

885216



UNIVERSIDAD AMERICANA DE ACAPULCO
EXCELENCIA PARA EL DESARROLLO

FACULTAD DE INGENIERÍA EN COMPUTACIÓN
INCORPORADA A LA UNIVERSIDAD NACIONAL
AUTÓNOMA DE MÉXICO

**RECONOCIMIENTO DE OBJETOS EN
IMÁGENES INFRARROJAS A TRAVÉS DEL
ANÁLISIS Y PROCESAMIENTO DE
IMÁGENES**

T E S I S

QUE PARA OBTENER EL TÍTULO DE:
INGENIERA EN COMPUTACIÓN

PRESENTA:
ELOISA MERCEDES VIVAS VILLASANA

DIRECTOR DE TESIS
ING. GONZALO TRINIDAD GARRIDO



ACAPULCO, GRO.

NOVIEMBRE DEL 2005

M. 339911



Universidad Nacional
Autónoma de México



UNAM – Dirección General de Bibliotecas
Tesis Digitales
Restricciones de uso

DERECHOS RESERVADOS ©
PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL

Todo el material contenido en esta tesis esta protegido por la Ley Federal del Derecho de Autor (LFDA) de los Estados Unidos Mexicanos (México).

El uso de imágenes, fragmentos de videos, y demás material que sea objeto de protección de los derechos de autor, será exclusivamente para fines educativos e informativos y deberá citar la fuente donde la obtuvo mencionando el autor o autores. Cualquier uso distinto como el lucro, reproducción, edición o modificación, será perseguido y sancionado por el respectivo titular de los Derechos de Autor.

Agradecimientos

Ing. Gonzalo Trinidad Garrido

Por el apoyo incondicional a lo largo de toda la carrera.

Ing. Juan Carlos Cañizares

Por la dedicación y consejos en la revisión de tesis.

Lic. Carolina Palacios

Por el apoyo incondicional, sus consejos y sobretodo por ser mi amiga.

Ing. Jorge Maisterrena

Por enseñarme que la vida no termina cuando uno tiene pequeños tropiezos.

Universidad Americana de Acapulco

Facultad de Ingeniería en Computación,

A los profesores.

Dedicatoria

A Dios

Por la bendición de la vida y la familia

A Mis Padres

Sr. Luis Vivas García y Sra. Elsa Villasana González

Por sus esfuerzos incansables para la realización de la carrera

A Mis Hermanos

Gaby, Carlos y Luis Eduardo

Por ser la inspiración de mi vida

A Mi Amado Esposo

Christian Alberto Ramírez Hernández

Por el amor, paciencia y apoyo brindado incondicionalmente durante la carrera

A Mi Cuñado

Juan Manuel Sánchez Morales

Por ser un pilar importante para la familia

A Mis Sobrinos

Claudia Yesenia y Luis Ángel

Por ser el motor de mi vida

A Mi Tío

Gonzalo Villasana González *

Siempre en mi corazón

Índice

	Página.
Portada	1
Agradecimientos	2
Dedicatoria	3
Índice	4
Lista de Figuras	8
Lista de Tablas	9
Introducción	10
Planteamiento del Problema	10
Justificación	12
Objetivo General	13
Objetivos Particulares	13
Hipótesis	14
Capítulo 1 Estado del Arte	
Definición del Sistema ATR	15
Factores de Influencia	16
Etapas principales y desempeño de un sistema ATR	
Preprocesamiento	17
Detección y Segmentación de Objetivos	17
Extracción de Características	17
Clasificador	18
Evolución de los Sistemas ATR	19
Sistemas de Reconocimiento Existentes	23

Aplicaciones

Detección de objetivos en imágenes co-registradas en el rango visual térmico	24
Identificación de objetivos usando momentos de Zernike y Redes Neuronales	25
Reconocimiento Bayesiano de objetivos por partes en segunda generación de imágenes FLIR's	27
Reconocimiento Automático de Placas de Matriculación de Vehículos	28
Estructura de la Tesis	30

Capítulo 2 Conceptos Básicos

Termografía	31
Formación de Imágenes Infrarrojas	33
Adquisición y Procesamiento Digital de Imágenes Infrarrojas	35
La Cámara	36
Especificaciones Técnicas	37
Filtros Morfológicos	
Erosión	38
Dilatación	39
Apertura	40
Descriptor de formas basadas en momentos	41
Momentos de Geométricos	42
Momentos de Zernike	43
Redes Neuronales	46
Función de Activación (Función de neurona)	48

Capítulo 3 Desarrollo e Implementación del Sistema

Metodología	49
Etapa de Preprocesamiento	53
Etapa de Segmentación	54
Discriminación de Falsos Objetivos	55
Etapa de Extracción	58
Etapa de Clasificación	61

Capítulo 4 Resultados del Sistema

Características

Algoritmo 1. Momentos Geométricos para Imágenes Sintéticas con objetivos de varias vistas	63
Algoritmo 2. Momentos geométricos para imágenes infrarrojas con objetivos de una sola vista	65
Algoritmo 3. Momentos de Zernike para imágenes infrarrojas de objetivos para una sola vista	67
Algoritmo 4. Momentos de Zernike para imágenes infrarrojas con objetivos de varias vistas	68

Resultados

Resultados del Algoritmo 1.

Imágenes Sintéticas varias vistas - Momentos Geométricos 71

Resultados del Algoritmo 2.

Imágenes Infrarrojas de una sola vista - Momentos Geométricos 72

Resultados del Algoritmo 3.

Imágenes Infrarrojas de una sola vista - Momentos de Zernike 73

Resultados del Algoritmo 4

Imágenes Infrarrojas de varias vistas - Momentos de Zernike 74

Capítulo 5 Conclusiones

Conclusiones 77

Trabajo Futuro 78

Bibliografía 80

Lista de Figuras

		Página.
Figura A.	Etapas de un sistema ATR	18
Figura B.	Problemática ATR típica	19
Figura C.	Sistema funcional	21
Figura D.	Imagen infrarroja de una escena y su proporcional en binaria	32
Figura E.	Espectro electromagnético	35
Figura F.	Proceso de adquisición de una imagen	36
Figura G.	Elemento estructurante de la operación morfológica de apertura	38
Figura H.	Ejemplos de operadores morfológicos de erosión y dilatación	39
Figura I.	Ejemplo gráfico de un operador de apertura	40
Figura J.	Arquitectura básica de una red neuronal	47
Figura K.	Ejemplo de Imágenes de diferentes distancias	51
Figura L.	Ejemplo de las diferentes vistas de un objeto	52
Figura M.	Funcionamiento General del Sistema	53
Figura N.	Imagen Binaria (lado izquierdo) e Imagen Filtrada (lado derecho)	55
Figura O.	Imagen Binaria Dilatada	57
Figura P.	Imagen de Extracción de Bordes	58
Figura Q.	Algoritmo de Extracción y Normalización de Características	59
Figura R.	Ejemplo de las imágenes utilizadas	64
Figura S.	Ejemplo de las imágenes infrarrojas (regiones).	67
Figura T.	Arquitectura de la Red para los momentos de Zernike	69
Figura U.	Arquitectura de la Red para los momentos de Zernike	70

Lista de Tablas

	Página.
Tabla 1. Lista los Momentos de Zernike	62
Tabla 2. Relación de Imágenes	64
Tabla 3. Relación de Regiones de Entrenamiento y Prueba	68
Tabla 4. Matriz de confusión con momentos geométricos para imágenes sintéticas	71
Tabla 5. Matriz de confusión con momentos geométricos para imágenes infrarrojas	72
Tabla 6. Matriz de confusión con momentos de Zernike para Imágenes Infrarrojas de una sola vista	73
Tabla 7. Matriz de confusión con momentos de Zernike para la fase de entrenamiento	75
Tabla 8. Matriz de confusión con momentos de Zernike para la fase de prueba	76

Introducción

El reconocimiento automático de objetivos (ATR, *Automatic Target Recognition*) en imágenes infrarrojas, es un problema de considerable interés en varias aplicaciones industriales, con algunas limitaciones tales como el trabajo en tiempo real, alta precisión en detección, capacidades adaptativas para ambientes con grandes cantidades de ruido, incremento en la complejidad de las tareas de detección del objetivo, etc.

Planteamiento del Problema

El reconocimiento automático de objetivos es una importante tarea en la explotación de las imágenes. Los sistemas de reconocimiento automático de objetivos deben ser capaces de detectar, clasificar, reconocer e /o identificar objetivos en un ambiente donde el fondo es confuso y los objetivos están a largas distancias y tal vez estos estén parcialmente ocluidos, degradados por el clima o aparezcan camuflajeados.

Normalmente en este tipo de escenarios los objetivos deben ser reconocidos en escenarios complejos al aire libre, bajo condiciones adversas, con factores adicionales tales como ruido en toda la escena, información equívocada ó engañosa que confunda el sistema de reconocimientos, haciendo el proceso de reconocimiento aún más difícil.

Se han usado varias metodologías para resolver el problema de ATR. Pero para este caso implementamos la metodología clásica del reconocimiento de patrones, que usan técnicas estadísticas y estructurales.

Estas técnicas son basadas en las características que ofrece cada objetivo. El sistema se desarrolla en una escena estática, donde el fondo contiene cuantiosa información en forma de ruido desordenado.

La mayoría de las investigaciones realizadas en los sistemas ATR para sistemas forward - looking infrared (FLIR) han sido propuestas para el área militar tales como el traqueo, navegación de aviones o misiles guiados a baja altitud [26][27].

La extracción de objetivos en las imágenes infrarrojas es una tarea compleja de resolver ya que estos son afectados por la temperatura y distribución de la energía radiante, además de la interacción entre la fuente y el fondo del medio ambiente en el que se desarrolla la escena [28].

En la actualidad existe una amplia variedad de programas comerciales destinados al reconocimiento, esto ha tenido un gran auge en el área militar, esto se ha implementado para detectar minas acuáticas bajo diversas condiciones, manejando técnicas de momentos de Zernike y redes neuronales, reportando un 91% del rango de clasificación correcta.

En general la mayoría de los experimentos para el reconocimiento de objetivos con imágenes infrarrojas FLIR reportan un rango del 89% al 95% de clasificación correcta, dependiendo de la técnica que se esté utilizando y la cantidad de imágenes que se están utilizando, estas técnicas pueden variar desde el módulo de preprocesamiento hasta el módulo de reconocimiento de un sistema ATR, técnicas como restricción convexa, reconocimiento bayesiano, redes neuronales, análisis de componentes principales, vecino más cercano, etc.

Las condiciones de la escena donde se realiza el experimento también influyen en el resultado del proceso, es decir, si la escena presenta condiciones climatológicas influyentes tales como lluvia, lluvia con granizo, temperaturas altas o bajas, presencia de árboles, construcciones, etc.; dificultando el proceso de reconocimiento de objetivos del sistema obteniendo malos resultados al producirse gran cantidad de ruido en la escena, en contra parte, las imágenes con escenas en condiciones normales, el proceso de reconocimiento es mayor.

Por lo expuesto anteriormente, el reconocimiento de objetivos en imágenes infrarrojas dista de ser una tarea fácil, ya que por lo general no existen parámetros de control en la escena bajo observación ya que las condiciones que presente la escena va a depender de las condiciones climáticas actuales que se estén dando en ese momento en el ambiente.

Justificación

Las imágenes infrarrojas son caracterizadas por contener pequeñas señales de ruido provocando que el proceso de extracción de objetivos de interés en la escena sea más complejo. Las imágenes infrarrojas también tienden a distorsionar las características de los objetivos debido al reflejo local y transferencia radiante del ambiente.

Esta distorsión hace difícil el proceso de descripción del objetivo basado sólo en las características de la forma del objetivo. En las imágenes infrarrojas el fondo contiene gran cantidad de ruido; así que los objetivos pueden presentarse en diversas caras y a diferentes distancias.

La mayoría de los sistemas de visión están restringidos a problemáticas específicas que se requieren de toda la apreciación de la escena donde va a interactuar el sistema.

Por otra parte, si tomamos en cuenta que el objetivo variaría con respecto a su forma, ya que este trabajo contempla el objetivo bajo diferentes vistas, así como debe ser reconocido a diferentes distancias y cualquier transformación afine.

Objetivo General

Desarrollar un Sistema de Visión que automatice el proceso de reconocimiento de objetivos reales en imágenes infrarrojas, haciendo énfasis en la extracción de características, donde los objetivos estarán basados en vistas[1]; aun cuando existan las variantes de escala, traslación, rotación, las vistas del objetivo registradas muy parecidas deben ser reconocidas por el sistema.

