizar y clasificar objetos (Rodriguez-Perez, 2012), puede ser implementada fácilmente en el dominio de la frecuencia mediante la transformada de Fourier (Rodriguez-Perez, 2012; Bolme, 2011) y puede ser entrenada tanto de forma *en línea* como *fuera de línea* (Savvides, Venkataramani, y Kumar, 2003). Aunque para el diseño de los filtros se usan las imágenes completas, pueden ser fácilmente extendidos al espacio de las características tal y como se propone en (Zhu, Liao, Lei, Liu, y Li, 2007). También, se pueden emplear las imágenes LBPi (Local Binary Pattern image) (B. Yang y Chen, 2013) que contienen características tanto locales como globales para la extensión de los filtros al espacio de las características. Otra ventaja de la correlación es que puede ser implementada ópticamente o digitalmente (Vijaya-Kumar y cols., 2005). En (Vijaya-Kumar y cols., 2005; Savvides y cols., 2007) se puede encontrar suficiente teoría sobre la implementación digital, mientras que en (Alfalou y Brosseau, 2010) se describen los últimos desarrollos en la correlación óptica.

El reconocimiento de patrones por correlación está basado en la selección o creación de una señal de referencia y entonces determinar el grado con el cual el objeto bajo análisis se acopla con la señal de referencia (Vijaya-Kumar y cols., 2005). El grado de acoplamiento o similitud es una simple estadística sobre el cual se basa la decisión acerca del objeto. La Figura 3.1 muestra el proceso gráfico de la correlación, donde un conjunto de entrenamiento es usado para sintetizar un filtro \mathbf{h} en el dominio de Fourier. Luego, la transformada de Fourier (TF) de la imagen de entrada es multiplicada elemento a elemento con el filtro. A este producto se le aplica la inversa de la TF (ITF) para obtener

el plano de correlación $g(x, y)$. Finalmente, el análisis del *pico* (valor más grande sobre $g(x, y)$) permite tomar una decisión sobre el objeto.

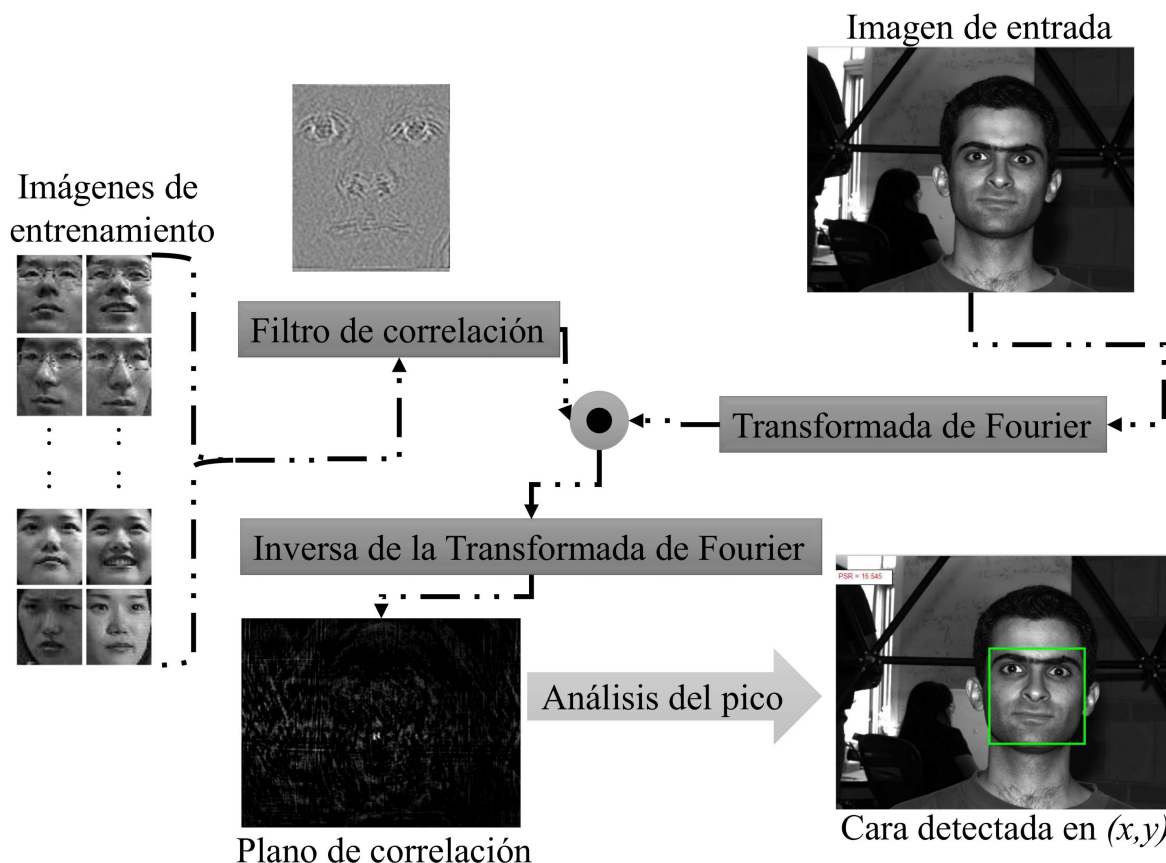


Figura 3.1: Proceso de correlación entre una imagen de entrada y un filtro de correlación en el dominio de Fourier.

El principal reto en la correlación es la construcción del filtro de correlación, el cual se construye usando un conjunto de imágenes de entrenamiento. El filtro de correlación \mathbf{h} es comparado con una imagen de entrada $f(x, y)$ mediante una correlación cruzada dada por:

$$g(x, y) = ITF\{F(u, v)H^*(u, v)\}, \quad (3.1)$$

donde $F(u, v)$, $H(u, v)$ y $*$ son respectivamente la TF de $f(x, y)$, la TF de $h(x, y)$ y la operación de complejo conjugado. La Figura 3.2 muestra los planos $g(x, y)$ de correlación para objetos verdaderos y objetos falsos. Si la imagen de entrada y el filtro son similares, entonces el plano de correlación debe exhibir un pico agudo y alto como el que se muestra en el plano izquierdo. En caso contrario, $g(x, y)$ no exhibe ningún pico sobresaliente como el plano de la derecha.

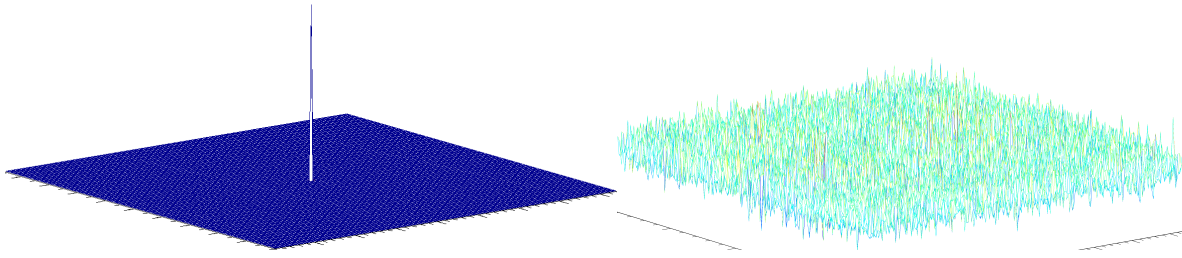


Figura 3.2: Planos de salida de correlación para un objeto verdadero (izquierda) y para un objeto falso (derecha).

El grado de similitud se obtiene analizando el *pico*. Algunas de las métricas más usadas son el *Peak to Side-lobe Ratio (PSR)* y la *capacidad de discriminación (CD)*. La métrica PSR mide el número de desviaciones estándar en el cual el pico se encuentra sobre el valor medio en $g(x, y)$, y está dado por:

$$PSR = \frac{(pico - \mu_{area})}{\sigma_{area}}, \quad (3.2)$$

donde $\mu_{\text{área}}$ y $\sigma_{\text{área}}$ son respectivamente la media y la desviación estándar del área o vecindad alrededor, pero sin incluirlo, del pico (Véase Figura 3.3). Los enfoques reportados en la literatura generalmente calculan el *PSR* sobre una área con resolución de 20×20 píxeles y de 5×5 píxeles para el área del pico.

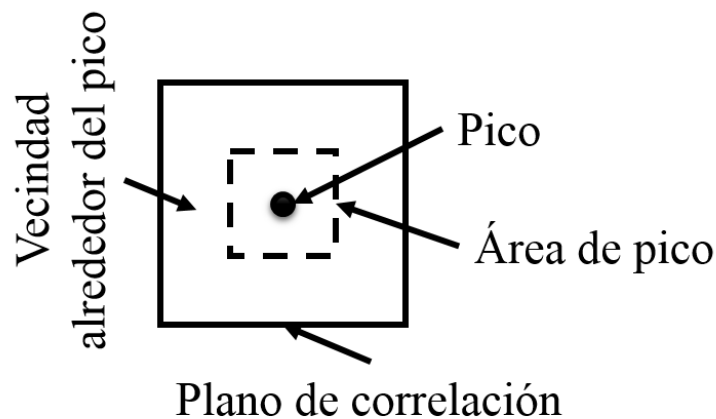


Figura 3.3: Plano de correlación en donde se delimitan el área del pico y su alrededor para el cálculo de *PSR*.

Por otra parte, la capacidad de discriminación (*CD*) mide la habilidad de un filtro para distinguir un objeto de otros diferentes y está dada por:

$$CD = 1 - \frac{C^R(0, 0)^2}{C^O(0, 0)^2} \quad (3.3)$$

donde $C^R(0,0)^2$ y $C^O(0,0)^2$ son los valores máximos del área de rechazo y área del objeto, respectivamente. En la Figura 3.4 se muestra un plano de correlación que delimita el área de rechazo y el área del objeto. Como se puede notar en las Ecuaciones 3.2 y 3.3, PSR no requiere conocer la ubicación del objeto de interés, mientras que en CD si lo requiere.

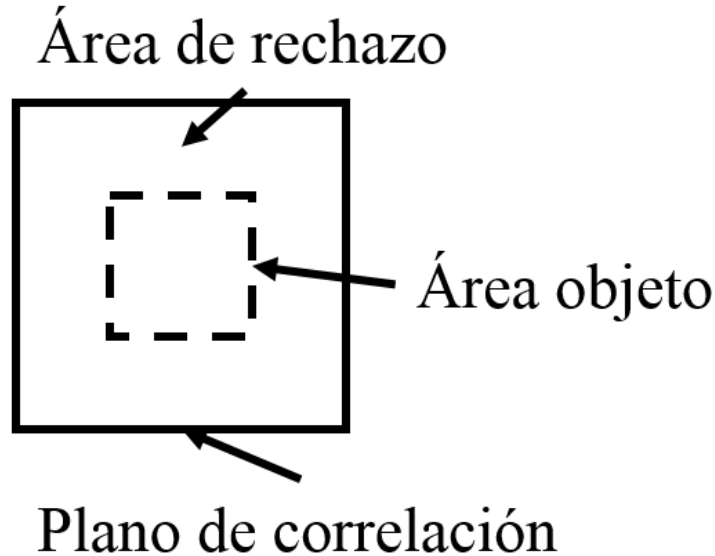


Figura 3.4: Área del plano de correlación sobre el cual se calcula la CD .

La calidad del plano de correlación es un aspecto importante a considerar, pues de ella depende de que se haga una decisión robusta sobre el objeto. En (Alfalou, Brosseau, Katz, y Alam, 2012; Elbouz, Alfalou, Brosseau, Yahia, y Alam, 2015) se propone pre-procesar $g(x, y)$ antes de aplicar alguna métrica. Este pre-procesamiento tiene por objetivo eliminar el ruido y minimizar los picos falsos mientras se maximiza el pico verdadero.

3.1. Acoplamiento de plantilla

En esta tesis una plantilla es una imagen de entrenamiento básica. El acoplamiento de plantillas es una técnica de reconocimiento de patrones que consiste en obtener la similitud mediante la comparación píxel a píxel entre una plantilla y una imagen de entrada. Sea $h(x, y)$ una plantilla y $f(x, y)$ una imagen, la similitud entre estas imágenes se calcula mediante el producto punto de los valores de los píxeles:

$$s(f(x, y), h(x, y)) = f(x, y) \cdot h(x, y) = \sum_x^d \sum_y^d f(x, y)h(x, y), \quad (3.4)$$

donde d es la cantidad de píxeles en las imágenes. Si ambas imágenes son similares, entonces el producto punto resultará en un valor grande. En caso contrario, el producto punto resultará en un valor pequeño. Por lo tanto, la decisión acerca del objeto puede determinarse comparando $s(\cdot)$ con un valor umbral τ previamente calculado. Cuando la imagen de entrada es más grande que la plantilla, entonces el cálculo de la similitud se hace de la siguiente manera. La plantilla es escaneada de izquierda a derecha y de arriba hacia abajo. En cada desplazamiento, con pasos de un píxel, se calcula para cada posición (x, y) el valor de similitud mediante la Ecuación 3.4.

Mediante el teorema de convolución, esta operación de escaneo se convierte en una multiplicación elemento a elemento en el dominio de la frecuencia mediante Fourier. El filtro de correlación más básico es el filtro de acoplamiento (FA) (North, 1963), el cual es óptimo para detectar una imagen de entrada conocida con ruido aditivo. Sea $r(x, y)$ una señal de referencia, el FA para el caso de ruido blanco está dado por:

$$H^*(u, v) = \alpha R(u, v), \quad (3.5)$$

donde $R(u, v)$ es la TF de $r(x, y)$, el símbolo $*$ es el complejo conjugado y α es cualquier constante compleja. El plano de correlación entre el FA y $s(x, y)$ está dada por:

$$g(x, y) = ITF\{S(u, v)H^*(u, v)\}, \quad (3.6)$$

donde $S(u, v)$ es la TF de $s(x, y)$. La Figura 3.5 muestra el plano de correlación de la operación en la Ecuación 3.6. La salida muestra un pico alto para el objetivo, pero también contiene varios picos falsos alrededor, incluso algunos más altos que el verdadero, producidos por otros objetos en la escena y por la iluminación intensa en la parte izquierda de la imagen. Esto evidencia la alta sensibilidad a la luz de los FA's. La presencia de picos falsos alrededor del pico verdadero no es deseable, ya que dificultan la localización exacta del objeto verdadero sobre la escena.



Figura 3.5: Resultado de correlacionar una imagen de entrada con una plantilla simple.

En resumen, los filtros de acoplamiento tienen las siguientes desventajas (Vijaya-Kumar, 1992):

- Son muy sensibles a cambios pequeños en la señal de referencia o imagen y
- Son ineficientes a la luz debido a que su magnitud de respuesta es mucho menor que uno en muchas frecuencias.

El FA tampoco se puede usar en el reconocimiento de múltiples objetos. Por ejemplo, si queremos reconocer las letras del alfabeto latino, entonces tendríamos que construir un FA por cada letra. Esto significa que para reconocer una letra, tenemos que realizar 26 correlaciones, lo cual resulta muy costoso y poco práctico en aplicaciones reales. Una forma de lograr el reconocimiento de objetos invariante a distorsiones es sintetizar un filtro usando varias perspectivas del objetivo. Este tipo de filtros se le conoce como filtros de correlación compuesto, y para su diseño se han propuesto varios algoritmos (Vijaya-Kumar, 1992; Vijaya-Kumar y cols., 2005; Savvides y cols., 2007). A continuación se describen y analizan algunos filtros de correlación compuestos que han sido usados en el procesamiento facial.

3.2. Filtros de correlación compuestos

Los filtros de correlación compuestos, o simplemente filtros compuestos (FC), fueron desarrollados para manejar distorsiones mediante la combinación de imágenes de entrenamiento. Este conjunto de entrenamiento contiene las principales distorsiones que se espera presente el objetivo. Existen diferentes algoritmos para sintetizar un FC capaz de reconocer objetos con distorsiones similares a los de las imágenes de entrenamiento. En esta sección se describen y analizan algunos FC que han sido aplicados en el procesamiento facial.

3.2.1. Funciones discriminantes sintéticas

Uno de los primeros enfoques propuestos para incorporar invariancia a distorsiones son las funciones discriminantes sintéticas (SDF, por sus siglas en inglés) (Casasent y Chang, 1986). SDF tiene como objetivo reconocer imágenes similares a las plantillas de entrenamiento, así como producir un pico fuerte para el objeto verdadero. Sea $T = \{f_1(x, y), f_2(x, y), \dots, f_N(x, y)\}$ el conjunto de N imágenes de entrenamiento. Cada plantilla $f_i(x, y) \in T$ se reorganiza en un vector columna \mathbf{x}_i de manera lexicográfica, es decir, de izquierda a derecha y de arriba hacia abajo. Luego, cada vector \mathbf{x}_i se agrega como columna a una matriz $X = [\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N]$. Así, un filtro SDF está dado por:

$$\mathbf{h} = X(X^+X)^{-1}\mathbf{u}, \quad (3.7)$$

donde $+$ es la transpuesta conjugada, -1 es la inversa de la matriz y \mathbf{u} es un vector que contiene los valores de correlación esperados en la salida de correlación para cada imagen

de entrenamiento. Generalmente, el vector \mathbf{u} contiene valores igual a 1 para objetos de la clase verdadera y valores igual a 0 para objetos de la clase falsa. El filtro SDF ha sido el punto de comparación o de partida para el desarrollo de nuevos algoritmos para la síntesis de nuevos filtros de correlación.

3.2.2. Filtrado no lineal

Aunque el filtro SDF es invariante a las distorsiones para las cuales fue entrenado, es muy sensible al ruido aleatorio y a la iluminación intensa, así como que el pico producido es muy amplio. La técnica de filtrado no lineal se aplica a los filtros compuestos como SDF para mejorar su desempeño en términos de capacidad de discriminación contra objetos similares al objetivo, agudeza del pico de correlación y robustez al ruido. Para esto, primero se aplica la transformación no lineal como la *k-ésima ley* a cada uno de las TF's de las imágenes de entrenamiento:

$$F_i^k(u, v) = |F_i(u, v)|^k \exp(i * \varphi_{F_i(u, v)}), \quad (3.8)$$

donde $|\cdot|$ y φ son el módulo y la fase de la TF, respectivamente. El factor de no linealidad $0 < k < 1$ eleva a la k -ésima potencia la magnitud mientras mantiene intacta la información de la fase que contiene la información más discriminante del objeto. Luego, estas imágenes filtradas se pueden usar en la síntesis de cualquier filtro compuesto, como el filtro SDF descrito previamente:

$$\mathbf{h} = \hat{X}(\hat{X}^+ \hat{X})^{-1} \mathbf{u}, \quad (3.9)$$

donde \hat{X} es la matriz cuyas columnas corresponden al vector columna de las imágenes filtradas por la Ecuación 3.8. La Figura 3.6 muestra el resultado del filtro SDF con el filtrado no lineal. Como se puede observar, se produce un pico fuerte y agudo a diferencia del obtenido por el filtro de acoplamiento en la Figura 3.5. La iluminación intensa ya no generó picos falsos ni los otros objetos en la escena, esto evidencia la alta capacidad de discriminación que se incorpora en los filtros compuestos mediante el filtrado no lineal.

3.2.3. Filtro del mínimo promedio de la energía de correlación

El pico producido por los FA se degrada rápidamente con distorsiones geométricas de la imagen. Como se mencionó previamente, los filtros SDF producen un pico fuerte pero amplio que dificulta la detección del objeto. Por tal motivo, el filtro del mínimo promedio de la energía de correlación (MACE, por sus siglas en inglés) (Mahalanobis, Kumar, y Casasent, 1987) tiene como objetivo producir un pico agudo y con valores de pico controlados como en el filtro SDF.

EL filtro MACE elimina los lóbulos laterales para producir distintos picos agudos. Una forma de lograr esto es minimizando la energía del plano de correlación, la cual

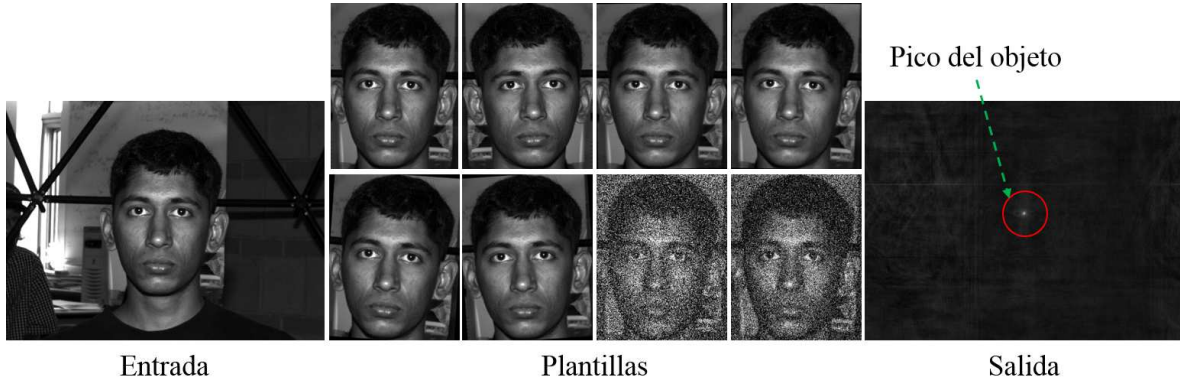


Figura 3.6: Resultado de correlacionar una imagen de entrada con un filtro SDF no lineal.

posteriormente suprimirá los lóbulos laterales. La minimización es posible con la función Average Correlation Energy (ACE), que para el conjunto T está dado por:

$$ACE = \frac{1}{N} \sum_{i=1}^N \sum_{x=1}^{d_1} \sum_{y=1}^{d_2} |g_i(x, y)|^2, \quad (3.10)$$

donde d_1 y d_2 corresponden a las dimensiones del plano de correlación $g_i(x, y)$. En el dominio de Fourier, la Ecuación 3.10 queda de la siguiente manera:

$$ACE = \frac{1}{N \cdot d} \sum_{i=1}^N \sum_{u=1}^{d_1} \sum_{v=1}^{d_2} |G_i(u, v)|^2, \quad (3.11)$$

donde $d = d_1 \times d_2$ y $G_i(u, v)$ es la TF de $g_i(x, y)$. Como $G_i(u, v)$ es el resultado de la correlación entre una imagen $f_i(x, y)$ y el filtro MACE $h(x, y)$, la Ecuación 3.11 puede reescribirse como:

$$ACE = \frac{1}{N \cdot d} \sum_{i=1}^N \sum_{u=1}^{d_1} \sum_{v=1}^{d_2} |F_i(u, v)|^2 |H(u, v)|^2. \quad (3.12)$$

La notación vectorial de ACE está dada por:

$$ACE = \mathbf{h}^+ D \mathbf{h}, \quad \text{donde } D = \frac{1}{N \cdot d} \sum_{i=1}^N X_i X_i^*. \quad (3.13)$$

El objetivo es minimizar ACE mediante el filtro MACE dado por:

$$\mathbf{h} = D^{-1} X (X^+ D^{-1} X)^{-1} \mathbf{u}. \quad (3.14)$$

El filtro MACE es óptimo para reconocer objetos afectados por ruido y genera un pico agudo y fuerte que facilita la detección del objeto. Como se puede ver en la Ecuación

3.14, se requiere de dos inversiones de matrices lo cual es costoso computacionalmente hablando.

3.2.4. Filtro del máximo promedio de la altura de correlación

Un aspecto importante a considerar es cómo incorporar a un filtro de correlación la invariancia a distorsiones. Para lograr esto, primero es necesario cuantificar la cantidad de distorsiones contenidas en T usando la función Average Similarity Measure (ASM), la cual es dado por:

$$ASM = \mathbf{h}^+ S \mathbf{h}, \quad (3.15)$$

donde:

$$S = \frac{1}{N \cdot d} \sum_{i=1}^N (X_i - \bar{X})(X_i - \bar{X})^*. \quad (3.16)$$

ASM mide la variación promedio entre un conjunto de planos de correlación expresado como una función del filtro y las imágenes de entrenamiento en T . Sin embargo, mientras minimizamos las distorsiones, es deseable maximizar los valores del pico. Esto se logra mediante el Average Correlation Height (ACH), el cual está dado por:

$$ACH = \mathbf{m}^+ \mathbf{h}, \quad (3.17)$$

donde \mathbf{m} es un vector que contiene el promedio de las imágenes de entrenamiento. Además de minimizar ASM y maximizar ACH, es deseable suprimir los efectos de la variación de la salida del filtro debido al ruido. Esto se logra mediante la función Output Noise Variance (ONV), la cual está dado por:

$$ONV = \mathbf{h}^t C \mathbf{h}, \quad (3.18)$$

donde C es la matriz de covarianza del ruido de entrada. La meta es suprimir ASM y ONV mientras se maximiza ONV. Esto se convierte en un problema de minimización-maximización que puede ser resuelto usando el cociente de Rayleigh:

$$\begin{aligned} J(\mathbf{h}) &= \frac{|ACH|^2}{ASM + ONV} \\ &= \frac{|\mathbf{m}^+ \mathbf{h}|^2}{\mathbf{h}^+ S \mathbf{h} + \mathbf{h}^t C \mathbf{h}} \\ &= \frac{|\mathbf{m}^+ \mathbf{h}|^2}{\mathbf{h}^+ (S + C) \mathbf{h}}. \end{aligned} \quad (3.19)$$

Finalmente, el filtro del máximo promedio de la altura de correlación (MACH, por sus siglas en inglés) está dado por (Mahalanobis, Vijaya-Kumar, Song, Sims, y Epperson, 1994):

$$\mathbf{h} = \alpha(S + C)^{-1}m, \quad (3.20)$$

donde $0 < \alpha < 1$ es un coeficiente de normalización.

3.2.5. Función discriminante sintética de compromiso óptimo

La función discriminante sintética de compromiso óptimo (OTSDF, por sus siglas en inglés) (Kumar, Carlson, y Mahalanobis, 1994) se obtiene minimizando una suma ponderada de ACE y ONV. El filtro OTSDF está dado por:

$$\mathbf{h} = (\alpha D + (\sqrt{1 - \alpha^2})C)^{-1}X[X^+(\alpha D + (\sqrt{1 - \alpha^2})C)^{-1}X]^{-1}\mathbf{u}, \quad (3.21)$$

donde $0 < \alpha < 1$ establece el nivel de compromiso entre el ruido y la agudeza del pico. El filtro OTSDF busca mantener un compromiso entre la agudeza del pico y la tolerancia al ruido.

3.2.6. Función discriminante sintética de compromiso óptimo no restringido

El filtro OTSDF no restringido (UOTSDF, por sus siglas en inglés) (Kumar y cols., 1994) también produce picos fuertes y agudos para imágenes afectadas por ruido. El filtro UOTSDF está dado por:

$$h = (\alpha D + \sqrt{1 - \alpha}C)^{-1}. \quad (3.22)$$

Este filtro es adecuado para reconocer imágenes afectadas por ruido en condiciones de baja iluminación.

3.2.7. Filtro compuesto no lineal no restringido

El filtro compuesto no lineal no restringido (UNCF, por sus siglas en inglés) (Santiago-Ramírez, González-Fraga, y Lázaro-Martínez, 2012) fue diseñado especialmente para reconocimiento facial. Una imagen facial está compuesta de cinco características faciales que son ojos (izquierdo y derecho), nariz, boca y mejillas. Por esta razón, $f_i(x, y) \in T$ puede ser considerado como la unión de M regiones faciales $r_i(x, y)$, es decir, $f_i(x, y) = \bigcup_{j=1}^M r_j(x, y)$. Cada región facial es considerada como un clasificador débil, por lo tanto, la unión de varios clasificadores débiles crea un clasificador fuerte. Una de las ventajas de usar regiones faciales es que cuando la entrada es una imagen parcial, es muy probable que esta se acople con alguna de las regiones de entrenamiento y así ser reconocida. También, la distorsión en una simple región facial tiene menos influencia sobre el reconocimiento. El filtro UNCF sintetizado con N imágenes faciales, cada uno con M regiones, está dado por:

$$H^*(u, v) = \frac{1}{N \times M} \sum_{i=1}^{N \times M} R_i(u, v), \quad (3.23)$$

donde $R_i(u, v)$ es la TF de $r_i(x, y)$. Como se puede notar, este filtro es un simple promediado que tiene la ventaja de enfatizar aquellas regiones comunes mientras elimina las características idiosincrásicas de una simple región facial.

3.2.8. Filtro promedio de filtros exactos sintéticos

Los filtros de correlación presentados previamente han sido diseñados para la tarea de reconocimiento. En esta tarea, se asume que ya se cuenta con la imagen recortada y alineada del objetivo. Sin embargo, en la tarea de detección se procesa una imagen arbitraria que puede o no contener el objetivo. Por lo tanto, un filtro para detección tiene que sintetizar aquellas características del objeto que le permitan ser detectado en cualquier escena y con fondo complejo. El filtro promedio de filtros exactos sintéticos (ASEF, por sus siglas en inglés) (Bolme, Draper, y Beveridge, 2009) logra este objetivo promediando una gran cantidad de filtros exactos.

Sea $f_i(x, y)$ una imagen de entrenamiento y $g_i(x, y)$ su salida de correlación esperada, la cual puede ser generada sintéticamente mediante una función Gaussiana como sigue:

$$g_i(x, y) = \exp\left(-\frac{(x - x_i)^2 + (y - y_i)^2}{\sigma^2}\right), \quad (3.24)$$

donde σ^2 es el radio de la Gaussiana en el centro del objeto. Cada par $f_i(x, y)$ y $g_i(x, y)$ es usado para construir un filtro exacto:

$$H_i^*(u, v) = \frac{G_i(u, v)}{F_i(u, v)}, \quad (3.25)$$

donde la división es elemento a elemento. $G_i(u, v)$ y $F_i(u, v)$ son las TF's de $g_i(x, y)$ y $f_i(x, y)$, respectivamente. Promediando todos los filtros exactos obtenemos el filtro ASEF (Bolme y cols., 2009):

$$H(u, v) = \frac{1}{N} \sum_{i=1}^N H_i^*(u, v). \quad (3.26)$$

El filtro ASEF es inestable cuando se usan pocas imágenes debido a que la división elemento a elemento en la Ecuación 3.26 es inestable cuando la frecuencia en la imagen de entrenamiento contiene muy poca energía. Por lo que el promediado de una gran cantidad de imágenes resuelve este problema.

3.2.9. Filtro de la mínima salida de la suma del error cuadrado

Otra tarea común en el procesamiento de imágenes es el seguimiento de objetos en una secuencia de imágenes. Un filtro adecuado para esta tarea requiere de la habilidad

de adaptarse a los cambios que sufra la apariencia del objeto en presencia de un fondo complejo y otros objetos en la escena. El filtro de la mínima salida de la suma del error cuadrado (MOSSE, por sus siglas en inglés) (Bolme, Beveridge, y Lui, 2010) es un algoritmo para producir filtros parecidos a ASEF pero con menos cantidad de imágenes de entrenamiento. El filtro MOSSE está dado por:

$$H_i^*(u, v) = \frac{A_i}{B_i}, \quad (3.27)$$

donde:

$$A_i = \eta G_i(u, v) \odot F_i^*(u, v) + (1 - \eta)A_{i-1}, \quad (3.28)$$

y:

$$B_i = \eta F_i(u, v) \odot F_i^*(u, v) + (1 - \eta)B_{i-1}, \quad (3.29)$$

donde η es la tasa de aprendizaje. Este factor pone mayor peso en los cuadros recientes y permite que el efecto de los cuadros anteriores decaigan exponencialmente sobre el tiempo. Los creadores de este algoritmo encontraron que $\eta = 0.125$ permite al filtro MOSSE adaptarse rápidamente a los cambios de apariencia mientras mantiene un filtro robusto a distorsiones. Además de la rapidez, los filtros compuestos en el seguimiento de objetos ofrecen algunas ventajas:

- **Actualización en línea.** Algunos filtros como ASEF y MOSSE puede actualizarse en línea. Esto permite que el filtro se adapte a los cambios en la apariencia del objeto.
- **Simplificación del seguimiento.** Usando filtros compuestos en una secuencia de vídeo, solo se requiere de la detección en el primer cuadro. En los posteriores cuadros, la tarea se simplifica a la localización del objeto. Esto reduce el costo y aumenta la velocidad del seguimiento.
- **Recuperación ante fallos.** Si el objeto desaparece de la escena, cuando vuelva a aparecer es detectado nuevamente por el filtro sin necesidad de otro nuevo seguimiento.

3.2.10. Filtro de solo fase segmentado asimétrico

Los filtros mencionados arriba sintetizan todas las imágenes de entrenamiento en una simple plantilla. Así, una imagen de entrada $f(x, y)$ requiere de solo una operación de correlación. Sin embargo, el filtro de solo fase segmentado asimétrico (ASPOF, por sus siglas en inglés) produce un filtro para cada imagen de entrenamiento. Por lo tanto, reconocer $f(x, y)$ requiere varias correlaciones. El filtro ASPOF está dado por (Leonard, Alfalou, y Brosseau, 2012):

- Las imágenes de entrenamiento en T son separadas en dos subclases: una, T_i^1 , que contiene imágenes de entrenamiento con $i = 1, 3, 5, \dots, N - 1$, la otro, T_j^2 con los otros índices $i = 2, 4, 6, \dots, N$.
- Enseguida, se construyen dos filtros segmentados, uno para cada subclase, $H_i^1(u, v)$ and $H_i^2(u, v)$, respectivamente. El filtro $H_i^1(u, v)$ está dado por:

$$H_i^1(u, v) = \begin{cases} F_i^1(u, v) & \text{if } \frac{E_i^1(u, v)}{\sum_{u,v} E_i^1(u, v)} \geq \alpha \frac{E_j^2(u, v)}{\sum_{u,v} E_j^2(u, v)} \\ \text{No asignado} & \text{en otro caso,} \end{cases} \quad (3.30)$$

donde $F_i^1(u, v)$ es la TF de $f_i(x, y) \in T_i^1$ mientras que $E_i^1(u, v)$ es su energía, y la constante α es un coeficiente de asignación de píxel. El filtro $H_i^2(u, v)$ es creado de forma similar.

- Después, los píxeles no clasificados, es decir, aquellos que no satisfacen la Ecuación 3.30, son asignado a uno de los dos filtros segmentados buscando la vecindad más próximos para evitar evitar píxeles aislados.
- Inmediatamente, los filtros $H_i^1(u, v)$ y $H_i^2(u, v)$ son multiplicados por filtros simétricos (Leonard y cols., 2012).
- Finalmente, se suman los dos filtros en el plano de Fourier para obtener el filtro ASPOF:

$$H_i(u, v) = p_1 H_i^1(u, v) + p_2 H_i^2(u, v), \quad (3.31)$$

donde p_1 y p_2 son dos coeficientes.

Este procedimiento genera un filtro ASPOF por cada imagen de entrenamiento en T . Por lo tanto, para una imagen dada, es necesario realizar $|T|$ correlaciones en el peor de los casos. Esto incrementa considerablemente el tiempo de ejecución.

3.3. Resumen

En este capítulo se ha descrito la técnica de reconocimiento de patrones por correlación y algunos de los algoritmos más usados en la literatura para la síntesis de filtros compuestos. Los filtros explotan eficientemente tanto la forma como el contenido del objeto, lo cual permite un procesamiento facial robusto bajo condiciones complejas como el de los ambientes reales. En este trabajo de investigación, los filtros son usados como *descriptores faciales* en la tareas de detección, reconocimiento y seguimiento facial.

Capítulo 4

Procesamiento facial

La cara es la parte frontal de la cabeza que en los humanos se extiende de la frente a la barbilla e incluye la boca, la nariz, las mejillas y los ojos (Webster, 2015). Este rasgo biológico tiene las siguientes propiedades:

- Universalidad. Todas las personas poseen una cara.
- Unicidad. Cada cara es única, incluso aquellas caras muy parecidas tiene algo que las diferencia. Por ejemplo, un lunar, una cicatriz, bello facial, distancia entre las características faciales, entre otros.
- Características medibles. Es posible obtener medidas de las diferentes características faciales, tales como la distancia entre ellas y sus dimensiones.
- Permanencia. La cara permanece estable por un mayor periodo de tiempo, por lo que es preferido para reconocer una persona sobre un periodo de tiempo más largo.

Estas propiedades hacen que la cara pueda ser usado como rasgo biométrico para el reconocimiento de la persona. Los seres humanos realizan el reconocimiento facial con gran habilidad y exactitud bajo ambientes tanto restringidos como no restringidos. Aunque para este reconocimiento, inconscientemente se usan otras características del cuerpo tales como estatura, complexión, color del cabello, entre otros. Sin embargo, el ser humano no puede reconocer a muchas persona al mismo tiempo y requiere tiempo para aprender a reconocer una nueva cara. Por esta razón, se han desarrollado algoritmos que hacen uso de las tecnologías de la visión artificial para automatizar el reconocimiento de la cara.

El reconocimiento facial automatizado tiene una amplia gama de aplicaciones incluyendo seguridad pública, control de acceso, multimedios y entretenimiento. La investigación en el reconocimiento facial inició en la década de 1960. Recientemente se ha tenido progresos significantes en esta área y se ha desplegado un gran número de sistemas de reconocimiento facial en ambientes restringidos (Holdren, 2011). Sin embargo, el reconocimiento facial al igual que el reconocimiento de objetos en general

bajo ambientes no restringidos y con personas no cooperativas es un problema que aún permanece abierto (Gong, McKenna, y Psarrou, 2000; Holdren, 2011; Andreopoulos y Tsotsos, 2013). Esto debido a que la apariencia de la cara sufre cambios que dificultan la recuperación de la información de las características faciales. Tales cambios de apariencia puede deberse a variaciones en iluminación, expresiones, oclusión, ruido, entre otros.

El reconocimiento facial es un conjunto de procesos cooperativos que realizan la detección, seguimiento y el reconocimiento de la cara tal como se muestra en la Figura 4.1. A ese conjunto tareas se les denomina procesamiento facial en esta tesis. En este capítulo se describe en qué consiste cada tarea de procesamiento facial y algunos de los algoritmos mas comunes.

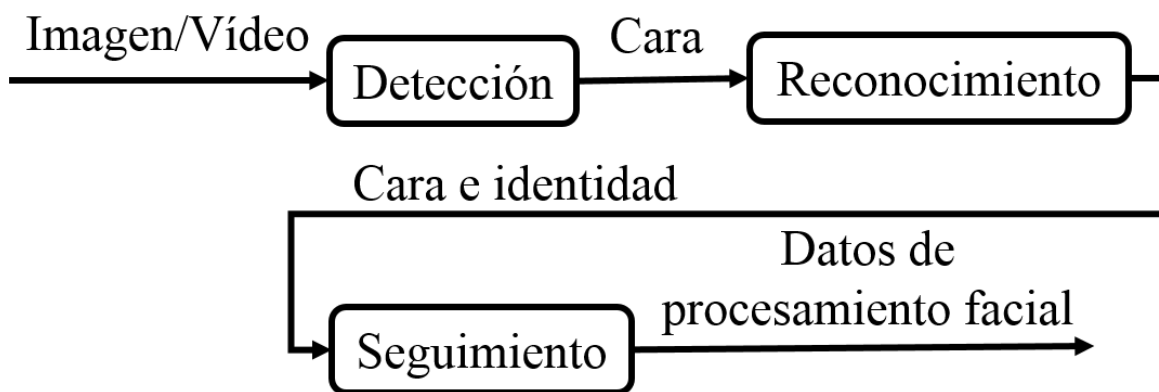


Figura 4.1: Esquema de un sistema de reconocimiento facial como un conjunto de procesos cooperativos.

4.1. Reconocimiento facial

Las caras humanas forman un único tipo de objetos debido a que comparte una estructura común: las características faciales siempre están configuradas en una forma similar (Gong y cols., 2000). En la Figura 4.2 se muestra un ejemplo de clase facial conformada por diferentes imágenes de una misma persona, las cuales presentan pose, expresiones y variaciones en iluminación. La apariencia y la forma de la cara, así como las medidas y distancias de las características faciales, nos permiten reconocer una persona bajo diferentes condiciones ambientales. Por lo tanto, estas condiciones plantean la necesidad de que los algoritmos de reconocimiento sean invariantes a distorsiones faciales.

El reconocimiento facial se clasifica en dos tareas principales: verificación e identificación. En la verificación facial, mostrado en la Figura 4.3, una persona reclama una identidad id y proporciona una muestra facial $f(x, y)$ al sistema. Entonces, el sistema recupera de una base de datos la plantilla biométrica B_i de la identidad reclamada.

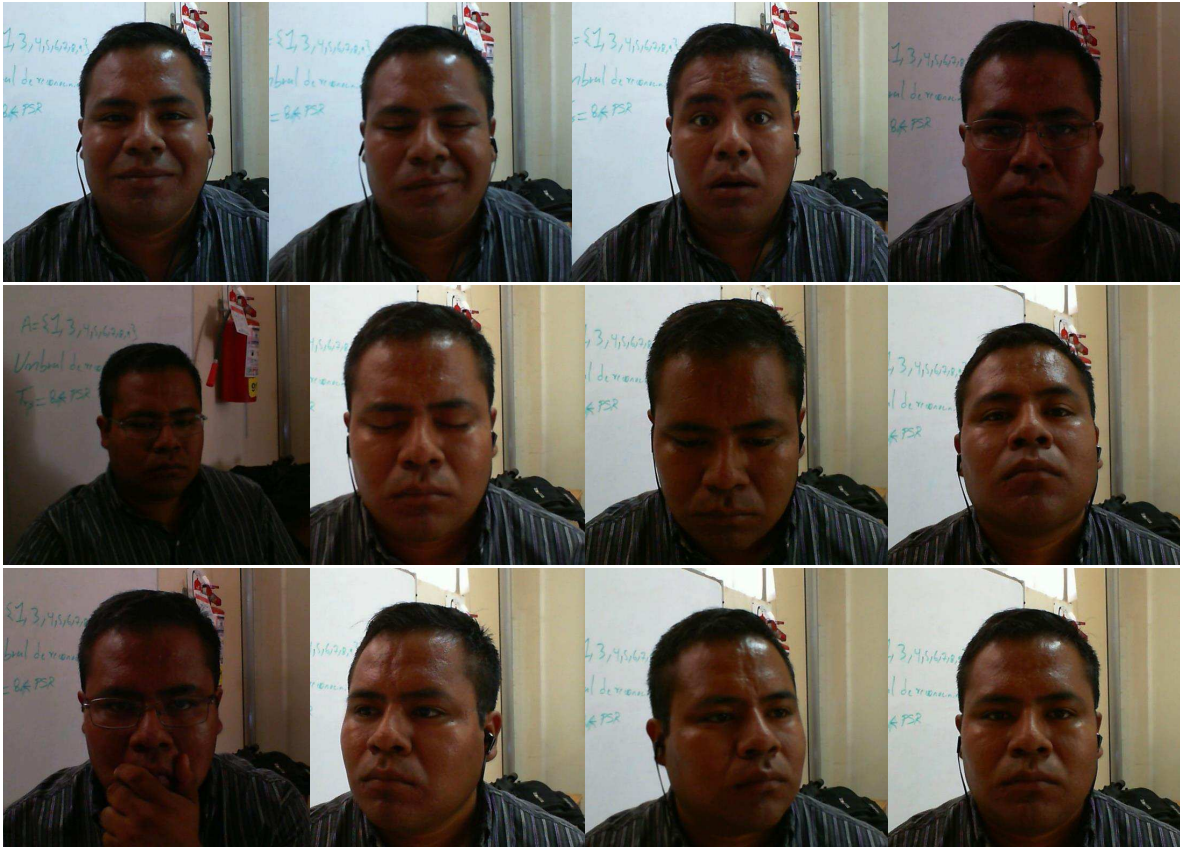


Figura 4.2: Diferentes imágenes de una misma persona que forman una misma clase facial.

Luego, el sistema compara B_i contra $f(x, y)$ para determinar el grado de similitud $s(f(x, y), B)$. Si $s(\cdot)$ es mayor que o igual a un umbral τ , entonces se acepta la identidad *id* reclamada. En caso contrario, se rechaza la identidad y se declara a la persona como un impostor. Los sistemas de verificación facial generalmente se despliegan en ambientes restringidos, como puede ser para abrir y cerrar puertas, acceder a un equipo de cómputo, desbloquear algún dispositivo inteligente, entre otros.

En la identificación facial, mostrada en la Figura 4.4, primero se obtiene la muestra facial $f(x, y)$ de una persona y se compara contra cada una de las plantillas biométricas B_i almacenadas en una base de datos G . Todos aquellos valores $s(f(x, y), B_i) \geq \tau$ se agregan a una lista de puntuaciones C , la cual es ordenada descendientemente al finalizar la comparación. De esta forma, la identidad de $f(x, y)$ corresponde a la plantilla con grado de similitud en $C(1)$. Es decir, la plantilla biométrica con el más alto grado de similitud mayor que o igual a τ . En caso de que C resulte vacío, la imagen facial $f(x, y)$ es declarada como desconocida. Los sistemas de identificación se despliegan generalmente en ambientes no restringidos. Por ejemplo; en lugares públicos, en aeropuertos, centrales de autobuses, universidades, hospitales, entre otros.

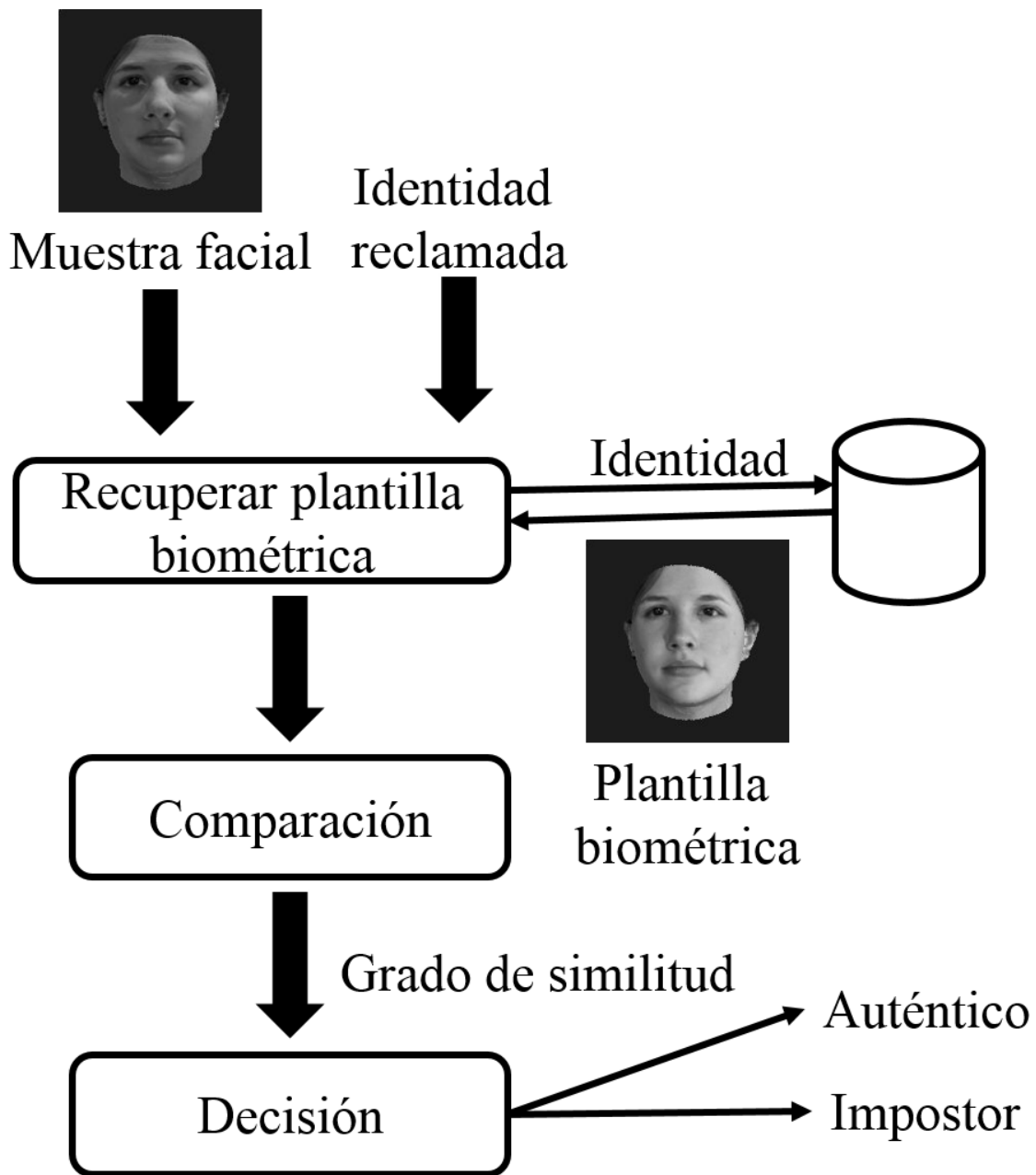


Figura 4.3: Esquema básico de la tarea de verificación facial.

Los primeros algoritmos de reconocimiento facial usaron simples modelos geométricos, pero el proceso de reconocimiento ha madurado en una ciencia de sofisticadas representaciones matemáticas y procesos de acoplamiento. El primer sistema semi-automatizado para el reconocimiento facial fue desarrollado en 1960. Este sistema requería que una persona localizara manualmente las características faciales sobre una imagen facial antes de calcular sus distancias y proporciones para un punto de referen-

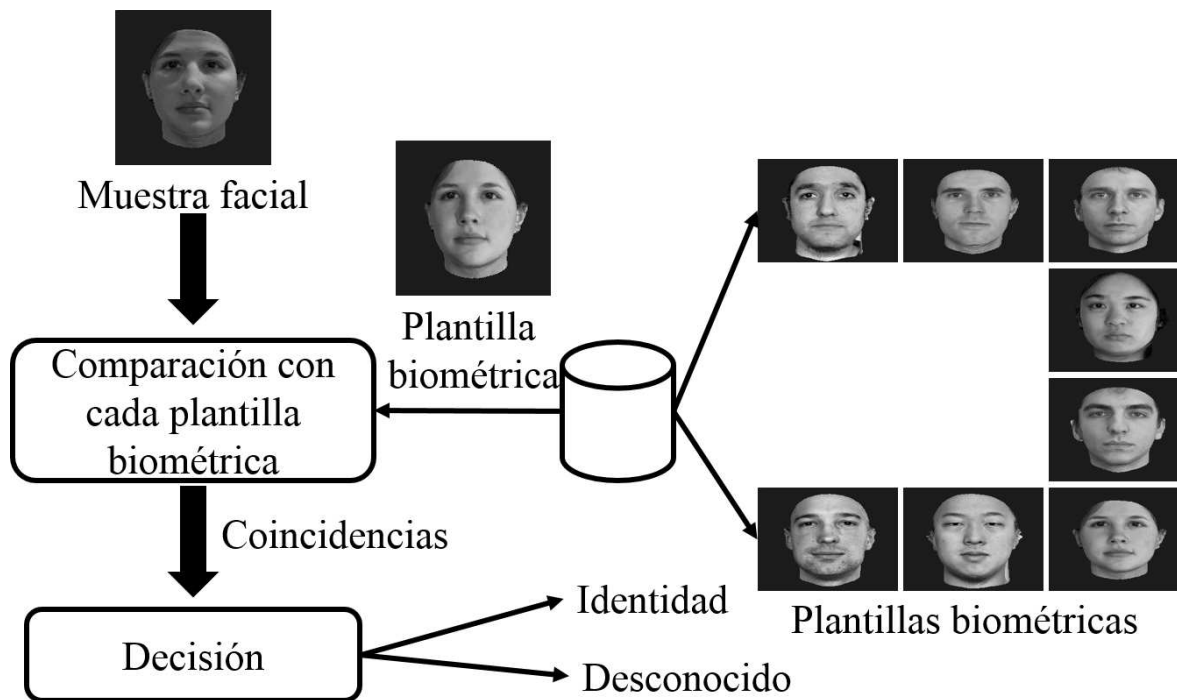


Figura 4.4: Esquema básico de la tarea de identificación facial.

cia en común. Posteriormente, en la década de 1970 se propuso en (Goldstein, Harmon, y Lesk, 1971) un conjunto de 22 características para el reconocimiento facial. Esas 22 características corresponden a las cejas, ojos, nariz, boca, barbilla, orejas, mejillas, frente y cabello. El problema con estos dos enfoques es que las características tenían que ser localizadas y medidas manualmente. Para resolver este problema, en (Sirovich y Kirby, 1987) se aplicó un método conocido como análisis de componentes principales (ACP) el cual se basa en representar las caras como una combinación lineal de ciertas caras básicas denominadas *Autocarar*s. Posteriormente, (Turk y Pentland, 1991) descubrieron que mientras se usaba la técnica de *Autocarar*s, el error residual puede ser usado para detectar caras en imágenes. Las imágenes faciales son proyectadas sobre un espacio de características (espacio facial) que codifica mejor la variación entre las imágenes faciales conocidas. El espacio facial está definido por las *Autocarar*s, las cuales son *eigenvectores* del conjunto de caras. Este descubrimiento permitió el desarrollo de sistemas que reconocían automáticamente y en tiempo real.

La mayoría de los sistemas de reconocimiento facial usan imágenes bidimensionales (2D) a color o escala de grises. Otro enfoque que ha surgido es el que utiliza las imágenes faciales en tercera dimensión (3D). Este enfoque representa la información tridimensional de diferentes formas: a) como una nube de puntos en el espacio, b) mapas de profundidad o imágenes de rango y c) la información tridimensional registrada con la imagen de textura. El enfoque de reconocimiento facial 3D inició al principio de la década de los noventa, específicamente en 1991, que es a partir de entonces cuando se

tienen trabajos reportados. Esto fue gracias al abaratamiento de los sensores de adquisición 3D y el aumento en su capacidad de procesamiento. (Wiskott, Fellous, Kruger, y V.-Malsburg, 1997) presentan un enfoque basado en el acoplamiento de grupos de mallas elásticas. Este enfoque extrae descripciones faciales concisas en la forma de *grafos de imágenes*. Los puntos fiduciales sobre la cara (ojos, boca, entre otros) son descritos por un conjunto de componentes wavelet (*jets*). Por su parte, (Kun, Lin, Li, y Liang, 2012) propone el reconocimiento facial 3D basado en el algoritmo de clasificador de cascada. Primero, el punto de los datos 3D son mapeados al plano $X - Y$ mediante un método de interpolación para obtener el rango de imagen. Luego, se emplea el ACP para encontrar el espacio facial de características de dimensiones bajas y entonces obtener el grado de similitud de acuerdo a la regla del vecino más próximo. Finalmente, las partes restantes de las caras son verificadas con base en el algoritmo de distancia Hausdorff. En (Hansen y Atkinson, 2010) se presenta un método inspirado biológicamente para extraer cierta información facial discriminatoria para el reconocimiento facial 3D. Este enfoque demuestra que descartando el 90 % del dato facial, mientras se mantiene el 10 % solo lleva a un 24 % de pérdida en el desempeño del reconocimiento 3D. En (Zappa, Mazzoleni, y Hai, 2010) se propone un sistema estereoscópico que adquiere dos imágenes de una persona los cuales son analizados separadamente por un modelo de apariencia que extrae de ellos 58 punto homólogos. La triangulación de esos puntos permite la comparación de máscaras tridimensionales contra una base de datos de máscaras de referencia. (Yin, Dai, Ouyang, Liu, y Wei, 2014) propone un método multimodal $2D + 3D$ para una aplicación en una ciudad inteligente. Los mapas de profundidad son explotados para las representaciones 3D, mientras que para la extracción de características se propone una nueva representación llamada patrón derivado local completo (PDLC). Esta representación se aplica de forma separada a características Gabor extraídas de imágenes 2D y mapas de profundidad. Entonces, se obtienen dos características adicionales: Gabor-PDLC y profundidad-PDLC. Ambas características son ponderadas por los coeficientes correspondientes que se combinan en el nivel de decisión para calcular la distancia de clasificación local. Al final, la identidad de la prueba facial corresponde a la distancia de clasificación más pequeña.

Las redes neuronales han sido aplicadas al reconocimiento facial automático. En (Linga-Reddy, Babub, y Kishore, 2010) se propone el uso de características *eigen/fisher* de componentes faciales multiescalados y redes neuronales artificiales. La idea básica consiste en construir vectores de características faciales mediante un muestreo hacia abajo de las características faciales con diferentes resoluciones de los componente faciales. Luego se aplica ACP o análisis discriminante lineal (ADL) para la reducción de la dimensionalidad y lograr una buena representación de las características faciales. Cada cara en la base de datos es proyectada en eigenspaces o eigenfaces para encontrar su vector de peso. Los vectores de peso de las imágenes faciales a ser entrenados se convierten en la entrada a un clasificador basado en una red neuronal, el cual usa funciones base de retropropagación/radial para reconocer las caras con variaciones en expresiones. En (Banerjee y AsitK.Datta, 2013) se propone un método que usa filtros de correlación para el reconocimiento bajo condiciones de iluminación variable y oclusión. El proce-

samiento se lleva a cabo en el dominio espacial mediante dos núcleos de convolución. Los núcleos se obtienen mediante el entrenamiento de una red neuronal de regresión generalizada usando características faciales generalizadas por ACP. También (Banerjee, Cahndra, y Datta, 2010) propone un método que combina las redes neuronales con los filtros de correlación. Este enfoque presenta una forma alternativa de toma de decisión sobre el objeto. Mientras la correlación se basa en valores umbrales, este enfoque propone tres vectores que caracterizan el plano de correlación en términos de la intensidad del pico, la energía total y PSR. Esos tres vectores son usados para entrenar una red neuronal de regresión generalizada. Por otra parte, (Xu, Zhang, y Gai, 2011) propone un método que usa una red neuronal cuántica con función de transferencia multi-nivel y clasificadores multi-capas. Primero, la imagen es preprocesada para eliminar información no relacionada con la cara, lo cual ayuda a localizar la posición de los ojos. Segundo, se extraen características mediante el método de eigenfaces. Finalmente, en la parte de reconocimiento facial, se usan las redes neuronales cuánticas basados funciones de transformación multinivel.

Actualmente, se requiere que los sistemas de reconocimiento facial procesen millones de comparaciones en busca de coincidencias y que traten con la pobre calidad de las imágenes en ambientes no restringidos. En (X. Chen y cols., 2013) se propone un algoritmo de reconocimiento facial acelerado mediante GPU para reducir el tiempo de procesamiento. Para imágenes faciales de baja calidad, en (S. Zhang, Zhang, Zhang, y Xue, 2010) se propone un algoritmo que aplica a las imágenes faciales la super-resolución basado en multi-escala para mejorar su calidad. Primero, se usan pirámides dirigibles para capturar características locales de bajo nivel. Entonces, esas características se combinan mediante estructuras similares a pirámides y parches sintéticos. Finalmente, la imagen facial de alta resolución se obtiene aplicando un operador de suavizado lineal global. Una forma simple de tratar con imágenes de baja calidad, es realizando un simple promediado para enfatizar las características faciales comunes. (Robertson, S.-Kramer, y Burton, 2015) propone un método que promedia varias imágenes faciales capturadas por un teléfono inteligente. Luego, la imagen facial promedio es usada en la verificación de la identidad de la persona. La mayoría de los enfoques usan imágenes en escala de grises para el reconocimiento facial, dejando a un lado la información que pudiera proporcionar las imágenes a color.

