CENTRO DE INVESTIGACIÓN CIENTÍFICA Y DE EDUCACIÓN SUPERIOR DE ENSENADA

PROGRAMA DE POSGRADO EN CIENCIAS EN CIENCIAS DE LA COMPUTACIÓN

Filtros morfológicos adaptativos para el reconocimiento de caracteres en imágenes degradadas de documentos

Tesis

para cubrir parcialmente los requisitos necesarios para obtener el grado de

Maestro en Ciencias

Presenta:

Julia Diaz Escobar

Ensenada, Baja California, México

2014

Tesis defendida por

Julia Diaz Escobar

y aprobada por el siguiente comité

Dr.Vitaly Kober

Director del Comité

Dr. Hugo Hidalgo Silva

Miembro del Comité

Dr. Josué Álvarez Borrego

Miembro del Comité

Dra. Ana Isabel Martínez García

Coordinador del

Posgrado en Ciencias de la Computación

Dr. Jesús Favela Vara

Director de Estudios de Posgrado

Octubre, 2014

Resumen de la tesis que presenta **Julia Diaz Escobar** como requisito parcial para la obtención del grado de Maestro en Ciencias en Ciencias de la Computación.

Filtros morfológicos adaptativos para el reconocimiento de caracteres en imágenes degradadas de documentos

Resumen elaborado por:

Julia Diaz Escobar

El Reconocimiento Óptico de Caracteres (OCR) en imágenes de documentos impresos digitalizados por medio de escáner es un tema muy estudiado, en donde las condiciones de captura tales como: la posición de la hoja, la iluminación, el contraste, la resolución, etc., suelen ser controladas y por lo tanto óptimas. En la actualidad se han propuesto diversos algoritmos para el reconocimiento de caracteres y existen diferentes sistemas OCR comerciales o de código libre (ABBYY, Tesseract, etc.), que tienen un buen desempeño. Sin embargo, hoy en día es más práctico utilizar un dispositivo móvil para la digitalización de un documento que el uso de un escáner; como consecuencia, la calidad de la imagen del documento se ve afectada, presentando distintas distorsiones geométricas, iluminación no homogénea, baja resolución, etc., y por lo tanto disminuyendo el desempeño de los sistemas OCR. Así que, para solucionar estos problemas, este trabajo propone el uso de varios filtros compuestos adaptativos basado en dos enfoques. El primer enfoque utilizado se basa en la descomposición por umbral y una correlación morfológica invariante a iluminación. Los filtros compuestos se basan en filtros de funciones discriminantes sintéticas no lineales diseñados mediante la incorporación de información de un conjunto de imágenes de entrenamiento y un valor dado de capacidad discriminación deseado. Para aquellos caracteres similares usamos un enfoque diferente basado en el bloqueo de los componentes espectrales de un filtro de sólo fase. Finalmente, los resultados obtenidos de las simulaciones realizadas con el sistema OCR propuesto se presentan y se comparan con el software comercial ABBYY, la comparación se hace midiendo los errores de clasificación.

Palabras Clave: **Descomposición por umbral, correlación morfológica, filtros no lineales.**

Abstract of the thesis presented by **Julia Diaz Escobar** as a partial requirement to obtain the Master in Sciences degree in Computer Science with orientation in

Adaptive morphological filtering for degraded image document character recognition

Abstract by:

Julia Diaz Escobar

Optical Character Recognition (OCR) in scanned printed documents is a well-studied task, where the captured conditions like sheet position, illumination, contrast, resolution, etc., are controlled. Many algorithms have been proposed and there are different systems, ABBYY for example, which have good performance. However, nowadays it is more practical to use a mobile device for document capture than using a scanner, as a consequence, the quality of the document images is often affected by geometric distortions, non-homogeneous illumination, low resolution, etc., and hence decreasing the performance of OCR engines. So, to better deal with these problems, this work propose to use multiple adaptive composite filters based on two different approaches for improvement of the detection and recognition performance. The first approach is based on threshold decomposition and an illumination-invariant morphological correlation. The composite filters are based on nonlinear and designed by incorporating information from a set of training images and a given value of discrimination capability. And, for those similar characters we use a different approach based on zero-masking of spectral components of a phase-only filter. Finally computed simulation results obtained with the proposed OCR system are presented and compared with those of the commercial software ABBYY, the comparison is made by counting the clasification errors.

Keywords: **Image Processing, pattern recognition, morphological filtering, binarization.**

Dedicatoria

A mi madre Ernestina, por todo su apoyo y sacrificio.

Agradecimientos

A mi hermana y sobrinas, por el apoyo que me brindaron.

A mi asesor Dr. Vitaly Kober por la oportunidad, paciencia y consejos brindados.

A Ricardo Cuesta por su amistad, tiempo y ayuda en cada duda a lo largo de mi carrera y maestría.

A los miembros de mi comité de tesis, Dr. Hugo Hidalgo y Dr. Josué Álvarez, por sus valiosos comentarios.

A los buenos maestros que me han enseñado y corregido a lo largo de mi vida.

A Daniel Miramontes por contestar todas mis dudas.

A mis amigas por estar conmigo en las buenas (fiestas) y entender cuando estoy ocupada.

Al CONACyT por la beca otorgada para el desarrollo de este trabajo.

Tabla de Contenido

Página

Tabla de Contenido (continuación)

Página

Tabla de Contenido (continuación)

Tabla de Contenido (continuación)

Página

Lista de Figuras

Lista de Figuras (continuación)

Lista de Figuras (continuación)

Lista de Figuras (continuación)

Lista de Tablas

Capítulo 1. Introducción

Hoy en día vivimos en la llamada era digital, donde el creciente número de computadoras y su fácil acceso hacen posible el intercambio de grandes cantidades de información de manera rápida y eficiente, permitiendo así su manipulación y almacenamiento en muy poco espacio. El Reconocimiento Óptico de Caracteres (OCR) se refiere al proceso de convertir imágenes de documentos impresos en archivos editables, mediante un proceso de digitalización, procesamiento y reconocimiento de caracteres.

En las últimas décadas, se han propuesto distintos sistemas OCR para imágenes de documentos digitalizados mediante el uso de un escáner, en el cual se controlan ciertas condiciones tales como la iluminación, posición, ruido y resolución. Entre los sistemas más populares se encuentran: ABBYY[1], Tesseract[2], OCRopus, entre otros. Sin embargo, hoy en día es más práctico utilizar dispositivos móviles como cámaras digitales, PC-cam, teléfonos celulares, tabletas, etc., para la captura de documentos que el uso de un escáner. Esto debido a que los dispositivos móviles pueden ser utilizados en cualquier parte, compartir información al mismo tiempo a través de Internet y a su bajo costo comparado con los escáneres tradicionales. Por otro lado la calidad de las imágenes de los dispositivos móviles a menudo suele ser pobre debido a la presencia de iluminación no homogénea, la resolución de la cámara, el ruido del sensor, etc., que afectan el desempeño de los sistemas OCR comúnes (Liang *et al.*, 2005; Doermann *et al.*, 2003).

Uno de los primeros esfuerzos para mejorar el desempeño de los sistemas OCR es preprocesar la imagen de entrada haciendo uso de técnicas de binarización, segmentación, detección de inclinación o sesgo, mejora de resolución, etc., sin embargo en algunos casos esto no es suficiente, ya que muchas técnicas no funcionan con imágenes contaminadas con ruido o distorsiones.

Los sistemas basados en extracción de características tales como Análisis de Componentes

[1] http://www.abbyy.com/
[2] http://code.google.com/p/tesseract-ocr/

Principales (PCA), Máquina de Soporte Vectorial (SVM) (Salehpour y Behrad, 2010), pero sobre todo Redes Neuronales (NN) (LeCun *et al.*, 1998; Simard *et al.*, 2003; Matei *et al.*, 2013; Kir *et al.*, 2013), entre otras, también han sido utilizados para el reconocimiento de caracteres; sin embargo, dichos sistemas pierden información importante durante la extracción de características y por lo tanto, producen errores de clasificación. Además, las redes neuronales requieren una complicada estructura de múltiples capas y necesitan un largo periodo de entrenamiento utilizando varios conjuntos de muestras.

En este trabajo se propone hacer uso de bancos de filtros compuestos no lineales adaptativos y filtros sólo fase para la detección y clasificación de los caracteres. El enfoque propuesto se basa en la descomposición por umbral, la correlación morfológica invariante a iluminación y la modificación del filtro sólo fase. Los filtros adaptativos compuestos fueron diseñados utilizando información de los caracteres y sus posibles distorsiones, así como de una típica escena de entrada; además en la etapa de entrenamiento, los filtros suelen alcanzar un determinado valor de capacidad de discriminación. Finalmente, los resultados obtenidos de las simulaciones realizadas por el sistema propuesto son presentados y discutidos.

1.1. Definición del problema

En los últimos años los sistemas OCR han sido muy populares y exitosos, sin embargo no todo está hecho. Cuando la digitalización se realiza mediante un escáner, los sistemas OCR (por ejemplo: ABBYY, Tesseract, etc.) suelen tener un buen desempeño; sin embargo, gracias al desarrollo de nuevas tecnologías, el uso de dispositivos móviles para la captura de imágenes de documentos ha sido mayor remplazando en su mayoría el uso de escáneres; pero con ello un nuevo problema ha surgido, ya que las imágenes de los dispositivos móviles suelen presentar degradaciones debido a la iluminación, perspectiva, resolución y componentes físicos del dispositivo, entre otros, dificultando el trabajo de los sistemas OCR actuales produciendo fallas en sus resultados y disminuyendo su precisión. En la actualidad los sistemas OCR cuentan con una precisión entre el 97 % en el mejor de los casos y suele disminuir hasta un 66 % en el peor. (Ozarslan y Eren, 2014) .

1.2. Justificación

Debido a la creciente disponibilidad y costo de los dispositivos móviles, el interés por sistemas OCR basados en dispositivos móviles ha aumentado debido a que pueden ser utilizados en lugares donde los escáneres tradicionales no.

Por lo que se requiere del diseño de nuevos algoritmos que sean capaces de reconocer caracteres de manera rápida y eficiente con el fin de no sólo realizar reconocimiento de caracteres, sino que también sean capaces de realizar reconocimiento de palabras completas mediante un análisis contextual y que la máquina sea capaz de reconocer y leer de la misma manera que el ser humano lo hace. Si se lograse esto, múltiples aplicaciones podrían ser desarrolladas, desde la automatización completa de la digitalización de documentos hasta el reconocimiento en tiempo real de texto para personas invidentes, por ejemplo.

1.3. Objetivo general

Desarrollar un sistema basado en filtros morfológicos adaptativos para el reconocimiento de caracteres en imágenes de documentos degradadas, con iluminación no uniforme y tolerante a ligeras distorciones geométricas en el que exista un grado de confiabilidad respecto a diversas métricas de desempeño.

1.4. Objetivos específicos

1. Estudiar las técnicas de reconocimiento de caracteres actuales.
2. Diseñar filtros morfológicos adaptativos compuestos capaces de reconocer caracteres tomando en cuenta ruido blanco e impulsivo, así como iluminación no uniforme y distorsiones geométricas.
3. Diseñar un sistema implementando los filtros morfológicos diseñados anteriormente.
4. Analizar el sistema propuesto respecto a iluminación no homogénea, distorsiones geométricas y ruido blanco e impulsivo.
5. Realizar un estudio comparativo entre el sistema propuesto y existentes en imágenes con

iluminación no homogénea, distorsiones geométricas, contaminadas con ruido blanco e impulsivo.

6. Evaluar estadísticamente el sistema propuesto en imágenes de prueba mediante el uso de intervalos de confianza, capacidad de discriminación y el número de errores (falsos positivos y falsos negativos).

7. Evaluar el sistema con imágenes reales.

1.5. Limitaciones y suposiciones

Muchos de los trabajos realizados anteriormente con respecto a reconocimiento de caracteres especifican el tipo de fuente o alfabeto. Para esta investigación se utilizará el alfabeto en inglés con 26 letras, sólo texto escrito a máquina y sin acentos, además de tomarse en cuenta los dígitos del 0 al 9.

Las imágenes presentarán ruido aditivo blanco e impulsivo, rotaciones no mayores de 15 grados, con escalamiento e iluminación no homogénea. Además, se supone que se cuenta con información *a priori* de los caracteres a reconocer, así como sus posibles distorsiones geométricas y las características estadísticas del ruido.

Se trabajará con imágenes en escala de grises, ya que existen diversos algoritmos capaces de convertir imágenes a color en monocromáticas, de ser necesario. Se supondrá que el sistema desarrollado se podrá adaptar para otros casos específicos.

1.6. Reconocimiento de caracteres

El proceso de OCR generalmente envuelve 3 pasos: pre-procesamiento, extracción de características y reconocimiento. A continuación se describe brevemente cada uno de ellos.

1.6.1. Pre-procesamiento

El objetivo del pre-procesamiento en los sistemas OCR es mejorar la calidad de la imagen y hacer más fácil el reconocimiento de los caracteres. El pre-procesamiento involucra distintas técnicas para la reducción de ruido, mejora del contraste, binarización, segmentación, mejora

de resolución, corrección de distorsiones geométricas, etc. Por ejemplo, en el trabajo de F. Shafait (Shafait *et al.*, 2008a) se presenta una técnica para retirar ruido marginal generado por componentes de la siguiente página, en un libro por ejemplo; además de variaciones de la posición de documentos digitalizados por medio de escáner y cámara digital. Esta técnica se basa en el reconocimiento del marco de la página mediante la aplicación del algoritmo de Voronoi (Kise *et al.*, 1998) utilizado para segmentar la imagen y extraer las líneas curvas de texto usando el enfoque de Ulges (Ulges *et al.*, 2005). Con la información obtenida del proceso anterior, se genera una función que optimiza la extracción del mínimo rectángulo que contiene al texto, suprimiendo todo aquello que se encuentre fuera de este marco y eliminado la variación debida a la posición del documento durante la captura de la imagen. Una vez realizado este proceso, se utiliza un sistema OCR comercial. Las limitantes de esta técnica surgen al momento de querer realizar la extracción de texto en imágenes que contienen fuente de tamaño grande o cuando el espacio entre palabras es muy amplio; además, no considera imágenes con bajo contraste o iluminación no uniforme.

Otra técnica comúnmente utilizada es la binarización de la imagen del documento. Entre las más comunes se encuentran la técnica global de Otsu (Otsu, 1975) la cual propone utilizar el histograma de la imagen para minimizar la varianza dentro de las clases y maximizar la varianza entre clases; la técnica local de Niblack (Niblack, 1985) la cual calcula un umbral por cada pixel de la imagen mediante el deslizado de una ventana rectangular sobre la imagen en escala de grises. El umbral es calculado utilizando la media m y desviación estándar s de todos los pixeles dentro de la ventana. Finalmente la técnica de Sauvola (Sauvola *et al.*, 1997) la cual es una modificación de la técnica de Niblack, con un mejor desempeño en documentos donde el fondo contiene ligera textura y variaciones en iluminación. La diferencia con la técnica de Niblack radica en la utilización del rango dinámico de la desviación estándar.

En los trabajos de B. Gatos *et al.* y J. He *et al.* (Gatos *et al.*, 2006; He *et al.*, 2005) se realiza una evaluación y comparación de las técnicas de binarización antes mencionadas entre otras, en imágenes degradadas de documentos con el fin de mejorar su calidad y obtener mejores resultados en el reconocimiento de caracteres. De igual forma, en el trabajo de W. Lund *et al.* (Lund *et al.*, 2013) realizan una combinación de técnicas de binarización y

posteriormente utilizan un sistema OCR comercial, con el fin de obtener mejores resultados de reconocimiento.

1.6.2. Reconocimiento por extracción de características

Una vez pre-procesada la imagen, el siguiente paso es realizar la extracción de características de la imagen, es decir, la extracción del menor número posible de características que capturen en lo mayor posible la escencia de la imagen. Dicho proceso suele ser el más importante, ya que no es tarea fácil saber cuáles características son las más importantes y depende en gran parte del conocimiento y dominio del problema y de la aplicación. Finalmente el conjunto de las mejores características previamente extraídas son utilizadas como la entrada de algún clasificador; entre los más utilizados se encuentran los clasificadores estadísticos, máquinas de soporte vectorial y redes neuronales. La elección del clasificador depende de la informacion *apriori* con la que se cuenta. Su objetivo es asignar las características a alguna clase previamente establecida.

O. Matei *et al.* (Matei *et al.*, 2013) proponen un sistema para el reconocimiento de dígitos en medidores de energía y/o gas mediante una cámara web. Este sistema funciona de la siguiente manera: se adquiere la imagen mediante la cámara web y se le realiza un preprocesamiento para convertir la imagen de color a imagen monocromática (escala de grises). Posteriormente, se realiza una ecualización del histograma para una mejor distribución de intensidades y finalmente se realiza una umbralización adaptativa para obtener el dígito. Una vez que la imagen se encuentra umbralizada, se procede a realizar una técnica llamada adelgazamiento la cual consiste en obtener los pixeles centrales del dígito y conectarlos mediante líneas para posteriormente obtener los ángulos entre segmentos adjuntos. Estos ángulos son los datos de entrada que se utilizan para entrenar una red neuronal (perceptrón multi-capa) y realizar el reconocimiento de los dígitos. Finalmente se confirma mediante la técnica del vecino más cercano[3], la cual utiliza casos positivos (verdaderos) y negativos (falsos) para el entrenamiento y clasifica una nueva muestra mediante el cálculo de la distancia más cercana a éstos.

[3]k-nearest neighbor

Por otro lado, el trabajo de Charles Jacobs *et al.* (Jacobs *et al.*, 2005) propone un método para el reconocimiento de caracteres en imágenes de documentos con baja resolución, utilizando una web-cam (1024×768). Asumen que el color del texto y el fondo es uniforme y que el texto es un tipo de fuente determinada. El sistema OCR propuesto está compuesto de dos partes: una red neuronal utilizada para predecir el caracter en cierta posición dada de la imagen de entrada y un reconocedor de palabras, el cual encuentra la palabra más probable en un rectángulo dado de la página. Finalmente el reconocimiento de palabras es un problema de optimización resuelto utilizando programación dinámica. Como resultados obtuvieron hasta un 87 % de éxito para el reconocimiento de letras, mientras que para el reconocimiento de números y letras obtuvieron un 83 % de éxito. Finalmente cuando incluyeron todo tipo de puntuaciones la precisión cayó a un 68 %.

B. Kir *et al.* (Kir *et al.*, 2013) realizan un sistema OCR utilizando la combinación de un conjunto de redes neuronales y el algoritmo de aprendizaje de correlación negativa. La combinación de un conjunto de redes neuronales es un sistema que utiliza más de una red neuronal para generar la solución de un mismo problema. Generalmente, cada red neuronal es entrenada de forma independiente, pero este tipo de estructura individual no contribuye a todo el sistema compuesto. El aprendizaje de correlación negativa (NCL) es una extensión del algoritmo de propagación hacia atrás, modificando la función de error y minimizándola durante el entrenamiento. La imagen se segmenta, se extrae y escala el caracter para finalmente ser clasificado por la ANN entrenada previamente obteniendo hasta un 83 % de éxito. El 17 % de error es debido a la resolución de la imagen y a una mala segmentación.

Finalmente, M. Salehpour y A. behrad (Salehpour y Behrad, 2010) proponen un sistema para el reconocimiento de dígitos manuscritos utilizando máquina de soporte vectorial. El sistema consiste en un pre-procesamiento de la imagen para convertirla en imagen binaria y mediante el uso de morfología matemática remover el ruido restante. Posteriormente se utilizan los algoritmos de PCA y PCA-LDA para la extracción de características que son utilizadas tanto para el entrenamiento como para la prueba. Finalmente se utiliza una máquina de soporte vectorial basada en pesos para la clasificación y reconocimiento de los dígitos.

1.6.3. Reconocimiento por apariencia

Un enfoque alternativo es aquel que se basa en la apariencia (valores de intensidad de la imagen). En este enfoque se utilizan los datos de entrenamiento en forma directa y no de características extraídas de ellos. La clasificación se hace mediante técnicas de correlación. El procedimiento general consiste en seleccionar una imagen de referencia para después determinar el grado de similitud que tiene con la escena observada. La imagen de referencia también se conoce como filtro. Los problemas de detección y localización se resuelven mediante correlación en dos pasos: se buscan los picos más altos en el plano de correlación tras aplicar el filtro, y posteriormente se utilizan las coordenadas de los picos como estimación de la posición de los objetos en la escena (Kumar *et al.*, 2005).

En este último enfoque se encuentra basado el siguiente trabajo de tesis. A continuación se presentan algunos de los trabajos previos relevantes.

1.7. Estado del arte

En el reconocimiento de objetos por correlación se involucran dos imágenes; una es la imagen de referencia, generalmente un filtro, y la otra una imagen de prueba, llamada escena de prueba. Para localizar y detectar la imagen de referencia dentro de la escena de prueba se aplica una operación de correlación de tal modo que si la imagen se encuentra en la escena de prueba, es de esperarse una correlación alta (Kumar *et al.*, 2005).

P. Maragos (Maragos, 1989) relaciona el criterio del error medio absoluto y una correlación no lineal llamada correlación morfológica. Además se prueba que este tipo de correlación produce picos de correlación mayores que los producidos por la correlación obtenida del error cuadrático medio.

En el trabajo de Y. Doh *et al.* (Doh *et al.*, 2004) se propone un algoritmo basado en la transformación sintética Hit-Miss (SHMT) para brindar invarianza a distorsiones en imágenes ruidosas. El elemento estructural está compuesto de imágenes del objetivo y objetos falsos basado en funciones discriminantes sintéticas. Primero definen el elemento estructural H_{SDF} (hit) como la combinación lineal de las imágenes de entrenamiento, y el elemento estructutal

M_{SDF} (miss) como la combinación lineal de las imágenes de los objetos a rechazar. Utilizando el elemento estructural hit, H_{SDF}, y el elemento estructural miss, M_{SDF}, se implementa una transformación sintética Hit-Miss.

En el documento de P. Maragos (Maragos, 2004) se presenta un resumen de técnicas para la mejora de imágenes haciendo uso de operaciones y filtros morfológicos, así como su relación con filtros de estadísticas de orden para imágenes en escala de grises.

Por otro lado, V. kober *et al.* (Kober *et al.*, 2001) proponen nuevos filtros no lineales haciendo uso de operaciones de estadísticas de orden sobre vecindarios espacialmente conectados. Además presentan algoritmos rápidos para la construcción de dichos filtros. Los filtros son evaluados obteniendo buenos resultados en imágenes contaminadas con ruido impulsivo, aditivo y una mezcla de ambos.

También los mismos autores (Kober *et al.*, 2004) proponen filtros de correlación no lineal adaptables localmente. El filtrado se basa en estadísticas de orden local tales como mediana y sumas truncadas en alfa[4]. Finalmente se reportan resultados favorables en el reconocimiento de objetos en imágenes contaminadas con ruido aditivo e impulsivo.

J. González *et al.* (González-Fraga *et al.*, 2006) proponen un algoritmo iterativo para el diseño de un filtro adaptativo basado en funciones discriminantes sintéticas con capacidad de discriminación, invariante a distorsiones geométricas y robusto a ruido aditivo. El filtro es construido con imágenes de distorsiones del objeto a reconocer, objetos falsos y el fondo a rechazar. Como resultado se obtiene un filtro de correlación el cual proporciona como salida un pico máximo que corresponde al objeto a identificar, suprimiendo al mismo tiempo los picos derivados de objetos falsos y el fondo. Su principal desventaja es su escasa robustez a ruido de naturaleza no Gaussiana; sin embargo, una ventaja del algoritmo es que se puede aplicar utilizando cualquier otro tipo de filtro compuesto, por ejemplo, filtros no lineales compuestos.

En los trabajos (Martínez-Díaz y Kober, 2008b) y (Martínez-Díaz y Kober, 2011) de

[4]Alpha trimmed

S. Martínez y V. Kober proponen filtros no lineales compuestos para el reconocimiento de objetos, en condiciones de iluminación no homogénea y distorsiones geométricas contaminadas con ruido no Gaussiano. Para el diseño del filtro utilizan descomposición binaria de la imagen por umbral, funciones discriminantes sintéticas no lineales, utilizando imágenes de los objetos reales y sus distintas distorsiones geométricas, así como imágenes de los objetos falsos; y definen una correlación no lineal normalizada tomando en cuenta la iluminación no homogénea. Por último, se propone un algoritmo para el proceso de reconocimiento del objeto.

Finalmente, P. M. Aguilar *et al.* (Aguilar-González *et al.*, 2014) proponen un algoritmo para el diseño de filtros adaptativos compuestos para reconocer objetos cuando se encuentran en imágenes ruidosas de entrenamiento, además de que su contorno y sus valores de iluminación no son explícitamente conocidos. Los filtros compuestos utilizan un conjunto de imágenes de entrenamiento, las cuales incluyen múltiples vistas del objeto a encontrar. Las funciones discriminantes sintéticas son una combinación lineal de imágenes de entrenamiento, pero pueden tener valores bajos de DC. Además, el criterio de SNR de los filtros compuestos se degrada gradualmente cuando un creciente número de imágenes son incluidas en el conjunto de entrenamiento. Para resolver este problema utilizan un banco de filtros compuestos, donde cada uno es diseñado con un subconjunto de imágenes del objeto. En este caso la detección es obtenida correlacionando la escena de entrada con cada uno de los filtros en el banco, después de obtiene la DC en cada resultado obtenido en los planos de correlación calculados y finalmente se elige aquel plano con el valor DC más alto. El desempeño del filtro puede ser mejorado mediante la búsqueda de estructuras en el fondo con una apariencia similar a la del objetivo. Estas estructuras son adicionadas al conjunto de entrenamiento como imágenes a ser rechazadas.

1.8. Algortimo de reconocimiento por características: SIFT

SIFT[5] es un método para la extracción de características distintivas invariantes a escala y rotación de una imagen, las cuales pueden ser utilizadas para la correspondencia confiable entre objetos o imágenes con distintas perspectivas, contaminadas con ruido y con cambios en la iluminación. A contiuación se describe brevemente el funcionamiento del algoritmo SIFT.

[5]Scale Invariant Feature Transform

- *Espacio de escalas.*

  Primero se construye un espacio de escalas, para esto se convoluciona la imagen de referencia del objeto a reconocer y el operador Gaussiano:

$$G(x, y, t) = \frac{1}{2\pi t^2} e^{\frac{-(x^2+y^2)}{2t^2}}, \tag{1}$$

  con t el parámetro de escalamiento; en este caso se puede considerar a t como la cantidad de desenfoque (entre mayor sea t mayor desenfoque presentará la imagen), obteniendo un conjunto de imágenes progresivas desenfocadas[6]. A este conjunto de imágenes se les llama escalas y se recomienda que sean cinco.

  Posteriormente la imagen de referencia se divide en la mitad de su tamaño y se generan nuevamente las cinco escalas. A estos grupos de imágenes se les llama octavas y se recomienda que sean cuatro.

- *El Laplaciano.*

  Una vez que se tiene el espacio de escalas lo siguiente es aplicar el operador Laplaciano (derivadas de segundo orden) para localizar las esquinas y los bordes de la imagen. La desventaja de utilizar el Laplaciano es que es computacionalmente costoso, por lo que se realiza un aproximación calculando la diferencia entre imágenes consecutivas en cada octava.

  Posteriormente se procede a localizar los máximos y mínimos en las imágenes obtenidas de las operaciones anteriores, para cada octava se toma el conjunto de 5 imágenes y se consideran sólo las tres imágenes centrales (2da, 3ra y 4ta imagen). Para cada pixel de cada una de estas imágenes se selecciona una vecindad de 26 pixeles. Los ocho pixeles vecinos alrrededor del pixel en cuestión, los nueve pixeles vecinos de la imagen anterior en la misma posición del pixel en cuestión y los nueve pixeles vecinos de la imagen posterior en la misma posición del pixel en cuestión. Para cada una de estas vecindades se busca saber si el pixel en cuestión es un máximo/mínimo comparado con los demás elementos, de no ser así se descarta.

[6]Blurred

- *Removiendo características con bajo contraste.*

  Hasta ahora se tiene un conjunto de características o puntos clave que cumplieron con el criterio de máximo/mínimo en la vecindad definida. No todas las características obtenidas son útiles, por lo que se eliminan aquellas características con bajo contraste definiendo un umbral y comparando la intensidad de los pixeles.

  También se eliminan aquellas características que son parte de un borde o una región, ya que lo que se busca es obtener sólo características pertenecientes a esquinas que son las que brindan mayor información. Esto es posible obteniendo dos gradientes perpendiculares a la característica. Si los dos gradientes son chicos entonces el punto clave pertenece a una región plana, ya que no hay mucha variación de intensidad; si un gradiente es chico y el otro es grande entonces el punto clave pertenece a un borde, ya que un gradiente esta sobre el borde y el otro es perpendicular a él. Finalmente si ambos gradientes son grandes entonces tenemos una esquina.

- *Asignación de orientaciones.*

  El siguiente paso es asignar una orientación a cada punto clave, esto para obtener invarianza a rotación. Para ello se obtiene el gradiente de orientación y magnitud utilizando los pixeles de alrrededor del punto clave. Posteriormente se crea un histograma y del histograma resultante se toma la orientación del pico más grande y todos los picos por arriba del 80 % se consideran como puntos clave asignándoles la misma magnitud y orientación que la del pico mayor.

- *Descriptor de características.*

  La idea general es crear una huella única para cada punto clave. Para esto, se define una ventana de 16×16 para cada punto clave y esta a su vez se divide en sub-ventanas de 4×4. Para cada sub-ventana se realiza un procedimiento semejante al anterior calculando los gradientes de magnitud y orientación formando un vector de 128 dimensiones.

Finalmente el reconocimiento se realiza mediante el algoritmo del vecino más cercano, seguido de la transformación Hough que se utiliza para identificar grupos de características pertenecientes a un mismo objeto y finalmente se verifica mediante mínimos cuadrados (Lowe, 2004).

Por otro lado SURF[7] también es un un algoritmo invariante a escala y rotación el cual se basa en el algoritmo SIFT. Este algoritmo es más rápido que el algoritmo SIFT ya que hace uso de imágenes integrales pero es menos preciso.

1.9. Sistemas OCR

Como ya se ha mencionado, en la actualidad existen distintos sistemas OCR, tanto comerciales como de código libre. A continuación se hará una pequeña descripción del sistema comercial ABBYY y el sistema de código libre Tesseract.

1.9.1. ABBYY

ABBYY surgió con la idea de crear un diccionario electrónico llamado Lingvo en el año de 1989, con el objetivo de traducir palabras (Ruso-inglés y viceversa) de manera fácil y rápida en pocos segundos. Años más tarde en 1993, se creó el primer sistema uni-fuente OCR en Rusia lladado ABBYY FineReader, pero fue hasta 1997 que se conoció dicho sistema como ABBYY a nivel mundial.[8]

Hoy en día ABBYY es un sistema comercial que ofrece distintos tipos de servicios, entre los cuales se encuentra el reconocimiento de caracteres en más de 150 idiomas[9]. A continuación se hace una breve descripción de su funcionamiento[10].