Objetivos Particulares

- Estudiar los conceptos básicos necesarios para el análisis, diseño y procesamiento de imágenes que serán de interés al sistema propuesto.
- Capturar el conjunto de imágenes con una cámara termal que no experimenta movimiento alguno, de tal manera que la escena sea estática.
- Implementar diversos algoritmos que contemplen la problemática de transformaciones afines en general.
- Realizar la clasificación y reconocimiento de objetivos a través de redes neuronales en las imágenes infrarrojas.
- El sistema reconocerá objetivos dentro de una escena estática.

Hipótesis:

Una de las principales características del sistema es poder llevar a cabo el reconocimiento en imágenes con alta cantidad de ruido, ya que este ocasiona que los objetivos de interés sean confundidos como el resto de la imagen o producir falsas alarmas.

Los sistemas actuales de visión por computadora para imágenes infrarrojas no implementan operadores morfológicos, para reducir la cantidad de ruido en la imagen, ni momentos de Zernike para la extracción de características, haciendo el sistema más robusto al ruido inherente del objetivo, ocasionando que el reconocimiento sea ineficiente.

El sistema planteado presenta características interesantes como son el reconocimiento de objetos bajo variaciones de traslación, de escala y de rotación. Además de robustez frente a imágenes con ruido presente.

Estado del Arte

Definición del Sistema ATR

La identificación automática de objetivos (*Automatic Target Recognition - ATR*) es una de las tecnologías que hace uso de la termografía infrarroja la cual ha tenido un auge en el área militar los últimos años, en especial en el ámbito de la aviación, por ejemplo a los pilotos les ayuda en la cabina a evaluar de una forma rápida y precisa cada vez mayor volumen de información

Los datos necesarios para el sistema ATR se obtienen mediante una interpretación de las imágenes generadas por los sensores, que, a ser posible, serán pasivos y, por tanto, indetectables, como por ejemplo, del tipo FLIR (*Forward Looking Infrared*, o visión infrarroja frontal).

El procedimiento ATR permite la identificación y clasificación de los objetos buscados o que hay que proteger. Los ámbitos de aplicación son tanto el reconocimiento táctico y estratégico en las misiones de combate como en aplicaciones civiles, por ejemplo, de la policía. En el futuro, el sistema ATR será imprescindible como tecnología de base para el éxito de las misiones de los aviones de combate y reconocimiento no tripulados (*Unmanned Air Vehicle - UAV*).

La tecnología ATR reducirá enormemente la cantidad de datos gráficos a transmitir, ya que podrá distinguir la información importante de la irrelevante. Sólo la información relevante se reproducirá con la máxima resolución.

En las imágenes térmicas, los objetivos de interés ocupan un pequeño número de píxeles a través de una secuencia de imágenes. Las imágenes son caracterizadas por contener altas cantidades de ruido, continuamente varía el fondo de la imagen, y la cantidad de información tiende a ser enorme. Típicamente las clases de objetivos contienen objetos tales como vehículos, pero estos pueden contener objetos tales como árboles, edificios, etc.

Factores de Influencia

El sistema depende de varios factores que determinarán el desempeño durante el desarrollo del sistema.

- La calidad de los resultados: La probabilidad de la clasificación errónea.
- La eficiencia del sistema: La cantidad de imágenes (objetivos) procesadas por segundo por el sistema.
- Recursos del Sistema: Los procesadores, memoria, E / S del ancho de banda, etc.

Etapas principales y desempeño de un sistema ATR

El sistema está constituido por varios subsistemas tales como preprocesamiento, segmentación, extracción de características, clasificación e identificación. Básicamente estos son agrupados en tres grupos, que se describirán brevemente como funciona de manera general cada uno de estas etapas:

Existen varios métodos para este fin, los esquemas basados en momentos son los más usados ya que estos proporcionan características invariantes ideales de traslación, rotación y escalamiento en imágenes 2D y 3D en aplicaciones de reconocimiento de patrones. Estos tipos de momentos pueden ser momentos ortogonales, Legendre y Zernike, pero siempre hay que tomar en cuenta que las características distinguen a cada una de las regiones que aparezcan en una imagen y el resultado de esta etapa es un vector de características que describa a cada una de las regiones segmentadas.

Clasificador: La etapa de clasificación es el proceso donde cada objetivo (región) se asigna a un grupo o clase dentro de un número de clases para determinar si el objetivo clasificado es el que se está buscando. En esta parte el proceso utiliza las características extraídas de cada una de las regiones contenidas en la imagen, para determinar a que clase pertenece el objetivo. A continuación en la Figura B, se presentan las etapas principales de un sistema ATR.

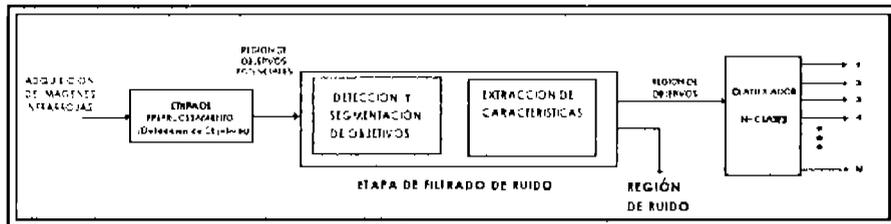


Figura A. Etapas de un Sistema ATR.

En cuanto al desempeño de los sistemas ATR estos están basados en modelos de objetivos con alguna complejidad determinada. Los modelos más complejos requieren más procesamiento para obtener mejores resultados. La complejidad del modelo y el poder de procesamiento computacional determinan el desempeño total del sistema.

Un caso tipo de los sistemas de reconocimiento automático es el de autos, donde se maneja una colección de imágenes térmicas como se aprecia en la Figura C, donde cuyos objetivos

a reconocer son los autos en imágenes térmicas, estas a su vez nos van a servir como base para comparar con las imágenes que se vaya adquiriendo en el sistema de reconocimiento, el cual debe ser capaz de determinar que tipo de automóvil basándose en una base de datos que tiene el sistema. Como se puede apreciar en la Figura C, un caso real donde cada una de las etapas anteriormente mencionadas son aplicadas.

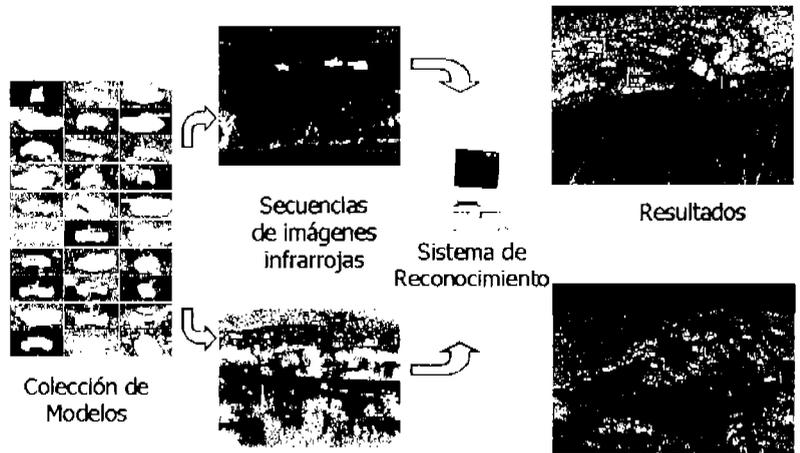


Figura B. Problemática ATR Típica

El propósito no es un factor potencial para el perjudicar el reconocimiento, que son invariantes a rotación, traslación y escala en comparación que los momentos Geométricos, asemeje a los modelos internos del mundo o dominio de interés.

Evolución de los Sistemas ATR.

El reconocimiento de objetos de tres dimensiones (3D) a imágenes de dos dimensiones (2D) es una importante rama de la visión por computación. El estudio y desarrollo experimental de sistemas de reconocimiento de objetos ha tenido considerable impacto en la dirección y contenido de la visión por computadora. A pesar de la existencia de varios paradigmas, algoritmos y sistemas que han sido propuestos a lo largo de las dos décadas pasadas.

Soluciones no versátiles han sido desarrolladas para este problema, solo soluciones parciales y logros limitados en ambientes muy restringidos son el estado del arte.

De hecho algunos investigadores creen que no es posible diseñar un sistema de reconocimiento de objetos que sea funcional para cualquier escena y ambiente el cual sea un sistema eficiente en una situación específica.

La dificultad en obtener una solución general y comprensiva a este problema puede ser atribuido a la complejidad del reconocimiento de objetos, esto involucra un procesamiento en todos los niveles de visión artificial (machine vision). Visión de nivel bajo, abarca lo que es detección de bordes y segmentación en imágenes. En el caso de la visión de nivel medio, se refiere a la descripción y representación de forma de patrones y extracción de características. Y en la visión de nivel alto, involucra a patrones de categoría de asignación, correspondencia y razonamiento.

El buen funcionamiento de un sistema de reconocimiento de objetos depende de los sucesos ocurridos en cada uno de estos niveles. La tarea es complicada por varios factores tales como él no saber cuantos objetos son presentados en la imagen, la posibilidad que los objetos tal vez estén ocluidos, la posibilidad de que objetos desconocidos aparezcan en la imagen, el movimiento de los objetos, variaciones en el muestreo del ambiente, las limitaciones de precisión del sensor.

Uno de los componentes importantes del presente y futuro son los sistemas de defensa que son usados en misiones de vehículos autónomos en un sistema de reconocimiento automático de objetivos (ATR). El sistema ATR eficazmente manejan los procesos de adquisición de objetivos y reconocimiento. Esto es deseable desde el punto de vista sistema, con la intervención del hombre esto se convertiría lento, inestable, vulnerable, y se limitaría a la actuación del sistema global o misiones en situaciones reales [5].

Una aplicación importante de los sistemas ATR en la actualidad, es guiar a pilotos de aviones de alto rendimiento que vuelan cerca de la tierra durante difíciles condiciones climáticas de tiempo o por la noche. Algunos ejemplos de sistemas que incorporan ATR son la navegación de altitud baja, sistemas de objetivos infrarrojos durante noche (LANTIRN), misil crucero y vehículos piloteados remotamente (RPV) en aplicaciones tales como el Águila RPV. Básicamente, el sistema de ATR realiza, adquisición automática de objetivos, identificación, y el procesamiento de rastreo en una secuencia de imágenes. Como podemos apreciar en la Figura D, un ejemplo de una aplicación real en el cual el sistema de reconocimiento detecta 3 objetos, de los cuales los clasifica como tanques, camiones, humvy (camión con condiciones para todo tipo de terreno, anfibio).



Figura C. Sistema funcional.

En general, el conjunto de objetivos podría consistir de tanques, camiones, barcos, autos, edificios, casas, etc. La meta es realizar estas funciones en tiempo real y poder adaptar a las situaciones tácticas dinámicas. El dominio del problema requiere las herramientas de imagen de procesamiento (IP), análisis de imagen (IA), reconocimiento de patrones (PR), e inteligencia artificial (AI). La investigación en esta área ha sido durante los últimos 25 años, pero sólo recientemente se han tenido algoritmos sofisticados, microprocesadores, y tecnología VLSI y VHSIC [6], lo cual ha originado que esta tecnología tenga mayor disponibilidad, así como también, ha tenido contribuciones en las ondas milimétricas de infrarrojos y la tecnología de sensor de láser que ahora ya es factible para lograr los objetivos de ATR. Sin embargo, el trabajo de los algoritmos en esta área sigue estando en sus inicios [7].

Los sistemas de ATR usan una amplia gama de algoritmos. Es importante desarrollar un criterio del desempeño cuantitativo para los sistemas de ATR por varias razones, 1) comparar varios sistemas de ATR y predecir su desempeño en un escenario dado, 2) estudiar el comportamiento de un sistema ATR y sus componentes bajo diferentes condiciones y conjunto de parámetros, de manera que se pueda encontrar el desempeño óptimo logrado y las tolerancias permisibles de sus componentes, 3) para entender las características del objetivo y el ambiente que afectan la actuación de algoritmos, 4) para encontrar elementos funcionales comunes actualmente entre los algoritmos en uso, 5) para ayudar al diseñador del algoritmo a escoger los algoritmos apropiados para su aplicación, y 6) para proporcionar una aceptable y completa evaluación de la metodología para propósitos de estandarización.