Algunos sistemas de reconocimiento facial han hecho uso del color de la piel, sin embargo, no han tenido impacto debido a la gran variedad de colores de piel en los humanos. Esto dificulta la segmentación de la cara. (Choi, Ro, y Plataniotis, 2011) presenta un estudio comparativo que aborda el impacto del desajuste del preprocesamiento en métodos de reconocimiento facial basado en imágenes a color. Por otra parte, es necesario que se exploten las técnicas de aprendizaje artificial en los algoritmo de reconocimiento facial para que estos no se limiten solo a la información con que fueron entrenados. En (Schroff, Kalenichenko, y Philbin, 2015) se presenta FaceNet, un sistema de reconocimiento facial, que aprende directamente de un mapeo de imagen facial a un espacio euclidiano compacto donde las distancias corresponden directamente a las medidas de similaridad facial. Una vez que este espacio ha sido generado, las tarea de

reconocimiento y agrupamiento con técnicas simples como los vectores de características.

En el presente trabajo de investigación se aborda el reconocimiento facial usando filtros de correlación compuestos, los cuales ofrecen varias ventajas:

- No es necesario que las imágenes faciales estén alineadas. Contrario a métodos basados en ACP y ADL que requieren que las características faciales estén en la misma posición y que la imagen esté centrada.
- Pueden usar tanto la apariencia como la forma de la cara, lo cual mejora el reconocimiento y la discriminación.

En (Savvides, Vijaya-Kumar, y Khosla, 2002) se propone el uso de filtro de correlación MACE en la verificación de imágenes faciales con expresiones. El uso de un umbral de reconocimiento universal mostró que MACE rechaza correctamente a las imágenes faciales impostoras. En (Kiat-Ng, 2005) se implementan los filtros de correlación tanto en la verificación como en la identificación sobre un dispositivo móvil. Este trabajo evidencia la posibilidad del reconocimiento facial mediante filtros de correlación sobre dispositivos con poder computacional y recursos de memoria limitados. En (Yan, Wang, Li, Yang, y Zhong, 2013) se propone el filtro Unconstrained Optimal Origin Trade off Filter (UOOTF) para el reconocimiento facial. El filtro UOOTF extrae las características faciales discriminativas mediante la eliminación de restricciones en el origen del plano de correlación y enfatiza el origen en la salida de correlación. En (Banerjee y Datta, 2012, 2014), primero se genera un par de filtros de correlación de la imagen de entrada proyectadas sobre subespacios específicos, llamados filtro de correlación de imagen de proyección óptima y filtro de correlación reconstruido. En (Yi, 2014) se propone un algoritmo de reconocimiento facial basado en filtros de correlación óptimos. Primero, una imagen facial es reconstruida usando el ACP y el filtro de correlación de imagen de proyección óptima. Segundo, el filtro de correlación de refactorización (FCR) para refactorizar las imágenes, las cuales son clasificadas de acuerdo a valores umbrales pre-establecidos de tolerancia a la iluminación. En (Yan, Wang, y Suter, 2014) se propone un algoritmo de extracción de características, llamado Multi-Subregion based Correlation FilterBank (MS-CFB), para el reconocimiento facial. El algoritmo MS-CFB combina los beneficios de algoritmos de extracción de características globales y locales, donde múltiples filtros de correlación corresponden a diferentes subregiones diseñadas para optimizar la salida de correlación. En (Arjun y Nair, 2014) se propone un método de reconocimiento facial basado en la correlación de solo fase. Este enfoque usa el concepto de histograma de patrones de fase Gabor (HPFG) complementados por la correlación de solo fase. Como se mencionó anteriormente, una ventaja de los filtros de correlación es que se pueden entrenar en línea. Esto permite que los filtros sean capaces de adaptarse a los cambios. (Lai, Ramanathan, y Wechsler, 2008) propone los filtros de correlación robustos y adaptativos (FCRA) que contiene información sobre la apariencia de la cara así como de su ubicación. La autenticación se realiza siguiendo un algoritmo de reconocimiento por partes. En (R. Chen, Yan, Liu, y Wang, 2014; Savvides

y cols., 2007) se describe la aplicación de los filtros de correlación en el reconocimiento facial. En dichos trabajos, los autores analizan la importancia del reconocimiento en el dominio de la frecuencia. Para imágenes faciales infrarrojas termales, en (Heo, Savvides, y Vijayakumar, 2005) se presenta una evaluación de su uso en el reconocimiento facial mediante filtros de correlación.

En la literatura se han propuesto varios algoritmos de reconocimiento facial. Sin embargo, aún no se ha logrado superar la capacidad de reconocimiento facial que tienen los expertos. Por ejemplo, en (White, Phillips, Hahn, Hill, y OToole, 2015) se realizó una comparación entre examinadores de identificación facial forense, personas no entrenadas y algoritmos computacionales. Este estudio reveló que los examinadores son los mejores en el reconocimiento facial. Para una detallada revisión de los algoritmos de reconocimiento facial existentes se recomienda revisar las Referencias (Choudhary y Goel, 2013; Zhao, Chellapa, Phillips, y Rosenfeld, 2003; Arya, Pratap, y Bhatia, 2015; Ding y Tao, 2015).

4.2. Detección facial

En la sección anterior se presentó el reconocimiento facial, donde se asume que la imagen de entrada corresponde a una cara humana. Esto significa que dicha cara fue previamente detectada y recortada. En esta sección se describe la detección facial y algunos de los más conocidos algoritmos reportados en la literatura.

La detección facial se define de la siguiente manera. Dada una imagen arbitraria, el objetivo es determinar si la imagen contiene alguna cara, y en caso de que sí contenga, se devuelve la ubicación y dimensión de cada cara presente en la imagen (M.-H. Yang, Kriegman, y Ahuja, 2002). Esta detección debe ser independiente de la posición, escala, orientación, edad y expresión, así como de las condiciones de iluminación y calidad de la imagen (S.-Z. Li y JianxinWu, 2011). Sin embargo, la detección facial es una tarea muy desafiante debido a factores tales como pose, ausencia de estructuras faciales, expresiones faciales, oclusión, orientación de la imagen y las condiciones de la misma (M.-H. Yang y cols., 2002). Un buen algoritmo de detección facial debe ser rápido y consumir la menor cantidad de recursos computacionales posible. Esto, debido a que una imagen de entrada puede contener varias caras y además de que cuando se procesan videos, hay una gran cantidad de imágenes disponibles a procesar.

La detección facial sobre una imagen puede realizarse usando el color de la piel, la forma y estructura de la cara, la presencia de alguna característica facial (en caso de una cara incompleta), la apariencia de la cara o la combinación de alguno de esos parámetros (M.-H. Yang y cols., 2002; Jiang, Fischer, Ekenel, y Shi, 2013). Los algoritmos que mejor desempeño han logrado son aquellos que se basan en la apariencia facial, los cuales extraen las características más relevantes de imágenes muestras que corresponden a caras y no caras. El algoritmo basado en apariencia más conocido es el Viola-Jones propuesto en (P. Viola y Jones, 2001; M. Viola, Jones, y Viola, 2003). La idea básica es escanear una imagen de entrada en diferentes posiciones y escala. En

cada posición, se toma una decisión si la región donde se ubica la ventana de escaneo corresponde o no a una cara. Este enfoque usa características rectangulares llamadas *Haar-Like*. Estas características se calculan sobre una representación de imagen denominada *imagen integral*. La *imagen integral* en la posición x, y contiene la suma de los píxeles arriba y a la izquierda de x, y , y está dado por:

$$ii(x, y) = \sum_{x' \leq x, y' \leq y} i(x', y'), \quad (4.1)$$

donde $ii(x, y)$ es la imagen integral e $i(x, y)$ es la imagen original.

Para el aprendizaje, se usa el algoritmo llamado Adaboost el cual selecciona las características y entrena el clasificador. Su forma de trabajar es muy simple, pero eficaz, pues lo que hace es combinar un conjunto de clasificadores débiles para formar un clasificador fuerte. En este contexto, cada característica *Haar-Likes* es un clasificador débil. El clasificador fuerte, al final del proceso de aprendizaje, toma la forma de un perceptrón de doble capa. El algoritmo Viola-Jones puede detectar una cara con una resolución mayor que o igual a 24×24 píxeles. Se sugiere la revisión de las Referencias (S.-Z. Li y JianxinWu, 2011; Y.-Q. Wang, 2013) para un análisis exhaustivo de este algoritmo y su implementación en sistemas de detección facial.

La correlación en la detección facial tiene la ventaja de que en una sola operación puede detectar múltiples objetos de la misma clase y con la misma dimensión. Para que un filtro de correlación pueda ser usado en la detección facial, éste tiene que contener información facial general y suficiente para producir un pico fuerte y agudo en el centro de cada cara en una imagen arbitraria. El proceso de detección mediante filtros de correlación es de la siguiente manera. Sea \mathbf{h} un filtro de correlación para la detección y $s(x, y)$ una imagen arbitraria. Si $s(x, y)$ contiene instancias de la cara humana, entonces $g(x, y) = ITF\{S(u, v)H^*(u, v)\}$ debe exhibir un pico en el centro de cada ocurrencia del objeto. La detección se lleva a cabo localizando los picos, los cuales son medidos por alguna métrica y si alguno es mayor que o igual a un valor umbral, entonces se considera como cara. La dimensión de cada cara corresponde a la dimensión del filtro con que fue detectada. El principal reto al que se enfrentan los filtros de correlación en la detección facial son las variaciones en escala y rotación (sobre el plano y fuera del plano).

El filtro ASEF ha sido aplicado en la detección facial (Bolme y cols., 2009). ASEF requiere de varios miles de imágenes de entrenamiento para lograr una detección robusta sobre fondos complejos. Una ventaja de los filtros de correlación es que pueden ser entrenados para detectar solo una persona en específica. Para este propósito, en (Yeom y Woo, 2013) se propone un método que usa un filtro óptimo no lineal y etapas de verificación subsecuentes. El filtro compuesto es entrenado con varias imágenes bajo modelado de distancias largas. Las regiones faciales candidatas son proporcionadas por las salidas de los filtros de la escena de entrada. Posteriormente, las falsas alarmas son eliminadas por etapas de prueba subsecuentes, los cuales consideran el color de la piel y pruebas de filtrados de máscaras de bordes. En (Majed y Arof, 2008) se presenta un sistema de detección facial basado en el filtro de correlación MACE.

En la actualidad existen distintas aplicaciones reales de la detección facial. En (Qian, Ma., Xu, Li, y Shen, 2012) se presenta un sistema de visión binocular para la detección y seguimiento facial en robots. La cara es detectada por AdaBost, el cual usa el color de la piel y datos de profundidad. Otra aplicación es la que se presenta en (Ptucha, Rhoda, y Mittelstaedt, 2013), la cual detecta caras en una imagen, luego realiza un autorecorte enfocándose en las personas y re-escalamiento. En las Referencias (Hjelmas, 2001; M.-H. Yang y cols., 2002; C. Zhang y Zhang, 2010; Degtyarev y Seredin, 2010; S.-Z. Li y JianxinWu, 2011) se pueden encontrar estudios completos sobre detección facial; retos, arquitecturas, algoritmos e implementación.

4.3. Seguimiento facial

Un vídeo puede ser considerado como una colección de imágenes estáticas que pueden registrar el movimiento de una cara en una escena. El vídeo tiene varias características que pueden ser explotadas en el seguimiento facial (Chellappa y cols., 2011): 1) correlación temporal, 2) información dinámica idiosincrásica y 3) disponibilidad de múltiples vistas de la cara humana. Estas características permiten el uso de algoritmos de procesamiento facial para imágenes estáticas en secuencias de vídeos.

El seguimiento facial consiste en estimar continuamente la localización y posiblemente la orientación de una cara en cada imagen estática de una secuencia de vídeo (M.-H. Yang y cols., 2002). Una solución simple de seguimiento pudiera consistir de aplicar el algoritmo de detección facial sobre cada imagen estática del vídeo. Sin embargo, esta solución es muy costosa computacionalmente y poco práctica cuando se sigue solamente un objeto. Esto porque en una solución basada en la detección, tendríamos que autenticar cada cara detectada para distinguir el objetivo. Por lo tanto, es necesario el uso de otras estrategias para el seguimiento facial sin la necesidad de autenticación y que se enfoque solo en la cara objetivo.

En (D. A. Ross, Lim, Lin, y Yang, 2008) se propone un método que aprende de forma incremental una representación de sub-espacio de baja dimensión, adaptándose eficientemente a los cambios en la apariencia del objetivo. El modelo actualiza el descriptor facial, basado en algoritmos incrementales para ACP, considerando la actualización de la media muestral y la minimización del efecto de anteriores observaciones. (Adam, Rivlin, y Shimshoni, 2006) propone un método que usa múltiples fragmentos de la imagen del objetivo. La estimación está basada en votos de cada fragmento sobre posibles posiciones y escalas. En (X. Zhang, Hu, Maybank, y Li, 2007) se propone un método de aprendizaje discriminativo basado en grafo embebido. Este método puede simultáneamente aprender el subespacio del objetivo y su estructura discriminativa local contra el fondo. En (Babenko y Belongie, 2011) se propone un método que entrena un clasificador discriminativo de manera *en-línea* para separar el objeto del fondo.

En (Painkras y Charoensak, 2005) se propone un framework para el seguimiento facial basado en wavelets de Gabor, el cual se compone de a) una unidad de detección facial, b) unidad de descomposición de Gabor y cálculo de similaridad y c) una unidad

de refinamiento facial. En (Saxena, Grover, y Joshi, 2008) se propone un método de seguimiento facial que primero detecta una cara usando un algoritmo de rango deficiente. Aunque este enfoque se basa en la detección, se aplica una estrategia para no realizar el reconocimiento en cada imagen. En su lugar, se hace una estimación de la ubicación de la cara objetivo en cada cuadro. Por su parte, (Myers y Megherbi, 2014) si aborda la autenticación de la cara en el seguimiento.

La mayoría de los enfoques descrito previamente, realizan el seguimiento sobre un solo vídeo, mientras que la vida real es necesario hacer el seguimiento de la cara de una misma persona en varios vídeos. Para este propósito, (Yoder, Medeiros, Park, y Kak, 2010) propone un sistema de seguimiento facial multi-cámara. Una ventaja de dicho enfoque es que no requiere de un servidor central, pues usa un protocolo que incluye una función de transferencia de los datos de seguimiento del objetivo. En (G. Wang, Alam Bhuiyan, Cao, y Wu, 2014) se propone un marco de trabajo que estima el movimiento de una cara hacia otra cara. Para este propósito, se introduce un algoritmo de detección de borde para generar caras que sean robustas a los movimientos de la cabeza.

Los filtros de correlación ofrecen dos ventajas principales en el seguimiento facial. Primero, se pueden entrenar para una persona en específica de tal forma que en cada imagen del vídeo, la meta sea localizar el objetivo. Segundo, pueden actualizarse *en-línea* lo que los posibilita adaptarse a los cambios de que sufra la cara objetivo. El filtro MOSSE (Bolme y cols., 2010) ha sido aplicado en esta tarea, resultando ser mucho más rápido que otros algoritmos existentes. Este algoritmo tiene la capacidad de recuperarse de fallos en el seguimiento, es invariante a escala y rotación. Sin embargo, después de procesar varias imágenes la ventana de seguimiento se desfasa del centro del objetivo. Otro algoritmo de seguimiento facial basado en filtros de correlación es el que se presenta en (Santiago-Ramírez y cols., 2012), donde un filtro se inicializa con solo una muestra facial para luego ser actualizado con cada cara localizada en subsecuentes imágenes del vídeo.

4.4. Resumen

En este capítulo se describieron algoritmos de procesamiento facial más comunes. Es fácil notar que la re-identificación de caras incluye implícitamente las tarea de detección, reconocimiento y seguimiento. También, es evidente que los filtros de correlación ofrecen varias ventajas en la re-identificación al poderse adaptar a los cambios.

Capítulo 5

Detección facial mediante filtros de correlación compuestos

La detección facial es el primer paso en un sistema de procesamiento facial automatizado. Los actuales algoritmos, algunos de los más comunes fueron mencionados en el Capítulo 4, son capaces de detectar rostros en imágenes de buena calidad. Generalmente, una imagen de buena calidad se obtiene en ambientes donde es posible controlar algún factor, como por ejemplo, la iluminación, el fondo y la cooperación del usuario. Sin embargo, el desempeño de esos algoritmos disminuye cuando procesan imágenes no restringidas (Lee, 2015). En estas imágenes la apariencia facial puede presentarse distorsionada debido a la iluminación variable, rotación fuera del plano (pose), rotación en el plano, expresiones, gestos, entre otros.

Otro aspecto importante a considerar en la detección es que la cara detectada tiene que ser adecuada para tareas subsecuentes de procesamiento facial. Por ejemplo, si la siguiente tarea es el reconocimiento facial, entonces la cara detectada debe contener todas las características faciales, centrada, alineada y frontal para que pueda ser reconocida eficientemente. Cuando la cara detectada no presenta las características ideales para un perfecto acoplamiento, entonces el desempeño de los algoritmos de reconocimiento facial disminuye considerablemente. Por esta razón, y debido al contexto de este trabajo de tesis, es necesario desarrollar un algoritmo de detección que permita encontrar caras en imágenes no restringidas.

En este capítulo se propone un algoritmo de detección facial basado en filtros de correlación compuesto. El filtro se diseña usando clasificadores fuertes sintetizados con imágenes seleccionadas de una base de datos mediante un algoritmo inteligente. El desempeño del algoritmo propuesto fue evaluado en términos de su habilidad para determinar la ubicación de una cara bajo condiciones de iluminación variable, pose, fondo complejo y algunas expresiones faciales.

5.1. Método de detección facial basado en correlación

El algoritmo propuesto en este trabajo explota tanto la forma como el contenido de la cara en un filtro de correlación compuesto no lineal tolerante a iluminación variable, ruido, ligeras variaciones en rotación (en el plano y fuera del plano) y algunas expresiones faciales. El algoritmo propuesto consiste principalmente de las siguientes etapas: *selección del conjunto de entrenamiento*, *síntesis de los filtros de correlación compuesto no lineales* y *detección facial*.

Estas etapas están representadas en la Figura 5.1. Dado un conjunto de imágenes faciales disponibles $D = \{f_1(x, y), f_2(x, y), \dots, f_N(x, y)\}$, en la etapa de *selección del conjunto de entrenamiento*, se selecciona un subconjunto óptimo de imágenes de entrenamiento $T \subseteq D$. En la etapa de *síntesis de los filtro de correlación compuesto no lineal*, el conjunto seleccionado T es sintetizado en un filtro de correlación compuesto no lineal. Finalmente, en la etapa de *detección facial* se realiza la detección facial sobre una imagen arbitraria $s(x, y)$. Las primeras dos etapas se ejecutan solo una vez, aunque en un futuro puede volverse a ejecutar cuando sea necesario una actualización. Sin embargo, la tercera etapa es la que se ejecuta cada vez que se desea detectar caras en una imagen. Estas etapas son descritas detalladamente a continuación.

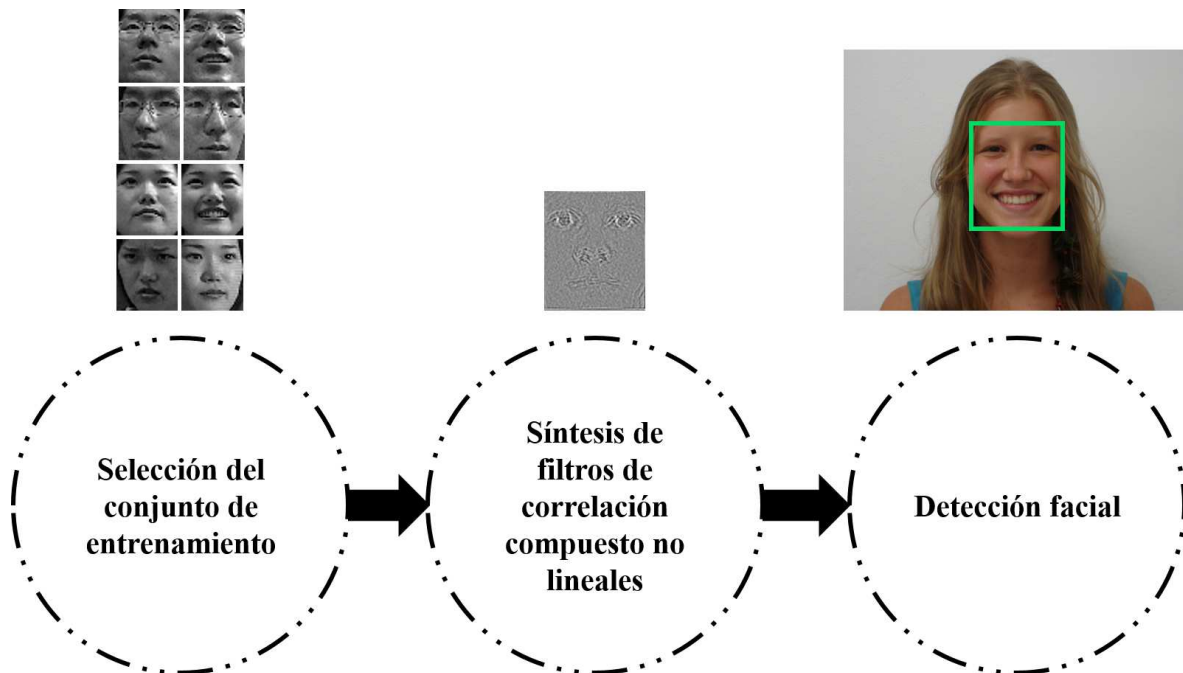


Figura 5.1: Principales etapas del método de detección facial propuesto.

5.1.1. Selección del conjunto de entrenamiento

El éxito de un algoritmo de detección facial basado en filtros de correlación depende grandemente del conjunto de entrenamiento T . Este conjunto debe contener las distorsiones más generales que se espera presente cualquier cara humana, por lo que es necesario usar una base de datos de imágenes faciales D con miles de imágenes. Sin embargo, no todas las imágenes faciales en una base de datos son útiles para la detección. Por lo tanto, es necesario seleccionar solo aquellas que contribuyen a una detección robusta.

Como el problema de selección del conjunto de entrenamiento consiste en escoger el mejor subconjunto de D , entonces puede plantearse como un problema de optimización numérica. Para que esto sea posible, es necesario definir un *operador* O , una *función objetivo* f y una estructura de datos \mathbf{x} para representar las posibles soluciones. El *operador* O modifica los valores de algunos elementos de \mathbf{x} para generar nuevas posibles soluciones \mathbf{x}' . Es fácil notar que al modificar solo una pequeña cantidad de elementos, la nueva solución hereda características de su antecesor. La calidad de cada estado, tanto de \mathbf{x} como de \mathbf{x}' , se calcula mediante f . De esta manera, la solución con una mejor calidad es la que se selecciona para generar nuevas y mejores soluciones.

En esta tesis se propone emplear el algoritmo de optimización (AO) Hill Climbing (HC) (X.-S. Yang, 2014a) para la selección, el cual es implementado en el Algoritmo 1. Dado el conjunto D que contiene varias imágenes faciales de diferentes personas, el objetivo es seleccionar aquellas que sean las más representativas para construir un filtro $H(u, v)$ y así obtener una detección facial robusta. También se requiere de un conjunto P que contiene imágenes de escenas con distorsiones a las cuales queremos que el filtro sea robusto. En otras palabras, solo se selecciona $f_i(x, y) \in D$ si éste contribuye a producir un pico agudo y fuerte en el centro de la cara que aparece en $s(x, y) \in P$.

Para el proceso de optimización numérica, se sugiere que \mathbf{x} sea un vector binario que indique el índice de la imagen $f_i(x, y) \in D$ a usar en la solución candidato T_i . De esta manera, si $\mathbf{x}(i) = 1$, entonces la imagen facial $f_i(x, y)$ se agrega a T_i . En caso contrario, $f_i(x, y)$ no es agregada a T_i .

En este trabajo, \mathbf{x} se crea de longitud igual a $|D|$ y poblado de valores reales pseudo-aleatorios. Redondeando los valores de \mathbf{x} se obtiene el vector binario, el cual es usado para generar $T_{i=0}$. El filtro $H(u, v)$ correspondiente a $T_{i=0}$ está dado por:

$$H(u, v) = \frac{1}{N} \sum_{i=1}^N F_i^k(u, v), \quad (5.1)$$

donde $F_i^k(u, v) = |F_i(u, v)|^k \exp(i\varphi_{F_i(u, v)})$ es el filtrado no lineal de la TF de $f_i(x, y) \in T_i$. Luego, el filtro $H(u, v)$ se correlaciona con cada imagen $s_i(x, y) \in P$. Cada plano de correlación $g_i(x, y)$ es examinado para buscar el pico y medir su agudeza mediante la métrica PSR . La calidad del filtro corresponde al promedio de los valores PSR (valor $p\hat{s}r$) obtenido con el conjunto de pruebas. Después de esta inicialización, iterativamente se realiza lo siguiente. El conjunto L se obtiene mediante la modificación de un simple elemento en \mathbf{x} . Luego, para cada instancia se obtiene T_i , se sintetiza su respectivo filtro

Algoritmo 1: Selección del conjunto de entrenamiento para el diseño de un filtro de correlación para detección facial.

Datos: Conjunto de imágenes faciales disponibles D , conjunto de imágenes de prueba P y valor umbral de similitud τ

Resultado: Conjunto de entrenamiento T

```

1  $\mathbf{x} \leftarrow$  Crear vector binario de posiciones con una longitud igual a  $|D|$ 
2  $T \leftarrow$  mapear  $\mathbf{x}$  a un conjunto de entrenamiento mediante la función  $m(x, D)$ 
3  $H(u, v) \leftarrow$  Entrenar filtro con  $T$ 
4 //Medir la calidad del filtro  $H(u, v)$ 
5  $p\hat{s}r \leftarrow$  Calcular  $psr$  promedio de la correlación de  $H(u, v)$  con imágenes de prueba en  $P$ .
6 Establecer el máximo de iteraciones  $maxIt$ 
7 mientras  $i \leq maxIt$  hacer
8    $i \leftarrow i + 1$ 
9    $L \leftarrow Vecinos(x)$ 
10   $sig\_p\hat{s}r \leftarrow -INF$ 
11   $sig\_x \leftarrow NULL$ 
12   $sig\_T \leftarrow NULL$ 
13  per  $j \leftarrow 1$  a  $|L|$  fai
14     $x_j \leftarrow L(j)$ 
15     $T_{actual} \leftarrow m(x_j, D)$ 
16     $H_{actual}(u, v) \leftarrow$  Construir filtro con  $T_{actual}$ 
17    //Medir la calidad del filtro  $H_{actual}(u, v)$ 
18     $p\hat{s}r_{actual} \leftarrow$  Calcular  $psr$  promedio de la correlación de  $H_{actual}(u, v)$  con imágenes de prueba en  $P$ .
19    si  $p\hat{s}r_{actual} > sig\_p\hat{s}r$  entonces
20       $sig\_x \leftarrow x_j$ 
21       $sig\_p\hat{s}r \leftarrow p\hat{s}r_{actual}$ 
22       $sig\_T \leftarrow T_{actual}$ 
23  si  $sig\_p\hat{s}r > p\hat{s}r$  entonces
24     $x \leftarrow sig\_x$ 
25     $T \leftarrow sig\_T$ 
26     $p\hat{s}r \leftarrow sig\_p\hat{s}r$ 
27  $T \leftarrow m(x, D)$ 
28 Devolver  $T$ 

```

$H(u, v)$ y se calcula su respectiva calidad $p\hat{s}r$. Si $p\hat{s}r$ de $x_j \in L$ es superior al de \mathbf{x} , entonces $\mathbf{x} = \mathbf{x}_j$. Al finalizar el ciclo, \mathbf{x} representa una solución candidata con una calidad superior o igual a la deseada, y además mejor que las anteriores.

En la Figura 5.2 se muestra un ejemplo de un conjunto disponible que contiene varias

imágenes imágenes faciales de N personas. Un círculo verde está sobre las imágenes seleccionadas de cada persona.

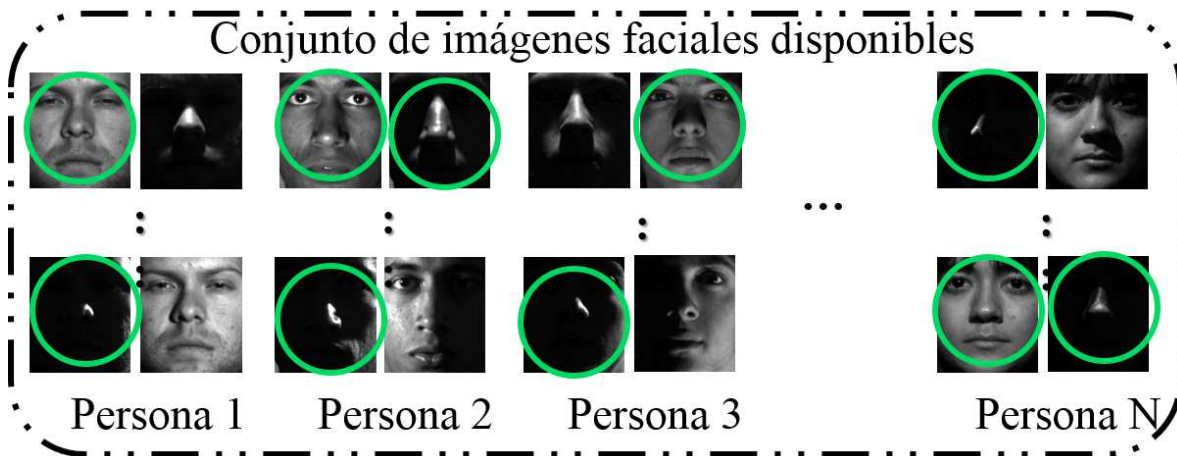


Figura 5.2: Conjunto de imágenes disponible D , donde un círculo verde indica las imágenes seleccionadas por el algoritmo propuesto para el conjunto T .

Es fácil notar que durante las iteraciones, el algoritmo de optimización prueba diferentes combinaciones de imágenes. Por lo que al final, se queda con la mejor combinación de todas las evaluadas. Por esta razón, el desempeño del conjunto de entrenamiento seleccionado es superior al que se pudiera obtener manualmente o mediante métodos convencionales, los cuales solo prueban una combinación.

5.1.2. Síntesis del filtro de correlación compuesto no lineal

Con el fin de producir un pico agudo y fuerte sobre un plano menos ruidoso, el método propuesto está basado en el filtro SDF no lineal y el filtro ASEF. Un filtro exacto es considerado como un clasificador débil porque solo se acopla con la imagen con que fue entrenado. Por esta razón, el acoplamiento falla con otras imágenes que no sean las del entrenamiento, incluso con los de la misma clase. Este problema se resuelve promediando una gran cantidad de filtros exactos para producir un clasificador fuerte, tal y como lo hace el filtro ASEF. Este clasificador fuerte es capaz de acoplarse tanto con las imágenes de entrenamiento como con imágenes que no se usaron en el entrenamiento.

El objetivo principal es sintetizar un filtro invariante a distorsiones de la apariencia facial. En este trabajo se consideraron aquellas distorsiones comunes en ambientes no restringidos, tales como rotación, ruido y forma de la cara. La iluminación fue abordada mediante preprocesamiento de las imágenes para eliminar la variabilidad en iluminación mientras se mejoran los bordes de las características faciales. La Tabla 5.1 muestra las transformaciones que se aplicaron a las imágenes para modelar aquellas distorsiones que no están en el conjunto de entrenamiento. Esto es importante debido a que aunque se usa una base de datos grande, generalmente no contiene tales distorsiones. Por tanto

es necesario generarlos de forma sintética. Las técnicas basadas en plantillas, como los filtros de correlación, son sensibles a la forma de la cara. Hay dos formas predominantes en las caras, la cara redonda y la cara alargada. Por lo tanto, es necesario analizar el conjunto D y generar el que hace falta. Por ejemplo, si en el conjunto disponibles predomina la cara redonda, entonces es aconsejable aplicar una transformación que reduzca su anchura para darle la apariencia de que es alargada. La Figura 5.3 muestra la aplicación de estas transformaciones a las imágenes originales para modelar las distorsiones no contenidas en T y así generar un clasificador fuerte (filtro ASEF) con cada transformación.

Tabla 5.1: Transformaciones de imágenes para incrementar el tamaño de T .

| Número | Transformación |
|--------|-------------------------------------|
| 0 | Imágenes originales |
| 1 | Rotación en el plano de $+z$ grados |
| 2 | Rotación en el plano de $-z$ grados |
| 3 | AWGN con media 0 y varianza s_1 |
| 4 | AWGN con media 0 y varianza s_2 |
| 5 | Escala en anchura |
| 6 | Escala en altura |
| 7 | Volteado de izquierda a derecha |

El clasificador fuerte es generado por el Algoritmo 2 que recibe como argumento el conjunto de entrenamiento T . A este conjunto se le aplican cada una de las transformaciones en la Tabla 5.1 para generar imágenes de entrenamiento sintéticas. Con cada conjunto generado se sintetiza un clasificador fuerte mediante el algoritmo ASEF. Al final, con estos clasificadores fuertes se sintetiza un filtro SDF no lineal con la suficiente información para detectar caras en imágenes no restringidas. Es fácil notar que el filtro de detección se sintetiza con $|T| \times l$, donde l es el número de transformaciones.

Algoritmo 2: Síntesis del filtro compuesto no lineal.

Datos: Conjunto de entrenamiento T

Resultado: Filtro de correlación compuesto $H(u, v)$ para la detección facial

```

1  $T_{cf} \leftarrow \{\}$ 
2 per  $i \leftarrow 0$  a 7 fai
3    $T_i \leftarrow$  Aplicar la  $i$ -ésima transformación de Tabla 5.1 al conjunto  $T$ 
4    $H_{ASEF}(u, v) \leftarrow$  Sintetizar un filtro ASEF usando  $T_i$  como conjunto de
   entrenamiento
5    $h_i(x, y) \leftarrow \mathcal{F}^{-1}\{H_{ASEF}(u, v)\}$ 
6    $h'_i(x, y) \leftarrow$  Rellenar  $h_i(x, y)$  con  $mean\{h_i(x, y)\}$ 
7    $T_{cf} \cup h'_i(x, y)$ 
8  $H(u, v) \leftarrow$  Diseñar un filtro SDF no lineal usando  $T_{cf}$ .
```

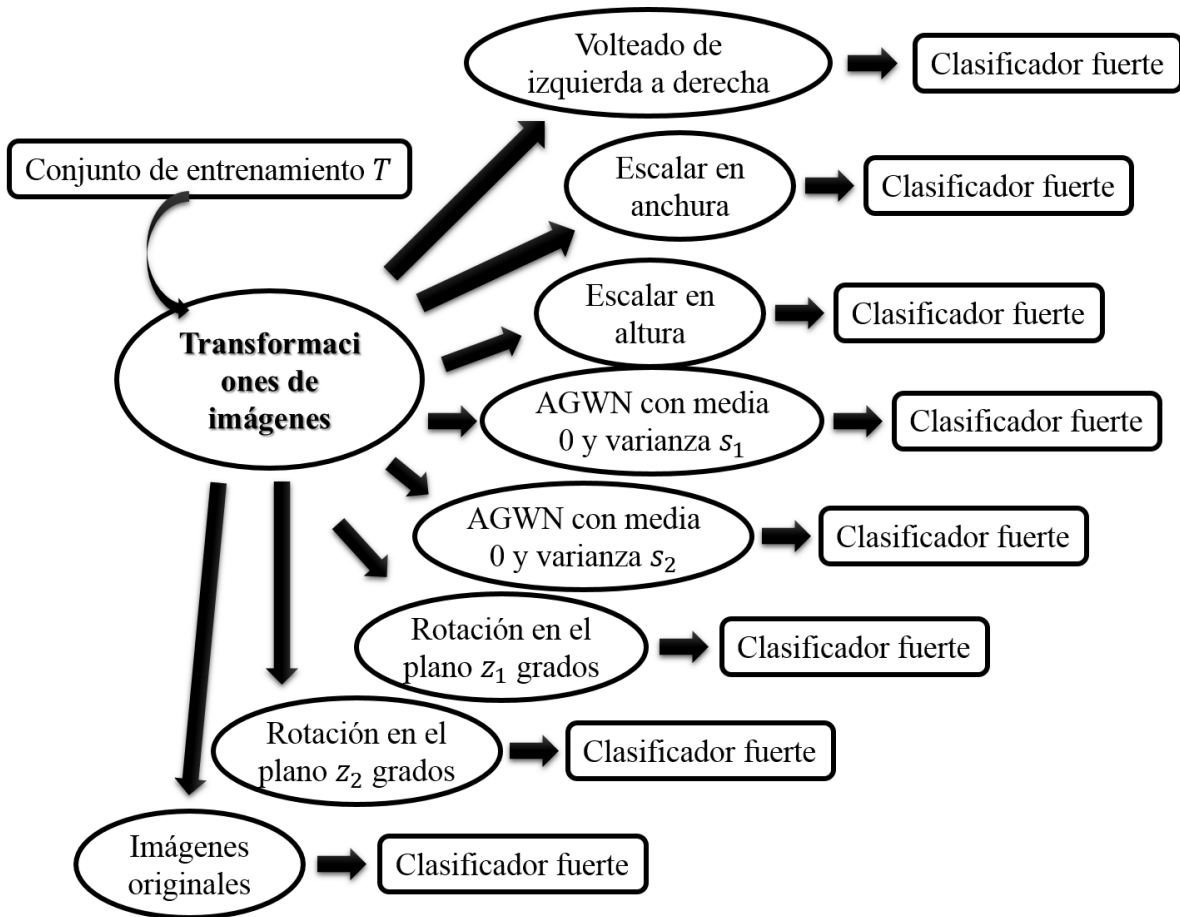


Figura 5.3: Aplicación de las transformaciones al conjunto original para construir los clasificadores fuertes.

El filtro compuesto generado por el Algoritmo 2 solo es capaz de detectar caras con la dimensión de las imágenes de entrenamiento. Es importante recalcar que en nuestros experimentos observamos que existe una tolerancia de ± 2 píxeles. Uno de los problemas más comunes en la detección es la escala, la cual se aborda en esta tesis usando un banco de filtros B de diferentes dimensiones. Sea $\mathbf{d}_{inferior} = [ancho, alto]$ la dimensión más pequeña para un filtro. Entonces se construyen los otros filtros escalando las imágenes con incrementos de cuatro píxeles hasta la dimensión más grande $\mathbf{d}_{superior} = [ancho, alto]$. Este rango de dimensiones indica la cara más chica y la cara más grande que puede ser detectada. El rango debe determinarse cuidadosamente con base al tamaño y la calidad de las imágenes originales. El banco de filtros se genera una sola vez para ser usado en una cantidad infinita de detecciones.

5.1.3. Detección facial

La detección facial se realiza mediante el Algoritmo 3. La entrada para este algoritmo es una imagen arbitraria $s(x, y)$ y un valor umbral τ para la detección. Primero, $s(x, y)$ es mejorado mediante la función $preprocesamiento(s(x, y))$ que aplica el modelado de retina (Vu y Caplier, 2009) de la siguiente manera. Se le aplica el filtrado de retina para mejorar la iluminación global. El filtrado de retina recupera mejor la información de bordes en regiones afectadas por sombras locales. Luego, se realiza la normalización de la energía de la imagen resultante, para después aplicar una ventana de cosenos para reducir los efectos de las frecuencias de los bordes cuando la imagen es transformada por la TF. Segundo, la imagen preprocesada $s'(x, y)$ es transformada por la TF y el filtrado no lineal para obtener $S^k(u, v)$. Tercero, $S^k(u, v)$ es correlacionado con cada filtro almacenado en B como se muestra en la Figura 5.4. Cada salida de correlación se examina en busca de picos para medir su agudeza. Si la agudeza $\geq \tau$, entonces la región del pico es considerada como cara por lo que su ubicación (x, y) y dimensión [*alto, ancho*] son agregados al vector de detecciones D_{det} . Finalmente, D_{det} es filtrado por la función $filtrarSubdetecciones()$ para combinar subdetecciones sobre una misma región. Esto porque los filtros con dimensiones cercanas pueden producir varias detecciones sobre una misma región. La Figura 5.5 muestra este proceso: A) se obtienen todas las subdetecciones, B) usando un clasificador de distancia se agrupan las subdetecciones que están muy próximas y C) de cada grupo se obtiene una sola detección promediando las dimensiones y coordenadas de las subdetecciones. En la Figura 5.5 vemos cómo varias subdetecciones sobre una misma vecindad producen una sola detección.

Algoritmo 3: Detección facial mediante filtros de correlación.

Datos: Imagen de entrada $s(x, y)$ y umbral de detección τ

Resultado: Vector D_{det} con la ubicación y dimensión de cada cara detectada $s(x, y)$.

```

1  $s'(x, y) \leftarrow preprocesamiento(s(x, y))$ 
2  $S(u, v) \leftarrow \mathcal{F}\{s'(x, y)\}$ 
3  $S^k(u, v) \leftarrow |S(u, v)|^k \exp(j\varphi(u, v))$ 
4 //Filtrado no lineal  $D_{det} \leftarrow []$ 
5  $count \leftarrow 1$ 
6 mientras  $dim\{s(x, y)\} \leq dim\{H_{count}(u, v)\}$  hacer
7    $H_{count}(u, v) \leftarrow$  Leer el filtro  $count$  de  $B$ 
8    $g(x, y) \leftarrow \mathcal{F}^{-1}\{S^k(u, v) \cdot H_{count}^*(u, v)\}$ 
9    $subdetecciones \leftarrow buscarDetecciones(g(x, y), \tau)$ 
10   $D_{det} \leftarrow [D_{det}; subdetecciones]$ 
11   $count \leftarrow count + 1$ 
12  $D_{det} \leftarrow filtrarSubdetecciones(D_{det})$ 

```

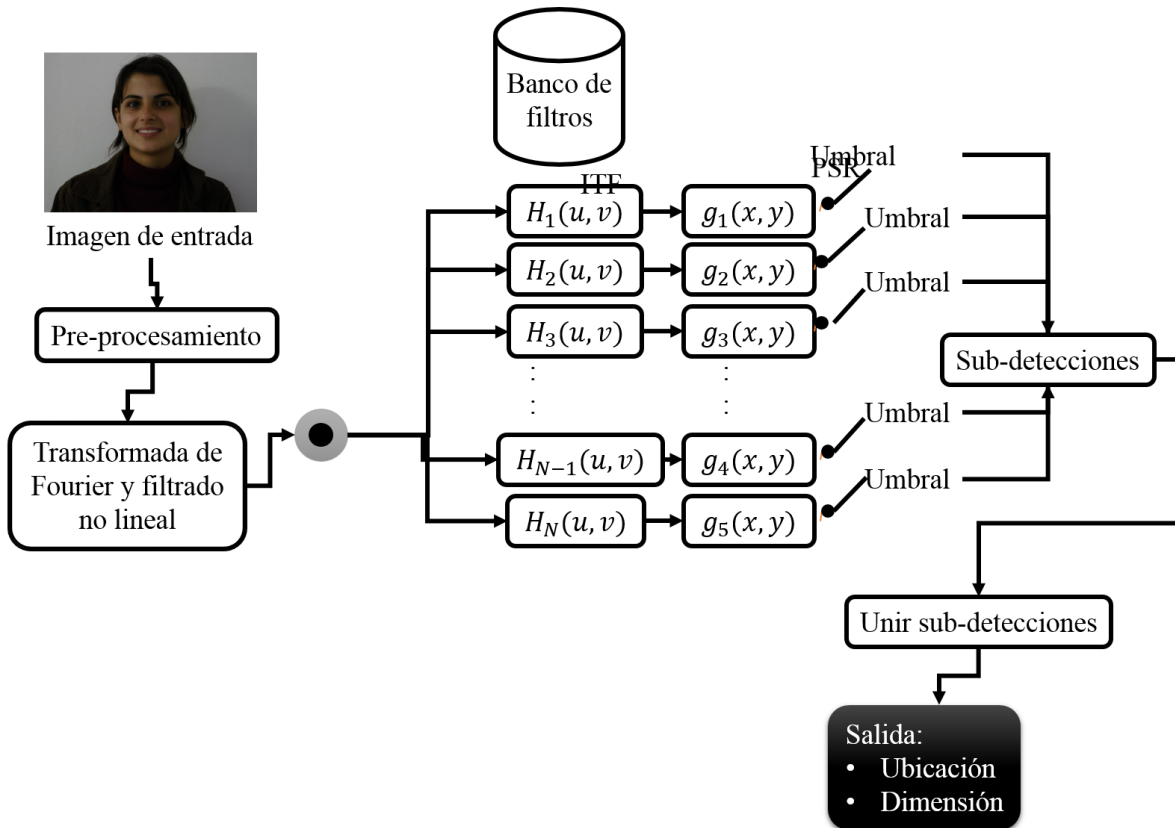


Figura 5.4: Proceso de correlación de la imagen de entrada con cada elemento del banco de filtros B .

En esta sección se ha presentado un método de detección facial que combina los principios de un filtro SDF no lineal y el filtro ASEF. Una característica importante del algoritmo propuesto es que toma ventaja del hecho que los filtros de correlación son invariante al desplazamiento y que permiten la detección paralela de objetos.

5.2. Resultados experimentales

En esta sección se presentan los resultados experimentales que obtuvo el método propuesto en la detección de caras en condiciones de iluminación variable, ligeras variaciones en rotación (fuera del plano y sobre el plano), expresiones faciales y con fondo tanto homogéneo como estructurado.

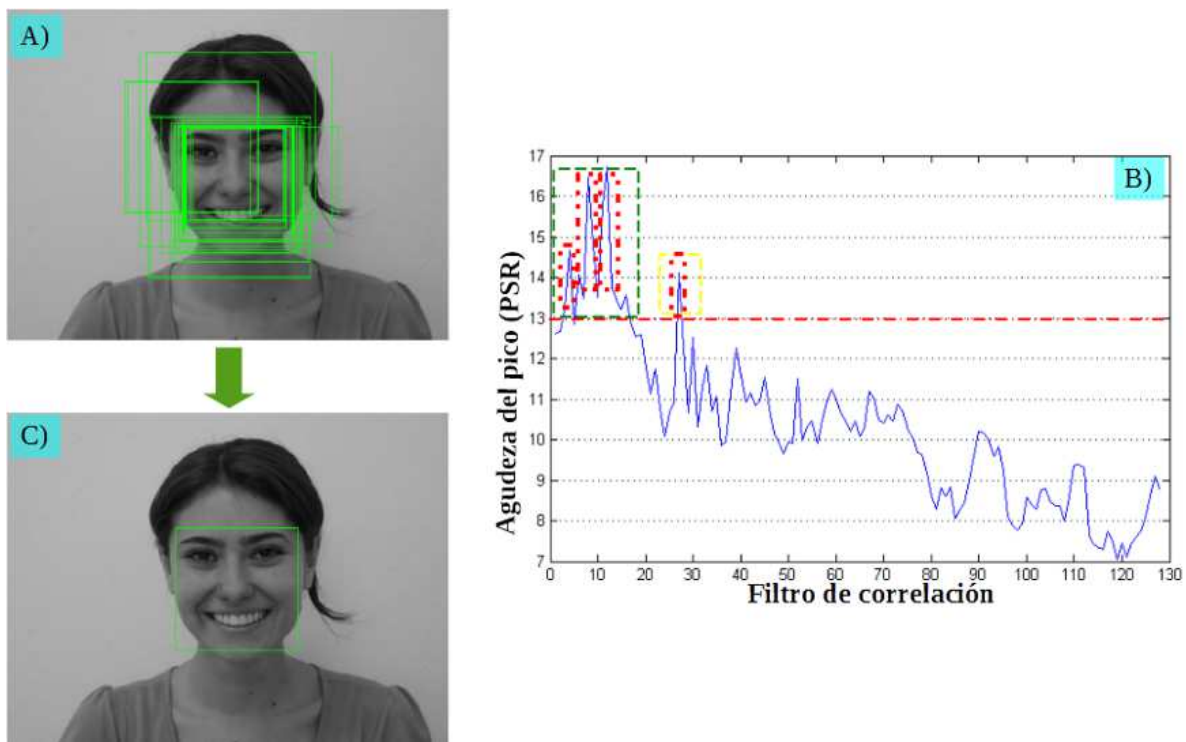


Figura 5.5: Combinación de varias subdetecciones sobre una misma región facial para una sola detección final verdadera.

5.2.1. Configuración del conjunto de imágenes para la evaluación

Para la evaluación del método propuesto se emplearon tres bases de datos de imágenes con rostros humanos. La primera, la base de datos faciales Yale B extendida (simplemente Yale B de ahora en adelante) que contiene 2414 imágenes faciales de 38 personas. Cada persona contiene 60, 63 o 64 diferentes imágenes faciales, donde cada imagen tiene una resolución de 168×192 píxeles y cada una con condiciones de iluminación diferente. Este conjunto se consideró como el conjunto D del cual se seleccionan las imágenes más representativas mediante el Algoritmo 1. La Figura 5.6 muestra las imágenes faciales de una persona en Yale B.

Las otras dos bases de datos formaron los conjuntos de prueba, es decir, imágenes que son procesadas por el método propuesto para determinar si contienen o no caras humanas. El primer conjunto de prueba se compone de imágenes seleccionados del la base de datos FEI (de Oliveira Junior y Thomaz, 2006). Hay 14 imágenes para cada uno de las 200 personas. Todas las imágenes son a color y fueron capturadas contra un fondo homogéneo con las caras en posición frontal y con rotación fuera del plano (pose) de hasta 180 grados. La escala de las caras en las imágenes varían mas o menos en un 10% y el tamaño de cada imagen original es de 640×480 píxeles. Este conjunto se dividió en dos subconjunto: el subconjunto Pose contiene para cada



Figura 5.6: Muestras faciales de una persona en la Yale B recortado.

persona cinco imágenes faciales con pose, mientras que el subconjunto Frontales contiene para cada persona cuatro caras con la mirada hacia la cámara. De esta forma, Pose contiene 1000 imágenes y Frontales contiene 800 imágenes. La Figura 5.7 presenta una muestra de una persona para cada uno de estos dos subconjuntos. El segundo

conjunto de prueba está conformado por imágenes seleccionados de la base de datos Yale B extendido, de la que se extrajeron las caras de Yale B, la cual contiene 16128 imágenes de 28 personas bajo 9 poses y 64 condiciones de iluminación (Georghiades, Belhumeur, y Kriegman, 2001a). De este conjunto solo se seleccionaron 9 personas con un pose diferente cada uno. Las personas son identificadas por la etiquetas YaleB11, YaleB22, YaleB27, YaleB28, YaleB29, YaleB30, YaleB32, YaleB33 y YaleB34, donde cada conjunto contiene 65 diferentes condiciones de iluminación. La Figura 5.8 muestra una imagen de persona con pose.

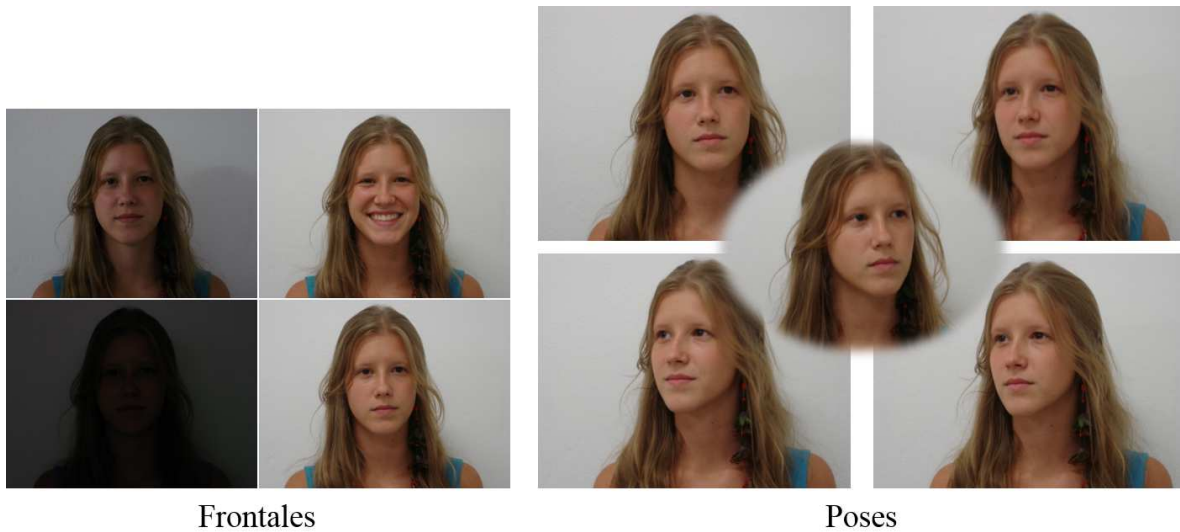


Figura 5.7: Imágenes de una persona para los subconjuntos Pose y Frontales.



Figura 5.8: Imágenes de las personas con pose en Yale B extendido.

Los conjuntos de prueba fueron configurados de tal forma que provean el desempeño del método propuesto para detectar caras a) sobre fondos homogéneos y estructurado, b) con pose, c) con diferentes condiciones de iluminación y d) no representados en el conjunto de entrenamiento.

5.2.2. Medidas de desempeño para el algoritmo de detección facial

Los algoritmos de detección facial, usualmente miden su desempeño mediante la Tasa de Detección (TD) y la Tasa de Falsa Alarma (TFA). También se han adoptados otras métricas para la evaluación como el Tiempo de Aprendizaje (TA), Tiempo de Ejecución (TE), Número de Muestras Requeridas en el Entrenamiento (NMRE) y la Razón entre TD y TFA (M.-H. Yang y cols., 2002). Aunque estas métricas funcionan bien para medir el desempeño de los algoritmos, los filtros de correlación siguen una dinámica de detección diferente. Es necesario un conjunto de métricas que nos permitan cuantificar detecciones exitosas, detecciones parciales, detecciones falsas y los falsos rechazos.

En (Santiago-Ramírez y cols., 2012) se propuso un conjunto de métricas que se adoptan en este trabajo para medir el desempeño de algoritmos de detección facial basado en filtros de correlación. Cuando un filtro se correlaciona con una imagen de entrada, entonces se genera un pico en el centro de la cara. La agudeza del pico comparado con un valor umbral permite determinar si la región corresponde realmente a una cara, mientras que las coordenadas del pico indican la ubicación de la misma. Las métricas mencionadas son descritas a continuación.

Sea S un conjunto de imágenes de prueba y $g_i(x, y)$ el plano de correlación generado por $s_i(x, y)$ y $H_i(u, v)$. Una localización y reconocimiento correcto sucede cuando la agudeza del pico en $g_i(x, y)$ es mayor a un umbral y su coordenada (x, y) está a una distancia $\leq d_\tau$ del punto central de la cara (x', y') previamente determinado de forma manual. Por lo tanto, la Tasa de Reconocimiento y Localización (TRL) está dada por:

$$TRL = \frac{|s_i(x, y) : d_{p,c} \leq d_\tau \wedge psr_{pico} \geq \tau_{deteccion}|}{|S|}, \quad (5.2)$$

donde $d_{p,c} = \sqrt{(x - x')^2 + (y - y')^2}$ es la distancia euclidiana entre el pico y el centro de la cara. Esta métrica es análoga a la tasa de detección, pues calcula la razón entre el número de caras detectadas por algoritmo entre el total de caras determinadas por un humano.

Cuando $d_{p,c} > d_\tau$ y $d_{p,c} \leq d'_\tau$, donde d'_τ se escoge de tal forma que la ventana de detección incluya una parte de la cara, entonces existe una localización desviada y es válido considerando que hay algoritmos basados en correlación para reconocer objetos parciales (Santiago-Ramírez y cols., 2012; Ángel González-Fraga, Kober, y Álvarez Borrego, 2006). La Tasa de Localización Desviada y Reconocimiento (TLDR) está dada por:

$$TLDR = \frac{|s_i(x, y) : d_\tau < d_{p,c} \leq d'_\tau \wedge psr_{pico} \geq \tau_{deteccion}|}{|S|}. \quad (5.3)$$

Cuando $d_{p,c} > d'_\tau$ y la agudeza del pico es menor que la agudeza umbral en una imagen que si contiene una cara, entonces sucede una falsa localización y un falso rechazo. Las veces que esto sucede está medida por la Tasa de Falsa Localización y Falso Rechazo (TFLFR):

$$TFLLFR = \frac{|s_i(x, y) : d_{p,c} > d'_\tau \wedge psr_{pico} < \tau_{deteccion}|}{|S|}. \quad (5.4)$$

Por último, cuando el pico se produce en las coordenadas correctas pero la agudeza del pico es menor que $\tau_{deteccion}$ entonces sucede un falso rechazo. Esto puede deberse a una distorsión fuerte en la apariencia facial, por lo que la agudeza de pico no da certeza al algoritmo de que realmente sea una cara. La Tasa de Falso Rechazo (TFR) está dada por:

$$TFR = \frac{|s_i(x, y) : d_{p,c} \leq d_\tau \wedge psr_{pico} < \tau_{deteccion}|}{|S|}. \quad (5.5)$$

Estas métricas asumen que todas las imágenes de prueba contienen una cara. Un buen desempeño del algoritmo en estas métricas confirmaría su desempeño en la práctica.

5.3. Resultados numéricos

En esta sección se presentan los resultados obtenidos por el método propuesto. El Algoritmo 1 fue ejecutado con $D = Yale B$. Como conjunto de imágenes de prueba P se usaron las que se muestran en la Figura 5.9. Estas escenas contienen iluminación variable, fondo homogéneo y estructurado, expresiones, sombra total y personas que no están en T . El objetivo es lograr que el filtro sea robusto a estas distorsiones. Experimentalmente se determinaron los parámetros $maxIt$ y τ a 800 y 8, respectivamente. En cada paso de la iteración se modificaban tres elementos, de los cuales se selecciona el mejor y se compara con su antecesor para seleccionar el que va a sobrevivir en la siguiente generación.