Primero, el sistema realiza una segmentación dividiendo las líneas de la imagen del documento en palabras; posteriormente, procesa los caracteres individualmente (letras, números y símbolos). Una vez que las palabras se encuentran divididas en caracteres, las imágenes de cada caracter encontrado por el programa son dadas a un clasificador para su reconocimiento.

ABBYY utiliza distintos tipos de clasificadores, entre los que se encuentran:

- Clasificador de cuadrícula[11]: compara la imagen del caracter con un conjunto de imágenes del posible caracter.

[7] Speeded Up Robust Features
[8] http://www.abbyy.com/company/history/
[9] http://finereader.abbyy.com/recognition_languages/
[10] http://www.abbyy-developers.com/en:tech:insideocr:classifier
[11] Raster classifier

- Clasificador de características: este tipo de clasificador trabaja con las características del caracter en cuestión, tales como su perímetro, número de puntos negros en ciertas áreas o a lo largo de líneas, etc. Su precisión depende de qué tan bien se hayan escogido las características a utilizar.

- Clasificador de contorno: éste es un tipo de clasificador de características; pero su diferencia radica en que las características obtenidas son sólo del contorno y no del caracter completo.

- Clasificador de estructura: este clasificador analiza la estructura mediante la descomposición en distintos componentes (líneas, arcos, círculos, etc.) recreando la estructura exacta del caracter analizado y posteriormente comparándolo con patrones de estructuras.

- Clasificador de diferenciador de características[12] y de estructuras[13]: finalmente estos clasificadores lo que hacen es obtener las características y el contorno respectivamente de dos caracteres similares para poder diferenciarlos poniendo especial atención en aquellas características que los hacen diferentes.

El sistema cuenta con reconocimiento de palabras, es decir, utiliza diccionarios con los cuales hace un análisis contextual para reconocer palabras completas.

Finalmente, ABBYY cuenta con una aplicación OCR para imágenes capturadas con teléfonos móviles[14], la cual analiza la imagen y mejora su calidad. En el caso de que la imagen capturada no presente ciertas condiciones, la aplicación determina si la imagen es apta para el procesamiento o indica si se debe tomar otra imagen.

1.9.2. Tesseract

Otro sistema OCR popular es el sistema Tesseract (Smith, 2007) el cual es de código libre desde 2005 por google[15] pero originalmente creado por un grupo de la compañía HP en los años 1985-1994. A continuación se presenta una breve descripción del funcionamiento del

[12]Feature differentiating classifier
[13]Structure differentiating classifier
[14]http://www.abbyy.com/mobile_imaging_sdk/
[15]http://code.google.com/p/tesseract-ocr/

sistema.

Primero se realiza una binarización local utilizando la técnica de F. Shafait *et al.* (Shafait *et al.*, 2008b) el cual usa el algoritmo de binarización de Sauvola (Sauvola *et al.*, 1997) y hacen uso de imágenes integrales con el propósito de disminuir el tiempo de ejecución. Posteriormente, utilizan análisis de componentes principales para extraer el contorno de los caracteres y con ello realizar un análisis de lineas y regiones de texto. Una vez terminado este proceso, se tiene un conjunto de palabras que entran a un clasificador previamente entrenado, el cual reconocerá cada una de ellas de una manera más precisa. Se realiza un análisis lingüístico eligiendo la mejor palabra según las siguientes categorías: las palabras más frecuentes, las palabras principales del diccionario, las palabras con el mismo número de letras, las principales palabras en mayúsculas y las principales palabras en minúsculas.

Finalmente la elección se realiza con aquella que tenga la menor distancia total, donde cada categoría antes mencionada es multiplicada por una constante diferente. Si del proceso anterior no se obtiene una buena clasificación o no se logra reconocer la palabra, entonces la palabra pasa a un proceso de corte, es decir, se utiliza un algoritmo para dividir la palabra en caracteres utilizando el contorno y revisando los posibles puntos de separación.

En un estudio realizado por C. Patel *et al.* (Patel *et al.*, 2012) el sistema Tesseract tiene una precisión del 70 % en promedio con imágenes en escala de grises y un 60 % de presición en imágenes a color. En general, los sistemas OCR en la actualidad cuentan con una presición que varía del 71 % al 98 % (Patel *et al.*, 2012).

1.10. Organización de la tesis

La tesis se encuentra distribuida de la siguiente manera: la segunda sección presenta los fundamentos teóricos como base para la comprensión del tema. En la tercera sección se presenta una breve descripción de los filtros lineales más populares. La cuarta sección presenta una descripción de filtros no lineales. Finalmente la quinta sección presenta el sistema propuesto, los resultados de los experimentos realizados, las conclusiones generales y trabajo futuro.

Capítulo 2. Marco Teórico

A continuación se definen una serie de conceptos importantes del área de procesamiento de imágenes y reconocimiento de patrones.

2.1. Sistemas lineales

2.1.1. Sistema

Se define un sistema como una unidad que convierte una función de entrada $f(\vec{x})$ en una función de salida (respuesta) $g(\vec{x})$, donde $\vec{x}$ es una variable independiente, tal como el tiempo o, en el caso de imágenes, la posición espacial. Se asume por simplicidad que $\vec{x}$ es una variable continua y D-dimensional, pero se aplica también para funciones discretas $f[n]$ D-dimensionales (Kumar *et al.*, 2005).

Se puede representar a la salida sistema como:

$$g(\vec{x}) = H[f(\vec{x})] \tag{2}$$

donde H es un operador que asigna a cada miembro del conjunto de posibles salidas $\{g(\vec{x})\}$ a cada miembro del conjunto de posibles entradas $\{f(\vec{x})\}$.

2.1.1.1. Linealidad

Se dice que H es un operador lineal si:

$$H[a_i f_i(\vec{x}) + a_j f_j(\vec{x})] = a_i H[f_i(\vec{x})] + a_j H[f_j(\vec{x})] = a_i g_i(\vec{x}) + a_j g_j(\vec{x}) \tag{3}$$

donde $a_i, a_j, f_i(\vec{x})$ y $f_j(\vec{x})$ son constantes arbitrarias y funciones respectivamente. A esta propiedad se le conoce como el principio de superposición.

2.1.1.2. Invarianza

Se dice que un sistema es invariante con respecto al espacio si:

$$f(\vec{x}) \longrightarrow g(\vec{x}),\ entonces\ f(\vec{x} - \vec{x_0}) \longrightarrow g(\vec{x} - \vec{x_0}) \tag{4}$$

para cualquier $f(\vec{x})$ y cualquier $\vec{x_0}$. Esto quiere decir, que si se conoce la salida para una entrada en particular, entonces es posible conocer la salida para cada versión desplazada de dicha entrada.

2.2. Definición de imagen

Podemos definir una imagen como una función continua bidimensional $f(\vec{x})$ donde cada par ordenado $\vec{x} = (x_1, x_2)$ representa una posición espacial de la imagen y el valor de f representa la intensidad lumínica de la imagen en dicho punto. Digitalmente, una imagen se puede representar mediante una matriz, donde los índices de sus renglones y columnas representan un punto de la imagen y el valor de cada elemento es representado por un escalar positivo llamado pixel, cuyo significado físico es determinado por la fuente de la imagen. Para imágenes en escala de grises este elemento representa un nivel de gris (González y Woods, 2006).

2.3. Modelos de escena

2.3.1. Traslapado

En este modelo, la escena de entrada $S(\vec{x})$ contiene un objeto de interés $t(\vec{x})$, en la posición desconocida $\vec{x_s}$ y degradado por ruido aditivo $n_s(\vec{x})$, tal como se muestra en la Eq.(5):

$$s(\vec{x}) = t(\vec{x} - \vec{x_s}) + n_s(\vec{x}) \tag{5}$$

2.3.2. No traslapado

En este modelo, se considera que un objeto opaco se encuentra sobre un fondo espacialmente disjunto y que toda la escena se corrompe con ruido aditivo. Formalmente el sistema se describe:

$$s(\vec{x}) = t(\vec{x} - \vec{x_s}) + b_s(\vec{x})\bar{w}(\vec{x} - \vec{x_s}) + n_s(\vec{x}) \tag{6}$$

donde $t(\vec{x} - \vec{x_s})$ y $n_s(\vec{x})$ son el objeto de interés y el ruido aditivo respectivamente; $b_s(\vec{x})$ representa el fondo disjunto y $\bar{w}(\vec{x})$ es la región inversa de soporte del objeto, la cual se define como 1 en las coordenadas de la señal donde el objeto no esta presente y cero en caso contrario.

2.3.3. Histograma

El histograma de una imagen digital con niveles de gris en el rango $[0, L-1]$ es una función discreta

$$h(r_k) = n_k \tag{7}$$

donde r_k es el k-ésimo nivel de gris con $k = 1, 2, \ldots, L-1$, n_k es el número de pixeles de la imagen con el nivel de gris k. En la práctica es común normalizar el histograma dividiendo cada uno de sus valores por el número total de pixeles n.

$$p(r_k) = n_k/n \tag{8}$$

En general, $p(r_k)$ da un estimado de la probabilidad de ocurrencia del nivel de gris r_k (González y Woods, 2006).

El histograma global de una imagen se obtiene de los niveles de gris de la imagen total, pero en algunas ocasiones es necesario sólo realzar pequeñas áreas de la imagen sin modificar la imagen por completo. Es entonces cuando se hace uso de histogramas locales, con la desventaja de contar con un menor número de pixeles para el procesamiento.

La Figura(1) presenta el histograma de una imagen digital.

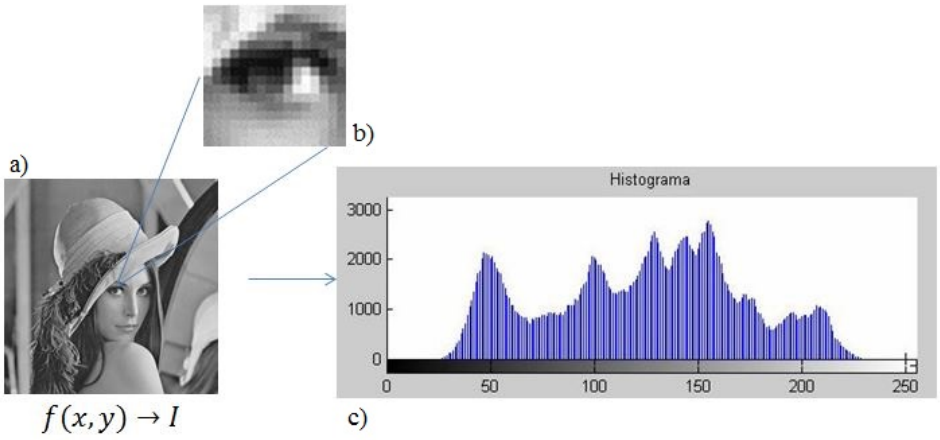

Figura 1: a) Imagen digital b) pixeles c) histograma de la imagen.

2.4. Transformadas espaciales

Sea $s = \{S_{n,m}\}$ el vector de los pixeles de la imagen a procesar, la cual tiene Q niveles de gris, n, m, son las coordenadas del pixel tal que $n = 1, 2, \ldots, N$ y $m = 1, 2, \ldots, M$; $L = N \times M$ es el tamaño de la matriz; $v = \{v_{n,m}\}$ es un vector de los pixeles de la imagen original libre de ruido; $\hat{v} = \{\hat{v}_{n,m}\}$ es el vector de los pixeles de la imagen resultante.

La vecindad de un pixel consiste de todos los pixeles espacialmente cercanos al pixel en cuestión y es llamada S-vecindad. Generalmente los pixeles pertenecientes a la vecindad coinciden con los pixeles de la ventana deslizante.

2.4.1. Renglón variacional

Otro concepto importante en estadísticas de orden es el renglón variacional, el cual se define como una secuencia 1-D $\{V(r)\}$ de H pixeles cuyos elementos se encuentran ordenados de manera ascendente con respecto a sus valores (Kober *et al.*, 2001):

$$\{V(r) : V(r) \leq V(r+1), r = 1, 2, \ldots, H\} \tag{9}$$

$V(r)$ es llamada la r-ésima estadística de orden y $r(V)$ es el rango de valor V. Para describir distintas estructuras en la imagen se define a continuación diferentes subconjuntos sobre la S-vecindad.

2.4.1.1. EV-vecindad

Se define como un subconjunto de pixeles $\{v_{m,n}\}$ donde sus valores se desvían del pixel central $v_{k,l}$ a lo más una determinada cantidad $-\epsilon_v$ y $+\epsilon_v$:

$$EV(v_{k,l}) = \{v_{m,n} : v_{k,l} - \epsilon_v \leq v_{m,n} \leq v_{k,l} + \epsilon_v\} \tag{10}$$

2.4.1.2. KNV-vecindad

Se define como un subconjunto de un número específico K de pixeles $\{v_{m,n}\}$ cuyos valores son cercanos al valor del pixel central $\{v_{k,l}\}$:

$$KNV(v_{k,l}) = \left\{ V(r) : \sum_{r=p}^{p+K-1} |v_{k,l} - V(r)| = MIN_p \right\} \tag{11}$$

2.4.1.3. ER-vecindad

Se define como un subconjunto de pixeles $\{v_{m,n}\}$ donde su rango se desvía del rango del pixel central $v_{k,l}$ a lo más una determinada cantidad $-\epsilon_r$ y $+\epsilon_r$:

$$ER(v_{k,l}) = \{v_{m,n} : r(v_{k,l}) - \epsilon_r \leq r(v_{m,n}) \leq r(v_{k,l}) + \epsilon_r\} \tag{12}$$

La elección de la vecindad (NBH) se define por la información *a priori* que se tenga de la imagen a procesar. Si se tiene el tamaño de la estructura que se desea preservar, entonces es conveniente trabajar con KNV-vecindad. La EV-vecindad nos ayuda a tomar en cuenta la propagación de la señal para preservarla o la fluctuación del ruido para suprimirlo.

Finalmente la ER-vecindad se utiliza para extraer bordes o eliminar ruido aditivo mezclado con otra clase de ruido (impulsivo por ejemplo) que contenga grandes colas en su distribución.

2.4.2. Ecualización de histograma

La ecualización se utiliza para realzar las diferencias entre el fondo y los detalles de una imagen. Esta puede ser global (utilizando todos los pixeles de la imagen) o local (utilizando vecindades). Haciendo uso del histograma local de una imagen, la ecualización se puede definir de la siguiente manera:

$$\hat{v_{m,n}} = A * RANK(NHB(v_{m,n})) + B = A \sum_{q=0}^{v_{m,n}} h_{NBH} + B \tag{13}$$

donde A y B son constantes de normalización y h_{NBH} es el histograma local de la vecindad NBH.

2.5. Descomposición por umbral

Una imagen $S(k,l)$ con Q niveles de gris puede ser representada por una suma de imágenes binarias (Fitch *et al.*, 1984)

$$S(k,l) = \sum_{q=0}^{Q-1} S^q(k,l) \tag{14}$$

donde $\{S^q(k,l), q = 0, 1, \ldots, Q-1\}$ es un conjunto de imágenes binarias obtenidas de la descomposición por umbral de la imagen original con un umbral q, de la siguiente manera:

$$S^q(k,l) = \begin{cases} 1, & \text{si } S(k,l) \geq q. \\ 0, & \text{en otro caso} \end{cases} \tag{15}$$

2.6. Correlación

Estadísticamente, se dice que dos variables aleatorias estan correlacionadas si conociendo la variación de una de ellas se puede saber algo de la otra. En procesamiento de imágenes y reconocimiento de patrones este concepto no es muy diferente. La correlación nos indica que tan similares son el objeto de prueba y un objeto de entrenamiento.

2.6.1. Error Cuadrático Medio (MSE) y correlación lineal

El Error Cuadrático Medio (MSE) entre dos señales esta dado por:

$$MSE(m,n) = \sum_{-\infty}^{\infty}\sum_{-\infty}^{\infty}(g(p,q) - h(m+p, n+q))^2 \tag{16}$$

$$= \sum_{-\infty}^{\infty}\sum_{-\infty}^{\infty}[g(p,q)^2 + h(m+p,n+q)^2 - 2g(p,q)h(m+p,n+q)], \tag{17}$$

siendo $\sum_{-\infty}^{\infty}\sum_{-\infty}^{\infty} g(p,q)h(m+p,n+q)$ la correlación lineal entre las dos señales. Es posible demostrar que maximizando este término, se minimiza el MSE entre las señales (Maragos, 1989).

2.6.2. Error Medio Absoluto (MAE) y correlación morfológica

Otra alternativa para el cálculo de la correlación, es derivar una expresión partiendo del Error Medio Absoluto (MAE) se puede obtener una medida de correlación no lineal. Este se puede calcular:

$$MAE(m,n) = \sum_{-\infty}^{\infty}\sum_{-\infty}^{\infty} |g(p,q) - h(m+p, n+q)|, \tag{18}$$

Desarrollando el término de valor absoluto, la ecuación anterior se puede escribir de la siguiente forma:

$$MAE(m,n) = \sum_{-\infty}^{\infty}\sum_{-\infty}^{\infty} [g(p,q) + h(m+p, n+q) - 2MIN(g(p,q), h(m+p, n+q))], \tag{19}$$

dado que las dos primeras sumatorias son constantes, la manera de minimizar la expresión anterior es maximizando el último término, lo cual da lugar a la siguiente medida de correlación no lineal llamada correlación morfológica:

$$c(m,n) = \sum_{-\infty}^{\infty}\sum_{-\infty}^{\infty} MIN(g(p,q), h(m+p, n+q)). \tag{20}$$

Es posible demostrar que al aplicar esta operación basada en la suma de mínimos para detectar una señal dentro de otra, el pico de correlación resultante es más agudo que al utilizar la correlación lineal basada en la suma de productos (Maragos, 1989).

2.7. Transformada de Fourier, espacio de frecuencias

La transformada de Fourier de una función se define como:

$$F(\omega) = \int_{-\infty}^{\infty} f(x) e^{-i\omega x} dx, \tag{21}$$

y su transformada inversa como:

$$F(\omega) = \frac{1}{2\pi} \int_{-\infty}^{\infty} F(\omega) e^{i\omega x} d\omega, \tag{22}$$

donde x y ω son las coordenadas en el dominio espacial y frecuencial respectivamente. La función $f(x)$ puede ser interpretada como una combinación lineal de funciones periódicas,

con seno y coseno sus funciones base.

2.7.1. Propiedades de la transformada de Fourier

- Linealidad

  Como la transformada de Fourier es un operador lineal entonces para dos funciones $f(x)$, $g(x)$ y dos constantes a, b, se cumple lo siguiente:

$$\mathscr{F}\{af(x) + bg(x)\} = aF(\omega) + bG(\omega). \tag{23}$$

- Desplazamiento en el dominio espacial

$$\mathscr{F}\{f(x-a)\} = F(\omega)e^{-i\omega a}. \tag{24}$$

- Desplazamiento en el dominio de la frecuencia

$$\mathscr{F}\{f(x)e^{iax}\} = F(\omega - a). \tag{25}$$

- Simetría conjugada

$$\mathscr{F}\{f^*(x)\} = F^*(-\omega). \tag{26}$$

  Si la función es real, entonces:

$$F(\omega) = F^*(-\omega). \tag{27}$$

2.7.2. Teorema de correlación

Es posible realizar la correlación de dos funciones en el dominio de la frecuencia mediante el uso de la transformada de Fourier de la siguiente manera. Sea $f(x)$ y $g(x)$ dos funciones complejas, el teorema de la correlación establece que:

$$\mathscr{F}\{f(x) \otimes g(x)\} = F^*(\omega)H(\omega) \tag{28}$$

$$\mathscr{F}\{f^*(x)g(x)\} = F(\omega) \otimes H(\omega) \tag{29}$$

$\otimes$ denota la correlación entre dos funciones y $F(\omega)$, $H(\omega)$ son las transformadas de Fourier de $f(x)$, $g(x)$ respectivamente. Es decir, la correlación en el dominio espacial es igual a la multiplicación de funciones en el dominio de las frecuencias.

2.7.3. Teorema de Parserval

El teorema de Parserval establece la preservación de la energía de una señal $f(x)$ en el dominio espacial y en el dominio de frecuencia. Es decir:

$$\int_{-\infty}^{\infty} |f(x)|^2 dx = \frac{1}{2\pi} \int_{-\infty}^{\infty} |F(\omega)|^2 d\omega \tag{30}$$

2.8. Calidad de la imagen

La calidad de una imagen depende de distintos factores tales como la resolución, el contraste, el ruido, la iluminación, etc. Además, se cree que entre mayor sea el número de pixeles mayor será la calidad de la imagen, pero esto no necesariamente en cierto.

El número y tamaño de los pixeles dependen totalmente del sensor de la cámara (CCD), a continuación se explica un poco su funcionamiento.

2.8.1. Sensor CCD

El dispositivo de carga acoplada [1] es un circuito integrado que contiene condensadores acoplados. Cada uno de estos condensadores, llamados también células fotoeléctricas, pueden convertir la intensidad de la luz en cargas eléctricas.

El funcionamiento básico del sensor es el siguiente. Podemos representar a cada pixel que compone el sensor CCD como un "balde" de tal manera que cada balde acumula la luz incidente, tal como se presenta en la Fig.(2). Posteriormente se mide la luz acumulada de cada balde (señal análoga) y es convertida por medio de un amplificador en una salida numérica (señal digital).

[1] charge-coupled divice

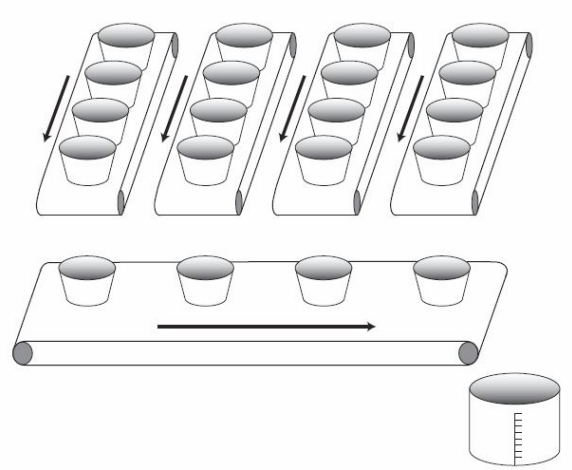

Figura 2: Analogía de sensor CCD (Russ, 2010).

La calidad del CCD se define por la habilidad de estas células de absorber fotones. Cuando el tamaño del CCD no cambia pero se aumenta el número de células fotoeléctricas entonces el número de fotones absorbidos en un tiempo determinado disminuye y por lo tanto, la información que se obtiene es menor. Por otro lado, si el tamaño de la célula fotoeléctrica es muy grande, el número de fotones absorbidos será mayor pero la probabilidad de ruido aumentará de igual forma. Por lo tanto existe un compromiso entre el tamaño del circuito y el número de células fotoeléctricas que lo componen (Russ, 2010).

2.8.2. Resolución

Por otro lado, la resolución de una imagen digital se mide por el número pixeles que contiene la imagen o el número de células fotoeléctricas que contiene el sensor CCD. Generalmente la resolución nos dice que tanto detalle es capaz de capturar la cámara en una imagen y su tamaño. Pero no define necesariamente la calidad de la imagen.

2.8.3. Contraste

El contraste se puede definir como la diferencia de intensidad entre un pixel y sus pixeles vecinos. El bajo contraste puede ser causado por una baja iluminación, el tamaño del sensor, la exposición (cantidad de luz recibida en un tiempo determinado) o la apertura de la lente durante la adquisición de la imagen. Distintas técnicas se han desarrollado para mejorar el contraste de las imágenes, entre las más utilizadas se encuentran la ecualización de histograma

y el estiramiento de contraste[2]. En muchas ocaciones se confunde el brillo con el contraste, pero no son lo mismo. Una imagen puede contener mucho brillo y poco contraste tal como muestra la Fig. (3) o viceversa.

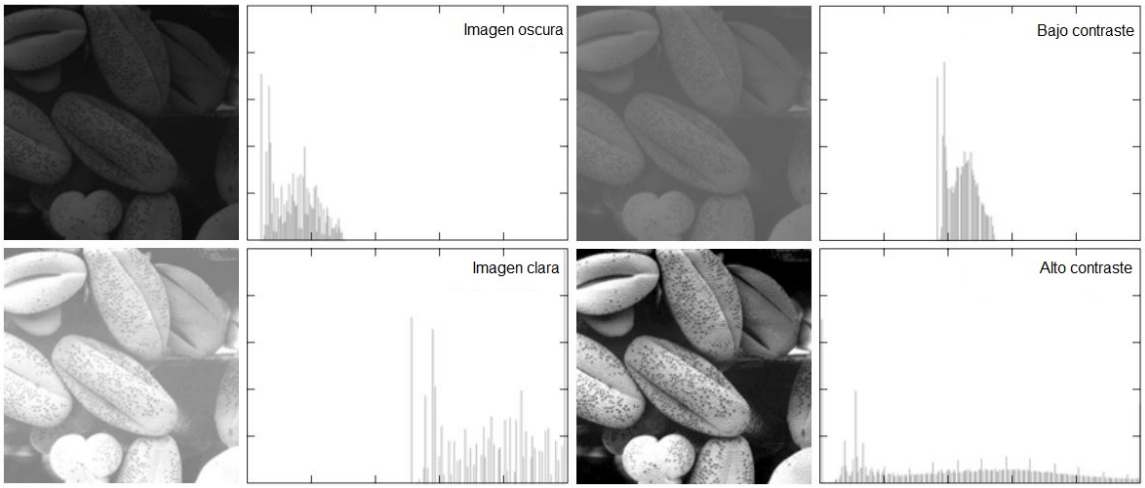

Figura 3: Diferencias de contraste y sus respectivos histogramas (Gonzalez y Woods, 2006).

2.8.4. Iluminación

La iluminación de una escena real depende de la fuente de iluminación usada y la forma de la superficie, un modelo de iluminación describe la relación entre éstas. en general se pueden describir tres modelos (Horn, 1990):

- Especular: en este modelo el ángulo de incidencia es el mismo que el ángulo reflejado.
- Lambertiano: en este modelo se supone que la superficie refleja la luz en todas direcciones.
- Mixto: este modelo es una combinación de los modelos anteriores.

El modelo que se utilizó en este trabajo es el modelo Lambertiano, el cual se describe a continuación (Díaz-Ramírez *et al.*, 2014).

2.8.4.1. Modelos de iluminación Lambertiano

Considérese la escena que se muestra en la Fig. 4 donde un objetivo se desplaza horizontalmente con respecto a un plano de dos dimensiones. La superficie es iluminada por una

[2]contrast stretching

fuente de luz puntual con los siguientes parámetros $I = [\rho, \phi, \varphi]$, donde ρ es la distancia entre un punto en la superficie y la fuente de luz, y ϕ, φ son los ángulos de inclinación entre la superficie normal y el punto de observación. Sea Ω el ángulo de incidencia de la luz, es decir el ángulo entre el vector de la superficie normal N y el vector de la dirección de la luz I; y la luz reflejada por una superficie Lambertiana esta dada por $R_L = cos(\Omega)$. De acuerdo con la Fig. 4, la luz reflejada por la superficie para una posición conocida de la fuente de iluminación y asumiendo que el punto de observación se encuentra sobre el eje z está dado por:

$$d(\vec{x}) = cos\left\{\frac{\pi}{2} - arctan\left[\frac{\rho}{cos(\phi)}\left[(\rho tan(\phi)cos(\varphi) - x_1)^2 + (\rho tan(\phi)sin(\varphi) - x_2)^2\right]^{-1/2}\right]\right\} \tag{31}$$

Nótese que $d(\vec{x})$ en la Eq. 31 es una función multiplicativa que depende de los parámetros ρ, ϕ y φ.

2.8.5. Ruido

Se le llama ruido a cualquier información indeseable que contamina una imagen, principalmente esta información es obtenida en el proceso digital de adquisición de las imágenes. A continuación se describen los dos más comunes (González y Woods, 2006):

2.8.5.1. Ruido Gaussiano

También llamado ruido aditivo, este tipo de ruido suele ser el más común (ruido electrónico en un sistema de captura de imagen) y tratable matemáticamente. La función de densidad de probabilidad (PDF) de una variable aleatoria Gaussiana se define como:

$$p(z) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{z-\mu^2}{2\sigma^2}} \tag{32}$$

donde z representa el nivel de gris, μ es la media y σ es la desviación estándar de los valores de z.

2.8.5.2. Ruido impulsivo

Generalmente en este modelo existen dos valores posibles a y b. Este tipo de ruido puede ser producto de un mal funcionamiento de los sensores en la cámara. La PDF del ruido

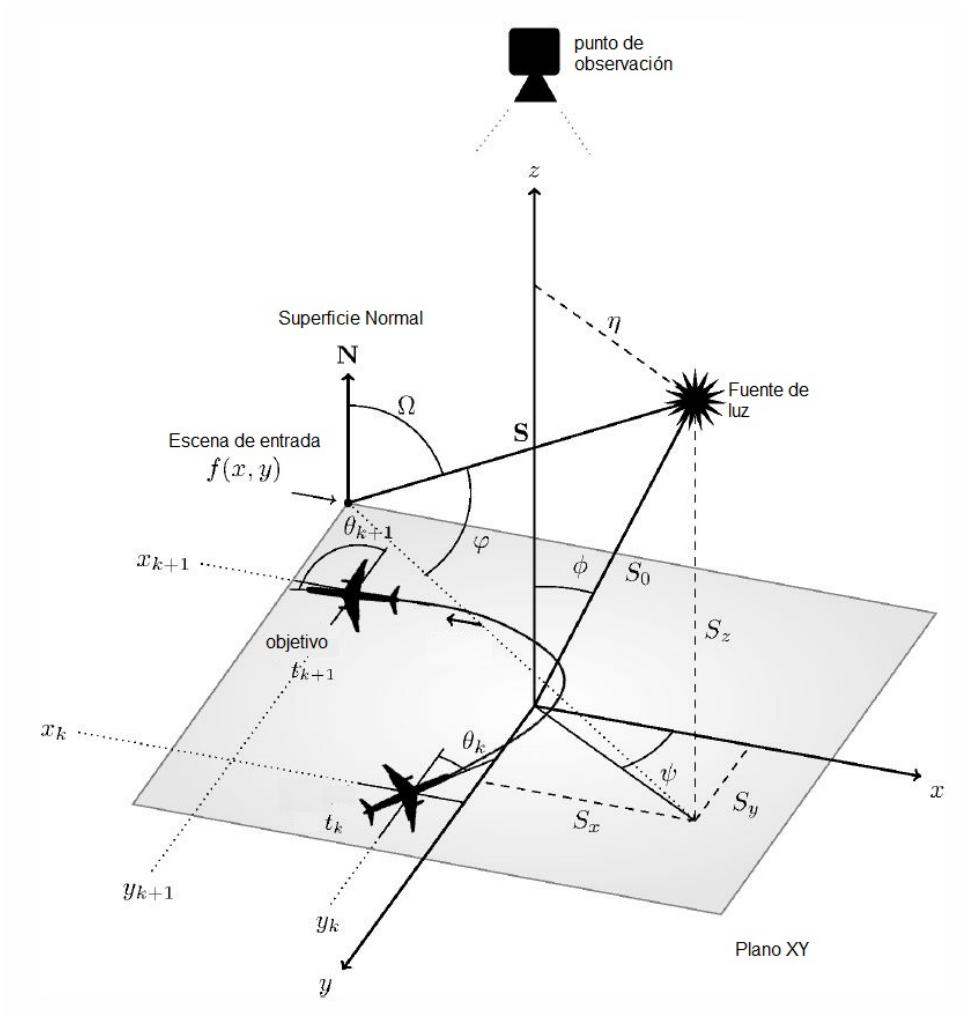

Figura 4: Modelo de iluminación Lambertiano (Díaz Ramírez *et al.*, 2014).

impulsivo bipolar se define como:

$$p(z) = \begin{cases} P_a, & \text{para } z = a \\ P_b, & \text{para} z = b \\ 0, & \text{en otro caso} \end{cases} \tag{33}$$

En la Fig.(5) se muestra una imagen contaminada con ruido Gaussiano e impulsivo.