Hasta ahora, los algoritmos ATR que se han evaluado no han tenido ninguna medida en contra de las pruebas que se han realizado con un conjunto de datos muy limitados, el desempeño de la clasificación ha sido generosa según los reportes. Sin embargo, en la práctica estos esfuerzos han sido sólo parcialmente exitosos y se han producido muchas falsas alarmas.

Algunas de las razones importantes de estos acontecimientos son la no repetibilidad de las firmas de estos objetivos, compitiendo con objetivos con ruido que tienen la misma forma como los objetivos reales, experiencia de una base de datos muy limitada, objetivos opacos, poco uso de información disponible relacionada y presentada en la imagen, tales como contexto, estructura, rango, etc.

Si se hiciera el uso de estas diversas fuentes de información, entonces se esperaría que las características de la firma del objetivo se extrajeran fiablemente y se mejoraría la efectividad de los sistemas actuales en adquisición y clasificación.

La mayor parte del trabajo en esta área de sistemas ATR se han realizado usando sensores FLIR, las imágenes FLIR's emiten radiaciones térmicas de los objetos. Generalmente las imágenes exhiben alto contraste y muestran pocas sombras. Para una perspectiva histórica de aplicaciones militares de técnicas infrarrojas refiérase a la siguiente nota [8]. Los temas relacionados al procesamiento de imágenes y arquitectura del reconocimiento de patrones aplicado a esta problemática no se discuten en detalle aquí, para un estudio refiérase a la siguiente nota [9].

Sistemas de Reconocimiento Existentes

Los sistemas de reconocimiento automático con imágenes térmicas, tienen un sin número de aplicaciones y por lo tanto es difícil mencionar y describir cada una de ellas, esta técnica puede detectar la variabilidad de la temperatura contenida en los objetos, entre las áreas principales de aplicación actual de visión computacional más importantes podemos mencionar:

Medicina: Interpretación de imágenes de rayos X, resonancia magnética, navegación automática en endoscopia, etc.

Manufactura: Robots flexibles, control de calidad, etc.

Imágenes aéreas: Análisis de imágenes de satélite, identificación de aviones, etc.

Vehículos autónomos: Vehículos para ambientes peligrosos, exploración espacial, aplicaciones militares, etc.

Texto y dibujos: Reconocimiento de texto escrito, interpretación de planos, etc.

Ejemplos de sistemas específicos:

- Reconocimiento de caracteres invariantes a rotación, traslación y escalamiento.
- Diagnóstico de fallas en radiografías de soldaduras.
- Interpretación de imágenes de endoscopia para navegación automática.

- Exploración en campos minados, para el reconocimiento de minas.
- Visión nocturna.
- Inspección de sistemas de calentamiento y refrigeración.
- Localización de fugas y emisión de gases.

En esta ocasión se han elegido algunos proyectos de la gran variedad de los sistemas existentes, de los cuales la descripción aquí mencionada será en ocasiones muy precisa, de modo que si usted desea ampliar detalles el interesado puede verificar en las referencias que se citan.

Aplicaciones

En la actualidad existen varias aplicaciones de sistema de reconocimiento de objetos en imágenes infrarrojas, a continuación se describen algunas aplicaciones implementadas en este campo:

Detección de objetivos en imágenes co-registradas en el rango visual térmico.

En [2] se analizan imágenes visuales, térmicas y de rango para el reconocimiento de objetivos, la razón del uso de múltiples imágenes es evitar la ambigüedad en la detección y reconocimiento de objetivos, explotando los diferentes tipos de información contenida en cada una de éstas. Las características usadas para distinguir los objetivos del ruido son: textura, brillo, temperatura, superficies planas y altura. Los resultados de detectores individuales son combinados para mejorar la detección. La operación morfológica denominada "erosión" es utilizada como una herramienta para remover información falsa. Los excelentes resultados obtenidos en la detección soportan convenientemente este método para otros problemas ATR. Se analizaron 3 tipos de imágenes (visual, térmica y una imagen de rango), representando tres escenas. La primera escena esta conformada por 2 tanques en un área árida sin vegetación. La segunda escena esta conformada por tres

tanques, incluyendo uno parcialmente ocluido por la vegetación y finalmente la tercera es parecida a la segunda sólo que se le agregó pequeñas construcciones, como edificios, puentes, etc. El proceso de generación en el caso de las imágenes térmicas fue primero generado con el uso de herramientas interactivas para definir los valores de temperatura de cada parte de las imágenes, y después se aplicó filtrado, interpolación y una correlación espacial aleatoria de datos adicional. Algunos objetivos no pudieron ser detectados de imágenes individuales pero en su mayoría fueron correctamente detectados por el sistema. Un total de ocho objetivos de siete fueron detectados correctamente y solo 16 alarmas falsas fueron producidas, casi la mitad de ellas fueron originadas de la escena tres donde había ruido tales como construcciones.

Identificación de objetivos usando momentos de Zernike y Redes Neuronales.

En [3] se desarrolló un algoritmo de identificación de objetivos bajo el agua capaz de identificar objetivos sumergidos en el agua. Algunos de los factores que influyen en el desempeño del sistema son: objetivos que varían en tamaño, forma y reflectividad. Los objetivos varían dependiendo de los entornos donde son colocados, es decir, estos pueden presentarse exaltados o parcialmente ocluidos, y a su vez las propiedades de los ambientes varían significativamente de una locación a otra, en este caso las características del suelo tales como arena, rocas, corales y vegetación pueden encubrir a un objetivo. En ocasiones cuando se trabaja con ruido, éste presenta patrones semejantes a objetivos verdaderos, lo cual ocasiona detectar objetivos falsos.

Se desarrolló y probó un esquema de extracción de características dependiente de la forma e invariante a rotación, tamaño de escala y traslación, propiedades que son extremadamente aplicables para cualquier problema de clasificación de objetivos. El desarrollo del esquema fue probado con un conjunto de imágenes ópticas-electrónicas recopiladas bajo diferentes condiciones ambientales con un fondo variable, tipo de rango y objetivos a reconocer.

El conjunto de datos óptico-electrónico fue recopilado usando un sensor LLS (Laser Line Scan) por el sistema de estación costera (CSS), localizado en la ciudad de Panamá, Florida. El conjunto de datos contiene una colección de imágenes (un total de 104) de 7 pruebas diferentes cubriendo un rango amplio de condiciones ambientales. La propuesta fue determinar la robustez de las características de Zernike y la habilidad de generalización del clasificador BPNN en situaciones difíciles donde está presente considerablemente la distorsión y falta de claridad. En el entrenamiento y prueba de los patrones fueron procesados los momentos de Zernike hasta el orden 7.

El sistema de clasificación de objetivos fue de dos capas BPNN con 18 entradas, 30 capas de neuronas escondidas y 3 salidas.

Las 3 salidas de la red corresponden a las 3 siguientes clases:

- Forma de bala (bullet-shape)
- Forma elíptica
- No Objetivo

La red fue entrenada basándose en los 18 momentos de Zernike extraídos. El número total de épocas entrenadas permitido fue de 10,000. El rango de la variable de entrenamiento fue seleccionado con el valor inicial de $\lambda = 0.1$ y el factor momento fue seleccionado en $\alpha = 0.95$.

También se implementaron 15 entrenamientos de prueba con diferentes pesos iniciales, los patrones del objetivo 1 fueron correctamente identificados, para el objetivo 2 uno de los patrones fue no clasificado como un no objetivo. Las clasificaciones incorrectas se dieron en su mayoría en un pequeño o quizás inadecuado entrenamiento de un conjunto de datos, así como, formas similares de objetivos falsos, particularmente cuando el patrón del objetivo es muy pequeño con distorsión considerable presente en las imágenes. El rango de clasificación correcta global (promedio de la diagonal de los elementos de la matriz de confusión) fue del 91% de aciertos, el cual es muy bueno.

Reconocimiento bayesiano de objetivos por partes en segunda generación de Imágenes FLIRs.

En [1] se presenta un sistema para el reconocimiento de objetivos en segunda generación en imágenes FLIR (Sistema de simulación térmica infrarrojo, donde se crea una imagen por medio del uso de sensores electro-ópticos que transforman la radiación electromagnética en señales). El reconocimiento de objetivos esta basado en la metodología. La metodología es basada en una jerarquía, es decir, una estructura modular para el reconocimiento del objeto.

En forma más general, el nivel más bajo de clasificadores que son entrenados para reconocer las clases de los objetos de entrada, mientras que el siguiente nivel, los clasificadores son entrenados para reconocer objetos específicos. En cada nivel los objetivos son reconocidos por sus partes, y de esa manera cada clasificador consiste de módulos, donde cada uno es un experto en una parte específica del objeto. Cada módulo experto es entrenado para reconocer una parte bajo diferentes ángulos y transformaciones. Una realización Bayesiana de la metodología propuesta es presentada en este trabajo. El cual los módulos expertos representan la probabilidad de las funciones de densidad de cada parte, modelado como una combinación de densidades para incorporar diferentes vistas (aspectos) de cada parte.

El reconocimiento depende de una presentación secuencial de las partes al sistema, sin usar ninguna información relacional entre las partes. En el marco de los resultados experimentales una vez que las partes de los objetivos son identificado bajo el criterio de restricción convexa, cada parte es normalizada para la traslación y rotación representada usando los momentos de Zernike hasta el orden 8 para el reconocimiento. El conjunto de entrenamiento consiste de 6 objetivos de 3 clases (tanques, camiones y APC's transporte blindado personal).

El sistema fue entrenado para reconocer 6 objetivos: camiones M35 y HMMWV, tanques M60, 2S1, 2SU y APC's M730. Para cada objetivo, un total de 72 vistas fueron consideradas (0-360°) con un intervalo de 5° dando un total de 432 imágenes. El conjunto de prueba incluye objetivos que el sistema no entreno para reconocer y algunos fondos (falsas alamas).

El conjunto de prueba fue manualmente dividido en 3 categorías basadas en los resultados de segmentación.

- Buena. 230 imágenes parecía el conjunto de entrenamiento,
- Defectuosa. Donde la segmentación fue pobre (inclula partes del fondo)
- Oclusión. 32 imágenes

El rango de reconocimiento global fue del 90.05% (391/432). Los errores de reconocimiento en la segmentación "Buena" se dieron cuando la vista se aproximaba a 0°.

En el "Defectuoso" se debió a la presencia de objetos en el fondo que se parecen a partes del objetivo.

Reconocimiento Automático de Placas de Matriculación de Vehículos

En esta ocasión hablaremos sobre un caso particular, un sistema que se realizó en el laboratorio de la Facultad de Informática de la Universidad Politécnica de Madrid, el cual se desarrolló para una empresa privada entre los años 1990 y 1991, consistió en un sistema automático para el reconocimiento de matrículas de vehículos industriales, en esta ocasión se describirá de manera superficial el sistema, si el lector está interesado en este tema puede consultar el artículo en [4].

Las escenas manejadas por el sistema han sido tomadas al aire libre, con los serios problemas que conlleva de cambios de intensidad luminosa, brillos deslumbramientos y oscurecimientos inesperados, etc.

Básicamente el proceso está dividido en tres fases claramente diferenciadas: la primera es la segmentación de la placa de la matrícula dentro de la escena.

Esto es, encontrar en la imagen el rectángulo de la matrícula para centrar el resto del trabajo en esa zona y no tener que tratar toda la imagen. La segunda fase es la segmentación de los caracteres de la placa; es decir, encontrar dentro de la zona que la fase anterior identificó como placa de matrícula las figuras que en ella aparecen que corresponden a letras o números. La última fase, es evidente, reconocer estos caracteres para decidir de qué matrícula se trata. Además de la clara secuencialidad hacia delante que existe en estas tres fases, se establece otra en orden inverso. Si al final de cada fase no se obtiene un resultado dentro de los márgenes admisibles, se retrocede para empezar de nuevo.

Este proyecto, totalmente acabado, ha demostrado una alta efectividad en el reconocimiento, obteniéndose cerca de un 90% de aciertos. Esta tasa se acerca al 100% de aciertos para matrículas en buen estado; es decir, sin manchas ni objetos extraños, tales como tornillos gruesos en medio de la placa.

El tiempo de proceso es de 6 segundos. Sin embargo con matrículas en buen estado puede alcanzarse la solución en tan sólo tres segundos, aunque en aquellas que están deformadas, incorrectamente iluminadas o en mal estado se consiguen buenos resultados en al menos 20 veinte segundos.