De un total de $|D| = 2414$ imágenes faciales, el algoritmo seleccionó 1224 para el conjunto de entrenamiento T . En la Figura 5.10, el gráfico de línea en la parte superior muestra el desempeño del filtro en cada paso del ciclo. Cuando ninguno de los elementos de la vecindad L ofrece una mejora, se mantiene la mejor solución hasta ese momento. En caso contrario, el valor psr promedio incrementaba. La gráfica de línea en la parte inferior muestra la variación en el tamaño del conjunto de entrenamiento. Como se puede observar, inicialmente el tamaño del conjunto tiende a aumentar. Esto debido que la solución inicial es aleatoria y usa aproximadamente la mitad de las imágenes disponibles. Por lo que hay una mayor cantidad de imágenes a seleccionar. Sin embargo, llega un punto en que la probabilidad de seleccionar elementos no usados en \mathbf{x} es pequeña. Entonces, el algoritmo empieza a seleccionar para L elementos de \mathbf{x} ocupados, es decir, aquellos con $\mathbf{x}(i) = 1$. Entonces inicia el proceso de depuración, es decir, descartar aquellas imágenes redundantes y que por sus características disminuyen el desempeño del filtro. Por esta razón se observa la disminución del tamaño de T_i mientras se mantiene o mejora la calidad. Como se puede notar, aplicar el operador O sobre \mathbf{x} implícitamente

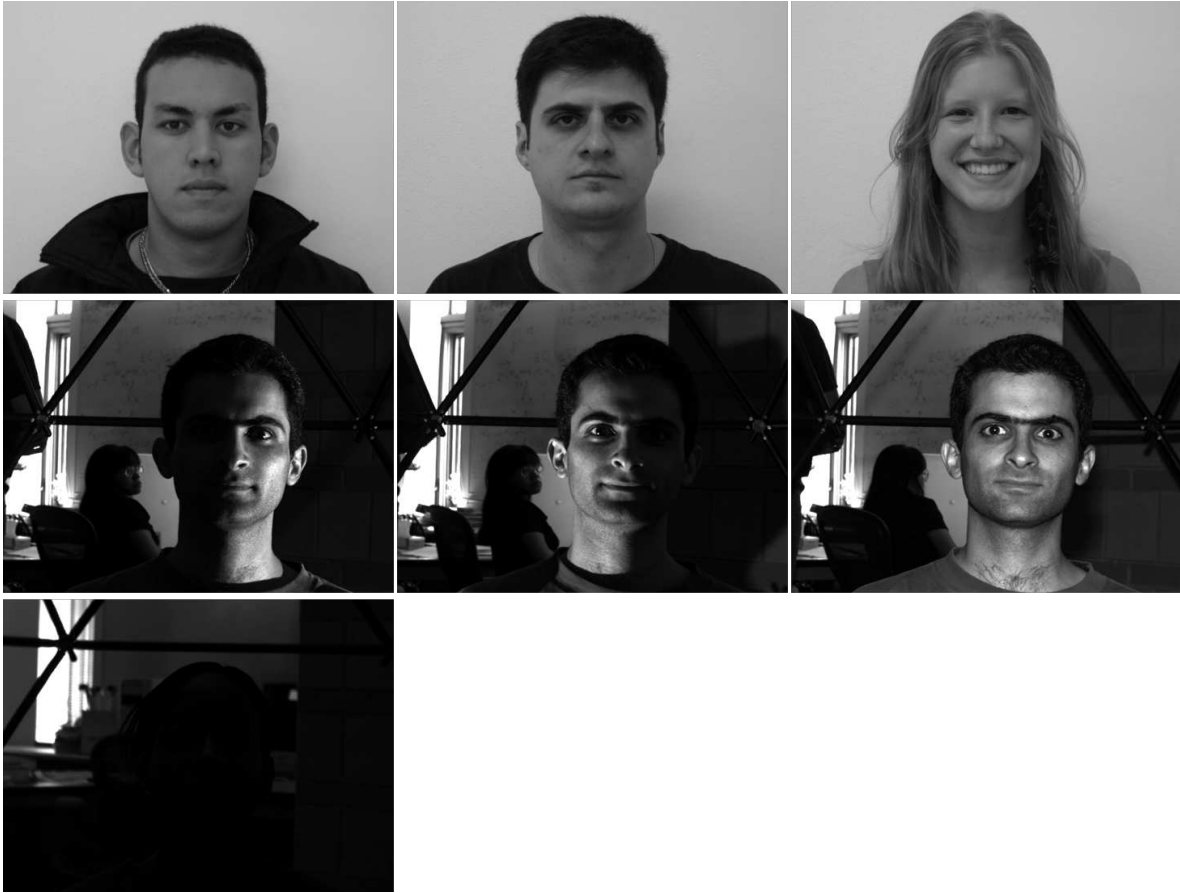


Figura 5.9: Conjunto de prueba P usado en la selección del conjunto de entrenamiento.

permite probar diferentes combinaciones de imágenes en D . Como resultado de esta operación, al final el algoritmo se queda con la mejor combinación de todas las evaluadas.

Con el conjunto T seleccionado se sintetizó un filtro de correlación mediante el Algoritmo 2. Se aplicaron seis transformaciones de rotación, dos de escala en anchura, dos de ruido aditivo blanco Gaussiano, una operación de volteado de izquierda a derecha y la imagen original. Por lo tanto, cada filtro fue entrenado con $|T| = 1224 \times 12 = 14688$ imágenes faciales. Idealmente, un filtro robusto a distorsiones produce en el centro de la cara un pico agudo y un plano de correlación menos ruidoso. Esto facilita la localización exacta y la determinación confiable de que dicha región corresponde a un rostro humano. En la Figura 5.11 se muestra el plano de correlación que produce el filtro diseñado con el método propuesto y el filtro ASEF para una misma imagen de entrada. Como se puede observar, el método propuesto produce un pico más agudo y fuerte, así como un plano menos ruidoso. En cambio, el pico producido por ASEF está rodeado de picos falso y el plano exhibe una mayor cantidad de ruido.

El método propuesto fue evaluado en términos de las métricas descritas en la Sub-

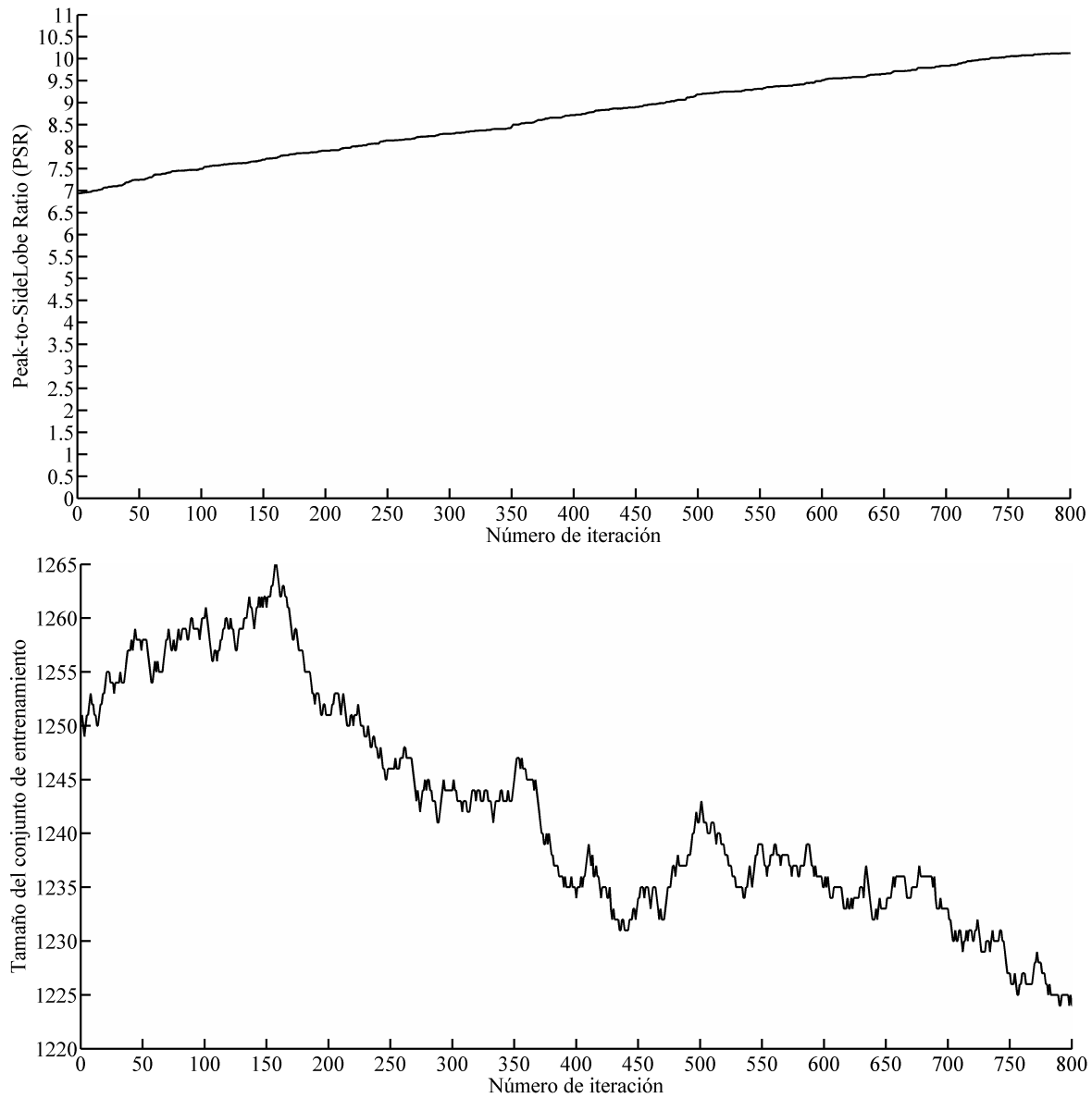


Figura 5.10: Valores PSR de las soluciones candidato en cada iteración (arriba) y disminución del tamaño del conjunto de entrenamiento debido a la optimización (abajo).

sección 5.2.2. La Tabla 5.2 muestra el desempeño obtenido por el método propuesto y el filtro ASEF sobre los conjunto de prueba. La Figura 5.12 muestra los resultados de las métricas TLR y $TLDR$ para el método propuesto (línea negra continua) y el filtro ASEF (línea negra punteada). Cuando los rostros no están en el conjunto de entrenamiento sobre escenas con fondo homogéneo, el método propuesto supera el desempeño del filtro ASEF tanto para caras frontales como con ligeras variaciones en pose. Cuando las imágenes tienen fondo estructurado y variaciones en iluminación el desempeño es

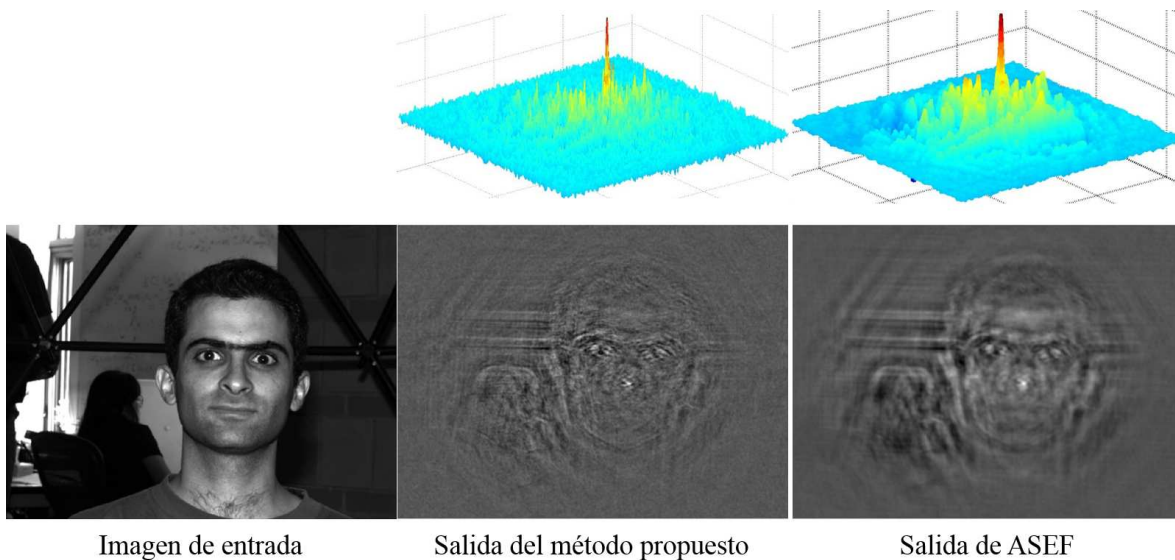


Figura 5.11: De izquierda a derecha: imagen original, plano de correlación del método propuesto y plano de correlación del filtros ASEF. Para cada plano de correlación, arriba se muestra la versión 3D y abajo la versión 2D.

el siguiente. Como se puede observar en la Figura 5.12, el método propuesto es mejor con los conjuntos YaleB22, YaleB27, YaleB30, YaleB32 y YaleB33. Los rostros en estos conjuntos presentan ligeras variaciones en pose, además de las variaciones en iluminación. El filtro ASEF, generalmente se desempeña mejor con rostros frontales o con una variación en pose menor al soportado por el método propuesto. En el caso de la TLDR, aunque el método propuesto obtiene valores altos, es aceptable considerando la existencia de algoritmos capaces de reconocer caras parciales. En términos de TFLFR, el método propuesto obtiene un buen desempeño donde supera al filtro ASEF. Mientras que con la TFR, ambas estrategias obtienen un desempeño óptimo.

Los resultados numéricos evidencian que el método propuesto produce el mejor desempeño cuando la cara en la escena es frontal o con ligeras variaciones en pose. Un aspecto a considerar para un buen desempeño es la construcción del banco de filtros. Experimentalmente se encontró que el método propuesto trabaja bien en el rango de dimensiones de 52×52 a 260×260 píxeles. Por otra parte, ASEF trabaja bien en el rango de dimensiones de 64×64 a 260×260 píxeles. En este trabajo de tesis, las caras detectadas son usadas para identificar a la persona e iniciar el filtro de seguimiento.

5.4. Conclusiones

En este capítulo se ha propuesto un algoritmo de detección facial basado en filtros de correlación compuesto no lineales. Dado un conjunto de imágenes faciales disponibles y un conjunto de escenas de prueba, se seleccionan solo las imágenes que contribuyen a

Tabla 5.2: Desempeño del método propuesto y del filtro ASEF en la detección facial. En verde las celdas con los mejores desempeños TLR y los peores para el resto de las métricas.

| Conjunto de prueba | Filtro | TLR | TLDR | TFLFR | TFR |
|--------------------|-----------|--------|-------|-------|------|
| Frontales | SDF K-Law | 96.00 | 3.63 | 0.38 | 0.00 |
| | ASEF | 87.75 | 1.38 | 0.00 | 0.00 |
| Pose | SDF K-Law | 85.80 | 13.60 | 0.60 | 0.00 |
| | ASEF | 83.20 | 4.50 | 0.00 | 0.00 |
| YaleB11 | SDF K-Law | 87.50 | 12.50 | 0.00 | 0.00 |
| | ASEF | 92.19 | 6.25 | 0.00 | 0.00 |
| YaleB22 | SDF K-Law | 100.00 | 0.00 | 0.00 | 0.00 |
| | ASEF | 98.46 | 0.00 | 0.00 | 0.00 |
| YaleB27 | SDF K-Law | 93.85 | 6.15 | 0.00 | 0.00 |
| | ASEF | 89.23 | 10.77 | 0.00 | 0.00 |
| YaleB28 | SDF K-Law | 83.08 | 4.62 | 12.31 | 0.00 |
| | ASEF | 89.23 | 3.08 | 0.00 | 0.00 |
| YaleB29 | SDF K-Law | 58.46 | 38.46 | 3.08 | 0.00 |
| | ASEF | 72.31 | 12.31 | 1.54 | 0.00 |
| YaleB30 | SDF K-Law | 86.15 | 13.85 | 0.00 | 0.00 |
| | ASEF | 81.54 | 7.69 | 0.00 | 0.00 |
| YaleB32 | SDF K-Law | 96.92 | 3.08 | 0.00 | 0.00 |
| | ASEF | 83.08 | 1.54 | 0.00 | 0.00 |
| YaleB33 | SDF K-Law | 43.08 | 47.69 | 9.23 | 0.00 |
| | ASEF | 20.00 | 24.62 | 16.92 | 0.00 |
| YaleB34 | SDF K-Law | 56.92 | 36.92 | 6.15 | 0.00 |
| | ASEF | 63.08 | 16.92 | 3.08 | 0.00 |

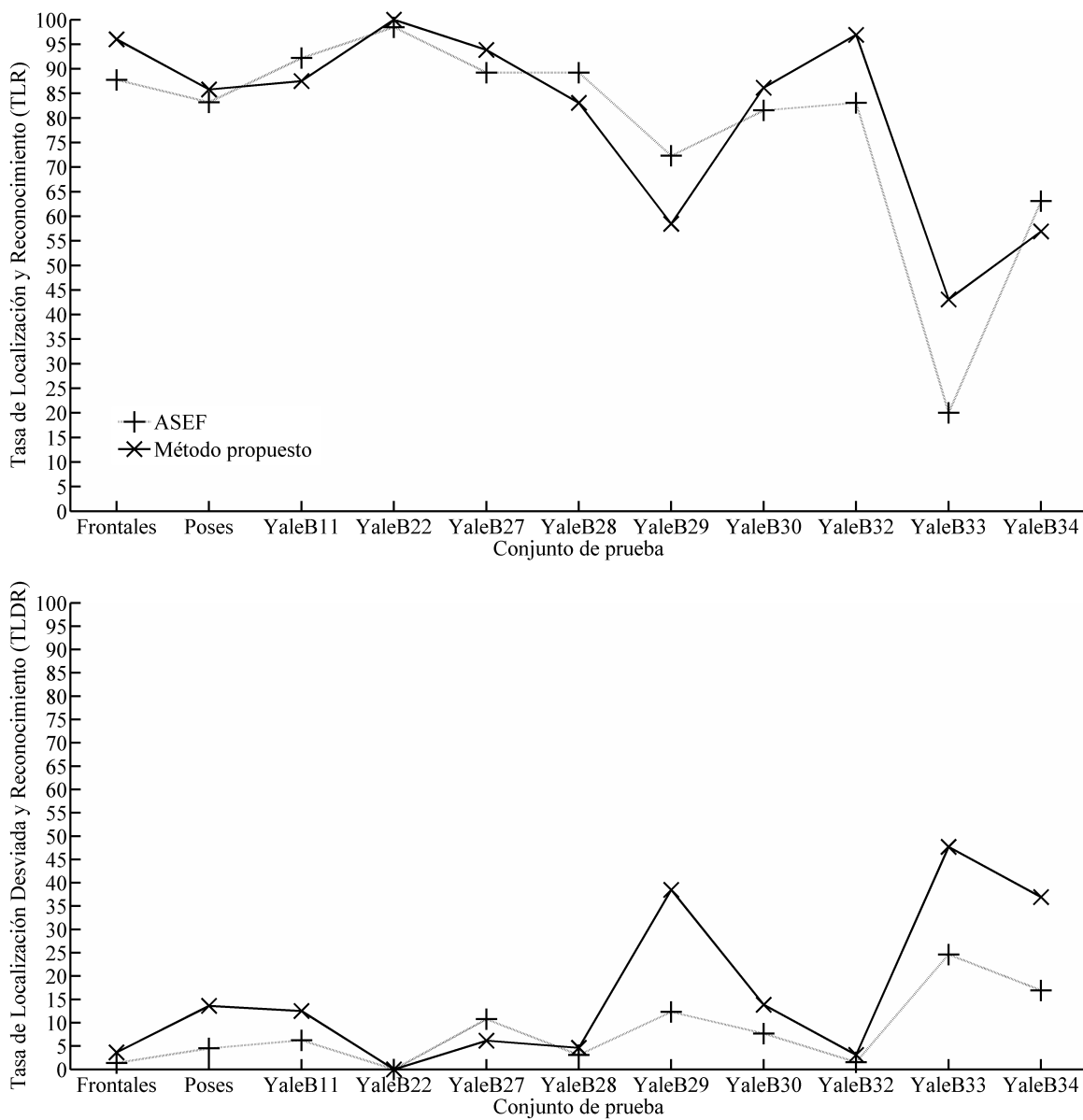


Figura 5.12: Desempeño del método propuesto contra el filtro ASEF en términos de las métricas TLR (arriba) y TLDR (abajo).

producir picos agudos y fuertes sobre las imágenes de prueba. Al conjunto seleccionado, se le aplica un conjunto de transformaciones para modelar aquellas distorsiones que no son muy fáciles de obtener con cámaras. Con el conjunto seleccionado y sus transformaciones, se construye un banco de filtros. Una imagen arbitraria se correlaciona con los filtros en el banco para determinar si esta contiene o no caras. Los experimentos y sus resultados evidencian varias ventajas del método propuesto. Primero, una selección inteligente del conjunto de entrenamiento minimiza la cantidad de imágenes de

entrenamiento mientras se maximiza el desempeño del filtro de correlación. Segundo, el promediado enfatiza las características comunes de las imágenes faciales y realza los bordes de las características faciales los cuales proveen a los filtros un mejor acoplamiento. Tercero, el filtro SDF entrenado con filtros ASEF ofrece una mayor tolerancia a rotación (sobre el plano y fuera del plano) que un filtro ASEF simple. Cuarto, el uso de un banco de filtros ayuda a lidiar con el problema de escala, sin embargo, es importante definir un buen rango de dimensiones de los filtros. Quinto, que un vector binario es una representación que permite la aplicación exitosa de algoritmos de optimización al problema de selección del conjunto de entrenamiento. Finalmente, el método propuesto permite detectar en escenas con fondos complejos caras factibles de ser identificadas por los algoritmos de reconocimiento facial existentes.

Capítulo 6

Seguimiento facial mediante filtros adaptativos

El seguimiento facial tiene como objetivo estimar la ubicación y dimensión de la cara de una persona en cada imagen de un vídeo. Ésta tarea permite la recopilación de información que puede ser usada para predecir futuros eventos no deseados o simplemente para mejorar la experiencia de la persona en el ambiente en que se encuentra. Sin embargo, el seguimiento facial es una tarea desafiante porque la apariencia de la cara sufre cambios debido a la iluminación variable causada por las diferentes fuentes de iluminación en el ambiente, ruido, oclusión (parcial y total), rotación (en el plano y fuera del plano), expresiones, escala, entre otros.

Aunque los algoritmos de seguimiento facial reportados en la literatura trabajan bien para ambientes restringidos, el seguimiento bajo ambientes no restringidos sigue siendo un problema abierto. En este capítulo se propone un algoritmo de seguimiento facial capaz de adaptarse a las variaciones en la apariencia facial y de recuperarse de fallos en el seguimiento. Para este propósito, se explota la continuidad en el tiempo de los vídeos y la capacidad de entrenamiento en línea de los filtros de correlación.

6.1. Seguimiento facial mediante filtros adaptativos

Sea S un vídeo capturado por una cámara donde aparece la cara $f(x, y)$ de una persona. El seguimiento consiste en detectar $f(x, y)$ cuando aparezca por primera vez y luego localizarla en las imágenes subsecuentes. En otras palabras, el seguimiento consiste en localizar instancias de $f(x, y)$ en subsecuentes imágenes de un vídeo.

Una forma simple de realizar el seguimiento sería ejecutar la tarea de detección facial en cada imagen $s_i(x, y) \in S$. Esto requiere de una operación adicional de clasificación de las diferentes caras detectadas para encontrar la cara objetivo. Esta estrategia, aunque simple, no es práctica por dos motivos. Primero, la tarea de detección sobre cada imagen tiene una complejidad computacional alta. Segundo, la tarea de clasificación requiere comparar la cara objetivo $f(x, y) \in s_i(x, y)$ contra todas las caras candidatas detectadas

en $s_{i+1}(x, y)$.

En este trabajo de investigación se propone un algoritmo de seguimiento facial que explota tres propiedades de los filtros de correlación. Primero, se explota la propiedad actualización en línea para que el filtro se adapte a los cambios de apariencia facial durante el seguimiento. Segundo, aprovecha la propiedad de que pueden ser entrenados para una cara específica. Tercero, la propiedad de localizar y clasificar simultáneamente, lo que evita la necesidad del uso de un clasificador.

La Figura 6.1 muestra el esquema general de seguimiento que se emplea en este trabajo. Primero, la imagen de entrada es pre-procesada para mejorar su calidad. En esta etapa, se mejora la iluminación global y elimina la variabilidad en la iluminación. Además, se busca mejorar los bordes de los objetos en la imagen de entrada. Segundo, se realiza la localización de la cara objetivo mediante la técnica de correlación. La región devuelta por la localización es verificada en el tercer paso para determinar si realmente corresponde a una cara. Si la región localizada realmente corresponde una cara, entonces en el paso cuatro se actualiza el filtro de seguimiento. Este proceso se repite hasta analizar la última imagen en S .

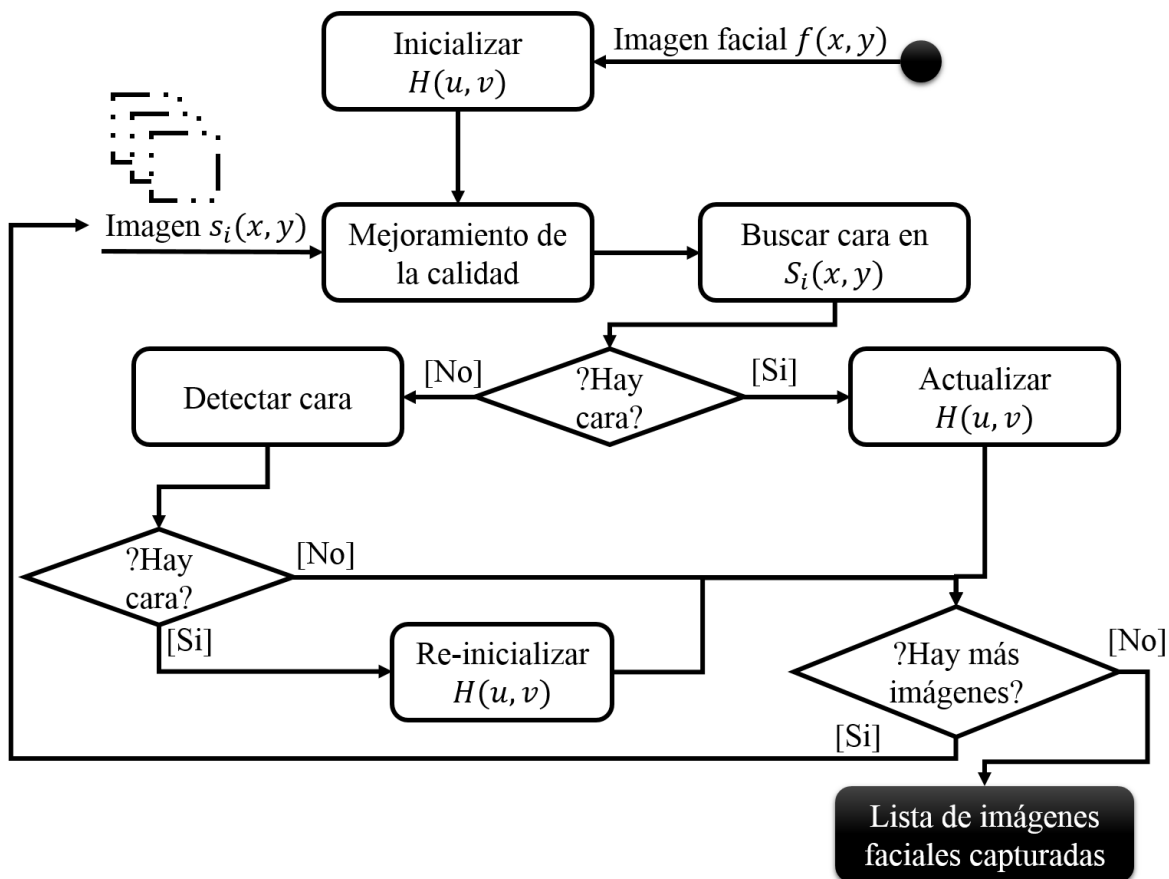


Figura 6.1: Esquema general del método de seguimiento propuesto.

Los métodos de seguimiento facial basados en filtros de correlación usan una ventana de seguimiento w con un tamaño constante. Algoritmos como MOSSE se adaptan a las variaciones faciales, incluso variaciones en escala. Sin embargo, al ser constante la dimensión de la ventana, no es posible capturar dentro de w la cara completa. Esto se entiende mejor con la Figura 6.2, donde se muestran dos imágenes diferentes de un mismo vídeo. Las cara en ambas imágenes son diferentes en tamaño, mientras que la dimensión de w es constante. Mientras en la imágenes de la izquierda w captura la cara completa, en la imagen de la derecha solo se captura una parte de la cara. Aunque el seguimiento es correcto, pues la ventana de seguimiento se adapta correctamente a la variación en escala, en algunas aplicaciones la región capturada no es útil debido a que requieren una imagen facial completa. Además, la variación en escala degrada suavemente el desempeño del seguidor hasta perder completamente el objetivo.



Figura 6.2: Ventana de seguimiento w con una dimensión constante sobre dos imágenes diferentes de un mismo vídeo S . A la izquierda la imagen $s_1(x, y)$ y a la derecha la imagen $s_{1066}(x, y)$.

El anterior problema plantea la necesidad de un algoritmo que pueda modificar la dimensión de w cuando la variación en escala ya no captura la imagen facial completa. Para lograr esto, en esta tesis se propone el uso de la detección facial cuando el seguimiento pierde la habilidad de capturar el objeto completo. La nueva dimensión de w es dada por la nueva detección. La Figura 6.4 muestra el diagrama básico de flujo del algoritmo propuesto para el seguimiento facial.

El algoritmo recibe como argumento un vídeo S y realiza el seguimiento de la siguiente manera. Primero, obtiene la imagen $s_i(x, y)$. Segundo, verifica si ya existe un filtro de seguimiento $H(u, v)$. Tercero, si $H(u, v)$ existe, entonces se correlaciona con la imagen de prueba $s_i(x, y)$. Cuarto, si la imagen de entrada contiene la cara objetivo, el plano de correlación exhibe un pico cuya agudeza psr es $\geq \tau$. Otras caras en $s_i(u, v)$ producirán picos con $psr \leq \tau$. En el quinto paso, la región del pico se extrae y en el sexto paso es verificada para determinar si realmente es una cara y corresponde a la persona objetivo. En (Burton, Jenkins, Hancock, y White, 2005) se demuestra que una

imagen facial promedio es un buen candidato para una representación facial robusta. Por esta razón, en esta tesis las imágenes faciales previamente capturadas en el mismo seguimiento son promediadas. En la Figura 6.3 se muestra el poder del promediado. Diez imágenes faciales con iluminación variable, algunas casi totalmente oscuras, son promediadas y la imagen promedio resulta con mejor iluminación global y no se pierden los bordes de las características faciales. El proceso de verificación se realiza aplicando correlación entre la imagen facial promedio y la región extraída.



Figura 6.3: En la parte inferior imagen facial promedio resultante de las diez imágenes faciales en la parte superior.

Si en la verificación se determina que la región si es la cara objetivo, entonces en el séptimo paso se guardan la ubicación y dimensión de la cara localizada. Finalmente, si se han realizado lc localizaciones exitosas, entonces se reinicia el filtro de seguimiento facial $H(u, v)$. En caso contrario, solo se realiza una actualización de $H(u, v)$. Si no existe $H(u, v)$ (segundo paso) o si $psr < \tau$ (cuarto paso) o si en el proceso de verificación resulta que la región extraída no es la cara objetivo (sexto paso); entonces se ejecuta la detección facial sobre $s_i(x, y)$. Si se detecta una cara, entonces se construye un nuevo filtro de seguimiento $H(u, v)$ con una nueva dimensión para w . Cada uno de estos pasos se repite mientras existan imágenes en el vídeo.

Cuando se detecta $f(x, y)$ por primera vez en $s_i(x, y)$, el objetivo es localizarlo en la imagen $s_{i+1}(x, y)$. Localizar una instancia de $f(x, y)$ sobre $s_{i+1}(x, y)$ es una tarea

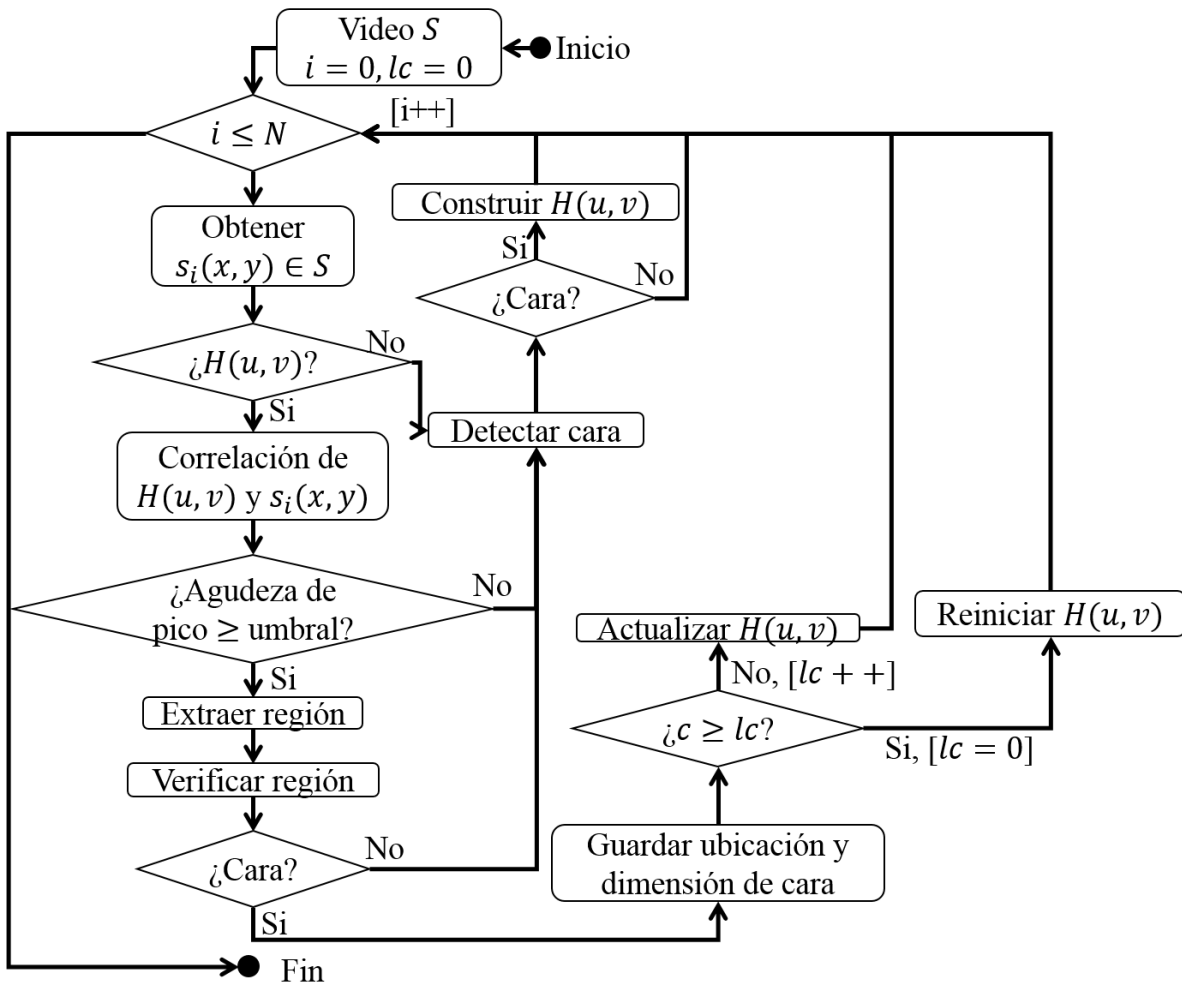


Figura 6.4: Diagrama de flujo del algoritmo propuesto para el seguimiento facial usando un filtro de correlación adaptable.

muy complicada considerando que éste pudiera presentar un fondo complejo, objetos similares, entre otros. Por esta razón, $H(u, v)$ se crea en un principio con un conjunto de entrenamiento T creado mediante la aplicación de transformaciones a $f(x, y)$. Algunas transformaciones que pueden ser aplicadas a $f(x, y)$ para generar T son rotación, ruido, escala y volteado. El filtro inicial está dado por (Santiago-Ramírez y cols., 2012):

$$H(u, v) = \frac{1}{N} \sum_{i=1}^N F_i^k(u, v), \quad (6.1)$$

donde $N = |T|$ y $F_i^k(u, v)$ son respectivamente el tamaño del conjunto de entrenamiento y la TF con el filtrado no lineal de $f_i(x, y)$. En la Figura 6.5 se muestra la salida de correlación de la imagen $s_{i+1}(x, y)$ con un filtro entrenado con $f(x, y)$, y ocho versiones generada aplicando transformaciones a $f(x, y)$, detectado en $s_i(x, y)$.

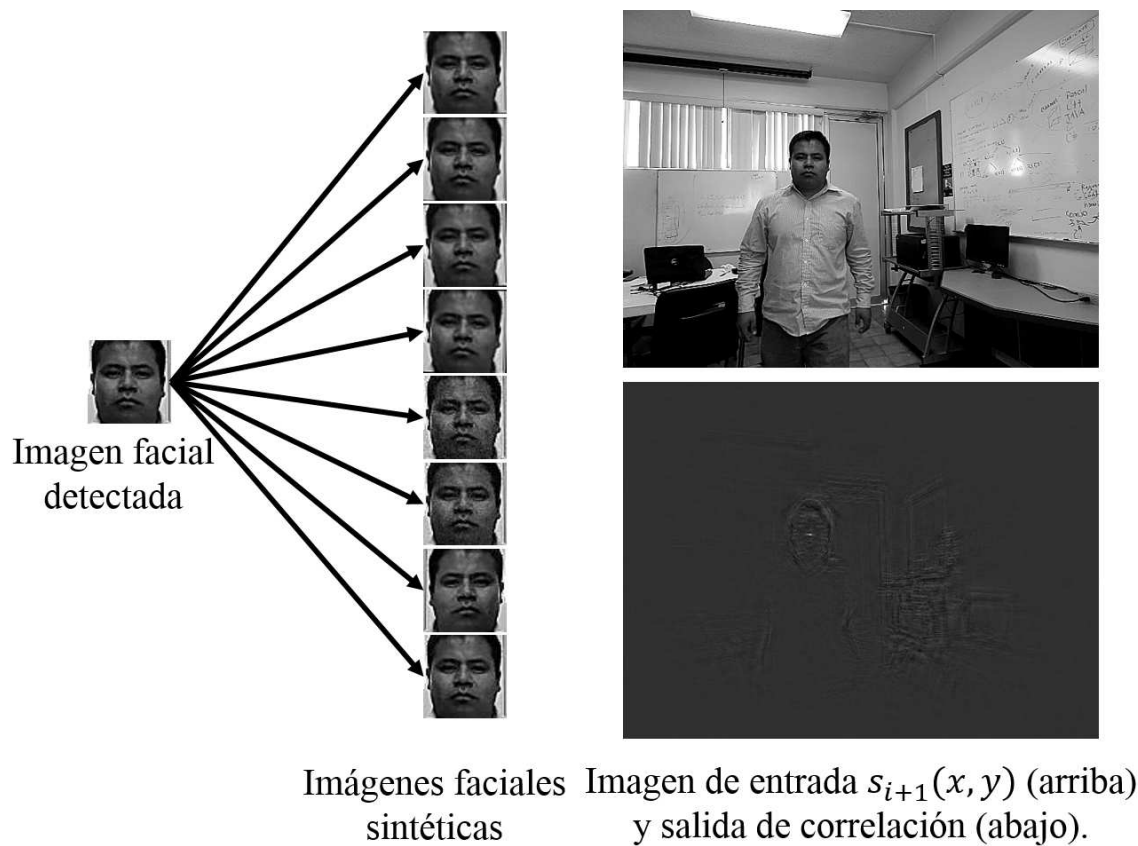


Figura 6.5: Salida de correlación de un filtro generado con una imagen facial y versiones generadas mediante la aplicación de transformaciones.

La actualización de $H(u, v)$ se logra mediante (Santiago-Ramírez y cols., 2012):

$$H(u, v)_{i+1} = (1 - \eta) \frac{n-1}{n} H_i(u, v) + \eta \frac{1}{n} F_i^k(u, v), \quad (6.2)$$

donde $0 \leq \eta \leq 1$ es la tasa de aprendizaje. Este valor controla qué tan rápido se adapta el seguidor a los cambios de la apariencia facial. Puede ser determinado considerando si el objeto permanece rígido o es un objeto flexible como la cara humana. Un valor de $\eta = 0$ no permite la adaptación, mientras que un valor de $\eta = 1$ causa que la apariencia del filtro se base solamente en la última imagen procesada. En este trabajo se determinó experimentalmente que un valor adecuado para η está entre 0.01 y 0.3. Los experimentos realizados en este capítulo usaron para η un valor de 0.0625 que provee al seguidor la habilidad de localizar solamente instancias que pueden ser reconocidas. Es decir, instancias frontales y con ligeras variaciones en rotación (en el plano y fuera del plano).

Como se puede notar en la Ecuación 6.2, en el proceso de actualización se resta peso a las muestras anteriores mientras se resalta el peso de la nueva cara en el filtro.

Esto causa que la apariencia del filtro se base tanto en las caras detectadas previamente como en la cara actual. El filtro actualizado contiene la suficiente información de la cara objetivo para producir un pico agudo sobre una imagen desconocida que contenga una instancia de la cara objetivo. El proceso de reiniciar ayuda a evitar el sobre-ajuste del filtro, que es cuando el filtro contiene información tan detallada que no puede reconocer instancias del objeto que presentan ligeras variaciones. En este trabajo, se determinó experimentalmente que en cada 60 localizaciones exitosas es necesario el re-inicio del filtro de seguimiento.

6.2. Resultados

En esta Sección se presentan los resultados numéricos obtenidos por el método propuesto al procesar vídeos grabados bajo condiciones no restringidos y personas no cooperativas.

6.2.1. Configuración de los datos de prueba

La Tabla 6.1 muestra los datos de los vídeos usados en la evaluación del método propuesto. Los vídeos *davidin300*, *trellis70*, *ming-hsuan* y *DudekData* fueron descargados de (D. Ross, Lim, Lin, y Yang, 2011), mientras que el vídeo *Ever02* fue obtenido de (Santiago-Ramírez y cols., 2012). El resto de los vídeos fueron grabados en el Laboratorio de Prácticas Avanzadas 2 (LPA2) de la Universidad Autónoma de Baja California. Estos vídeos fueron grabados a 30 cuadros por segundo, cada imagen con una resolución de 480×640 píxeles. También, se puede observar la cantidad de imágenes contenidas en cada vídeo, así como cuantos de ellos contienen la cara objetivo y cuantos no la contienen. En cuanto a la iluminación se consideró iluminación intensa (II), sombra total (ST) y sombra local (SL). Otras fuentes de variación caracterizada en los vídeos son la oclusión, expresiones, pose y escala. También se considera si la cámara y la persona se mueven o no.

La Figura 6.6 muestra una imagen de muestra de cada vídeo usado en la evaluación del algoritmo propuesto. En cada imagen de cada vídeo, se etiquetó manualmente la ubicación y dimensión de cada cara. También, se etiquetó aquellas imágenes que no contenían la cara objetivo para fines de evaluación. Esto con el fin de evaluar el desempeño del método propuesto usando las métricas de la Sección 5.2.2 del Capítulo 5.

6.2.2. Resultados numéricos

Los resultados obtenidos por el método propuesto se muestran en la Tabla 6.2. En los vídeos *ming-hsuan*, *DudekData*, *Ever02* y *DSCN0973* la iluminación global no afecta la visibilidad del rostro de las personas. Por esta razón, a pesar de la presencia de oclusión parcial y expresiones faciales, el método propuesto obtuvo una *TLR* superior al 97%, mientras que se obtuvo un desempeño óptimo en las métricas *TLDR* y *TFLFR*.

Tabla 6.1: Características de los vídeos usados en la evaluación. Notación: con cara (CC), sin cara (SC), iluminación intensa (II), sombra total (ST), sombra local (SL), oclusión (Oc) y expresiones (Ex).

| Nombre | Imágenes | | | Iluminación | | | Oc | Ex | Cámara | | Pose | Escala |
|------------|----------|-----|-------|-------------|----|----|----|----|--------|------|------|--------|
| | CC | SC | Total | II | ST | SL | | | Móvil | Fijo | | |
| davidin300 | 462 | 0 | 462 | × | | × | × | × | | | × | × |
| trellis70 | 500 | 0 | 500 | × | × | × | | | × | | × | × |
| ming-hsuan | 1680 | 124 | 1804 | × | | | × | × | | × | × | × |
| DudekData | 1127 | 18 | 1145 | × | | × | × | × | × | | × | × |
| Ever02 | 800 | 0 | 800 | × | | | × | × | × | | × | × |
| DSCN0971 | 33 | 87 | 120 | × | | × | | | | × | × | × |
| DSCN0972 | 773 | 277 | 1050 | × | | × | | | | × | × | × |
| DSCN0973 | 905 | 115 | 1020 | × | | × | × | × | | × | × | × |
| DSCN0974 | 41 | 79 | 120 | × | | × | | × | | × | × | × |
| DSCN0975 | 101 | 139 | 240 | × | | × | | | | × | × | × |
| DSCN0976 | 443 | 217 | 660 | × | | × | | | | × | × | × |
| DSCN0977 | 98 | 112 | 210 | × | | × | | | | × | × | × |
| DSCN0978 | 462 | 80 | 542 | × | × | × | | | | × | × | × |
| DSCN0979 | 395 | 445 | 840 | × | × | × | | × | | × | × | × |
| DSCN0980 | 28 | 92 | 120 | × | × | | | | | × | × | × |
| DSCN0982 | 161 | 19 | 180 | × | | × | | | | × | × | × |
| DSCN0983 | 86 | 4 | 90 | × | | × | | | | × | × | × |

En el caso de los vídeos davidin300, DSCN0972, DSCN0975, DSCN0978, DSCN0979 y DSCN0982, las imágenes principalmente son afectados por iluminación intensa, sombras locales, expresiones faciales y pose. Estas variaciones causan una disminución en el desempeño del método propuesto, aunque superior al 80%. Sin embargo, de ese desempeño aceptable, los vídeos DSCN0978, DSCN0979 y DSCN0982 obtuvieron altos valores para la *TFR*. Esto debido a que la cara no era nítida en dichos vídeos debido a la iluminación intensa, sombra total, poses y expresiones más fuertes que los otros vídeos.

Los vídeos trellis70 y DSCN0980 en donde el rostro de la persona está sombreado en su totalidad y dificulta el mejoramiento de los bordes y otros detalles faciales que son importantes para el acoplamiento perfecto se obtuvo un bajo desempeño. El peor desempeño lo obtuvieron los vídeos DSCN0971 y DSCN0974, que aunque la iluminación global es buena, la variación en pose fue lo que dificultó la detección y el seguimiento correcto de las caras.

En los vídeos davidin300, trellis70, DSCN0972, DSCN0976 y DSCN0979 el rostro de la persona desapareció de la escena respectivamente 1, 2, 3, 4 y 3 veces causando fallos en el seguimiento. Sin embargo, el algoritmo propuesto fue capaz de recuperarse en cuanto la cara volvía a aparecer en la escena. Cuando la cara aparece en la escena pero del mismo tamaño que cuando desapareció, entonces la recuperación es con el mismo filtro de seguimiento. En otro caso, cuando la cara aparece con un tamaño diferente,



Figura 6.6: Imagen de muestra de los vídeos evaluados.

éste ya no va a ser localizado por el filtro actual. Por lo tanto, es necesario realizar una nueva detección sobre la imagen actual asegurándose que la identidad se mantenga. El que la cámara fuera móvil o fija no influyó en el desempeño del algoritmo, pues la principal fuente de distorsión que afecta el desempeño proviene de la iluminación, las expresiones faciales y la rotación (en el plano y fuera del plano).

6.3. Conclusión

Se ha propuesto un algoritmo de seguimiento facial usando un filtro de correlación capaz de adaptarse a las variaciones faciales y recuperarse de los fallos en el seguimiento. Los ambientes no controlables y personas no cooperativas representan un gran desafío en el seguimiento facial, pues la iluminación variable, rotación (en el plano y fuera del plano), expresiones, entre otros, modifican la apariencia facial de una imagen a otra. Los resultados experimentales demuestran que el algoritmo propuesto ofrece un desempeño superior al 97 % bajo condiciones de iluminación global que permiten obtener una

Tabla 6.2: Desempeño del método propuesto en la tarea de seguimiento facial. En verde las celdas con los mejores desempeños TLR y los peores para el resto de las métricas.

| Conjunto | TLR | TLD | TFL | TFR |
|------------|-------|------|-------|-------|
| davidin300 | 87.23 | 0.00 | 0.87 | 1.08 |
| trellis70 | 61.60 | 9.60 | 8.60 | 7.60 |
| ming-hsuan | 98.82 | 0.00 | 0.00 | 0.70 |
| DudekData | 97.69 | 0.00 | 0.00 | 1.15 |
| Ever02 | 99.63 | 0.00 | 0.00 | 0.38 |
| DSCN0971 | 48.48 | 0.00 | 0.00 | 45.45 |
| DSCN0972 | 89.65 | 1.03 | 6.47 | 2.33 |
| DSCN0973 | 98.12 | 0.00 | 0.00 | 0.88 |
| DSCN0974 | 58.54 | 0.00 | 0.00 | 39.02 |
| DSCN0975 | 84.16 | 0.00 | 0.00 | 3.96 |
| DSCN0976 | 79.68 | 0.23 | 3.16 | 14.67 |
| DSCN0977 | 78.57 | 0.00 | 0.00 | 21.43 |
| DSCN0978 | 86.80 | 0.00 | 0.00 | 12.77 |
| DSCN0979 | 89.37 | 0.00 | 0.00 | 9.11 |
| DSCN0980 | 60.71 | 0.00 | 17.86 | 10.71 |
| DSCN0982 | 88.24 | 0.00 | 0.00 | 5.88 |
| DSCN0983 | 72.55 | 0.00 | 0.00 | 9.80 |

imagen facial de buena calidad. En otro caso, cuando la cara es afectado por sombras locales, sombras totales y locales, aunque el desempeño obtenido es menor que en el caso anterior, es superior al 80 %. El uso de correlación en el seguimiento, reduce la cantidad de detecciones requeridas. Pues dado una cara detectada, ésta es usada para construir un filtro de correlación para localizar instancias de la misma en subsecuentes imágenes de un vídeo. Además, el uso de la detección adapta la dimensión de la ventana de seguimiento a una cara escalada.

Capítulo 7

Selección automática del conjunto de entrenamiento

Como se vio en el Capítulo 3, los filtros de correlación se generan de un conjunto de entrenamiento T el cual contiene las imágenes más representativas de una clase facial. Por esta razón, su desempeño depende en mayor medida de una adecuada selección de T . En este capítulo, se presenta una metodología basada en algoritmos de optimización numérica para la selección automática del conjunto de entrenamiento T . Dado un algoritmo de optimización, la metodología propuesta usa sus principales mecanismos para iterativamente examinar un conjunto de imágenes disponibles con el fin de encontrar el mejor sub-conjunto para la construcción del filtro. Para este fin, se proponen tres funciones objetivo como criterios de optimización para la selección del conjunto de entrenamiento.

7.1. Trabajo relacionado

El conjunto de entrenamiento T debe ser pequeño por conveniencia computacional (Vijaya-Kumar, 1992) y debe contener las principales distorsiones que se espera estén presentes en el objetivo (Vijaya-Kumar, 1992; Rodriguez-Perez, 2012). Además, es importante considerar que usar pocas o muchas imágenes puede disminuir el desempeño del filtro (Rodriguez-Perez, 2012). El conjunto de entrenamiento T puede ser seleccionado mediante métodos semi-supervisados, métodos convencionales o métodos basados en algoritmos de optimización numérica.

En los métodos semi-supervisados un diseñador de filtros analiza visualmente un conjunto de imágenes disponibles T_{id} para seleccionar las imágenes más representativas para el conjunto T . Esta forma de selección tiene dos principales desventajas. Primero, no ofrece ningún procedimiento óptimo para seleccionar las imágenes más representativas, y segundo, no es práctico porque la cantidad de esfuerzo humano requerido es proporcional al número de filtros a diseñar. En (Santiago-Ramírez y cols., 2012) se presenta un trabajo de procesamiento facial basado en filtros de correlación donde el

conjunto T fue seleccionado manualmente. En dicho trabajo se usó la base datos Yale B que contiene 64 diferentes variaciones en iluminación. Para cada persona se seleccionaron las cuatro imágenes más representativas para el conjunto T ; uno con sombra en el perfil derecho, otro con sombra en el perfil izquierdo, otro con sombra en toda la cara y por último una cara con iluminación homogénea.

Los métodos convencionales son aquellos métodos que no requieren de la supervisión de un diseñador de filtros pero que siguen una estrategia de selección muy básica que no logra capturar las imágenes más representativas de T_{id} o únicamente prueban una sola combinación de imágenes. En (Vijaya-Kumar, 1992) se describe un método que consiste en distribuir N imágenes uniformemente sobre el rango de distorsiones esperados. En dicho trabajo, este método de selección de T se ejemplifica con el siguiente ejemplo. Si la distorsión esperada es la rotación en el plano, entonces rotando una simple imagen facial en incrementos de 6° se obtienen 60 imágenes faciales que pueden ser usadas como conjunto de entrenamiento. Esas 60 imágenes faciales se usan para sintetizar un filtro compuesto invariante a rotación. En (Hassebrook y Rahmati, 1993) se propone un enfoque similar para la rotación fuera del plano (pose) en el reconocimiento de objetos $3D$. Sin embargo, hay distorsiones en donde no es posible obtener un conjunto de muestras uniformemente distribuido, por ejemplo, el causado por las sombras locales presentes en las imágenes faciales capturadas en ambientes no restringidos. En este caso, un enfoque más apropiado consiste en capturar de una persona varias muestras faciales con sombras locales. Todas las imágenes capturadas son consideradas como el conjunto T_{id} . Entonces, usando un umbral de correlación τ_c se seleccionan las imágenes de entrenamiento tal y como se propone en (Vijaya-Kumar, 1992; Rodriguez-Perez, 2012). Primero, se selecciona una única imagen $f_i(x, y) \in T_{id}$ que es usada para entrenar un filtro \mathbf{h} . Después \mathbf{h} es correlacionado contra todas las imágenes en T_{id} . La imagen $f_j(x, y) \in T_{id}$ con el cual \mathbf{h} obtuvo el peor desempeño es seleccionado como una imagen de entrenamiento adicional. El conjunto de entrenamiento ahora consiste de dos imágenes, es decir, $T = \{f_i(x, y), f_j(x, y)\}$, los cuales son usados para entrenar un nuevo filtro \mathbf{h} . El nuevo filtro se correlaciona otra vez contra todas las imágenes restantes en T_{id} , y una vez más, la imagen con el cual el filtro obtuvo su peor desempeño se selecciona como imagen de entrenamiento adicional. Este proceso continúa hasta que las imágenes restantes alcancen un grado de similitud $\geq \tau_c$. En (Liebowitz, 1989) se propone un método similar, donde al inicio se selecciona de forma aleatoria un conjunto de entrenamiento T . Entonces, se añaden más imágenes de entrenamiento a dicho conjunto usando el mismo proceso del método descrito previamente.

El problema de selección del conjunto de entrenamiento T consiste en seleccionar el mejor subconjunto de T_{id} . Por lo tanto, este problema puede ser planteado como un problema de optimización. En (Diaz-Ramirez, Cuevas, Kober, Trujillo, y Awwal, 2014) se propone un método que combina dos algoritmos de optimización para la síntesis de filtros de correlación compuestos. Primero, la administración de la población se realiza mediante Strength Pareto Evolutionary Algorithm (SPEA2) (Zitzler, Laumanns, y Thiele, 2001). Segundo, los individuos en la población son representados como cadenas de longitud variable. Entonces, se aplican operadores genéticos a los individuos me-

diante Speciation Adaption Genetic Algorithm (SAGA) (Harvey, 1992). Este algoritmo basado en optimización diseña solo un tipo específico de filtros llamado Multi-Objective Synthetic Discriminant Function (MOSDF). Tampoco considera el uso de otros algoritmos de optimización más simples y más rápidos.

Los métodos semi-supervisados y métodos convencionales seleccionan un conjunto de entrenamiento con desempeño limitado. Esto debido a que analizan solo una combinación de la muchas que pudiera obtenerse del conjunto T_{id} . Por otra parte, el método basado en optimización no considera el uso de otros algoritmos de optimización ni la síntesis de otros filtros de correlación existentes. Además, estos algoritmos no han sido evaluados sistemáticamente para el reconocimiento facial en el contexto biométrico. Para resolver este problema, en esta tesis se propone una metodología basada en algoritmos de optimización numérica para la selección de conjuntos de entrenamiento para la síntesis de filtros compuestos como los reportados en la literatura.

7.2. Fundamentos de los algoritmos de optimización

La optimización es un problema que consiste en seleccionar la mejor alternativa de entre un número de posibles *estados* (Schwefel, 1995). En (Nocedal y Wright, 1999) se describe como un proceso iterativo. Se inicia con una suposición inicial de valores óptimos de las variables y genera una secuencia de estimaciones mejoradas hasta alcanzar una *solución* x^* . Para moverse de una iteración a otra, las estrategias hacen uso de los valores de la *función objetivo* f y las restricciones Ω . Algunos algoritmos de optimización (AO) acumulan información reunida de iteraciones previas, mientras que otras solo usan información del punto actual.

Un buen AO debe poseer las siguientes propiedades (Nocedal y Wright, 1999):

- Robustez. Deben desempeñarse bien sobre una amplia variedad de problemas en su clase, para todas las selecciones razonables de las variables iniciales.
- Eficiencia. Pudieran no requerir mucho tiempo de cómputo o almacenamiento.
- Exactitud. Pudieran ser capaces de identificar una solución con precisión, sin ser demasiado sensitivo a los errores en el dato o a los errores de redondeo aritmético.

Los algoritmos de optimización puede resolver problemas de minimización o de maximización. En un problema de minimización, el objetivo es encontrar un *minimizador global* de f , es decir, un punto donde la función alcance sus más pequeños valores posibles. Formalmente, la minimización se define como:

Definición 1 (Minimizador global) Un punto x^* es un *minimizador global* si $f(x^*) \leq f(x)$ para toda x .

El vector x de la Definición 1 contiene valores reales en el rango de \mathbb{R}^n . El *minimizador global* puede ser difícil de encontrar, porque nuestro conocimiento de f es solo local. Este es un aspecto importante a considerar, pues al no tener información global de f , es posible que se llegue solo a un *minimizador local* definido formalmente como:

Definición 2 (Minimizador local) Un punto x^* es un *minimizador local* si existe una vecindad N de x^* tal que $f(x^*) \leq f(x)$ para toda $x \in N$.