2.8.6. Distorsiones geométricas

Las distorsiones geométricas se pueden producir por irregularidades en la perspectiva de la cámara, el movimiento del objeto durante la captura o condiciones ambientales presentes.

Figura 5: a) Imagen original, b) imagen contaminada con ruido aditivo Gaussiano, c) imagen contaminada con ruido impulsivo.

Entre las distorsiones más comunes se encuentran: el desplazamiento, la rotación, escalamiento, etc., y en muchas ocasiones se pueden encontrar combinaciones de éstas.

Las transformaciones geométricas modifican la relación espacial entre los pixeles. En términos del procesamiento de imágenes digitales una transformación geométrica consiste de dos operaciones básicas: una transformación espacial que define la re-ubicación de los pixeles en el plano imagen; y una interpolación de los niveles de gris, los cuales tienen que ver con la asignación de los valores de intensidad de los pixeles en la imagen transformada.

Una transformación afín es aquella en la que las nuevas coordenadas son expresadas linealmente en términos del punto original, es decir:

$$x' = ax + by + m \tag{34}$$

$$y' = cx + dy + n. \tag{35}$$

Las ecuaciones anteriores pueden representarse en forma matricial de la siguiente manera:

$$\begin{pmatrix} x' \\ y' \end{pmatrix} = \begin{pmatrix} a & b & m \\ c & d & n \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} x \\ y \\ 1 \end{pmatrix} \tag{36}$$

El sistema anterior tiene solución si la matriz $\left(\begin{smallmatrix} a & b & m \\ c & d & n \\ 0 & 0 & 1 \end{smallmatrix}\right)$ es no singular, es decir, si su

determinante es distinto de cero. A esta matriz se le conoce como matriz de transformación y dependiendo de sus coeficientes se define una transformación afín.

2.8.7. Transformaciones afines

2.8.7.1. Traslación

Sea $a = d = 1$ y $b = c = 0$, la traslación se define por la matriz:

$$\begin{pmatrix} x' \\ y' \end{pmatrix} = \begin{pmatrix} 1 & 0 & m \\ 0 & 1 & n \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} x \\ y \\ 1 \end{pmatrix} = \begin{pmatrix} x+m \\ y+n \\ 1 \end{pmatrix} \tag{37}$$

2.8.7.2. Escalamiento

Sea $m = n = 0$ (centrado en el origen) y $b = c = 0$, el escalamiento se define por la matriz:

$$\begin{pmatrix} x' \\ y' \end{pmatrix} = \begin{pmatrix} S_x & 0 & 0 \\ 0 & S_y & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} x \\ y \\ 1 \end{pmatrix} = \begin{pmatrix} S_x x \\ S_y y \\ 1 \end{pmatrix} \tag{38}$$

2.8.7.3. Rotación

Sea $m = n = 0$, la rotación se define por la matriz:

$$\begin{pmatrix} x' \\ y' \end{pmatrix} = \begin{pmatrix} cos(\theta) & -sen(\theta) & 0 \\ sen(\theta) & cos(\theta) & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} x \\ y \\ 1 \end{pmatrix} = \begin{pmatrix} xcos(\theta) - ysen(\theta) \\ xsen(\theta) + ysen(\theta) \\ 1 \end{pmatrix} \tag{39}$$

2.8.7.4. Shearing

Sea $m = n = 0$, el estiramiento se define por la matriz:

$$\begin{pmatrix} x' \\ y' \end{pmatrix} = \begin{pmatrix} 1 & Sh_h & 0 \\ Sh_v & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} x \\ y \\ 1 \end{pmatrix} = \begin{pmatrix} x + ySh_h \\ Sh_v + y \\ 1 \end{pmatrix} \tag{40}$$

Si $Sh_v = 0$, entonces el estiramiento es horizontal (dirección x); si $Sh_h = 0$, entonces el estiramiento es vertical (dirección y). Finalmente, al aplicar una transformación afín en una imagen los pixeles cambian de posición respecto a su posición inicial y es necesario aplicar una interpolación numérica para asignar valores de intensidad a las nuevas posiciones en el plano de la imagen. Entre las interpolaciones más comunes se encuentran: el vecino más cercano, interpolación bilineal e interpolación bicúbica.

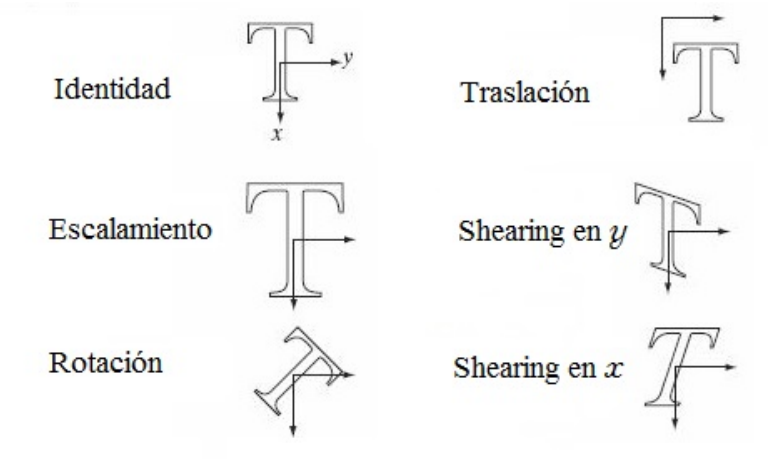

Figura 6: Transformaciones afines.

2.9. Métricas de desempeño

Existen diferentes criterios para medir la respuesta de un filtro de correlación. Entre las más comunes se encuentran (Kumar *et al.*, 2005):

2.9.1. Capacidad de Discriminación (DC)

La DC se define como la habilidad del filtro para distinguir entre patrones similares. Matemáticamente se expresa como:

$$DC = 1 - \frac{|C^B|^2}{|C^0|^2} \tag{41}$$

donde C^B es el pico máximo en el plano de correlación sobre el área del fondo a ser rechazada y C^0 es el pico máximo en el plano de correlación sobre el área del objeto que va a ser reconocido. Cuando el valor DC es menor o igual a cero indica que el objeto a localizar no

proporcionó el máximo en el plano de correlación. Los valores negativos indican que el filtro no fue capaz de reconocer el objeto buscado.

2.9.2. Razón de Discriminación (DR)

Este tipo de discriminación es útil cuando se desea que el pico de salida del filtro sea poco sensible a las distorsiones geométricas. Para un problema de discriminación entre dos clases, sea Ω_1 el conjunto de todas las posibles imágenes de la clase 1 y Ω_2 el conjunto de imágenes de la clase 2. Una medida de la invariabilidad a la distorsión es la razón de discriminación (DR) definida como:

$$DR = \frac{min_{i,\Omega_1}|y_i(0)|}{max_{i,\Omega_2}|y_i(0)|} \tag{42}$$

donde $y_i(0)$ es la salida en el origen cuando la entrada es la i-ésima señal del conjunto.

2.9.3. Razón Señal-Ruido(SNR)

Esta metrica de desempeño caracteriza la variación del pico de correlación deseado en la salida del filtro con respecto al ruido en la entrada. Se define formalmente como:

$$SNR = \frac{|E\{y(0)\}|^2}{var\{y(0)\}} \tag{43}$$

Donde $E\{\}$ representa el promedio, $var\{\}$ representa la varianza del pico y $y(0)$ es el resultado de la correlación (asumiendo que ocurre en el origen).

Mayores valores de SNR indican una mayor tolerancia al ruido y por lo tanto una menor probabilidad de error en el proceso de detección.

2.9.4. Razón Pico-Lóbulo lateral(PSR)

Esta métrica de desempeño nos dice la agudeza del pico de correlación. Formalmente se define como:

$$PSR = \frac{|E\{y(0)\}|^2}{var\{y(\tau)\}} \tag{44}$$

donde $\tau >> 0$ representa un punto en el plano de salida de correlación, alejado del origen, suponiendo que el pico del objeto ocurre en el origen.

2.9.5. Razón Pico a Energía de Correlación (POE)

Para minimizar la probabilidad de falsas alarmas, es conveniente que en la salida del filtro las regiones del objeto y fondo sean lo más diferentes posible. Para lograr esto, se propuso el criterio POE, definido formalmente como:

$$POE = \frac{|E\{y(0)\}|^2}{E\{|y(\tau)|^2\}} \tag{45}$$

donde la barra indica promedio espacial, el denominador denota la energía promedio de la salida del filtro.

2.10. Resumen

En este capítulo se presentaron algunos conceptos teóricos importantes que ayuden a la comprensión de este trabajo. Si el lector desea profundizar más en algún concepto mencionado, puede seguir las referencias aquí mencionadas.

Capítulo 3. Filtros lineales clásicos

Supongamos que estamos tratando de localizar todas las apariciones de la imagen de referencia u objetivo (el caracter C en este ejemplo) en una imagen de prueba (llamada también escena de entrada). Una forma de lograr esto es cross-correlacionar la imagen de destino con la escena de entrada. La imagen del objetivo se coloca en la parte superior izquierda de la escena de entrada y se lleva a cabo la multiplicación de píxeles entre las dos matrices, todos los valores se suman para producir un valor de salida de correlación. Este proceso se repite desplazando la imagen objetivo de derecha y hacia abajo, produciendo de este modo una matriz de dos dimensiones como salida, llamada plano de correlación. Idealmente, esta salida de correlación debería tener valores altos correspondientes a el caracter " C " en la escena de entrada y ceros para otras letras. Por lo tanto, los valores grandes de correlación indican la presencia y ubicación del caracter que se está buscando. Pero esto no será siempre alcanzable ya que algunos caracteres diferentes tienen una alta correlación cruzada entre ellos; por ejemplo, la letra "C" y la letra " O " (Kumar *et al.*, 2005).

La correlación involucra dos señales o imágenes. Una imagen de referencia es correlacionada con una imagen de prueba (escena de prueba) para detectar y localizar la imagen de referencia. Por lo tanto, la correlación se puede considerar como un sistema con una entrada (la escena), una plantilla o filtro (derivado de la imagen de referencia), y una salida (la correlación)(Kumar *et al.*, 2005).

El trabajo pionero propuesto para el reconocimiento de objetos utilizando correlación fue hecho por Vander Lugt en el año de 1964 haciendo uso de un método óptico. La popularidad de los métodos de correlación para el reconocimiento de patrones se debe mucho a el papel que desempeñan los filtros de correspondencia en la detección de señales de ecos de radar corrompidas por ruido aditivo. A continuación se describe el modelo.

3.1. Filtro de correspondencia

Los filtros de correlación tienen su popularidad debido a su uso en sistemas de radares. Sea $s(t)$ la señal transmitida y $r(t)$ la señal recibida. Para un modelo de ruido aditivo, el problema de detección se simplifíca al escoger entre dos hipótesis:

$$H_1 : r(t) = n(t), \tag{46}$$

$$H_2 : r(t) = s(t) + n(t), \tag{47}$$

donde $n(t)$ denota el ruido y se asume que está modelado como un proceso estacionario en sentido amplio (débil) con media cero y la densidad del espectro de potencias (PSD) $P_n(\vec{v})$.

El enfoque básico que se utiliza para el problema de la detección de la señal binaria es el siguiente. La señal recibida pasa por un sistem LSI con respuesta al impulso $h(t)$, se busca el máximo valor de la señal de salida $y(t)$ y este máximo es comparado con un umbral pre-definido T. Si y_{max} excede el umbral T, entonces la señal recibida contiene la señal transmitida, de lo contrario se declara como sólo ruido. Si el umbral T es bajo, entonces la probabilidad de una pérdida es pequeña, pero la probabilidad de falsas alarmas es grande; es aquí donde entra la medida de desempeño.

En este enfoque, lo más importante es el diseño del filtro $H(\vec{v})$. Un buen filtro debe de hacer el promedio de y_{max} grande y la varianza del ruido lo más pequeña posible; por lo que es deseable que el filtro $H(\vec{v})$ maximice la SNR definida a continuación:

$$SNR = \frac{|E\{y_{max}|H_1\}|^2}{var\{y_{max}\}}. \tag{48}$$

Como la media del ruido se asume es cero, $E\{y_{max}|H_1\}$ es el valor máximo de la salida del filtro cuando $s(t)$ es la señal de entrada. Con el propósito de determinar el filtro óptimo $H(f)$, se puede suponer sin perdida de generalidad que la salida $y(t)$ tiene su máximo en el origen, por lo que tenemos:

$$\begin{aligned} |E\{y_{max}|H_1\}|^2 &= |E\{y(0)_{max}|H_1\}|^2 \\ = |\int s(t)h(-t)|^2 &= |\int S(\vec{v})H(\vec{v})d\vec{v}|^2. \end{aligned} \tag{49}$$

Se supone que $s(t), h(t)$ son reales. De la misma manera tenemos que el ruido es independiente de la señal $s(t)$. Como el ruido de entrada $n(t)$ es un proceso estacionario en sentido amplio con PSD igual a $P_n(\vec{v})$, la salida del ruido del sistema LSI es también estacionario en el sentido amplio con PSD igual a $P_n(\vec{v})|H(\vec{v})|^2$. Finalmente, como la varianza de un proceso aleatorio con media cero es igual al área bajo la curva de PSD, se puede expresar el denominador como:

$$var\{y_{max}\} = \{\int P_n(\vec{v})|H(\vec{v})|^2 d\vec{v}\}. \tag{50}$$

Por lo que podemos expresar SNR como:

$$SNR = \frac{|\int S(\vec{v})H(\vec{v})d\vec{v}|^2}{\int P_n(\vec{v})|H(\vec{v})|^2 d\vec{v}}. \tag{51}$$

Utilizando la desigualdad de Cauchy-Schwarz se llega a que:

$$SNR_{max} = \int \frac{|S(\vec{v})|^2}{P_n(\vec{v})} d\vec{v}, \tag{52}$$

y finalmente la restricción para la igualdad da como resultado que:

$$\left[\frac{S(\vec{v})}{\sqrt{P_n(\vec{v})}}\right] = \beta[H(\vec{v})\sqrt{P_n(\vec{v})}]^* \Rightarrow H(\vec{v}) = \alpha\frac{S^*(\vec{v})}{P_n(\vec{v})}, \tag{53}$$

donde α es cualquier constante compleja.

3.1.1. Ruido blanco

Si el ruido con el que la señal se encuentra contaminada es ruido blanco, entonces el filtro óptimo está dado por:

$$H(\vec{v}) = \alpha\frac{S^*(\vec{v})}{N_0}, \tag{54}$$

donde la PSD $P_n(\vec{v}) = N_0$ constante.

3.1.2. Ruido coloreado

El filtro de correspondencia puede ser expresado como dos filtros en cascada de la siguiente manera:

$$H(\vec{v}) = \alpha\frac{S^*(\vec{v})}{P_n(\vec{v})} = H_{pre}(\vec{v})H_{MF}(\vec{v}), \tag{55}$$

donde

$$H_{pre}(\vec{v}) = \frac{1}{\sqrt{P_n(\vec{v})}} \tag{56}$$

y

$$H_{MF}(\vec{v}) = \alpha \frac{S^*(\vec{v})}{\sqrt{P_n(\vec{v})}}. \tag{57}$$

Lo que significa que primero se "pre-blanquea" el ruido de la señal de entrada por el primer filtro y finalmente pasa por el segundo filtro que es el filtro de correspondencia.

3.2. Filtro Sólo Fase (POF)

El filtro sólo fase propuesto por Horner y Gianino (Horner y Gianino, 1984) el cual se define de la siguiente manera:

$$H_{POF}(\vec{v}) = \frac{S^*(\vec{v})}{|S(\vec{v})|} = e^{-j\theta(\vec{v})}, \tag{58}$$

donde $\theta(\vec{v})$ denota la fase de $S(\vec{v})$. De la ecuación anterior podemos ver que el filtro tiene magnitud 1 para todas las frecuencias, dejando pasar toda la luz a través de él. Además la fase contiene la mayor cantidad de información de la imagen que la magnitud.

La eficiencia óptica, también conocida como eficiencia Horner (η_H) se define como el radio entre la energía de correlación entre la escena de entrada ($f(\vec{x})$) y cualquier filtro ($h(\vec{x})$) y la energía total de la escena de entrada. Matemáticamente esto es:

$$\eta_H = \eta_M \frac{\int\int |f(\vec{x}) \otimes h^*(\vec{x})|^2 d\vec{x}}{\int\int |f(\vec{x})|^2 d\vec{x}}, \tag{59}$$

con η_M la eficiencia media. Horner y Gianino obtuvieron como resultado una mejora en la eficiencia óptica comparado con el filtro de correspondencia. Otra ventaja del filtro sólo fase es la ausencia de lóbulos laterales, generando así una mejor discriminacón entre objetos distintos en comparación con el filtro de correspondencia. Sin embargo sus desventajas con respecto al filtro de correspondencia radican en su pobre capacidad de discriminación para objetivos de bajo contraste incorporado en un fondo complicado y su bajo SNR en imágenes ruidosas.

3.3. Filtros de correlación avanzados

Para el procesamiento de imágenes, la mayor ventaja de los filtros de correlación recae en la habilidad de producir picos invariantes al desplazamiento y la simplicidad del procesamiento ya que se evita la necesidad de segmentación. Desafortunadamente, el filtro de correspondencia no es adecuado para el reconocimiento de patrones en la práctica, ya que su respuesta se degrada rápidamente cuando existen variaciones del objetivo. Por lo que la solución es diseñar filtros robustos que tomen en consideración estos problemas y tengan un mejor desempeño. El diseño de un buen filtro debe cumplir los siguientes tres objetivos:

- Reconocer distintas versiones distorsionadas del objeto de referencia.
- Tener un comportamiento robusto en presencia de ruido o confusión.
- Mantener una alta probabilidad de reconocimiento correcto manteniéndose el rango de error bajo.

3.4. Invarianza a distorsiones

Las distorsiones en imágenes debidas a cambios de escala y rotaciones en el plano pueden ser descritas matemáticamente en términos de transformaciones de coordenadas utilizando una sola imagen. A continuación se presentan algunos casos.

3.4.1. Una transformación de coordenadas básica

Sea $f(x, y)$ una imagen en coordenadas cartesianas. En la transformada ln-polar (LPT) $f(\rho, \theta)$ es calculada usando el mapeo (Kumar *et al.*, 2005):

$$\rho = \ln(r) = \ln\{\sqrt{(x - x_0)^2 + (y - y_0)^2}\} \tag{60}$$

y

$$\theta = tan^{-1}((y - y_0), (x - x_0)). \tag{61}$$

Si la imagen original esta escalada y rotada, la LPT esencialmente se desplaza y puede ser expresada $f(\rho + \tau_\rho, \theta + \tau_\theta)$ donde τ_ρ es el ln de factor de escalamiento y τ_θ es el ángulo de rotación. Por lo que LPT convierte los cambios de traslación y rotación en desplazamientos,

y después se puede aplicar el filtro MF Eq.(53) en el sistema de coordenadas de la transformación, pero en la práctica se vuelve muy elaborado.

La LPT se puede utilizar también para obtener un filtro tal que su respuesta pueda ser controlada en la presencia de cambios de rotaciones y traslaciones. Recordando que la salida del filtro es la correlación en el origen, que es lo mismo que el producto interno de la imagen y la función del filtro tenemos:

$$\int\int f(\rho+\tau_\rho,\theta+\tau_\theta)h(\rho,\theta)e^{2\rho}d\rho d\theta = c(\tau_\rho,\tau_\theta), \tag{62}$$

donde $J = e^{2\rho}$ es el jacobiano (matriz de derivadas) debido al cambio de coordenadas. Sea $\hat{h}(\rho,\theta) = h(\rho,\theta)e^{2\rho}$ tenemos:

$$f(\rho,\theta) \otimes \hat{h}(\rho,\theta) = c(\rho,\theta). \tag{63}$$

Finalmente el filtro se define como:

$$h(\rho,\theta) = e^{-2\rho} IFT \left\{ \frac{FT\{c(\rho,\theta)\}}{FT\{f(\rho,\theta)\}} \right\}. \tag{64}$$

Su limitación radica en que como el filtro está definido de tal manera que sólo existe una solución para $h(\rho,\theta)$ tal que cumpla las restricciones impuestas por $c(\rho,\theta)$ (el valor deseado del pico de correlación determinado por el usuario), el filtro no puede ser optimizado mediante otros tipos de criterios.

3.5. Filtros de correlación compuestos

Los filtros de correlación compuestos fueron desarrollados para abarcar un mayor número de distorsiones que no pueden ser modeladas matemáticamente mediante transformaciones de coordenadas. Los filtros compuestos se basan en varias imágenes de entrenamiento, las cuales son puntos de vista representativos del objeto o patrón a reconocer, por lo que es importante la elección adecuada de dichas imágenes para el diseño del filtro.

El objetivo de todo filtro compuesto es ser capaz de reconocer el objeto para el cual

fueron entrenados, desde cualquier punto de vista, y al mismo tiempo ser capaz de rechazar cualquier otra cosa distinta.

3.5.1. Funciones Discriminantes Sintéticas (SDF)

En este enfoque, el filtro es diseñado para dar un valor específico en el origen del plano de correlación en respuesta de cada imagen de entrenamiento. Por ejemplo, en el problema de dos clases, los valores de correlación en el origen deben ser 1 para las imágenes de una clase, y 0 para las imágenes de la otra (Kumar *et al.*, 2005).

Para desarrollar el marco teórico de los filtros SDF se asume que se tiene N imágenes de entrenamiento disponibles. Sea $u_i,\ 1 \leq i \leq N$ el valor en el origen del plano de correlación producido por el filtro $h(\vec{n})$ el cual se obtiene como una combinación lineal del conjunto de imágenes de entrenamiento como:

$$h(\vec{n}) = \sum_{i=1}^{N} w_i t_i(\vec{n}) \tag{65}$$

donde $w_i,\ 1 \leq i \leq N$ son los coeficientes de peso, los cuales son seleccionados de modo que satisfagan:

$$u_i = t_i(\vec{n}) \otimes h(\vec{n}). \tag{66}$$

Para obtener los N valores de los pesos w_i simultáneamente, se puede sustituir $h(\vec{n})$ Eq.(65) en la ecuación anterior Eq.(66), de la siguiente manera:

$$\sum_{i=1}^{N} w_i R_{ij} = u_i,\ j = 1, 2, \ldots, N \tag{67}$$

donde $R_{ij} = t_i(\vec{n}) * t_j(\vec{n})$. Este sistema de ecuaciones lineales se puede representar en notación matricial de la siguiente manera: sea R una matriz de $N \times d$. La i-ésima columna de R está dada por la i-ésima imagen de entrenamiento convertida en un vector en forma lexicográfica. Sea W y U vectores columna que representan w_i y u_i respectivamente. Las

ecuaciones anteriores pueden reescribirse como:

$$h = RW, \tag{68}$$

$$U = R^{*}h, \tag{69}$$

donde el superíndice * representa la transpuesta conjugada. Sustituyendo la Eq.(68) en Eq.(69) se obtiene:

$$U = (R^{*}R)W. \tag{70}$$

El elemento (i, j) de la matriz $S = R * R$ es el valor de la correlación cruzada en el origen de las imágenes $t_i(\vec{n})$ y $t_j(\vec{n})$. Si la matriz S es no singular, la solución del sistema está dada por:

$$W = (S)^{-1}U. \tag{71}$$

Finalmente el filtro SDF en forma de vector puede expresarse en el dominio espacial como:

$$h = R(S)^{-1}U. \tag{72}$$

El proceso para determinar la clase del patrón de prueba requiere la localización de los picos en la superficie de correlación. La posición del objetivo en la entrada es indicada por medio de la localización del pico. Se dice que el filtro "reconoció" el objetivo cuando el valor del pico excede un cierto umbral. Como sea, el pico se encuentra rodeado de lóbulos laterales grandes, los cuales pueden causar errores si exceden el pico principal.

3.5.2. Filtro de Mínimo Promedio de Energía de Correlación (MACE)

En la práctica, es necesario eliminar los lóbulos laterales para asegurar un pico de correlación agudo y reducir las oportunidades de error. Una forma de lograr esto es minimizar la energía en el plano de correlación. El promedio de energía de correlación (ACE) para N imágenes de entrenamiento se define como (Kumar *et al.*, 2005):

$$ACE = \frac{1}{d \cdot N} \sum_{i=1}^{N} \sum_{k}^{d} \sum_{l}^{d} 2|G_i(\vec{k})|^2, \tag{73}$$

donde $G_i(\vec{k}) = H(\vec{k})T_i^*(\vec{k})$ sustituyendo en la Eq.(73) tenemos:

$$ACE = \frac{1}{d \cdot N} \sum_{i=1}^{N} \sum_{k}^{d} \sum_{l}^{d} 2|H(\vec{k})|^2 |T_i(\vec{k})|^2, \tag{74}$$

en representación matricial tenemos:

$$ACE = h^+ Dh, \tag{75}$$

donde $D = \frac{1}{d \cdot N} \sum_{i=1}^{N} T_i^* T_i$ es una matriz diagonal. Finalmente para obtener el filtro MACE que minimice el ACE sujeto a la condición

$$T^+ h = d \cdot u, \tag{76}$$

se obtiene:

$$h = D^{-1}T(T^+ D^{-1} T)^{-1} U. \tag{77}$$

Aún cuando los filtros MACE reducen los lóbulos laterales del pico de correlación, existen ciertas desventajas de su uso. Primero, son sensibles a la presencia de ruido; y segundo, son sensibles a la variación entre-clases.

3.6. Filtros adaptativos compuestos

Como ya se mencionó anteriormente, el diseño de los filtros de correlación comúnmente se realiza optimizando un algún criterio de desempeño con respecto un modelo de señal para la escena de entrada. Dos modelos son comúnmente utilizados: aditivo (ruido traslapado) y no traslapado. El filtro que maximiza el criterio SNR para el modelo aditivo es el filtro de correspondencia (MF). Minimizando la probabilidad de falsas alarmas obtenemos el filtro óptimo (OF).

Debido a que los filtros de correlación utilizan una representación explícita, su desempeño disminuye significativamente cuando los objetos de interés sufren de algún cambio en su apariencia en la escena de entrada (debido a distorsiones geométricas, presencia de ruido, degradaciones). Cuando las distorsiones geométricas pueden ser modeladas matemáticamen-

te, entonces se puede utilizar esta información para el diseño de los filtros.

Otra forma de lidiar con distorsiones geométricas, tales como rotaciones y traslaciones, es el uso de filtros de correlación compuestos. Los filtros compuestos utilizan un conjunto de imágenes de entrenamiento las cuales incluyen múltiples puntos de vista del objeto a reconocer en la escena. Las funciones discriminantes sintéticas son una combinación lineal de estas imágenes de entrenamiento. Lamentablemente este tipo de filtros tienen un valor DC bajo ya que sólo controlan el valor en el pico de correlación, apareciendo lóbulos laterales en el área del fondo.

En el trabajo de J. González *et al.* (González-Fraga *et al.*, 2006) se propone un nuevo algoritmo para diseñar filtros SDF con una capacidad de discriminación dada. El filtro es adaptativo a la escena de entrada, el cual es diseñado a partir del objetivo, objetos falsos y el fondo a ser rechazado. El filtro es capaz de suprimir los lóbulos laterales generados por el fondo y objetos falsos.

El algoritmo propuesto consiste de los siguientes pasos:

1. Diseñar un filtro de funciones discriminantes sintéticas adaptativo (A-SDF) de la forma convencional SDF entrenándolo sólo con la imagen de referencia.

2. Llevar a cabo la correlación entre la imagen del fondo y el filtro A-SDF diseñado.

3. Calcular el valor DC.

4. Si el valor DC es mayor o igual al valor deseado, entonces el proceso del diseño del filtro es terminado; de otro modo continuar con el siguiente paso.

5. Crear un nuevo objeto a ser rechazado del fondo. El origen del objeto debe de estar en la posición del pico más alto del lóbulo lateral en el plano de correlación. El objeto es incluído en la clase de objetos a rechazar.

6. Diseñar un nuevo filtro A-SDF utilizando las dos clases (verdadera y falsa), volver al paso dos.

En cada iteración el algoritmo escoge entre todos los lóbulos laterales como un pico a ser suprimido, en el siguiente paso para asegurar el decrecimiento monótono de la función DC

contra el índice de iteración durante el diseño del filtro.

Sin embargo, se ha demostrado que el valor SNR de los filtros compuestos gradualmente se degrada cuando un número creciente de imágenes se incluyen en el conjunto de entrenamiento. Este problema puede ser resuelto utilizando un banco de filtros compuestos, donde cada uno es diseñado con un subconjunto de imágenes de entrenamiento del objetivo. En este caso, la detección se lleva a cabo correlacionando la escena de entrada con cada uno de los filtros en el banco. Posteriormente, los DC resultantes en cada uno de los planos de correlación son comparados y el plano con el valor DC más alto es escogido. A continuación se presenta un algoritmo que utiliza este enfoque.

3.7. Filtros adaptativos compuestos y banco de filtros

En el trabajo de P. Aguilar *et al.* (Aguilar-González *et al.*, 2014) se propone un algoritmo para el diseño de filtros adaptativos compuestos para la detección y localización de un objeto en una escena no traslapada. Estos filtros están diseñados como una combinación lineal de filtros óptimos y de correspondencia. Las imágenes de entrenamiento del objeto a reconocer se encuentran contaminadas de ruido y su forma y valores de intensidad no se conocen explicitamente.

Utilizando el algoritmo de González (González-Fraga *et al.*, 2006) se construyen los filtros adaptativos compuestos, posteriormente cada uno de estos filtros diseñados se encuentran en un banco de filtros. La imagen de entrada se correlaciona con cada uno de los filtros en el banco y finalmente, el plano de correlación con un valor DC mayor es escogido como la salida del sistema.