Estructura de la Tesis

En esta sección de la tesis, se proporciona una perspectiva general de la organización de la tesis, presentando una pequeña descripción de cada uno de los capítulos que componen la tesis.

En el *primer capítulo* se maneja la problemática general de los sistemas de reconocimiento automático existentes, así como las aplicaciones en cada una de las áreas y se describen algunas de las aplicaciones de los sistemas a detalle.

En el *segundo capítulo* se establecen los conceptos básicos se utilizaron para la solución al problema planteado. En este capítulo se describe todo lo referente a la termografía infrarroja, así como, la región del infrarrojo en el espectro electromagnético y la formación de las imágenes infrarrojas. Se detalla como se llevo a cabo el proceso de la adquisición de las imágenes infrarrojas con la cámara térmica PUMA II. Se define el proceso de obtención de los momentos de Zernike en cada uno de los objetivos, así como, la normalización efectuada a estos, antes del cálculo de los momentos.

Finalmente se menciona la descripción de la Red Neuronal y el algoritmo de entrenamiento implementados en este sistema.

En el *tercer capítulo* se enfoca a la propuesta de solución para el reconocimiento automático de objetivos en imágenes infrarrojas. Se explica a detalle el proceso y resultado de cada uno de los módulos de preprocesamiento, segmentación, extracción de características, clasificación y reconocimiento del sistema de visión.

En el *cuarto capítulo* se presentan los resultados obtenidos en los diferentes módulos, con imágenes infrarrojas, así como un apartado donde se muestran bajo que características funciona el sistema.

Y por último, el *quinto capítulo* las conclusiones a las que se llegó en este trabajo de investigación, además del trabajo futuro por desarrollar para el sistema.

Conceptos Básicos

El objetivo de este capítulo es proporcionar al lector los conceptos y definiciones básicas que ayudaron a la solución del problema planteado.

Termografía

Una técnica que le permite a la visión humana, poder observar escenas a distancia, sin contacto alguno y que mida la distribución de la temperatura en la superficie emitida por objetos animados e inanimados con gran precisión, se le llama Termografía o Termograma. De la misma forma se le conoce a la imagen producida por una cámara infrarroja en la cual, para esto se utilizan detectores de infrarrojos, lo que permite extraer una imagen cuantificable en temperatura mediante cálculos [10]. las cámaras miden la temperatura de cualquier objeto o superficie de la imagen y producen una imagen que interpretan el diseño térmico con facilidad. Los instrumentos de detección térmica infrarroja son usados para medir y registrar, sin contacto alguno, las temperaturas contenidas en una diversidad de objetos.

La termografía se extiende más allá de nuestra visión de las longitudes de onda visible, gracias a la innovación de los sistemas de adquisición de imágenes térmicas, se ha podido trabajar en el rango de visión de la porción infrarroja del espectro electromagnético.

La función principal de los sistemas de adquisición de imágenes térmicas es convertir la información obtenida de los cuerpos como consecuencia de su condición térmica a imágenes bidimensionales correspondientes a su misma temperatura.

Cada píxel correspondientes a la imagen o termograma es asociado a un color o nivel de gris, el cual representará una temperatura dada de acuerdo a una escala predeterminada [11]

En la Figura E, en (a) se muestra un ejemplo de una imagen infrarroja RGB (Red-Green-Blue), capturada de la cámara infrarroja y su respectivo histograma, en (b) la imagen correspondiente a una imagen binaria en el cual se aprecia la escala de valores de 0 a 1 y el histograma que relaciona los niveles de gris con un valor de temperatura en la imagen.

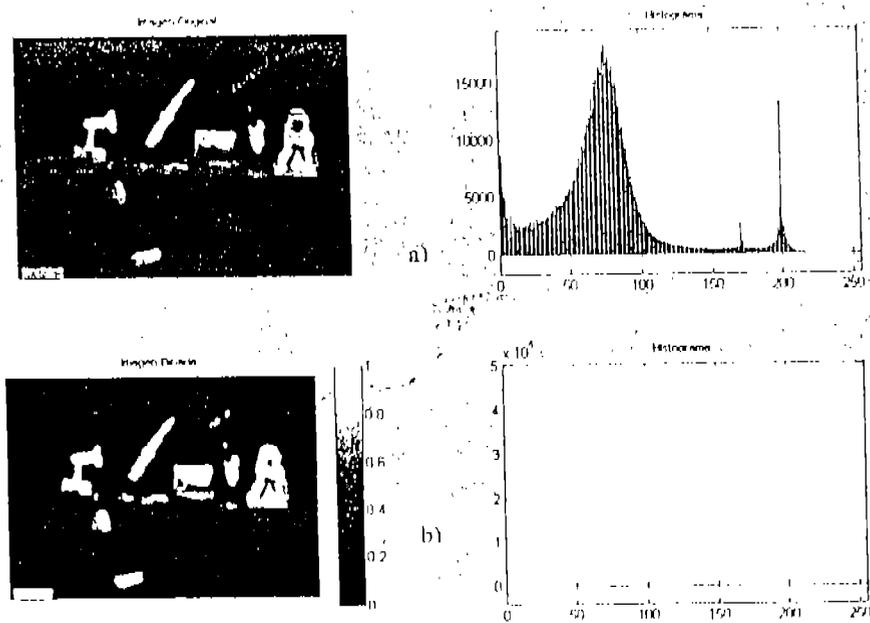


Figura D. Imagen infrarroja de una escena y su proporcional en binaria.

Toda molécula contenida en un objeto, que tenga una temperatura superior al cero absoluto (-273° K) emite rayos infrarrojos y estos serán mayores entre más temperatura tenga el objeto [12].

Las imágenes térmicas también son conocidas como FLIRs (Forward Looking Infrared) estas igual de igual manera hacen un análisis de las temperaturas del objeto rastreado cuantificando sus valores con un alto grado de confiabilidad, adicionalmente se obtienen los diferentes niveles isotérmicos que permiten un análisis cualitativo del objeto en estudio. Una imagen FLIR puede detectar un diferencial de temperatura de hasta 0.18 grados centígrados.

Formación de Imágenes Infrarrojas

Una Imagen Infrarroja (IR) es aquella que es tomada en el canal o banda infrarroja. Las longitudes de onda IR, a las cuales son sensibles los sensores. El canal IR se basa en un principio básico de la física que nos dice:

"Todo cuerpo que se encuentre a una temperatura determinada emite una radiación electromagnética que depende de dicha temperatura de emisión. A esta temperatura se le denomina temperatura de brillo (T_b) del cuerpo emisor."

El Sol emite un T_b equivalente del orden de 6.000 °C. La Tierra, y para un observador en el espacio como es un satélite emite por término medio, como si tuviera una T_b del orden de 17 °C a 20 °C. Lo mismo le sucede al ser humano, por poseer una temperatura dada, nuestro cuerpo emite radiación térmica que nuestros ojos, sensible a la luz visible o "blanca", no pueden ver. Otro principio físico y básico es el siguiente:

"A mayor T_b , mayor es la energía emitida por el cuerpo emisor y mayor es el intervalo de longitudes de onda de emisión".

Esto significa que en un cuerpo como el Sol (a 6.000 °C) emite más energía que la Tierra (entre 17 y 20 °C), como era de esperar, esta energía está repartida en un espectro de longitudes de ondas mayor: El Sol emite energía en longitudes de onda visible o luz blanca, infrarroja, ultravioleta, rayos X, γ etc. Por el contrario, la Tierra emite menos energía y ésta se concentra, sobretudo, en longitudes de onda del IR.

La radiación infrarroja esta localizada entre la región de lo visible y de microondas del espectro electromagnético y abarca la región de 2 a 1000 μm . Esta región es conocida también como región térmica del espectro. La mayor parte del espectro de emisión infrarrojo no es útil para los sensores debido a que la radiación es absorbida por agua y dióxido de carbono en la atmósfera. La banda infrarroja se divide en tres secciones como se aprecia en la Figura F:

- La banda infrarroja de longitudes de onda largas (LWIR) que cubre la región de 8 a 15 μm , con aproximadamente un 100% de transmisión en la banda de 9 a 12 μm . La banda LWIR ofrece excelente visibilidad para la mayoría de los objetos terrestres.
- La banda infrarroja de longitudes de onda medias (MWIR) cubre el rango de 2 a 6 μm .
- La banda infrarroja de longitudes de ondas cortas (SWIR) que cubre el rango de 0.7 a 2 μm .

La región entre 6 y 8 μm no es utilizable en la atmósfera terrestre puesto que la radiación es absorbida por vapor de agua y sólo logra cubrir distancias cortas desde la fuente de emisión. Un micrón o micrómetro (μm), es un millón de metro.

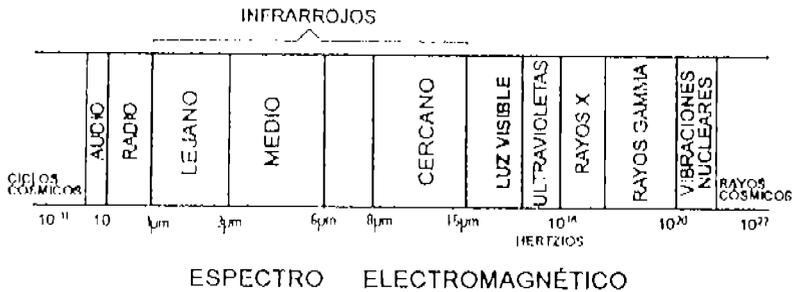


Figura E Espectro Electromagnético

Adquisición y Procesamiento Digital de Imágenes Infrarrojas

Para la adquisición digital de las imágenes infrarrojas se requiere de un dispositivo físico que sea sensible a una banda del espectro de energía electromagnética (visible, infrarrojo, ultravioleta) y que produce una señal eléctrica de salida proporcional al nivel de energía percibido. en este proyecto se utilizó la Cámara Térmica que trabaja en el infrarrojo lejano. Una vez que la cámara adquiere la imagen, la tarjeta de adquisición procesa la señal analógica (señal de video), la cual es llevada a un convertidor analógico-digital, este proceso es ejecutado por un digitalizado (frame grabber). Una señal típica de video o analógica, contiene información de un frame donde cada frame corresponde a una pantalla completa de información visual. En sentido amplio existen tres tipos de tarjetas:

- Digitalizadores de imágenes o frame grabbers.
- Digitalizadores de imágenes con un procesador específico
- Tarjetas de procesamiento paralelo

En este proyecto se utilizó un frame grabber, que muestrean la señal de la cámara y guardan la imagen digital en una zona de memoria, que pueda ser accedida por el microprocesador del computador para su tratamiento. Básicamente están constituidos por los puestos de entrada de la señal de video, y el conversor analógico / digital para obtener la imagen [13].

Una vez capturada la imagen es transmitida al computador. Cada imagen transmitida es almacenada en la memoria RAM, procesada por la CPU a través de un programa de edición de imágenes fotográficas y proyectadas en el monitor del ordenador, éste proceso se ejemplifica en la Figura F.

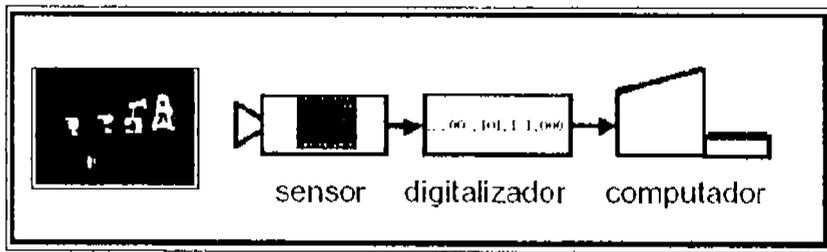


Figura F. Proceso de adquisición de una imagen.

La Cámara

Una cámara térmica infrarroja tiene aspecto semejante a una cámara de video convencional, la diferencia de esta radica en el rango del espectro que trabaja, la cámara que se utilizó funciona en el rango del infrarrojo lejano, es decir lo que se observa a través de ella es radiación térmica de un cuerpo. En este proyecto se trabajó con una cámara térmica infrarroja PUMA II que despliega imágenes de vídeo en tiempo real con una energía radiante dentro del rango espectral de 7 a 14 micrón.

La cámara PUMA II trabaja con una fuente de 12 VDC y consume un máximo de 4.5 watts.

La ganancia del video es automáticamente ajustable basándose en la información que contiene la escena.