EL problema de maximización se define cambiando los signos a las definiciones anteriores:

Definición 3 (Maximizador global) Un punto x^* es un *maximizador global* si $f(x^*) \geq f(x)$ para toda x .

Definición 4 (Maximizador local) Un punto x^* es un *maximizador local* si existe una vecindad N de x^* tal que $f(x^*) \geq f(x)$ para toda $x \in N$.

Para evitar los máximos/mínimos locales, en la literatura se han propuesto diferentes estrategias, como por ejemplo, la aleatorización de algunos componentes de la estrategia. En esta tesis se revisaron los algoritmos de optimización Hill Climbing (HC), Simulated Annealing (SA), (1 + 1)-Evolution Strategy (ES) y Gray Wolf Optimizer (GWO). Estos algoritmos se describen a continuación.

7.2.1. Hill Climbing

Los algoritmos Hill Climbing (HC) buscan un máximo global de forma intuitiva como lo hace un escalador para subir desde un valle hasta la cima de una montaña. Para el problema de minimización, el sentido del desplazamiento es simplemente invertido. Este optimizador comienza con una solución arbitraria al problema, luego intenta encontrar una mejor solución variando incrementalmente un único elemento a la solución. Si el cambio produce una solución mejor, entonces otro cambio incremental se le realiza a la nueva solución. Este proceso continúa hasta que no se puedan encontrar mejoras. Este proceso es representado por el Algoritmo 4 para un problema de maximización. La función *Vecinos()* obtiene los vecinos de x_i y evalúa cada uno de ellos mediante la función objetivo f . Si hay un vecino mejor, entonces éste sustituye a x_i .

El algoritmo HC es de búsqueda local, por lo tanto si no se tiene cuidado puede caer en un mínimo/máximo local. Para resolver este problema, se han propuesto diferentes enfoques, por ejemplo, el re-inicio aleatorio (X.-S. Yang, 2014a) o simplemente la aleatorización de algunos componentes (Schwefel, 1995).

7.2.2. Simulated Annealing

El algoritmo Simulated Annealing (SA) es una técnica de búsqueda aleatoria y basado en trayectoria para la optimización global (X.-S. Yang, 2014b). SA imita el proceso de recocido en el procesamiento de materiales cuando un metal se enfría y congela en un estado cristalino con la mínima energía y cristales de tamaño más grande para reducir los defectos en estructuras metálicas. El proceso de recocido involucra el cuidadoso control de la temperatura y su tasa de enfriamiento, comúnmente conocido como *calendario de recocido*.

Algoritmo 4: Algoritmo Hill Climbing.

```

1 Obtener solución arbitraria  $x$ 
2 mientras Mientras no exista un mejor nodo hacer
3    $N = \text{Vecinos}(x_i)$ 
4    $\text{siguienteEval} = -INF$ 
5    $\text{siguienteNodo} = NULL$ 
6   per  $j \leftarrow 1$  a  $|N|$  fai
7      $x_j \leftarrow N(j)$ 
8     si  $f(x_j) > \text{siguienteEval}$  entonces
9        $\text{siguienteNodo} = x_j$ 
10       $\text{siguienteEval} = f(x_j)$ 
11  si  $\text{siguienteEval} \geq f(x)$  entonces
12     $x = \text{siguienteNodo}$ 
13 devolver  $x^* = x$ 

```

El Algoritmo 5 resume los principales pasos del Simulated Annealing. Se inicia con una solución arbitraria x , la cual puede ser usada como argumento para f y se calcula la temperatura inicial Z . El proceso de optimización iterativamente intenta aproximar la temperatura Z a un valor ≤ 0 . Se recomienda iniciar con temperaturas muy altas y luego reducir la temperatura rápidamente hasta un 50 % o aceptar 60 % de los peores movimientos (X.-S. Yang, 2014b). Después de esto, el enfriamiento inicia a hacerse más lento hasta llegar a la solución óptima. Se puede observar que si no hay una mejora, es decir, $\Delta E > 0$, entonces se decide aleatoriamente si se acepta o no la solución x' . Este simple mecanismo evita quedar estancado en los mínimos locales, ya que funciona también como un mecanismo de salida. La disminución de la temperatura puede lograrse mediante la expresión:

$$Z_{k+1} = \alpha Z_k, \quad (7.1)$$

donde $0 < \alpha < 1$ es un factor constante para la disminución de la temperatura.

7.2.3. (1 + 1) Evolution Strategy

El algoritmo de optimización (1+1) *Evolution Strategy* (ES) pertenece a una subclase de métodos de búsqueda directa basado en los principios de la evolución orgánica. El algoritmo ES se basa principalmente en la mutación, recombinación y selección aplicado a una población de soluciones candidatas para iterativamente encontrar la solución óptima x^* . El algoritmo básico consiste de los siguientes pasos (Schwefel, 1995):

Inicialización: Se define $x_E^{(0)} = \{x_{E,i}^{(0)}, i = 1(1)n\}^T$, tal que $G_j(x_E^{(0)}) \geq 0$ para toda $j = 1(1)m$. Establecer $g = 0$.

Algoritmo 5: Algoritmo Simulated Annealing.

```

1 Crear solución inicial  $x$ 
2 Establecer temperatura inicial  $Z$ 
3 Definir  $maxK$ 
4 mientras No termina hacer
5   per  $k = 1 : maxK$  fai
6      $x' =$  perturbar  $x$ 
7      $\Delta E = E(x') - E(x)$ 
8     si  $\Delta E \leq 0$  entonces
9        $x = x'$ 
10    en otro caso
11      si  $rnd(0, 1) < exp(\frac{-\Delta E}{Z})$  entonces
12         $x = x'$ 
13    Disminuir  $Z$ 
14 devolver  $x^* = x$ 

```

Mutación: Construir $x_N^g = x_E^g + z^g$ con componentes de $x_{N,i}^g = x_{E,i}^g + z_i^g$ para toda $i = 1(1)n$.

Slección: Decidir:

$$x_E^{g+1} = \begin{cases} x_N^g & \text{si } f(x_N^g) \leq f(x_E^g) \text{ y } G_j = (x_N^g) \geq 0 \text{ para toda } j = 1(1)m \\ x_E^g & \text{en otro caso.} \end{cases} \quad (7.2)$$

Incrementar $g = g + 1$ e ir a la *mutación* hasta que se alcance el criterio de parada.

En la inicialización se obtiene el punto E^g en el espacio de soluciones que es caracterizado por un vector de posiciones x_E^g . Luego, en la iteración g se obtiene un segundo punto N^g caracterizado por el vector de posiciones x_N^g . Este nuevo vector x_N^g se genera mediante una ligera variación de los elementos de x_E^g . Esta variación se obtiene a través de la adición de valores pseudo-aleatorios z^g . Entonces, de x_E^g y x_N^g se selecciona el mejor individuo que sobrevivirá a la siguiente generación, es decir, x_E^{g+1} . Esta decisión se basa en la calidad de cada individuo, la cual es medida por la función objetivo f . Uno de los aspectos a considerar en este algoritmo, es cómo escoger los valores de z_g la cual juega el papel de mutación. Generalmente, en la literatura se ha propuesto modificar un porcentaje pequeño de los elementos de x^g pero suficiente para esperar una mejora en el nuevo individuo.

7.2.4. Grey Wolf Optimizer

El algoritmo de optimización Grey Wolf Optimizer (GWO) imita el mecanismo de liderazgo jerárquico y cacería de los lobos grises en la naturaleza (Mirjalili, Mirjalili, y

Lewis, 2014). Los pasos básicos de este optimizador son los siguientes (Mirjalili y cols., 2014):

Acorralamiento de la presa. El acorralamiento se modela mediante las siguientes expresiones:

$$\vec{D} = |\vec{C} \cdot \vec{X}_p(t) - \vec{X}(t)|, \quad (7.3)$$

$$\vec{X}(t+1) = \vec{X}_p(t) - \vec{A} \cdot \vec{D}, \quad (7.4)$$

donde t indica la iteración actual, \vec{A} y \vec{C} son vectores de coeficientes, \vec{X}_p es el vector de posición de la presa, y \vec{X} indica el vector de posiciones de un lobo gris.

Los vectores \vec{A} y \vec{C} se calculan de la siguiente manera:

$$\vec{A} = 2\vec{a} \cdot \vec{r}_1 - \vec{a} \quad (7.5)$$

$$\vec{C} = 2 \cdot \vec{r}_2, \quad (7.6)$$

donde los componentes de \vec{a} se disminuyen linealmente de 2 a 0 sobre el curso de las iteraciones y r_1, r_2 son vectores aleatorios en $[0, 1]$.

Cacería. Los lobos grises tiene la habilidad de reconocer la ubicación de la presa y acorralarlo. En un espacio de búsqueda abstracta, no se tiene idea acerca de la ubicación de la presa. Para modelar este comportamiento en la cacería, se asume que se tiene un conocimiento de una ubicación posible de la presa. Por lo tanto, se obtienen los siguientes tres expresiones:

$$\vec{D}_\alpha = |\vec{C}_1 \cdot \vec{X}_\alpha - \vec{X}|, \vec{D}_\beta = |\vec{C}_2 \cdot \vec{X}_\beta - \vec{X}|, \vec{D}_\delta = |\vec{C}_3 \cdot \vec{X}_\delta - \vec{X}|, \quad (7.7)$$

$$\vec{X}_1 = \vec{X}_\alpha - \vec{A}_1 \cdot \vec{D}_\alpha, \vec{X}_2 = \vec{X}_\beta - \vec{A}_2 \cdot \vec{D}_\beta, \vec{X}_3 = \vec{X}_\delta - \vec{A}_3 \cdot \vec{D}_\delta, \quad (7.8)$$

$$\vec{X}(t+1) = \frac{\vec{X}_1 + \vec{X}_2 + \vec{X}_3}{3}. \quad (7.9)$$

Con estas expresiones se estima la posición de la presa y se actualizan las posiciones del resto de los lobos.

Atacar presa (Explotación). La cacería de los lobos termina cuando se ataca la presa que ha dejado de moverse. El acercamiento a la presa se modela matemáticamente disminuyendo los valores de \vec{a} .

Búsqueda de la presa (Exploración). Los lobos deben de realizar la búsqueda de su presa basándose en las posiciones de sus líderes. Esto significa que hay una divergencia con respecto a los líderes, mientras que con respecto a la presa hay una convergencia. La divergencia está dado por \vec{A} con valores aleatorios mayores que 1 o menores que -1. Esta exploración, que se logra con la divergencia, permite la búsqueda global. También,

apoyan a la exploración el vector C que emula los efectos de los obstáculos para el acorralamiento de la presa en la naturaleza.

El acorralamiento, la cacería, el ataque y la exploración se resumen en el Algoritmo 6. Como se puede ver, al finalizar la iteración se devuelve el individuo X_α el cual se considera como la solución óptima x^* .

Algoritmo 6: Pseudo-código para el algoritmo GWO.

```

1 Inicializar la población de lobos grises  $X_i$ 
2 Inicializar  $a$ ,  $A$  y  $C$ 
3 Calcular la calidad de cada agente de búsqueda=
    ■  $X_\alpha$ : EL mejor agente de búsqueda.
    ■  $X_\beta$ : El segundo mejor agente de búsqueda.
    ■  $X_\delta$ : El tercer mejor agente de búsqueda.

mientras  $t < \text{número máximo de iteraciones}$  hacer
    per Para cada agente de búsqueda fai
        Actualizar la posición del actual agente de búsqueda mediante la
        Ecuación 7.9.
        Actualizar  $a$ ,  $A$  y  $C$ 
        Calcular la calidad de todos los agentes de búsqueda
        Actualizar  $X_\alpha$ ,  $X_\beta$  y  $X_\delta$ 
    devolver  $X_\alpha$ 

```

En esta sección se ha descrito brevemente algunos algoritmos de optimización numérica que son usados en este capítulo para la selección automática de conjuntos de entrenamiento. Con estos conjuntos se sintetizan filtros de correlación compuesto para el reconocimiento facial robusto bajo condiciones no restringidos. En este trabajo se seleccionaron los algoritmos HC, SA, ES y GWO para evaluar la metodología propuesta porque son fáciles de implementar, requieren menos recursos computacionales y obtienen buenas respuestas en un tiempo corto.

7.3. Selección automática del conjunto de entrenamiento

En esta sección se propone una metodología basada en optimización numérica para la selección automática del conjunto de entrenamiento para la síntesis de filtros compuestos para el reconocimiento facial robusto.

Sea T_{id} un conjunto dado de imágenes faciales disponibles con distorsiones esperadas para la cara objetivo. Como un subconjunto específico de T_{id} es adecuado para sintetizar un filtro compuesto óptimo \mathbf{h} para reconocimiento facial, el objetivo es seleccionar dicho

subconjunto $T \subseteq T_{id}$. La decisión de cual es el mejor subconjunto puede ser tomado usando un algoritmo de optimización, el cual considera la tripleta (C, f, Ω) , donde C es el conjunto de todas las combinaciones posibles de imágenes en T_{id} (soluciones candidatos), f es la función objetivo que evalúa la calidad de cada solución candidato $T_i \in C$ y Ω es el conjunto de restricciones. La meta es seleccionar una solución óptima T_i^* que satisfaga las restricciones Ω .

Las soluciones candidatos exhiben dos propiedades interesantes. Primero, el espacio de solución C contiene soluciones candidatos de diferentes tamaños, que van desde 1 hasta $|T_{id}|$. Segundo, las funciones objetivos que se proponen en la Subsección 7.3 incrementan su valor cuando se encuentra una solución mejor. Además, una imagen facial $f_i(x, y) \in T_{id}$ puede o no ser usada en una solución candidato, el cual se convierte en un problema de si/no, que puede ser planteado como un problema de decisión binario. Esto permite a una solución candidato ser codificada como un vector binario de posiciones $\mathbf{x}(i) \in [0, 1]$ con $i = 1, 2, \dots, |T_{id}|$. El vector binario mantiene su definición del Capítulo 6, y a continuación se detalla su aplicación al proceso de optimización.

Optimización mediante una representación de vector binario de posiciones

El objetivo de un AO es seleccionar la mejor alternativa de entre un conjunto de soluciones candidatos (Schwefel, 1995). En general, un AO inicia con un *estado inicial* generado usando un proceso estocástico. Este *estado inicial* es entonces usado para iterativamente generar *nuevos estados* mediante la realización de un conjunto de cambios a través de lo que se le conoce como *operador* de los AO. Después de crear *nuevos estados* en cada iteración, una *estrategia de administración de población* es usada para seleccionar el estado o conjunto de estados que sobrevivirán para la siguiente iteración. Normalmente, esta selección se basa en la calidad de cada estado medido por una *función objetivo*. Este proceso continúa hasta obtener el estado óptimo o deseado. A continuación se describe cómo se aplica el vector binario de posiciones a este proceso para la selección del conjunto de entrenamiento.

Para aplicar con éxito un AO al problema de selección de imágenes de entrenamiento, es necesario definir un espacio de búsqueda B , un *operador* O y una *estrategia de administración de población*. El espacio de búsqueda B representa una esquema codificado para las soluciones candidatos en C tal que $\forall T_i \in C, \exists \mathbf{x}_i \in B : m(\mathbf{x}_i) \rightarrow T_i$, donde m es una función de mapeado. El espacio de búsqueda B está definido como el conjunto codificado de todos los subconjuntos que se pueden generar con T_{id} , lo cual implica que $m : B \rightarrow C$. En este trabajo \mathbf{x}_i es binario tal que $m_2(T_{ai}) = \mathbf{x}_i$. Esto significa que la función de mapeado m_2 indexa todas las imágenes faciales $f_i(x, y) \in T_{id}$ en un vector \mathbf{x}_i que pudiera ser poblado dependiendo de un AO particular para generar un *estado inicial*. Dado el *estado inicial* \mathbf{x}_i , el AO pudiera iniciar el proceso de optimización mediante la generación de *nuevos estados* usando el *operador* O de tal forma que $O(\mathbf{x}_i) \rightarrow \mathbf{x}_{i+1}$. Dependiendo del AO seleccionado, O pudiera ser una operación nombrada como mutación para *ES*, perturbación para *SA*, caza para *GWO* y adición de un simple elemento para *HC*. Dados los *nuevos estados* y el *estado inicial*, la *estrategia*

de administración de la población selecciona de ellos el que tenga la mejor calidad para que sobreviva en la siguiente iteración. La calidad se calcula de acuerdo a las siguientes funciones objetivo.

Funciones objetivo para el cálculo de la calidad de las soluciones candidatas

Aquí se proponen las funciones objetivo f para evaluar la calidad de las soluciones candidatas $T_i \in C$. Es importante recordar que una solución candidata se obtiene mediante $m(\mathbf{x}_i) \rightarrow T_i$.

Tasa de Reconocimiento del Filtro (TRF): dado un conjunto de prueba P , la *TRF* cuantifica el número de imágenes faciales $f_i(x, y) \in P$ que son reconocidos por el filtro \mathbf{h} . La *TRF* está dada por:

$$TRF = \frac{|\{f_i(x, y) \in P : s_i \geq \tau\}|}{|P|}, \quad (7.10)$$

donde s_i es el grado de similitud de $f_i(x, y)$ y \mathbf{h} . Un filtro óptimo se obtiene cuando la *TRF* es igual a 1.

Capacidad de Discriminación del Filtro (CDF): Dado un conjunto de salidas de correlación, *CDF* mide la variación promedio de esas superficies usando *ASM* y la altura promedio de sus picos de correlación usando *ACH*. Para una solución candidata $T_i \in C$ y su correspondiente filtro \mathbf{h} , la *CDF* está dado por:

$$CDF = 1 - \frac{ASM^2}{ACH^2}. \quad (7.11)$$

Como se puede notar, *CDF* tiende a 1 con *ASM* bajo y *ACH* alto.

Agudeza Promedio de Picos (APP): *APP* mide la agudeza promedio de los picos:

$$APP = \frac{1}{|T_{id}|} \sum_{i=1}^{|T_{id}|} s_i, \quad (7.12)$$

donde s_i es obtenido mediante la correlación de $f_i(x, y) \in T_{id}$ y \mathbf{h}_s . El filtro \mathbf{h}_s es sintetizado de un conjunto T_s que contiene caras generadas de la siguiente manera. Sea $R_i = \{r_1(x, y), r_2(x, y), r_3(x, y), r_4(x, y), r_5(x, y)\}$ el conjunto de regiones faciales de $f_i(x, y) \in T_i$, de esta forma, una nueva imagen facial se obtiene mediante:

$$f'_i(x, y) = r_1(x, y) \cup r_2(x, y) \cup r_3(x, y) \cup r_4(x, y) \cup r_5(x, y), \quad (7.13)$$

donde $r_i(x, y)$ es seleccionado aleatoriamente. Las nuevas caras se obtienen con el fin de incorporar a los filtros una mayor capacidad de discriminación. Además, calcular el desempeño del filtro sobre T_{id} permite conocer el desempeño con caras no conocidas pero de la misma clase que T_i .

Note que la función objetivo *TRF* requiere un conjunto de prueba P , mientras que *CDF* y *APP* requieren respectivamente T_i y T_s . Por lo tanto, *TRF* es apropiado cuando se dispone de muchas imágenes faciales por persona.

7.3.1. Metodología basada en optimización para seleccionar conjuntos de entrenamiento

La Figura 7.1 muestra la metodología propuesta para la selección del conjunto de entrenamiento T . Dado un conjunto de imágenes faciales disponibles T_{id} y alguna de las funciones objetivo f propuestos previamente, el conjunto de entrenamiento es seleccionado de la siguiente manera. El primer paso consiste en generar una población inicial de soluciones candidatas obtenidas de T_{id} . Una solución candidato es generada mediante $m_2(T_{id}) \rightarrow \mathbf{x}_i$ y poblada con valores reales pseudo-aleatorios. Luego, los valores en \mathbf{x}_i son redondeados para obtener el vector binario de posiciones. El segundo paso consiste en sintetizar un filtro compuesto con $T_i \leftarrow m(\mathbf{x}_i)$. En el tercer paso, se calcula la calidad de las soluciones candidato mediante la *función objetivo* f . El cuarto paso consiste de un proceso iterativo en el cual el *operador* O se aplica a los padres para generar descendientes, es decir, $O(\mathbf{x}_i) = \mathbf{x}_{i+1}$. Es deseable usar valores aleatorios en O para evitar máximos locales. Los descendientes \mathbf{x}_{i+1} son usados para la síntesis de nuevos filtros de correlación compuesto. Basado en el nivel de calidad individual, la *estrategia de administración de población* selecciona el mejor estado de entre el estado inicial y los descendientes para que sobreviva a la siguiente iteración. Finalmente, se devuelve una solución óptima T_i^* cuando se alcance un criterio de parada. Analizar C en su totalidad es muy caro debido a la gran cantidad de subconjuntos a explorar, pues $|C| = |P(T_{id})| - 1 = 2^{|T_{id}|} - 1$. Por esta razón, se sugiere que el criterio de parada sea un número de iteraciones que provean un desempeño deseable.

Para usar la metodología propuesta, es necesario escoger un algoritmo de optimización e a): identificar el *operador* y la *estrategia de administración de la población* para aplicarlo en la estrategia de selección del conjunto de entrenamiento, b) escoger alguna *función objetivo* propuesto para evaluar la calidad de cada *solución candidato* y c) establecer un criterio de parada.

7.4. Resultados experimentales y discusión

En esta sección se presenta y discute el desempeño en el reconocimiento facial de la metodología propuesta. Para propósitos de evaluación, los algoritmos de optimización ES, HC, SA y GWO se usaron en la selección de imágenes de entrenamiento para la síntesis de los filtros UNCF, MACH y UOTSDF. Se evaluó el impacto de la selección de los conjuntos de entrenamiento sobre el desempeño de los filtros compuestos en el reconocimiento facial. Los parámetros determinados experimentalmente para la metodología propuesta son los siguientes: 20 iteraciones cuando los algoritmos de optimización usan TRF o APP, y 10 cuando se usa CDF; una población inicial de un individuo para ES, HC y SA, y 30 individuos para GWO; $\frac{1}{5}$ de los elementos en \mathbf{x}_i son modificados por el operador O ; y un umbral de reconocimiento $\tau = 10$.

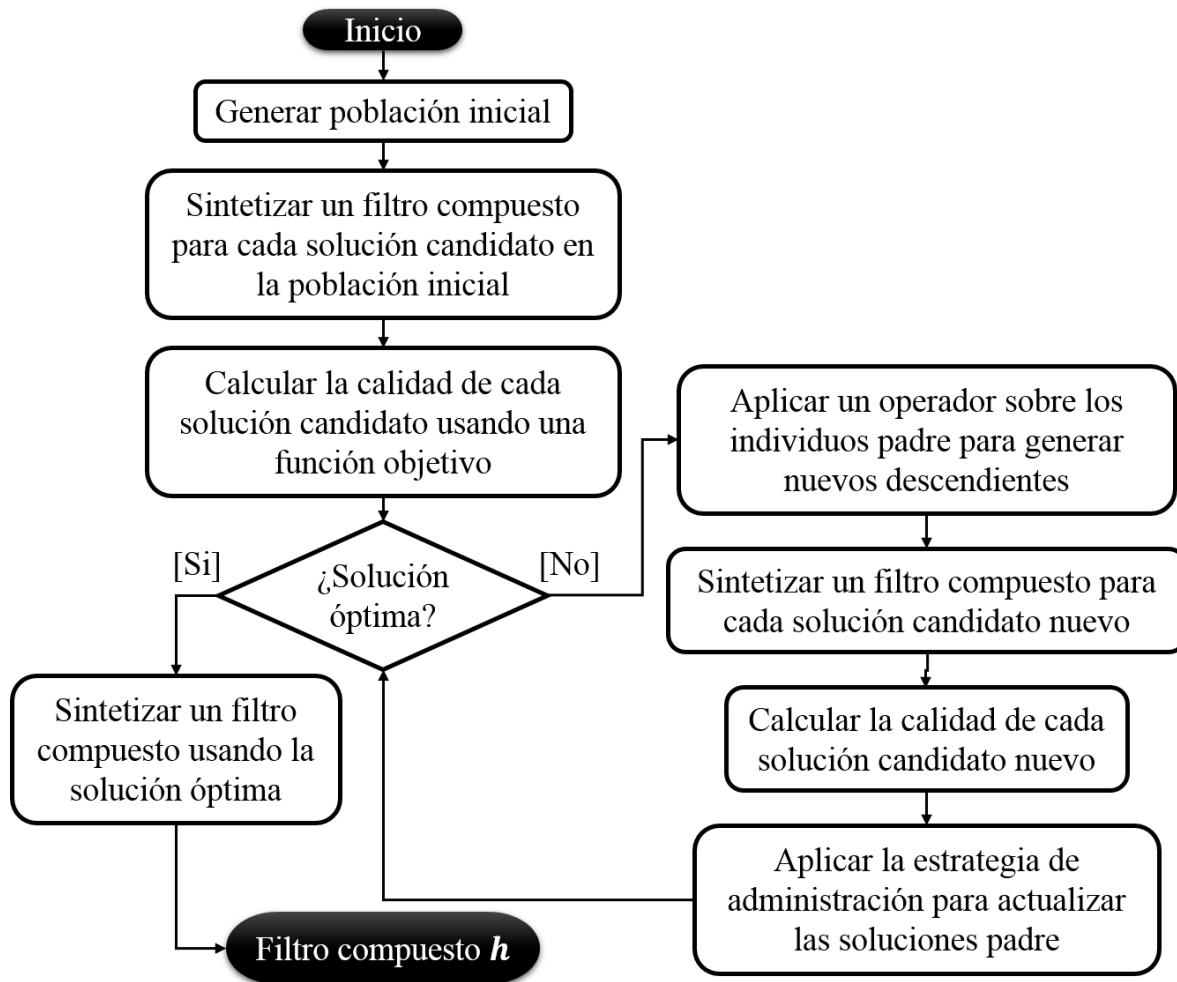


Figura 7.1: Diagrama de bloque de la metodología propuesta para la selección de conjuntos de entrenamiento para filtros compuestos.

7.4.1. Configuración de los datos

Las bases de datos faciales CMU (Lab, 2001) y Yale B (Georghiades, Belhumeur, y Kriegman, 2001b) fueron usadas para evaluar la metodología propuesta. CMU contiene 13 personas, las cuales cada una contiene 75 imágenes en escala de grises mostrando diferentes expresiones. Esas imágenes tienen una resolución de 64×64 píxeles y fueron capturadas bajo una misma condición de iluminación. Yale B contiene 38 personas, cada una con nueve poses tomadas bajo 64 condiciones de iluminación variable. Cada persona contiene entre 60 y 64 imágenes en escalas de grises con una resolución de 192×168 píxeles.

Sea A_i la i -ésima clase facial en alguna de las bases de datos. Cada conjunto A_i fue dividido en dos subconjuntos $T_{id,i} \cup P_i = A_i$. Primeramente, 20 imágenes fueron seleccionadas aleatoriamente de A_i para $T_{id,i}$. Se asume que esas imágenes son conocidas,

por lo tanto pueden ser exploradas por la metodología propuesta para la selección del sub-conjunto óptimo T_i^* . Segundo, el resto de las imágenes fueron puestas dentro de un conjunto de prueba P_i . Se asume que las imágenes de este conjunto son desconocidas, por lo que pueden ser usadas para probar los filtros compuestos sintetizados con el conjunto T_i^* . Con esta división, el conjunto P_i para cada persona contiene 55 imágenes faciales para CMU y 40-44 imágenes faciales para Yale B. En las Figuras 7.2 y 7.3 se presentan ejemplos de los conjuntos $T_{id,i}$ y P_i para CMU y Yale B, respectivamente.



Figura 7.2: Muestra del conjunto $T_{id,i}$ (izquierda) y P_i (derecha) para la i -ésima persona en CMU.



Figura 7.3: Muestra del conjunto $T_{id,i}$ (izquierda) y P_i (derecha) para la i -ésima persona en Yale B.

Con el fin de mejorar el desempeño de los filtros, cada imagen fue procesada de la siguiente manera. Primero, el filtrado de retina (Vu y Caplier, 2009) fue aplicado a las imágenes para mejorar la iluminación global. Segundo, a la imagen resultante se le aplicó la diferencia de Gaussianas (Davidson y Abramowitz, 2014; Vu y Caplier, 2009) para remover las variaciones en iluminación y mejorar el borde de las características faciales. Finalmente, la imagen resultante fue transformado a una imagen LBPi (Ojala,

Pietikainen, y Maenpaa, 2002; Huang, Z.-Li, y Wang, 2004; Ao, Yi, Lei, y Li, 2009). Las imágenes LBPI tienen buenas propiedades para el reconocimiento facial, tales como: robustez a distorsiones locales, gran capacidad de discriminación y son capaces de preservar tanto la apariencia intrínseca como las relaciones espaciales (B. Yang y Chen, 2013). La Figura 7.4 muestra ejemplos de las imágenes resultantes de estas operaciones.



Figura 7.4: Imágenes resultantes del preprocesamiento. De izquierda a derecha: imagen original, filtrado de retina, diferencias de Gaussianas e imagen LBPI.

7.4.2. Resultados numéricos

Las Tablas [7.1, 7.2, 7.3, 7.4] muestran los resultados estadísticos calculados con las métricas FERET. Para cada base de datos faciales, todos los conjuntos P_i fueron puestos en un conjunto P_G , resultando 715 para CMU y 1654 para Yale B. De acuerdo a las Tablas 7.1 y 7.2, la metodología propuesta es capaz de seleccionar conjuntos de entrenamientos para la síntesis de filtros compuestos con resultados muy competitivos en términos de las métricas TID y TV . El filtro MACH obtuvo el mejor desempeño para imágenes con iluminación homogénea y expresiones faciales contenidas en el conjunto CMU, mientras que el filtro UNCF fue mejor en imágenes con iluminación variable contenidas en el conjunto Yale B. Las sombras locales en las imágenes faciales no permiten una adecuada recuperación de la información de borde. Por esta razón, en la mayoría de los casos, el desempeño obtenido para imágenes con iluminación homogénea es mejor que el obtenido por las imágenes con iluminación variable. El filtro UNCF sintetizado en (Santiago-Ramírez y cols., 2012) con imágenes seleccionados manualmente de Yale B, obtuvo una $TID = 95.60\%$ y una $TV = 95.72\%$. Esos resultados son superados por el mismos filtros sintetizado con conjuntos de entrenamiento seleccionados por el método propuesto usando para la selección cualquiera de los AO's evaluados y TRF como función objetivo. El filtro MOSDF (Diaz-Ramirez y cols., 2014) obtuvo una $TR = 98.4\%$ (contra el mejor resultado obtenido en este trabajo de 97.16%) en una evaluación parcial con solo 10 de los 38 personas en Yale B. Esos filtros MOSDF fueron sintetizados con 44 imágenes de la clase verdadera y 268 de la clase falsa, mientras que la metodología propuesta sintetizó los filtros con un promedio de 12 imágenes. En (Qiao, Chen, y Tan, 2010) el reconocimiento facial se realiza usando un método de reducción de dimensionalidad no supervisado logrando una $TR = 95.18\%$, la cual es superada

por la metodología propuesta usando para la selección cualquier AO con la *TRF*. Para el conjunto CMU, el filtro MACH sintetizado con conjuntos de entrenamientos seleccionados usando *ES* y *TRF* superó la $TR = 99.16\%$ obtenido en (Pinto-Fernández y Díaz-Ramírez, 2012) usando el mismo filtros compuesto.

Tabla 7.1: Desempeño de los filtros evaluados en términos de la Tasa de Identificación y Detección (TID) (%) con el mejor desempeño para cada función objetivo en negrita.

| | | ES | | | HC | | | SA | | | GWO | | |
|--------|--------|--------------|--------------|--------------|--------------|--------------|--------------|---------------|--------------|--------------|--------------|--------------|--------------|
| | | TRF | CDF | APP | TRF | CDF | APP | TRF | CDF | APP | TRF | CDF | APP |
| UNCF | CMU | 99.30 | 95.66 | 97.34 | 97.76 | 86.29 | 94.41 | 98.74 | 96.50 | 97.48 | 99.44 | 86.01 | 98.18 |
| | Yale B | 97.16 | 88.24 | 92.49 | 96.22 | 70.27 | 77.36 | 96.69 | 93.03 | 86.82 | 96.99 | 77.72 | 93.50 |
| MACH | CMU | 99.72 | 99.72 | 99.02 | 99.72 | 99.72 | 98.18 | 99.58 | 99.44 | 95.66 | 99.72 | 99.58 | 99.58 |
| | Yale B | 95.21 | 94.86 | 78.78 | 96.16 | 93.50 | 90.00 | 96.51 | 91.49 | 79.73 | 95.74 | 96.16 | 94.44 |
| UOTSDF | CMU | 99.72 | 98.88 | 98.88 | 99.44 | 80.56 | 98.18 | 100.00 | 97.48 | 97.20 | 99.86 | 94.97 | 99.86 |
| | Yale B | 96.28 | 88.12 | 82.86 | 95.27 | 70.00 | 75.32 | 96.63 | 90.13 | 82.80 | 96.69 | 89.24 | 93.26 |

Tabla 7.2: Desempeño de los filtros evaluados en términos de la Tasa de Verificación (TV) (%) con el mejor desempeño para cada función objetivo en negrita.

| | | ES | | | HC | | | SA | | | GWO | | |
|--------|--------|--------------|--------------|--------------|--------------|--------------|--------------|---------------|--------------|--------------|--------------|--------------|--------------|
| | | TRF | CDF | APP | TRF | CDF | APP | TRF | CDF | APP | TRF | CDF | APP |
| UNCF | CMU | 99.30 | 95.66 | 97.34 | 97.76 | 86.29 | 94.41 | 98.74 | 96.50 | 97.48 | 99.44 | 86.01 | 98.18 |
| | Yale B | 97.16 | 88.24 | 92.55 | 96.22 | 70.27 | 77.36 | 96.69 | 93.09 | 86.82 | 96.99 | 77.72 | 93.50 |
| MACH | CMU | 99.72 | 99.72 | 99.02 | 99.72 | 99.72 | 98.46 | 99.58 | 99.44 | 95.66 | 99.86 | 99.72 | 99.72 |
| | Yale B | 95.21 | 94.86 | 78.84 | 96.22 | 93.50 | 90.00 | 96.51 | 91.49 | 79.73 | 95.74 | 96.16 | 94.44 |
| UOTSDF | CMU | 99.72 | 98.88 | 98.88 | 99.44 | 80.56 | 98.32 | 100.00 | 97.48 | 97.20 | 99.86 | 95.10 | 99.86 |
| | Yale B | 96.28 | 88.12 | 82.86 | 95.27 | 70.00 | 75.39 | 96.63 | 90.13 | 82.80 | 96.69 | 89.24 | 93.26 |

Las Tablas 7.3 y 7.4 contienen los resultados en términos de las métricas *TFA* y *TFAc*. Para la métrica *TFA*, P_Q es el conjunto de todas las imágenes en P_i de CMU cuando se está evaluando Yale B, y viceversa. A pesar del buen desempeño del filtro MACH en términos de las métricas *TID* y *TV*, en términos de las métricas *TFA* y *TFAc* no fue buena en comparación con los otros filtros evaluados. El filtro UNCF en (Santiago-Ramírez y cols., 2012) obtuvo una $TFA = 1.25\%$ y una $TFAc = 0.25\%$ para Yale B. Esos resultados son superados por el obtenido por el filtro UNCF sintetizado por la metodología propuesta, incluso por los peores valores, los cuales son $TFA = 0.10\%$ y $TFAc = 0.07\%$.

En esta investigación se buscó estudiar el efecto del algoritmo de optimización (AO), función objetivo (FO) y el filtro compuesto (FC) sobre el desempeño de la metodología propuesta. Para este objetivo, se realizó un diseño factorial (Montgomery, 2013) de $4 AO's \times 3 FO's \times 3 FC's$ con $n = 2$ réplicas. Una réplica es un ensayo completo del experimento donde se investigan todos los posibles combinaciones de los niveles de los factores. En este trabajo, una prueba fue sobre CMU y el otro sobre Yale B. Los resultados numéricos para el diseño factorial, generalmente se resumen en una tabla de Análisis de Varianza (ANOVA, por siglas en inglés).

Sea cada una de las Tablas [7.1, 7.2, 7.3, 7.4] una *batería de diseño* obtenido en $n \times (4 \times 3 \times 3) = 72$ ejecuciones experimentales para cada métrica FERET. En las Tablas

Tabla 7.3: Desempeño de los filtros evaluados en términos de la Tasa de Falsa Alarma (TFA) (%) con el peor desempeño para cada función objetivo en negrita.

| | | ES | | | HC | | | SA | | | GWO | | |
|--------|--------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|
| | | TRF | CDF | APP | TRF | CDF | APP | TRF | CDF | APP | TRF | CDF | APP |
| UNCF | CMU | 0.00 | 0.00 | 0.00 | 0.08 | 0.00 | 0.04 | 0.08 | 0.00 | 0.00 | 0.08 | 0.00 | 0.04 |
| | Yale B | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.10 |
| MACH | CMU | 1.20 | 1.08 | 0.79 | 0.62 | 0.70 | 0.79 | 0.95 | 0.99 | 0.08 | 1.53 | 1.37 | 1.74 |
| | Yale B | 0.82 | 1.03 | 0.21 | 0.82 | 1.33 | 0.00 | 1.64 | 0.72 | 0.00 | 1.33 | 2.67 | 1.03 |
| UOTSDF | CMU | 1.02 | 0.33 | 0.75 | 0.61 | 0.00 | 0.33 | 1.06 | 0.00 | 0.04 | 1.43 | 0.29 | 0.00 |
| | Yale B | 0.51 | 0.41 | 0.41 | 0.62 | 0.00 | 0.00 | 0.31 | 0.62 | 0.00 | 0.41 | 0.10 | 0.82 |

Tabla 7.4: Desempeño de los filtros evaluados en términos de la Tasa de Falsa Aceptación (TFAc) (%) con el peor desempeño para cada función objetivo en negrita.

| | | ES | | | HC | | | SA | | | GWO | | |
|--------|--------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|
| | | TRF | CDF | APP | TRF | CDF | APP | TRF | CDF | APP | TRF | CDF | APP |
| UNCF | CMU | 0.26 | 0.17 | 0.19 | 0.06 | 0.03 | 0.09 | 0.33 | 0.19 | 0.11 | 0.36 | 0.01 | 0.29 |
| | Yale B | 0.06 | 0.05 | 0.07 | 0.04 | 0.01 | 0.02 | 0.04 | 0.03 | 0.02 | 0.04 | 0.01 | 0.06 |
| MACH | CMU | 1.43 | 1.33 | 1.06 | 0.58 | 1.07 | 0.99 | 1.43 | 1.01 | 0.11 | 1.48 | 1.16 | 1.96 |
| | Yale B | 0.11 | 0.13 | 0.06 | 0.10 | 0.10 | 0.02 | 0.07 | 0.09 | 0.01 | 0.12 | 0.22 | 0.15 |
| UOTSDF | CMU | 1.31 | 0.45 | 0.95 | 0.36 | 0.03 | 0.65 | 0.93 | 0.14 | 0.12 | 1.19 | 0.29 | 1.26 |
| | Yale B | 0.07 | 0.05 | 0.03 | 0.07 | 0.00 | 0.01 | 0.08 | 0.08 | 0.02 | 0.08 | 0.05 | 0.14 |

[7.5, 7.6, 7.7, 7.8] se presenta el ANOVA para las métrica TID , TV , TFA y $TFAc$, respectivamente. En dichas tablas, las variable FV, GL, SC y CM son respectivamente la fuente de variabilidad, grados de libertad, suma cuadrada y cuadro medio. Un valor $p < \alpha = 0.05$ en la ANOVA indica que el factor es estadísticamente significativo. El análisis de los efectos estadísticamente significativos revela qué factores son importantes en el uso de la metodología propuesta.

Solo los efectos de FO son estadísticamente significativos en las métricas TID y TV , y como se puede ver en las gráfica de efectos principales ubicado en la primera fila de la Figura 7.5, el mejor desempeño se obtiene usando la TRF . Este resultado confirma los resultados obtenidos en el protocolo FERET, donde la misma función objetivo provee los mejores desempeños. Para la métrica TFA ; AO, FO, FC, AO \times FC, y FO \times FC tiene los efectos más significativos. El filtro UNCF sintetizado con conjuntos de entrenamiento seleccionados por la metodología propuesta usando HC y la APP (ver la gráfica de efectos principales en la esquina inferior izquierdo de la Figura 7.5) es el mejor para rechazar impostores verdaderos. En el caso de la métrica TFAc, solo el efecto de FC es estadísticamente significativo. La gráfica de la esquina inferior derecha de la Figura 7.5 muestra que el filtro UNCF obtuvo el mejor desempeño para rechazar impostores falsos.

Tabla 7.5: ANOVA para la métrica TID.

| FV | GL | SC | CM | F_0 | <i>valorp</i> |
|----------|----|---------|--------|-------|---------------|
| AO | 3 | 305.00 | 101.65 | 2.28 | 0.096 |
| FO | 2 | 714.90 | 357.45 | 8.03 | 0.001 |
| FC | 2 | 156.00 | 77.98 | 1.75 | 0.188 |
| AO×FO | 6 | 389.40 | 64.90 | 1.46 | 0.221 |
| AO×FC | 6 | 305.50 | 50.52 | 1.14 | 0.358 |
| FO×FC | 4 | 308.50 | 77.12 | 1.73 | 0.164 |
| AO×FO×FC | 12 | 201.60 | 16.80 | 0.38 | 0.963 |
| Error | 36 | 1603.30 | 44.54 | | |
| Total | 71 | 3984.10 | | | |

Tabla 7.6: ANOVA para la métrica TV

| FV | GL | SC | CM | F_0 | <i>p-value</i> |
|----------|----|---------|--------|-------|----------------|
| AO | 3 | 303.30 | 101.10 | 2.26 | 0.098 |
| FO | 2 | 712.30 | 356.16 | 7.97 | 0.010 |
| FC | 2 | 158.50 | 79.24 | 1.77 | 0.184 |
| AO×FO | 6 | 390.10 | 65.02 | 1.46 | 0.221 |
| AO×FC | 6 | 308.20 | 51.37 | 1.15 | 0.354 |
| AO×FC | 4 | 306.60 | 76.65 | 1.72 | 0.168 |
| AO×FO×FC | 12 | 202.90 | 16.91 | 0.38 | 0.963 |
| Error | 36 | 1607.80 | 44.66 | | |
| Total | 71 | 3989.70 | | | |

Tabla 7.7: ANOVA para la métrica TFA

| FV | GL | SC | CM | F_0 | <i>valorp</i> |
|----------|----|---------|---------|-------|---------------|
| AO | 3 | 1.6839 | 0.56129 | 5.36 | 0.004 |
| FO | 2 | 1.3235 | 0.66176 | 6.31 | 0.004 |
| FC | 2 | 11.1361 | 5.56805 | 53.12 | 0 |
| AO×FO | 6 | 0.4315 | 0.07192 | 0.69 | 0.662 |
| AO×FC | 6 | 1.9987 | 0.33311 | 3.18 | 0.013 |
| FO×FC | 4 | 1.9316 | 0.48289 | 4.61 | 0.004 |
| AO×FO×FC | 12 | 0.7718 | 0.06431 | 0.61 | 0.816 |
| Error | 36 | 3.7734 | 0.10482 | | |
| Total | 71 | 23.0504 | | | |

Tabla 7.8: ANOVA para la métrica TFAc.

| FV | GL | SC | CM | F_0 | <i>valorp</i> |
|----------|----|---------|---------|-------|---------------|
| AO | 3 | 0.8467 | 0.28223 | 0.94 | 0.433 |
| OF | 2 | 0.3188 | 0.15938 | 0.53 | 0.593 |
| FC | 2 | 3.1289 | 1.56444 | 5.2 | 0.01 |
| AO×FO | 6 | 0.7395 | 0.12326 | 0.41 | 0.868 |
| AO×FC | 6 | 0.299 | 0.04983 | 0.17 | 0.984 |
| FO×FC | 4 | 0.3662 | 0.09155 | 0.3 | 0.873 |
| AO×FO×FC | 12 | 0.2546 | 0.02121 | 0.07 | 1 |
| Error | 36 | 10.8382 | 0.30106 | | |
| Total | 71 | 16.7918 | | | |

7.4.3. Discusión acerca de la importancia del conjunto de entrenamiento

Como se ha mencionado previamente, se requiere que T sea pequeño por conveniencia computacional, pero que contenga todas las distorsiones que se espera presente la cara objetivo. Por lo tanto, la calidad y cantidad de las imágenes de entrenamiento influyen directamente en el desempeño del filtro de correlación. La Figura 7.6 muestra la gráfica de intervalos del tamaño promedio del conjunto seleccionado por el método propuesto, mostrando que los filtros son sintetizados con un promedio de 60% de imágenes en T_{id} . Como se puede observar, existe más desviación con el conjunto CMU debido al hecho de que las expresiones cambian la forma de las características faciales. Esto dificulta un acoplamiento perfecto, y en consecuencia, se requieran más imágenes de entrenamiento para poder reconocer todas las imágenes en $T_{id,i}$. Comparado con los conjuntos de entrenamiento usados en la literatura, se puede resaltar lo siguiente: (Diaz-Ramirez y cols., 2014) requiere más imágenes de la clase verdadera e imágenes de la clase falsa, (Qiao y cols., 2010) requiere más imágenes y obtiene un desempeño menor, mientras que (Santiago-Ramírez y cols., 2012) requiere de menos imágenes pero su desempeño también es superado por la metodología propuesta con todas las combinaciones y la *función objetivo* TFR.

Para mostrar la importancia de la cantidad y la calidad de las imágenes de entrenamiento, los filtros MACH, UOTSDF y UNCF fueron sintetizados con conjuntos de

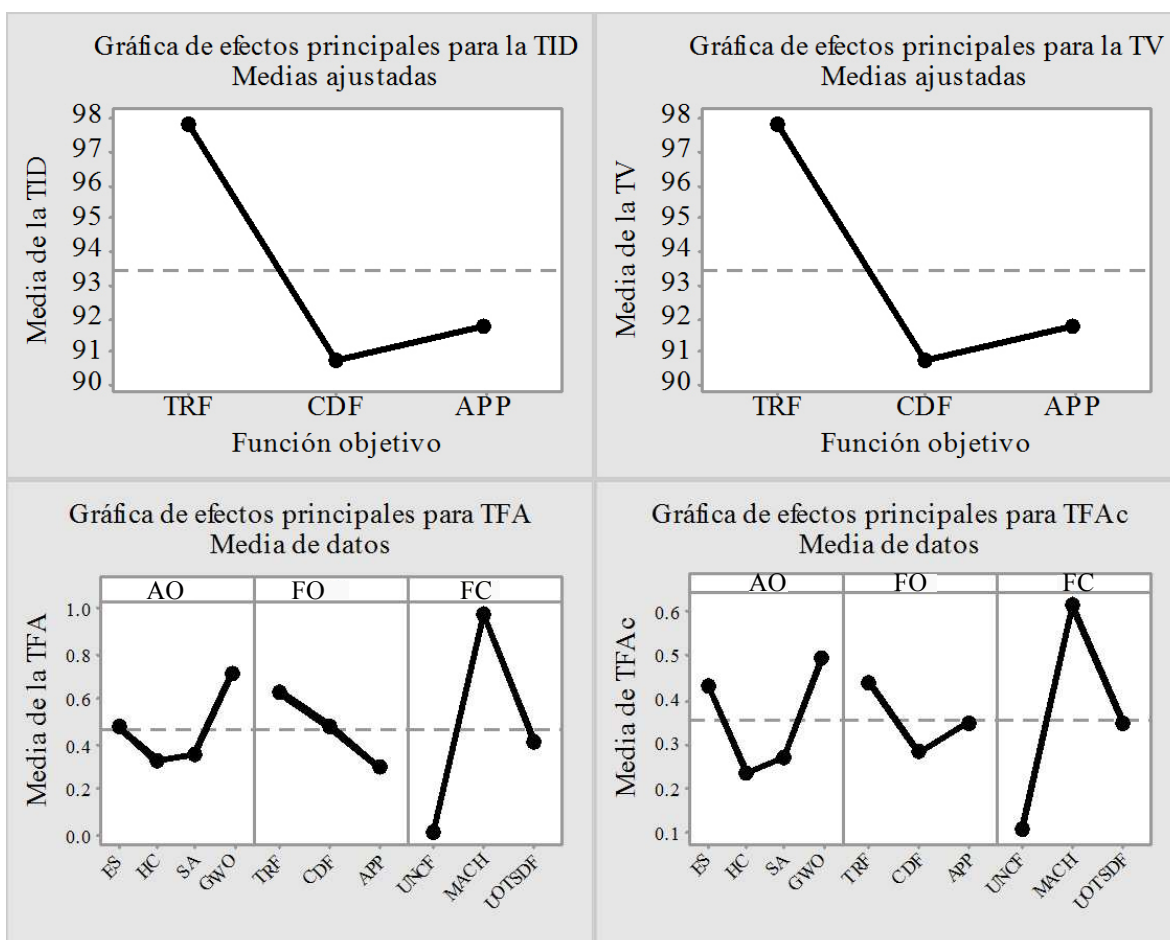


Figura 7.5: Gráficas de efectos principales para las métricas FERET con medias ajustadas.

entrenamiento seleccionados de tres formas diferentes: a) método convencional usado en el experimento previo, b) selección aleatoria y c) usando la metodología propuesta. La la Figura 7.7 muestra los resultados obtenidos por conjuntos seleccionados convencional y aleatoriamente. El método convencional reconoce las 63 imágenes usando un conjunto de entrenamiento más pequeño que el seleccionado de forma aleatoria. Porque en cada paso del método convencional se selecciona la imagen con el peor desempeño, existe un mejora incremental, aunque no siempre se obtiene este desempeño debido a que no existe garantía de que siempre se seleccionen las imágenes más representativas. Por otra parte, existe una variabilidad en el desempeño con la selección aleatoria debido al hecho de que no considera la contribución individual de las imágenes en el diseño del filtro. El problema asociado con usar esos enfoques de selección de imágenes de entrenamiento es abordado por la metodología propuesta. Los resultados en la Figura 7.8 muestra que nuestra propuesta selecciona más imágenes que el método convencional y menos que el método aleatorio para reconocer todas las imágenes faciales de una misma clase.

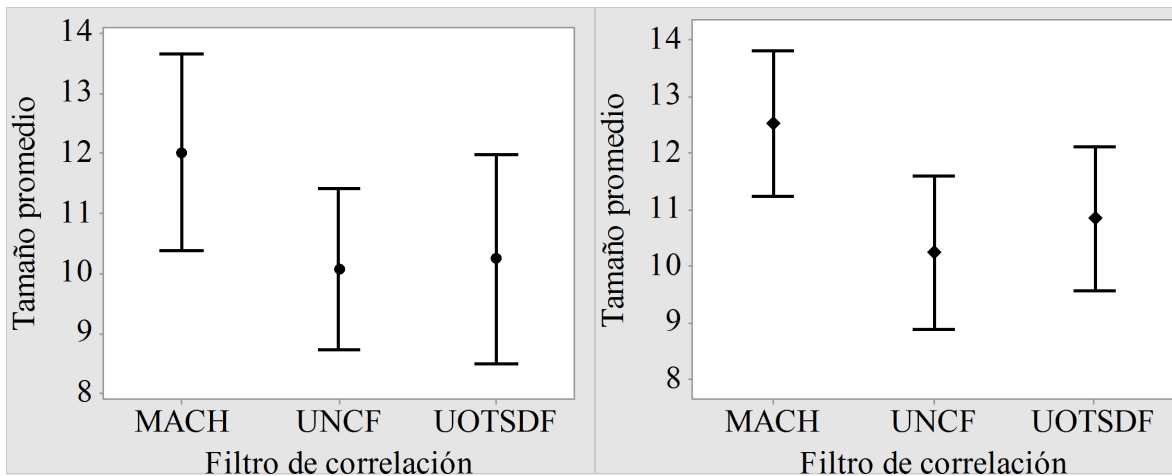


Figura 7.6: Gráficas de intervalos del tamaño promedio del conjunto de entrenamiento seleccionado por la metodología propuesta para CMU (izquierda) y Yale B (derecha).

7.4.4. Comparación de la selección convencional contra la selección hecha por el método propuesto

Aquí se compara el desempeño de los conjuntos de entrenamiento seleccionados por el método convencional en (Vijaya-Kumar, 1992; Rodriguez-Perez, 2012) contra aquellos seleccionados por el método propuesto usando GWO y TRF. Sea T_i^c y T_i^p los conjuntos de entrenamientos seleccionados respectivamente por el método convencional y el método propuesto. Con estos conjuntos se implementó el filtro ASPOF para reconocer personas en CMU y Yale B. Los resultados obtenidos se muestran en la Tabla 7.9. El método propuesto es mejor en las métricas TID y TV, mientras que desempeña igual que el método convencional en la métrica TFA. Solo en la métrica TFAc el método propuesto obtuvo el peor desempeño, aunque muy bajo, pues es solo el 0.02%. El buen desempeño obtenido por el método propuesto se debe a que analiza varias combinaciones de imágenes en T_{id} antes de determinar la solución. En cambio, el método convencional analiza solo una combinación posible.

Tabla 7.9: Desempeño de conjuntos de entrenamiento seleccionados convencionalmente y mediante el método propuesto usando GWO y TRF.

| | Selección convencional | | Método propuesto | |
|------|------------------------|--------|------------------|--------|
| | CMU | Yale B | CMU | Yale B |
| TID | 87.13 | 66.94 | 99.16 | 99.38 |
| TV | 87.13 | 66.94 | 99.16 | 94.50 |
| TFA | 0.00 | 0.00 | 0.00 | 0.00 |
| TFAc | 0.00 | 0.01 | 0.00 | 0.02 |

En las Figuras 7.9 y 7.10 se muestra el tamaño de los conjuntos de entrenamiento y el desempeño de sus respectivos filtros para los conjuntos CMU y Yale B, respectivamente.

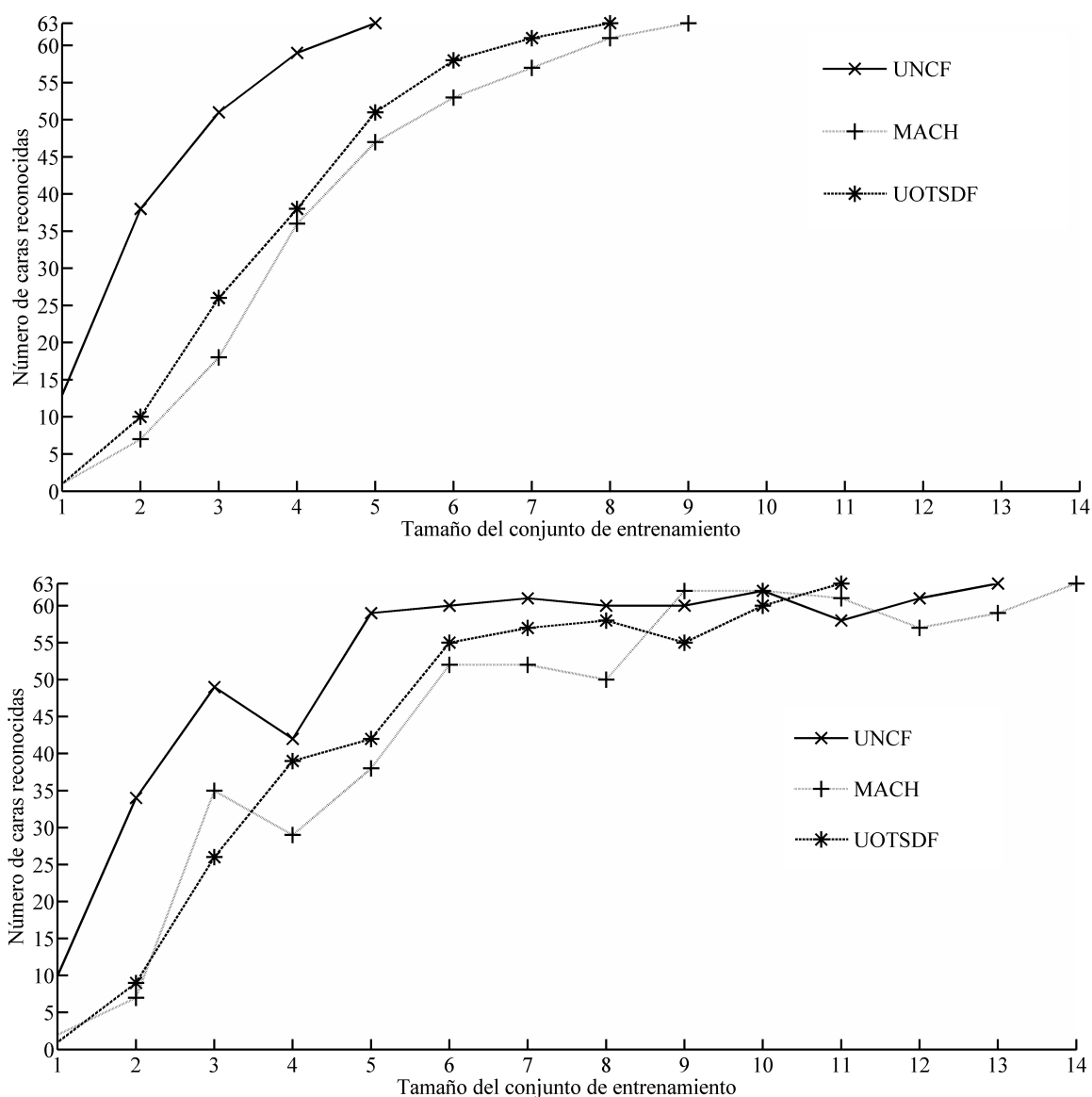


Figura 7.7: Desempeño de filtros compuestos sintetizados con conjuntos de entrenamiento de diferentes tamaños. Arriba para conjuntos seleccionados convencionalmente y abajo para conjuntos seleccionados aleatoriamente.

En ambas bases de datos, el análisis extensivo realizado por el método propuesto causa que el conjunto T_i^p sea más grande que T_i^c . Porque T_i^p contiene más y mejores imágenes, es lógico que el filtro producido obtenga un mejor desempeño TRF tal y como se muestra en ambas figuras.