El problema con la correlación lineal es que funciona bien sólo si la escena de prueba contiene replicas exactas de la imagen de referencia (objetivo), y si no existen otros objetos similares a la imagen de referencia.

Otra deficiencia del desempeño de la correlación es su sensibilidad hacia el ruido, arrojando salidas erróneas. Los picos agudos de correlación son muy importantes en la estimación

de la localización de la imagen de referencia en la escena de prueba. Es más fácil localizar el objetivo si la plantilla es diseñada para producir picos agudos. Desafortunadamente, la tolerancia al ruido y la agudeza de los picos son criterios conflictivos, y es necesario el diseñar técnicas que optimicen este compromiso entre los dos criterios en conflicto.

Además la correlación no funciona bien si la imagen de referencia aparece en la escena de prueba con algunos cambios en la apariencia (distorsiones), así como cambios en la iluminación, distintas perspectivas, etc.

3.8. Resumen

En este capítulo se hace una breve descripción del planteamiento matemático de algunos filtros lineales de correlación. Esto con el fin de entender en que se basan y como se definen, así como su diferencia con los filtros no lineales utilizados en este trabajo.

Capítulo 4. Filtros no lineales

El filtrado no lineal es un procesamiento que puede ser global o local utilizando una ventana con desplazamiento. La ventana es una vecindad que contiene los pixeles de alrededor del pixel central de forma geométrica (W-vecindad). La forma y tamaño de la W-vecindad es similar a la región de soporte del objetivo. Se asume que la W-vecindad es lo suficientemente pequeña como para que la señal y el ruido puedan ser considerados estacionarios sobre el área de la ventana.

Los filtros no lineales tales como filtros pila, filtros de estadísticas de orden, morfológicos, etc., resultan muy efectivos en la eliminación de ruido aditivo e impulsivo, realce y restauración de imágenes. Además se ha demostrado que son más robustos y proveen soluciones en muchos casos donde los filtros lineales son inapropiados (Kober *et al.*, 2001).

4.1. Filtros de estadísticas de orden

Tres tipos de estimadores derivados de la de la estimación robusta deben ser usados en la estimación del pixel central de las vecindades: el L-estimador basado en la combinación lineal de estadísticas de orden; el R-estimador que se deriva de cálculos basados en el rango de los pixeles de una vecindad; y el M-estimador o estimador máxima verosimilitud. Todos los estimadores pueden ser implementados usando unas cuantas operaciones sobre los vecindarios (NBH) (Kober *et al.*, 2002).

Algunos filtros de estadísticas de orden son los siguientes:

- $SIZE(NBH)$: es la cantidad de pixeles que componen la vecindad.
- $MEAN(NBH)$: es el promedio de los pixeles sobre la vecindad.
- $MED(NBH)$: es la mediana de los pixeles sobre la vecindad.
- $MIN(NBH)$: es el mínimo de los pixeles sobre la vecindad.
- $MAX(NBH)$: es el máximo de los pixeles sobre la vecindad.

En general, si el ruido es de tipo Gaussiano, entonces la mejor estimación está dada por la operación promedio de la muestra; si la distribución del ruido tiene colas pesadas, entonces la operación mediana dará mejor resultado. En el caso de distribuciones de ruido de un sólo lado, las operación mínimo o máximo será más apropiada.

4.1.1. Algoritmos de realce y mejora local de imagen

El objetivo del diseño de filtros de estadísticas de orden es el de preservar o modificar en una forma deseada las estructuras de la imagen original, mientras se elimina ruido y estructuras no deseadas (Kober *et al.*, 2001).

A continuación se describen algunos algoritmos utilizando filtros de estadísticas de orden y vecindarios (NBH).

4.1.1.1. Algoritmo de suavizado

El algoritmo de suavizado basado en la operación promedio y la EV-vecindad (Eq.(10)) es el siguiente:

$$\hat{v}_{m,n}^{i+1} = MEAN(EV[\hat{v}_{m,n}^{i}]) \tag{78}$$

donde el primer elemento $\hat{v}_{m,n}^{1} = s_{m,n}$ y se recomienda $\epsilon_v = 1{,}5\sigma$.

4.1.1.2. Algoritmo para eliminar ruido impulsivo y aditivo mezclado

$$\hat{v}_{n,m}^{i+1} = \begin{cases} MEAN(EV[\hat{v}_{m,n}^{i}]), & \text{si } SIZE(EV[\hat{v}_{m,n}^{i}]) \geq umbral^{i} \\ MED(S[\hat{v}_{m,n}^{i}] - EV[\hat{v}_{m,n}^{i}]), & \text{otra forma} \end{cases} \tag{79}$$

lo que hace este algoritmo es suavizar aquellas áreas donde la variación de las intensidades sea menor y eliminar aquellos pixeles que pertenecen a algún tipo de ruido.

4.1.1.3. Algoritmo para realce de imagen

Una imagen puede considerarse como la suma dos señales, detalle (D) y fondo (F):

$$v_{m,n} = v_{m,n}^{D} + v_{m,n}^{F} \tag{80}$$

Un método directo para reducir la pequeña variación local de una imagen, y al mismo tiempo incrementar la variación en los detalles, es sustrayendo una estimación local del fondo de la imagen. Por lo tanto, el algoritmo puede ser descrito de la siguiente manera (Kober *et al.*, 2001)

$$\hat{v}_{m,n} = C(v_{m,n} - SMTH\lfloor NBH(v_{m,n})\rfloor) \tag{81}$$

donde SMTH es una operación local de suavizado descrita previamente, y C es una constante de ganancia.

4.2. Filtros morfológicos

los filtros morfológicos son transformaciones no lineales que modifican localmente las características geométricas de señales/imágenes. Se basan en operadores de un conjunto de métodos para el análisis de imágenes, llamado morfología matemática. En esta metodología, cada señal es vista como un conjunto en un espacio Euclidiano, y los filtros morfológicos son un conjunto de operaciones que transforman la señal y provee una descripción cuantitativa de su estructura geométrica. Para señales binarias (vistas como conjuntos), la erosión, dilatación, apertura y cierre son las operaciones morfológicas más simples, y en base a éstas se pueden diseñar un mayor número de filtros morfológicos (Serra, 1986).

Dada una imagen binaria $f[x]$ con valores 1 para los objetos y 0 para el fondo de la imagen. Sea $W = y_1, y_2, \ldots, y_n$ una ventana deslizante de tamaño n, entonces

$$\varphi_b(f)[x] = b(f[x - y_1], \ldots, f[x - y_n]) \tag{82}$$

donde $b(v_1, v_2, \ldots, v_n)$ es una función booleana de n variables, el mapeo $f \to \varphi_b(f)$ es llamado filtro morfológico. La aplicación principal de éste tipo de filtros se encuentran en el procesamiento de imágenes biomédicas, reconocimiento de caracteres, detección de objetos, entre otros (Maragos, 2004).

De los conceptos dados de la morfología matemática, podemos utilizar conjuntos para representar los objetos de una imagen y las operaciones entre conjuntos para representar las transformaciones binarias en las imágenes. Específicamente, dada una imagen binaria, se

pueden representar los objetos de la imagen por un conjunto X y el fondo por el complemento del conjunto, denotado por X^c. Cuando se aplica la operación logica OR entre el objeto X y un elemento estructural B se obtiene la dilatación ($\oplus$) de X por B definida de la siguiente manera:

$$X \oplus B = \{x + y : x \in X, y \in B\} = \bigcup_{y \in B} X_{+y} \tag{83}$$

donde $X_{+y} = \{x + y : x \in X\}$ es la traslación de X sobre el vector y. De manera similar, aplicando la transformación AND se obtiene la erosión ($\ominus$) de X por B definida de la siguiente manera:

$$X \ominus B = \{x : B_{+x} \subseteq X\} = \bigcap_{y \in B} X_{-y} \tag{84}$$

Aplicando las operaciones de erosión y dilatación en cascada se obtienen las operaciones de apertura($\circ$) y cierre($\bullet$)

$$X \circ B = (X - B) \oplus B \tag{85}$$

$$X \bullet B = (X \oplus B) - B \tag{86}$$

respectivamente. La combinación de las operaciones anteriores permite efectuar diferentes labores de filtrado en una imagen, por ejemplo, suavizar los bordes de la imagen o remover ruido impulsivo.

La operación de erosión se puede utilizar para efectuar reconocimiento de objetos dentro de una imagen binaria, esto se debe a que dicha operación da como salida los puntos centrales en los cuales el elemento estructural coincide con algún objeto dentro de la imagen.

4.2.1. Propiedades de los operadores morfológicos

- **Inclusión**: Aunque las señales acústicas son sumadas, las señales visuales no están compuestas de esta manera. El mundo que nos rodea no es traslucido; al contrario, está compuesto de objetos opacos que se ocultan unos detrás de otros. El término inclusión es utilizado para expresar esta ley fundamental. (*Si el contorno del objeto más cercano contiene el contorno de objetos más alejados, éstos estarán completamente fuera de la vista, de lo contrario no podrán ocultarse*). Esta relación de inclusión en el

universo visual remplaza la aditividad de la percepción acústica. Por lo tanto el primer pre-requisito para cualquier filtro morfológico Φ es que satisfaga la siguiente condición:

$$F \leq g \Rightarrow \Phi(f) \leq \Phi(g) \tag{87}$$

Donde f y g son funciones reales en $\Re^n$ y Φ es un mapeo de funciones con valores reales de $\Re^n$ a $\Re^n$. Si cumple esto decimos que Φ es un transformador creciente. Lo cual simplemente indica que si f es el fondo y g es un objeto en primer plano (el cual, por lo tanto, cubre una parte de f) entonces $\Phi(g)$ cubre las porciones correspondientes de $\Phi(f)$.

- **Idempotencia**: Las relaciónes de contraste (intersección, inclusión y disyunción) son preservadas por el filtrado morfológico debido a que es una transformación creciente; pero esto a su vez produce una pérdida de información, la cual se trata de controlar sumando un segundo axioma a la definición de filtrado morfológico denominado idempotencia, es decir, que al aplicar el mismo operador dos veces se obtiene el mismo resultado.

$$\Phi[\Phi(f)] = \Phi(f) \tag{88}$$

En el caso general de un espacio E no existen requerimientos adicionales, pero en el caso de que E sea un espacio Euclidiano se suman continuidad e invarianza a la traslación tal como se da en los filtros lineales. Basándose en estas dos condiciones de ser creciente e idempotente, se puede construir una teoría de filtrado muy poderosa

- **Dualidad**: Las transformaciones morfológicas vienen dadas en pares, tan pronto como se define una transformación Φ, se debe introducir otra transformación Φ^* llamada transformación dual de Φ; y se define de la siguiente manera:

$$\phi^*(X) = [\phi(X^c)]^c \tag{89}$$

donde X^c es el **complemento** del conjunto X tal que $X \cap X^c = \emptyset$ y $X \cup X^c = E$.

- **Extensividad**: Finalmente, sea X un conjunto de un espacio E y Φ un operador

morfológico que actúa sobre el conjunto X, tenemos que $\Phi(X)$ es **extensiva** si:

$$Si\ X \subset \Phi(X) \tag{90}$$

y **anti-extensiva** si:

$$Si\ \Phi(X) \subset X \tag{91}$$

La teoría de los filtros morfológicos, hace énfasis en las propiedades de crecimiento e idempotencia, así como en las reglas de orden de las transformaciones (Serra, 1986).

Existe una familia de transformaciones que estudia la diferencia entre dos (o más) transformaciones básicas, llamada **diferencia** o **residuo**. Los residuos más utilizados en la práctica se clasifican básicamente en tres tipos:

1. Residuos de dos primitivas.
2. Residuos de dos familias de primitivas.
3. Residuos basados en transformaciones Hit-Miss.

4.2.2. Hit-Miss

La operación de erosión se puede utilizar para efectuar reconocimiento de objetos dentro de una imagen binaria, esto se debe a que dicha operación da como salida los puntos centrales en los cuales el elemento estructural coincide con algún objeto dentro de la imagen. Sea X un objeto binario dentro de una imagen, la cual se erosiona con el elemento estructural A y sea X^c el complemento de X, el cual lo erosiona por el elemento estructural B; el conjunto de puntos en los cuales la versión desplazada de (A, B) coincide dentro de la imagen X se llama la transformada de acierto-fallo (hit-miss $\odot$) de X por (A, B) y se denota por:

$$X \odot (A, B) = \{x : A_{+x} \subseteq X, B_{+x} \subseteq X^c\} \tag{92}$$

Esta operación se ha utilizado ampliamente para la detección de objetos en imágenes binarias.

4.3. Plantilla dual

La correspondencia de imágenes y la detección de objetos son dos de los principales temas en muchos problemas de la visión por computadora. Muchos de los esfuerzos por resolver estos problemas se basan en la minimización del error cuadrático medio entre ambas imágenes, lo que conduce a la maximización de la correlación entre ellas. Su popularidad se basa en su tratabilidad matemática y su fácil implementación. Pero a pesar de esto, la norma l_1 la cual se define como la suma de mínimos (Eq.18) tiene algunas ventajas importantes; las cuales se verán a continuación (Maragos, 1988).

4.3.1. Correspondencia de imágenes bajo la norma l_1

Sea $f(\vec{n})$y $g(\vec{n})$ dos imágenes reales a ser comparadas. Ambas imágenes se encuentran definidas en el plano cartesiano discreto con pares de coordenadas $\vec{n} = (n_1, n_2) \in Z^2$. Como una medida de correspondencia entre f y g se considera la suma de las diferencias absolutas:

$$E(\vec{k}) = \sum_{\vec{n} \in W} |f(\vec{n} + \vec{k}) - g(\vec{n})| \tag{93}$$

Donde $\vec{k}$ es un vector $\vec{k} = (k_1, k_2)$ que representa el desplazamiento de f y W es un subconjunto de Z^2. Si W es igual a todo el plano Z^2 entonces E se convierte en la norma l_1 de la diferencia entre las señales f y g.

Minimizar la norma l_1 en lugar de la norma l_2 da lugar a tres ventajas. Primero, la norma l_1 es más rápida y fácil de calcular (computacionalmente) que la norma l_2. La norma l_1 es más robusta en presencia de ruido no gaussiano comparada con la norma l_2. Y finalmente, debido a que la norma l_1 es mayor a la norma l_2, minimizar el error basado en la norma l_1 puede resultar en una mejor correspondencia entre f y g que minimizar el error de la norma l_2.

Ahora bien, sea $|a - b| = a + b - 2min(a, b)$ para cualquier par de reales a, b entonces,

$$E(\vec{k}) = \sum_{\vec{n} \in W} f(\vec{n} + \vec{k}) + \sum_{\vec{n} \in W} g(\vec{n}) - 2 \sum_{\vec{n} \in W} min[f(\vec{n} + \vec{k}), g(\vec{n})] \tag{94}$$

Existen distintas formas de hacer que la suma $\sum_{\vec{n} \in W} f(\vec{n} + \vec{k}) + \sum_{\vec{n} \in W} g(\vec{n})$ no afecte la

minimización del error $E(\vec{k})$. En dichos casos, minimizar $E(\vec{k})$ es equivalente a maximizar la correlación no lineal

$$M_{fg} = \sum_{\vec{n}\in W} min[f(\vec{n}+\vec{k}), g(\vec{n})] \tag{95}$$

llamada correlación morfológica. Por ejemplo, si f y g son imágenes del mismo tamaño, mayor al posible desplazamiento $\vec{k}$ y cero fuera de sus dominios, entonces la ventana W será igual al conjunto Z^2. En este caso el factor $\sum_{\vec{n}\in W} f(\vec{n}+\vec{k}) + \sum_{\vec{n}\in W} g(\vec{n})$ es igual a el área bajo f y g; por lo que es constante y no afecta la minimización de $E(\vec{k})$.

Ahora, asumiendo que el dominio G de g es menor que el de f se puede considerar g como una plantilla, la cual se desea detectar en distintas posiciones $\vec{k}$ en la imagen mayor f.

4.3.2. Correspondencia de plantilla

Sea g una plantilla[1], entonces $\sum_{\vec{n}\in W} g(\vec{n})$ es un factor constante igual al área bajo g, por lo que podemos ver que minimizar $E(\vec{k})$ es equivalente a minimizar Eq.(94)

$$E_g(\vec{k}) = \sum_{\vec{n}\in W} f(\vec{n}+\vec{k}) - 2\sum_{\vec{n}\in W} min[f(\vec{n}+\vec{k}), g(\vec{n})]. \tag{96}$$

Realizando descomposición por umbral de f y g utilizando la Eq.(14), y definiendo una plantilla dual $g^*(\vec{n}) = T - g(\vec{n})$ con T el número de niveles de gris de la escena y $\vec{n} \in W/G$, obtenemos:

$$E_g(\vec{k}) = \sum_{\vec{n}\in W}\sum_{q}[f_q(\vec{n}+\vec{k}) - 2f_q(\vec{n}+\vec{k})g_q(\vec{n})] \tag{97}$$

$$= \sum_{\vec{n}\in W}\sum_{q}[f_q(\vec{n}+\vec{k})[g_q(\vec{n}) + g_q^*(\vec{n})] - 2f_q(\vec{n}+\vec{k})g_q(\vec{n})] \tag{98}$$

$$= \sum_{\vec{n}\in W}\sum_{q}[f_q(\vec{n}+\vec{k})g_q^*(\vec{n}) - f_q(\vec{n}+\vec{k})g_q(\vec{n})] \tag{99}$$

$$= -\sum_{\vec{n}\in W}\sum_{q} f_q(\vec{n}+\vec{k})g_q(\vec{n}) + \sum_{\vec{n}\in W}\sum_{q} f_q(\vec{n}+\vec{k})g_q^*(\vec{n}) \tag{100}$$

Ahora, sea $g(\vec{n}) = T \geq 1 \forall \vec{n} \in G$ y 0 en otro caso. Entonces todas las capas binarias de g se

[1] template

vuelven idénticas, de la misma forma para su dual, es decir, $g^*(\vec{n}) = T - g(n) \; \forall \; \vec{n} \in W/G$ y 0 en otro caso. Finalmente obtenemos que:

$$E_g(\vec{k}) = -M_{fg}(\vec{k}) + M_{fg^*}(\vec{k}) \tag{101}$$

es decir,

$$-E_g(\vec{k}) = M_{fg^*}(\vec{k}) - M_{fg}(\vec{k}). \tag{102}$$

Por lo tanto, minimizar $E_g(\vec{k})$ equivale a maximizar la diferencia entre las correlaciones morfológicas de f con g y su dual g^*.

4.4. Máscaras binarias

4.4.1. Máscara binaria de anillos concéntricos

J. Álvarez y colaboradores (Álvarez-Borrego *et al.*, 2013) proponen el uso de máscaras binarias para el reconocimiento de objetos invariantes a rotaciones y posición. El método consiste en obtener la parte imaginaria $f_i(x, y)$ y real $f_r(x, y)$ de la transformada de Fourier de la imagen de referencia I para posteriormente generar cuatro máscaras binarias de anillos concéntricos. Estas máscaras se obtienen utilizando tanto la parte real como la imaginaria de la siguiente manera: primero se fija la posición $x = c_x$ donde (c_x, c_y) es el pixel central de la imagen I y y varía de 0 hasta N donde $M \times N$ es el tamaño de I. Utilizando lo anterior, se definen las siguientes funciones univariables:

$$Z_r(y) = \begin{cases} 1, & \text{si } f_r(c_x, y) > 0 \\ 0, & \text{otro caso} \end{cases} \tag{103}$$

$$Z_{r\_inv}(y) = \begin{cases} 1, & \text{si } f_r(c_x, y) \leq 0 \\ 0, & \text{otro caso} \end{cases} \tag{104}$$

$$Z_i(y) = \begin{cases} 1, & \text{si } f_i(c_x, y) > 0 \\ 0, & \text{otro caso} \end{cases} \tag{105}$$

$$Z_{i\_inv}(y) = \begin{cases} 1, & \text{si } f_i(c_x, y) \leq 0 \\ 0, & \text{otro caso} \end{cases} \tag{106}$$

Posteriormente, la gráfica de Z_r es rotada 360° sobre el eje x generando cilindros concéntricos y finalmente transformando estos cilindros al espacio de dos dimensiones se obtiene la máscara de anillos binarias asociadas a la imagen I. Lo mismo sucede para las gráficas restantes.

Una vez que se obtienen las cuatro máscaras binarias se procede a obtener una firma de la imagen con el objetivo de identificar la imagen de referencia sin importar el ángulo de rotación y su posición en el plano. La firma se obtiene de multiplicar el módulo de I por cada una de las máscaras binarias y sumar los valores de cada uno de los anillos asignando el valor resultante al índice del anillo correspondiente, obteniendo así 4 firmas diferentes. Finalmente los autores recomiendan el uso de 10 y 18 imágenes para obtener el mejor filtro firma promedio.

4.4.2. Filtro sólo fase y bloqueo de frecuencias

En el trabajo realizado por Kober y Mozerov (Kober y Mozerov, 2000) se propone un nuevo método para mejorar el desempeño de los filtros sólo fase haciendo uso de máscaras binarias. A continuación se describe el método.

Sean

$$\begin{aligned} T(\vec{v}) &= |T(\vec{v})| exp[i\Phi_t(\vec{v})], \\ B(\vec{v}) &= |B(\vec{v})| exp[i\Phi_b(\vec{v})], \end{aligned} \tag{107}$$

las transformadas de Fourier del objeto a reconocer $t(\vec{x})$ y del objeto a rechazar $b(\vec{x})$ respectivamente. Aquí, $\Phi_t(\vec{v})$ y $\Phi_b(\vec{v})$ son la distribución de sus fases, y $|T(\vec{v})|, |B(\vec{v})|$ sus magnitudes respectivamente. Cuando el filtro POF es utilizado, la capacidad de discrimina-

ción 41 puede rescribirse de la siguiente manera:

$$DC = 1 - \frac{|\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}|B(\vec{v})|exp[i\Delta\Phi(\vec{v})]d\vec{v}|^2}{|\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}|T(\vec{v})|d\vec{v}|^2}, \tag{108}$$

donde $\Delta\Phi(\vec{v}) = \Phi_t(\vec{v}) - \Phi_b(\vec{v})$ es la diferencia de fases.

Ahora, se desea que el filtro diseñado por la técnica de bloqueo de frecuencias tenga el mínimo número elementos de frecuencias bloqueadas. Por lo que el diseño del filtro puede ser fácilmente implementado mediante el siguiente algoritmo con un mínimo de elementos bloqueados.

Se supone que la escena de entrada es real y que contiene la imagen de referencia $t(\vec{x})$ y el objeto a rechazar $b(\vec{x})$. Debido a la simetría de $T(\vec{v})$ y $B(\vec{v})$, podemos reescribir la Eq. 108 como:

$$DC = 1 - \frac{|2\int_{-\infty}^{\infty}\int_{0}^{\infty}|B(\vec{v})|cos[i\Delta\Phi(\vec{v})]d\vec{v}|^2}{|2\int_{-\infty}^{\infty}\int_{0}^{\infty}|T(\vec{v})|d\vec{v}|^2}. \tag{109}$$

Sustituyendo un valor predefinido de DC, como por ejemplo $1-\epsilon^2$, en la Eq.109 y haciendo uso de la fórmula $(|a|^2 - |b|^2) = (a^2 - b^2) = (a-b)(a+b)$ obtenemos:

$$(\int_{-\infty}^{\infty}\int_{0}^{\infty}|B(\vec{v})|cos[i\Delta\Phi(\vec{v})] - \epsilon|T(\vec{v})|d\vec{v}) \times (\int_{-\infty}^{\infty}\int_{0}^{\infty}|B(\vec{v})|cos[i\Delta\Phi(\vec{v})] + \epsilon|T(\vec{v})|d\vec{v}) = 0. \tag{110}$$

El objetivo es forzar el producto a ser cero. Nótese que el producto es igual a cero si en la Eq.110 al menos uno de los dos productos es cero. Esto puede lograrse mediante el bloqueo de algunas frecuencias. A continuación se describe el algoritmo utilizado para lograr esto.

Primero, las expresiones del primer producto de la Eq.110 son ordenadas en un renglón variacional de acuerdo a su magnitud; después, se escoge el mínimo número de elementos a ser bloqueados por ceros para satisfacer la condición de igualdad en la Eq.110. Entonces, si el cálculo del primer paréntesis da mayor a cero, hacemos los primeros elementos del renglón variacional iguales a cero, empezando por el valor más grande, bloqueando así el mínimo número de elementos de la primera expresión en Eq.110. Posteriormente, este mismo procedimiento se repite para los elementos del segundo paréntesis. Generalmente, para el caso

de imágenes positivas y reales es necesario solamente tomar en consideración los elementos de la primera expresión en la Eq.110.

4.5. Filtros de Funciones Discriminantes Sintéticas No Lineales (NSDF)

Los filtros NSDF están basados en operaciones lógicas y la operación de correlación morfologica(20). Los filtros incorporan información de varios objetos de la clase falsa y verdadera. Finalmente, la salida se normaliza para obtener un valor de correlación deseado. A continuación se define formalmente los filtros SDF no lineales compuestos (Martínez-Díaz y Kober, 2008b).

Sea $\{T_i(\vec{k}), i = 1, \ldots, N\}$ un conjunto de N imágenes de entrenamiento en escala de grises pertenecientes a la clase verdadera (la clase que se desea reconocer) y sea $\{P_i(\vec{k}), i = 1, \ldots, M\}$ un conjunto de M imágenes pertenecientes a la clase falsa (la clase que se desea rechazar). El fitro NSDF se define como:

$$H_{NSDF}(\vec{k}) = \sum_{i=1}^{Q-1} [\bigcap_{i=1}^{N} T_i^q(\vec{k})] \bigcap \overline{[\bigcup_{i=1}^{M} P_i^q(\vec{k})]}, \tag{111}$$

donde $\{T_i^q(\vec{k}), q = 1, \ldots, Q-1, i = 1, \ldots, N\}$ y $\{P_i^q(\vec{k}), q = 1, \ldots, Q-1, i = 1, \ldots, M\}$ son las imágenes binarias obtenidas de la descomposición por umbral de las imágenes de la clase verdadera y falsa, respectivamente. Aquí $\bigcap_{i=1}^{N}$ representa la intersección lógica entre las imágenes binarias y $\bigcup_{i=1}^{M}$ representa la unión lógica entre las imágenes binarias. Una vez sintetizado el filtro, se calcula la correlación usando la Eq.(20). El resultado se normaliza multiplicando por $\frac{u}{\bar{t}}$ donde u es el valor de correlación deseado en el origen de las coordenadas y $\bar{t} = \sum_{\vec{k} \in W} H_{NSDF}(\vec{k})$. Al calcular la correlación, la región de soporte del filtro es tomada como la intersección lógica de las regiones de soporte de las imágenes de entrenamiento pertenecientes a la clase verdadera.

4.5.1. Reconocimiento entre dos clases de objetos

Los filtros NSDF se pueden utilizar para reconocer objetos de dos clases diferentes. Primero se asignan los objetos de una de las clases al conjunto $\{T_i(\vec{k}), i = 1, \ldots, N\}$ y los de la otra clase al conjunto $\{P_i(\vec{k}), i = 1, \ldots, M\}$. Posteriormente se diseña un filtro NSDF y

se aplica el proceso de correlación con la escena de prueba. Como resultado, al encontrarse un objeto del primer conjunto el valor de correlación es u mientras que para los objetos del segundo conjunto el valor es cero.

4.5.2. Reconocimiento de objetos de la misma clase

para reconocer los objetos de la misma clase se puede utilizar una versión de los filtros propuestos en la cual se incluya sólo los objetos de la clase verdadera. Sea $\{T_i(\vec{k}), i = 1, \ldots, N\}$ un conjunto de N imágenes de entrenamiento pertenecientes a la clase verdadera. El filtro NSDF se construye de la siguiente manera:

$$H_{NSDF}(\vec{k}) = \sum_{i=1}^{Q-1} \bigcap_{i=1}^{N} T_i^q(\vec{k}) \tag{112}$$

El resultado de la intersección en las coordenadas $\vec{k}$ es uno si los pixeles correspondientes son uno, de otro modo el resultado es cero. Nuevamente el filtro se correlaciona con la escena de entra utilizando la Eq.(20). Finalmente el resultado es normalizado por $u/\overline{t}$ y la región de soporte del filtro es tomada como la intersección lógica de las regiones de soporte de las imágenes de entrenamiento.

4.5.3. Correlación compuesta

Cuando el conjunto de imágenes de entrenamiento es muy grande, el desempeño del filtro suele disminuir. En esta situación puede ser preferible utilizar varios filtros compuestos para el reconocimiento. La correlación compuesta se define de la siguiente manera. Sea $\{T_i(\vec{k}), i = 1, \ldots, N\}$ un conjunto de N imágenes de entrenamiento en escala de grises pertenecientes a la clase verdadera y sea $\{P_i(\vec{k}), i = 1, \ldots, M\}$ un conjunto de M imágenes pertenecientes a la clase falsa. El fitro NSDF se define como:

$$H_{NSDF}^i(\vec{k}) = \sum_{q=1}^{Q-1} [\bigcap_{i=1}^{N} T_i^q(\vec{k})] \bigcap \overline{[\bigcup_{i=1}^{M} P_i^q(\vec{k})]}. \tag{113}$$

La correlación conpuesta se define como:

$$\widehat{C}(\vec{k}) = MAX(\{\frac{u}{\overline{t}_i} C_i(\vec{k}), i = 1, \ldots, N\}) \tag{114}$$

donde $\widehat{C}(\vec{k})$ es la correlación compuesta y $C_i(\vec{k})$ es la i-ésima correlación entre la escena de prueba y el i-ésimo filtro $H^i_{NSDF}(113)$, calculada con la Eq.(20).

4.6. Resumen

En este capítulo se presentaron algunos filtros no lineales. Los filtros no lineales suelen tener mejores resultados en situaciones donde los filtros lineales fallan, como por ejemplo la eliminación de ruido impulsivo conservando bordes o lineas delgadas. Algunos de los filtros mencionados aquí, serán utilizados para el desarrollo de este trabajo.

Capítulo 5. Sistema OCR propuesto

Un problema frecuente en el uso de técnicas de correlación para el reconocimiento de objetos es que cuando la iluminación o el objeto de referencia varía en la escena de entrada, el desempeño del filtro suele disminuir. Como una alternativa se encuentra el uso de filtros no lineales, los cuales son más robustos a ruido no Gaussiano y en algunos casos cuentan con un mejor desempeño.

En este capítulo se describe el sistema OCR propuesto utilizando el enfoque de descomposición por umbral y correlación morfológica; el sistema se compone de dos etapas. La primera etapa de pre-procesamiento, en la cual se busca mejorar la calidad de la imagen suprimiendo el ruido presente y homogéneizando la iluminación; y la segunda etapa de reconocimiento y clasificación de los caracteres.