Especificaciones Técnicas:

Sensor:	Bario Estroncio Titanate (Barium Strontium Titanate, BST)
Rango Espectral:	7 – 14 micrón,
Resolución:	320 x 240,
Sensibilidad:	0.1°C,
Estabilización de Temperatura:	Sistema de Enfriamiento Termoeléctrico,
Tiempo de encendido:	30 Segundos (Típico),
Control de la Ganancia:	Automático o Manual,
Control de Nivel:	Automático o Manual,
Rango de Frame:	30 Hz.,
Imagen:	Escala de Grises, RGB
Salida de Video:	RS-170,
Voltaje de Operación:	9 a 32 VCD,
Poder:	2.5 Watts (típico) 4.5 Watts (Max.),
Protección de Sobrevoltaje:	-12 a 40 VCD,
Temperatura de Operación:	-40°C a 70°C,
Temperatura de Almacenamiento:	-40°C a 70°C,

Especificaciones Ópticas:

100 mm f/1.0 (con el foco motorizado)	9°H x 6.8° V FOV
IFOV (mr)	0.6
Rango del Foco	15 pies al Infinito
Lentes de tipo montura	Bayoneta Cerrada

Filtros Morfológicos

La aplicación de transformaciones morfológicas a la imagen, es con el fin de modificar la estructura o forma de los objetos que están presentes en la imagen, esta herramienta además de ser útil para la extracción de características, permiten la eliminación de ruido y falsas alarmas que se producen en todo el proceso de segmentación. La morfología matemática esta basada en la teoría de conjuntos y en la topología [13]. Para efectos de este proyecto se implementó una erosión seguida de una dilatación mejor conocida como *Apertura*.

Erosión

La erosión es la degradación progresiva de uno de los campos (0 o 1). Es decir un elemento del campo a degradar seguirá perteneciendo al mismo si está rodeado de elementos iguales a ella. En caso contrario, pasará al otro campo, este proceso iterativo terminaría por destruir la imagen. Matemáticamente se expresará de la siguiente manera: si se toma el elemento estructural ver Figura G, simétrico respecto al origen de B , $\overset{\vee}{B}$, la erosión de un conjunto X respecto al elemento B es:

$$X \ominus \overset{\vee}{B} = \{x \mid B_x \subset X\}$$

Dependiendo del elemento estructurante, la erosión será más pronunciada o no.

1	1	1
1	1	1
1	1	1

Figura G. Elemento estructurante aplicado al proceso de la operación morfológica de apertura.

Dilatación

En el caso de la dilatación es el crecimiento progresivo de uno de los campos (0 o 1). Un elemento del campo contrario a crecer será convertido si posee algún vecino perteneciente al campo que se expandiona. En caso contrario, permanecerá igual. Los elementos pertenecientes al campo a expandionar evidentemente no se modifican. Si se aplica un número elevado de veces terminaría por destruir la imagen, ya que todos los píxeles estarían a nivel alto. La dilatación consistiría en pasar el elemento estructural por el borde del objeto, pasando a valer la unidad todos aquellos puntos por los que pase, matemáticamente la dilatación es la operación dual de la erosión:

$$X \oplus \check{B} = \left(X^c - \check{B} \right)^c$$

En la Figura H., se muestra un ejemplo de las operaciones morfológicas de erosión y dilatación.

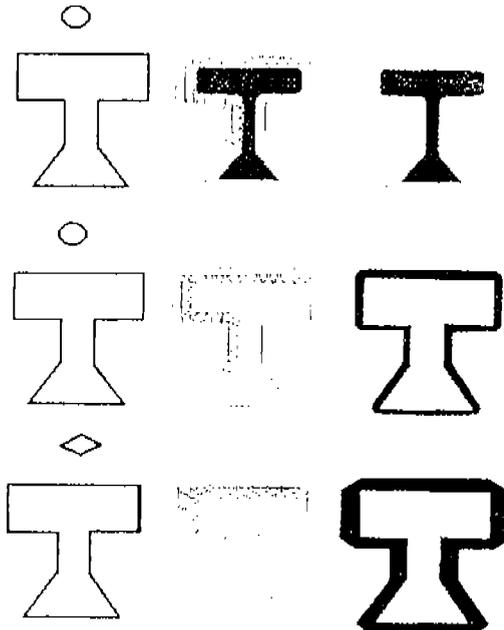


Figura H. Ejemplos de operadores morfológicos de erosión y dilatación.

Apertura

Las dos transformaciones antes vistas no son conmutativas, después de una erosión no se puede recuperar la imagen original con una dilatación ya que se habrán perdido pequeños detalles. Sin embargo esto mismo tiene una gran utilidad ya que la información que se mantiene corresponderá a las partes más importantes de los objetos y además se habrá eliminado pequeños debidos al ruido. La apertura (opening conocida en la literatura anglosajona) se definirá entonces como una combinación de erosiones y dilataciones en las primero es una erosión y la última una dilatación siempre con el mismo elemento estructural, normalmente se le denomina apertura ya que al empezar por una erosión se tiende a romper las piezas en sus partes constitutivas.

$$X \circ B = \left(X \ominus B \right) \oplus B$$

En la Figura G, se puede ver como funciona la apertura suavizando los contornos del objeto, rompe enlaces delgados, y elimina pequeñas protusiones.



Figura I. Ejemplo gráfico de un operador de apertura.

Descriptor de formas basadas en momentos

Las funciones momentos de la distribución de intensidad de una imagen son usadas para caracterizar varias características de la forma de una imagen. Las funciones momento comúnmente usadas son los momentos geométricos momentos complejos, momentos Legendre y momentos de Zernike. Los momentos también han sido usados ampliamente en el reconocimiento de patrones y aplicaciones de identificación de objetivos, como descriptores de características de una imagen que son invariantes bajo traslación rotación y variaciones de escala [14][15][16].

En este proyecto se trabajó con imágenes binarias, las cuales para aplicaciones de este proyecto f va ser una imagen binaria y $S = \{(x, y) | f(x, y) = 1\}$ la representación bidimensional de una imagen. Entonces el (j, k) -ésimo momento digital de S está dado por:

$$M_{jk}(S) = \sum_{(x,y) \in S} x^j y^k \quad \text{ec. 1.1}$$

Donde $M_{00}(S)$ representa el área de la imagen binaria.

El centro de gravedad (\bar{x}, \bar{y}) de S , esta expresado en función de algunos momentos, como lo muestra la ecuación 1.2.

$$x = \frac{M_{10}(S)}{M_{00}(S)} \quad y = \frac{M_{01}(S)}{M_{00}(S)} \quad \text{ec. 1.2}$$

Usando el centro de gravedad, es posible definir el (j, k) -ésimo momento central de S , como:

$$\mu_{jk}(S) = \sum_{(x,y) \in S} (x - \bar{x})^j (y - \bar{y})^k \quad \text{ec. 1.3}$$

Los momentos centrales son invariantes a traslaciones. Si

$$S^* = \{(x^*, y^*) \mid x^* = x + a, y^* = y + b, (x, y) \in S\}$$

Momentos de Geométricos

Los momentos básicos requeridos para este sistema fueron los momentos geométricos, estos han sido utilizados para normalizar la región contra variantes de traslación y el cálculo de M_{00} para determinar el factor de escala β (ver momentos de Zernike), antes del cálculo de momentos de Zernike, los momentos geométricos de orden $(p + q)$ son definidos como:

$$M_{pq} = \sum_{y=1}^m \sum_{x=1}^n x^p y^q I(x, y)$$

ec. 1.4

Donde $I(x, y)$ es el valor de intensidad del píxel (x, y) , y $p, q = 0, 1, 2, \dots$.

Si $I(x, y) = 1$ para todas las x, y y si la ecuación 1.4 es correctamente escalada entonces M_{00} da el área (en píxeles) del patrón de interés de la imagen 2D y M_{10}, M_{01} da las coordenadas del centro de gravedad del patrón como en la ecuación 1.2., entonces los momentos Geométricos en este proyecto se utilizaron para el cálculo del centroide y M_{00} para el factor de escala.

Desafortunadamente, mientras la función básica (x^p, y^q) es completada, esta no es ortogonal.

Esto significa que es ineficiente para usar momentos Geométricos para el representación de patrones porque ellos contienen información redundante [17][18].

Momentos de Zernike

La evaluación de los momentos de Zernike es computacionalmente más compleja comparada con los momentos Geométricos y Legendre. La evaluación de los momentos integrales para una imagen binaria puede ser más rápida utilizando sólo los puntos del borde.

Los métodos de Integración del contorno para un rápido procesamiento de momentos Geométricos en una imagen binaria usando sólo la información del borde se ha analizado ya en la siguiente literatura [17][18][19].

La característica de ortogonalidad en los momentos de Zernike, ayuda a obtener características de forma que representen independencia de las características geométricas, conduce a una mínima cantidad de información redundante en un conjunto de momentos de Zernike. El alto grado de información contenida facilita la reconstrucción de la imagen de un conjunto finito de momentos, usando el teorema de expansión de Fourier en funciones ortogonales, para estos fines no se llevó a cabo la reconstrucción de la Imagen.

Los momentos de Zernike también han mostrado ser robustos a la presencia de ruido y cuantización en una imagen y han mostrado ser mejores que los momentos geométricos en varias aplicaciones de reconocimiento de patrones [20][21].

Los momentos de Zernike están definidos en las funciones ortogonales de los polinomios radiales de Zernike [22][23]. Los momentos de Zernike son momentos ortogonales complejos cuya magnitud tiene una propiedad invariante a rotación. Las características principales [3][24] de los momentos de Zernike con respecto al desempeño de otros momentos son:

- La capacidad de representación de una imagen
- Aspectos de redundancia de información
- Susceptibilidad al ruido en la imagen

Los momentos de Zernike ecuación 1.5, son formas derivadas de un conjunto complejo de polinomios, el cual forman un conjunto ortogonal dentro de un círculo unitario $x^2 + y^2 = 1$ de la forma:

$$Z_{mn}(x, y) = V_{mn}(\rho, \theta) = R_{mn} \exp(jm\theta) \quad \text{ec. 1.5}$$

Donde:

n = entero positivo o negativo $n - |m| \quad |m| \leq n$

ρ = tamaño del vector del origen a (x, y)

θ = ángulo entre el vector ρ y eje x

$R_{mn}(\rho)$ = polinomio radial el cual es definido

$$R_{mn}(\rho) = \sum_{s=0}^{n-|m|/2} (-1)^s \frac{(n-s)!}{s! \left(\frac{n+|m|}{2} - s\right)! \left(\frac{n-|m|}{2} - s\right)!} \rho^{n-2s} \quad \text{ec. 1.6}$$

$$R(\rho) = \left\{ R_{mn}(\rho) \mid n = 0, 1, 2, \dots, \infty, \quad |m| \leq n, \quad y \quad n - |m| \text{ es par} \right\} \quad \text{ec. 1.7}$$

Entonces los momentos de Zernike de orden n con repeticiones m , en una imagen continua de dos dimensiones es definida como en la ecuación 1.8, pero para cuestiones de imágenes digitales la integral es reemplazada por la sumatoria como se aprecia en la ecuación 1.9.

$$A_{nm} = \frac{n+1}{\pi} \iint_{x^2+y^2 \leq 1} I(x, y) Z_{nm}^*(\rho, \theta) dx dy \quad \text{ec. 1.8}$$

$$A_{nm} = \frac{n+1}{\pi} \sum_x \sum_y I(x, y) Y_{nm}^*(\rho, \theta) \quad x^2 + y^2 \leq 1 \quad \text{ec. 1.9}$$

Donde A_{nm} representa los momentos asociados con la imagen.

La magnitud de los momentos de Zernike es por lo tanto invariante a la rotación. Los momentos ortogonales incluyendo los momentos de Zernike, representan formas más estrechamente que otros tipos de momentos. Se ha demostrado que los momentos de Zernike representan a las curvas más exactamente que otros momentos [1].

Los momentos de Zernike son rotaciones invariantes pero no invariantes de traslación y escala. Para esto se utilizan los momentos geométricos como se analizó en la sección anterior (momentos Geométricos), la invariante de traslación es lograda por el movimiento del origen de la imagen al centroide del objetivo.