En esta sección se ha presentados un análisis detallado de los resultados obtenidos por el método propuesto. Con base a esos resultados, el tratamiento dado por GWO, TRF y MACH es el mejor para identificar y verificar imágenes faciales bajo iluminación

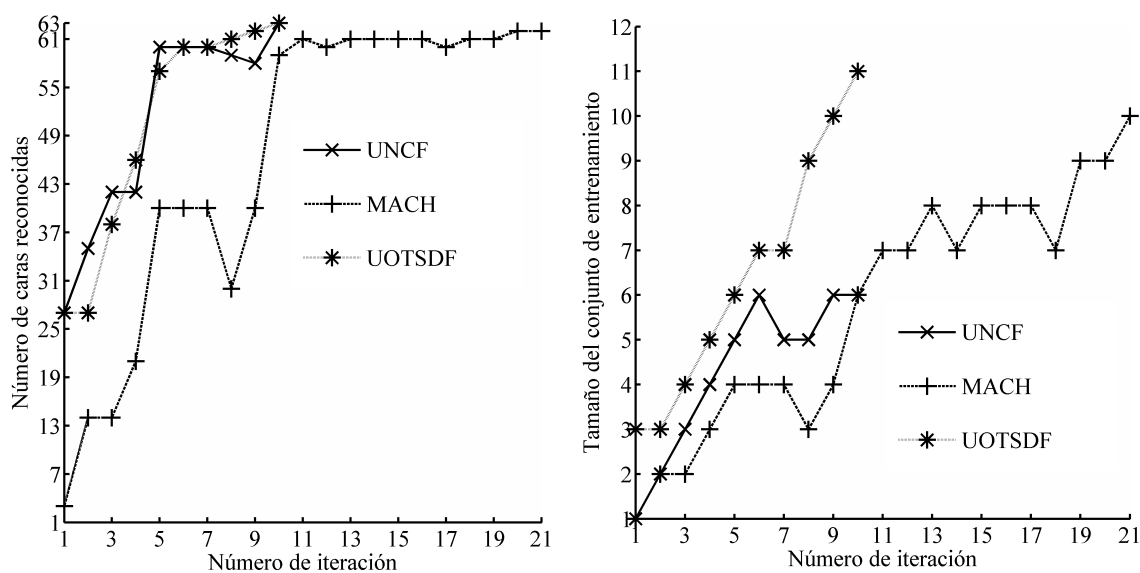


Figura 7.8: Desempeño de filtros compuestos (izquierda) y tamaño del conjunto de entrenamiento en cada iteración de la metodología propuesta (derecha).

homogénea y expresiones faciales. Sin embargo, si lo que importa es rechazar impostores y reconocer eficientemente imágenes con iluminación variable, entonces el tratamiento dado por UNCF, GWO o ES y APP es el recomendado pues mostró el mejor desempeño bajo estas condiciones. Además, los resultados obtenidos muestran que la metodología propuesta selecciona conjuntos de entrenamiento tal y como lo haría un experto pero con un desempeño optimizado comparado con aquellos reportados en la literatura de reconocimiento facial.

Los algoritmos de optimización HC, SA, ES y GWO fueron seleccionados debido a que son fáciles de implementar, requieren menos recursos computacionales y obtienen buenas respuestas en corto tiempo. Aunque es posible que el método propuesto pueda usar algoritmos de optimización similares a los evaluados en este trabajo. Por ejemplo, Particle Swarm Optimization (PSO) (Kennedy y Eberhart, 1995) y Ant Colony Optimization (ACO) (Dorigo y Stützle, 2004). No existe garantía de que el método propuesto obtenga la solución óptima, sin embargo, los resultados numéricos indican que los algoritmos probados tienden a una mejor solución.

7.5. Conclusiones

En este capítulo se propuso un método basado en optimización para seleccionar automáticamente conjuntos de entrenamiento para filtros compuestos. Para este fin, se propusieron tres funciones objetivo como criterios de optimización en la selección. La metodología propuesta fue evaluada con los algoritmos de optimización (1 + 1)-

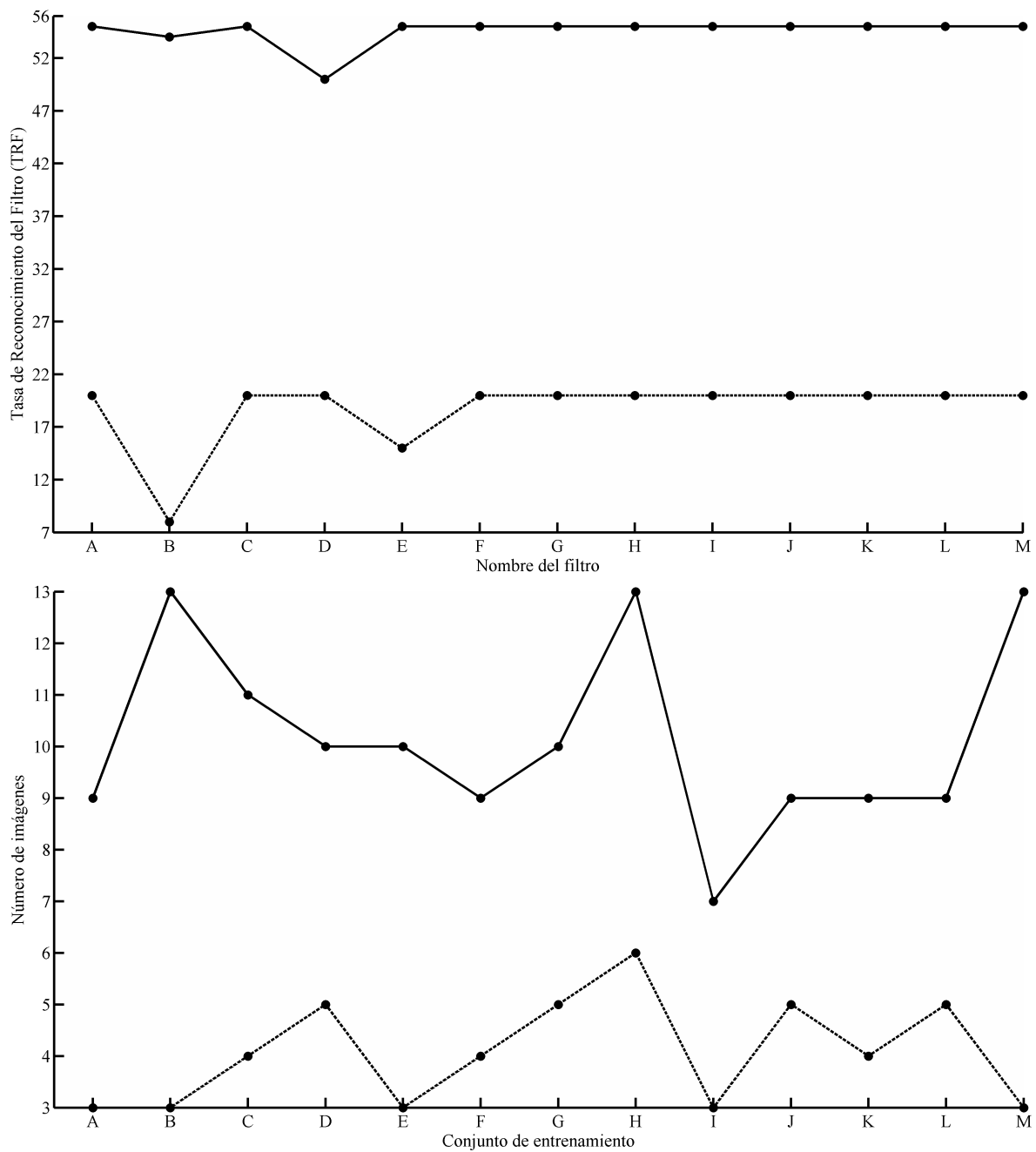


Figura 7.9: Desempeño TRF del filtro ASPOF (arriba) y tamaño del conjunto de entrenamiento (abajo) usando la el conjunto CMU. Línea continua para el resultado obtenido por el método propuesto y línea segmentada para el resultado del método convencional.

Evolution Strategy, Hill Climbing, Simulated Annealing and Grey Wolf Optimizer y los filtros de correlación UNCF, MACH y UOTSDF. Mientras que el filtro MACH diseñado con GWO y TRF obtuvo el mejor desempeño en la identificación y verifi-

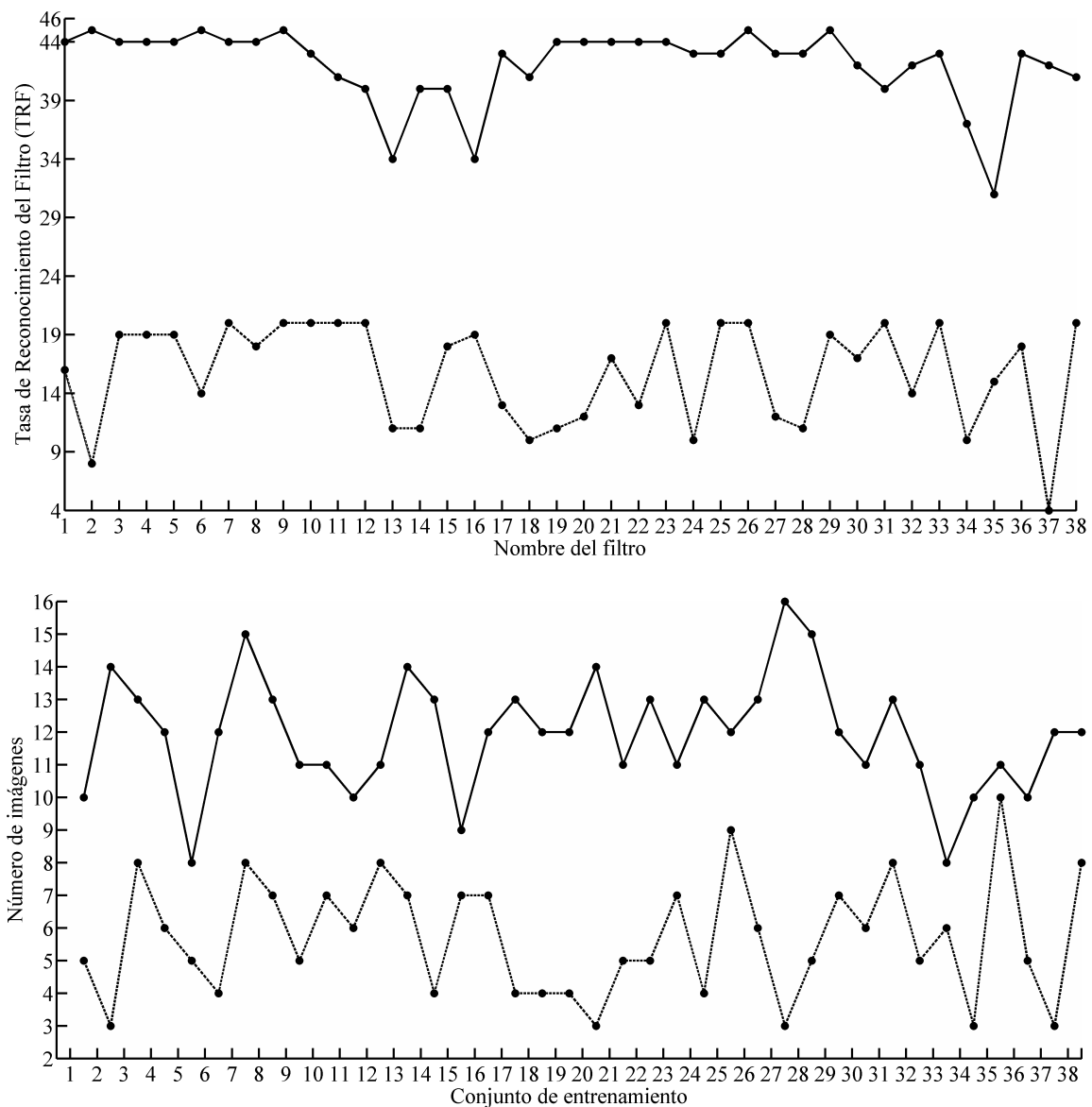


Figura 7.10: Desempeño TRF del filtro ASPOF (arriba) y tamaño del conjunto de entrenamiento (abajo) usando la el conjunto CMU. Línea continua para el resultado obtenido por el método propuesto y línea segmentada para el resultado del método convencional.

cación de caras con iluminación homogénea y expresiones, este filtro obtuvo el peor desempeño para rechazar caras impostoras. Por otra parte, el filtro UNCF diseñado con GWO o ES y APP obtuvo el mejor desempeño bajo condiciones de iluminación variable, y más importante aún, obtuvo el mejor desempeño en el rechazo exitoso de impostores. Se pueden resaltar cuatro fortalezas de la metodología propuesta. Primero, los filtros diseñados por la metodología propuesta logran resultados competitivos en el

reconocimiento facial. La metodología propuesta puede usar algoritmos de optimización que van de soluciones basadas en un solo individuo a soluciones basadas en una población de individuos. Tercero, las funciones objetivo propuestas pueden ser usadas como criterios de optimización adecuados para automáticamente seleccionar conjuntos de entrenamiento. Cuarto, la metodología propuesta selecciona conjuntos de entrenamiento menores que la mayoría de los enfoques establecidos en la literatura. Debido a esos cuatro puntos, la metodología propuesta es una buena alternativa para la selección de conjuntos de entrenamiento para la síntesis de filtros compuestos con desempeño optimizado en términos de las métricas FERET, tolerancia a distorsión y número de imágenes de entrenamiento.

Capítulo 8

Re-identificación facial con predicción del campo de vista destino

El problema de la re-identificación facial consiste en determinar si una cara que aparece en un vídeo capturado por una cámara, coincide con las caras que aparecen en vídeos capturados por otras cámaras. En otras palabras, una cara detectada y reconocida en una ubicación, es comparada con un conjunto factible de candidatos detectados en otras ubicaciones y sobre el tiempo en busca de coincidencias.

El uso de la cara en la re-identificación tienen varias ventajas. Primero, la cara permanece estable por un intervalo mayor de tiempo, por lo que puede ser aprovechado para buscar personas en vídeos capturados en tiempos diferentes (Bauml y cols., 2010; Gandhi y Trivedi, 2007). Segundo, la cara es única para cada persona y es universal por lo que puede ser usado como rasgo biométrico para la identificación de la persona. Tercero, debido a que es un rasgo visible de la persona, su uso posibilita el reconocimiento discreto, no intrusivo y a distancia (Stallkamp, Ekenel, y Stiefelhagen, 2007).

Una solución al problema de re-identificación pudiera incluir las tareas de procesamiento facial tales como detección, reconocimiento y seguimiento. Los avances en el procesamiento de caras permiten realizar estas tareas en presencia de ruido, oclusión parcial, ligeros movimientos laterales y variabilidad en iluminación. Sin embargo, la re-identificación en campos de vista (CDV) disjuntos y bajo condiciones no restringidas es un problema que permanece abierto a la investigación. Las condiciones no restringidas causan que las caras a reconocer y las plantillas biométricas almacenadas presenten diferencias en iluminación, pose, expresión, edad, entre otros. Estas distorsiones confunden a los algoritmos de procesamiento facial, por lo que su desempeño disminuye considerablemente.

Cuando una persona entra a una red de cámaras a través del CDV i , se detecta e identifica inmediatamente la cara para iniciar el proceso de re-identificación. La cara es seguida dentro de CDV i hasta que la persona abandone el lugar. Cuando esta persona aparezca en otro campo de vista j (CDV destino), inmediatamente su cara es detectada

y se determina si coincide con la cara que apareció en i . Este proceso se repite mientras la persona esté dentro de la red de cámaras.

Es fácil notar que un conocimiento anticipado del CDV j permite la optimización de los recurso y la anticipación a posibles eventos no deseados. La predicción del CDV j puede lograrse mediante el uso de una simple estrategia que use la frecuencia con que un CDV es visitado. Sin embargo, tal método es incapaz de adaptarse a cambios y/o exploración de nuevas rutas. Por esta razón, en este capítulo se propone un algoritmo de re-identificación facial que usa el modelo descrito en el Apéndice A para predecir el campo de vista destino de la persona.

8.1. Método de re-identificación facial

La Figura 8.1 muestra el esquema de los componentes del método de re-identificación facial propuesto. Como se puede observar, en cada campo de vista se detecta, reconoce y se da un seguimiento a la cara de la persona. Con las muestras faciales capturadas en el seguimiento, se construye un descriptor facial el cual se comparará con descriptores generados en otros campos de vista para determinar si existe coincidencia. Esta correspondencia se determina mediante correlación y la métrica PSR para estimar la agudeza del pico de correlación producido.

Los principales módulos del método propuesto son la *detección facial*, *registro facial*, *reconocimiento facial*, *seguimiento facial*, *predicción del destino*, generación del *descriptor facial* y el *establecimiento de correspondencia*. A continuación se describen los módulos de *Registro facial*, *descriptor facial*, *establecimiento de la correspondencia* y la *predicción del campo de vista destino*.

8.1.1. Registro facial

Como cualquier sistema de reconocimiento facial, es necesario contar con una galería G de plantillas biométricas de las personas conocidas. En este trabajo, el *registro facial* se realiza tal y como se muestra en la Figura 8.2. Primero, se obtienen varias y diferentes imágenes faciales de la persona objetivo. Estas muestras se obtienen capturando imágenes mientras la persona mueve su rostro de la siguiente manera. Partiendo de una posición de mirada directa a la cámara (posición neutral), la cara se mueve lentamente hacia arriba y luego hacia abajo. Entonces se regresa a la posición neutral para girar la mirada hacia la izquierda y la derecha para obtener muestras faciales con rotación fuera del plano (pose). Posteriormente, manteniendo fijamente la mirada hacia la cámara se mueve la cabeza hacia la izquierda y derecha para capturar caras con rotación sobre el plano. Finalmente, se le pide a la persona realizar expresiones faciales. Segundo, el conjunto de imágenes faciales capturados $f_i(x, y)$ se asume como el conjunto $T_{id,i}$ que será examinado por la metodología presentada en el Capítulo 7 para seleccionar solamente las imágenes más representativas como conjunto de entrenamiento T . Para esta selección se sugiere usar el algoritmo de optimización GWO, la función objetivo

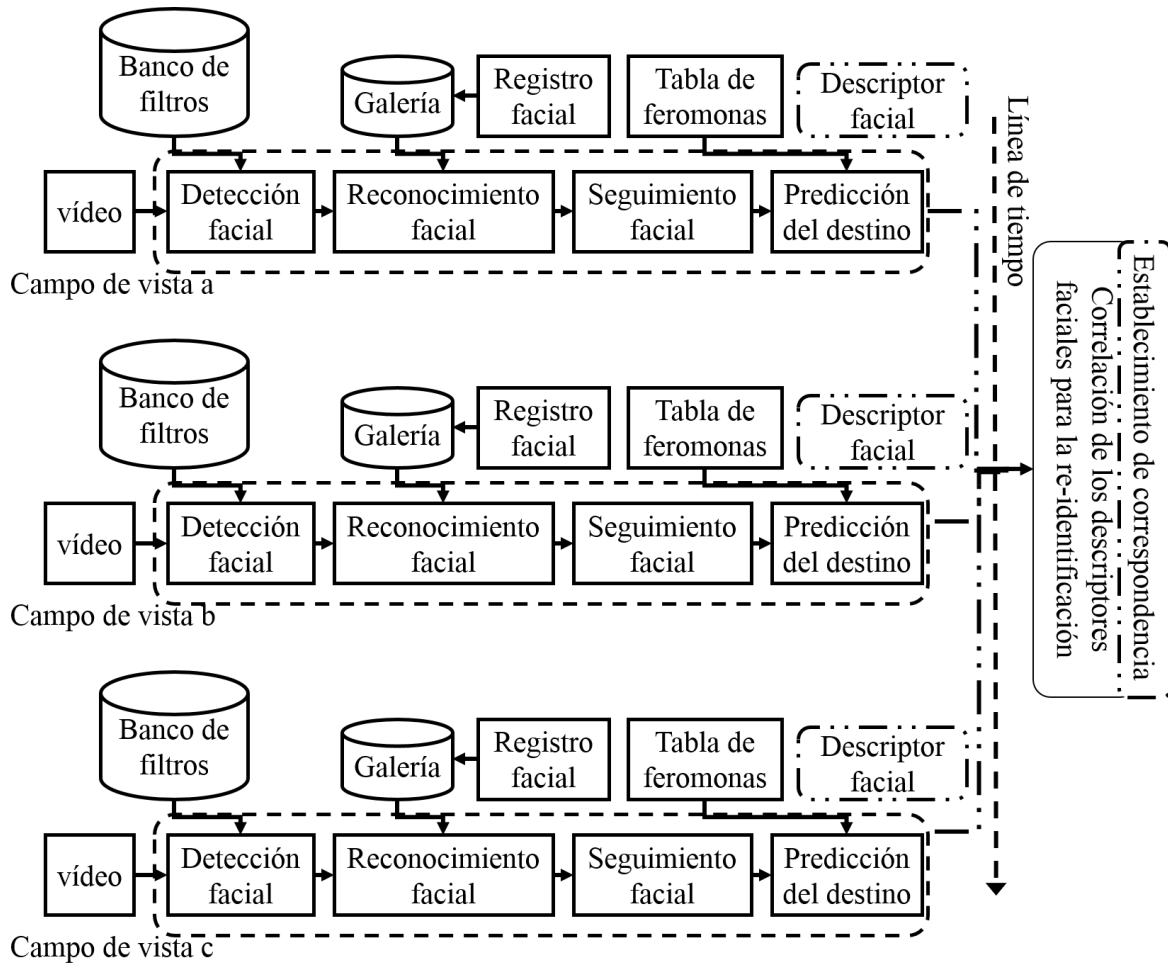


Figura 8.1: Diagrama del método de re-identificación facial propuesto.

FRR y el filtro *UNCF*. Esta combinación se recomienda debido a que obtuvo el mejor desempeño bajo condiciones de iluminación variable que es muy común en ambientes no restringidos. Tercero, con el conjunto de entrenamiento T seleccionado se construye un filtro de correlación compuesto. Cuarto, se capturan los datos personales del objetivo para la construcción de su identidad. Finalmente, el filtro \mathbf{h}_i se asocia con los datos personales capturados y se almacenas en G para ser usados en la re-identificación de la persona.

Para este fin se implementó el módulo de *Registro facial facial* mostrado en la Figura 8.3. La interfaz muestra la escena de entrada $s_i(x, y)$, la imagen facial $f(x, y)$ capturado en $s_i(x, y)$, la actualización del filtro de seguimiento \mathbf{h}_i , el plano de correlación $g_i(x, y)$ y los controles de operación. El sistema opera de la siguiente manera. El objetivo introduce su número de seguridad social, el cual servirá como identificador único y personal. Posteriormente puede iniciar con la captura oprimiendo el botón *Iniciar captura*. Para esto, solo se pide al objetivo que dirija su mirada hacia la cámara y realice los movi-

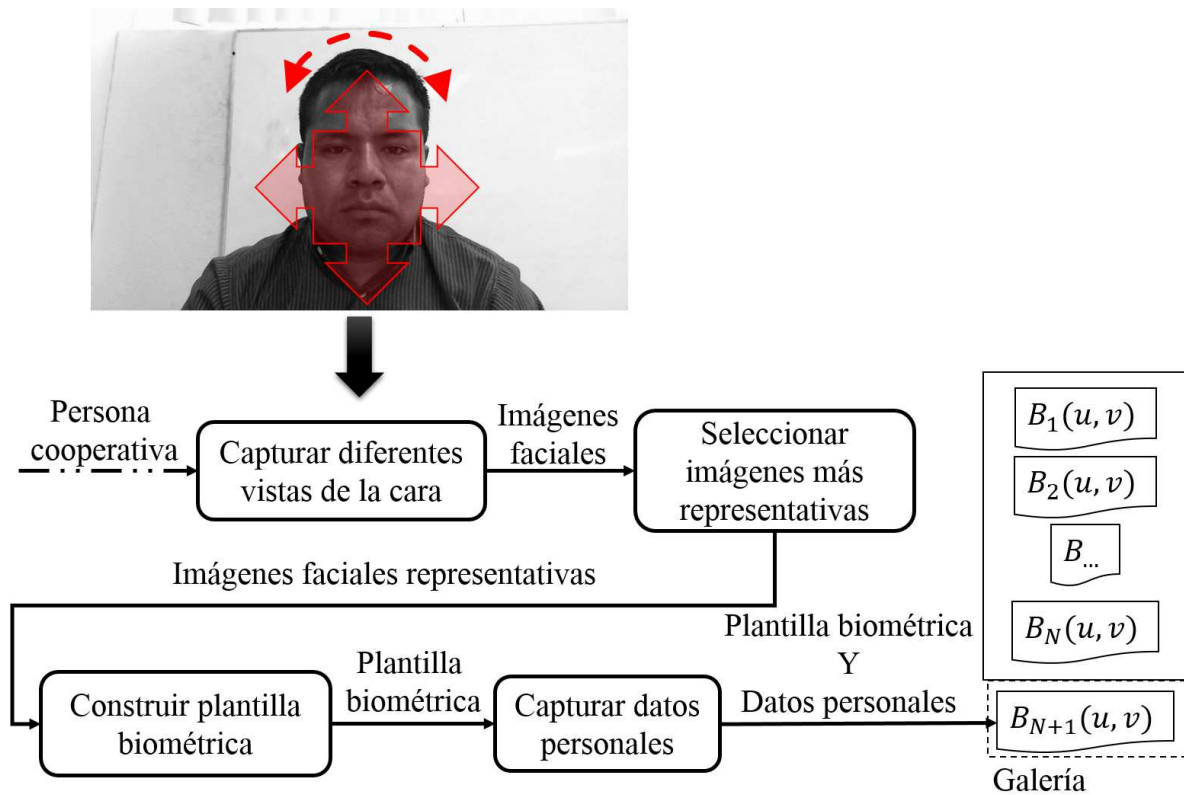


Figura 8.2: Diagrama básico de bloques del proceso de registro facial para objetivos nuevos.

miento previamente descritos. El sistema está abierto a cualquier cantidad de capturas. Sin embargo, en los experimentos se observó que un mínimo de 60 imágenes genera un buen conjunto de entrenamiento. El usuario puede terminar la captura cuando lo desee mediante el botón *Terminar captura*. Una vez terminada la captura, procede a crear el filtro de forma automática como se describió en el Capítulo 7. Finalmente, después de que se haya creado el filtro, se capturan los datos personales del usuario que se muestran en la Figura 8.4. Los datos personales a ingresar son nombre completo, fecha y lugar de nacimiento, direcciones, entre otros. El sistema admite dos tipos de usuario, *objetivo* y *operador*. El *objetivo* es la persona a re-identificar, mientras que el *operador* es la persona que opera y mantiene el sistema.

Para probar la funcionalidad de este sub-módulo, se ejecutó el módulo de *Registro facial* para un *objetivo* en específico. La Figura 8.5 muestra 110 imágenes faciales diferentes de una misma persona capturada por el sistema en una sesión. Como se puede observar, este conjunto contiene imágenes frontales, con pose, rotación y expresiones. Es importante notar que en un ambiente no restringido, estas distorsiones faciales son las más comunes. Por lo que es importante que estén representados en el conjunto de entrenamiento. Las imágenes capturadas, son divididas en los conjuntos $T_{id,i}$ y P_i los cuales se muestran en la Figura 8.6. El sistema seleccionó aleatoriamente $\frac{1}{5}$ del total

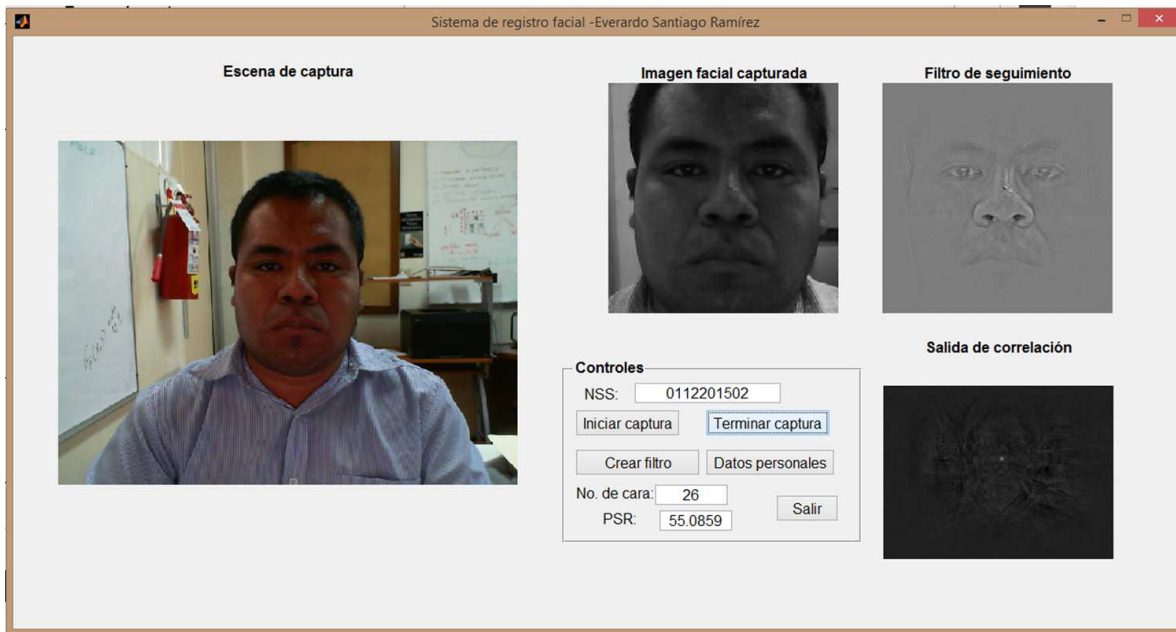


Figura 8.3: Interfaz gráfica de usuario del módulo de registro facial.

de imágenes faciales capturadas para el conjunto $T_{id,i}$, mientras que las restantes 88 imágenes fueron puestas en el conjunto P_i . De las 22 imágenes disponibles, el sistema seleccionó 13 imágenes para el conjunto de entrenamiento T_i (véase la Figura 8.7). Como se puede observar, el conjunto seleccionado contiene imágenes faciales frontales, con pose, rotación y expresiones, las cuales son las más representativas.

El conjunto de entrenamiento T_i seleccionado por el módulo de *Registro facial* es adecuado para construir un filtro de correlación compuesto \mathbf{h}_i . Este filtro es suficiente para reconocer a la persona objetivo en un ambiente no restringido. Es importante indicar que aunque el registro facial se realiza una sola vez, es necesario una actualización después de varios años de uso. Esto debido a que la cara sufre cambios principalmente causados por la edad.

8.1.2. Descriptor facial

Un *descriptor facial* puede definirse como una señal que describe eficientemente las características generales y particulares de la cara de una persona. Sea $F = \{f_1(x, y), f_2(x, y), \dots, f_N(x, y)\}$ el conjunto de N imágenes faciales capturadas durante el seguimiento facial en el CDV i . Dichas imágenes se promedian para generar una simple imagen facial promedio. A la imagen facial promedio se le aplica la transformada de Fourier y luego el filtrado no lineal para producir el *descriptor facial* \mathbf{h}_i .

Sistema de registro facial-Everardo Santiago Ramírez

Datos personales

Nombre nombre:

Fecha de nacimiento: Día Mes Año:

Lugar de nacimiento:

CURP:

Dirección:

Correo electrónico:

Teléfono: Celular:

Ocupación:

Tipo usuario:

Contraseña:

Controles

Figura 8.4: Interfaz gráfica de usuario del módulo de captura de registro de datos personales.

8.1.3. Establecimiento de correspondencia

El establecimiento de la correspondencia tiene como objetivo determinar coincidencias entre los diferentes descriptores generados en los seguimientos. Sean \mathbf{h}_i y \mathbf{h}_j respectivamente los descriptores faciales generados por seguimientos en los campos de vista i y j . Por lo tanto, el objetivo es determinar el grado de similitud de \mathbf{h}_i y \mathbf{h}_j , el cual se obtiene con una correlación. En el plano de correlación, se mide la agudeza del pico usando la métrica PSR . Si el valor $PSR \geq \tau$, entonces se declara una correspondencia y se asume que la persona es la misma en ambos campos de vista.

8.1.4. Predicción del campo de vista destino

Las personas habitualmente llegan a un CDV siguiendo una ruta previamente aprendida. El recorrido desde el CDV origen hasta el CDV destino exhibe un patrón de movimientos que pueden ser analizados para predecir automáticamente el CDV destino en futuros recorridos. Esta predicción permite la optimización de recursos al enfocar la atención únicamente en los campos de vista con mayor probabilidad de que aparezca la



Figura 8.5: Diferentes imágenes faciales de una misma persona capturadas por el módulo de *Registro facial*.

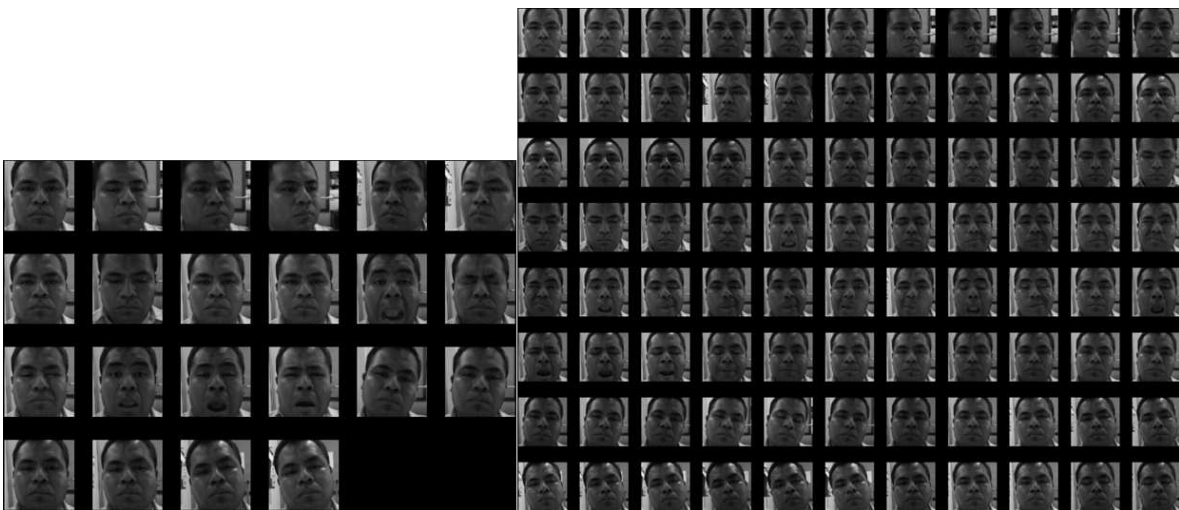


Figura 8.6: Conjuntos $T_{id,i}$ (izquierda) y P_i (derecha) correspondientes a la i -ésima persona.

persona. Además, con esta predicción se mejora la experiencia del usuario en un sistema biométrico prácticamente ubicuo en el ambiente.

En este trabajo de investigación se propone un modelo de predicción del CDV destino mediante los principios de evaporación y actualización del nivel de feromonas. Considere una red de cámaras fijas como un grafo conectado $G = (V, A)$, donde V es el conjunto



Figura 8.7: Conjunto de entrenamiento T_i seleccionado por el módulo de *Registro facial*.

de $N = |V|$ nodos (campos de vista) y A es el conjunto de aristas (dirigidos y no dirigidos) que conectan a los nodos. El arco formado por los nodos i y j es al que se le desea asignar un valor de peso que pueda ser usado para probabilísticamente predecir el campo de vista destino.

Los valores de peso de las aristas del grafo se mantiene en una matriz F para cada persona. La posición $F(i, j)$ contiene el nivel de feromonas (peso) que indican la deseabilidad de la persona por dicho camino y en consecuencia al destino que conduce.

La predicción está dada por:

$$P(j) = \frac{F(i, j)}{\sum_{l \in V_i} F(i, l)}, \quad (8.1)$$

donde P es la probabilidad de que se dirija al CDV j , V_i es la vecindad de i y l es uno de sus miembros. De todas las probabilidades calculadas, se selecciona como destino el vecino con mayor probabilidad. Cuando una persona selecciona un CDV s , hay que actualizar el nivel de feromonas mediante:

$$F_{i,s} = F_{i,s} + \frac{F_{i,s}}{(\sum_{l \in V_i} T_{i,l})^m}. \quad (8.2)$$

Si el CDV s es diferente al CDV j estimado mediante la Ecuación 8.1, entonces se realiza una disminución del nivel de feromonas de la arista (i, j) . Esta penalización impulsa al modelo a explorar un nuevo CDV destino y está dada por:

$$F(i, j) = F(i, j) - \beta \frac{F(i, j)}{(\sum_{l \in V_i} F(i, l))^m}. \quad (8.3)$$

Cuando la persona abandona la red de cámaras, entonces se realiza una evaporación de los niveles de feromonas mediante:

$$F_{i,j} = (1 - \frac{1}{\rho^r}) F_{i,j}, \quad (8.4)$$

donde ρ es el número de campos de vista visitados durante el recorrido y r la cantidad de recorridos en la zona vigilada. En el Apéndice A se presentan evidencias que validan el funcionamiento de este modelo de predicción.

8.1.5. Algoritmo de re-identificación facial propuesto

El Algoritmo 7 integra todos los módulos mencionados previamente para la tarea de re-identificación facial. Se recibe como argumentos la matriz de feromonas F_i de la i -ésima persona y una lista de adyacencia del grafo $G(V, A)$ correspondiente a la red de cámaras. El algoritmo funciona de la siguiente manera. Partiendo del CDV $i \in G$, primero se detecta la cara de la persona. La *detección facial* se realiza mediante el método reportado en el Capítulo 5. Segundo, la cara detectada es identificada usando las plantillas biométricas creadas en el *Registro facial*. Si se encuentra alguna coincidencia, entonces se devuelve su identidad. Tercero, mediante el método propuesto en el Capítulo 6 se hace el seguimiento de la cara detectada y reconocida en i . Durante el seguimiento se obtienen las muestras faciales para actualizar el filtro de seguimiento. Al finalizar el seguimiento, con las muestras capturadas se construye el *descriptor facial* para la re-identificación. Cuarto, cuando se detecta que la persona ha abandonado el CDV i , entonces se emplea el modelo propuesto en el Apéndice A para predecir el CDV j destino

Algoritmo 7: Algoritmo de re-identificación facial con predicción del campo de vista destino.

Datos: Matriz de feromonas F_i y lista de adyacencia $G(N, A)$
Resultado: T_i y lista de campos de vista recorridos

- 1 $centinela \leftarrow 0$
- 2 $CDV_i \leftarrow$ CDV que es entrada/salida ($CDV_i \in G$)
- 3 $f(x, y) \leftarrow DetectarCara(CDV_i)$
- 4 $id_i \leftarrow IdentificarCara(f(x, y))$
- 5 $listaCarasCapturadas \leftarrow SeguirCara(f(x, y), CDV_i)$
- 6 $h_i \leftarrow$ Construir descriptor facial usando $listaCarasCapturadas$
- 7 $CDV_j \leftarrow PredecirNodoDestino(V_i, F_i)$
- 8 **mientras** $centinela \neq 1$ **hacer**
- 9 $f(x, y) \leftarrow DetectarCara(CDV_j)$
- 10 $id_j \leftarrow IdentificarCara(f(x, y))$
- 11 $listaCarasCapturadas2 \leftarrow SeguirCara(f(x, y), CDV_j)$
- 12 $h_j \leftarrow$ Construir descriptor facial usando $listaCarasCapturadas2$
- 13 Medir similitud entre h_i y h_j
- 14 **si** Hay similitud **entonces**
- 15 | Agregar CDV_j al recorrido
- 16 **en otro caso**
- 17 | Iniciar nueva re-identificación de $f(x, y)$ en CDV_j con una nueva instancia del Algoritmo 7
- 18 **fin**
- 19 Actualizar F_i
- 20 Borrar CDV_j
- 21 $CDV_j \leftarrow PredecirNodoDestino(V_j, F_i)$
- 22 **si** CDV_j es nodo salida **entonces**
- 23 | Evaporar F_i
- 24 | $centinela = 1$
- 25 **fin**
- 26 **fin**

y se actualiza la tabla de feromonas F_i . Finalmente, cuando la persona ha abandonado la zona vigilada por la red de cámaras, se realiza una evaporación de feromonas.

El método propuesto está basado en filtros de correlación. Sin embargo, es factible el uso de otras técnicas de procesamiento facial. Por ejemplo, el algoritmo Viola-Jones (P. Viola y Jones, 2001; M. Viola y cols., 2003) para la detección facial, PCA para el reconocimiento y MOSSE para el seguimiento facial. Para ejemplificar la obtención del grafo G , considere la Figura 8.8 que muestra una red de cámaras. Cada cámara representa un nodo y los caminos que enlazan dichas cámaras forman las aristas.

Uno de los requisitos del método propuesto es que las cámaras tienen que estar instaladas de tal forma que puedan capturar sin mucho problema la cara de la persona.

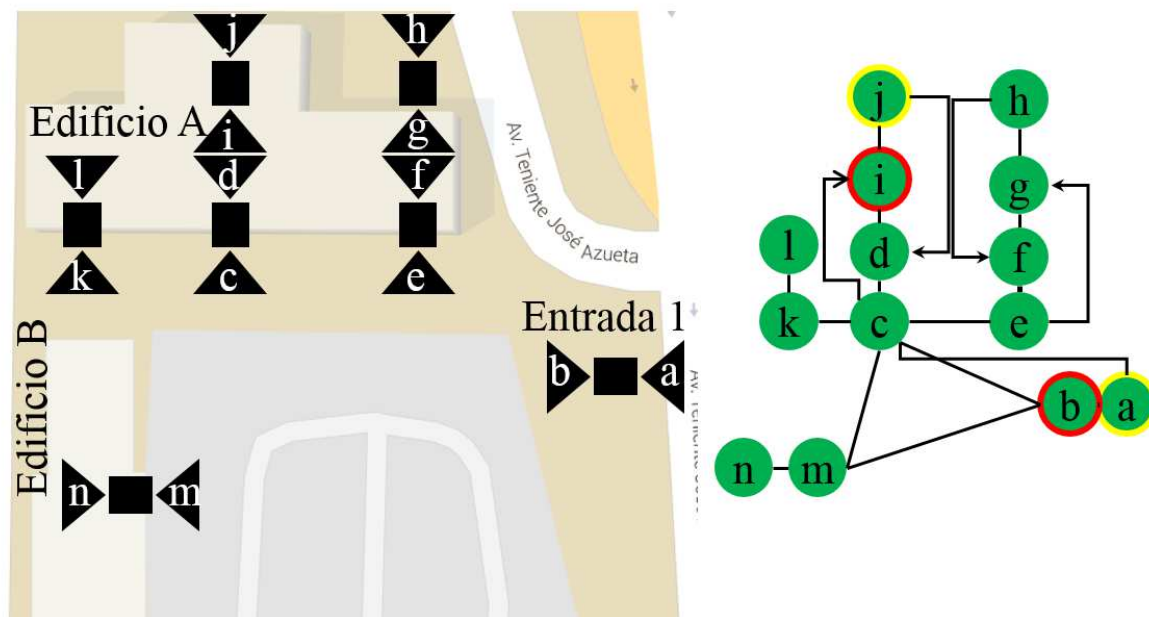


Figura 8.8: Red de cámaras (izquierda) y su respectivo grafo G (derecha).

La Figura 8.9 muestra una configuración deseada de cada cámara, de tal forma que sea posible capturar una imagen de la cara de la persona con las características faciales visibles. Esto requiere que la implementación del método propuesto requiera de la reubicación de las cámaras existentes o la instalación de una nueva red. Aunque esto pareciera una desventaja, lo cierto es que trae como beneficio una mayor exactitud en la re-identificación y un aprovechamiento real de la infraestructura. Pues en las redes de cámaras existentes, generalmente las cámaras se ubican en lugares altos donde no es posible capturar la cara de las personas, lo cual hace imposible su uso en las tareas de visión artificial pues no se puede capturar características más distintivas y detalladas de la persona.

8.2. Resultados experimentales

El algoritmo de re-identificación facial propuesto fue evaluado en un escenario real. El objetivo de este experimento es validar que el algoritmo propuesto es capaz de predecir el CDV j destino, así como descubrir y aprender nuevas rutas.

8.2.1. Configuración de los datos de prueba

La Figura 8.10 muestra el escenario de prueba, el cual consiste de los campos de vista $\{a, b, c, d\}$, donde la entrada a la red de cámaras es a través del CDV a , mientras la salida es por b o incluso por a . En un entorno real, existen barreras o restricciones sobre el área vigilada que permiten a la persona circular sobre un par de campos de

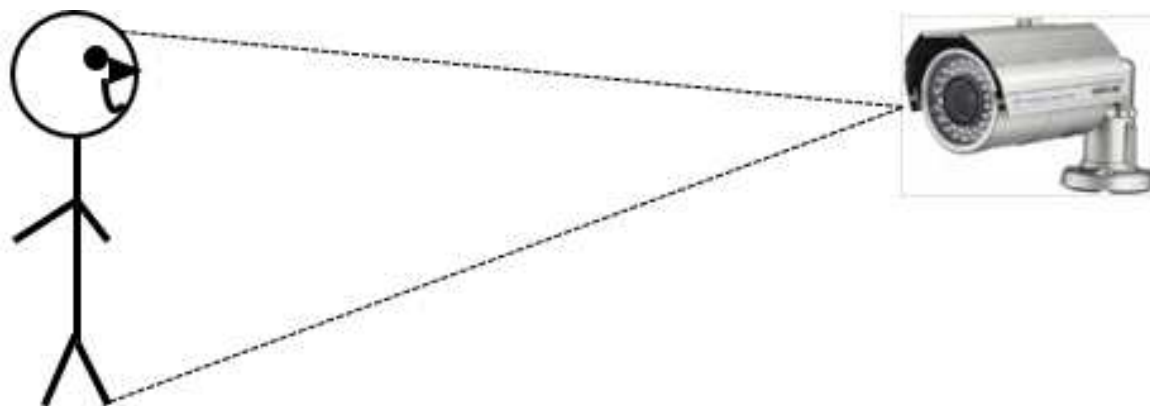


Figura 8.9: Posición de la cámara con respecto a la cara de la persona (arriba) e imagen facial deseada (abajo).

vista en ambas direcciones o en una sola dirección. Por ejemplo, en el grafo de la Figura 8.10 se puede ir directamente del CDV *a* hacia el CDV *c* pero no viceversa. Por tal motivo, el grafo contiene aristas no dirigidas y aristas dirigidas que representan estas restricciones.

Sea R_i el conjunto de campos de vista que una persona acostumbra recorrer, el cual caracteriza un patrón de movimientos que pueden ser aprendidos y usado para predecir el CDV en futuros recorridos. La Figura 8.11 muestra tres patrones de movimientos que se ejecutarán de forma secuencial sobre el escenario LPA2. Como la red se conforma de cuatro nodos y considerando los resultados presentados en la Figura A.5 del Apéndice A, para cada patrón de movimientos se realizaron diez recorridos, obteniendo así un total de 30 recorridos en el experimento.

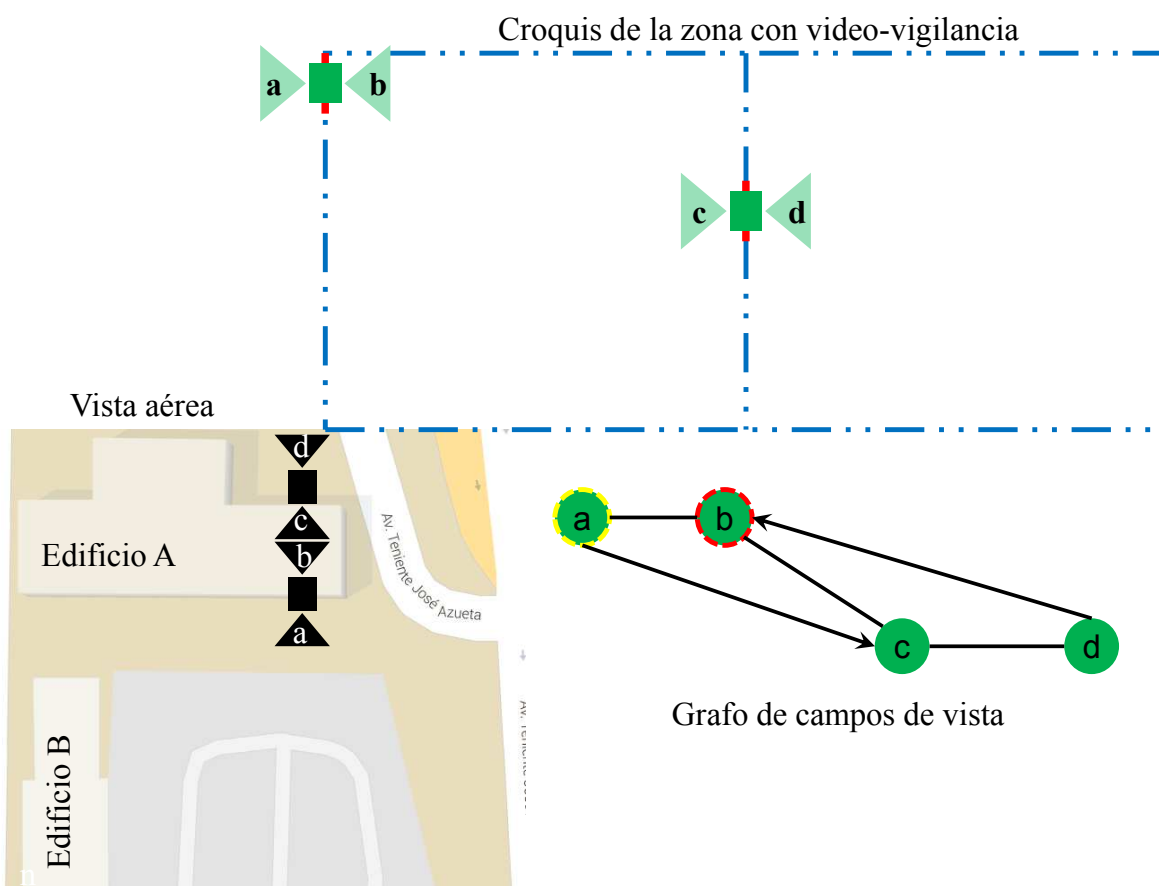


Figura 8.10: Vista aérea, croquis y grafo de campos de vista del área vigilada por cámaras denominado LPA2.

Los vídeo de los recorridos fueron grabados a 30 cuadros por segundo. Las imágenes capturadas fueron transformadas a escala de grises y con una resolución de 480×640 píxeles. La Tablas [8.1, 8.2, 8.3] muestran la cantidad de imágenes en cada vídeo grabado. En dichos vídeos, la cara de la persona presenta variaciones en iluminación, pose,

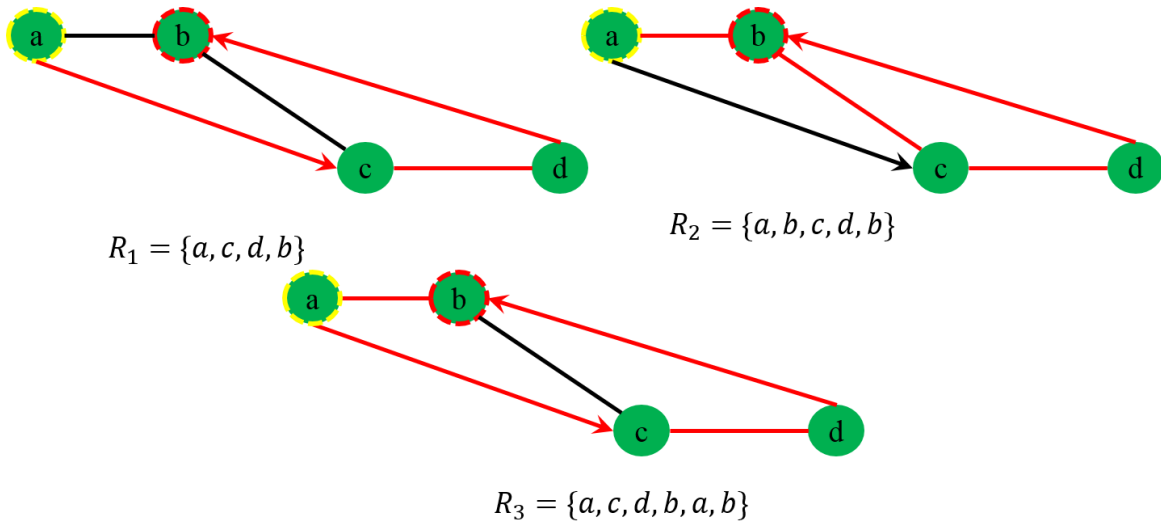


Figura 8.11: Patrones de movimientos para la evaluación del algoritmo de re-identificación sobre el escenario LPA2.

rotación, escala, oclusión parcial y expresiones faciales. Además de estas variaciones, la persona giraba lo que causaba que su cara apareciera/desapareciera.

Tabla 8.1: Datos de los vídeos capturados para el patrón de movimiento 1. Notación: número de recorrido (NR), campo de vista (CDV) y total de imágenes (TI).

| | | | | | | | | | | | | | | | | | | | | |
|-----|-----|-----|----|-----|-----|-----|----|----|----|-----|-----|-----|----|----|----|-----|----|-----|----|-----|
| NR | 1 | 1 | 1 | 1 | 2 | 2 | 2 | 2 | 3 | 3 | 3 | 3 | 4 | 4 | 4 | 4 | 5 | 5 | 5 | 5 |
| CDV | a | c | d | b | a | c | d | b | a | c | d | b | a | c | d | b | a | c | d | b |
| TI | 85 | 152 | 69 | 168 | 270 | 159 | 77 | 83 | 83 | 81 | 64 | 87 | 51 | 91 | 97 | 150 | 83 | 356 | 41 | 231 |
| NR | 6 | 6 | 6 | 6 | 7 | 7 | 7 | 7 | 8 | 8 | 8 | 8 | 9 | 9 | 9 | 9 | 10 | 10 | 10 | 10 |
| CDV | a | c | d | b | a | c | d | b | a | c | d | b | a | c | d | b | a | c | d | b |
| TI | 170 | 71 | 88 | 73 | 56 | 61 | 82 | 82 | 58 | 351 | 190 | 138 | 58 | 42 | 71 | 58 | 89 | 42 | 71 | 87 |

Tabla 8.2: Datos de los vídeos capturados para el patrón de movimiento 2. Notación: número de recorrido (NR), campo de vista (CDV) y total de imágenes (TI).

| | | | | | | | | | | | | | | | | | | | | | | | | | | |
|-----|-----|-----|-----|-----|-----|----|----|-----|-----|----|----|-----|-----|-----|-----|----|-----|-----|-----|----|----|----|-----|----|-----|----|
| NR | 11 | 11 | 11 | 11 | 11 | 12 | 12 | 12 | 12 | 12 | 13 | 13 | 13 | 13 | 13 | 14 | 14 | 14 | 14 | 14 | 15 | 15 | 15 | 15 | 15 | |
| CDV | a | b | c | d | b | a | b | c | d | b | a | b | c | d | b | a | b | c | d | b | a | b | c | d | b | |
| TI | 58 | 83 | 96 | 42 | 61 | 62 | 83 | 39 | 126 | 76 | 72 | 124 | 433 | 279 | 107 | 90 | 73 | 128 | 127 | 85 | 87 | 73 | 128 | 86 | 91 | |
| NR | 16 | 16 | 16 | 16 | 16 | 17 | 17 | 17 | 17 | 17 | 18 | 18 | 18 | 18 | 18 | 19 | 19 | 19 | 19 | 19 | 19 | 20 | 20 | 20 | 20 | 20 |
| CDV | a | b | c | d | b | a | b | c | d | b | a | b | c | d | b | a | b | c | d | b | a | b | c | d | b | |
| TI | 111 | 106 | 112 | 101 | 106 | 72 | 53 | 103 | 109 | 93 | 71 | 68 | 174 | 130 | 244 | 80 | 188 | 126 | 72 | 68 | 77 | 68 | 90 | 56 | 267 | |

8.2.2. Resultados numéricos y discusión

En esta sección se presentan los resultados numéricos obtenidos sobre el conjunto de datos descritos previamente. Debido a que cada algoritmo de procesamiento facial fue extensamente evaluado en su respectivo capítulo, aquí nos enfocaremos únicamente en

Tabla 8.3: Datos de los vídeos capturados para el patrón de movimiento 3. Notación: número de recorrido (NR), campo de vista (CDV) y total de imágenes (TI).

| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
|-----|----|-----|----|----|-----|-----|-----|-----|----|-----|----|-----|-----|----|----|----|----|-----|----|----|----|----|----|----|----|-----|----|-----|----|----|----|---|
| NR | 21 | 21 | 21 | 21 | 21 | 21 | 22 | 22 | 22 | 22 | 22 | 22 | 23 | 23 | 23 | 23 | 23 | 23 | 24 | 24 | 24 | 24 | 24 | 24 | 24 | 25 | 25 | 25 | 25 | 25 | 25 | |
| CDV | a | c | d | b | a | b | a | c | d | b | a | b | a | c | d | b | a | b | a | c | d | b | a | b | a | c | d | b | a | b | a | b |
| TI | 81 | 171 | 65 | 94 | 68 | 63 | 250 | 171 | 65 | 148 | 72 | 90 | 65 | 63 | 55 | 72 | 93 | 279 | 55 | 66 | 64 | 72 | 66 | 43 | 73 | 142 | 64 | 120 | 59 | 49 | | |
| NR | 26 | 26 | 26 | 26 | 26 | 26 | 27 | 27 | 27 | 27 | 27 | 27 | 28 | 28 | 28 | 28 | 28 | 28 | 29 | 29 | 29 | 29 | 29 | 29 | 29 | 30 | 30 | 30 | 30 | 30 | 30 | |
| CDV | a | c | d | b | a | b | a | c | d | b | a | b | a | c | d | b | a | b | a | c | d | b | a | b | a | c | d | b | a | b | a | b |
| TI | 66 | 149 | 53 | 66 | 197 | 100 | 63 | 27 | 38 | 66 | 45 | 100 | 139 | 63 | 38 | 82 | 67 | 55 | 89 | 63 | 79 | 87 | 80 | 91 | 88 | 54 | 79 | 62 | 88 | 91 | | |

evaluar en la practica el modelo de predicción propuesto en el Apéndice A. Sin embargo, es importante considerar que un buen desempeño en la predicción es producto de un buen procesamiento facial.

Antes de presentar los resultados, vamos a explicar el comportamiento del modelo de predicción ante diferentes patrones de movimiento. Considere el grafo de la Figura 8.10 y los patrones de movimientos $R'_1 = \{a, c, d, b\}$ y $R'_2 = \{a, c, b\}$. La Figura 8.12 muestra el desempeño RES obtenido. Para R'_1 , al inicio el modelo obtiene un desempeño nulo. Sin embargo, a partir del segundo recorrido se obtiene un $RES = 1$. En cuando el modelo se encuentra con R'_2 , el desempeño RES cae hasta 0.67 y se mantiene estable hasta en el recorrido 13. Posteriormente, el modelo aprende a distinguir bien el nuevo patrón y el desempeño sube hasta $RES = 1$.