5.1. Pre-procesamiento

El objetivo de esta primera etapa de pre-procesamiento es mejorar la calidad de la imagen de manera que aumente el desempeño en la etapa de reconocimiento. Para lograr esto, se utilizan distintos tipos de filtros no lineales, a continuación se describe cada uno de ellos.

5.1.1. Filtro adaptativo de estadísticas de orden

Primero, para suprimir la mezcla de ruido aditivo e impulsivo de la escena de entrada $S(\vec{x})$, se utilizó un filtro iterativo no lineal de estadísticas de orden[1], el cual se describe a continuación:

$$\hat{p}_{\vec{n}}^{i+1} = \begin{cases} MEAN(CEV[\hat{p}_{m,n}^{i}]), & \text{si } SIZE(CEV[\hat{p}_{\vec{n}}^{i}]) \geq umbral^{i} \\ MED(S[\hat{p}_{\vec{n}}^{i}] - CEV[\hat{p}_{\vec{n}}^{i}]), & \text{otra forma} \end{cases} \tag{115}$$

donde i es el número de iteraciones, con $\hat{p}_{\vec{n}}^{1} = \hat{S}_{\vec{n}}$ para la primera iteración ; $umbral^{i}$ es un umbral previamente definido; finalmente CEV es un subconjunto de pixeles $\{p_{\vec{n}}\}$ de la vecindad los cuales están espacialmente conectados Δ con el pixel central $\{p_{\vec{x}}\}$ y cuyos valores

[1] Rank-order filter

se desvían del valor del pixel central a lo más por determinadas cantidades $-\epsilon_p$ y ϵ_p,

$$CEV(p_{\vec{x}}) = CON_{\Delta}(\{p_{\vec{n}} : p_{\vec{x}} - \epsilon \leq p_{\vec{n}} \leq p_{\vec{x}} + \epsilon\}). \tag{116}$$

Para nuestros experimentos se utilizó $i = 3$, $umbral = 2$, y $\Delta = 1$.

5.1.2. Filtro de realce

Posteriormente, se utiliza un filtro local de realce[2] para aumentar el contraste entre el fondo y los caracteres de la escena de entrada $S(\vec{x})$, de la siguiente manera:

$$\widehat{p}_{\vec{x}} = p_{\vec{x}} + C\lfloor p_{\vec{x}} - MEAN(NBH(p_{\vec{x}}))\rfloor \tag{117}$$

donde $p_{\vec{x}} = (x_1, x_2)$ son las coordenadas en la escena de entrada; $p_{\vec{x}}$ y $\widehat{p}_{\vec{x}}$ son los $\vec{x}$-ésimos pixeles de la escena de entrada y de la escena resultante respectivamente; C es una constante de realce y NBH es el vecindario centrado en cada pixel $p_{\vec{x}}$. Para este trabajo se utilizó $C = 5$ y una vecindad de 9×9 pixeles.

Finalmente se utiliza una ecualización local para homogéneizar localmente los valores de la imagen.

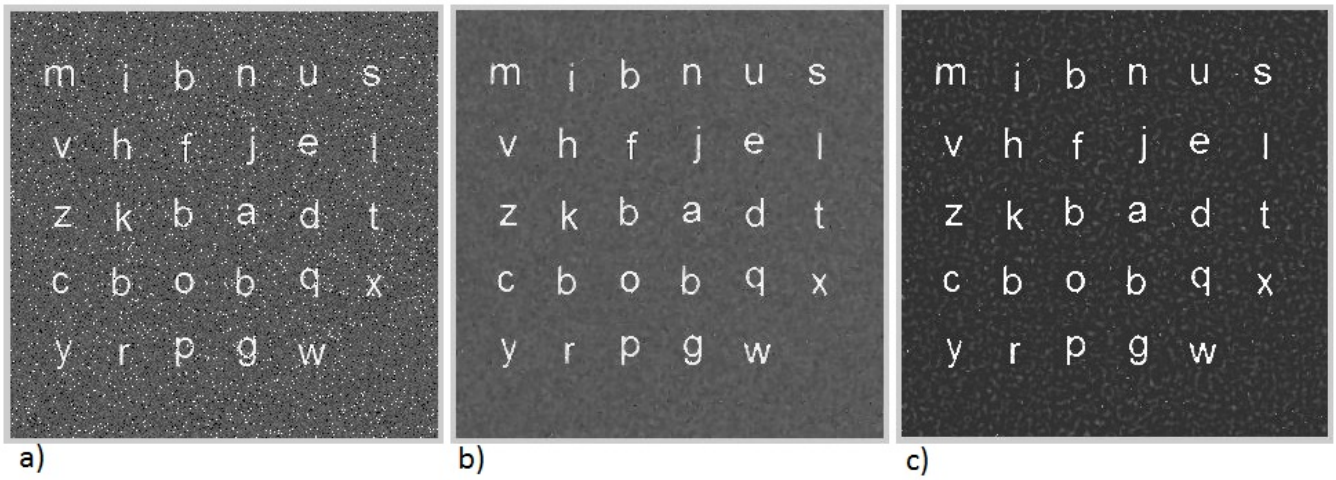

Figura 7: a) Imagen sintética degradada con mezcla de ruido aditivo e impulsivo b) imagen resultante al aplicar filtro adaptativo de estadísticas de orden c) imagen resultante después de aplicar a imagen (b) filtro de realce.

[2]Unsharp filter

5.1.3. Normalización de escena de entrada

Por último, para corregir los efectos de la iluminación se procedió a normalizar la escena de entrada. Esto se puede lograr utilizando la escena de entrada $S(\vec{x})$ y el objeto de referencia $T(\vec{x})$ suponiendo que la ventana W es lo suficientemente pequeña tal que la escena dentro de la ventana puede considerarse uniformemente iluminada.

Sea $\widehat{S}(\vec{x})$ la escena normalizada de la siguiente manera:

$$\widehat{S}(\vec{x}) = a(\vec{x})S(\vec{x}) + b(\vec{x}), \tag{118}$$

donde $a(\vec{x})$ y $b(\vec{x})$ son coeficientes locales de normalización, los cuales toman en concideración la iluminación desconocida y desviaciones de la escena en la ventana W alrededor del $\vec{x}$-ésimo pixel, respectivamente. Para obtener los coeficientes óptimos es necesario minimizar el error medio absoluto (MAE) entre la escena normalizada $\widehat{S}(\vec{x})$ y la imagen de referencia $T(\vec{x})$, pero no es posible realizar esta minimización en forma analítica por lo que los coeficientes son estimados minimizando el error cuadrático medio (MSE) entre las escenas de la siguiente manera (Martínez-Díaz y Kober, 2011). Sea $E(\vec{x})$ el MSE entre la escena de prueba normalizada y la imagen de referencia,

$$E(\vec{x}) = \sum_{\vec{n}\in W} |[a(\vec{x})S(\vec{n}+\vec{x}) + b(\vec{x})] - T(\vec{n})|^2 \tag{119}$$

obteniendo las derivadas parciales de Eq.(119) con respecto a y b, e igualándolas a cero se obtiene:

$$a(\vec{x}) = \frac{\sum_{\vec{n}\in W} T(\vec{n}) \cdot S(\vec{n}+\vec{x}) - |W| \cdot \overline{T} \cdot \overline{S}(\vec{x})}{\sum_{\vec{n}\in W}[S(\vec{n}+\vec{x})]^2 - |W| \cdot [\overline{S}(\vec{x})]^2} \tag{120}$$

$$b(\vec{x}) = \overline{T} - a(\vec{x})\overline{S}(\vec{x}), \tag{121}$$

donde $\overline{T}$ y $\overline{S}(\vec{x})$ son los promedios del objeto de referencia y de la escena de entrada sobre la ventana W en la $\vec{x}$-ésima posición, respectivamente.

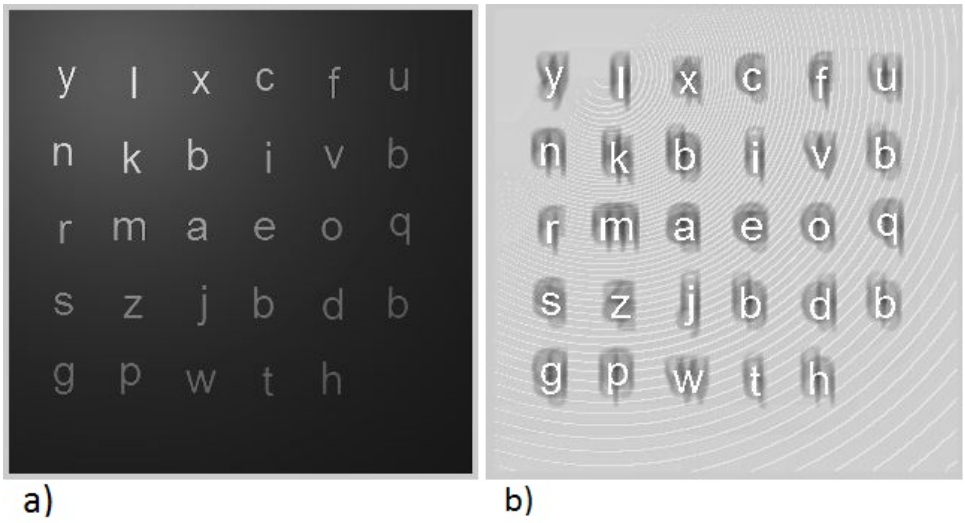

Figura 8: a) Imagen sintética degradada con iluminación no homogénea b) imagen resultante al normalizar escena de entrada.

Una vez finalizada la etapa de pre-procesamiento, se continúa con la etapa de reconocimiento y clasificación.

5.2. Reconocimiento y clasificación

En esta etapa, el objetivo es reconocer los caracteres y estimar la localización de cada uno de ellos en la imagen del documento. Para lograr esto, se utilizaron filtros de funciones discriminantes sintéticas no lineales y correlación morfológica.

5.2.1. Descomposición por umbral y Filtros NSDF

De acuerdo al enfoque de descomposción por umbral, una imagen con Q niveles de gris $S(\vec{x})$ puede ser representada como la suma de capas binarias de la siguiente manera Eq.(14):

$$S(\vec{x}) = \sum_{q=1}^{Q-1} S^q(\vec{x}). \tag{122}$$

Por otro lado, sea $\{T_i(\vec{x}), i = 1, \ldots, N\}$ un conjunto de N imágenes de entrenamiento en escala de grises pertenecientes a la clase verdadera. El filtro no lineal puede expresarse de la siguiente manera Eq.(112):

$$H_{NSDF}(\vec{x}) = \sum_{q=1}^{Q-1} \bigcap_{i=1}^{N} T_i^q(\vec{k}), \tag{123}$$

Para generar las imágenes de entrenamiento utilizadas para la composición de cada filtro se utilizaron las transformadas afines (Sec.2.8.7) considerando las posibles distorsiones geométricas presentes en la imagen del documento. A continuación se hace una breve descripción.

5.2.1.1. Imágenes de entrenamiento

Para generar las imágenes de entrenamiento, se utilizó una imagen de referencia de tamaño 25×25 por cada caracter en letra Arial 14 y se definieron cinco transformaciones afines, las cuales se describen a continuación.

- **Rotación**

  Utilizando la imagen de referencia del caracter y la Eq.(39) se realizaron rotaciones de -15 a 15 grados con un paso de tres grados.

- **Estiramiento en y**

  Utilizando la imagen de referencia del caracter y la Eq.(38) se realizó un escalamiento en dirección y por un factor $S_y = 0{,}8$ manteniendo fijo el valor en x (S_x), y posteriormente rotando la imagen obtenida de -15 a 15 grados con un paso de tres grados.

- **Estiramiento en x**

  Utilizando la imagen de referencia del caracter y la Eq.(38) se realizó un escalamiento en dirección x por un factor $S_x = 0{,}8$ manteniendo fijo el valor en y (S_y), y posteriormente rotando la imagen obtenida de -15 a 15 grados con un paso de tres grados.

- **Shearing en x**

  Utilizando la imagen de referencia del caracter y la Eq.(40) se realizó la transformación shearing en dirección x variando Sh_h de -0.5 a 0.5 con paso de 0.1, e igualando a cero Sh_v.

- **Escalamiento**

  Finalmente cada una de las imágenes generadas anteriormente fueron escaladas utilizando la Eq.(38), por un factor de 0.8 a 1.1 con un paso de 0.1.

En la Fig.(9) se muestra un ejemplo de las imágenes de entrenamiento generadas para el caracter “a”. Este mismo procedimiento se realizó para el resto de los caracteres.

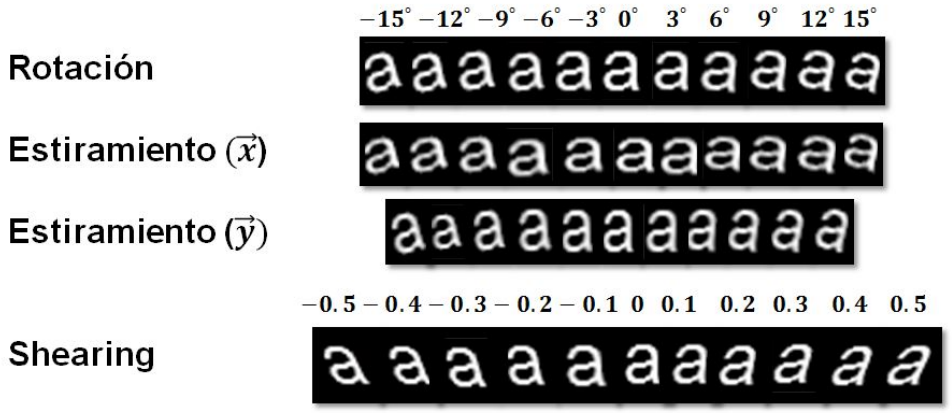

Figura 9: Imágenes de entrenamiento generadas por transformaciones afines para caracter "a".

En total se obtuvieron 176 imágenes de entrenamiento por caracter.

Una vez que se tienen las imágenes de entrenamiento, se procede a generar los filtros NSDF.

5.3. Sistema multi-nivel

El objetivo de diseñar un sistema multi-nivel es agrupar aquellos caracteres que se encuentren fuertemente correlacionados en subconjuntos de tal manera que podamos ir descartando otros que no lo estén.

5.3.1. Diseño de filtros NSDF adaptativos

Para poder separar los caracteres en subconjuntos se utilizó el algoritmo iterativo descrito en la Sec.(3.6) pero con una pequeña modificación en el penúltimo paso, ya que lo que se busca es crear un filtro que reconozca aquellos caracteres cuya correlación sea mayor, a continuación se describe el algoritmo:

1. Diseñar un filtro de funciones discriminantes sintéticas no lineal adaptativo (A-NSDF) entrenándolo sólo con la imagen del caracter.

2. Llevar a cabo la correlación entre la imagen del fondo y el filtro A-NSDF diseñado.

3. Calcular el valor DC.

4. Si el valor DC es mayor o igual al valor deseado, entonces el proceso del diseño del filtro es terminado; de otro modo continuar con el siguiente paso.

5. Crear un nuevo objeto a ser aceptado del fondo. El origen del objeto debe de estar en la posición del pico más alto del lóbulo lateral en el plano de correlación. El objeto es incluído en la clase de objetos verdaderos.

6. Diseñar un nuevo filtro A-NSDF utilizando la clase verdadera, volver al paso 2.

Al finalizar el algoritmo, se obtiene un conjunto de imágenes pertenecientes a la clase verdadera del filtro. Se espera que el filtro compuesto que se obtiene logre un DC cercano al deseado. Posteriormente utilizamos el enfoque de plantilla dual[3] para obtener un mejor resultado y evitar falsos positivos. A continuación se describe este último enfoque.

5.3.2. Diseño de plantilla dual

Sea $\{T_i(\vec{n}), i = 1, \ldots, N\}$ el conjunto de N imágenes de entrenamiento en escala de grises obtenidas como resultado del algoritmo adaptativo descrito anteriormente, con $H_{A-NSDF}(\vec{n})$ el filtro resultante. Utilizando el enfoque de descomposición por umbral (Eq.14) definimos una plantilla dual $H^{dual}_{A-NSDF}(\vec{n})$ de la siguiente manera:

$$H^{dual}_{A-NSDF}(\vec{n}) = \sum_{q=1}^{Q-1} \bigcap_{i=1}^{N} T^q_{dual_i}(\vec{n}), \tag{124}$$

$$con\ T^q_{dual_i}(\vec{n}) = 1 - T^q_i(\vec{n}), \vec{n} \in W, q = 1, \ldots, Q-1. \tag{125}$$

Ahora, se define una plantilla de dos niveles: $H_{2N} = 255$ para $\vec{n} \in RS$ y 0 otro lado, donde RS es la región de soporte de H_{A-NSDF}; de igual forma para H^{dual}_{A-NSDF} obteniendo H^{dual}_{2N}.

Posteriormente, la escena es correlacionada utilizando la Eq.(20) con H_{2N} y H^{dual}_{2N}. Nótese que bajo ciertas condiciones, descritas en la Sec.(4.3.2), la maximización de la diferencia entre las dos correlaciones morfológicas es equivalente a la minimización del error $E(\vec{x})$ (Eq.93), es decir,

$$-E_H(\vec{n}) = M_{SH^{dual}}(\vec{n}) - M_{SH}(\vec{n}). \tag{126}$$

[3]Dual Template

En la Fig.(10) se presenta el plano de correlación resultante después de utilizar el algoritmo iterativo descrito en (5.3.1) generando dos clases: $C_1 = \{b, p, d, g, q, o, e, c, s, a, m, n, h, u\}$ y $C_2 = \{r, t, f, z, k, x, y, w, v, l, j, i\}$, obteniendo un mínimo DC de 0.30. El segundo plano de correlación muestra el resultado al utilizar el enfoque de plantilla dual mejorando la capacidad de discriminación a más del doble del logrado anteriormente.

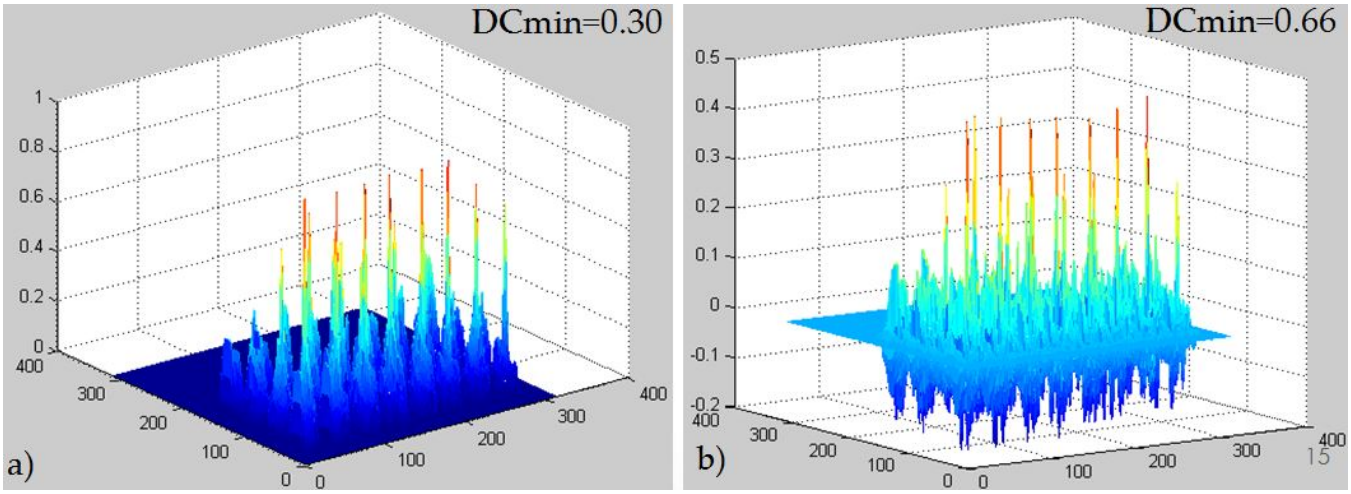

Figura 10: a) Plano de correlación final utilizando algoritmo adaptativo b) plano de correlación utilizando algoritmo adaptativo y el enfoque de plantilla dual.

5.3.2.1. Diseño de sistema multi-nivel

Utilizando lo anterior se dividió el sistema en 8 niveles, cada nivel contiene un número diferente de clases, tal como muestra la Fig.(11). Por ejemplo, el nivel 1 se encuentra conformado por 2 clases; la clase verdadera, $C_1 = \{b, p, d, q, g, o, e, c, s, a, m, h, n, u\}$ y la clase falsa, $C_2 = \{r, t, f, z, k, x, y, w, v, l, j, i\}$. El filtro A-NSDF que se obtuvo del algoritmo anterior (5.3.1) está formado por la intersección de algunas de las imágenes de entrenamiento pertenecientes a la clase C_1, en este caso: $T_{N1C1} = \{b, p, d, q, g, o, e, a\}$.

Ahora, cada nivel contiene distintas clases $\{C_i\}$ y cada clase esta compuesta por distintos filtros A-NSDF llamados banco de filtros. Por ejemplo, en el caso del subconjunto de imágenes T_{N1C1} se tienen 11 filtros por cada transformación geométrica, excepto escalamiento; es decir,

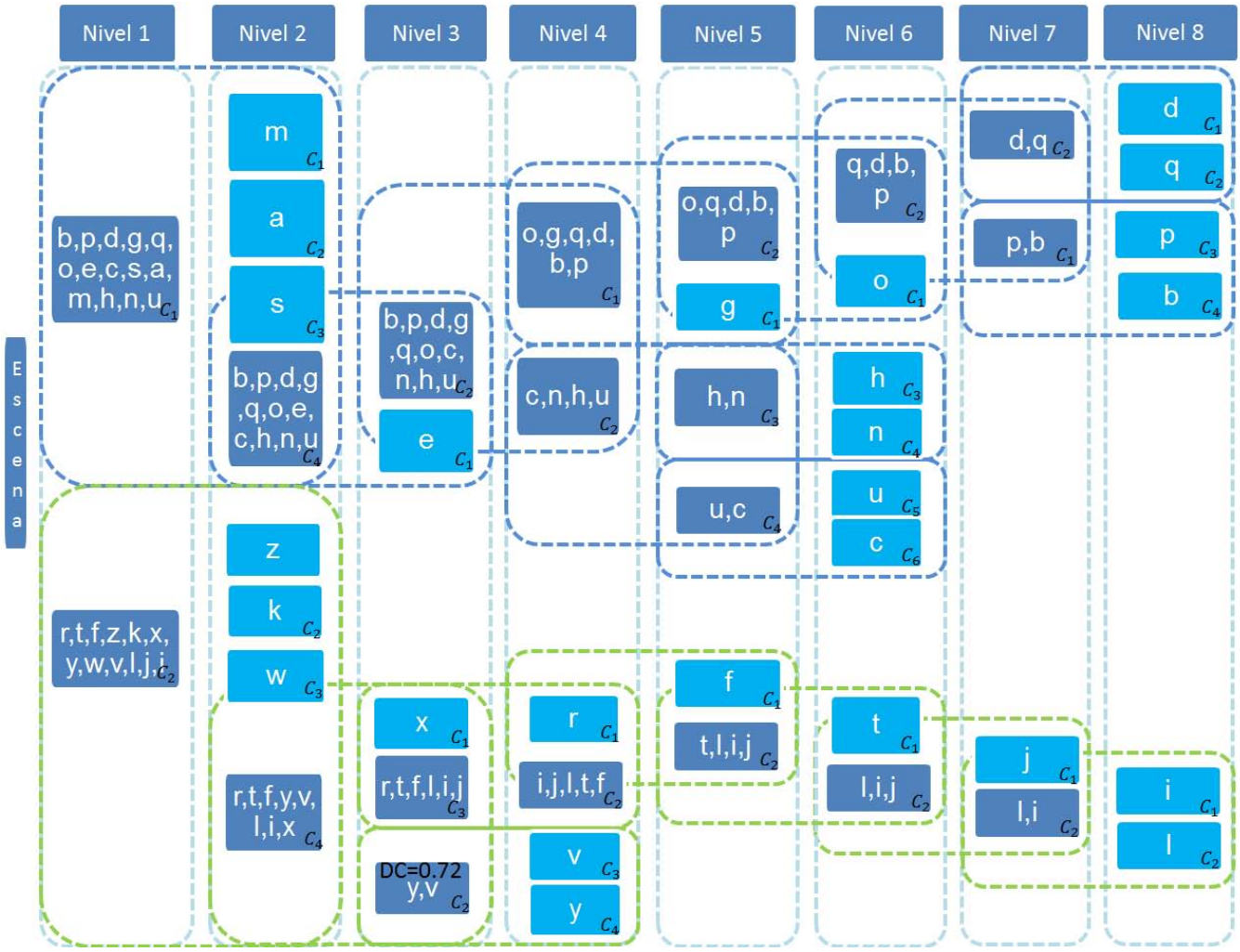

Figura 11: Sistema multi-nivel.

44 filtros. Y posteriormente las imágenes de cada uno de estos filtros son escaladas por los factores 0.8, 0.9 y 1.1 obteniendo así un total de 176 filtros por clase. A este conjunto de filtros se le define como banco de filtros, en la Fig.(12) se muestra un ejemplo. Cada banco de filtros funciona de la siguiente manera: la escena de entrada $S(\vec{x})$ es correlacionada utilizando la Eq.(20) con cada uno de los filtros H_n, posteriormente se escoge el plano de correlación que sea mayor a cierto umbral previamente definido por medio de los experimentos realizados y finalmente se estima su localización mediante las coordenadas del pico máximo en el plano de correlación seleccionado.

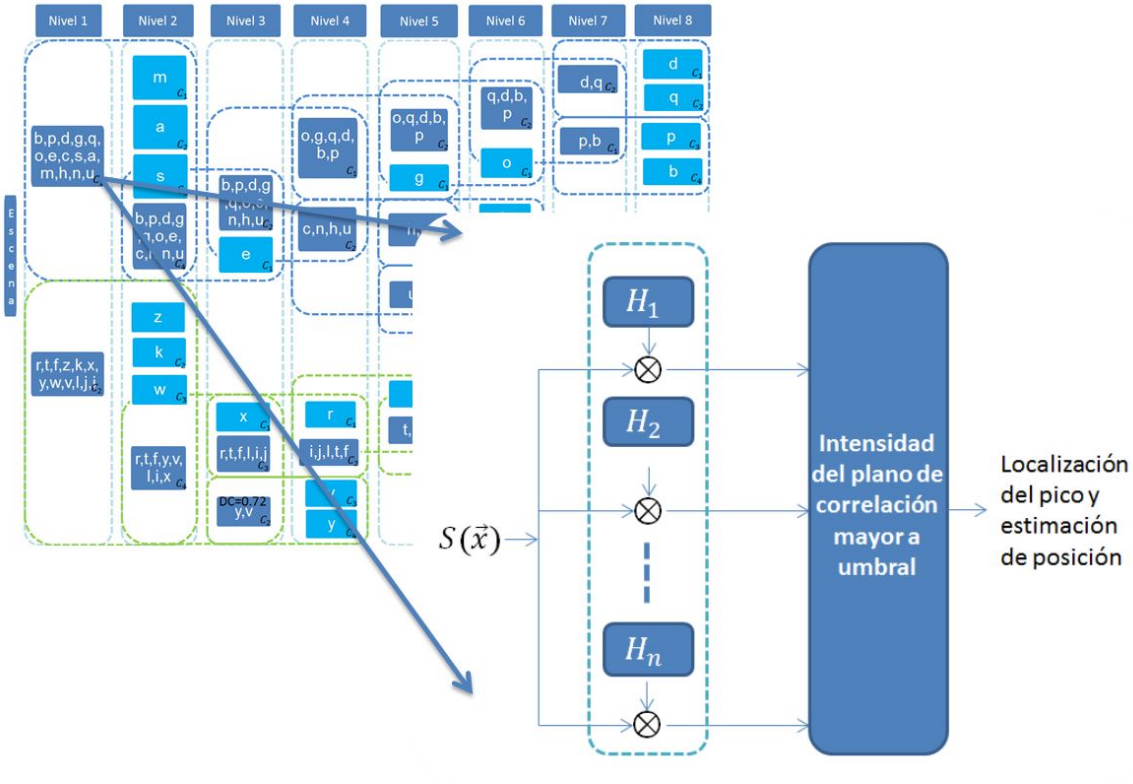

Figura 12: Banco de filtros.

Finalmente algunos caracteres son muy similares, como consecuencia, el valor DC del filtro compuesto A-NSDF es cercano o igual a cero. Para estos casos utilizamos un enfoque diferente basado en el filtro sólo fase (POE) y bloqueo de frecuencias descrita en la Sec.(4.4.2).

5.4. Resultados

En esta sección se presentan los resultados de las simulaciones realizadas. El desempeño de los filtros propuestos es evaluado en términos de capacidad de discriminación (DC) y

errores de clasificación.

El tamaño de todas las imágenes sintéticas utilizadas en los experimentos es de 310 × 310 pixeles. Cada imagen contiene diferentes distorsiones geométricas generadas utilizando transformaciones afines (Fig.2.8.7) descritas en la Tabla 1.

Cuadro 1: Distorsiones geométricas presentes en las imágenes sintéticas de prueba.

Distorsión geométrica	**Factor**	**Rango**	**Tamaño de paso**
Rotación	–	$[-15°, 15°]$	$1°$
Estiramiento($\vec{x}$) y rotación	$S_x = 0{,}8$	$[-15°, 15°]$	$1°$
Estiramiento($\vec{y}$) y rotación	$S_y = 0{,}8$	$[-15°, 15°]$	$1°$
Shearing($\vec{x}$)	–	$[-0{,}5, 0{,}5]$	0,05
Escalamiento	–	$[0{,}8, 1{,}1]$	0,1

5.4.1. Descripción de los experimentos

Para cada experimento se utilizaron 88 imágenes sintéticas, se repitió cada experimento tres veces utilizando las mismas condiciones de ruido e iluminación y sólo se varió la posición de los caracteres en la escena y las distintas distorsiones geométricas.

En cuanto a errores de clasificación, se dirá que se tiene un falso positivo si el sistema dice reconocer un objeto como verdadero pero no lo es; y se dirá que se tiene un falso negativo si el sistema dice no reconocer un objeto verdadero pero si se encuentra presente.

5.4.1.1. Conjunto de experimentos 1

El primer conjunto de experimentos se realizó utilizando 88 escenas degradadas con ruido aditivo con Desv. Est. igual a 0, 5, 10 y 15 Fig.(13), evaluando todos los filtros del sistema (Fig.11). Los resultados se muestran de la Fig.(14) hasta la Fig.(26).