Este es realizado usando momentos de primer orden M_{10} y M_{01} de una imagen $f(x, y)$,

donde
$$m_{pq} = \sum_x \sum_y x^p y^q f(x, y)$$

El centroide de la imagen está dado por $\bar{x} = (m_{10} / m_{00})$ y $\bar{y} = (m_{01} / m_{00})$. Es logrado transformando la imagen $f(x, y)$ a otra imagen $f_1(x, y) = f(x + \bar{x}, y + \bar{y})$. Después la imagen es centralizada alrededor de su centroide. La imagen es normalizada contra los

cambios de escala usando $a = \sqrt{\frac{\beta}{m_{00}}}$, donde β es una constante real positiva dada y

m_{00} es el área de la imagen. Finalmente la imagen obtenida de todo este proceso es

$$f_1(x, y) = f\left(\frac{x}{a} + \bar{x}, \frac{y}{a} + \bar{y}\right)$$

Como en muchas ocasiones la imagen obtenida de todo el proceso normalmente no coincide con la imagen original esta se ajusta por medio de la interpolación del vecino más cercano.

Esto es por que si:

- $m_{00} > \beta$ La imagen la reduce.
- $m_{00} < \beta$ La imagen la agranda.
- $m_{00} = \beta$ La imagen no cambia.

El cálculo se realizo para el orden 8, obteniendo un vector de 25 características pero normalizado a magnitud y ángulo obtenemos un vector de 50 características, pero en este caso el momento A_{00} es el mismo para todos y el momento A_{11} es cero[1][25], solo tomamos en cuenta los momentos del orden 2 al 8 para una discriminación relevante.

Redes Neuronales

Las redes neuronales artificiales emulan las redes neuronales biológicas y se utilizan para aprender estrategias de control observando la forma como una persona lo hace. Son sistemas que aprenden con ejemplos, es decir, que no requieren que la tarea a ejecutar se programe. También se utilizan para aprender a controlar un sistema aplicando entradas, evaluando la calidad de las respuestas para estas entradas, y ajustando la fórmula que genera las acciones de control para mejorar la respuesta del sistema.

Las redes neuronales artificiales han demostrado su capacidad para resolver problemas donde el cómputo tradicional ha tenido gran dificultad durante años.

Algunas aplicaciones exitosas de las redes neuronales son:

- Procesamiento de imagen y voz.
- Reconocimiento de patrones.
- Planeamiento, predicción, control y optimización de procesos, etc.

El análisis realizado a diferentes arquitecturas, dio como resultado la elección del perceptron multicapa conocida también como *BackPropagation* (en referencia al algoritmo más popular de entrenamiento), que por sus características propias, es la red más empleada para la clasificación de patrones, adaptándose por lo tanto a las condiciones de reconocimiento de este tipo.

Estas particularidades la hacen sobresalir por que representan series de datos complejos como los momentos de Zernike, apoyándose en los niveles de neuronas ocultas que la habilitan para aprender progresivamente las propiedades representativas de los vectores de características, que al ser ingresados paralelamente proporcionan la capacidad de realizar el proceso de reconocimiento prácticamente en tiempo real.

En la siguiente figura J, se muestra a la neurona como es activada, por la fuerza combinada de las señales de entrada siendo estas superior a un cierto nivel, en el caso general el valor de activación de la neurona viene dado por una función de activación $f(.)$.

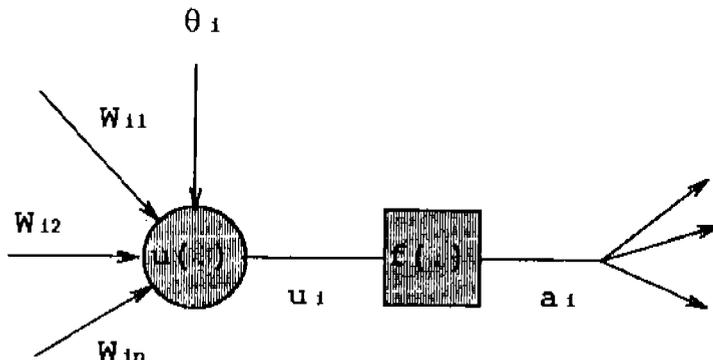


Figura J. Arquitectura básica de una Red Neuronal.

Función de Activación (Función de neurona)

El valor de red, expresado por la función de base, $u(w, x)$, será inmediatamente transformada por una función de activación no lineal. Por ejemplo, las funciones de activación más comunes son la función Paso, Rampa o Sigmoidal y Gaussiana.

En este proyecto se utilizó la función Sigmoidal en particular, la cual está dada por la siguiente ecuación.

$$f(u_i) = \frac{1}{1 + e^{-u_i/\sigma}}$$

Desarrollo e Implementación del Sistema

Metodología

Planteadas la problemática general de los sistemas de reconocimiento automático de objetivos, en el capítulo uno, el objetivo principal para este proyecto es desarrollar un sistema de reconocimiento automático de objetivos en imágenes infrarrojas bajo invariantes de traslación, rotación y escala.

Se realizaron varios prototipos con diversas imágenes en las cuales más adelante hablaremos de cada uno de ellos.

El sistema propuesto realizaría las siguientes disposiciones:

- Obtener las imágenes infrarrojas de los diversos objetivos a reconocer,
- Realizar el proceso de filtrado de las imágenes para disminuir la presencia de ruido,
- Segmentar los posibles objetivos a candidatos potenciales,
- Calcular el centroide de cada una de las regiones detectadas,
- Normalización de la región contra escala y traslación,
- Extraer características de los objetivos normalizados,
- Almacenar la información en una base de datos para la fase de clasificación,
- Entrenar esa información hasta que los pesos de la red se ajusten,

- Probar todo el proceso con objetivos que no fueron utilizados en la fase de entrenamiento,
- Evaluar resultados con base en la base de datos y las regiones que no fueron utilizadas para el entrenamiento.

El sistema cuenta con una variedad de imágenes, los cuales sirvieron para la realización de este proyecto, en la Figura K., se muestra 5 imágenes de toda una colección de imágenes tomadas de diferentes distancia como 5, 6, 7, 9 y 10 metros, en el cual se puede apreciar el objetivo principal a reconocer en cada una de las imágenes y como las regiones a reconocer van perdiendo la consistencia en cuanto a forma, conforme se van alejando de la cámara.

Las imágenes fueron tomadas en un ambiente estático, no cambiante, el cual nos proporcionaba las mejores condiciones para las tomas, estas imágenes fueron tomadas en un horario donde la luz de día era nula.

Los objetos contenidos en las imágenes consisten básicamente en fuentes de energía, tales como lámparas, reguladores, cautines, no breaks, etc., objetos que emanan suficiente cantidad de calor para ser detectada por una cámara infrarroja.

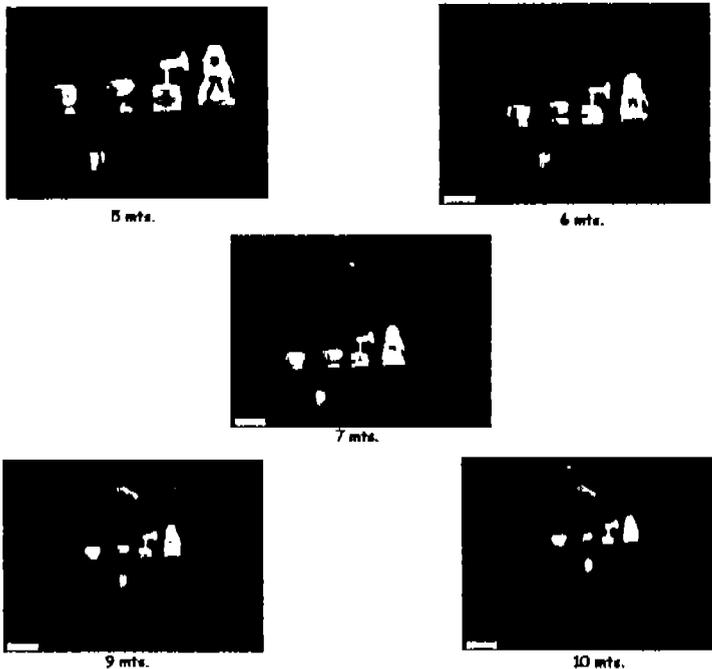


Figura K. Ejemplo de Imágenes de diferentes distancias

Los objetos contenidos en las imágenes presentan diversas vistas, es decir, estos se muestran en las imágenes con sus diferentes vistas, cada uno de los objetos son rotados de manera manual para su posterior captura.

Para la realización de este proyecto se utilizaron dos colecciones de imágenes, la primera colección de imágenes contiene imágenes con objetos en cuatro vistas (frente-atrás-derecho-izquierdo); la segunda colección de imágenes contiene imágenes con objetos con todas las vistas posibles de 0 a 360°, se dice posibles por que los objetos fueron manipulados de forma manual y experimental, ocasionando en varios casos la obtención de una cantidad total de vistas diferentes por cada objeto.

En la Figura L, se presentan algunos ejemplos de imágenes con objetos que se muestran en sus cuatro diferentes vistas, estas imágenes fueron tomadas de una distancia de 5 metros, y pertenecen a la colección de imágenes recopiladas de cuatro vistas (frente-atrás-derecho-izquierdo) de enfoque regulado, en la siguiente Figura 8 observe las cuatro diferentes vistas del objeto-lámpara, en la primera imagen se aprecia el objeto-lámpara visto de frente, en la imagen derecha superior se aprecia al objeto-lámpara visto por la parte de atrás, en la imagen inferior izquierda se observa al objeto-lámpara su vista derecha y por último en la imagen inferior derecha se aprecia al objeto-lámpara su vista izquierda, en general en la Figura 8 se pueden apreciar sus cuatro diferentes vistas de un solo objetivo.

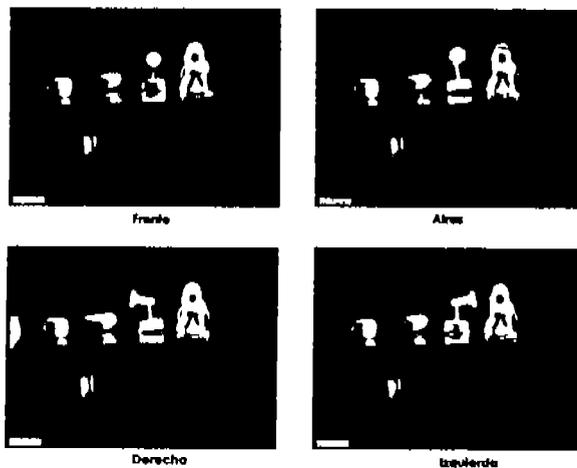


Figura L. Ejemplo de las diferentes vistas de un objeto.

En la Figura M, se muestra un esquema general del funcionamiento del sistema mostrando en cada una de las etapas el avance del sistema.

A continuación en la página 45 se describe el desarrollo detallado efectuado en cada una de las etapas del Sistema de Reconocimiento de Objetivos implementado en este proyecto.

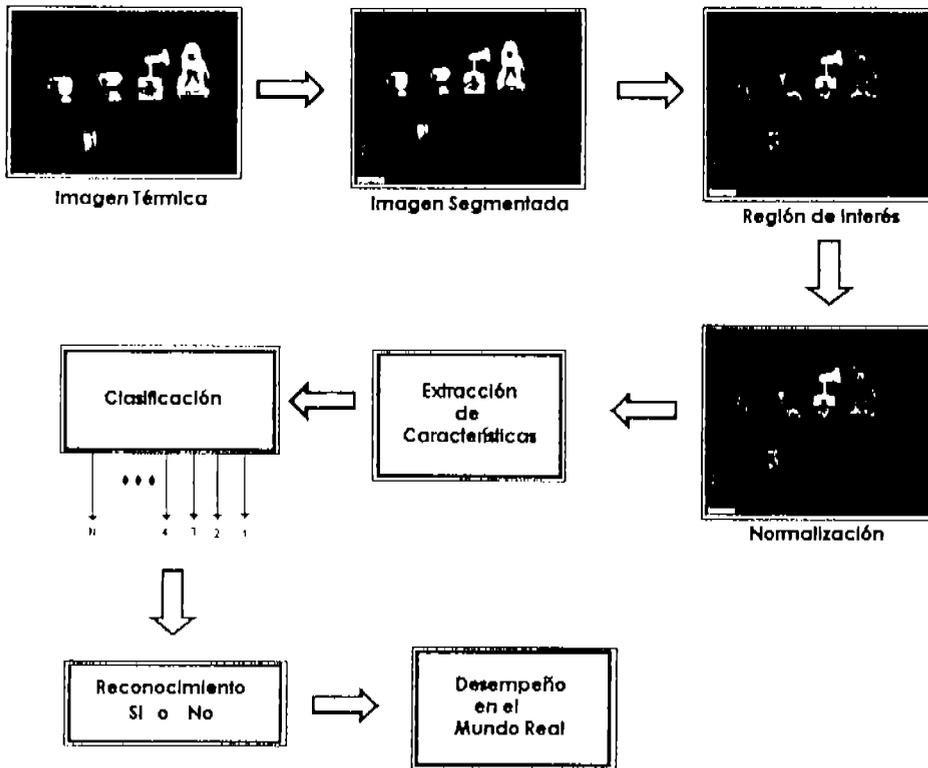


Figura M. Funcionamiento General del Sistema.