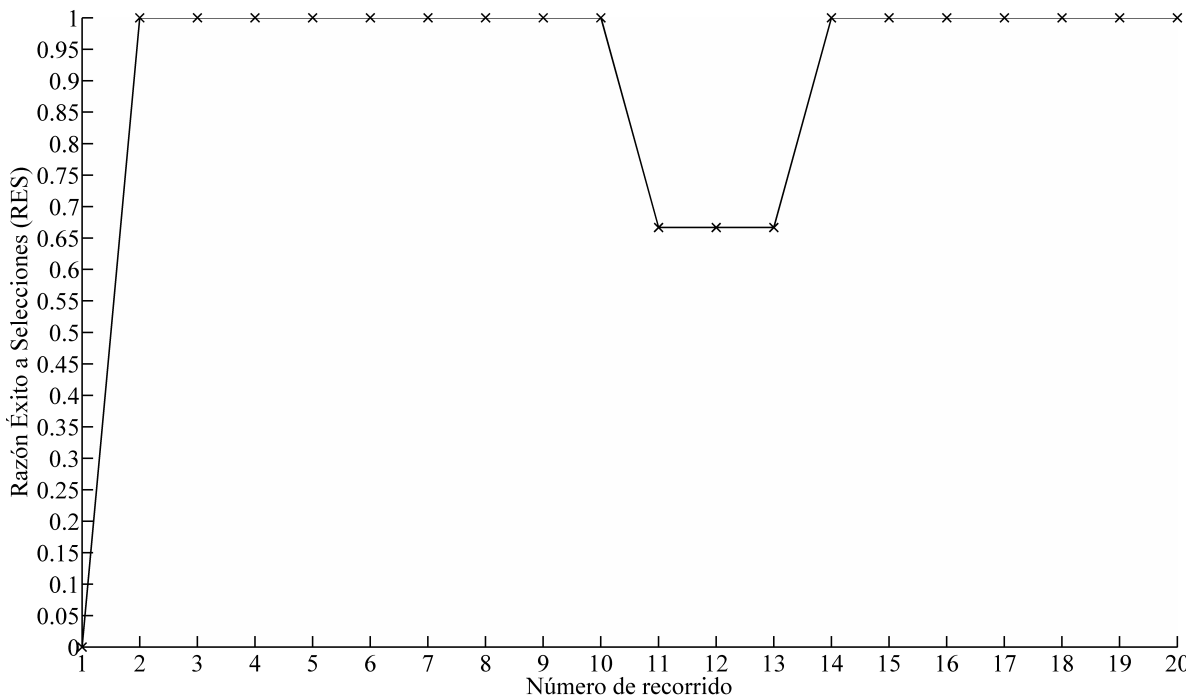


Figura 8.12: Desempeño RES de los patrones de movimiento $R'_1 = \{acdb\}$ y $R'_2 = \{acb\}$ sobre el grafo de la Figura 8.10.

Las Tablas [8.4, 8.5, 8.6] muestran los resultados del establecimiento de la correspondencia para los patrones de movimientos R_1 , R_2 y R_3 , respectivamente. Con un

umbral de reconocimiento $\tau = 8$ se determinan las coincidencias. No es necesario que todos los CDV's coincidan con todos, es suficiente con que se encuentre una coincidencia para establecer la re-identificación. Para R_1 , los *descriptores faciales* de los campos de vista b, a y b respectivamente de los recorridos 2, 5 y 9 no coincidieron con ningún otro CDV. Para R_2 , los CDV's d y a de los recorridos 11 y 17, respectivamente, no se encontró coincidencias. Finalmente, para R_3 los CDV's b y a de los recorridos 27 y 29, respectivamente, tampoco coincidieron con otros campos de vista. Esto se debió principalmente a la iluminación variable sobre la cara lo que dificultó un acoplamiento perfecto. También, se observó que la estandarización del tamaño de las imágenes capturadas en diferentes tamaños generaba una pérdida de información de la textura facial.

La Tabla 8.7 muestra la cantidad de caras capturadas en cada campo de vista. Se capturaron en promedio 43, 42 y 38 imágenes faciales respectivamente en los patrones de movimientos R_{1-3} . Las caras capturadas contenían diferentes distorsiones, sin embargo, la escala fue la distorsión que más dificultó el seguimiento. En la Figura 8.13 se muestra el desempeño *RES* del método de re-identificación facial propuesto sobre el escenario real LPA2. Como se puede observar, en las primeras dos predicciones se obtiene un valor $RES = 0.65$. En los siguientes 8 recorridos con el patrón de movimiento R_1 , el modelo aprende a distinguir el destino de la persona. Sin embargo, cuando la persona cambia su patrón de movimientos, a partir del recorrido 11, el valor de la predicción cae hasta $RES = 0.5$. Luego incrementa a $RES = 0.75$ en las siguientes dos predicciones. Posteriormente, el modelo aprende a distinguir el nuevo destino de la persona. Sin embargo, vuelve a disminuir cuando se enfrenta al nuevo patrón de movimiento R_3 donde alcanza un máximo de $RES = 0.6$. El desempeño de R_3 fue afectado por el gran nivel de feromonas acumulado por los patrones de movimientos anteriores, por lo que los 10 recorridos no fueron suficientes para aprender a distinguir bien el nuevo destino. Los patrones de movimientos evaluados tenían en común varios campos de vista.

Para el *operador* del sistema se desarrolló un módulo cuya interfaz se muestra en la Figura 8.14. El *operador* puede administrar personas, ya sean nuevos operadores o nuevos objetivos. También puede consultar los datos de re-identificaciones pasadas y el nivel de feromonas acumulado para cada *objetivo*. Además, puede elegir la zona de vigilancia y los recorridos disponibles para la re-identificación de un objetivo en específico. La escena de la re-identificación se muestra en la interfaz para que sea visualizada por el *operador*. Al finalizar un recorrido; el sistema muestra los campos de vista recorridos, el valor *RES* obtenidos y el resultado del establecimiento de la correspondencia entre los descriptores faciales generados en los diferentes seguimientos.

El resultado obtenido en la evaluación sobre un ambiente real confirma el funcionamiento del método propuesto en la práctica. En un ambiente real, suceden de forma concurrente varios factores que afectan el acoplamiento perfecto de las imágenes faciales. Principalmente, la escala, iluminación variable y el pose son los factores que más impactan en el desempeño del método de re-identificación facial. La iluminación variable puede ser removida mediante operaciones de procesamiento como el modelado de retina mientras que la escala puede ser abordado mediante una combinación de la

Tabla 8.4: Resultado de comparaciones de imágenes faciales promedio del patrón de movimientos 1. En verde las celdas de los valores donde no se pudo establecer correspondencia.

| Recorrido 1 | | | | | Recorrido 2 | | | | |
|-------------|---|-------|-------|-------|--------------|---|-------|-------|-------|
| - | a | c | d | b | - | a | c | d | b |
| a | - | 18.77 | 10.21 | 10.91 | a | - | 5.31 | 11.41 | 6.00 |
| c | - | - | 10.82 | 9.32 | c | - | - | 8.29 | 6.75 |
| d | - | - | - | 12.26 | d | - | - | - | 5.04 |
| b | - | - | - | - | b | - | - | - | - |
| Recorrido 3 | | | | | Recorrido 4 | | | | |
| - | a | c | d | b | - | a | c | d | b |
| a | - | 8.30 | 7.40 | 11.06 | a | - | 12.96 | 7.86 | 9.55 |
| c | - | - | 11.13 | 8.02 | c | - | - | 8.05 | 12.25 |
| d | - | - | - | 10.07 | d | - | - | - | 9.47 |
| b | - | - | - | - | b | - | - | - | - |
| Recorrido 5 | | | | | Recorrido 6 | | | | |
| - | a | c | d | b | - | a | c | d | b |
| a | - | 7.67 | 5.46 | 4.23 | a | - | 10.20 | 6.50 | 7.58 |
| c | - | - | 11.80 | 7.34 | c | - | - | 5.49 | 5.24 |
| d | - | - | - | 8.85 | d | - | - | - | 11.43 |
| b | - | - | - | - | b | - | - | - | - |
| Recorrido 7 | | | | | Recorrido 8 | | | | |
| - | a | c | d | b | - | a | c | d | b |
| a | - | 15.41 | 8.37 | 9.21 | a | - | 7.54 | 10.65 | 8.18 |
| c | - | - | 10.84 | 7.05 | c | - | - | 8.16 | 9.18 |
| d | - | - | - | 11.53 | d | - | - | - | 9.83 |
| b | - | - | - | - | b | - | - | - | - |
| Recorrido 9 | | | | | Recorrido 10 | | | | |
| - | a | c | d | b | - | a | c | d | b |
| a | - | 7.32 | 14.02 | 6.76 | a | - | 5.11 | 7.93 | 8.64 |
| c | - | - | 10.63 | 7.43 | c | - | - | 12.82 | 7.61 |
| d | - | - | - | 7.84 | d | - | - | - | 12.82 |
| b | - | - | - | - | b | - | - | - | - |

Tabla 8.5: Resultado de comparaciones de imágenes faciales promedio del patrón de movimientos 2. En verde las celdas de los valores donde no se pudo establecer correspondencia.

| Recorrido 11 | | | | | Recorrido 12 | | | | | | |
|--------------|---|-------|-------|-------|--------------|---|---|-------|-------|-------|-------|
| - | a | b | c | d | b | - | a | b | c | d | b |
| a | - | 10.98 | 13.16 | 9.86 | 6.22 | a | - | 12.22 | 8.94 | 6.10 | 9.47 |
| b | - | - | 14.32 | 7.61 | 16.30 | b | - | - | 11.69 | 6.06 | 14.11 |
| c | - | - | - | 12.16 | 14.06 | c | - | - | - | 7.73 | 11.79 |
| d | - | - | - | - | 6.22 | d | - | - | - | - | 9.08 |
| b | - | - | - | - | - | b | - | - | - | - | - |
| Recorrido 13 | | | | | Recorrido 14 | | | | | | |
| - | a | b | c | d | b | - | a | b | c | d | b |
| a | - | 8.07 | 6.05 | 5.17 | 7.16 | a | - | 8.98 | 5.08 | 7.62 | 9.74 |
| b | - | - | 10.33 | 7.60 | 16.79 | b | - | - | 12.69 | 9.31 | 15.64 |
| c | - | - | - | 14.68 | 10.13 | c | - | - | - | 10.11 | 11.76 |
| d | - | - | - | - | 7.74 | d | - | - | - | - | 8.72 |
| b | - | - | - | - | - | b | - | - | - | - | - |
| Recorrido 15 | | | | | Recorrido 16 | | | | | | |
| - | a | b | c | d | b | - | a | b | c | d | b |
| a | - | 9.95 | 8.78 | 15.15 | 16.46 | a | - | 11.19 | 10.80 | 8.43 | 6.68 |
| b | - | - | 9.67 | 7.97 | 13.37 | b | - | - | 9.64 | 5.35 | 15.44 |
| c | - | - | - | 7.61 | 8.74 | c | - | - | - | 7.72 | 10.21 |
| d | - | - | - | - | 12.44 | d | - | - | - | - | 6.48 |
| b | - | - | - | - | - | b | - | - | - | - | - |
| Recorrido 17 | | | | | Recorrido 18 | | | | | | |
| - | a | b | c | d | b | - | a | b | c | d | b |
| a | - | 7.29 | 4.93 | 5.82 | 12.83 | a | - | 8.29 | 13.70 | 7.15 | 8.23 |
| b | - | - | 8.95 | 13.71 | 13.38 | b | - | - | 8.46 | 6.45 | 4.82 |
| c | - | - | - | 9.50 | 8.17 | c | - | - | - | 10.28 | 10.91 |
| d | - | - | - | - | 9.48 | d | - | - | - | - | 8.17 |
| b | - | - | - | - | - | b | - | - | - | - | - |
| Recorrido 19 | | | | | Recorrido 20 | | | | | | |
| - | a | b | c | d | b | - | a | b | c | d | b |
| a | - | 6.17 | 11.18 | 9.53 | 9.06 | a | - | 9.41 | 13.25 | 5.33 | 8.71 |
| b | - | - | 12.42 | 9.99 | 17.54 | b | - | - | 15.54 | 9.85 | 10.71 |
| c | - | - | - | 15.96 | 9.96 | c | - | - | - | 10.03 | 13.61 |
| d | - | - | - | - | 9.06 | d | - | - | - | - | 11.39 |
| b | - | - | - | - | - | b | - | - | - | - | - |

detección y la propiedad de localización de los filtros de correlación.

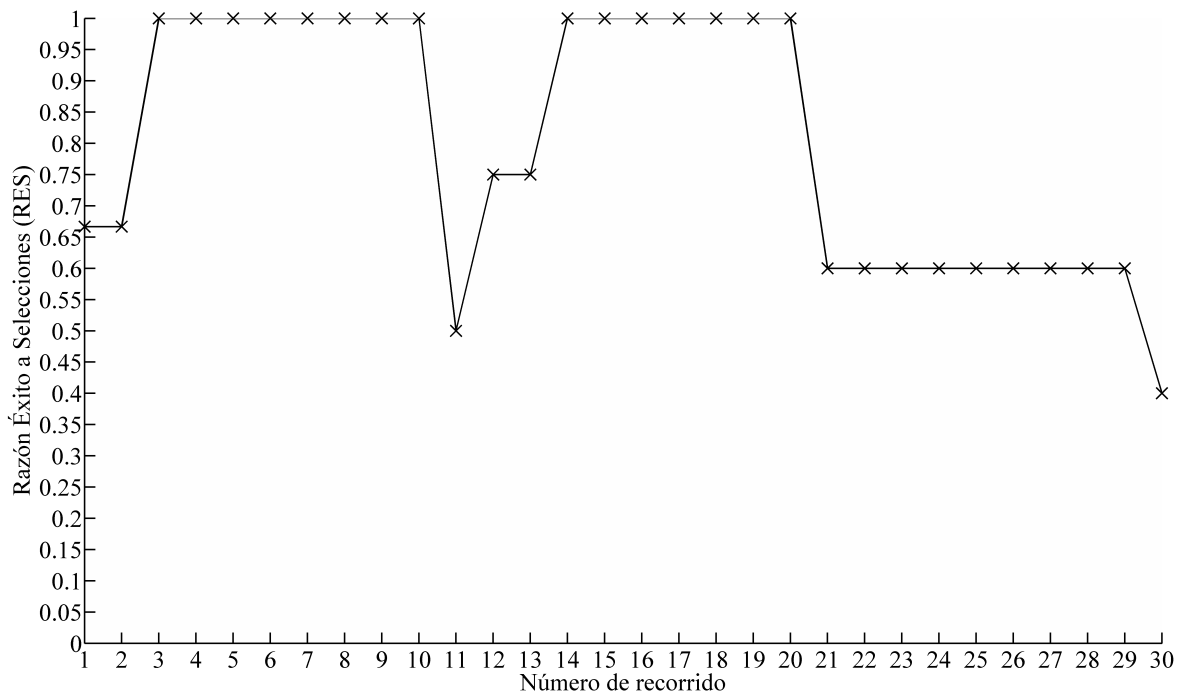


Figura 8.13: Desempeño RES del método de re-identificación facial propuesto sobre el escenario LPA2 de la Figura 8.10.



Figura 8.14: Interfaz gráfica de usuario para el operador del sistema de re-identificación facial.

8.3. Conclusiones

En este capítulo se ha propuesto un método de re-identificación facial basado en filtros de correlación y con predicción del campo de vista destino. Dado una red de

cámaras, con sus respectivas entradas y salidas, el proceso de re-identificación es como sigue. Se detecta e identifica una imagen facial en el campo de vista i . Luego la persona es seguida y se capturan sus muestras faciales. Cuando la persona abandona el campo de vista i , se construye su descriptor facial e inmediatamente se estima el campo de vista j destino. Este proceso se repite hasta que la persona abandone la zona vigilada por la red de cámaras. Los resultados obtenidos indican lo siguiente. Primero, los filtros de correlación diseñados específicamente para el reconocimiento como por ejemplo MACH, UOTSDF, SDF no lineal y UNCF son adecuados para ser usados como descriptores faciales. Segundo, los principios de los algoritmos de hormigas pueden ser usados en la predicción del campo de vista destino con una eficiencia superior al 96.15% en simulación y un 80% en la práctica. Tercero, la selección automática del conjunto de entrenamiento T permite producir en el registro facial un filtro robusto para identificar a la persona bajo condiciones reales. Finalmente, el método propuesto permite el desarrollo de sistemas de re-identificación facial para vídeo vigilancia con inteligencia computacional.

Tabla 8.6: Resultado de comparaciones de imágenes faciales promedio del patrón de movimientos 3. En verde las celdas de los valores donde no se pudo establecer correspondencia.

| Recorrido 21 | | | | | | Recorrido 22 | | | | | | | |
|--------------|---|-------|-------|-------|-------|--------------|---|---|-------|-------|-------|-------|-------|
| - | a | c | d | b | a | b | - | a | c | d | b | a | b |
| a | - | 6.82 | 6.22 | 8.36 | 9.82 | 6.82 | a | - | 4.68 | 5.59 | 8.30 | 11.27 | 9.09 |
| c | - | - | 15.42 | 6.78 | 10.16 | 9.47 | c | - | - | 17.60 | 8.98 | 10.44 | 8.29 |
| d | - | - | - | 9.47 | 9.96 | 9.32 | d | - | - | - | 11.68 | 14.63 | 7.18 |
| b | - | - | - | - | 7.76 | 10.53 | b | - | - | - | - | 12.56 | 13.96 |
| a | - | - | - | - | - | 7.94 | a | - | - | - | - | - | 8.71 |
| b | - | - | - | - | - | - | b | - | - | - | - | - | - |
| Recorrido 23 | | | | | | Recorrido 24 | | | | | | | |
| - | a | c | d | b | a | b | - | a | c | d | b | a | b |
| a | - | 15.39 | 13.52 | 8.45 | 10.16 | 6.21 | a | - | 12.19 | 9.56 | 8.67 | 16.68 | 10.70 |
| c | - | - | 14.75 | 10.91 | 10.28 | 8.11 | c | - | - | 13.64 | 13.37 | 10.40 | 8.46 |
| d | - | - | - | 11.69 | 10.29 | 7.27 | d | - | - | - | 12.65 | 12.89 | 6.19 |
| b | - | - | - | - | 10.85 | 7.14 | b | - | - | - | - | 9.65 | 11.46 |
| a | - | - | - | - | - | 9.11 | a | - | - | - | - | - | 7.56 |
| b | - | - | - | - | - | - | b | - | - | - | - | - | - |
| Recorrido 25 | | | | | | Recorrido 26 | | | | | | | |
| - | a | c | d | b | a | b | - | a | c | d | b | a | b |
| a | - | 11.75 | 17.43 | 8.01 | 17.82 | 9.60 | a | - | 9.18 | 6.09 | 6.32 | 9.39 | 4.22 |
| c | - | - | 11.67 | 7.14 | 9.90 | 6.53 | c | - | - | 10.64 | 14.32 | 10.30 | 8.63 |
| d | - | - | - | 9.79 | 11.51 | 11.18 | d | - | - | - | 9.84 | 8.40 | 6.26 |
| b | - | - | - | - | 6.70 | 12.36 | b | - | - | - | - | 8.37 | 10.38 |
| a | - | - | - | - | - | 7.44 | a | - | - | - | - | - | 8.21 |
| b | - | - | - | - | - | - | b | - | - | - | - | - | - |
| Recorrido 27 | | | | | | Recorrido 28 | | | | | | | |
| - | a | c | d | b | a | b | - | a | c | d | b | a | b |
| a | - | 6.07 | 6.84 | 8.88 | 4.67 | 5.10 | a | - | 8.31 | 5.56 | 6.03 | 11.67 | 5.19 |
| c | - | - | 13.47 | 10.93 | 9.26 | 5.23 | c | - | - | 13.48 | 9.24 | 12.14 | 8.58 |
| d | - | - | - | 11.19 | 8.70 | 5.47 | d | - | - | - | 7.22 | 10.49 | 9.44 |
| b | - | - | - | - | 9.19 | 6.75 | b | - | - | - | - | 11.01 | 16.35 |
| a | - | - | - | - | - | 6.67 | a | - | - | - | - | - | 11.52 |
| b | - | - | - | - | - | - | b | - | - | - | - | - | - |
| Recorrido 29 | | | | | | Recorrido 30 | | | | | | | |
| - | a | c | d | b | a | b | - | a | c | d | b | a | b |
| a | - | 4.92 | 6.83 | 4.14 | 7.18 | 6.90 | a | - | 12.65 | 11.49 | 11.49 | 76.44 | 8.08 |
| c | - | - | 12.28 | 8.92 | 8.72 | 6.35 | c | - | - | 13.28 | 13.87 | 11.82 | 8.66 |
| d | - | - | - | 5.68 | 9.98 | 6.25 | d | - | - | - | 6.05 | 12.12 | 6.12 |
| b | - | - | - | - | 8.86 | 16.03 | b | - | - | - | - | 11.06 | 14.71 |
| a | - | - | - | - | - | 9.53 | a | - | - | - | - | - | 8.36 |
| b | - | - | - | - | - | - | b | - | - | - | - | - | - |

Tabla 8.7: Cantidad de caras capturadas en cada seguimiento.

| CDV | Número de recorrido para R'_1 | | | | | | | | | |
|-----|---------------------------------|-----|----|----|----|----|----|-----|----|----|
| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
| a | 57 | 171 | 67 | 33 | 60 | 25 | 37 | 34 | 35 | 61 |
| c | 111 | 5 | 43 | 51 | 61 | 41 | 35 | 220 | 16 | 11 |
| d | 44 | 52 | 45 | 76 | 14 | 25 | 46 | 51 | 42 | 43 |
| b | 61 | 18 | 59 | 66 | 81 | 24 | 54 | 57 | 37 | 40 |
| CDV | Número de recorrido para R'_2 | | | | | | | | | |
| | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 |
| a | 36 | 44 | 36 | 59 | 76 | 94 | 54 | 44 | 52 | 56 |
| b | 64 | 64 | 29 | 28 | 50 | 60 | 26 | 42 | 55 | 27 |
| c | 64 | 17 | 65 | 37 | 37 | 80 | 11 | 93 | 69 | 50 |
| d | 10 | 21 | 18 | 65 | 51 | 61 | 13 | 28 | 40 | 28 |
| b | 13 | 47 | 66 | 44 | 58 | 38 | 65 | 80 | 39 | 85 |
| CDV | Número de recorrido para R'_3 | | | | | | | | | |
| | 21 | 22 | 23 | 24 | 25 | 26 | 27 | 28 | 29 | 30 |
| a | 64 | 30 | 52 | 38 | 52 | 47 | 45 | 65 | 61 | 58 |
| c | 26 | 15 | 45 | 42 | 49 | 58 | 12 | 32 | 15 | 34 |
| d | 35 | 42 | 29 | 26 | 42 | 36 | 26 | 15 | 44 | 41 |
| b | 36 | 79 | 45 | 48 | 46 | 49 | 49 | 52 | 34 | 31 |
| a | 9 | 33 | 70 | 45 | 45 | 31 | 24 | 52 | 46 | 58 |
| b | 37 | 62 | 18 | 29 | 25 | 28 | 14 | 24 | 58 | 58 |

Capítulo 9

Conclusiones

La re-identificación de personas es una herramienta muy importante para la vídeo vigilancia automatizada e inteligente. Esta tesis propone varios algoritmos de procesamiento facial que en su conjunto permiten re-identificar a una persona en una red de cámaras. Se usaron como descriptores faciales los filtros de correlación permitiendo que el establecimiento de la correspondencia se simplifique a un proceso de acoplamiento.

En la primera parte de esta tesis se propuso un algoritmo de detección facial basado en filtros de correlación. La escala es una variación muy común en imágenes no restringidas. Este problema fue abordado con un banco de filtros de diferentes tamaños. Debido a que el escalamiento genera pérdida de información, el rango de dimensiones fue determinado experimentalmente manteniendo el compromiso entre la capacidad de reconocimiento y de discriminación. El método fue evaluado sobre conjuntos de prueba con fondo complejo, iluminación variable, pose, rotación y expresiones. Se obtuvieron resultados muy competitivos comparado con métodos también basado en filtros de correlación.

En la segunda parte se propuso un algoritmo de seguimiento facial. Uno de los retos en el seguimiento es cómo abordar el problema de escala y los cambios de apariencia facial. Estos problemas se abordaron mediante la propuesta de un algoritmo que usa un filtro de correlación que se entrena en línea y la ejecución esporádica de la detección facial. Esto permitió que cada que se detecte una cara objetivo, ésta es usada para actualizar el filtro para así adaptarse a los cambios de apariencia. La variación en escala es tratado mediante una ventana de seguimiento w que se adapta al tamaño de la cara. Este algoritmo fue evaluado sobre vídeos capturados en un ambiente no restringido, obteniendo resultados que confirman su buen desempeño en la práctica.

En la tercera parte, se propuso una metodología basada en optimización numérica para la selección automática del conjunto de entrenamiento. La metodología propuesta usa los principales mecanismos de los algoritmos de optimización para probar diferentes combinaciones de imágenes y seleccionar el mejor de ellos como conjunto de entrenamiento. La metodología propuesta fue evaluada sobre dos bases de datos, cuatro algoritmos de optimización y tres filtros de correlación. Los resultados obtenidos en el reconocimiento facial indican que el método es capaz de seleccionar las imágenes más

representativa para producir filtros óptimos en términos de agudeza del pico, tolerancia a distorsiones y cantidad de imágenes usadas.

En la cuarta parte se propuso un modelo de predicción del campo de vista destino. En los sistemas de re-identificación convencionales, cuando una persona abandona algún campo de vista la buscan en los restantes campos de vista vecinos. Esto involucra un alto consumo de recursos y resulta poco práctico. Para resolver este problema, se propone un modelo que en base al patrón de movimiento de la persona, predice el campo de vista destino. Este modelo se basa en los principios de la evaporación y actualización de los algoritmos de hormigas para aprender a distinguir un destino y adaptarse a los cambios de patrones de movimientos.

Finalmente, en la quinta parte se integran todos los algoritmos y métodos mencionados arriba en un método de re-identificación facial. Una de las ventajas del método propuesto es que no requiere de la interacción de un operador en el proceso de re-identificación. El operador solo interviene en la administración de personas nuevas y el mantenimiento en operación del sistema.

9.1. Trabajo futuro

Aunque con lo propuesto en esta tesis se ha logrado mucho, aún hay muchas cosas por hacer para mejorar el desempeño del método de re-identificación facial propuesto. A continuación se describen brevemente algunos puntos como trabajo futuro.

- Dentro de la literatura revisada en el Capítulo 2 no se encontró ninguna metodología para validar los métodos de re-identificación. Tampoco se encontró disponible algún *benchmark* que pudiera ser usado para evaluar la calidad o el nivel de los métodos existentes. Por lo tanto, se propone como trabajo futuro la creación de una metodología y un *benchmark* para evaluar y validar la efectividad de los métodos existentes y nuevos.
- Paralelización de los módulos de procesamiento facial. La re-identificación de personas es una tarea que requiere del procesamiento concurrente de varias imágenes. Por esta razón, es necesario la adecuación de los algoritmos propuesto para su implementación paralela.
- El modelo de predicción propuesto requiere de muchos recorridos para aprender un nuevo patrón de movimientos cuando entre los patrones existen varios campos de vista en común. Esto requiere que en un futuro se perfeccione el modelo para minimizar la cantidad de recorridos necesarios.

Apéndice A

Predicción del campo de vista destino

En este capítulo se propone un modelo matemático el cual explota los principios de los rastros de feromonas en colonias de hormigas para predecir el CDV destino de una persona. Este modelo se desarrolló con el propósito de usarlo en un algoritmo de re-identificación facial para minimizar la cantidad de recursos computacionales y humanos requeridos. Los resultados numéricos muestran que el modelo predice el destino con una buena probabilidad de éxito, la cual mejora cuando se incrementa la cantidad de recorridos que realiza las personas en la red de cámaras.

A.1. Trabajos relacionados

El ser humano es un ser de hábitos. Por ejemplo, para llegar a un lugar frecuentado, generalmente sigue una ruta previamente aprendida. Este comportamiento permite analizar los patrones de movimientos de una persona para predecir su destino, y así mejorar la experiencia del usuario en aplicaciones computacionales. Los actuales enfoques para predecir el destino están basados en dispositivos móviles (teléfonos inteligentes, tabletas, dispositivos vestibles, entre otros) que cuentan con algún sistema de posicionamiento geográfico que permiten registrar los movimiento de la persona como una secuencia de ubicaciones sobre un espacio geográfico.

En la literatura se han propuesto varios enfoques para predecir el destino y/o la ruta de una persona. En (Kostov, Ozawa, Yoshioka, y Kudoh, 2005) se presenta un método de predicción de destino de viaje. Primero, se almacena la información histórica de viajes del usuario tales como ubicación, fecha y hora obtenidos mediante un navegador para carros con GPS. Luego, con la información histórica se determinan las categorías día/hora que tienen correlación con los destinos del usuario. Finalmente, con esas categorías se predice el destino. También, en (Alvarez-Garcia, Ortega, Gonzalez-Abril, y Velasco, 2010) se propone un sistema de predicción de destino de viaje basado en la generación de cadenas de un Modelo de Markov Oculto (CMO) de registros de GPS con

ubicaciones actuales. Los trabajos en (Tanaka, Kishino, Terada, y Nishio, 2009; Krumm, 2006) proponen métodos basados en el contexto y trayectoria de manejo medido por sistemas de navegación para predecir automáticamente el destino de un automovilista. Además, (Krumm, 2006) apoya su predicción en la intuición común de que los choferes tienden a escoger rutas eficientes. En (Filev, Tseng, Kristinsson, y McGee, 2011) se trata el problema de aprendizaje de ubicaciones de señales de alto para la predicción de destinos para vehículos. Este enfoque está basado en datos de tiempo real y en un modelo de aprendizaje de decisión que combina modelos Fuzzy y Markov. En (Davami y Sukthankar, 2012) se propone un mecanismo para la actualización en línea de modelos de predicción del destino de un usuario específico. Mientras que (Nadembega, Taleb, y Hafid, 2012) propone un método para predecir el destino de un usuario mediante la filtración de patrones históricos de movimientos y también el uso de conocimiento contextual. En (Nakahara y Murakami, 2012) se propone un método para identificar los patrones de movimiento del usuario a través de registros de posiciones no periódicas mediante GPS. Con estos patrones de comportamiento se predicen los destinos que ha de tomar la persona.

Una aplicación poco común es el que se propone (Scott y He, 2012), el cual usa el área de una ruta potencial para predecir un destino de compra. El método propuesto en (Xue y cols., 2013), descompone trayectorias históricas en subtrayectorias que comprenden dos ubicaciones de vecindad. Luego conecta las subtrayectorias en trayectorias sintetizadas. Un método para predecir destino mediante mensajes de Twitter que contengan información de ubicación se propone en (Shinmura, Zhu, Ota, y Fukazawa, 2014). La predicción se realiza mediante el análisis de la información contenida en los mensajes. En ocasiones, solo es necesario predecir la ruta para la definición del más eficiente en términos de optimización de recurso. En (J. Chen, Meng, Guo, Grumbach, y Sun, 2006) se propone el uso de grafos de autómatas celulares para predecir trayectorias. Mientras que (Eisner, Funke, Herbsty, Spillnery, y Storandt, 2011) propone usar datos imprecisos de ubicaciones medidos mediante GSM para la misma tarea. Por otra parte, la predicción tanto del destino como de la trayectoria mejora considerablemente el desempeño de los algoritmos de predicción. Un enfoque para predecir tanto el destino como la ruta de una persona es propuesto en (L. Chen, Lv, y Chen, 2010). Para este fin, se recolecta información de movimientos mediante un dispositivo GPS y luego los datos se clasifican en pares origen-destino. De estos patrones, se construye un árbol de patrones de movimientos para hacer la predicción.

Es fácil notar que la mayoría de los algoritmos reportados en la literatura para la predicción del destino y/o ruta dependen de un dispositivo de posicionamiento geográfico. Sin embargo, resulta muy costoso o simplemente no es posible implementarlos en todas las áreas. Por ejemplo, en lugares donde no es posible portar objetos por cuestiones de seguridad o higiene. Además, al basarse en dispositivos es posible la suplantación de identidad, pérdida, daños, inhabilitación por falta de energía, entre otros.

Aunque los algoritmos propuestos en la literatura son capaces de predecir tanto el destino como la trayectoria de una persona; la predicción robusta, no intrusiva y consciente de la identidad de la persona sin el uso de un dispositivo externo es un

problema que permanece abierto. Esta predicción debe considerar aspectos importantes tales como la exactitud de la medición de la localización, exploración de rutas alternas, identidad de la persona para evitar suplantaciones, entre otros.

A.2. Principios teóricos del algoritmo de hormigas

Las hormigas exhiben comportamientos sociales que han llamado la atención de los científicos computacionales. Uno de los principales comportamientos que los investigadores han estudiado, es su habilidad para encontrar la ruta óptima entre su nido y una fuente de comida. Esto lo logran mediante la comunicación basada en feromonas; una sustancia química y olorosa que segregan las hormigas. Este patrón de comportamiento es el que ha inspirado a los científicos a desarrollar algoritmos, conocidos como algoritmos de hormigas, para la solución de problemas de optimización (Dorigo y Stützle, 2004).

El campo de los algoritmos de hormigas estudia modelos derivados de la observación del comportamiento de hormigas reales. Las hormigas coordinan sus actividades mediante la *stigmergy*, una forma de comunicación indirecta mediada por modificaciones del ambiente. Por ejemplo, una hormiga forrajera deposita feromonas sobre el suelo durante un recorrido de su nido a una fuente de comida. Este rastro de feromonas incrementa la probabilidad de que otras hormigas sigan la misma ruta para llegar a la fuente de comida descubierta. La idea principal de este tipo de algoritmos es usar una forma de *stigmergy artificial* para coordinar sociedades de agentes artificiales.

Para comprender mejor los algoritmos de hormigas, veamos el siguiente experimento descrito en (Dorigo y Stützle, 2004). En el experimento se usó un puente doble conectando un nido de hormigas a una fuente de comida. Se realizaron varias pruebas variando la razón $r = \frac{l_l}{l_c}$, donde l_l es la longitud de la rama más larga y l_c es la longitud de la más corta. En el primer experimento, el puente tenía dos ramas de longitud igual a uno, resultando $r = 1$. Inicialmente, las hormigas fueron libres de moverse entre el nido y la fuente de comida y el porcentaje de hormigas que escogió uno u otro de las dos ramas fue observado sobre el tiempo. En la fase inicial sucedió una selección aleatoria, pero eventualmente todas las hormigas usaron la misma rama. Esto se debió a que cuando iniciaron las pruebas, las dos ramas no tenían feromonas por lo que las hormigas no tenían preferencias y seleccionaban con la misma probabilidad cualquiera de las ramas. Las hormigas al caminar depositan feromonas, una gran cantidad de hormigas sobre una rama resulta en una gran cantidad de feromonas sobre dicha rama, esta gran cantidad de feromonas a su vez estimulan a más hormigas a seleccionar dicha rama otra vez, y así hasta que finalmente, las hormigas convergen solo a una de las dos rutas.

En el segundo experimento, la longitud del razón entre las dos ramas fue establecida a $r = 2$. Esto indica que la longitud de la rama larga fue el doble que la longitud de la rama corta. En este caso, después de algún tiempo todas las hormigas seleccionaron solo la ruta más corta. Inicialmente las dos ramas les parecen idénticos a las hormigas por lo que la selección fue aleatoria. Algo importante que hay que recalcar de las feromonas es

que con el tiempo estas se evaporan. En la rama larga las hormigas tardaban más en su viaje, esto en conjunto con la evaporación de las feromonas, provoca una acumulación más lenta de feromonas. Por otra parte, en la rama más corta, debido a que las hormigas tardaban menos en su recorrido el nivel de feromonas incrementaba rápidamente. Por esta razón, pasado un tiempo el gran nivel de feromonas en la rama más corta sesgaba la decisión a su favor. Sin embargo, también se observó que un pequeño grupo de hormigas seguían tomando la rama más larga, lo cual puede interpretarse como una exploración de rutas.

En un experimento adicional, inicialmente solo se le ofreció a las hormigas la rama larga y después de 30 minutos se le ofreció una rama corta. Algo interesante que se notó es que la rama corta fue seleccionada esporádicamente y la colonia quedó atrapada en la rama más larga. Esto se debe a la gran concentración de feromonas sobre la rama larga y la lenta evaporación de feromonas. Una lenta evaporación de las feromonas permite a la colonia de hormigas olvidar rutas sub-óptimas de tal forma que puedan descubrir y aprender rutas nuevas.

Los principios de actualización y evaporación del nivel de feromonas es el que se explota en este trabajo para predecir el CDV destino de una persona que ha sido detectada y reconocida.

A.3. Modelo de predicción del CDV destino

Como se mencionó anteriormente, las hormigas coordinan sus actividades mediante la *stigmergy*. Las hormigas salen de su nido en busca de comida y vuelven a ella guiados por el olor de las feromonas que depositan durante su recorrido. Las feromonas al evaporarse lentamente permiten la exploración de nuevas rutas. Sin embargo, las personas tienen características diferentes a pesar de que también son seres sociales. La diferencia consiste principalmente en la capacidad de pensar que poseen las personas, la cual les permite hacer un análisis más exhaustivo de su ambiente y tomar la ruta más eficiente hacia su destino. Aunque es difícil saber la ruta que ha decidido tomar una persona, es posible predecirlo explotando la premisa de que el ser humano es un ser de hábitos. Cuando una persona ha decidido que una ruta es eficiente, en términos de sus necesidades, entonces es muy probable que en futuras ocasiones use esa misma ruta para ir a su destino. Cuando es necesario un cambio de ruta para llegar al mismo destino, lo hace hacia la siguiente ruta más eficiente.

Es importante recalcar tres aspectos de las personas y las hormigas. Primero, que las personas definen sus rutas mediante sus propias necesidades, aunque ésta decisión pueda ser influenciada por el comportamiento de otras personas o el ambiente. Segundo, la ruta para la entrada/salida a un espacio vigilado puede variar según las necesidades de la persona y el contexto en que se encuentra. Tercero, el rastro de feromonas es usado por las hormigas para encontrar la ruta más corta entre el nido y la fuente de comida. Por esta razón, los principios de los algoritmos de hormigas no se pueden aplicar directamente al problema de la predicción de destino de una persona. Por lo tanto, al

igual que en (Cornejo, Dornhaus, Lynch, y Nagpal, 2014; Caro y Dorigo, 1998), en esta tesis únicamente se usan los principios de la actualización y evaporación de las feromonas. Estos tres aspectos fueron considerados en el desarrollo del modelo que se presenta a continuación.

A.3.1. Predicción del campo de vista destino

El objetivo en este trabajo es definir un modelo que pueda ser usado por algoritmos de re-identificación facial para predecir el CDV destino de una persona. Con este objetivo en mente, considere una red de cámaras como un grafo conectado $G = (V, A)$, donde V es el conjunto de $N = |V|$ nodos (campos de vista) y A es el conjunto de aristas (dirigidos y no dirigidos) que conectan a los nodos. El arco formado por los nodos i y j es al que se le desea asignar un valor de peso como medida de la deseabilidad. Este peso, posteriormente es usado para probabilísticamente predecir el campo de vista destino. La Figura A.1 muestra un ejemplo de un grafo con $N = |V| = 16$ nodos, 17 arcos y una ruta seguida por una persona indicada por las flechas de color rojo.

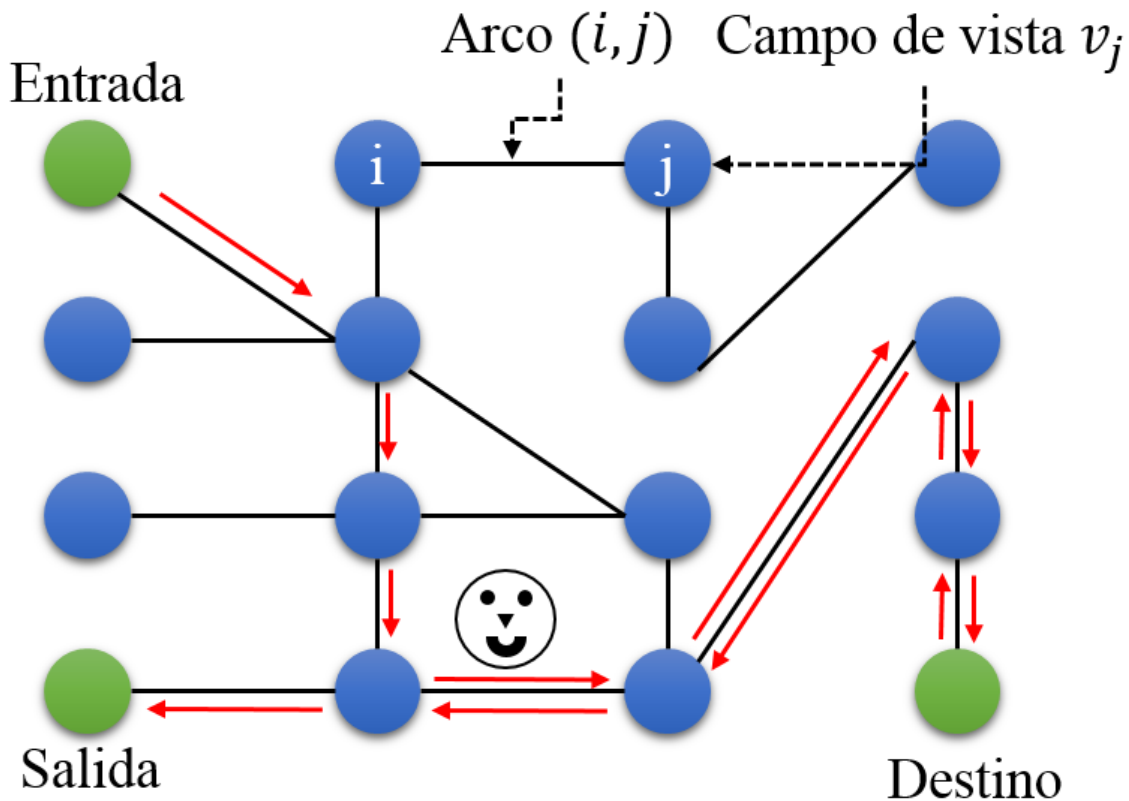


Figura A.1: Red de cámaras representadas como un grafo conectado $G = (V, A)$ con una ruta de un nodo entrada a un nodo destino y del nodo destino al nodo salida.

Estructura de datos

El modelo propuesto mantiene los niveles de feromonas en una matriz F_i asociados a la i -ésima persona. En la Figura A.2 se muestra una representación gráfica de una matriz de feromonas, donde el nivel de feromonas indica la deseabilidad de un CDV origen i hacia un CDV destino j . Una persona en el CDV i tienen cualquiera de los $|V_i|$ campos de vista vecinos como destino. Cada arco (i, j) se inicializa con un nivel de feromonas $F_{i,j} = \frac{1}{|V_i|}$ tal que:

$$\sum_{j \in V_i} F_{i,j} = 1. \quad (\text{A.1})$$

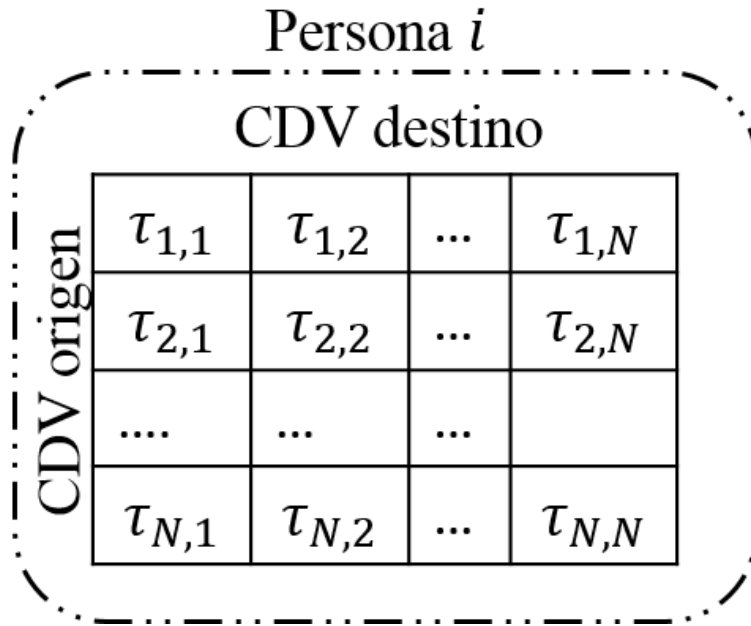


Figura A.2: Tabla de feromonas de la i -ésima persona. Esta estructura de datos contiene los niveles de feromonas del camino que une a los campos de vista i y j .

Modelo para la predicción del CDV destino

El CDV j el cual se espera sea el destino de la persona se calcula probabilísticamente usando los datos en F_i de cada persona. El modelo de predicción propuesto está dado por:

$$P(j) = \frac{F_{i,j}}{\sum_{l \in V_i} F_{i,l}}, \quad (\text{A.2})$$

donde $P(j)$ indica la probabilidad de que la persona vaya al CDV j cuando se encuentre en el CDV i . Cuando una persona se encuentra en el CDV i , su destino se calcula mediante la Ecuación A.2 de la siguiente manera. Primero, se calcula $P(j \in V_i)$. Segundo, se selecciona el CDV con mayor probabilidad como el campo de vista esperado CDV e , es decir, el CDV que el modelo determina como destino.

En la vida real, la persona decide por su propia cuenta hacia qué CDV dirigirse. Por lo tanto, es importante distinguir entre el campo de vista esperado e y el campo de vista seleccionado por la persona s . La actualización del nivel de feromonas $F_{i,s}$ en F_i está dado por:

$$F_{i,s} = F_{i,s} + \frac{F_{i,s}}{(\sum_{l \in V_i} F_{i,l})^m}. \quad (\text{A.3})$$

Cuando el CDV seleccionado sea diferente al CDV esperado, entonces se realiza una disminución del nivel de feromonas de la arista (i, e) . Esta penalización impulsa al modelo a explorar un nuevo CDV destino. Esta disminución está dado por:

$$F_{i,e} = F_{i,e} - \beta \frac{F_{i,e}}{(\sum_{l \in V_i} F_{i,l})^m}. \quad (\text{A.4})$$

La variable m en las Ecuaciones A.3 y A.4 es la cantidad de veces que se ha visitado en el mismo recorrido un CDV. Esta variable tiene la función de disminuir el efecto de los ciclos como se explica a continuación. Los ciclos en un grafo pueden concentrar una gran cantidad de feromonas sobre una vecindad y provocar predicciones fallidas. En el modelo propuesto no se considera la eliminación de ciclos, sin embargo, la variable m permite que el nivel de feromonas a depositar disminuya gradualmente en cada visita a un mismo nodo en un recorrido. Es decir, la primera vez que se visita el CDV i , el nivel de feromonas a depositar es igual a $\frac{1}{|V_i|}$. Posteriormente, cuando cuando se visite otra vez el CDV i , entonces la cantidad de feromonas a agregar en la actualización tiende a disminuir, tal y como lo muestra la línea punteada en la Figura A.3. Este simple mecanismo provee al modelo de predicción la capacidad de salirse de los ciclos. La Figura A.3 muestra el incremento del nivel de feromonas cada vez que se visite el mismo CDV (línea negra continua) en un recorrido, donde se puede observar que las primeras visitas, el crecimiento es rápido hasta que después llega a estabilizarse.

Cuando se termina un recorrido, se aplica una evaporación de feromonas a todo el grafo mediante:

$$F_{i,j} = (1 - \frac{1}{\rho^r}) F_{i,j}, \quad (\text{A.5})$$

donde ρ es el número de campos de vista visitados durante el recorrido y r la cantidad de recorridos en la zona vigilada. Un recorrido es el conjunto de campos de vista visitados por el objetivo. En los primeros recorridos, la evaporación es rápida debido a que el agente inicialmente realiza una exploración del grafo. Después de varios recorridos, el

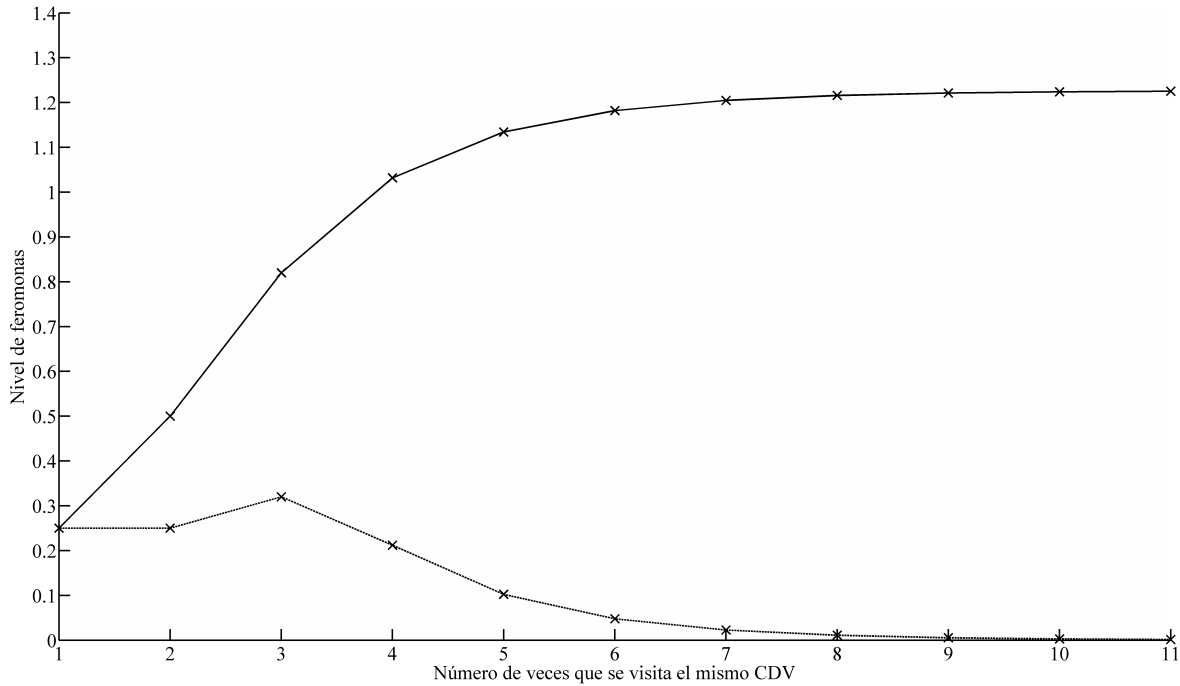


Figura A.3: Aplicación del modelo de actualización sobre la arista (i, j) cuando es visitado varias veces. La línea continua indica el nivel de feromonas actualizado en cada visita, mientras que la línea segmentada indica la cantidad de feromonas agregado en la actualización.

agente aprende a distinguir los campos de vista destinos por lo que la evaporación es lenta.

Con el fin de determinar la cantidad de recorridos que necesita el modelo para aprender a distinguir el CDV destino, el modelo de la Ecuación A.2 fue evaluado sobre vecindades de tamaño $t = \{2, 3, 4, 5, 6, 10, 20, 30, 100, 200\}$. Una red de cámaras se mapea a un grafo sobre el cual se ejecuta el modelo de predicción. Un ejemplo de este mapeo se muestra en la Figura A.4, donde los campos de vista CDV_i con borde amarillo son las entradas y los que tienen borde rojo las salidas. En un ambiente real puede existir restricciones que permiten a una persona circular de un campo de vista a otros en solo una dirección. Por esta razón, el grafo muestra aristas dirigidos y doblemente dirigidos. Del dado izquierdo, se muestra la zona de vigilancia y la ubicación de la cámara.

Para cada vecindad de tamaño t_i , se realizaron 1000 predicciones. Después de cada predicción se actualizó el nivel de feromonas (Ecuación A.3) y posteriormente se realizó la evaporación de feromonas (Ecuación A.5). Las Figuras [A.5, A.6, A.7, A.8, A.9] muestran el resultado obtenido. En dichas figuras, el número de recorrido desplegado fue truncado para una mejor visualización. De los resultados en las gráficas se deduce lo siguiente. Para vecindades de tamaño $t_i \leq 30$, el modelo requiere un promedio de 10 recorridos para aprender a distinguir bien el destino de la persona. Para vecindades

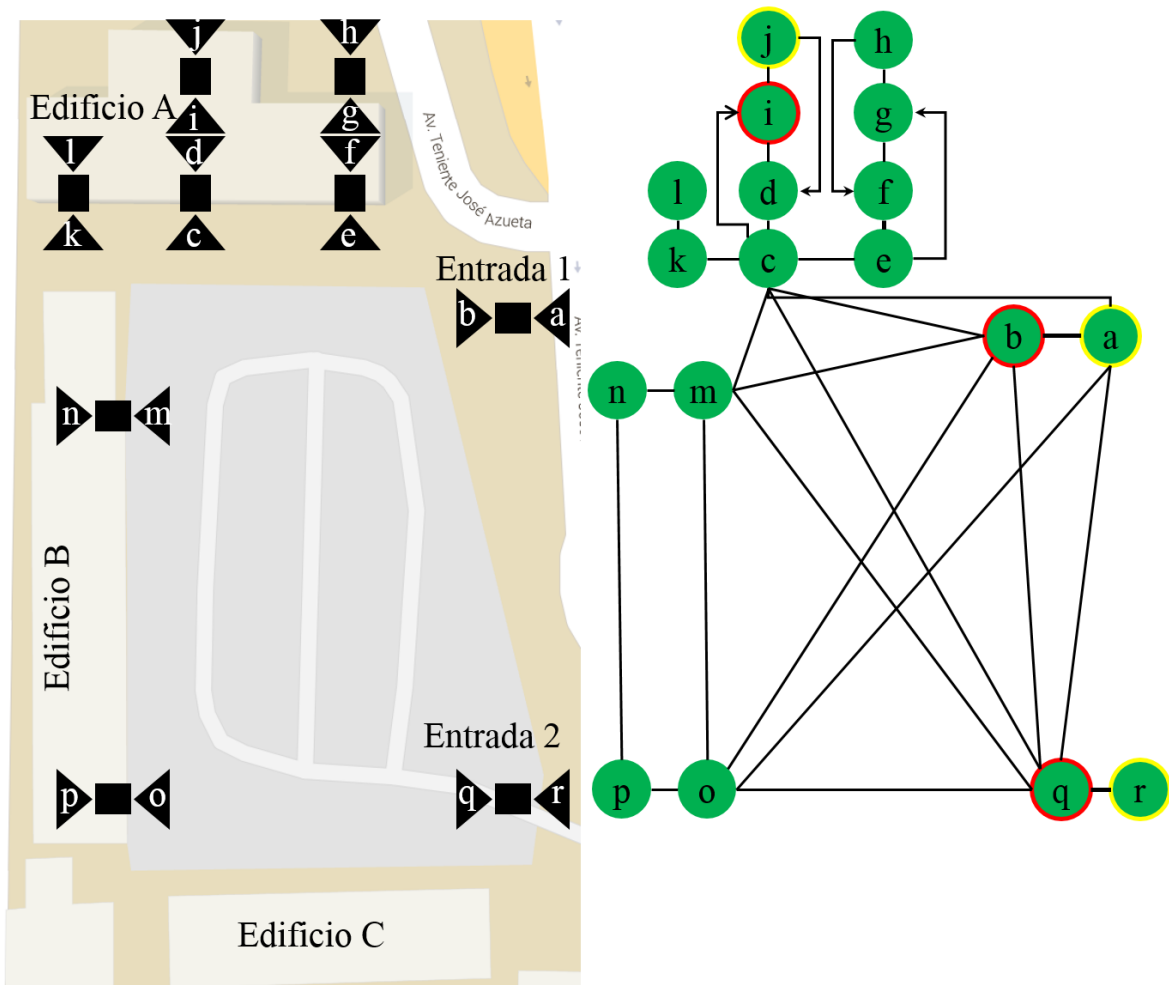


Figura A.4: Red de cámara y su correspondiente grafo.

de tamaño 100 y 200, el modelo aprende a distinguir el destino después de 38 y 110 predicciones, respectivamente. Esto debido a la gran cantidad de opciones disponibles, lo cual causa que la probabilidad por un CDV i incremente lentamente. También, las gráficas muestran que una vez que el algoritmo aprende a distinguir el destino, la probabilidad de los otros CDV's disminuye hasta estabilizarse en un valor mayor que cero. Tampoco la probabilidad del destino aprendido llega a uno, evitando así un acaparamiento y permitiendo la exploración de nuevos destino.