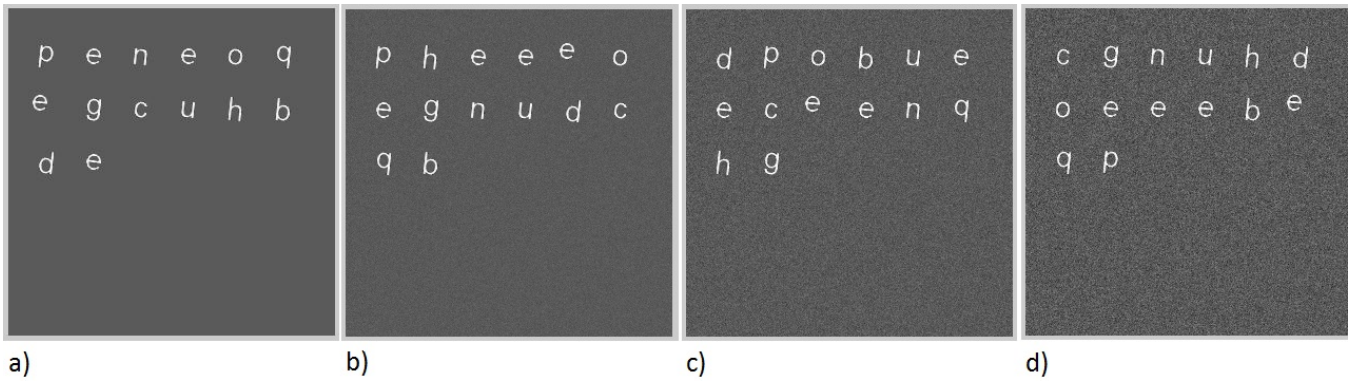

Figura 13: Ejemplo de las imágenes sintéticas utilizadas para los experimentos realizados degradadas con ruido aditivo: a) $\sigma = 0$, b) $\sigma = 5$, c) $\sigma = 10$ y d) $\sigma = 15$.

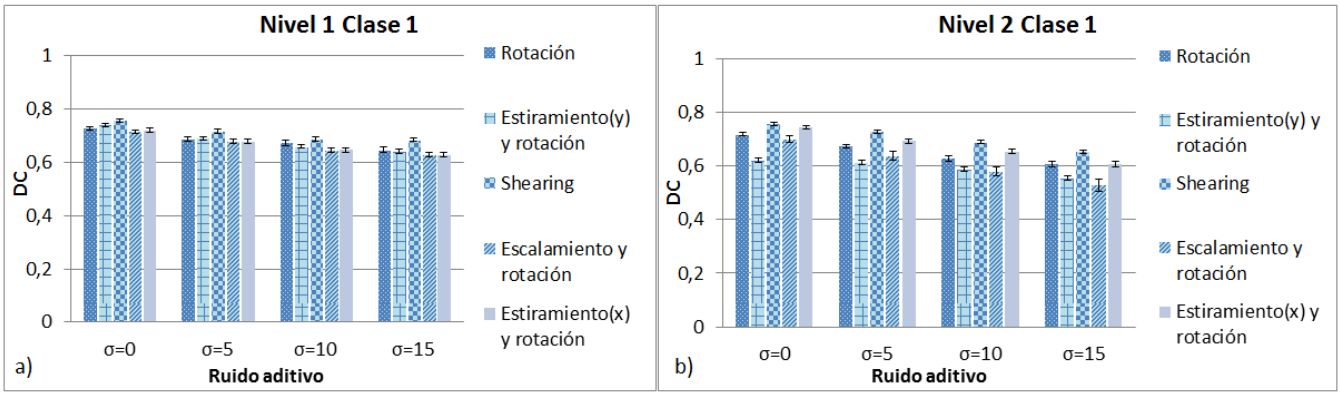

Figura 14: Desempeño de banco de filtros en 88 imágenes degradadas con ruido aditivo, DC con un nivel de confianza del 95 % a) Nivel 1: $C_1 = \{b, p, d, g, q, o, e, c, s, a, m, n, h, u\}$ b) Nivel 2: $C_1 = \{m\}$.

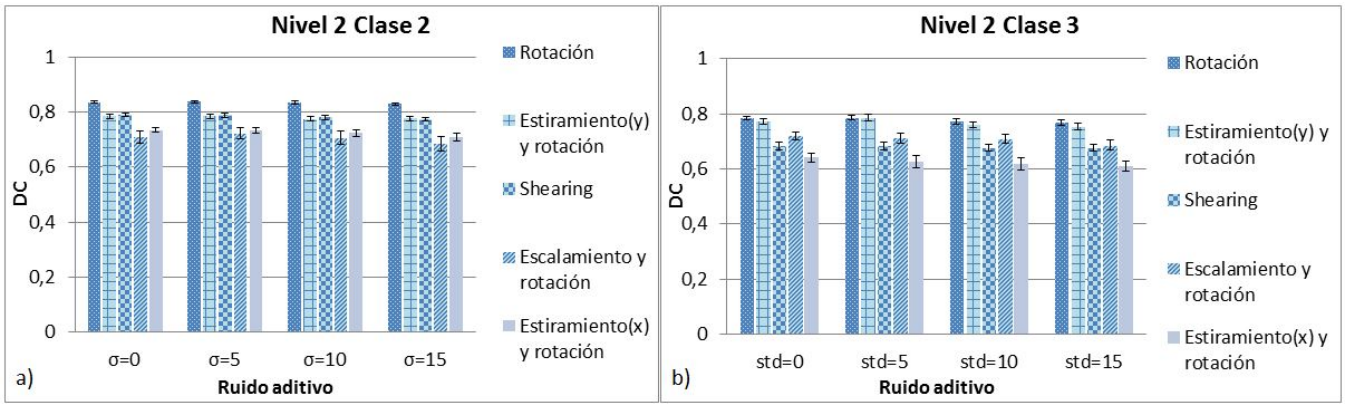

Figura 15: Desempeño de banco de filtros en 88 imágenes degradadas con ruido aditivo. DC con un nivel de confianza del 95 % a) Nivel 2: $C_2 = \{a\}$ b) Nivel 2: $C_3 = \{s\}$.

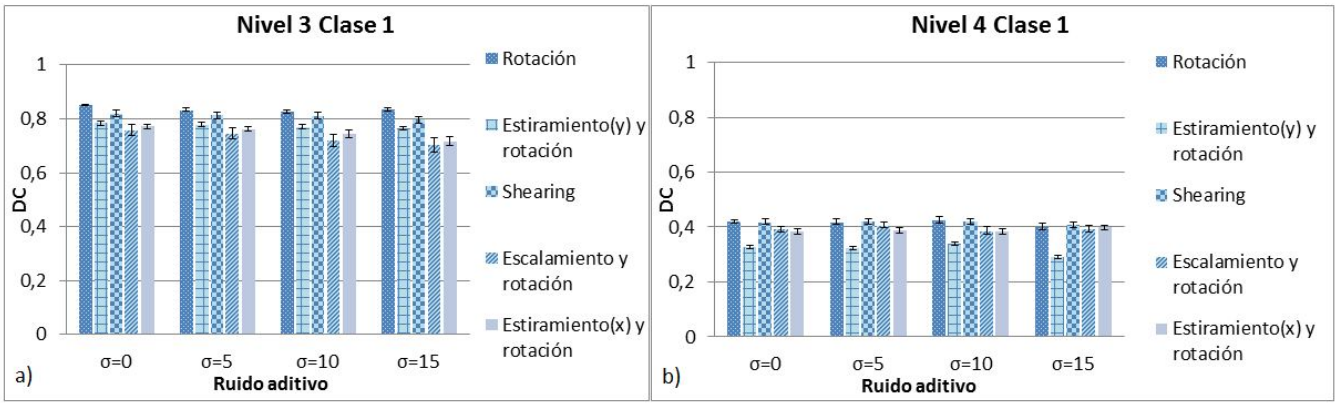

Figura 16: Desempeño de banco de filtros en 88 imágenes degradadas con ruido aditivo. DC con un nivel de confianza del 95 % a) Nivel 3: $C_1 = \{e\}$ b) Nivel 4: $C_1 = \{o, q, g, d, b, p\}$.

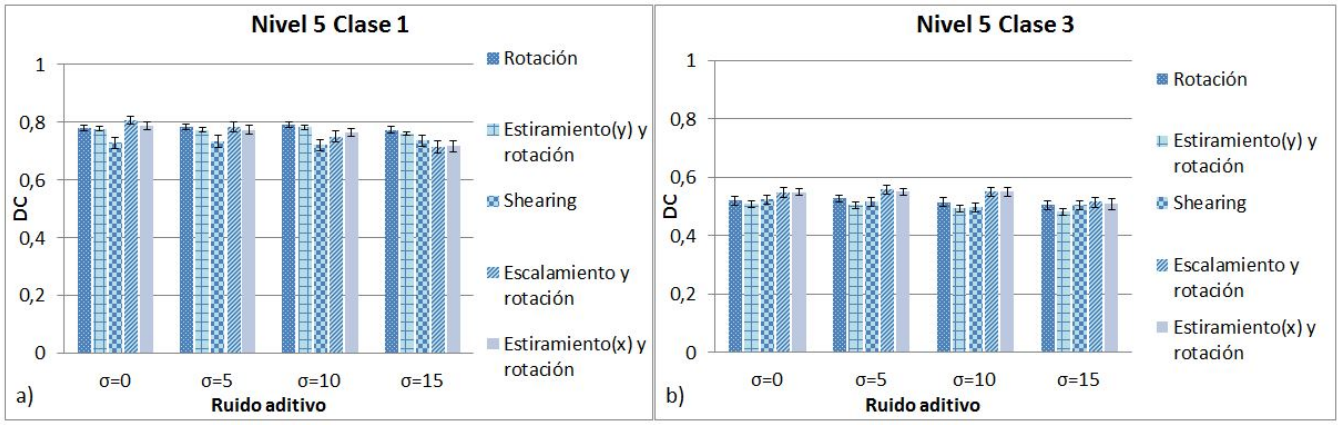

Figura 17: Desempeño de banco de filtros en 88 imágenes degradadas con ruido aditivo. DC con un nivel de confianza del 95 % a) Nivel 5: $C_1 = \{g\}$ b) Nivel 5: $C_3 = \{h, n\}$.

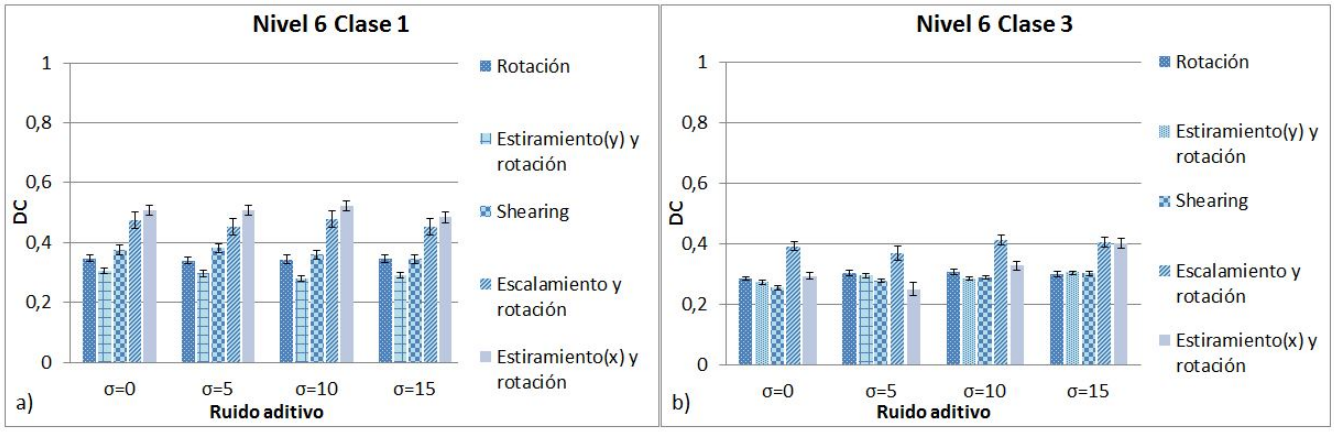

Figura 18: Desempeño de banco de filtros en 88 imágenes degradadas con ruido aditivo. DC con un nivel de confianza del 95 % a) Nivel 6: $C_1 = \{o\}$ b) Nivel 6: $C_3 = \{h\}$.

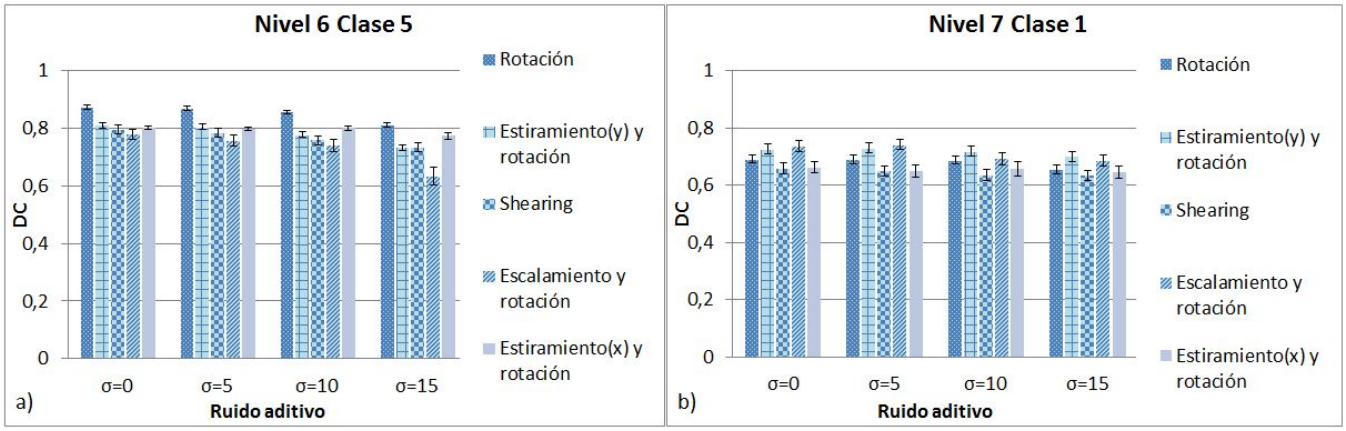

Figura 19: Desempeño de banco de filtros en 88 imágenes degradadas con ruido aditivo. DC con un nivel de confianza del 95 % a) Nivel 6: $C_5 = \{u\}$ b) Nivel 7: $C_1 = \{p, b\}$.

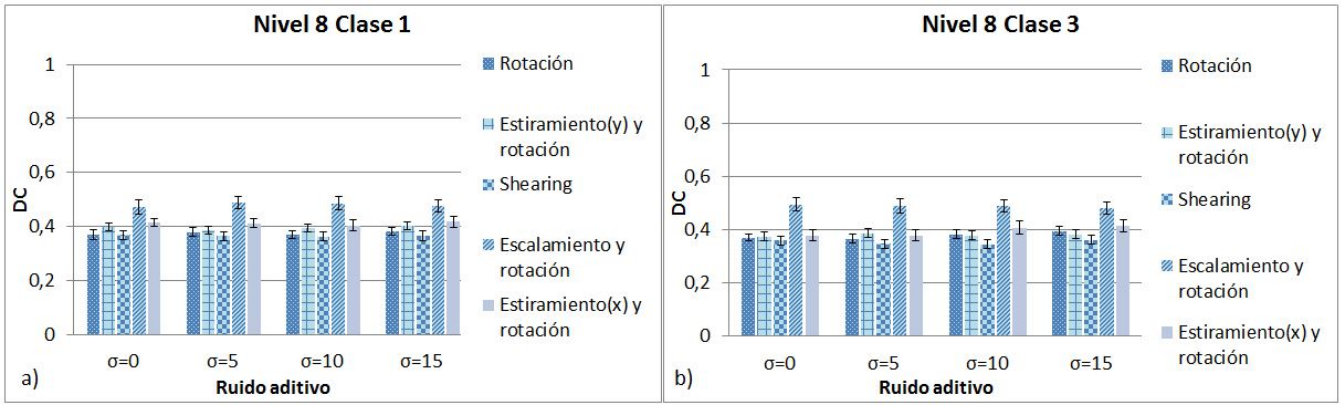

Figura 20: Desempeño de banco de filtros en 88 imágenes degradadas con ruido aditivo. DC con un nivel de confianza del 95 % a) Nivel 8: $C_1 = \{d\}$ b) Nivel 8: $C_3 = \{b\}$.

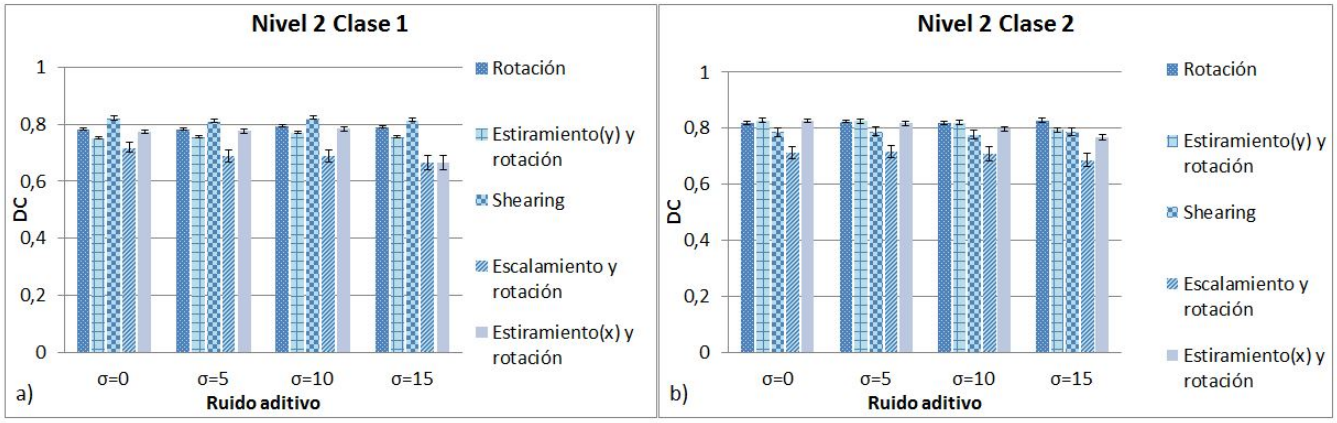

Figura 21: Desempeño de banco de filtros en 88 imágenes degradadas con ruido aditivo. DC con un nivel de confianza del 95 % a) Nivel 2: $C_1 = \{z\}$ b) Nivel 2: $C_2 = \{k\}$.

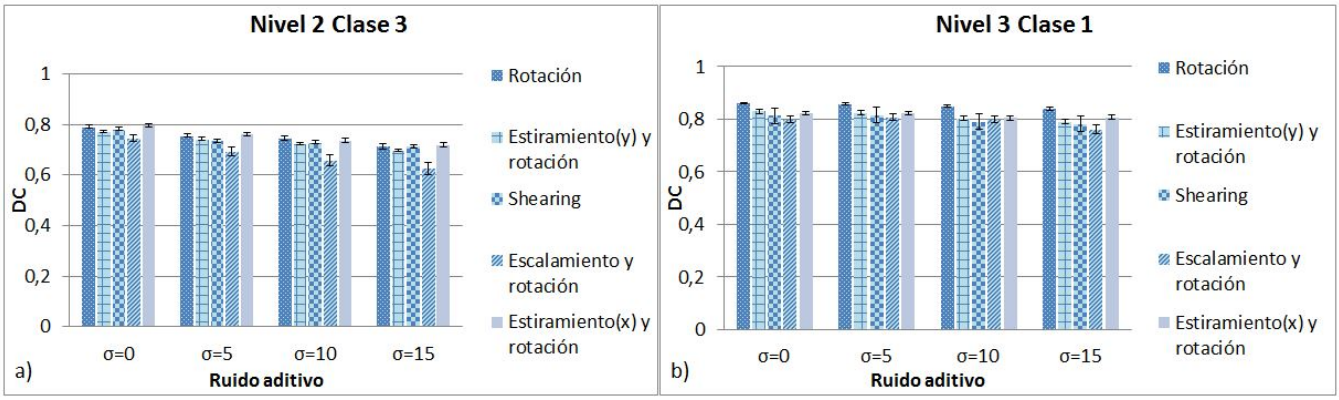

Figura 22: Desempeño de banco de filtros en 88 imágenes degradadas con ruido aditivo. DC con un nivel de confianza del 95 % a) Nivel 2: $C_3 = \{w\}$ b) Nivel 3: $C_1 = \{x\}$.

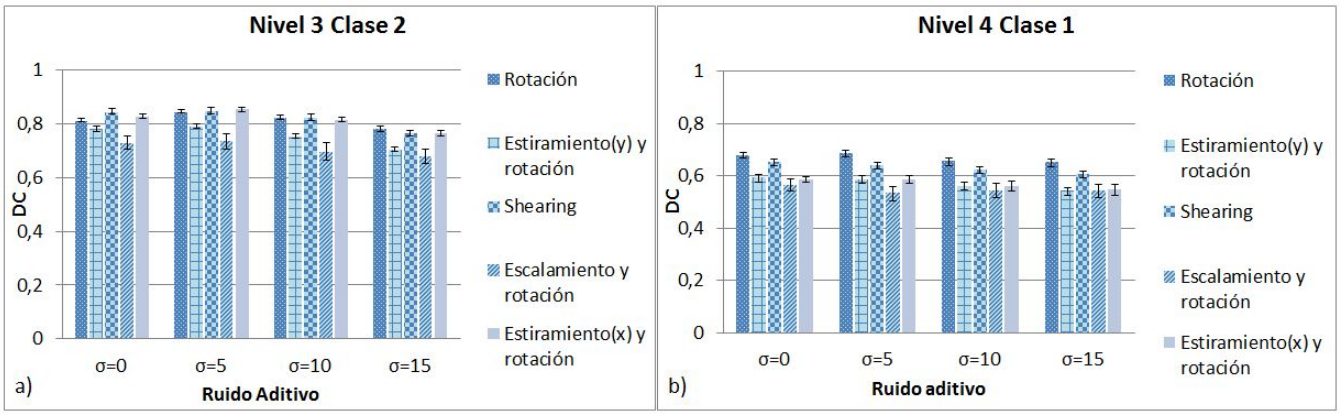

Figura 23: Desempeño de banco de filtros en 88 imágenes degradadas con ruido aditivo. DC con un nivel de confianza del 95 % a) Nivel 3: $C_2 = \{y, v\}$ b) Nivel 4: $C_1 = \{r\}$.

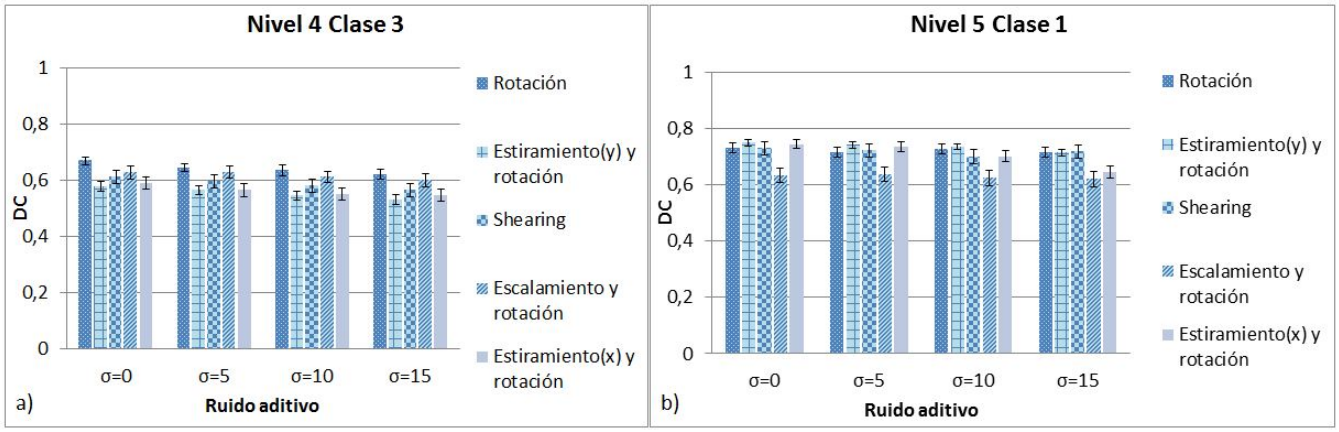

Figura 24: Desempeño de banco de filtros en 88 imágenes degradadas con ruido aditivo. DC con un nivel de confianza del 95 % a) Nivel 4: $C_3 = \{v\}$ b) Nivel 5: $C_1 = \{f\}$.

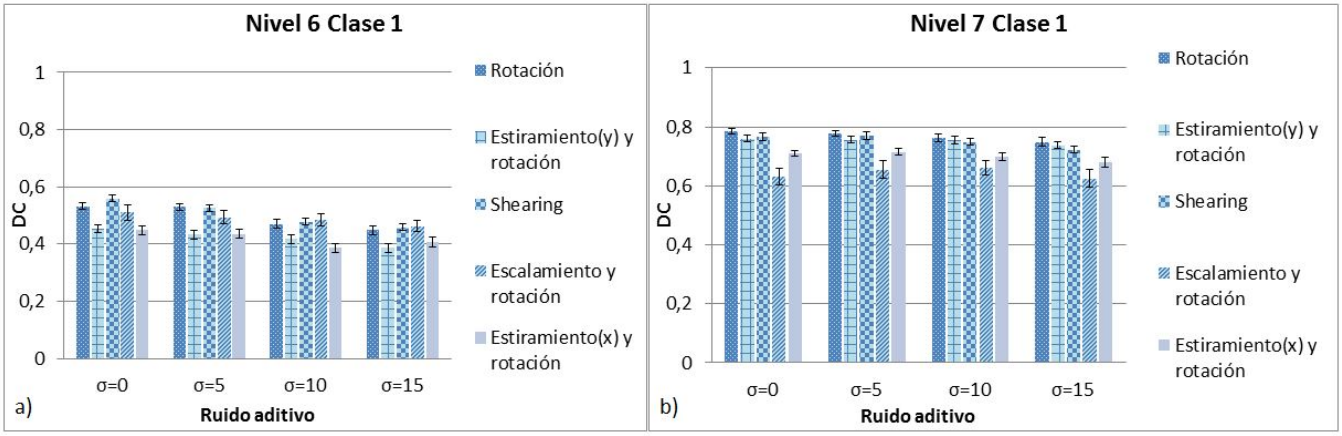

Figura 25: Desempeño de banco de filtros en 88 imágenes degradadas con ruido aditivo. DC con un nivel de confianza del 95 % a) Nivel 6: $C_1 = \{t\}$ b) Nivel 7: $C_1 = \{j\}$.

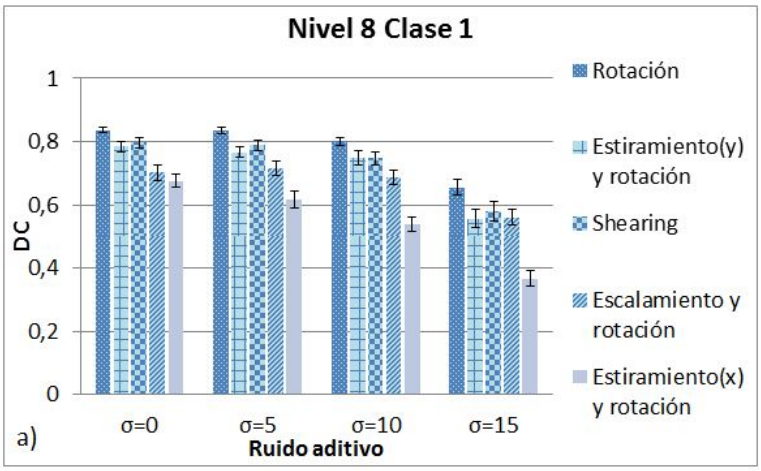

Figura 26: Desempeño de banco de filtros en 88 imágenes degradadas con ruido aditivo. DC con un nivel de confianza del 95 % a) Nivel 8: $C_1 = \{i\}$.

Como se puede apreciar en las gráficas, los resultados arrojan un buen desempeño en cuanto a capacidad de discriminación, obteniendo en todos los casos un DC mayor a 0.3 en promedio y en la mayoría de los casos un DC mayor a 0.6. En tanto a errores de clasificación, la mayoría se encuentran por debajo del 10 % de error de falsos negativos y 0 % en errores de falsos positivos. Las Fig.(27),(28) y (29) muestran los errores de clasificación más altos obtenidos.

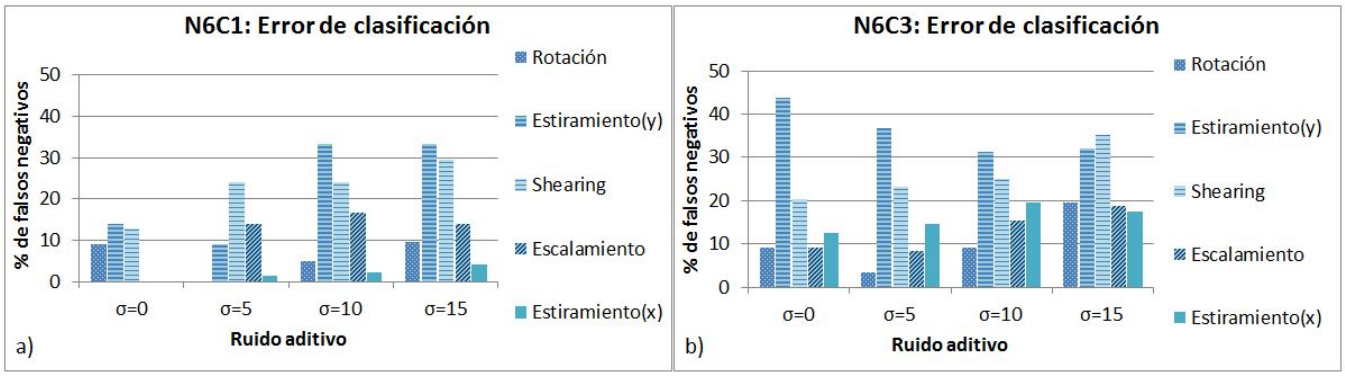

Figura 27: Errores de clasificación en 88 imágenes sintéticas degradadas con ruido aditivo. a) Nivel 6: $C_1 = \{o\}$ b) Nivel 6: $C_3 = \{h\}$.