Etapa de Preprocesamiento

Una vez que la imagen se ha adquirido y digitalizado el siguiente paso es mejorar la calidad de la imagen bajo los criterios subjetivos del ojo humano en la etapa de preprocesamiento para sistemas de reconocimiento de patrones. El objetivo final es que tenga mayores posibilidades de éxito para procesos posteriores.

El preprocesamiento trata normalmente de técnicas que mejoren el contraste, suprimir ruido y aislar regiones que por sus características indiquen la posibilidad de información del objetivo deseado.

En el proceso de generación de imágenes, algunos ruidos son introducidos en las imágenes. Generalmente, los píxeles con ruido aparecen como puntos con niveles de gris bien diferentes de su vecindad (oscuros -negros- o saturados -blancos-). Estos puntos pueden aparecer distribuidos aleatoriamente o en forma sistemática (fajas verticales y horizontales). Las causas de dicho ruido pueden ser fallas de los detectores, limitaciones del sistema electrónico del sensor, entre otros.

No existe una regla (receta) que mejor se aplique para mejorar la calidad de una imagen, ya que depende de las características de la escena; época de adquisición, ángulo de iluminación, altura del sensor y bandas. En el caso de un aumento de contraste no revelará nunca una información nueva que no esté contenida en la imagen. El contraste solamente presentará la misma información existente en los datos brutos, pero de una forma más clara para el usuario

En este caso la imagen RGB (Red-Green-Blue) adquirida por la cámara la convertimos en una imagen binaria para tener una mejor manipulación sobre los objetivos como se vio en el capítulo 2 en el tema de filtros morfológicos.

Etapa de Segmentación

El siguiente paso es la segmentación, definida en un sentido amplio, su objetivo es dividir la imagen en las partes que la constituyen o los objetos que la forman. En general la segmentación autónoma es uno de los problemas más difíciles en el procesamiento de la imagen. Por una parte, una buena segmentación nos facilitará mucho la solución del problema, por otra parte la segmentación errónea nos conducirá al fallo.

En este trabajo de reconocimiento, el objetivo de la segmentación es determinar la presencia de todas las posibles regiones a ser objetivos candidatos en una imagen infrarroja con ruido original de la escena. Para este caso todas las imágenes son transformadas en imágenes binarias (como se menciona en la Etapa de Procesamiento), con un umbral específico aplicado a la intensidad de los píxeles que conforman la imagen. Se maneja imágenes binarias con el fin de eliminar ruido de manera más abrupta a un umbral determinado y así evitar la primera etapa y obtener la forma de las regiones reales, es decir, con el ruido incorporado que provee una imagen en un entorno real.

Discriminación de Falsos Objetivos

En este caso aplicamos un operador morfológico a la imagen binaria Figura N, llamada apertura binaria (Binary Opening), cuya función principal es aplicar una erosión seguida de la dilatación a cada una de las regiones de la imagen binaria, usando el método de vecino más cercano, esto es con el fin de eliminar falsas alarmas [23] [24] [25], obteniendo una imagen con menor cantidad de ruido como se aprecia en la Figura Ñ.

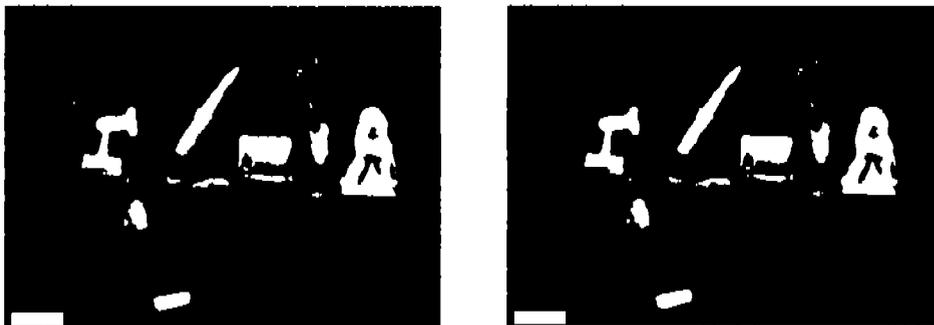


Figura N. Imagen Binaria (lado izquierdo) e Imagen Filtrada (lado derecho).

Ya localizados los objetivos potenciales dentro de la imagen y algunos píxeles pertenecientes a cada una de estas que han sido correctamente identificados. El siguiente paso será crecer las regiones, donde los otros píxeles que pertenecen a las regiones deberán ser identificados para completar la silueta de las regiones. Algunas técnicas de crecimiento de regiones examinan píxeles vecinos donde los píxeles con similar propiedades son puestos en la misma región.

Esta es una técnica simple y es de uso limitado donde la información contiene ruido y esto hace un problema complejo para mayor información refiérase al segundo capítulo.

Los límites de las regiones pueden también ser difíciles de encontrar el más destacado para crecerlo innecesariamente. Otra técnica examina las propiedades de los píxeles vecinos bajo consideraciones.

Esto reduce los problemas experimentados por la examinación de un solo píxel para extender la certeza pero estos métodos tal vez no produzcan bordes suavizados. La imagen obtenida del proceso anterior nos proporciona una imagen con menor cantidad de ruido, el siguiente paso es crecer las regiones para tratar de obtener su tamaño y forma original, por lo tanto, se procede aplicar a la imagen un operador morfológico de dilatación [26] [27] con un elemento estructural de la siguiente forma:

$$\begin{matrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{matrix}$$

Aplicado el elemento estructurante a en toda la imagen obtenemos la siguiente imagen binaria dilatada como se muestra en la Figura O.



Figura O. Imagen Binaria Dilatada

Para obtener la dimensión y forma real de las regiones en forma de bordes realizamos una resta de imágenes, es decir, la Imagen Dilatada menos la Imagen Filtrada y así obtenemos los bordes de cada una de las regiones.

La extracción de bordes es el método más empleado en la segmentación de imágenes, la efectividad desde el punto de vista del reconocimiento de objetos, de una operación cuyo resultado sea la obtención de los bordes de un objeto, cual es la segmentación, es evidente, puesto que la mayor cantidad de información se encuentra en los bordes.

De hecho, la mayoría de los objetos pueden reconocerse a partir de su forma, que esta fijada por sus bordes. En la Figura P., se muestra el resultado de la resta, en esta imagen se aprecia el borde de todas las regiones detectadas, entre ellas el objetivo potencial y ruido.

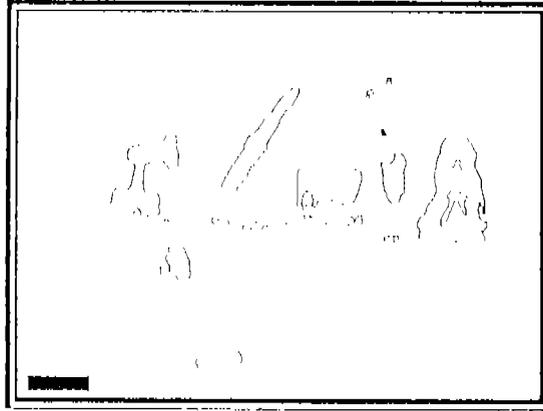


Figura 9. Imagen de Extracción de Bordes

Detectadas y seleccionadas cada una de las regiones mayores a un arreglo de 12×15 , extraemos características de cada una de ellas, es decir, calculamos los momentos de Zernike y/o los momentos Geométricos a cada una de las regiones identificadas en la imagen esto se verá más a detalle en la siguiente etapa.

Etapa de Extracción

En la etapa de segmentación se apartan las regiones de interés de todo el entorno de la imagen original, a cada una de estas regiones de interés se le realiza el calculo de características a través de momentos de Zernike. Los momentos permiten manipular la forma de los contornos de las regiones de interés de forma cuantitativamente.

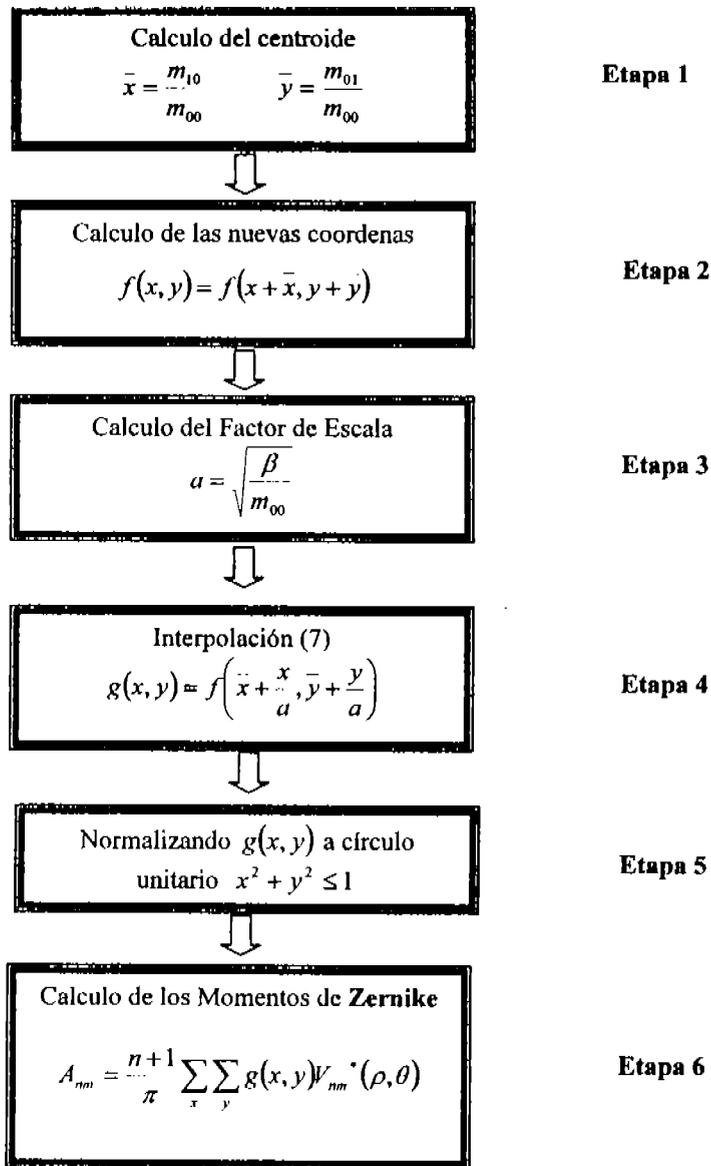


Figura Q. Algoritmo de Extracción y Normalización de Características.

En la Figura Q, presentamos el algoritmo implementado para la obtención y normalización de características implementado en este proyecto.

En la Etapa 1 se realiza el cálculo del centro de masa de la región de interés, para mayor información de la Etapa 1 refiérase al segundo capítulo, en la Etapa 2 se realiza el cálculo de las nuevas coordenada con base al centroide y de esta forma que sean invariantes a traslación.

Uno de los problemas que afectan al buen desempeño de un sistema ATR son los cambios de escala, para esto en la Etapa 3 se aplica el factor de escala para convertir todas las regiones de interés invariante a cualquier cambio de tamaño, para mayor información de la Etapa 2 y 3 refiérase al segundo capítulo.

En la Etapa 4 se realiza interpolación del vecino más cercano de las coordenadas originales, las nuevas coordenadas y el factor de escala, obteniendo una nueva función llamada $g(x, y)$.

Las coordenadas finales son normalizadas a círculo unitario es decir que los valores obtenidos van a estar en un rango de 0 ~ 1, esto lo podemos apreciar en la Etapa 5. Y finalmente en la Etapa 6 se realiza el cálculo de momento de Zemike, para mayor información de todo el procedimiento refiérase al capítulo 2.