Con el fin de ver la importancia del uso de m y la penalización de las predicciones incorrectas de la Ecuación A.4, se experimentó sobre el grafo en la parte izquierda de la Figura A.10 con una persona. En la parte derecha de la misma figura se muestra un ejemplo de una cadena de los campos de vista recorridos, en el cual se indican los ciclos que contiene. El objetivo de este experimento fue obtener el número de veces en que se cayó en ciclos con a) las variables m y la penalización, b) sin la variable m y con penalización y c) sin la variable m y sin la penalización. El resultado obtenido con

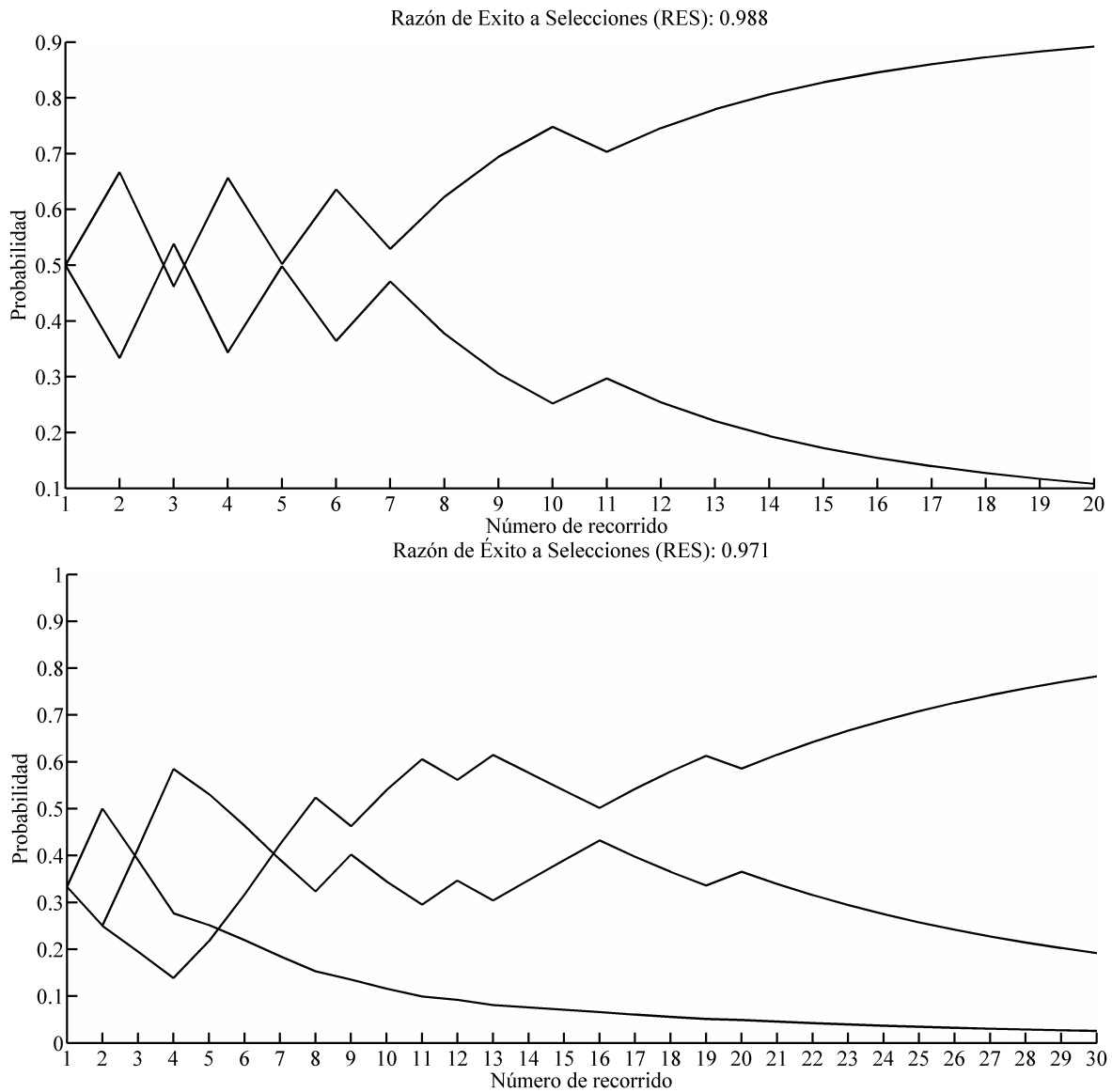


Figura A.5: Probabilidad de selección del CDV_i en 1000 predicciones para las vecindades de tamaño 2 y 3.

10000 recorridos fueron 90, 725039 y 2133 para a), b) y c), respectivamente. El mejor desempeño se obtiene usando m y la penalización de predicciones fallidas. Esto confirma lo que se mencionó previamente, el modelo de actualización no elimina los ciclos pero provee al modelo una forma de minimizar su ocurrencia.

A.3.2. Resultados experimentales del modelo de predicción

Aquí se presentan los resultados de dos experimentos que prueban la validez del modelo de predicción propuesto. El primero, muestra cómo en un simple recorrido se

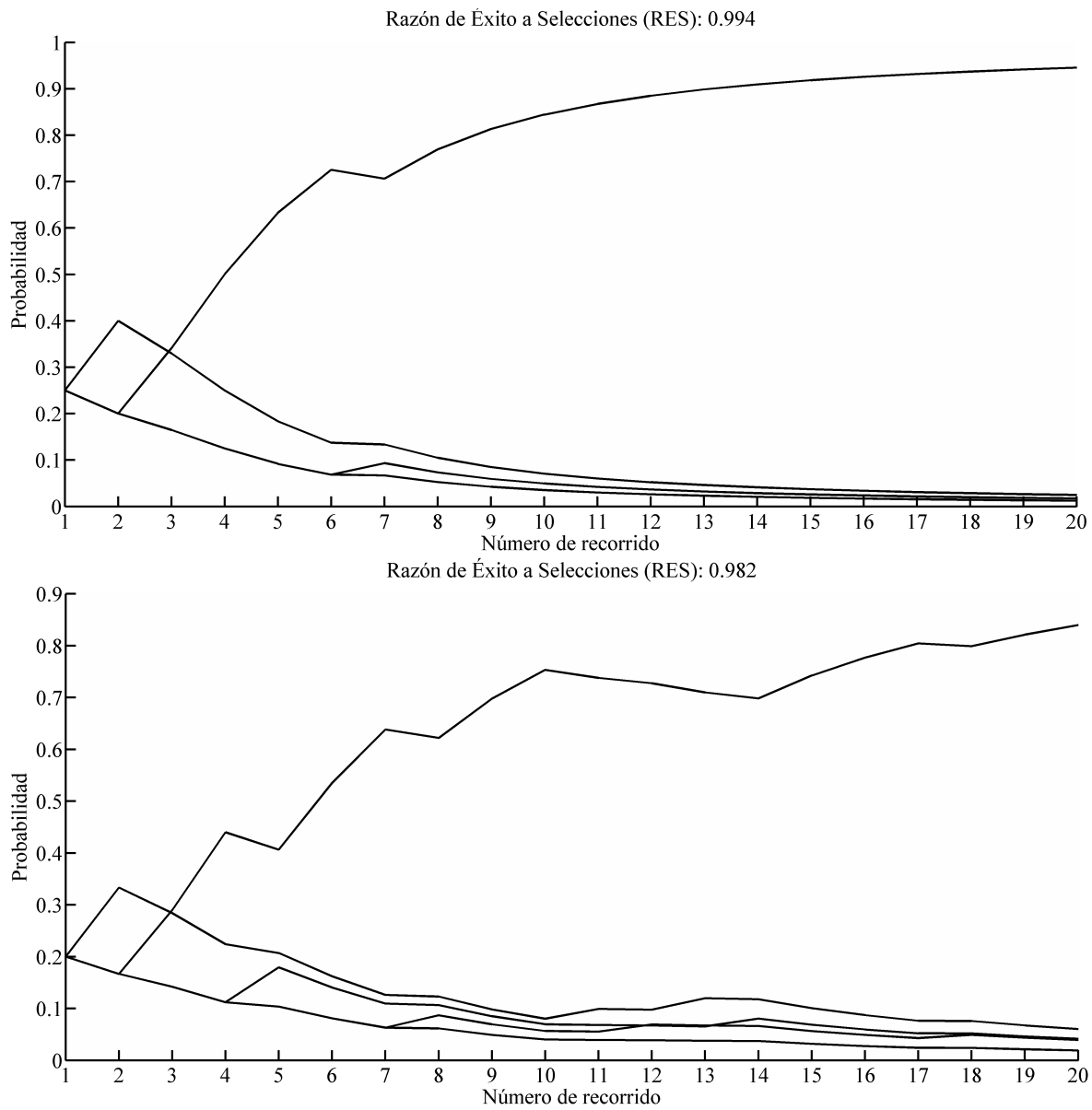


Figura A.6: Probabilidad de selección del CDV_i en 1000 predicciones para las vecindades de tamaño 4 y 5.

seleccionan los nodos destino hasta abandonar el área bajo observación, así como la actualización/evaporación de feromonas. El segundo experimento consiste de una simulación extensa mediante el método de Monte Carlo (J. S. Liu, 2001). Esta simulación extensa evalúa el modelo propuesto sobre varios grafos, desde simples hasta complejos en términos de la cantidad de nodos y la configuración de los enlaces entre ellos.

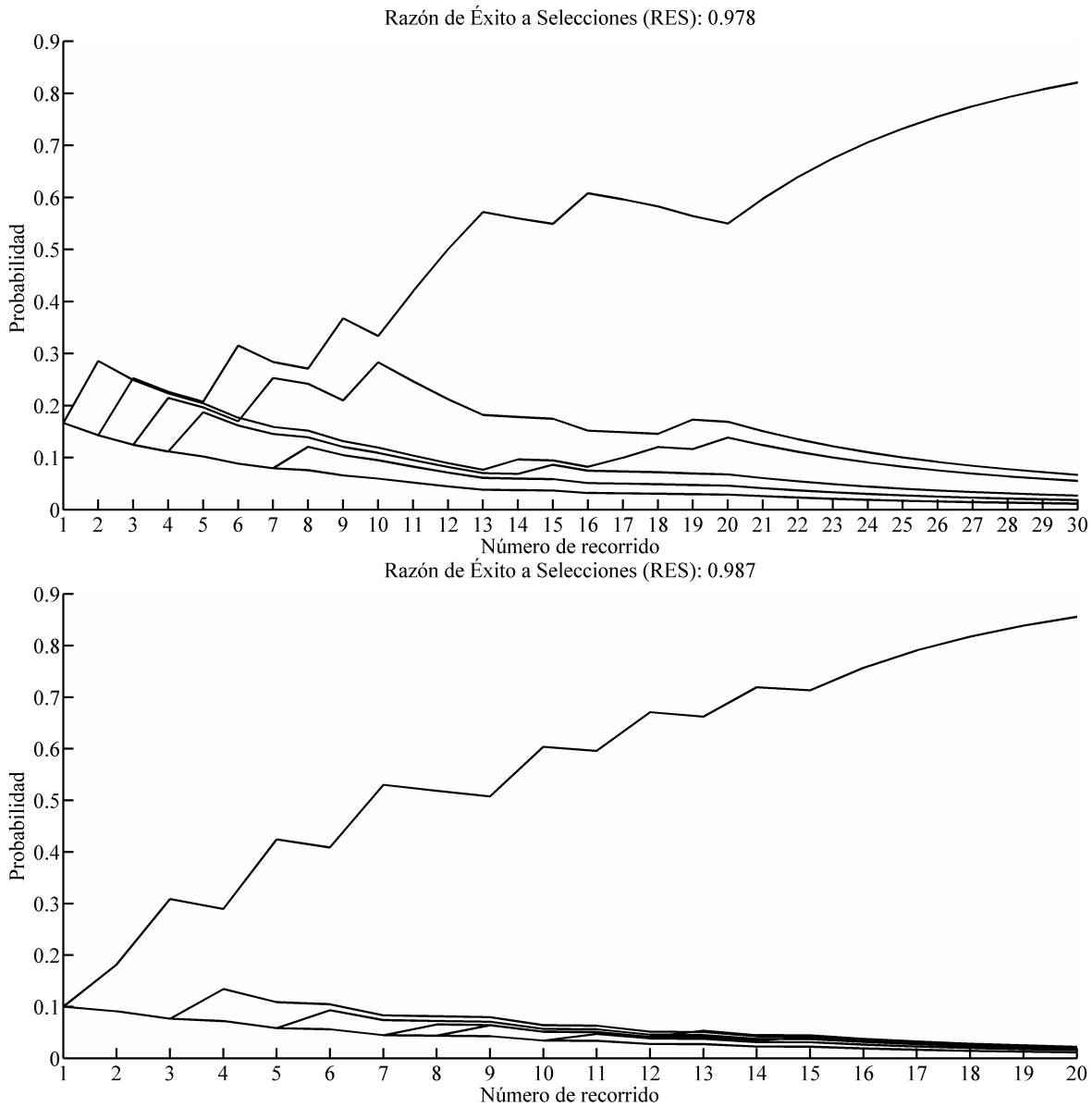


Figura A.7: Probabilidad de selección del CDV_i en 1000 predicciones para las vecindades de tamaño 6 y 10.

Experimento simple

Para ejemplificar el funcionamiento del modelo propuesto, considere una área vigilada por una red de cinco cámaras (Véase la Figura A.11). Los CDV's son identificadas con las letras $\{a, b, c, d, e\}$, los caminos existentes son $\{(a, b), (a, c), (b, c), (b, d), (c, a), (c, e), (e, d)\}$ y se usan nodos ocultos para las entradas/salidas las cuales son etiquetadas con ES.

La Figura A.12 muestra el funcionamiento del modelo propuesto para una persona que recorre los campos de vista $\{a, c, e, d\}$. Para el CDV i se muestra la probabilidad

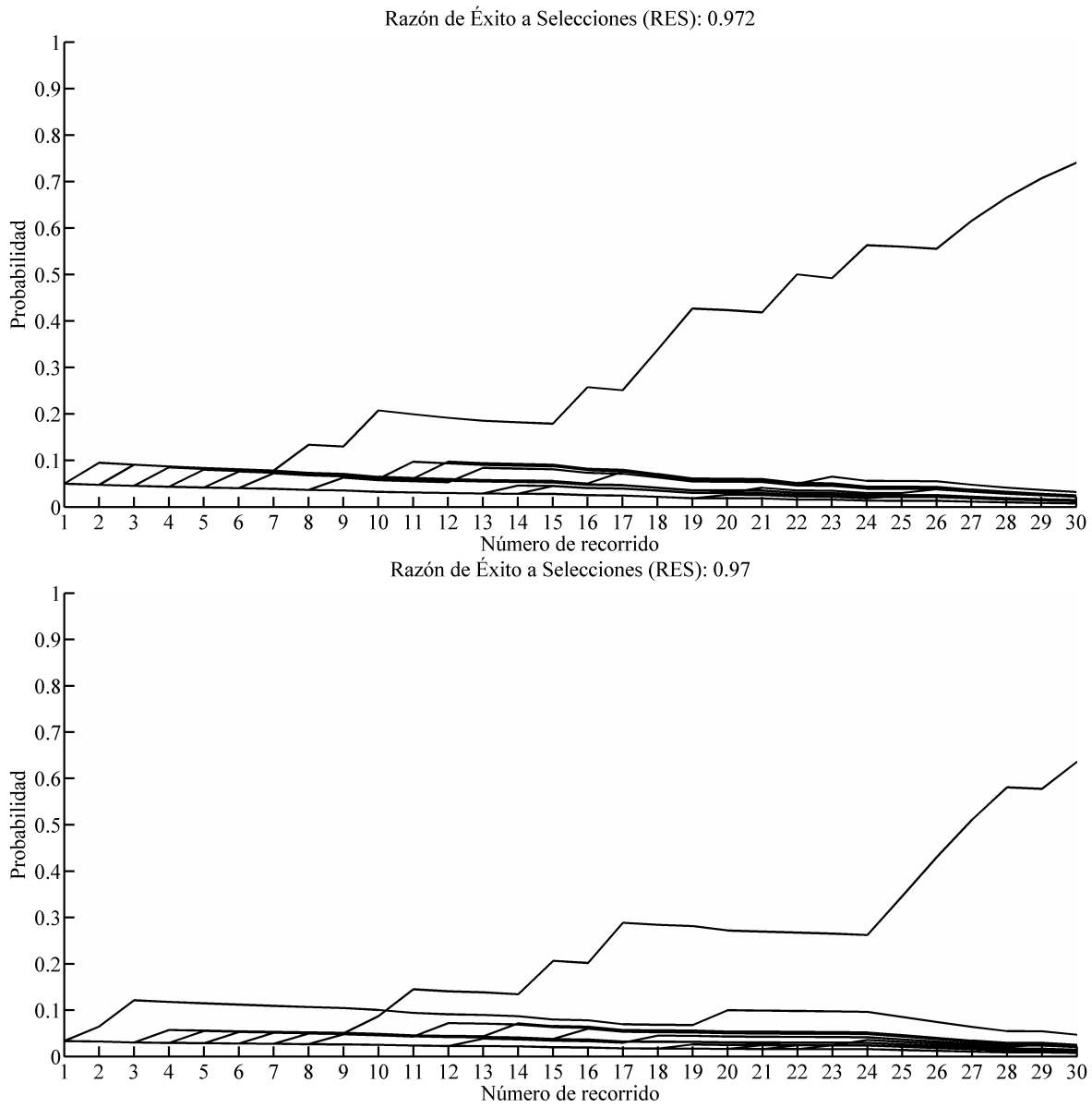


Figura A.8: Probabilidad de selección del CDV_i en 1000 predicciones para las vecindades de tamaño 20 y 30.

P de tomar el CDV j como destino, así como la actualización del nivel de feromonas $\tau_{i,j}$. El campo de vista a tiene como vecinos el conjunto $V_a = \{b, c, ES\}$. Cada nodo en V_a tiene la misma probabilidad de ser seleccionado, sin embargo, la persona toma como destino el nodo c . Por lo tanto, el nivel de feromonas del arco (a, c) se actualiza mediante la Ecuación A.3. Este proceso se repite hasta que la persona abandone la zona de vigilancia. Esto sucede cuando la persona llega al nodo g y toma como destino el nodo oculto ES .

Como se puede ver en la Ecuación A.5, la evaporación de feromonas depende de la

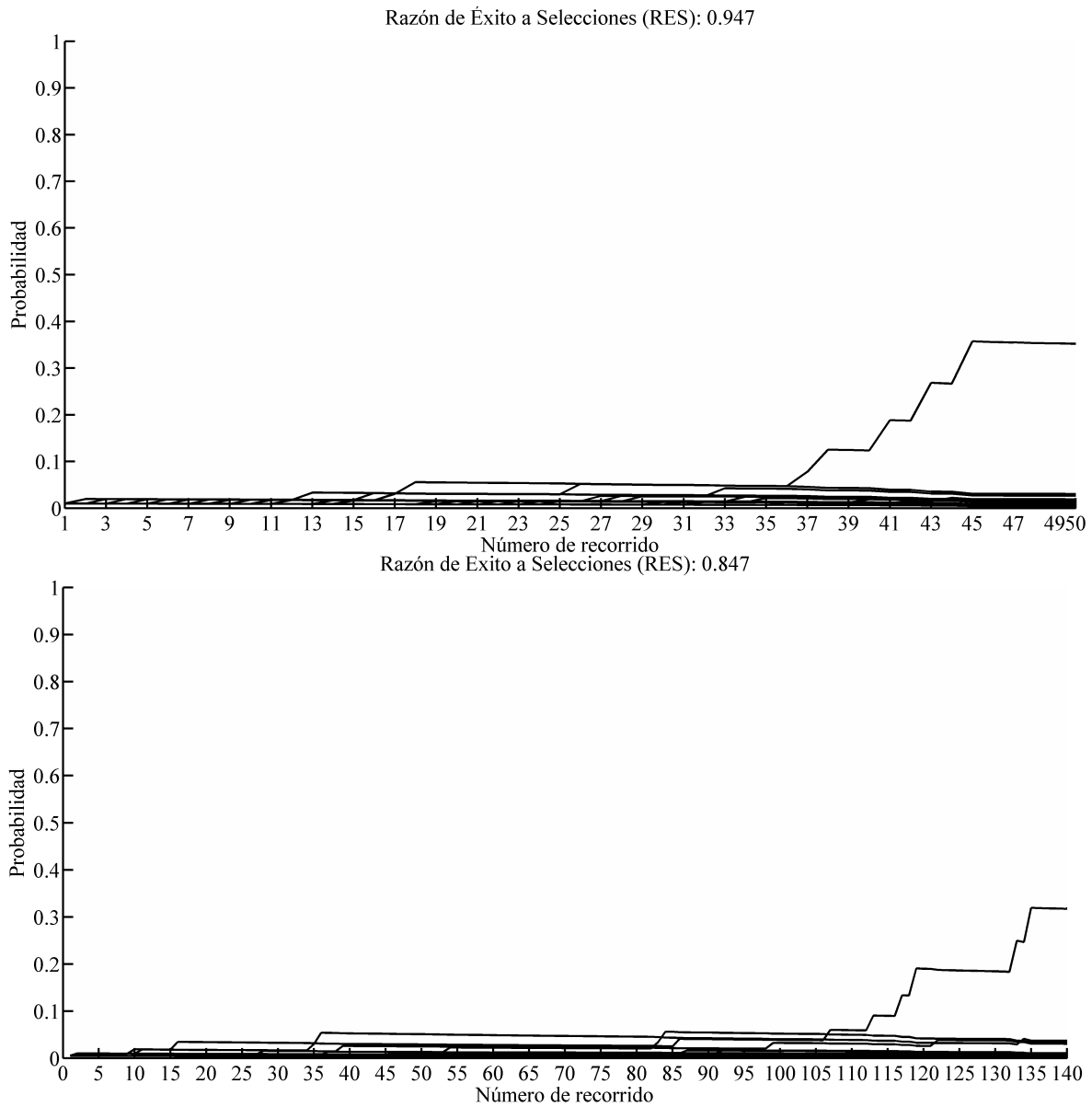


Figura A.9: Probabilidad de selección del CDV_i en 1000 predicciones para las vecindades de tamaño 100 y 200.

cantidad de CDV's cruzados ρ . Las Tablas [A.1, A.2, A.3] muestran respectivamente la F_i inicial, el actualizada durante el recorrido y el actualizado con la evaporación. Como se puede observar en la Tabla A.3, la evaporación es rápida debido a que es el primer recorrido que ha hecho la persona. Al inicio conviene que la evaporación sea rápida, pues es cuando la persona en el mundo real explora su ambiente por lo que aún no define bien su destino.

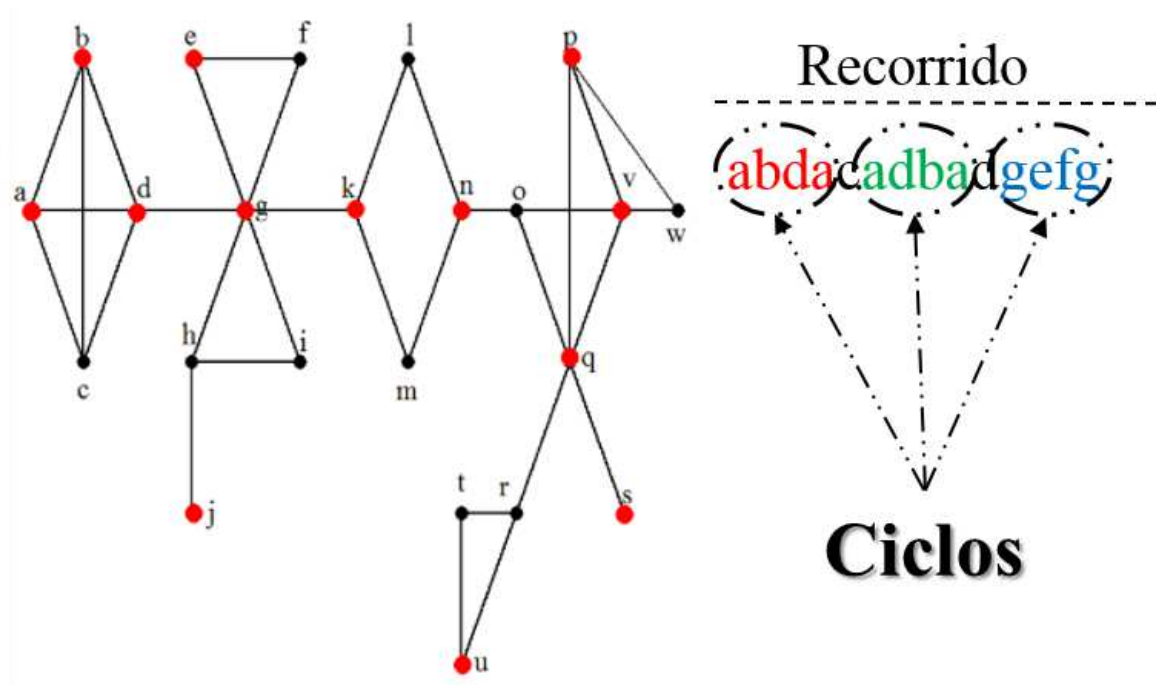


Figura A.10: En la parte izquierda se muestra el grafo de campos de vista con entradas y salidas de color rojo, mientras que en el lado derecho se muestra un recorrido con tres ciclos.

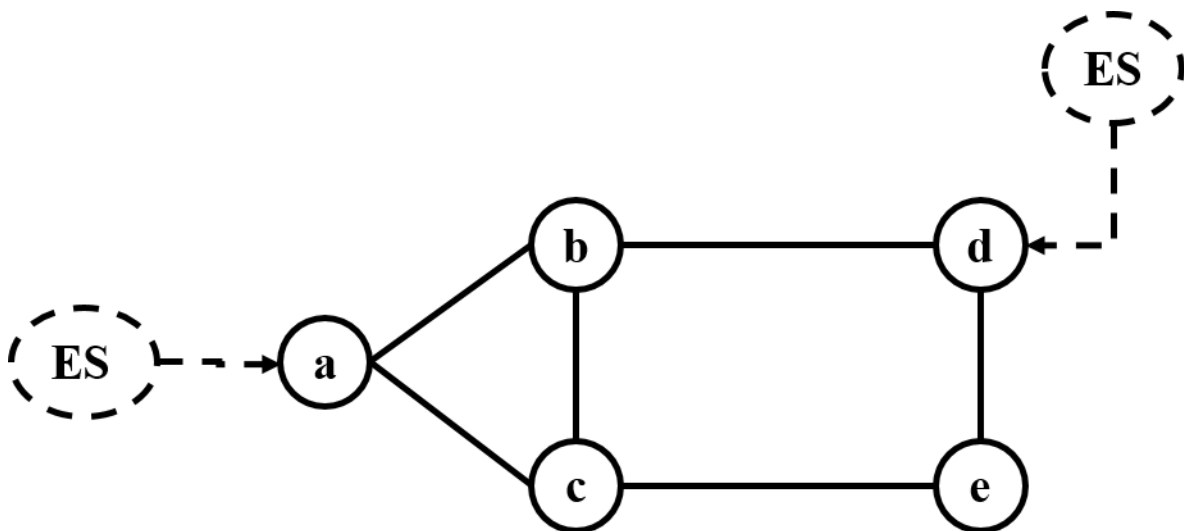


Figura A.11: Grafo una red de cinco cámaras.

Experimento extenso sobre diferentes grafos

La simulación por el método de Monte Carlo es una técnica que combina conceptos estadísticos (muestreo aleatorio) con la capacidad de los ordenadores para imitar, me-

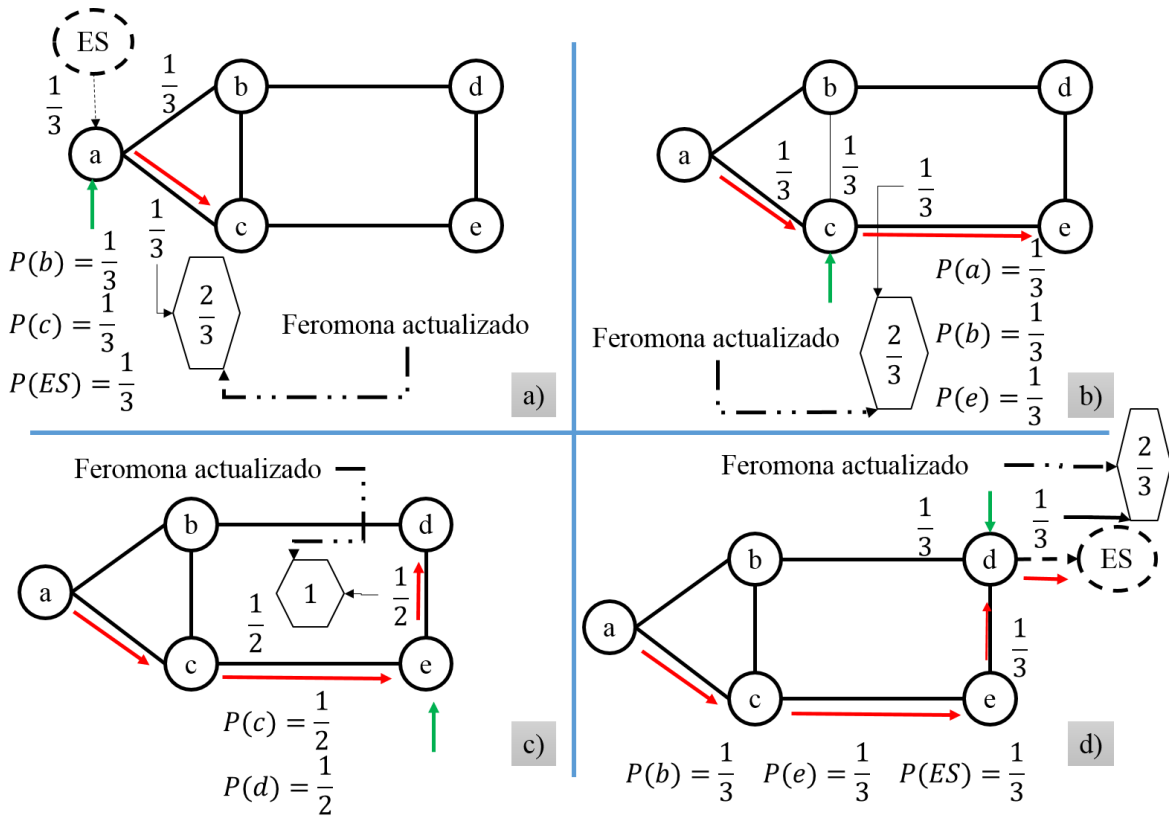


Figura A.12: Recorrido de una persona en una red de cámaras. Las flechas rojas indican la trayectoria tomada, mientras que las flechas verdes señalan el CDV en que se encuentra el agente.

Tabla A.1: Tabla de feromonas F_i inicial

| | a | b | c | d | e | ES |
|---|------|------|------|------|------|------|
| a | 0.00 | 0.33 | 0.33 | 0.00 | 0.00 | 0.33 |
| b | 0.33 | 0.00 | 0.33 | 0.33 | 0.00 | 0.00 |
| c | 0.33 | 0.33 | 0.00 | 0.00 | 0.33 | 0.00 |
| d | 0.00 | 0.33 | 0.00 | 0.00 | 0.33 | 0.33 |
| e | 0.00 | 0.00 | 0.50 | 0.50 | 0.00 | 0.00 |

diante modelos matemáticos, el comportamiento aleatorio de sistemas reales (J. S. Liu, 2001). Este método se usó para simular el comportamiento del método propuesto con el fin de responder a las siguientes preguntas:

1. ¿Cuál es el porcentaje promedio de éxito del modelo propuesto?.
2. ¿Cuál fue recorrido seleccionado por la persona?

Tabla A.2: Tabla de feromonas F_i actualizado después del un recorrido

| | a | b | c | d | e | ES |
|---|------|------|------|------|------|------|
| a | 0.00 | 0.33 | 0.67 | 0.00 | 0.00 | 0.33 |
| b | 0.33 | 0.00 | 0.33 | 0.33 | 0.00 | 0.00 |
| c | 0.33 | 0.33 | 0.00 | 0.00 | 0.67 | 0.00 |
| d | 0.00 | 0.33 | 0.00 | 0.00 | 0.33 | 0.67 |
| e | 0.00 | 0.00 | 0.50 | 1.0 | 0.00 | 0.00 |

Tabla A.3: Tabla de feromonas F_i después de la evaporación

| | a | b | c | d | e | ES |
|---|------|------|------|------|------|------|
| a | 0.00 | 0.28 | 0.50 | 0.00 | 0.00 | 0.25 |
| b | 0.25 | 0.00 | 0.25 | 0.25 | 0.00 | 0.00 |
| c | 0.25 | 0.25 | 0.00 | 0.50 | 0.00 | 0.00 |
| d | 0.00 | 0.25 | 0.00 | 0.00 | 0.25 | 0.50 |
| e | 0.00 | 0.00 | 0.38 | 0.75 | 0.00 | 0.00 |

3. ¿El recorrido seleccionado por la persona, coincide con el recorrido que contiene el mayor nivel de feromonas?

Para la simulación, el CDV s se selecciona de la siguiente manera:

1. Se crea un mapa de probabilidades con las feromonas de los CDV's vecinos del CDV i . Para este fin, se obtiene un vector normalizado del nivel de feromonas $t_{i,j}$ de V_i . Así, cada mapa de probabilidad m_p contiene los límites superior e inferior (li_i, li_s) del rango de probabilidades para el nodo correspondiente.
2. Se calcula un número pseudo-aleatorio ns con una distribución uniforme sobre el intervalo abierto $(0, 1)$.
3. El CDV s corresponde al CDV asociado al elemento de m_p donde $li_i \leq ns < li_s$.

La Figura A.13 muestra un ejemplo de mapa de probabilidades para el conjunto V_c . Considere $ns = 0.23$, como $0.17 \leq ns < 0.34$, entonces se selecciona como CDV seleccionado a b . En este mismo mapa de probabilidades se puede observar que el CDV esperado es e el cual tiene mayor nivel de feromonas.

El modelo fue probado sobre 31 grafos no dirigidos mostrados en las Figuras [A.14, A.15, A.16, A.17]. Los nodos rojos indican las entradas y salidas al área de vigilancia. Los grafos son diferentes en cuando a la cantidad de nodos y configuración de los enlaces entre ellos. En los grafos podemos observar nodos con grados que van desde uno hasta seis. Para cada grafo se consideraron 30 agentes artificiales (personas) con 2000 recorridos cada uno. La entrada a usar en cada recorrido fue seleccionado probabilísticamente

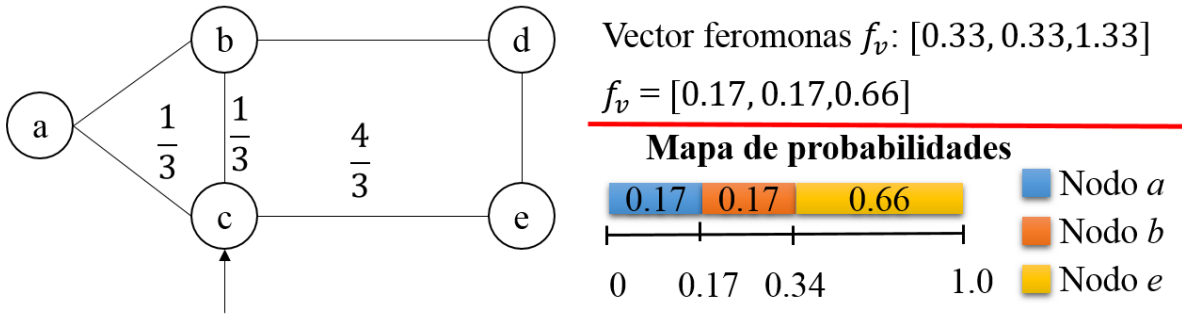


Figura A.13: Mapa de probabilidades para los vecinos del CDV c , los cuales son $V_c = \{a, b, e\}$.

considerando el nivel de feromonas en ellos. Luego se inicia el recorrido, estimando en cada paso el CDV destino usando el modelo propuesto. Un recorrido termina cuando el agente artificial estando en un CDV i , decide probabilísticamente tomar la salida.

Si el CDV s es igual al CDV e , entonces existe una *selección exitosa*. La Razón de Éxitos a selecciones (RES) se propone en esta tesis como medida de desempeño del modelo propuesto, la cual está dada por:

$$RES = \frac{\text{Total de selecciones exitosas}}{\text{total de selecciones}}. \quad (\text{A.6})$$

La Figura A.18 muestra el diagrama de cajas del desempeño RES del modelo propuesto. Como se puede observar, se obtiene una RES promedio superior a 0.9615 con una varianza de 0.0014. En la Figura A.19 muestra el diagrama de cajas de la cantidad de trayectorias seleccionadas por los agentes durante los 2000 ejecuciones. Los agentes tomaron una media de 66 trayectorias diferentes. Los grafos $G1$ y $G28$ tuvieron una varianza grande. Durante los experimentos no se observó alguna condición o características que motivara este desempeño de estos dos grafos. Puede que se deba a la presencia de varios ciclos y la ubicación de entradas/salidas. Sin embargo, no hay evidencia suficiente para sustentar esto, pues otros grafos similares tuvieron pocas variaciones.

Para cada grafo, se determinó la ruta más seleccionada por la persona (RS: Ruta Seleccionada) y la ruta con mayor cantidad de feromonas (RE: Ruta Esperada) al finalizar los 2000 recorridos. Para un buen modelo de predicción, se espera que RS y RE sean iguales. Las Tablas [A.4, A.5, A.6, A.7, A.8, A.9] muestran para cada grafo ambos tipos de rutas y la frecuencia total de la RS. De un total de $31 \text{ grafos} \times 30 \text{ persona} = 930$ ejecuciones, solo en el 6.7% de las veces la RS y la RE fueron diferentes.

El modelo de predicción propuesto muestra resultados muy prometedores para la reidentificación facial. El experimento extenso muestra que el modelo obtiene un desempeño RES superior al 96%. En la mayoría de los casos, el recorrido seleccionado por el agente artificial fue igual al de mayor contenido de nivel de feromonas. Durante los experimentos, no se observó algún indicio de que el grado de los nodos influyera sig-

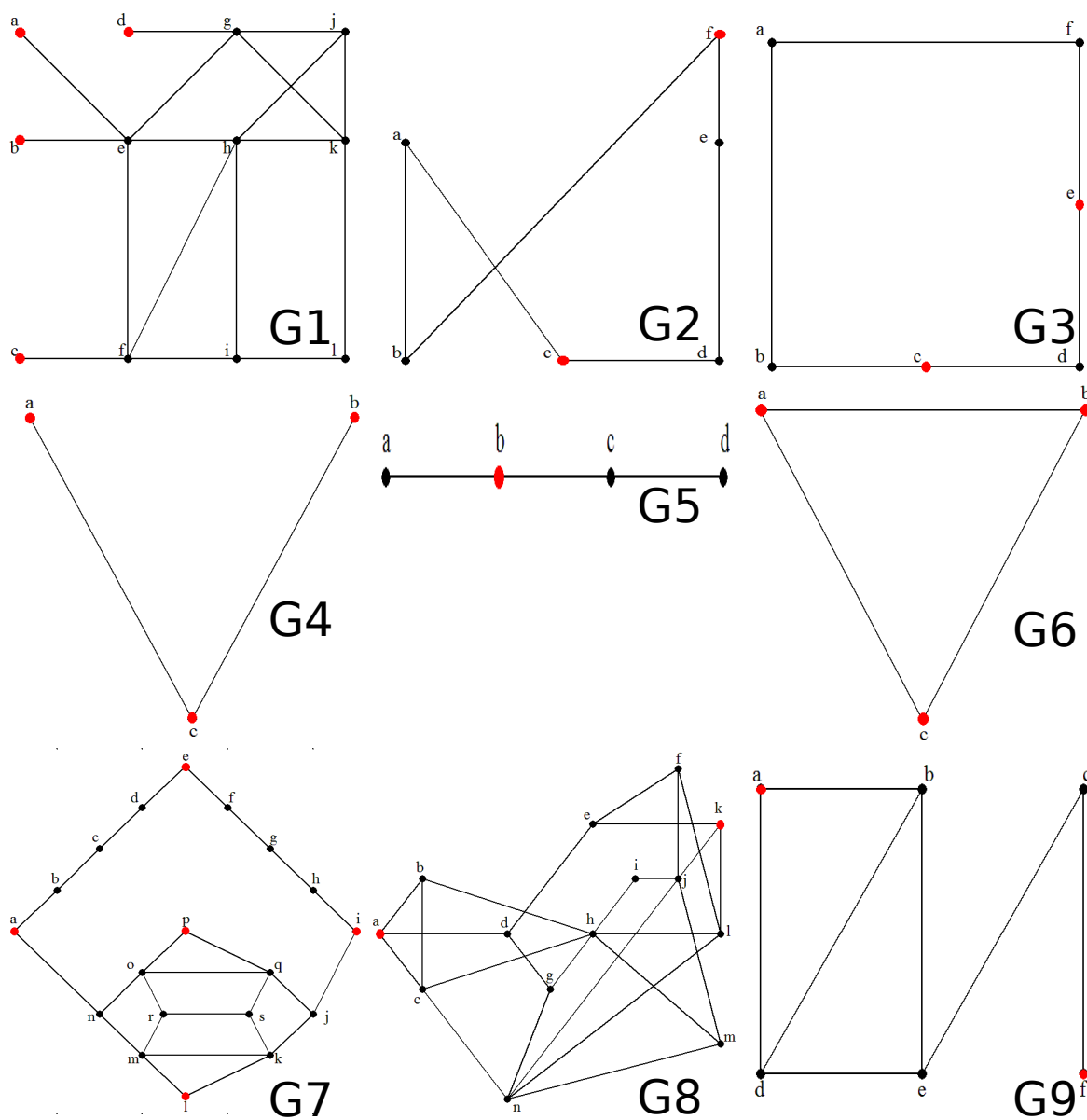


Figura A.14: Primer grupo de grafos usados en el experimento.

nificativamente en el desempeño del modelo. Sin embargo, si se observó que mientras más grande sea el grafo, es necesario de mayor cantidad de entradas/salidas. A pesar de esto, los resultados obtenidos indican el modelo es capaz de encontrar la salida en la mayoría de los casos.

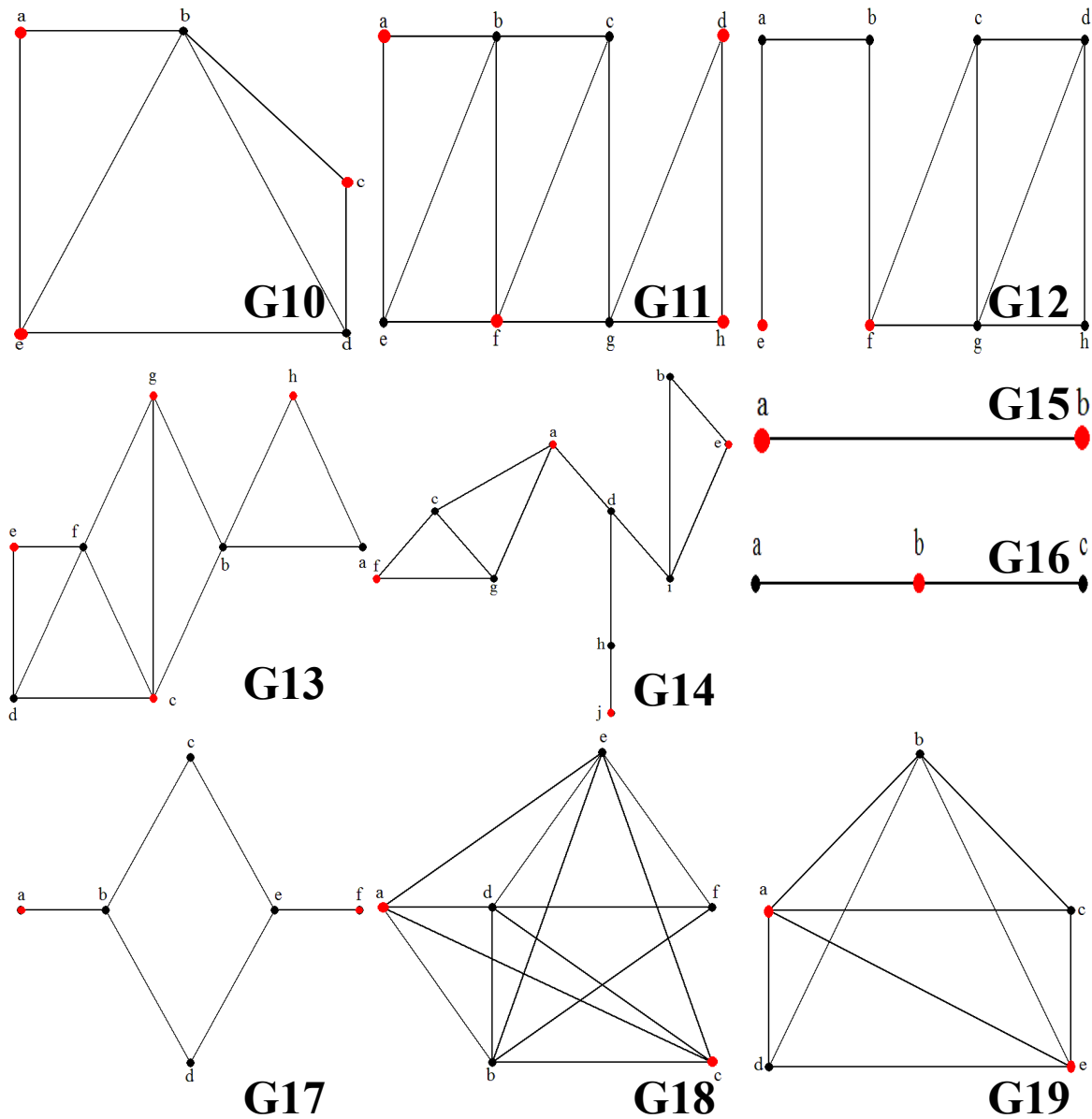


Figura A.15: Segundo grupo de grafos usados en el experimento.

A.4. Conclusiones

En este capítulo se ha presentado un modelo basado en el nivel de feromonas para la predicción del CDV destino. Los resultados obtenidos de la simulación muestran una efectividad media superior al 96 %. Los resultados muestran también que en la mayoría de los casos el recorrido más frecuentado por la persona es igual al que contiene mayor cantidad de feromonas. Esto indica que el modelo es confiable para la predicción del destino en base al patrón de movimiento de un agente artificial.

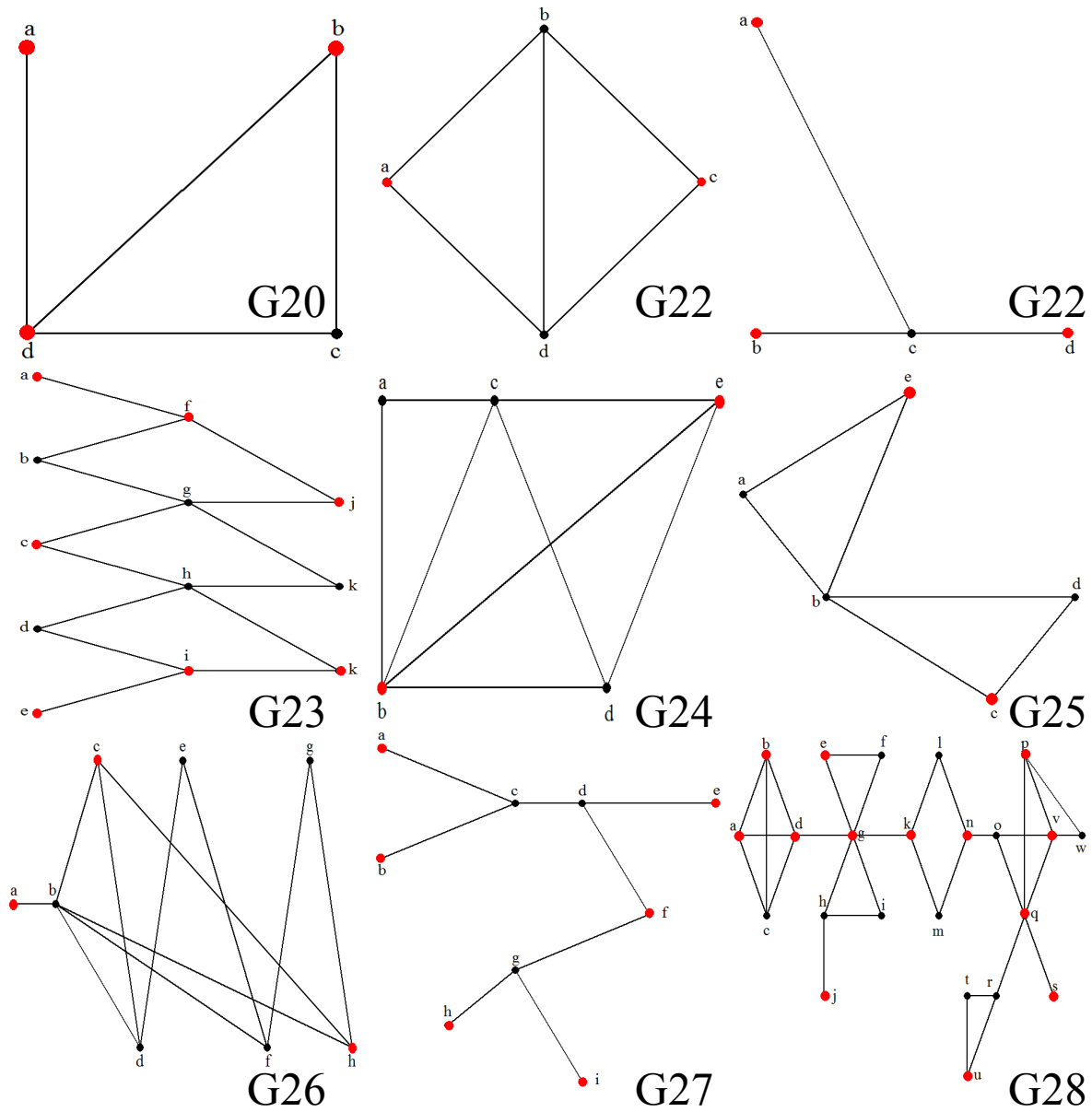


Figura A.16: Tercer grupo de grafos usados en el experimento.

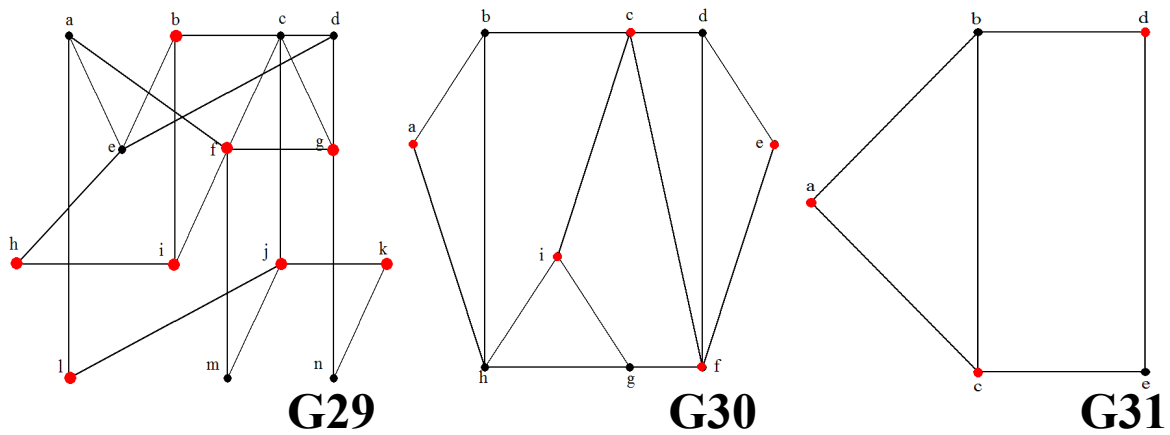


Figura A.17: Cuarto grupo de grafos usados en el experimento.

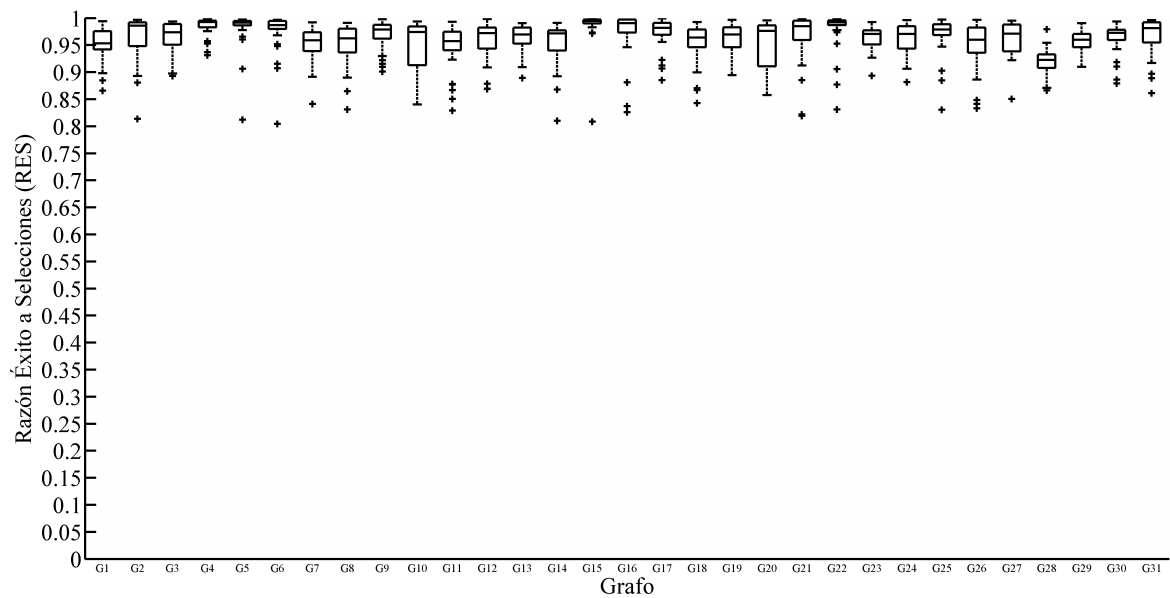


Figura A.18: Diagrama de cajas de la Razón de Éxito a Selecciones (RES) obtenido en cada grafo.

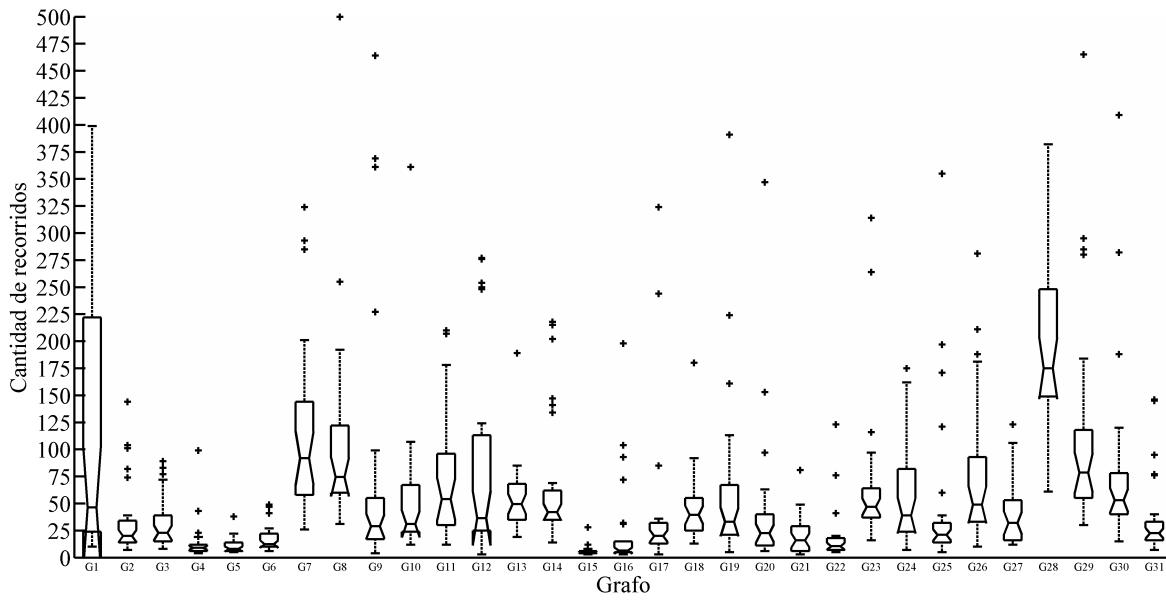


Figura A.19: Diagrama de cajas de la cantidad de rutas tomadas por agente artificial sobre cada grafo.