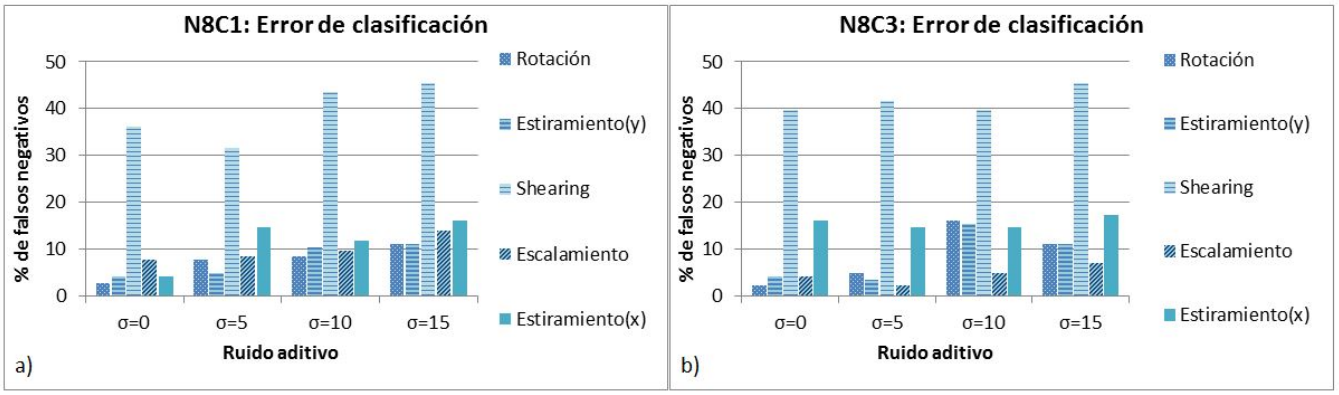

Figura 28: Errores de clasificación en 88 imágenes sintéticas degradadas con ruido aditivo. a) Nivel 8: $C_1 = \{d\}$ b) Nivel 8: $C_3 = \{b\}$.

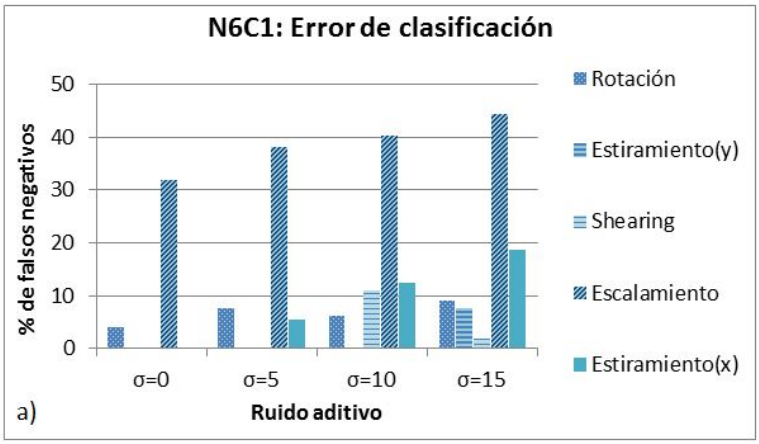

Figura 29: Errores de clasificación en 88 imágenes sintéticas degradadas con ruido aditivo. Nivel 6: $C_1 = \{t\}$.

5.4.1.2. Conjunto de experimentos 2

El segundo conjunto de experimentos se realizó utilizando escenas degradadas con iluminación no homogénea utilizando el modelo de iluminación Lambertiano (Eq.4) con $\phi = 65$ y $\varphi = 60$, y variando el parámetro $\rho = 30, 40, 50$ y 70 (Fig.30); evaluando algunos filtros del sistema. Los resultados del desempeño de los filtros en cuanto a DC y errores de clasificación se muestran de la Fig.(31) hasta la Fig.(36).

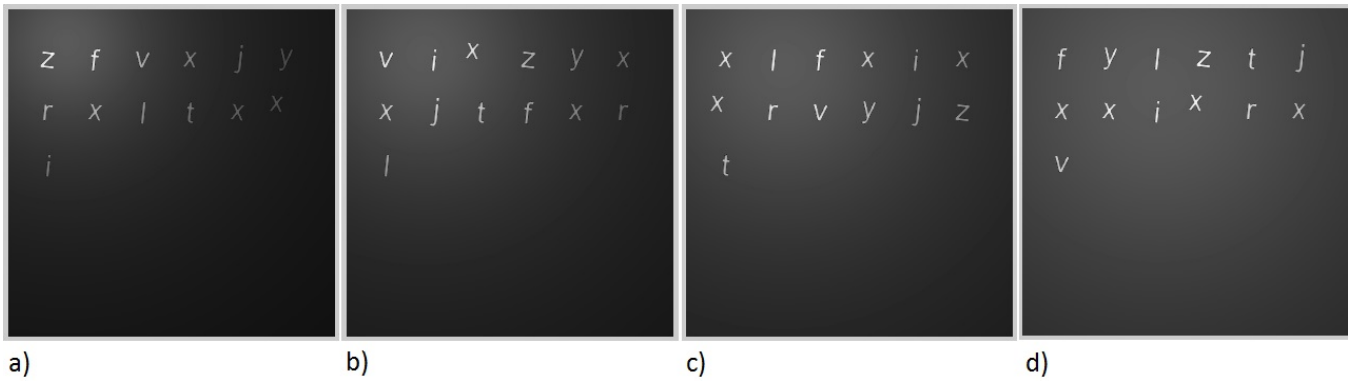

Figura 30: Ejemplo de las imágenes sintéticas utilizadas para los experimentos realizados degradadas con iluminación no homogénea: a) $\rho = 30$, b) $\rho = 40$, c) $\rho = 50$ y d) $\rho = 70$

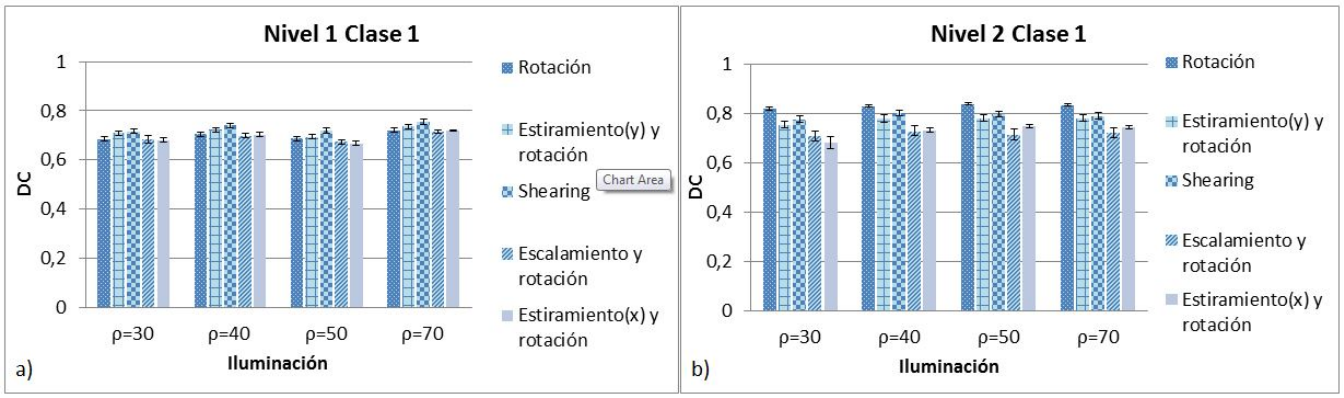

Figura 31: Resultado de evuación de 88 imágenes sintéticas degradadas con iluminación no homogénea. Capacidad de discriminación del banco de filtros con un 95 % de confianza a) Nivel 1: $C_1 = \{b, p, d, g, q, o, e, c, s, a, m, n, h, u\}$ b) Nivel 2: $C_1 = \{m\}$.

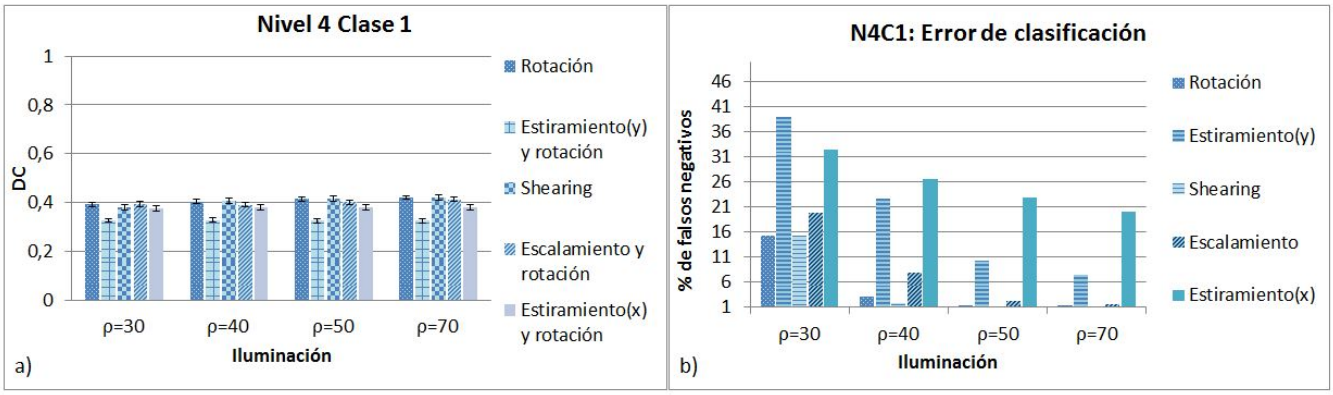

Figura 32: Resultado de evuación de 88 imágenes sintéticas degradadas con iluminación no homogénea. Nivel 4: $C_1 = \{o, g, q, d, b, p\}$ a) Capacidad de discriminación del banco de filtros con un 95 % de confianza b) errores de clasificación.

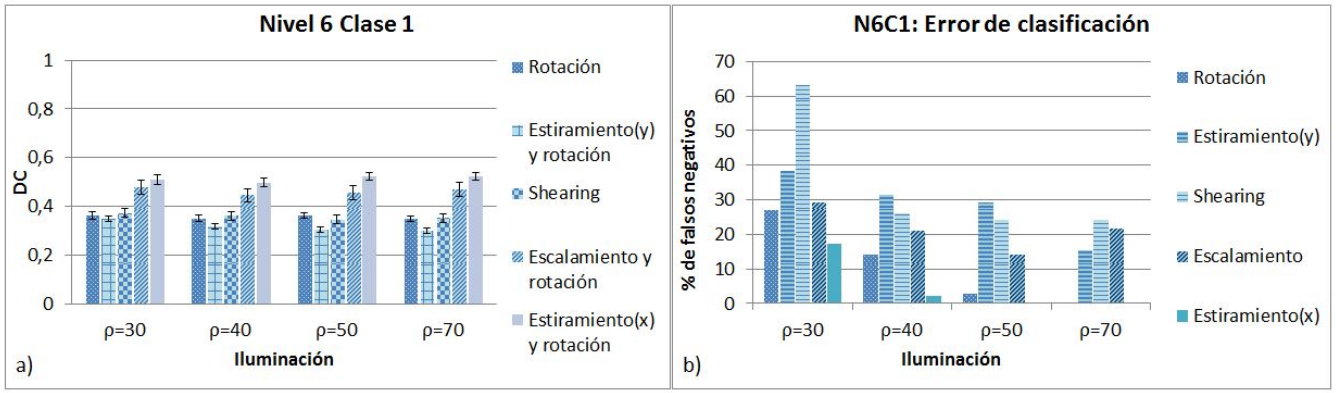

Figura 33: Resultado de evuación de 88 imágenes sintéticas degradadas con iluminación no homogénea. Nivel 6: $C_1 = \{o\}$ a) Capacidad de discriminación del banco de filtros con un 95 % de confianza b) errores de clasificación.

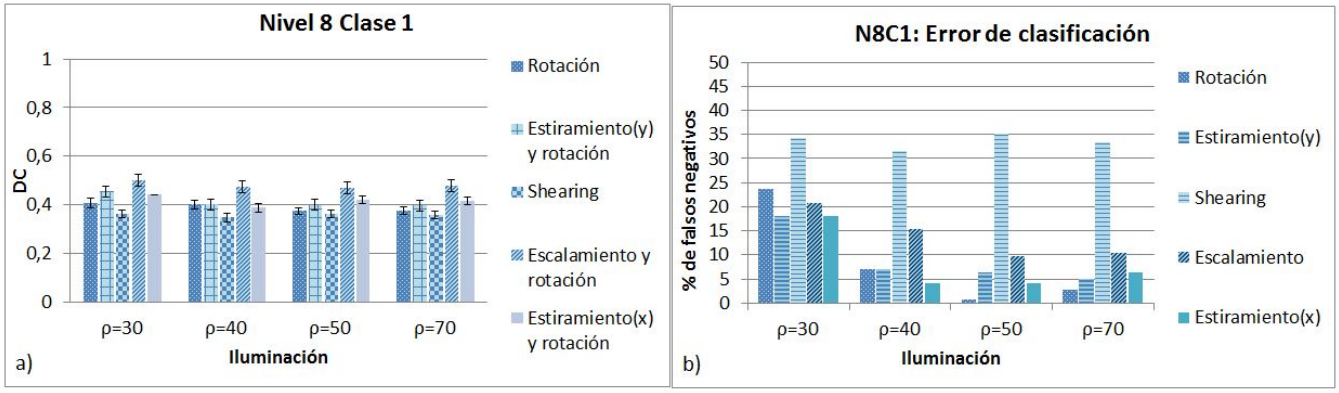

Figura 34: Resultado de evuación de 88 imágenes sintéticas degradadas con iluminación no homogénea. Nivel 8: $C_1 = \{d\}$ a) Capacidad de discriminación del banco de filtros con un 95 % de confianza b) errores de clasificación.

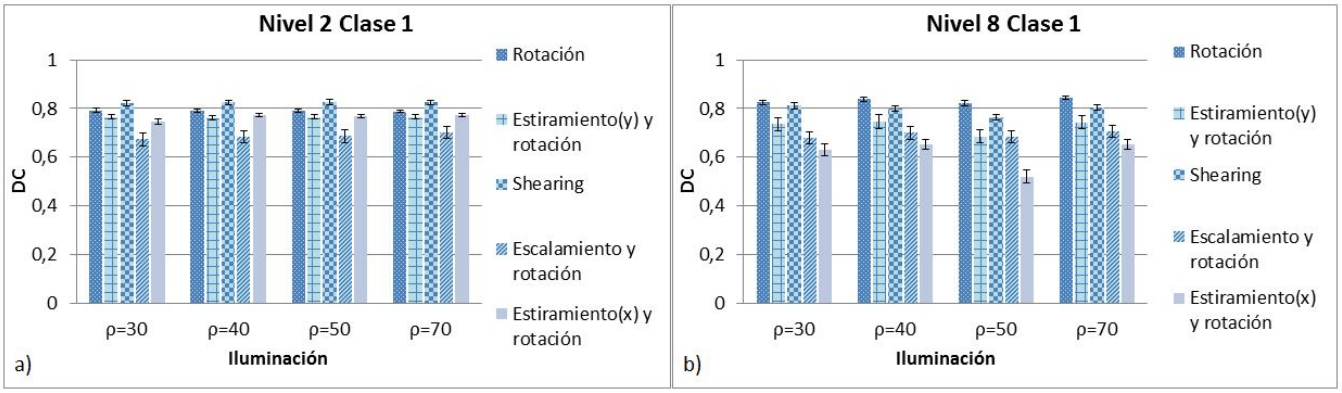

Figura 35: Resultado de evuación de 88 imágenes sintéticas degradadas con iluminación no homogénea. Capacidad de discriminación del banco de filtros con un 95 % de confianza a) Nivel 2: $C_1 = \{z\}$ b) Nivel 6: $C_1 = \{i\}$.

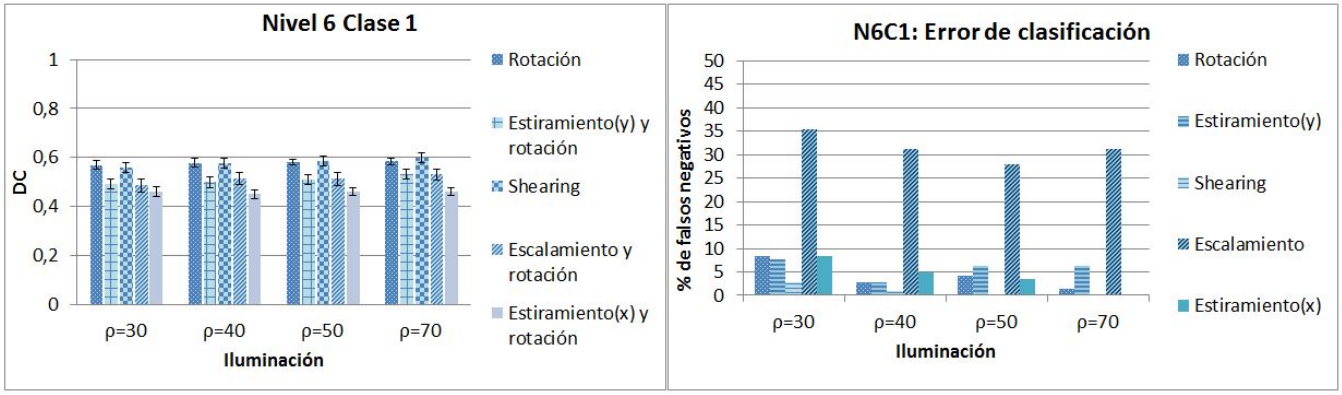

Figura 36: Resultado de evuación de 88 imágenes sintéticas degradadas con iluminación no homogénea. Nivel 6: $C_1 = \{t\}$ a) Capacidad de discriminación del banco de filtros con un 95 % de confianza b) errores de clasificación.

5.4.1.3. Conjunto de experimentos 3

El tercer conjunto de experimentos se realizó utilizando escenas degradadas con iluminación no homogénea utilizando el modelo de iluminación Lambertiano (Eq.4) con $\phi = 65$, $\varphi = 60$ y $\rho = 50$; se agregó una mezcla de ruido aditivo e impulsivo a la escena de entrada con Desv. Std. igual a 0, 5, 10 y 15;, y probabilidad de 0.05, respectivamente. Al final de cada gráfica se muestra el peor caso generado, utilizando mezcla de ruido aditivo con Desv. Std igual a 15, ruido impulsivo con probabilidad 0.05 e iluminación no homogénea con $\phi = 65$, $\varphi = 60$ y $\rho = 30$. Los resultados del desempeño de algunos de los filtros en cuanto DC y errores de clasificación se muestran de la Fig.(38) hasta la Fig.(42).

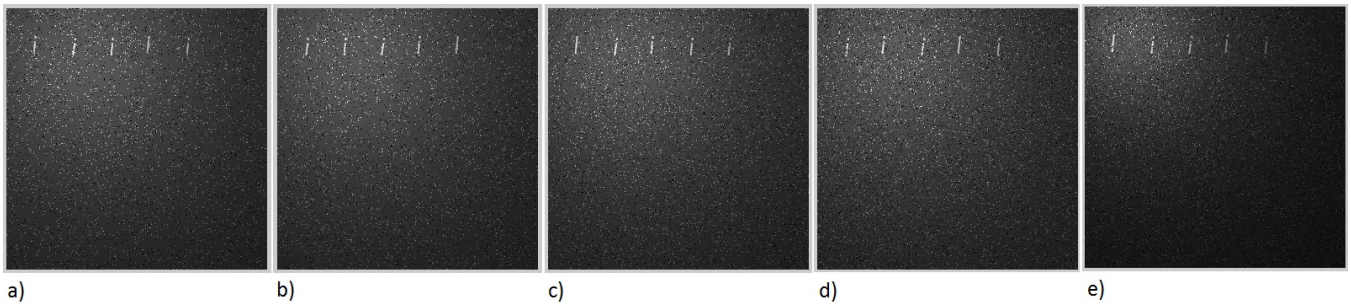

Figura 37: Ejemplo de las imágenes sintéticas utilizadas para los experimentos realizados degradadas con iluminación no homogéneay mezcla de ruido aditivo e impulsivo con: a) $\sigma = 0,\ \rho = 50\ y\ prob = 0{,}05$, b) $\sigma = 5,\ \rho = 50\ y\ prob = 0{,}05$, c) $\sigma = 10,\ \rho = 50\ y\ prob = 0{,}05$ d)$\sigma = 15,\ \rho = 50\ y\ prob = 0{,}05$ y e) $\sigma = 15,\ \rho = 30\ y\ prob = 0{,}05$

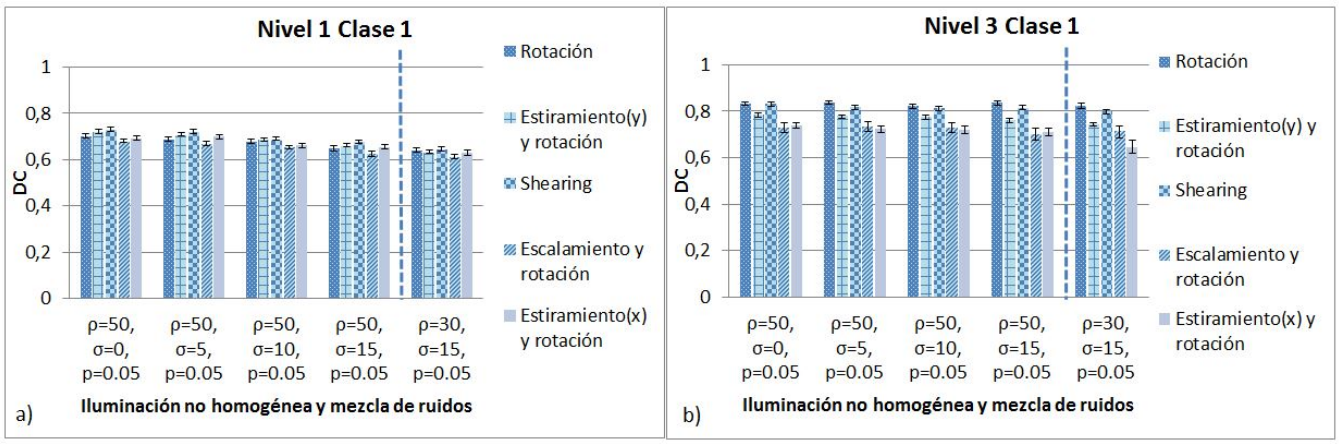

Figura 38: Desempeño de banco de filtros en 88 imágenes degradadas con iluminación no homogénea y mezcla de ruido aditivo e impulsivo. DC con un nivel de confianza del 95 % a) Nivel 1: $C_1 = \{b, p, d, g, q, o, e, c, s, a, m, n, h, u\}$ b) Nivel 2: $C_1 = \{e\}$.

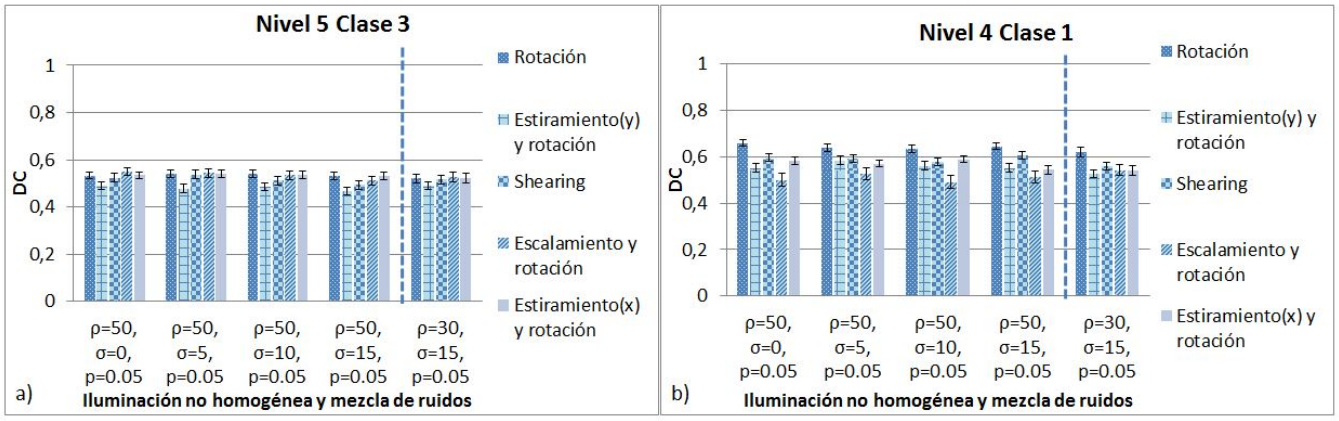

Figura 39: Desempeño de banco de filtros en 88 imágenes con iluminación no homogénea y mezcla de ruido aditivo e impulsivo. DC con un nivel de confianza del 95 % a) Nivel 5: $C_3 = \{h, n\}$ b) Nivel 3: $C_1 = \{r\}$.

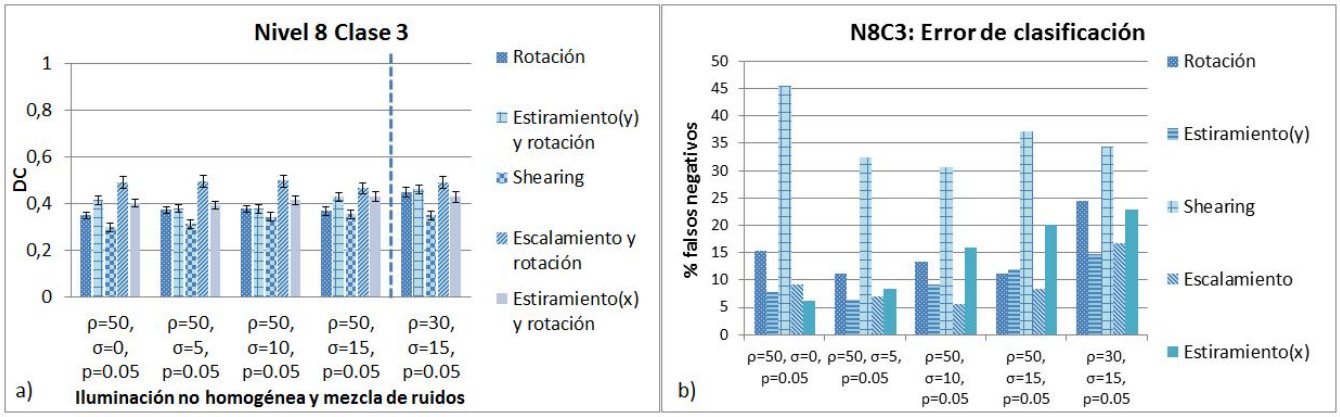

Figura 40: Resultado de evuación de 88 imágenes sintéticas degradadas con iluminación no homogénea y mezcla de ruido aditivo e impulsivo. Nivel 5, $C_1 = \{b\}$ a) Capacidad de discriminación del banco de filtros con un 95 % de confianza b) errores de clasificación.

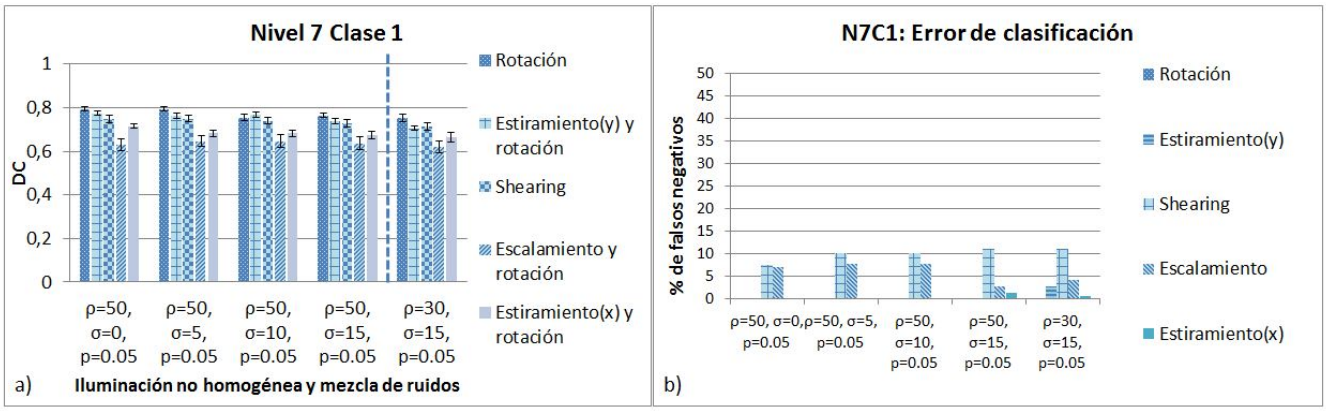

Figura 41: Resultado de evuación de 88 imágenes sintéticas degradadas con iluminación no homogénea y mezcla de ruido aditivo e impulsivo. Nivel 7, $C_3 = \{j\}$ a) Capacidad de discriminación del banco de filtros con un 95 % de confianza b) errores de clasificación.

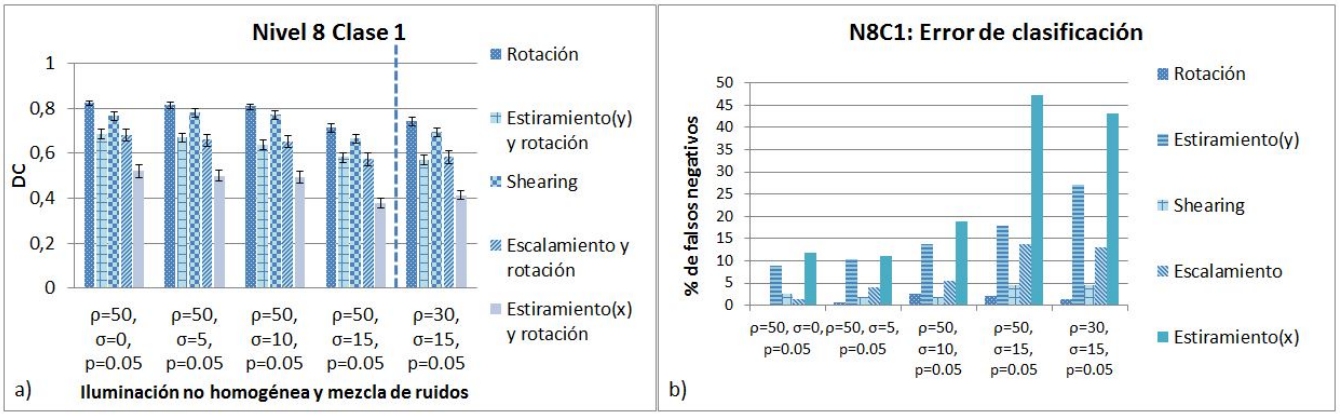

Figura 42: Resultado de evuación de 88 imágenes sintéticas degradadas con iluminación no homogénea y mezcla de ruido aditivo e impulsivo. Nivel 8, $C_1 = \{i\}$ a) Capacidad de discriminación del banco de filtros con un 95 % de confianza b) errores de clasificación.

En general, los filtros muestran un buen desempeño. En los casos en donde existe un número alto de errores de clasificación, como por ejemplo, en el caso del caracter i (Fig.42) los filtros compuestos por las imágenes generadas por las transformaciones de estiramiento pueden ser eliminados, ya que la transformación no afecta la forma del caracter.

Otro caso en el que se generaron errores altos de clasificación fue en el Nivel 6 Clase 1 (Fig.36) al momento de escalar las imágenes de entrenamiento. Esto debido a que sólo era escalado el caracter t y los demás caracteres permanecían del mismo tamaño, lo que producía que algún caracter de la clase falsa tomara la forma del caracter verdadero. Esto podría solucionarse creando un filtro adaptativo para cada filtro de escalamiento.

Finalmente, los caracteres que tuvieron peores resultados fueron los caracteres b y p, y d y q. Es necesario buscar otro criterio que haga posible una mejor clasificación de estos caracteres.

5.4.2. Comparación contra ABBYY y el algoritmo SIFT

A continación se presenta algunas pruebas realizadas a imágenes reales con el sistema propuesto, el algoritmo SIFT(Sec.1.8) y el sistema comercial ABBYY (Sec.1.9.1).

5.4.2.1. Descripción imágenes reales

El tamaño de la imagen real es de 310×310 formato .jpg en escala de grises. El tipo de fuente utilizado fue Times new roman, Arial, Comic sans y Verdana y el tamaño varía de 12 a 16. Los resultados se compararán a través del número de errores de clasificación.

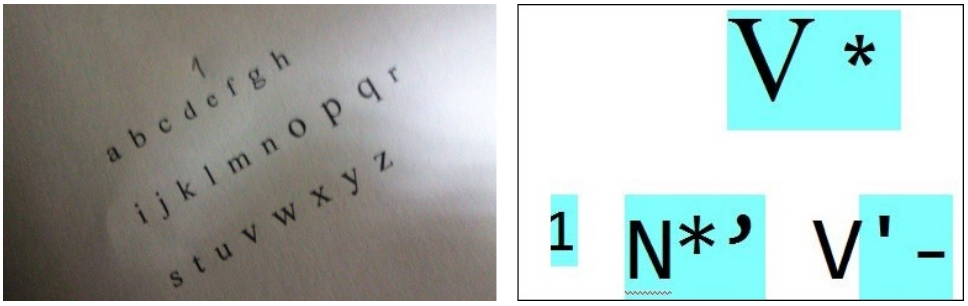

Figura 43: A la izquierda se muestra la imagen real (Times New Roman) y a la derecha el resultado con el sistema OCR ABBYY.

Cuadro 2: Errores de clasificación, Fig(43).

	Falsos negativos	Falsos positivos
ABBYY	26	9
SIFT	18	7
Sistema propuesto	12	8

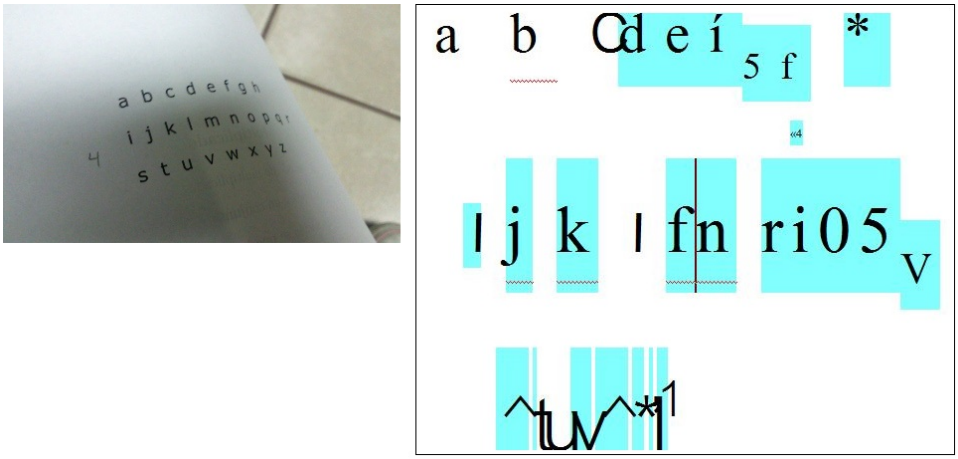

Figura 44: A la izquierda se muestra imagen real (Verdana), y a la derecha resultado con OCR ABBYY.

Cuadro 3: Errores de clasificación, Fig(44).

	Falsos negativos	Falsos positivos
ABBYY	9	15
SIFT	18	14
Sistema propuesto	13	7

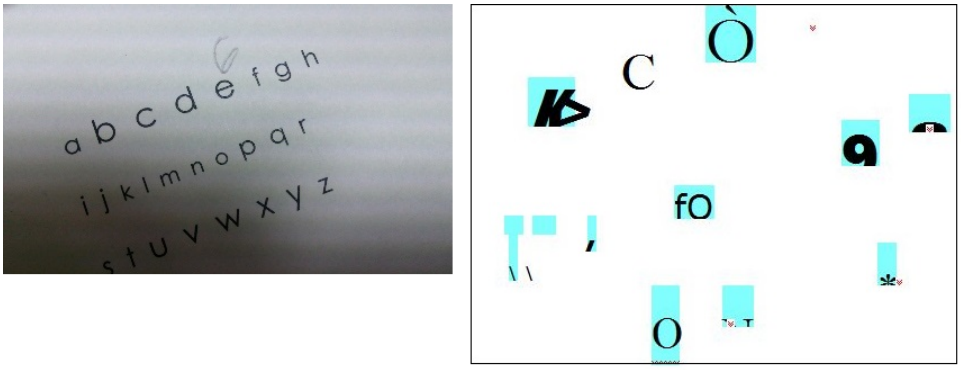

Figura 45: A la izquierda se muestra la imagen real (Times New Roman), y a la derecha el resultado con el sistema OCR ABBYY.

Cuadro 4: Errores de clasificación, Fig(45).

	Falsos negativos	Falsos positivos
ABBYY	26	12
SIFT	11	7
Sistema propuesto	11	10

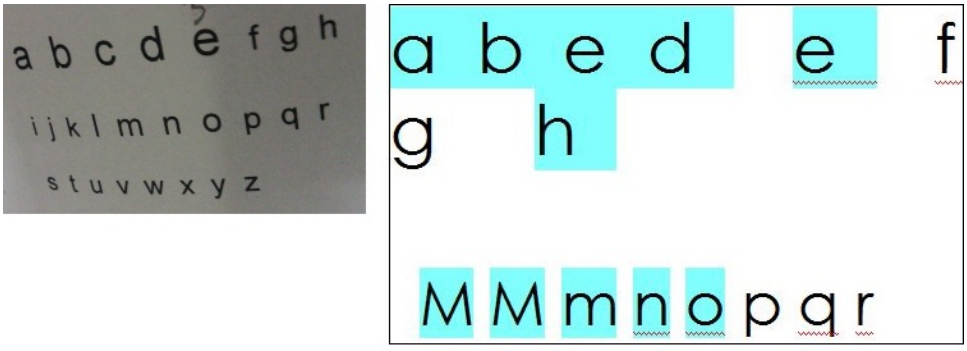

Figura 46: A la izquierda se muestra la imagen real (Arial), y a la derecha el resultado con el sistema OCR ABBYY.

Cuadro 5: Errores de clasificación, Fig(46).

	Falsos negativos	Falsos positivos
ABBYY	13	3
SIFT	11	16
Sistema propuesto	6	14

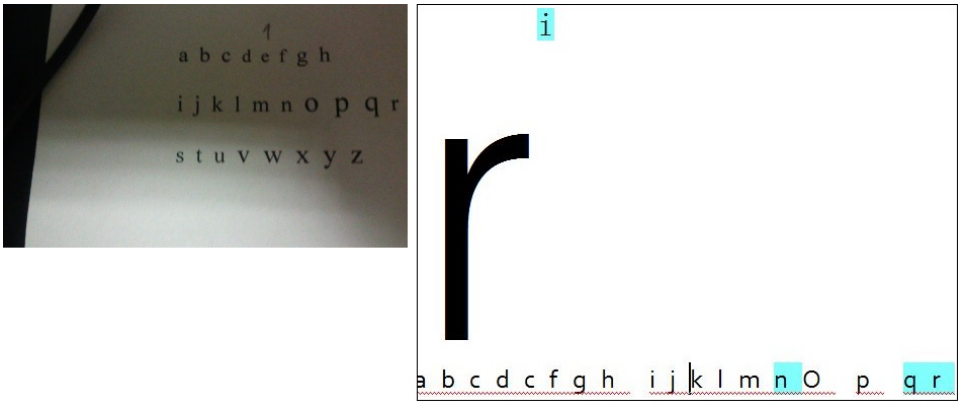

Figura 47: A la izquierda se muestra la imagen real (Times New Roman), y a la derecha el resultado con el sistema OCR ABBYY.

Cuadro 6: Errores de clasificación, Fig(47).

	Falsas negativos	Falsos positivos
ABBYY	8	2
SIFT	16	9
Sistema propuesto	11	8

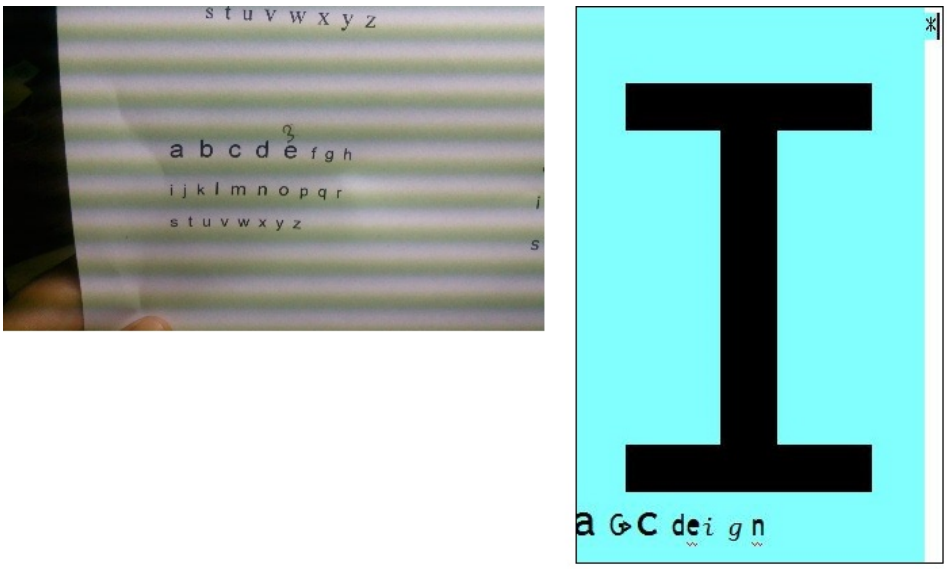

Figura 48: A la izquierda se muestra la imagen real (Arial), y a la derecha el resultado con el sistema OCR ABBYY.

Cuadro 7: Errores de clasificación, Fig(48).

	Falsas negativos	Falsos positivos
ABBYY	21	5
SIFT	22	8
Sistema propuesto	14	13

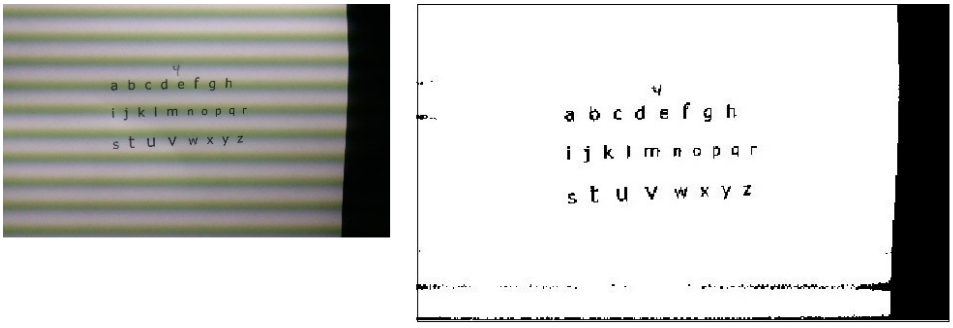

Figura 49: A la izquierda se muestra la imagen real (Verdana), y a la derecha el resultado con el sistema OCR ABBYY.

Cuadro 8: Errores de clasificación, Fig(49).

	Falsas negativos	Falsos positivos
ABBYY	26	-
SIFT	24	13
Sistema propuesto	10	15

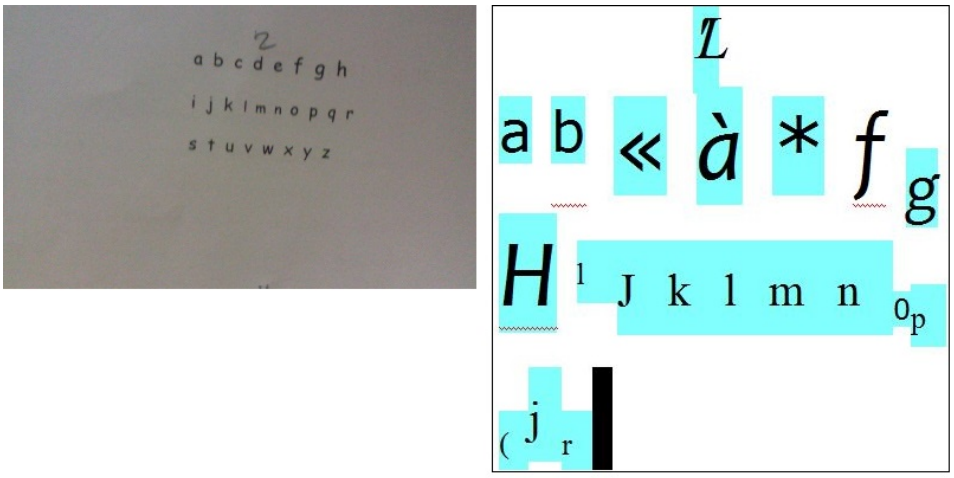

Figura 50: A la izquierda se muestra la imagen real (Comic Sans), y a la derecha el resultado con el sistema OCR ABBYY.

Cuadro 9: Errores de clasificación, Fig(50).

	Falsas negativos	Falsos positivos
ABBYY	14	8
SIFT	19	8
Sistema propuesto	18	9

5.4.3. Discusión de resultados

Los experimentos realizados con imágenes sintéticas se dividieron en 3 conjuntos. El primer conjunto llamado *Conjunto de experimentos 1* se consideraron todos los niveles del sistema. El objetivo era demostrar la efectividad de los filtros en cada nivel aún cuando las imágenes se encuentren contaminadas con ruido Gaussiano. Esto es importante ya que si los filtros que se encuentran en los primeros niveles fallan entonces el reconocimiento está destinado a fallar ya que los últimos niveles dependen totalmente de la eficiencia de los primeros filtros. Los casos con mayores errores de clasificación son los siguientes: reconocimiento del caracter "o" del conjunto de caracteres "q, d, b, p"; reconocimiento del caracter "h" de el caracter "n"; diferenciar el caracter "d" del caracter "q" y el caracter "p" del caracter "b"; finalmente reconocer el caracter "t" del conjunto de caracteres "l,i, y j". La DC en el mejor de los casos se encuentra por arriba del 0.8 y en la mayoría estuvo alrededor del 0.6; finalmente en todos los casos estuvo por arriba del 0.3.

En el *segundo conjunto de experimentos* se consideró la iluminación no homogénea y se realizaron pruebas con sólo algunos de los filtros del sistema, presentando los mejores y enfocándonos en los peores resultados. Como era de esperarse, los errores aumentaron al incrementar la diferencia en la iluminación de la imagen ($\rho = 30$). Nuevamente los filtros con los errores más altos son aquellos pertenecientes a los caracteres "o, q, b, d, p , g". En el caso del caracter "t" los errores de clasificación son mayores en el escalamiento, pero en las otras transformaciones los errores son menores al 10 %. La DC nuevamente estuvo alrededor del 0.6 en los mejores casos y por arriba del 0.35 en los peores casos presentados.

El último conjunto de experimentos realizado fue el *conjunto de experimentos 3*. Aquí se consideraron la mezcla de ruido aditivo e impulsivo así como iluminación no homogénea. Las pruebas se realizaron sólo a algunos filtros del sistema que no fueron considerados en el experimento anterior. Los errores de clasificación más grandes los vuelve a presentar el caracter "b" en la transformación "shearing", en las otras transformaciones los errores se mantienen por debajo del 25 % y la DC se encuentra por arriba del 0.3. En el caso del filtro para el caracter "i", los errores de clasificación aumentaron conforme el ruido y la iluminación varió y sólo fueron mayores del 30 % para la tranformación de estiramiento en y. La DC en todos

los casos se encuentra alrededor del 0.4 y si retiramos la transformación de estiramiento (lo cual no afectaría, ya que dicha transformación no modifica demasiado la forma del caracter en cuestión), la DC se encuentra alrededor del 0.5.

En general los filtros mostraron un buen desempeño al utilizar imágenes sintéticas las cuales presentaban distintas distorsiones que se deseaban evaluar. Posteriormente se realizó un estudio comparativo entre el sistema propuesto, el algoritmo SIFT y el sistema comercial ABBYY, utilizando imágenes reales capturadas por una cámara de teléfono móvil de 5 megapixeles en condiciones de baja iluminación. El documento capturado contenía 26 caracteres y se mantuvo un espacio entre cada caracter, esto debido a que en el sistema propuesto no se consideró ningún tipo de segmentación y por lo tanto el sistema suele fallar con aquellos caracteres que tienen poca distancia entre ellos; por ejemplo, los caracteres “rn” son confundidos con el caracter “m”. Además el tipo de fuente usado en las pruebas afecta el comportamiento del sistema propuesto debido a que en la composición de los filtros se utilizó un sólo tipo de fuente sin considerar algunos cambios que existen entre caracteres de fuentes distintas; por ejemplo, el caracter “a” en Times New Roman no es el mismo símbolo que se utiliza en Arial. Finalmente en las imágenes reales utilizadas no se incluyeron palabras completas, ya que el sistema ABBYY realiza reconocimiento por contexto, haciendo uso de diccionarios, lo que daría como resultado una comparación injusta debido a que posiblemente pueda no reconocer un caracter, pero si la palabra completa.

Aún con estos inconvenientes, logramos alcanzar resultados favorables comparándolos con los resultados del sistema comercial ABBYY y el algoritmo basado en características SIFT. El sistema ABBYY tiene deficiencias en cuanto a la resolución de la imagen, ya que si ésta es muy baja, el programa no logra reconocer absolutamente nada. Otro inconveniente del sistema es la técnica de binarización que utiliza, ya que si la imagen tiene un cambio muy abrupto de iluminación el sistema simplemente no lo binariza correctamente dando como resultado sólo un área en negro. En el caso del algoritmo SIFT sabemos que es un algoritmo basado en características y que su éxito depende totalmente de las características extraídas de la imagen de referencia por lo que en este caso la imagen del caracter no proporciona las suficientes características que ayuden a reconocerlo.

En general, los mayores errores de clasificación obtenidos fueron los generados por falsos negativos, es decir, los filtros no alcanzaron a reconocer algunos caracteres; el mayor inconveniente de esto, es que el diseño del sistema se basa en niveles y si el caracter no es reconocido desde el primer nivel entonces no podrá ser reconocido por el filtro que lo clasifique al final; por lo que es necesario lograr un nivel de confianza mayor en cada uno de los niveles.

5.5. Conclusiones

El reconocimiento de caracteres en imágenes de documentos es un problema muy complejo, ya que se deben considerar diversos factores que pueden interferir y/o alterar la imagen. Factores físicos propios del sistema de captura, factores ambientales presentes en la toma de la imagen, y hasta factores externos, como lo son el tipo de hoja del documento o su estado físico. Además, se tiene que lidiar con distintos tipos de símbolos, fuentes y caracteres lo que vuelve un verdadero reto el reconocerlos y clasificarlos.

En este trabajo se abordó dicha problemática, proponiendo un sistema multi-nivel que reconozca y clasifique con cierto nivel de confianza. Se utilizaron filtros no lineales para el diseño del sistema, intentando lograr un nivel de confianza alto para cada uno de los niveles. Como era de esperarse, el mayor reto se encuentra en la clasificación de caracteres muy similares. Se podría pensar que sólo es necesario identificar desde un principio aquellos caracteres similares y buscar la manera de diferenciarlos, pero no es tan simple. El problema es que un caracter puede ser similar a otros caracteres no sólo por su forma inicial si no por distorsiones que pueden aumentar su similitud con otros caracteres. Un ejemplo de esto es la diferencia de tamaños, es decir, dos caracteres no necesariamente son similares a la misma escala, pero si uno es más grande que el otro podría ser que la similitud aumente; también afectan las rotaciones o distorsiones que deforman al caracter y hacen que parezca otro. Otro inconveniente son los caracteres que se encuentran a los lados del caracter a reconocer, ya que suelen en ocasiones unirse y tomar la apariencia de otro caracter. Incluir todos estos detalles, vuelven al sistema mucho más complejo y por lo tanto mucho más lento.

En este trabajo se logró generar un sistema que cumple con cierto nivel de desempeño en

reconocimiento y clasificación bajo condiciones de iluminación no homogénea, ruido y distorsiones geométricas, pero se encuentra un poco limitado al tipo de fuente. En comparación con el sistema comercial ABBYY y el algoritmo SIFT se logró obtener resultados bastante favorables.

Finalmente, es necesario trabajar mucho más en el sistema propuesto y mejorar algunos aspectos, los cuales se describen a continuación.

5.6. Trabajo futuro

- Reducir el número de filtros utilizados por transformación geométrica. En algunos casos, dos filtros distintos que se encuentran en el mismo banco de filtros pueden reconocer la misma imagen, por lo que se podría prescindir de alguno de ellos (filtros dentro de la misma transformación).
- Disminuir el número de filtros verificando si una transformación reconoce otra transformación geométrica retirando así dicho filtro (filtros entre distintas transformaciones).
- Para aquellos filtros con bajo DC, utilizar alguna variación que aumente el DC obtenido como por ejemplo, separar en más clases el conjunto y con ello considerar mayores variaciones.
- Mejorar la división de los caracteres en niveles. Verificar si la clasificación realizada es la óptima o existe alguna otra clasificación que arroje un mayor DC.
- Implementar el algoritmo en GPU, disminuyendo así su tiempo de ejecución.
- Sintetizar nuevos filtros para imágenes escaladas, ya que al escalar una imagen puede que se correlacione con un nuevo caracter al variar su tamaño.

Lista de referencias

Aguilar-González, P. M., Kober, V., y Díaz-Ramírez, V. H. (2014). Adaptive composite filters for pattern recognition in nonoverlapping scenes using noisy training images. *Pattern Recognition Letters*, **41**: 83–92.

Álvarez-Borrego, J., Solorza, S., y Bueno-Ibarra, M. A. (2013). Invariant correlation to position and rotation using a binary mask applied to binary and gray images. *Optics Communications*, **294**: 105–117.

Díaz-Ramírez, V. H., Picos, K., y Kober, V. (2014). Target tracking in nonuniform illumination conditions using locally adaptive correlation filters. *Optics Communications*, **323**: 32–43.

Doermann, D., Liang, J., y Li, H. (2003). Progress in camera-based document image analysis. En: *Document Analysis and Recognition, Proceedings of the Seventh International Conference on*, pp. 606–616. IEEE.

Doh, Y.-H., Kim, J.-C., Kim, J.-W., Choi, K.-H., Kim, S.-J., y Alam, M. S. (2004). Distortion-invariant pattern recognition based on a synthetic hit-miss transform. *Optical Engineering*, **43**(8): 1798–1803.

Esser, D., Muthmann, K., y Schuster, D. (2013). Information extraction efficiency of business documents captured with smartphones and tablets. En: *Proceedings of the 2013 ACM symposium on Document engineering*, pp. 111–114. ACM.

Fitch, J., Coyle, E. J., y Gallagher Jr, N. C. (1984). Median filtering by threshold decomposition. *Acoustics, Speech and Signal Processing, IEEE Transactions on*, **32**(6): 1183–1188.

Gatos, B., Pratikakis, I., y Perantonis, S. J. (2006). Adaptive degraded document image binarization. *Pattern recognition*, **39**(3): 317–327.

González, R. y Woods, R. (2006). *Digital Image Processing, (3a ed.)*. Prentice-Hall Inc. Upper Saddle River, NJ, USA.

González-Fraga, Kober, V., Angel, J., y Álvarez-Borrego, J. (2006). Adaptive synthetic discriminant function filters for pattern recognition. *Optical Engineering*, **45**(5): 057005–057005.

He, J., Do, Q., Downton, A. C., y Kim, J. (2005). A comparison of binarization methods for historical archive documents. En: *Document Analysis and Recognition, Proceedings of the Eighth International Conference on*, pp. 538–542. IEEE.

Holley, R. (2009). How good can it get? analysing and improving ocr accuracy in large scale historic newspaper digitisation programs. *D-Lib Magazine*, **15**(3/4).

Horn, B. K. (1990). Height and gradient from shading. *International journal of computer vision*, **5**(1): 37–75.

Horner, J. L. y Gianino, P. D. (1984). Phase-only matched filtering. *Applied optics*, **23**(6): 812–816.

Jacobs, C., Simard, P. Y., Viola, P., y Rinker, J. (2005). Text recognition of low-resolution document images. En: *Document Analysis and Recognition, Proceedings of the Eighth International Conference on*, pp. 695–699. IEEE.

Kir, B., Oz, C., y Gulbag, A. (2013). The application of optical character recognition for mobile device via artificial neural networks with negative correlation learning algorithm. En: *Electronics, Computer and Computation (ICECCO), 2013 International Conference on*, pp. 220–223. IEEE.

Kise, K., Sato, A., y Iwata, M. (1998). Segmentation of page images using the area voronoi diagram. *Computer Vision and Image Understanding*, **70**(3): 370–382.

Kober, V. I. y Mozerov, M. G. (2000). Phase-only filter with improved filter efficiency and correlation discrimination. *Pattern Recognition and Image Analysis*, **10**(4): 514–519.

Kober, V. I., Mozerov, M. G., y Álvarez-Borrego, J. (2001). Nonlinear filters with spatially connected neighborhoods. *Optical Engineering*, **40**(6): 971–983.

Kober, V. I., Mozerov, M. G., Álvarez-Borrego, J., y Ovseyevich, I. A. (2002). Unsharp masking by the rank-order filters with spatially adaptive neighborhoods. *Pattern Recognition and Image Analysis*, **12**(1): 46–56.

Kober, V. I., Mozerov, M. G., Álvarez-Borrego, J., y Ovseyevich, I. A. (2004). Adaptive rank-order correlations. *Pattern Recognition and Image Analysis*, **14**(1): 33–39.

Kumar, B. V., Mahalanobis, A., y Juday, R. D. (2005). *Correlation pattern recognition*, Vol. 27. Cambridge University Press Cambridge.

LeCun, Y., Bottou, L., Bengio, Y., y Haffner, P. (1998). Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, **86**(11): 2278–2324.

Liang, J., Doermann, D., y Li, H. (2005). Camera-based analysis of text and documents: a survey. *International Journal of Document Analysis and Recognition (IJDAR)*, **7**(2-3): 84–104.

Lowe, D. G. (2004). Distinctive image features from scale-invariant keypoints. *International journal of computer vision*, **60**(2): 91–110.

Lund, W. B., Kennard, D. J., y Ringger, E. K. (2013). Combining multiple thresholding binarization values to improve ocr output. En: *IS&T/SPIE Electronic Imaging*, pp. 86580R–86580R. International Society for Optics and Photonics.

Ma, D. y Agam, G. (2013). A super resolution framework for low resolution document image ocr. En: *IS&T/SPIE Electronic Imaging*, pp. 86580P–86580P. International Society for Optics and Photonics.

Maragos, P. (1988). Optimal morphological approaches to image matching and object detection. En: *Computer Vision., Second International Conference on*, pp. 695–699. IEEE.

Maragos, P. (1989). Morphological correlation and mean absolute error criteria. *ASSP, Int. Conf. on. IEEE*, pp. 1568–15721.

Maragos, P. (2004). Morphological filtering for image enhancement and feature detection. *analysis*, **19**: 18.

Martínez-Díaz, S. y Kober, V. (2008a). Distortion-invariant pattern recognition with nonlinear correlation filters. En: *Optical Engineering+ Applications*, pp. 707327–707327. International Society for Optics and Photonics.

Martínez-Díaz, S. y Kober, V. (2008b). Nonlinear synthetic discriminant function filters for illumination-invariant pattern recognition. *Optical Engineering*, **47**(6): 067201–067201.

Martínez-Díaz, S. y Kober, V. (2011). Morphological correlation for robust image recognition. En: *Computational Science and Its Applications (ICCSA), International Conference on*, pp. 263–266. IEEE.

Martínez-Díaz, S., Kober, V., y Ovseyevich, I. (2008). Adaptive nonlinear composite filters for pattern recognition. *Pattern Recognition and Image Analysis*, **18**(4): 613–620.

Matei, O., Pop, P. C., y Vălean, H. (2013). Optical character recognition in real environments using neural networks and k-nearest neighbor. *Applied intelligence*, **39**(4): 739–748.

Niblack, W. (1985). *An Introduction to Digital Image Processing*. Strandberg Publishing Company. Birkeroed, Denmark, Denmark, pp. 112–116.

Otsu, N. (1975). A threshold selection method from gray-level histograms. *Automatica*, **11**(285-296): 23–27.

Ozarslan, S. y Eren, P. E. (2014). Text recognition and correction for automated data collection by mobile devices. En: *IS&T/SPIE Electronic Imaging*, pp. 902706–902706. International Society for Optics and Photonics.

Patel, C., Patel, A., y Patel, D. (2012). Optical character recognition by open source ocr tool tesseract: A case study. *International Journal of Computer Applications*, **55**(10): 50–56.

Russ, J. C. (2010). *The image processing handbook*. CRC press. pp. 12–6.

Salehpour, M. y Behrad, A. (2010). Cluster based weighted svm for the recognition of farsi handwritten digits. En: *Neural Network Applications in Electrical Engineering (NEUREL), 2010 10th Symposium on*, pp. 219–223. IEEE.

Sauvola, J., Seppanen, T., Haapakoski, S., y Pietikainen, M. (1997). Adaptive document binarization. En: *Document Analysis and Recognition, Proceedings of the Fourth International Conference on*, Vol. 1, pp. 147–152. IEEE.

Serra, J. (1986). Introduction to mathematical morphology. *Computer vision, graphics, and image processing*, **35**(3): 283–305.

Shafait, F., Keysers, D., y Breuel, T. M. (2008a). Efficient implementation of local adaptive thresholding techniques using integral images. *Proc. SPIE*, **6815**(8): 681510–681510.

Shafait, F., Keysers, D., y Breuel, T. M. (2008b). Efficient implementation of local adaptive thresholding techniques using integral images. En: *Electronic Imaging 2008*, pp. 681510–681510. International Society for Optics and Photonics.

Simard, P. Y., Steinkraus, D., y Platt, J. C. (2003). Best practices for convolutional neural networks applied to visual document analysis. En: *12th International Conference on Document Analysis and Recognition*, Vol. 2, pp. 958–958. IEEE Computer Society.

Smith, R. (2007). An overview of the tesseract ocr engine. En: *ICDAR*, Vol. 7, pp. 629–633.

Ulges, A., Lampert, C. H., y Breuel, T. M. (2005). Document image dewarping using robust estimation of curled text lines. En: *Document Analysis and Recognition, Proceedings of the Eighth International Conference on*, pp. 1001–1005. IEEE.

Zhang, H., Zhao, K., Song, Y.-Z., y Guo, J. (2013). Text extraction from natural scene image: A survey. *Neurocomputing*, **122**: 310–323.