Tabla A.4: Ruta Seleccionada (RS), Ruta Esperada (RE) y Frecuencia de la Ruta Seleccionada (FRS) de cada persona sobre los grafos 1 al 5. En verde las celdas con los recorridos donde RS y RE son diferente.

| | G1 | | | G2 | | | G3 | | | G4 | | | G5 | | |
|-----------|-----------|------|----------|----------|------|---------|---------|------|-------|-------|------|-------|-------|------|--|
| RS | RE | FRS | RS | RE | FRS | RS | RE | FRS | RS | RE | FRS | RS | RE | FRS | |
| aefhjgdES | aefhjgdES | 1081 | fES | fES | 1367 | edcES | edcES | 1429 | acbES | acbES | 991 | dES | dES | 878 | |
| cfEbES | cfEbES | 788 | acES | acES | 922 | cbafeES | cbafeES | 1211 | aES | aES | 1281 | dES | dES | 839 | |
| aegdES | aegdES | 1030 | caES | caES | 1341 | efabcES | efabcES | 1535 | bcES | bcES | 1165 | dES | dES | 960 | |
| aES | aES | 984 | cdefES | cdefES | 847 | cdeES | cdeES | 1228 | cES | cES | 889 | baES | baES | 793 | |
| aES | aES | 735 | aES | aES | 934 | cdeES | cdeES | 1241 | bES | bES | 1118 | baES | baES | 971 | |
| cfeaES | cfeaES | 650 | caES | caES | 821 | cES | cES | 1112 | bES | bES | 957 | aES | aES | 956 | |
| dES | dES | 1033 | caES | caES | 809 | cdeES | cdeES | 1446 | bcES | bcES | 897 | bES | bES | 1241 | |
| aES | aES | 701 | caES | caES | 1015 | cdeES | cdeES | 1169 | cES | cES | 1358 | dES | dES | 838 | |
| bES | bES | 1069 | caES | caES | 834 | eES | eES | 1533 | acES | acES | 1033 | debES | debES | 1178 | |
| bES | bES | 1188 | abfES | abfES | 837 | cES | cES | 1509 | aES | aES | 1104 | dES | dES | 915 | |
| cES | cES | 911 | fES | fES | 1334 | cbafeES | cbafeES | 1447 | caES | caES | 1301 | dES | dES | 1302 | |
| cfeaES | cfeaES | 791 | aES | aES | 1330 | cbafeES | cbafeES | 1480 | cES | cES | 887 | bES | bES | 1245 | |
| bES | bES | 653 | cdefES | cdefES | 1369 | cdeES | cdeES | 1480 | bES | bES | 1312 | debES | debES | 810 | |
| bES | bES | 1220 | fedcES | fedcES | 904 | eES | eES | 1243 | aES | aES | 901 | aES | aES | 1128 | |
| cES | cES | 1158 | fedcES | fedcES | 1188 | cES | cES | 1428 | acbES | acbES | 740 | abES | abES | 1228 | |
| aES | aES | 721 | cdefbaES | cdefbaES | 957 | cdeES | cdeES | 1350 | caES | caES | 1112 | dcbES | dcbES | 1350 | |
| dES | dES | 697 | caES | caES | 1362 | cdeES | cdeES | 985 | cES | cES | 996 | bES | bES | 1221 | |
| bES | bES | 622 | caES | caES | 1283 | cES | cES | 1471 | aES | aES | 1076 | dES | dES | 949 | |
| aES | aES | 1222 | fbES | fbES | 1055 | cES | cES | 1451 | caES | caES | 1289 | bES | bES | 1098 | |
| aES | aES | 1170 | cdefES | cdefES | 1334 | edcES | edcES | 1236 | cES | cES | 1215 | abES | abES | 807 | |
| cES | cES | 1072 | fES | fES | 865 | cES | cES | 1521 | cES | cES | 1247 | bES | bES | 1162 | |
| cES | cES | 745 | cES | cES | 1101 | eES | eES | 1297 | caES | caES | 1296 | dES | dES | 829 | |
| aES | aES | 1202 | abfES | abfES | 1197 | edcES | edcES | 1553 | bcES | bcES | 907 | baES | baES | 1331 | |
| dES | dES | 661 | aES | aES | 1308 | cES | cES | 1565 | bES | bES | 1182 | dcbES | dcbES | 1017 | |
| aES | aES | 1172 | fedcES | fedcES | 1075 | cES | cES | 1397 | cES | cES | 838 | baES | baES | 940 | |
| dES | dES | 772 | fES | fES | 860 | efabcES | efabcES | 1090 | cES | cES | 898 | debES | debES | 846 | |
| aegdES | aegdES | 1178 | acES | acES | 905 | eES | eES | 1027 | aES | aES | 1177 | dES | dES | 840 | |
| dES | dES | 823 | fedcaES | fedcaES | 1010 | eES | eES | 1234 | cbES | cbES | 1334 | bES | bES | 1358 | |
| aES | begkjhif | 492 | aES | aES | 876 | eES | eES | 965 | bcaES | bcaES | 954 | aES | aES | 850 | |
| cES | cES | 1166 | aES | aES | 769 | cES | cES | 1268 | bES | bES | 815 | baES | baES | 1083 | |

Tabla A.5: Ruta Seleccionada (RS), Ruta Esperada (RE) y Frecuencia de la Ruta Seleccionada (FRS) de cada persona de los grafos 6 al 10. En verde las celdas con los recorridos donde RS y RE son diferente.

| G6 | | | G7 | | | G8 | | | G9 | | | G10 | | | |
|-------|-------|------|----------------|----------------|-----|---------------|---------------|------|----------|----------|-------|---------|---------|------|-----|
| RS | RE | FRS | RS | RE | FRS | RS | RE | FRS | RS | RE | FRS | RS | RE | FRS | |
| acbES | acbES | 1329 | ijkmrES | ijkmrES | 896 | dES | dES | 630 | fES | abecfES | 1018 | aeES | aeES | 820 | |
| bES | bES | 1143 | sqpES | sqpES | 693 | abcnlfedES | abcnlfedES | 1026 | aES | aES | 1304 | eabdcES | eabdcES | 1183 | |
| bES | bES | 755 | IES | IES | 716 | ngdES | ngdES | 1066 | fES | fES | 1536 | cES | cES | 849 | |
| acES | acES | 1033 | rsES | rsES | 547 | hlfedgnES | hlfedgnES | 537 | aES | aES | 1312 | cbdeES | cbdeES | 872 | |
| bES | bES | 855 | pES | rmkjhgfES | 353 | nlkES | nlkES | 1046 | abecfES | abecfES | 1222 | cES | cES | 1049 | |
| acES | acES | 1212 | pqiES | pqiES | 591 | dES | dES | 1078 | fES | fES | 1426 | eaES | eaES | 1014 | |
| bcES | bcES | 718 | anoqsrES | anoqsrES | 386 | nlkES | nlkES | 659 | abdecfES | abdecfES | 1421 | abdeES | abdeES | 1313 | |
| cbES | cbES | 988 | lmrES | lmrES | 536 | kjfedacbhES | kjfedacbhES | 934 | aES | aES | 1484 | eaES | eaES | 1066 | |
| cbES | cbES | 578 | anorsES | anorsES | 854 | aES | aES | 727 | fES | fES | 1587 | aES | aES | 1200 | |
| acES | acES | 1275 | IES | IES | 485 | adES | adES | 1017 | fES | fES | 1037 | abcES | abcES | 826 | |
| aES | aES | 967 | iES | iES | 793 | aES | aES | 1082 | abdecfES | abdecfES | 1249 | cheaES | cheaES | 761 | |
| abES | abES | 927 | pES | pES | 407 | dghES | dghmjihES | 275 | abecfES | abecfES | 1496 | ebaES | ebaES | 850 | |
| aES | aES | 1275 | pES | pES | 536 | nES | nES | 563 | adecfES | adecfES | 1438 | ebcES | ebcES | 1223 | |
| aES | aES | 1129 | ijklmnaES | ijklmnaES | 695 | dES | dES | 928 | fES | febad** | febad | 1075 | eES | eES | 924 |
| bES | bES | 1305 | pqjknnabcedeES | pqjknnabcedeES | 425 | ngdES | ngdES | 965 | aES | aES | 1121 | eaES | eaES | 758 | |
| aES | aES | 1182 | rmlES | rmlES | 743 | dES | dES | 729 | fES | adbecfES | 959 | cES | cES | 1290 | |
| cabES | cabES | 1299 | skjiES | skjiES | 450 | hgDES | hgDES | 982 | feedaES | feedaES | 1512 | eES | eES | 871 | |
| acbES | acbES | 1353 | rmlksES | rmlksES | 825 | nES | klfekES | 552 | aES | aES | 1450 | edeES | edeES | 1315 | |
| bES | bES | 1129 | ronmksES | ronmksES | 438 | nlkES | nlkES | 1027 | fES | fES | 1497 | ebdcES | ebdcES | 808 | |
| cbaES | cbaES | 1306 | pES | sES | 405 | abcnlfjihgdES | abcnlfjihgdES | 1027 | fES | fES | 1463 | ebcES | ebcES | 789 | |
| caES | caES | 1341 | ijqopES | ijqopES | 846 | dghlnES | dghlnES | 539 | abecfES | abecfES | 982 | aeES | aeES | 948 | |
| baES | baES | 1179 | eES | eES | 740 | abhES | abhES | 764 | fES | fES | 1237 | aebcES | aebcES | 405 | |
| cES | cES | 907 | ihgfedcbaES | ihgfedcbaES | 592 | acbhES | acbhES | 590 | fES | fES | 1175 | aebdcES | aebdcES | 1020 | |
| bcES | bcES | 1094 | eES | eES | 553 | hmgdES | hmgdES | 959 | fES | fES | 1450 | edbaES | edbaES | 1025 | |
| cbES | cbES | 1244 | sES | sES | 599 | aES | aES | 1016 | fES | fES | 1515 | abcES | abcES | 934 | |
| baES | baES | 1045 | sES | aES | 366 | nmjihES | nmjihES | 911 | aES | aES | 1132 | cdbeES | cdbeES | 1129 | |
| bcaES | bcaES | 1224 | lmrES | lmrES | 587 | hmnlfedabhm** | hmnlfedabhES | 202 | adbecfES | adbecfES | 1372 | ebaES | ebaES | 779 | |
| bcES | bcES | 1205 | pqiES | pqiES | 400 | daES | daES | 697 | febaES | febaES | 1017 | eES | eES | 747 | |
| abES | abES | 1149 | pqskmrES | pqskmrES | 520 | dES | dES | 616 | aES | aES | 1451 | cdbaES | cdbaES | 1000 | |
| abES | abES | 874 | eES | eES | 569 | nES | nES | 654 | aES | aES | 1380 | cES | cES | 1340 | |

Tabla A.6: Ruta Seleccionada (RS), Ruta Esperada (RE) y Frecuencia de la Ruta Seleccionada (FRS) de cada persona sobre los grafos 11 al 15. En verde las celdas con los recorridos donde RS y RE son diferente.

| G11 | | | G12 | | | G13 | | | G14 | | | G15 | | |
|---------|---------|------|------------|------------|------|---------|----------|------|------------|------------|------|------|------|------|
| RS | RE | FRS | RS | RE | FRS | RS | RE | FRS | RS | RE | FRS | RS | RE | FRS |
| fES | fES | 741 | eES | eES | 1545 | degES | degES | 616 | jhdaES | jhdaES | 952 | aES | aES | 1480 |
| cES | cES | 1164 | gES | gES | 1471 | hES | hES | 546 | fES | fES | 874 | aES | aES | 1500 |
| cfeaES | cfeaES | 888 | eabfgES | eabfgES | 1433 | hbgES | hbgES | 805 | ebiaES | ebiaES | 978 | bES | bES | 1484 |
| aES | aES | 600 | eES | eES | 1534 | hbES | hbES | 649 | fegaES | fegaES | 944 | aES | aES | 1515 |
| cES | cES | 895 | gES | gES | 1165 | bES | eES | 425 | jhdagfES | jhdagfES | 869 | baES | baES | 1295 |
| hES | hES | 847 | eES | eES | 1541 | cdES | cdES | 883 | ebidagcfES | ebidagcfES | 710 | bES | bES | 1520 |
| hgdES | hgdES | 574 | gfbaeES | gfbaeES | 1303 | dfgbES | dfgbES | 430 | adhjES | adhjES | 1016 | abES | abES | 1569 |
| hbeaES | hbeaES | 633 | eabfgES | eabfgES | 938 | gfcbeES | gfcbeES | 532 | jES | jES | 551 | bES | bES | 1532 |
| cES | cES | 1018 | gES | gES | 1175 | cfdeES | cfdeES | 660 | eES | eES | 1075 | abES | abES | 1199 |
| fES | fES | 444 | eabfcdhgES | eabfcdhgES | 1322 | bahES | bahES | 696 | fcgadhjES | fcgadhjES | 784 | aES | aES | 1544 |
| dhES | dhES | 750 | gcfbaeES | gcfbaeES | 1525 | hbgES | hbgES | 931 | eES | eES | 823 | bES | bES | 1034 |
| fbaeES | fbaeES | 577 | eES | eES | 1525 | bahES | dfcbahES | 415 | aES | aES | 1226 | bES | bES | 1393 |
| fES | fES | 812 | eabfgES | eabfgES | 1199 | gbcES | gbcES | 538 | jES | ebidhjES | 613 | bES | bES | 1141 |
| hES | hES | 696 | eES | eES | 1545 | cdES | cdES | 835 | jES | agfeaES | 307 | abES | abES | 1473 |
| hES | hES | 538 | gES | gES | 1521 | efdES | efdES | 833 | fES | fES | 750 | aES | aES | 1553 |
| fES | fES | 590 | gES | gES | 1117 | cES | cES | 589 | fES | fES | 1073 | baES | baES | 1531 |
| cbaES | feaES | 438 | eES | eES | 1461 | hES | hES | 498 | jES | adibeES | 687 | bES | bES | 1422 |
| cES | cES | 556 | gES | gES | 1511 | eES | eES | 503 | eidhjES | eidhjES | 758 | abES | abES | 959 |
| fgcES | fgcES | 1066 | eES | eES | 1536 | dfcgbES | dfcgbES | 494 | fES | fES | 778 | abES | abES | 1563 |
| fES | fES | 555 | gES | gES | 1266 | cdES | cdES | 713 | jES | eidaES | 460 | bES | bES | 1256 |
| aebcES | ldhES | 491 | eES | eES | 1470 | cES | cES | 656 | jhdaES | jhdaES | 756 | abES | abES | 971 |
| cES | cES | 943 | eES | eES | 1503 | habES | habES | 546 | jhdaibeES | jhdaibeES | 851 | bES | bES | 1276 |
| aES | aES | 723 | gcfbaeES | gcfbaeES | 1134 | dES | dES | 587 | fES | fES | 894 | abES | abES | 1525 |
| cfbaES | cfbaES | 664 | gES | gES | 1543 | eES | eES | 467 | eidhjES | eidhjES | 610 | aES | aES | 1487 |
| hES | hES | 523 | gES | gES | 1443 | cES | cES | 1042 | fES | fES | 692 | baES | baES | 1511 |
| cES | cES | 1054 | gES | gES | 1215 | dfgES | dfgES | 788 | eES | eES | 721 | aES | aES | 1461 |
| dghbcES | dghbcES | 686 | gES | gES | 1466 | cES | cES | 1072 | aES | aES | 744 | aES | aES | 1578 |
| cES | fES | 583 | gES | gES | 1190 | hbgfcES | hbgfcES | 672 | eES | eES | 1236 | baES | baES | 1279 |
| aES | aES | 768 | eabfgES | eabfgES | 1012 | cdES | bES | 507 | jES | jES | 1034 | abES | abES | 1407 |
| cfES | cfES | 801 | eES | eES | 1491 | dES | dES | 581 | fES | fES | 1211 | bES | bES | 1109 |

Tabla A.7: Ruta Seleccionada (RS), Ruta Esperada (RE) y Frecuencia de la Ruta Seleccionada (FRS) de cada persona sobre los grafos 16 al 20. En verde las celdas con los recorridos donde RS y RE son diferente.

| G16 | | | G17 | | | G18 | | | G19 | | | G20 | | |
|------|-------|------|---------|---------|------|----------|----------|------|---------|-----------|------|-------|---------|------|
| RS | RE | FRS | RS | RE | FRS | RS | RE | FRS | RS | RE | FRS | RS | RE | FRS |
| baES | baES | 1537 | abcefES | abcefES | 1077 | cES | cES | 1454 | aES | abcedaES | 445 | bES | bES | 995 |
| baES | baES | 1525 | fES | fES | 1333 | cebaES | cebaES | 1353 | adeES | adeES | 1098 | aES | aES | 1343 |
| bES | bES | 1202 | aES | aES | 1296 | aedfbcES | aedfbcES | 1275 | adbceES | adbceES | 1250 | dES | dES | 1261 |
| baES | baES | 1481 | fecbaES | fecbaES | 1082 | aES | aES | 974 | eaES | eaES | 1444 | adbES | adbES | 1111 |
| bES | bES | 1558 | fecbaES | fecbaES | 1481 | cdbaES | cdbaES | 1490 | eaES | eaES | 1500 | dES | dES | 928 |
| baES | baES | 1432 | abdefES | abdefES | 1417 | acES | acES | 1488 | eES | eES | 1010 | adES | adcbdES | 317 |
| bES | bES | 1126 | fES | fES | 1229 | caES | caES | 1455 | eES | eES | 1289 | aES | aES | 1139 |
| baES | baES | 1565 | abcefES | abcefES | 1061 | cES | cES | 1019 | aES | aES | 1206 | aES | aES | 822 |
| abES | abES | 1046 | abdefES | abdefES | 1242 | cdaES | cdaES | 1393 | eaES | eaES | 1574 | aES | aES | 1309 |
| aES | aES | 1190 | aES | aES | 1505 | acES | acES | 1282 | aES | aES | 1575 | bES | bES | 794 |
| aES | aES | 1470 | aES | aES | 1595 | aES | aES | 1308 | aceES | aceES | 1434 | bES | bES | 1273 |
| aES | aES | 1556 | abcefES | abcefES | 1512 | cbefdaES | cbefdaES | 1484 | eaES | eaES | 1398 | bdES | bdES | 901 |
| bES | bES | 1570 | aES | aES | 1470 | caES | caES | 1511 | aES | aES | 1295 | adES | adES | 1209 |
| bES | bES | 1538 | fES | fES | 1298 | cES | cES | 1275 | eaES | eaES | 1298 | bcdES | bcdES | 876 |
| bES | bES | 1112 | fES | fES | 1561 | ceaES | ceaES | 1485 | acbeES | acbeES | 1421 | adES | adES | 1125 |
| aES | aES | 1521 | aES | fecbd | 455 | aES | aES | 1107 | eES | eES | 1454 | dcbES | dcbES | 968 |
| bES | bES | 1555 | fES | fES | 1270 | aedfbcES | aedfbcES | 1376 | abceES | abceES | 1467 | dES | dES | 980 |
| aES | bebES | 635 | aES | aES | 1167 | acES | acES | 1521 | abdeES | abdeES | 1193 | dES | dES | 1060 |
| abES | abES | 1484 | aES | aES | 1508 | cES | adbef | 1014 | eaES | eachbeES | 148 | bES | bES | 1178 |
| bES | bES | 1464 | aES | aES | 1547 | aES | aES | 1270 | eES | eES | 1411 | dbES | dbES | 1310 |
| baES | baES | 1476 | fES | fES | 1525 | aES | aES | 1043 | eES | eES | 1307 | dbES | dbES | 699 |
| abES | abES | 1544 | abdefES | abdefES | 1507 | afbecES | afbecES | 1248 | aES | aES | 986 | bES | bES | 1080 |
| bES | bES | 1541 | fES | fES | 1154 | cES | cES | 1043 | aES | aES | 1511 | aES | aES | 1336 |
| aES | aES | 1512 | fES | fES | 1263 | cES | cES | 1517 | ecaES | ecadbceES | 178 | aES | aES | 1020 |
| baES | baES | 1124 | fecbaES | fecbaES | 1419 | aES | cdbfeaES | 827 | ecaES | ecaES | 1319 | dbES | dbES | 868 |
| abES | abES | 1128 | aES | aES | 1555 | aES | aES | 1552 | adeES | adeES | 1071 | adES | adES | 816 |
| abES | abES | 1562 | fES | fES | 1313 | cES | cES | 1257 | eES | eES | 1262 | bES | bES | 1325 |
| bES | bES | 1556 | aES | aES | 1256 | cES | cES | 1075 | ecaES | ecaES | 1140 | bdES | bdES | 1296 |
| aES | bebES | 684 | aES | aES | 1276 | aebdcES | aebdcES | 1450 | adeES | adeES | 1210 | dES | dES | 900 |
| aES | bebES | 436 | fES | fES | 1051 | aES | aES | 1561 | eES | eES | 1533 | aES | aES | 1022 |

Tabla A.8: Ruta Seleccionada (RS), Ruta Esperada (RE) y Frecuencia de la Ruta Seleccionada (FRS) de cada persona sobre los grafos 21 al 25. En verde las celdas con los recorridos donde RS y RE son diferente.

| G21 | | | G22 | | | G23 | | | G24 | | | G25 | | |
|--------|--------|------|-------|-------|------|----------|-----------|-----|---------|---------|------|----------|---------|------|
| RS | RE | FRS | RS | RE | FRS | RS | RE | FRS | RS | RE | FRS | RS | RE | FRS |
| cES | cES | 1467 | bcaES | bcaES | 1326 | eES | eES | 548 | baceES | baceES | 1245 | ebcES | ebcES | 1524 |
| abcES | abcES | 1227 | dES | dES | 1199 | jES | jES | 626 | abES | abES | 1331 | eES | eES | 1514 |
| cdbaES | cdbaES | 1027 | bcdES | bcdES | 1258 | jES | fbgkhdil | 372 | aES | aES | 1173 | ebcES | ebcES | 1504 |
| abdcES | abdcES | 1455 | dES | dES | 1364 | jkhdilES | jkhdilES | 900 | bES | bES | 1317 | cES | cES | 1214 |
| cES | cES | 1213 | dcaES | dcaES | 1321 | jgbfES | fES | 382 | eES | beES | 765 | cbaeES | cbaeES | 1188 |
| aES | aES | 912 | acdES | acdES | 720 | afbgjES | afbgjES | 634 | abceES | abceES | 877 | cES | cES | 1085 |
| cES | cES | 1142 | bES | bES | 1114 | aES | aES | 470 | edcbaES | edcbaES | 967 | eabcES | eabcES | 1491 |
| cES | cES | 1467 | dcaES | dcaES | 1033 | aES | aES | 564 | baES | baES | 769 | cdbaeES | cdbaeES | 1169 |
| aES | aES | 1455 | bES | bES | 909 | iES | iES | 831 | bdcaES | bdcaES | 835 | cbaeES | cbaeES | 1398 |
| cES | cES | 1047 | aES | aES | 902 | jfaES | jfaES | 903 | edbcaES | edbcaES | 1161 | eES | eES | 1567 |
| cES | cES | 1205 | bES | bES | 830 | eES | eES | 501 | eES | eES | 767 | cbeES | cbeES | 1419 |
| aES | aES | 1060 | bES | bES | 1314 | cgjfES | cgjfES | 499 | eES | eES | 1308 | cbaeES | cbaeES | 1344 |
| cES | cES | 1508 | aES | aES | 844 | jgbfES | jgbfES | 609 | bdeES | bdeES | 735 | cbeES | cbeES | 1481 |
| cES | cES | 1557 | acbES | acbES | 813 | iES | iES | 932 | abeES | abeES | 777 | ebdcES | ebdcES | 1269 |
| cES | cES | 1146 | acdES | acdES | 1337 | eES | afbgkhlid | 227 | bES | bES | 1175 | cdbaeb** | cdbae | 1051 |
| aES | aES | 1527 | bcaES | bcaES | 1349 | jfaES | jfaES | 397 | acedbES | acedbES | 995 | eES | eES | 1548 |
| cES | cES | 1540 | acdES | acdES | 1082 | fES | fES | 834 | abeES | abeES | 1192 | eES | eES | 1343 |
| abcES | abcES | 1529 | bcdES | bcdES | 804 | jkhcES | jkhcES | 379 | aES | aES | 1302 | cES | cES | 1534 |
| aES | aES | 1457 | bES | bES | 1298 | aES | aES | 622 | eES | eES | 962 | cES | cES | 1525 |
| aES | aES | 1493 | bES | bES | 763 | idhcgjES | idhcgjES | 785 | bdeES | bdeES | 467 | cbeES | cbeES | 1286 |
| adbCES | adbCES | 1451 | dcbES | dcbES | 1292 | afES | afES | 597 | abeES | abeES | 1212 | cES | cES | 1552 |
| cES | cES | 1198 | bES | bES | 1327 | eilES | eilES | 911 | ebaES | ebaES | 1134 | cES | cES | 1502 |
| aES | aES | 1592 | bES | bES | 1368 | faES | faES | 527 | beES | beES | 868 | cbeES | cbeES | 1543 |
| aES | aES | 1468 | bcaES | bcaES | 1073 | jgbfES | jgbfES | 390 | eES | abcaES | 756 | eES | eES | 762 |
| cbaES | cbaES | 1195 | dcbES | dcbES | 1222 | cES | cES | 647 | eES | eES | 1247 | cES | cES | 1547 |
| aES | aES | 1546 | dcbES | dcbES | 835 | eES | eES | 435 | eES | abcaES | 422 | eES | eES | 1566 |
| cES | cES | 1040 | aES | aES | 881 | fES | fES | 697 | beES | acdbeES | 878 | cES | cES | 1263 |
| aES | aES | 1219 | dES | dES | 1196 | jES | jES | 635 | aES | aES | 1334 | cES | cES | 1534 |
| abcES | abcES | 1277 | aES | aES | 997 | iES | iES | 249 | acbES | acbES | 1045 | eES | eES | 1535 |
| cES | cES | 1245 | aES | aES | 1055 | IES | IES | 547 | bcdeES | bcdeES | 1143 | eES | eES | 1172 |

Tabla A.9: Ruta Seleccionada (RS), Ruta Esperada (RE) y Frecuencia de la Ruta Seleccionada (FRS) de cada persona sobre los grafos 26 al 31. En verde las celdas con los recorridos donde RS y RE son diferente.

| G26 | | | G27 | | | G28 | | | G29 | | | G30 | | | G31 | | |
|------------|------------|------|---------|---------|-----|----------|----------|-----|--------------|--------------|-----|----------|-----------|------|---------|---------|------|
| RS | RE | FRS | RS | RE | FRS | RS | RE | FRS | RS | RE | FRS | RS | RE | FRS | RS | RE | FRS |
| chES | chES | 1244 | aES | aES | 535 | xpES | xpES | 210 | lfgcjkndebES | lfgcjkndebES | 384 | edfES | edfES | 823 | abdES | abdES | 1142 |
| cdcfghES | cdcfghES | 1103 | hES | hES | 626 | wES | wES | 544 | gdnkES | gdnkES | 663 | cES | cES | 760 | acedES | acedES | 1070 |
| chES | chES | 1050 | iES | iES | 513 | wvqonES | wvqonES | 260 | IES | IES | 362 | iES | iES | 561 | abdES | abdES | 1179 |
| abhgfdedES | abhgfdedES | 1011 | hES | hgfES | 441 | bES | bES | 179 | kndehibES | kndehibES | 365 | edfES | edfES | 914 | chaES | chaES | 1315 |
| abcES | abcES | 1226 | edcbES | edcbES | 681 | dES | dES | 483 | headjES | headjES | 581 | iES | iES | 566 | dbaES | dbaES | 1253 |
| hES | hES | 754 | fES | fES | 511 | kES | kES | 305 | fibES | fibES | 420 | fedES | fedES | 608 | dES | acvIES | 750 |
| chES | chES | 1104 | igfdeES | igfdeES | 588 | wES | kmnoqrES | 202 | ibES | ibES | 440 | abcES | abcES | 726 | caES | caES | 1299 |
| aES | aES | 987 | fES | hgES | 419 | gES | bcaES | 144 | IES | jmnES | 331 | eES | ihbES | 488 | aES | aES | 1273 |
| cES | cES | 845 | bcaES | bcaES | 869 | IES | urtuES | 173 | ljcgES | ljcgES | 712 | aES | aES | 554 | cES | cES | 903 |
| abfedeES | abfedeES | 800 | hES | hES | 851 | wES | wES | 412 | lafgES | lafgES | 405 | ahbcES | ahbcES | 865 | dES | dES | 990 |
| aES | cde | 594 | fES | fES | 504 | dbaES | dbaES | 784 | hES | lcb | 323 | fES | fES | 582 | acES | acES | 846 |
| hES | hES | 1025 | fES | fES | 560 | aES | aES | 244 | bES | bES | 583 | fES | fES | 1124 | dES | dES | 1319 |
| cES | cES | 863 | fES | fES | 531 | nmkES | nmkES | 251 | hES | fajmES | 280 | aES | aES | 1041 | cES | cES | 1253 |
| hES | hES | 1162 | eES | eES | 795 | eES | eES | 483 | kjmfaehES | kjmfaehES | 430 | ahbcES | ahbcES | 675 | abdES | abdES | 974 |
| cbfghES | cbfghES | 1232 | iES | iES | 763 | ihgkmmES | ihgkmmES | 437 | hiES | hiES | 511 | cfghaES | cfghaES | 748 | debaES | debaES | 1215 |
| aES | aES | 1314 | bES | bES | 950 | bES | bES | 281 | hES | hES | 392 | icfES | icfES | 949 | dbaES | dbaES | 1014 |
| cES | cES | 1333 | edfES | edfES | 920 | qES | qES | 288 | fES | fES | 362 | icbhgfES | icbhgfES | 644 | dbacES | dbacES | 762 |
| chES | chES | 1289 | hgfES | hgfES | 923 | iES | iES | 305 | kjES | headjES | 352 | cighbaES | cighbaES | 554 | abdecES | abdecES | 1030 |
| chES | chES | 1277 | hES | hES | 469 | jES | jES | 219 | ibeaES | ibeaES | 458 | efghbES | efghbES | 843 | dES | acedES | 654 |
| cES | cES | 1366 | eES | eES | 772 | wES | pvpES | 201 | hES | hES | 412 | aES | aES | 1011 | abdES | abdES | 1268 |
| cdcfghcl** | cdcfghcl** | 553 | acbES | acbES | 577 | sES | sES | 264 | jmfcbES | jmfcbES | 626 | edfES | edfES | 837 | caES | caES | 1301 |
| aES | aES | 910 | acdeES | igfdeES | 447 | vES | vES | 399 | hiES | hiES | 645 | eES | eES | 600 | cES | cES | 816 |
| cES | cES | 1291 | bES | bES | 534 | kmnoqvES | kmnoqvES | 396 | fgES | fgES | 554 | icfES | icfES | 586 | acbdES | acbdES | 1355 |
| cES | cES | 633 | acbES | acbES | 697 | jES | jES | 327 | gES | gES | 611 | cfghbaES | cfghbaES | 632 | caES | caES | 1258 |
| aES | aES | 1102 | acbES | acbES | 773 | bcaES | bcaES | 239 | jcdcaES | jcdcaES | 528 | cES | abgtdedES | 417 | caES | caES | 644 |
| hbdcES | hbdcES | 1156 | acdfES | acdfES | 496 | nlkES | nlkES | 266 | kjES | kjES | 529 | fdeES | fdeES | 790 | decES | decES | 1326 |
| aES | cbdefgh | 701 | bES | bES | 780 | uES | uES | 374 | gES | gES | 471 | cdfeES | cdfeES | 825 | acES | decES | 833 |
| abdchES | abdchES | 1032 | acdeES | acdeES | 817 | vqES | vqES | 370 | bES | bES | 618 | ihES | ihES | 812 | cES | cES | 1326 |
| hES | hES | 894 | aES | aES | 738 | dbES | ntnuES | 222 | jES | jES | 431 | cES | cES | 564 | decES | decES | 1311 |
| aES | chcES | 383 | iES | iES | 617 | jhgfES | jhgfES | 218 | fgcjkES | fgcjkES | 760 | ahideES | ahideES | 637 | dES | dES | 1158 |

Referencias

- Adam, A., Rivlin, E., y Shimshoni, I. (2006). Robust fragments-based tracking using the integral histogram. En *Proceedings of the 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition* (p. 798-805). Washington, DC, USA: IEEE Computer Society.
- Alfalou, A., y Brosseau, C. (2010). Understanding correlation techniques for face recognition: From basics to applications. En M. Oravec (Ed.), *Face recognition*. InTech.
- Alfalou, A., Brosseau, C., Katz, P., y Alam, M. S. (2012, May). Decision optimization for face recognition based on an alternate correlation plane quantification metric. *Opt. Lett.*, 37(9), 1562–1564.
- Alvarez-Garcia, J., Ortega, J., Gonzalez-Abril, y Velasco, F. (2010). Trip destination prediction based on past gps log using a hidden markov model. *Expert Systems with Applications*.
- An, L., Kafai, M., Member, S., y Bhanu, B. (2013, Junio). Dynamic bayesian network for unconstrained face recognition in surveillance camera networks. *IEEE Journal on Emerging and Selected Topics in Circuits and Systems*, 10(10), 1-10.
- An, L., Kafai, M., Yang, S., y Bhanu, B. (2013, Aug). Reference-based person re-identification. En *Advanced video and signal based surveillance (avss), 2013 10th IEEE International Conference on* (p. 244-249).
- Andreopoulos, A., y Tsotsos, J. K. (2013). 50 years of object recognition: Directions forward. *Computer Vision and Image Understanding*, 117(8), 827 - 891.
- Ao, M., Yi, D., Lei, Z., y Li, S. (2009). Face recognition at a distance: System issues. En M. Tistarelli, S. Li, y R. Chellappa (Eds.), *Handbook of remote biometrics* (p. 155-167). Springer London.
- Arjun, J., y Nair, M. (2014). Robust face recognition technique using gabor phase pattern and phase only correlation. *CSI Transactions on ICT*, 2(2), 85-95.
- Arya, S., Pratap, N., y Bhatia, K. (2015). Future of face recognition: A review. *Procedia Computer Science*, 58, 578-585.
- Ayedi, W., Snoussi, H., y Abid, M. (2011, Septiembre). A fast multi-scale covariance descriptor for object re-identification. *Pattern Recognition Letters*, 33(2012), 1902-1907.
- Babenko, B., y Belongie, M.-H. Y. S. (2011). Robust object tracking with online multiple instance learning..

- Banerjee, P.-K., y AsitK.Datta. (2013). Generalized regression neural network trained preprocessing of frequency domain correlation filter for improved face recognition and its optical implementation. *Optics and LaserTechnology*, 45, 217-227.
- Banerjee, P.-K., Cahndra, J.-K., y Datta, A.-K. (2010). A frequency domain face recognition technique based on correlation plane features as input to a regression neural network. *Procedia Computer Science*.
- Banerjee, P.-K., y Datta, A. K. (2012). An illumination tolerant class specific 2d subspace based face recognition technique using optimum correlation filter. *Optik*.
- Banerjee, P.-K., y Datta, A.-K. (2014). Class specific subspace dependent nonlinear correlation filtering for illumination tolerant face recognition. *Pattern Recognition Letters*, 36, 177-185.
- Bauml, M., Bernardin, K., Fischer, M., Ekenel, H., y Stiefelhagen, R. (2010, Aug). Multi-pose face recognition for person retrieval in camera networks. En *Advanced video and signal based surveillance (avss), 2010 seventh ieee international conference on* (p. 441-447).
- Bedagkar-Gala, A., y Shah, S. K. (2011, Septiembre). Part-based spatio-temporal model for multi-person re-identification. *Pattern Recognition Letters*, 33(2012), 1908-1915.
- Bedagkar-Gala, A., y Shah, S.-K. (2014). A survey of approaches and trends in person re-identification. *Image and Vision Computing*.
- Berdugo, G., Soceanu, O., Moshe, Y., Rudoy, D., y Dvir, I. (2010). Object reidentification in real world scenarios across multiple non-overlapping cameras. En *European signal processing conference* (p. 1806-1810).
- Bolme, D.-S. (2011). *Theory and applications of optimized correlation output filters* (Tesis Doctoral no publicada). Colorado State University.
- Bolme, D. S., Beveridge, J. R., y Lui, B. A. D. Y. M. (2010). Visual object tracking using adaptive correlation filters. *IEEE Conference on Computer Vision and Pattern Recognition*, 2544 - 2550.
- Bolme, D. S., Draper, B. A., y Beveridge, J. R. (2009). Average of synthetic exact filters. *IEEE Conference on Computer Vision and Pattern Recognition*, 2105 - 2112.
- Burton, A. M., Jenkins, R., Hancock, P. J., y White, D. (2005). Robust representations for face recognition: The power of averages. *Cognitive Psychology*, 51(3), 256 - 284.
- Caro, G. D., y Dorigo, M. (1998). Antnet: Distributed stigmergetic control for communications networks. *Journal of Artificial Intelligence Research*.
- Casasent, D., y Chang, W.-T. (1986). Correlation synthetic discriminant functions. *Appl. Opt.*, 25, 2343-2350.
- Chan, T.-S., y Kumar, A. (2011, Noviembre). Reliable ear identification using 2-d quadrature filters. *Pattern Recognition Letters*, 33(2012), 1870-1881.
- Chellappa, R., Ni, J., y Patel, V. M. (2011, Diciembre). Remote identification of faces: Problems, prospects, and progress. *Pattern Recognition Letters*, 33(2012), 1849-1859.

- Chen, J., Meng, X., Guo, Y., Grumbach, S., y Sun, H. (2006). Modeling and predicting future trajectories of moving objects in a constrained network. *Proceedings of the 7th International Conference on Mobile Data Management*.
- Chen, L., Lv, M., y Chen, G. (2010). A system for destination and future route prediction based on trajectory mining. *Pervasive and Mobile Computing*.
- Chen, R., Yan, Y., Liu, M., y Wang, H. (2014). Comparative study: Face recognition via the correlation filter technique. En *Proceedings of international conference on internet multimedia computing and service* (pp. 251:251–251:255). New York, NY, USA: ACM.
- Chen, X., Huang, K., y Tan, T. (2013, Junio). Object tracking across non-overlapping views by learning inter-camera transfer models. *Pattern Recognition*, 47(2014), 1126-1137.
- Choi, J.-Y., Ro, Y.-M., y Plataniotis, K.-N. (2011). A comparative study of preprocessing mismatch effects in color image based face recognition. *Pattern Recognition*, 44, 412-430.
- Choudhary, K., y Goel, N. (2013). A review on face recognition techniques. En *International conference on communication and electronics system design*.
- Cornejo, A., Dornhaus, A., Lynch, N., y Nagpal, R. (2014). Distributed computing: 28th international symposium, disc 2014, austin, tx, usa, october 12-15, 2014. proceedings. En F. Kuhn (Ed.), (pp. 46–60). Berlin, Heidelberg: Springer Berlin Heidelberg.
- Crihalmeanu, S., y Ross, A. (2011, Noviembre). Multispectral scleral patterns for ocular biometric recognition. *Pattern Recognition Letters*, 33(2012), 1860-1869.
- Davami, E., y Sukthankar, G. (2012). Online learning of user-specific destination prediction models. *International Conference on Social Informatics*.
- Davidson, M. W., y Abramowitz, M. (2014). *Molecular expressions microscopy primer: Digital image processing-difference of gaussians edge enhancement algorithm*. Olympus America Inc., and Florida State University.
- Degtyarev, N., y Seredin, O. (2010). *Comparative testing of face detection algorithms* (Inf. Téc.). Tula State University.
- de Oliveira Junior, L. L., y Thomaz, C. E. (2006, june). *Captura e alinhamento de imagens: Um banco de faces brasileiro* (Inf. Téc.). Department of Electrical Engineering, FEI, Sao Bernardo do Campo, São Paulo, Brazil,.
- Diaz-Ramirez, V., Cuevas, A., Kober, V., Trujillo, L., y Awwal, A. (2014). Pattern recognition with composite correlation filters designed with multi-objective combinatorial optimization. *Optics Communications*, 338, 77 - 89.
- Ding, C., y Tao, D. (2015). A comprehensive survey on pose-invariant face recognition. *CoRR*, abs/1502.04383.
- Doretto, G., Sebastian, T., Tu, P., y Rittscher, J. (2011, Enero). Appearance-based person reidentification in camera networks: problem overview and current approaches. *Ambient Intell Human Comput*, 2(2011), 127-151.
- Dorigo, M., y Stützle, T. (2004). *Ant colony optimization*. Scituate, MA, USA: Bradford Company.

- Eisner, J., Funke, S., Herbsty, A., Spillnery, A., y Storandt, S. (2011). Algorithms for matching and predicting trajectories. *OJSA*.
- Elbouz, M., Alfalou, A., Brosseau, C., Yahia, N. B. H., y Alam, M. (2015). Assessing the performance of a motion tracking system based on optical joint transform correlation. *Optics Communications*, 349, 65 - 82.
- Filev, D., Tseng, F., Kristinsson, J., y McGee, R. (2011). Contextual on-board learning and prediction of vehicle destinations. *IEEE*.
- Fischer, M., Ekenel, H., y Stiefelhagen, R. (2010, June). Interactive person re-identification in tv series. En *Content-based multimedia indexing (cbmi), 2010 international workshop on* (p. 1-6).
- Gandhi, T., y Trivedi, M. M. (2007, Febrero). Person tracking and reidentification: Introducing panoramic appearance map (pam) for feature representation. *Machine Vision and Applications*, 18(2007), 207-220.
- Georghiades, A., Belhumeur, P., y Kriegman, D. (2001a). From few to many: Illumination cone models for face recognition under variable lighting and pose. *IEEE Trans. Pattern Anal. Mach. Intelligence*, 23(6), 643-660.
- Georghiades, A., Belhumeur, P., y Kriegman, D. (2001b). From few to many: Illumination cone models for face recognition under variable lighting and pose. *IEEE Trans. Pattern Anal. Mach. Intelligence*, 23(6), 643-660.
- Goldstein, A., Harmon, L. D., y Lesk, A. (1971, May). Identification of human faces. *Proceedings of the IEEE*, 59(5), 748-760.
- Gong, S., Cristani, M., Loy, C., y Hospedales, T. (2014). The re-identification challenge. En S. Gong, M. Cristani, S. Yan, y C. C. Loy (Eds.), *Person re-identification* (p. 1-20). Springer London.
- Gong, S., McKenna, S. J., y Psarrou, A. (2000). Dynamic vision: From images to face recognition. En S. Gong, S. J. McKenna, y A. Psarrou (Eds.), (1.^a ed., p. 3-6). London, UK, UK: Imperial College Press.
- Gupta, S., Kapoor, S., y Gupta, P. (2012, Marzo). Synthesis of a face image at a desired pose from a given pose. *Pattern Recognition Letters*, 33(2012), 1942-1950.
- Hamdoun, O., Moutarde, F., Stanculescu, B., y Steux, B. (2008, Septiembre). Person re-identification in multi-camera system by signature based on interest point descriptors collected on short video sequences. En *2nd acm/ieee international conference on distributed smart cameras*. Stanford, Palo Alto, United States.
- Hansen, M.-F., y Atkinson, G.-A. (2010). Biologically inspired 3d face recognition from surface normals. En *Proceedings of the international conference and exhibition on biometrics technology* (Vol. 2, p. 26-34).
- Harvey, I. (1992). The saga cross: The mechanics of recombination for species with variable-length genotypes. *Parallel Problem*, 269-278.
- Hassebrook, L.-G., y Rahmati, M. (1993). Training set selection with multiple out-of-plane rotation parameters. En *Optical pattern recognition iv* (Vol. 1959, p. 32-42).
- Heo, J., Savvides, M., y Vijayakumar, B. (2005). Performance evaluation of face recognition using visual and thermal imagery with advanced correlation filters. *Computer Society Conference on Computer Vision and Pattern Recognition*.

- Hjelmas, E. (2001). Face detection: A survey. *Computer Vision and Image Understanding*, 83, 236-274.
- Holdren, J.-P. (2011). *The national biometrics challenge* (Inf. Téc.). National Science and Technology Council Subcommittee on Biometrics and Identity Management.
- hsun Chang, T., y Gong, S. (2001). Tracking multiple people with a multi-camera system. En *Ieee workshop on multi-object tracking*.
- Huang, X., Z.-Li, S., y Wang, Y. (2004, Dec). Shape localization based on statistical method using extended local binary pattern. En *Multi-agent security and survivability, 2004 ieee first symposium on* (p. 184-187).
- Javed, O., Rasheed, Z., Shafique, K., y Shah, M. (2003, Oct). Tracking across multiple cameras with disjoint views. En *Computer vision, 2003. proceedings. ninth ieee international conference on* (p. 952-957 vol.2).
- Jiang, F., Fischer, M., Ekenel, H. K., y Shi, B.-E. (2013). Combining texture and stereo disparity cues for real-time face detection. *Signal Processing: Image Communication*, 28, 1100-1113.
- Kang, W. (2012, Marzo). Vein pattern extraction based on vectorgrams of maximal intra-neighbor difference. *Pattern Recognition Letters*, 33(2012), 1926-1923.
- Karaman, S., Lisanti, G., Bagdanov, A.-D., y Del-Bimbo, A. (2014). Leveraging local neighborhood topology for large scale person re-identification. *Pattern Recognition*.
- Kenk, V.-S., Mandeljc, R., Kovacic, S., Kristan, M., Hajdinjak, M., y Pers, J. (2014). Visual re-identification across large, distributed camera networks. *Image and Vision Computing*.
- Kennedy, J., y Eberhart, R. (1995, Nov). Particle swarm optimization. En *Neural networks, 1995. proceedings., ieee international conference on* (Vol. 4, p. 1942-1948 vol.4). doi: 10.1109/ICNN.1995.488968
- Kiat-Ng, C. (2005). *Pda face recognition system using advanced correlation filters* (Tesis Doctoral no publicada). Carnegie mellon university.
- Kostov, V., Ozawa, J., Yoshioka, M., y Kudoh, T. (2005). Travel destination prediction using frequent crossing pattern from driving history. *IEEE Conference on Intelligent Transportation Systems*.
- Krumm, J. (2006). Real time destination prediction based on efficient routes. *SAE International*.
- Kumar, B. V., Carlson, D. W., y Mahalanobis, A. (1994, Oct). Optimal trade-off synthetic discriminant function filters for arbitrary devices. *Opt. Lett.*, 19(19), 1556-1558.
- Kun, B., Lin, L., Li, Z., y Liang, F.-S. (2012). 3d face recognition method based on cascade classifier. *Procedia Engineering*, 29, 705-709. (2012 International Workshop on Information and Electronics Engineering)
- Lab, A. M. P. (2001). *CMU face expression database*. Descargado de <http://chenlab.ece.cornell.edu/projects/FaceAuthentication/>
- Lai, H., Ramanathan, V., y Wechsler, H. (2008). Reliable face recognition using adaptive and robust correlation filters. *Computer Vision and Image Understanding*.

- Lee, J. (2015, Diciembre). *Msu licenses facial recognition system to nec*. Web. <http://www.biometricupdate.com>.
- Leonard, I., Alfalou, A., y Brosseau, C. (2012, May). Spectral optimized asymmetric segmented phase-only correlation filter. *Appl. Opt.*, *51*(14), 2638–2650.
- Li, S.-Z., y JianxinWu. (2011). Handbook of face recognition. En S.-Z. Li y A.-K. Jain (Eds.), (p. 277-303). Springer.
- Li, W., Wu, Y., Mukunoki, M., Kuang, Y., y Minoh, M. (2015). Locality based discriminative measure for multiple-shot human re-identification. *Neurocomputing*.
- Liebowitz, S. (1989). *Optical range image processing for object recognition* (Tesis Doctoral no publicada). Carnegie Mellon University.
- Linga-Reddy, K.-R., Babub, G.-R., y Kishore, L. (2010). Face recognition based on eigen features of multi scaled face components and an artificial neural network. *Procedia Computer Science*, *2*, 62-74.
- Liu, C., Gong, S., y Loy, C. C. (2013). On-the-fly feature importancemining for person re-identification. *Pattern Recognition*.
- Liu, J. S. (2001). *Monte carlo strategies in scientific computing*. Springer.
- Ma, B., Sub, Y., y Jurie, F. (2014). Covariance descriptor based on bio-inspired features for person re-identification and face verification. *Image and Vision Computing*.
- Mahalanobis, A., Kumar, B. V. K. V., y Casasent, D. (1987, Sep). Minimum average correlation energy filters. *Appl. Opt.*, *26*(17), 3633–3640.
- Mahalanobis, A., Vijaya-Kumar, B., Song, S., Sims, S., y Epperson, J. (1994). Unconstrained correlation filters. *Applied Optics*, *33*, 3751-3759.
- Majed, S., y Arof, H. (2008). *Pattern correlation approach towards face detection system framework*.
- Marsico, M. D., Doretto, G., y Riccio, D. (2012, Marzo). M-vivie: A multi-thread video indexer via identity extraction. *Pattern Recognition Letters*, *33*(2012), 1882-1890.
- Mazzon, R., Tahir, S. F., y Cavallaro, A. (2012, Marzo). Person re-identification in crowd. *Pattern Recognition Letters*, *33*(2012), 1828-1837.
- Mirjalili, S., Mirjalili, S. M., y Lewis, A. (2014). Grey wolf optimizer. *Advances in Engineering Software*, *69*, 46 - 61.
- Mittal, A., y Davis, L. S. (2003). M 2 tracker: a multi-view approach to segmenting and tracking people in a cluttered scene. *International Journal of Computer Vision*, *51*, 189 - 203.
- Montgomery, D. C. (2013). Design and analysis of experiments. En L. Ratts y A. Melhorn (Eds.), (p. 183-233). Wiley.
- Myers, A., y Megherbi, D. (2014, May). An efficient computational intelligence technique for affine-transformation-invariant image face detection, tracking, and recognition in a video stream. En *Computational intelligence and virtual environments for measurement systems and applications (civemsa), 2014 ieee international conference on* (p. 88-93).
- Nadembega, A., Taleb, T., y Hafid, A. (2012). A destination prediction model based on historical data, contextual knowledge and spatial conceptual maps. *Communication QoS, Reliability and Modeling Symposium*.

- Nakahara, F., y Murakami, T. (2012). A destination prediction method based on behavioral pattern analysis of nonperiodic position logs. *Information Processing Society of Japan*.
- Nanni, L., Munaro, M., Ghidoni, S., Menegatti, E., y Brahnam, S. (2015). Ensemble of different approaches for a reliable person re-identification system. *Applied Computing and Informatics*.
- Nappi, M., y Wechsler, H. (2012, Febrero). Robust re-identification using randomness and statistical learning: Quo vadis. *Pattern Recognition Letters*, 33(2012), 1820-1827.
- Naresh-Boddeti, V. (2012). *Advances in correlation filters: Vector features, structured prediction and shape alignment* (Tesis Doctoral no publicada). Carnegie Mellon University.
- Ángel González-Fraga, J., Kober, V., y Álvarez Borrego, J. (2006). Adaptive sdf filters for recognition of partially occluded objects. En *Application of digital image processing xxix*.
- Nocedal, J., y Wright, S.-J. (1999). Numerical optimization. En P. Glynn y S.-M. Robinson (Eds.), (p. 7-8). Springer.
- North, D.-. (1963). An analysis of the factors which determine signal/noise discriminations in pulsed carrier systems. En *Proc. ieee* (Vol. 51, p. 1016-1027).
- Ojala, T., Pietikainen, M., y Maenpaa, T. (2002, Jul). Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 24(7), 971-987.
- Painkras, E., y Charoensak, C. (2005, Nov). A framework for the design and implementation of a dynamic face tracking system. En *Tencon 2005 2005 ieee region 10* (p. 1-6).
- Pinto-Fernández, S., y Díaz-Ramírez, V. (2012). Improvement of facial recognition with composite correlation filters designed with combinatorial optimization. En *Optics and photonics for information processing*.
- Prakash, S., y Gupta, P. (2012, Marzo). A rotation and scale invariant technique for ear detection in 3d. *Pattern Recognition Letters*, 33(2012), 1924-1931.
- Prosser, B., Zheng, W.-S., Gong, S., y Xiang, T. (2010). Person re-identification by support vector ranking. En *Proceedings of the british machine vision conference* (pp. 21.1–21.11). BMVA Press. (doi:10.5244/C.24.21)
- Ptucha, R., Rhoda, D., y Mittelstaedt, B. (2013). Auto zoom crop from face detection and facial features. En C.-A. Bouman, I. Pollak, y P.-J. Wolfe (Eds.), *Computational imaging xi* (Vol. 8657). Electronic Imaging.
- Qian, J., Ma, S., Xu, Y., Li, X., y Shen, Y. (2012). Application of binocular vision system to face detection and tracking in service robot. En Q.-D. Hai-Guo (Ed.), *International workshop on image processing and optical engineering* (Vol. 8335).
- Qiao, L., Chen, S., y Tan, X. (2010). Sparsity preserving projections with applications to face recognition. *Pattern Recognition*, 43(1), 331 - 341.
- Robertson, D.-J., S.-Kramer, R.-S., y Burton, A.-M. (2015). Face averages enhance user recognition for smartphone security. *PLoS ONE*.

- Rodriguez-Perez, A.-F. (2012). *Maximum margin correlation filters* (Tesis Doctoral no publicada). Carnegie Mellon University.
- Ross, D., Lim, J., Lin, R.-S., y Yang, M.-H. (2011). Descargado de <http://www.cs.toronto.edu/~dross/ivt/>
- Ross, D. A., Lim, J., Lin, R.-S., y Yang, M.-H. (2008, mayo). Incremental learning for robust visual tracking. *Int. J. Comput. Vision*, 77(1-3), 125–141.
- Roy, A., Sural, S., y Mukherjee, J. (2012, Febrero). A hierarchical method combining gait and phase of motion with spatiotemporal model for person re-identification. *Pattern Recognition Letters*, 33(2012), 1891-1901.
- Santiago-Ramírez, E., González-Fraga, J.-A., y Lázaro-Martínez, S. (2012). Face recognition and tracking using unconstrained non-linear correlation filters. *Procedia Engineering*, 35, 192-201. (International Meeting of Electrical Engineering Research 2012.)
- Satta, R., Fumera, G., y Roli, F. (2012, Abril). Fast person re-identification based on dissimilarity representations. *Pattern Recognition Letters*, 33(2012), 1838-1848.
- Savvides, M., Bhagavatula, R., hui Li, Y., y Abiantun, R. (2007). Face recognition. En K. Delac y M. Grgic (Eds.), (p. 495-526). InTech.
- Savvides, M., Venkataramani, K., y Kumar, B. V. (2003). Incremental updating of advanced correlation filters for biometric authentication systems. *International Conference on Multimedia and Expo*, 3, 229-232.
- Savvides, M., Vijaya-Kumar, B., y Khosla, P. (2002). Face verification using correlation filters. *Carnegie Mellon University*.
- Saxena, V., Grover, S., y Joshi, S. (2008, Sept). A real time face tracking system using rank deficient face detection and motion estimation. En *Cybernetic intelligent systems, 2008. cis 2008. 7th ieee international conference on* (p. 1-6).
- Schroff, F., Kalenichenko, D., y Philbin, J. (2015). Facenet: A unified embedding for face recognition and clustering. *CoRR*, abs/1503.03832.
- Schwefel, H. P. (1995). *Evolution and optimum seeking* (1.^a ed.; H. P. Schwefel, Ed.). New York: Wiley Interscience.
- Scott, D. M., y He, S. Y. (2012). Modeling constrained destination choice for shopping: a gis-based, time-geographic approach. *Journal of Transport Geography*.
- Shi, S.-C., Guo, C.-C., Lai, J.-H., Chen, S.-Z., y Hu, X.-J. (2015). Person re-identification with multi-level adaptive correspondence models. *Neurocomputing*.
- Shinmura, T., Zhu, D., Ota, J., y Fukazawa, Y. (2014). Destination prediction considering both tweet contents and location transition history. *Seventh International Conference on Mobile Computing and Ubiquitous Networking*.
- Sirovich, L., y Kirby, M. (1987, Mar). Low-dimensional procedure for the characterization of human faces. *J. Opt. Soc. Am. A*, 4(3), 519–524.
- Soori, U., Yuen, P., Han, J. W., Ibrahim, I., Chen, W., Hong, K., ... Richardson, M. (2013, Abril). Target recognitions in multiple-camera closed-circuit television using color constancy. *Optical Engineering*, 52(4), 1-14.
- Stallkamp, J., Ekenel, H. K., y Stiefelhagen, R. (2007). *Video-based face recognition on real-world data*.

- Tanaka, K., Kishino, Y., Terada, T., y Nishio, S. (2009). A destination prediction method using driving contexts and trajectory for car navigation systems. *ACM*.
- Toyama, K., y Horvitz, E. (2000). Bayesian modality fusion: Probabilistic integration of multiple vision algorithms for head tracking. En *Fourth asian conference on computer vision (accv)*.
- Turk, M., y Pentland, A. (1991, Jun). Face recognition using eigenfaces. En *Computer vision and pattern recognition* (p. 586-591).
- Utsumi, A., y Tetsutani, N. (2004, May). Human tracking using multiple-camera-based head appearance modeling. En *Automatic face and gesture recognition, 2004. proceedings. sixth ieee international conference on* (p. 657-662).
- Vijaya-Kumar, B. (1992). Tutorial survey of composite filter design for optical correlators. *Applied Optics*, 31, 4773-4801.
- Vijaya-Kumar, B., Mahalanobis, A., y Juday, R. (2005). *Correlation pattern recognition* (1.^a ed.; B. Vijaya-Kumar, A. Mahalanobis, y R. Juday, Eds.). New York: Cambridge University Press.
- Viola, M., Jones, M. J., y Viola, P. (2003). Fast multi-view face detection. En *Proc. of computer vision and pattern recognition*.
- Viola, P., y Jones, M. (2001). Robust real-time object detection. *Second International Workshop on Statistical and Computational Theories of Vision-Modeling, Learning, Computing, and Sampling*, 1-25.
- Vu, N.-S., y Caplier, A. (2009, Nov). Illumination-robust face recognition using retina modeling. En *Image processing (ictp), 2009 16th ieee international conference on* (p. 3289-3292). doi: 10.1109/ICIP.2009.5413963
- Wang, G., Alam Bhuiyan, M., Cao, J., y Wu, J. (2014, April). Detecting movements of a target using face tracking in wireless sensor networks. *Parallel and Distributed Systems, IEEE Transactions on*, 25(4), 939-949.
- Wang, Y.-Q. (2013). An analysis of viola-jones face detection algorithm. *Image Processing On Line*.
- Webster, M. (2015). *face definition*.
- White, D., Phillips, P.-J., Hahn, C.-A., Hill, M., y OToole, A.-J. (2015). Perceptual expertise in forensic facial image comparison. *Proceedings Royal Society B*.
- Wiskott, L., Fellous, J.-M., Kruger, N., y V.-Malsburg, C. (1997). Face recognition by elastic bunch graph matching. *IEEE Trans. Pattern Anal. Mach. Intell.*, 19(7), 775-779.
- Xu, Y., Zhang, X., y Gai, H. (2011). Quantum neural networks for face recognition classifier. *Procedia Engineering*, 15, 1319-1323.
- Xue, A. Y., Zhang, R., Zheng, Y., Xie, X., Huang, J., y Xu, Z. (2013). Destination prediction by sub-trajectory synthesis and privacy protection against such prediction. *IEEE International Conference on Data Engineering*.
- Yan, Y., Wang, H., Li, C., Yang, C., y Zhong, B. (2013). An effective unconstrained correlation filter and its kernelization for face recognition. *Neurocomputing*, 119, 201-211.

- Yan, Y., Wang, H., y Suter, D. (2014). Multi-subregion based correlation filter bank for robust face recognition. *Pattern Recognition*, 47, 3487-3501.
- Yang, B., y Chen, S. (2013). A comparative study on local binary pattern (LBP) based face recognition: LBP histogram versus LBP image. *Neurocomputing*, 120, 365-379.
- Yang, M.-H., Kriegman, D.-J., y Ahuja, N. (2002). Detecting faces in images: a survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24, 34-58.
- Yang, X.-S. (2014a). Nature inspired optimization algorithms. En X.-S. Yang (Ed.), *Nature-inspired optimization algorithms* (p. 67 - 75). Oxford: Elsevier.
- Yang, X.-S. (2014b). Nature inspired optimization algorithms. En X.-S. Yang (Ed.), *Nature-inspired optimization algorithms* (p. 67 - 75). Oxford: Elsevier.
- Yeom, S., y Woo, Y.-H. (2013). Person-specific face detection in a scene with optimum composite filtering and colour-shape information. *International Journal of Advanced Robotic Systems*, 10(70), 1-8.
- Yi, X. (2014). Face recognition of illumination tolerance in 2d subspace based on the optimum correlation filter. *Journal of Multimedia*, 9(1), 8391.
- Yin, S., Dai, X., Ouyang, P., Liu, L., y Wei, S. (2014). A multi-modal face recognition method using complete local derivative patterns and depth maps. *Sensors*, 14, 19561-19581.
- Yoder, J., Medeiros, H., Park, J., y Kak, A. C. (2010, octubre). Cluster-based distributed face tracking in camera networks. *Trans. Img. Proc.*, 19(10), 2551-2563.
- Zappa, E., Mazzoleni, P., y Hai, Y. (2010). Stereoscopic based 3d face recognition system. *Procedia Computer Science*, 1(1), 2521-2528.
- Zhang, C., y Zhang, Z. (2010). *A survey of recent advances in face detection* (Inf. Téc.). Microsoft Research.
- Zhang, S., Zhang, H., Zhang, L., y Xue, Y. (2010). A multi-scale-based super-resolution method for face image. *International Congress on Image and Signal Processing*(3).
- Zhang, X., Hu, W., Maybank, S., y Li, X. (2007, Oct). Graph based discriminative learning for robust and efficient object tracking. En *Computer vision, 2007. iccv 2007. ieee 11th international conference on* (p. 1-8).
- Zhao, W., Chellapa, R., Phillips, P., y Rosenfeld, A. (2003). Face recognition: A literature survey. *ACM Computing Surveys*, 35(4), 399-458.
- Zhou, T., Qi, M., Jiang, J., Wanga, X., Hao, S., y Jin, Y. (2014). Person re-identification based on nonlinear ranking with difference vectors. *Information Sciences*.
- Zhu, X., Liao, S., Lei, Z., Liu, R., y Li, S. Z. (2007). Feature correlation filter for face recognition. En S.-W. Lee y S. Li (Eds.), *Advances in biometrics* (Vol. 4642, p. 77-86). Springer Berlin Heidelberg.
- Zitzler, E., Laumanns, M., y Thiele, L. (2001). *Spea2: Improving the strength pareto evolutionary algorithm* (Inf. Téc. n.º 103). Swiss Federal Institute of Technology.