Etapa de Clasificación

El reconocimiento en general es el proceso de encontrar una correspondencia entre ciertas características en la imagen y características similares de un modelo del objetivo [28]. La cuestión más importante involucrada en este proceso es:

- a) Identificar el tipo de características a usar,
- b) Determinar el mejor procedimiento para establecer correspondencia entre la imagen y las características del modelo.

La consistencia y eficiencia del sistema de reconocimiento de objetivos depende directamente en que tan cuidadosamente son manejadas estas características. En este sistema el reconocimiento seguirá una metodología de abajo-hacia-arriba, es decir las características extraídas de una imagen son clasificadas en uno de varios tipos de objetivos.

En este estudio se procesa cada imagen hasta el orden 8, en la Tabla 1, se muestra la lista de los momentos de Zernike y sus correspondientes características. Cada imagen es normalizada para la escala y traslación antes de procesar los momentos de Zernike, el momento $|A_{00}|$ permanece constante y el momento $|A_{11}|$ es cero.

Solo los momentos de orden 2-8 son relevantes para la discriminación. Cada objetivo fue representada por un vector de 20 puntos característicos, que finalmente terminaron siendo 40, por la normalización implementada para el manejo de las características.

Para mayor información de este tema refiérase al segundo capítulo.

Orden	Momentos	No. de Momentos
0	A00	1
1	A11	1
2	A20,A21	2
3	A31,A33	2
4	A40,A42 A44	3
5	A51,A53,A55	3
6	A60,A62 A64,A66	4
7	A71,A73 A75,A77	4
8	A80,A82,A83,A84,A86,A88	5

Tabla 1. Lista los Momentos de Zemike

Resultados del Sistema

Características

En este capítulo se muestran los resultados obtenidos de los distintos algoritmos implementados en este proyecto en el cual se evaluará la eficiencia de los algoritmos propuestos con base en la problemática y el tipo de imágenes, se muestra la matriz de confusión así como el porcentaje de clasificación correcta entre otros.

Algoritmo 1. Momentos Geométricos para Imágenes Sintéticas con objetivos de varias vistas.

Para este experimento se utilizó una colección de autos de 360 imágenes RGB, las cuales son tratadas antes del cálculo de características para convertirlas en imágenes binarias, 5 autos diferentes se manejaron en este experimento ver Figura R; donde cada auto contiene 72 imágenes, 36 son de alta resolución y las otras 36 restantes son de baja resolución, cada imagen representa una vista del auto cada 10 grados, formando un recorrido del auto de 0 a 360 grados, ver Tabla 2.

El primer conjunto de 36 imágenes corresponde al mismo conjunto de las 36 imágenes restantes pero a menor escala. Las características extraídas de cada una de estas regiones son normalizadas a magnitud y ángulo, en la cual se muestran los coeficientes del orden 0 al orden 8, cabe mencionar que los coeficientes A00 y A11 son discriminados para el entrenamiento, como se discutió en la sección de extracción de características.

Auto	Resolución Alta	Resolución Baja	Total	Total de Imágenes
HMVV	36	36	72	
M1	36	36	72	
M2	36	36	72	
TRUCK	36	36	72	
TRACTOR	36	36	72	360

Tabla 2. Relación de Imágenes.

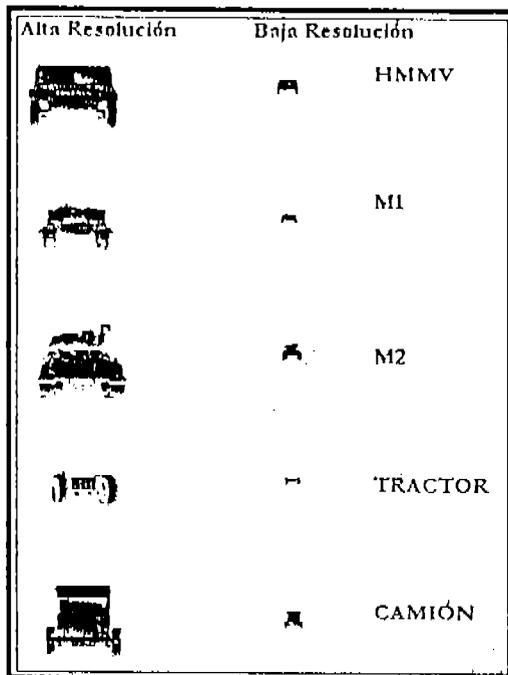


Figura R. Ejemplo de las imágenes utilizadas

Algoritmo 2. Momentos geométricos para imágenes infrarrojas con objetivos de una sola vista.

Para la implementación de este algoritmo de procesamiento rápido, con momentos geométricos invariantes a escala, traslación y rotación, se utilizaron 5 objetos de una sola vista y a diferentes escalas, a continuación se muestran algunos ejemplos de ellos a continuación:



Sin embargo cabe resaltar que estas aproximaciones no obtienen descriptores que son invariantes bajo transformaciones afines en general. Un conjunto completo de cuatro momentos invariantes afine derivadas de los momentos del segundo y tercer orden fueron implementadas, los detalles y discusiones de este algoritmo pueden ser encontrados en [29]. Todos los momentos característicos son dependientes de las transformaciones lineales de los niveles de grises de las regiones, para describir las propiedades de la forma de la región, se trabajo con datos de imágenes binarias $[f(i,j)=1]$ y la dependencia de las transformaciones de niveles de grises desaparece.

$$I1 = \frac{((m20*m02)-(m11^2))}{(m00^4);} \quad (3.1)$$

$$I2 = \frac{((m30^2) * (m03^2) - 8 * m30 * m21 * m12 * m03 + 4 * (m30 * (m12^3) + (m21^3) * m03) - 3 * (m21^2) * (m12^2))}{(m00^{10})};$$

(3.2)

$$I3 = \frac{((m20 * (m21 * m03 - (m12^2))) - m11 * (m30 * m03 - m21 * m12) + m02 * (m30 * m12 - (m21^2)))}{(m00^7)};$$

(3.3)

$$I4 = \frac{((m20^3) * (m03^2) - 6 * (m20^2) * m11 * m12 * m03 - 6 * (m20^2) * m02 * m21 * m03 + 9 * (m20^2) * m02 * (m12^2) + 12 * m20 * (m11^2) * m21 * m03 + 6 * m20 * m11 * m02 * m30 * m03 - 18 * m20 * m11 * m02 * m21 * m12 - 8 * (m11^3) * m30 * m03 - 6 * m20 * (m02^2) * m30 * m12 + 9 * m20 * (m02^2) * (m21^2) + 12 * (m11^2) * m02 * m30 * m12 - 6 * m11 * (m02^2) * m30 * m21 + (m02^3) * (m30^2))}{(m00^{11})};$$

(3.4)

Se implementaron las cuatro características anteriores (3.1) (3.2) (3.3) (3.4), pero sin normalizar las regiones obteniendo los siguientes resultados.

En la fase de Entrenamiento fue entrenada para reconocer 5 objetivos diversos, se utilizaron

- 80 ejemplos para el entrenamiento y,
- 55 ejemplos para la fase de prueba.

La arquitectura propuesta se muestra en la Figura 5, la cual consiste de 3 capas: la primera capa de entrada, la cual consta de 4 nodos de entrada, una capa escondida o intermedia la cual tiene 4 nodos y una capa de salida, con 5 nodos de salida.

La función de transferencia utilizada es la sigmoide, y el error RMS para el conjunto de entrenamiento fue de 0.19632.

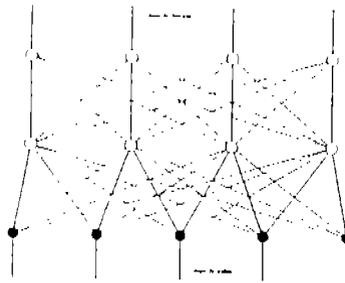


Figura 5. Arquitectura de la Red para los momentos Geométricos no Normalizados

Algoritmo 3. Momentos de Zorniko para Imágenes Infrarrojas de objetivos para una sola vista.

Entrenamiento

En la fase de Entrenamiento igual que en el proyecto anterior fue entrenada para reconocer 5 objetivos diversos, se utilizaron 80 ejemplos para el entrenamiento. En la Figura H, se muestra la grafica del error RMS para el conjunto de entrenamiento este fue de 0.00004

Pruoba

En esta fase se utilizaron 55 ejemplos diferentes a los ejemplos de entrenamiento de igual manera se reconocieron 5 objetivos diversos.

Algoritmo 4. Momentos de Zernike para imágenes infrarrojas con objetivos de varias vistas.

La teoría de los momentos proporciona una interesante y útil alternativa para la representación de formas de objetos. Una ventaja de los momentos es su gran facilidad de cálculo, sobre todo aplicándolos a imágenes binarias, en el caso de los momentos de Zernike el tiempo de proceso es mayor que cualquier otro, pero son más robustos al ruido en la imagen y de mayor representación de características de cada región.

Cada una de las regiones para obtener sus respectivas características, en este caso los momentos implementados fueron los momentos geométricos y momentos de Zernike, este último invariante a rotación, traslación y escala. Para esta implementación se utilizaron imágenes binarias, por el fácil manejo de aplicar operaciones en estas y por que en esta aplicación disminuye la presencia de falsas alarmas en las imágenes.

En este experimento se utilizó un total de 605 imágenes algunos ejemplos de estas se observan en la Figura T, 455 regiones se utilizaron para la parte de entrenamiento y 150 regiones para la fase de prueba ver Tabla 3, todas las imágenes utilizadas en este experimento fueron tratadas con operadores morfológicos para manipularlas como regiones binarias. Igual que en el experimento anterior se manejaron 5 objetos térmicos diferentes, a diferentes distancias, 5, 6,7, 8 y 9 metros, y con todas las vistas posibles de cada uno de estos en un giro total de 0 a 360 grados.

ENTRENAMIENTO				PRUEBA			
Región	Clase	Total	Total de imágenes	Región	Clase	Total	Total de imágenes
Virgen	1	87	455	Virgen	1	47	148
Lámpara	2	140		Lámpara	2	42	
Regulador-Cautín	3	99		Regulador-Cautín	3	23	
Cautín	4	63		Cautín	4	22	
Eliminador	5	65		Eliminador	5	14	

Tabla 3. Relación de Regiones de Entrenamiento y Prueba.

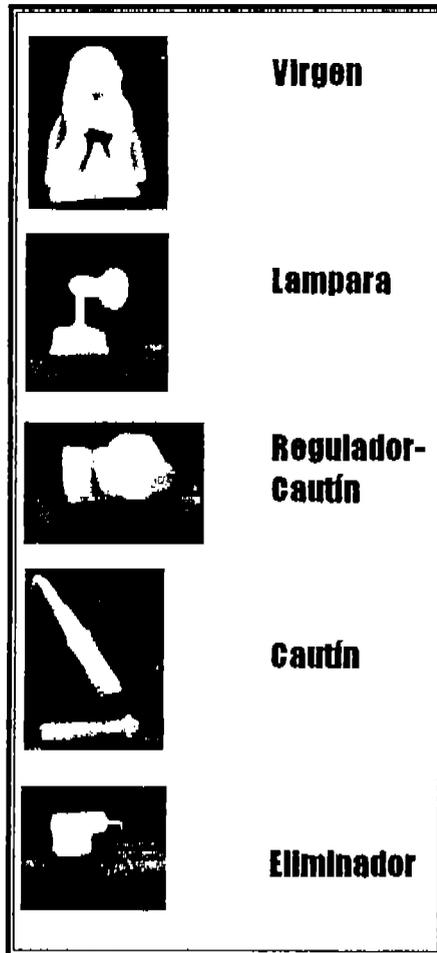


Figura T. Ejemplo de las imágenes infrarrojas (regiones).

Para efectos de este proyecto se utilizó una red neuronal feed forward con el algoritmo de entrenamiento Back propagación para los datos originales, la arquitectura de la red contempla 3 capas, capa de entrada, capa intermedia y capa de salida, la cual en la capa de entrada maneja 46 atributos de entrada y en la capa de salida maneja 6 nodos de salida, la función de transferencia utilizada fue la tangente sigmoideal (-1, 1), se utilizó la técnica de validación k-fold –cross validation con una $k = 10$ para el entrenamiento.

La arquitectura propuesta se muestra en la Figura U, la cual consiste de 3 capas:

- Una capa de entrada, la cual tiene 40 nodos de entrada
- Una capa escondida, la cual tiene 40 nodos intermedios
- Una capa de salida, con 5 nodos de salida.

La función de transferencia utilizada es la *sigmoide*.

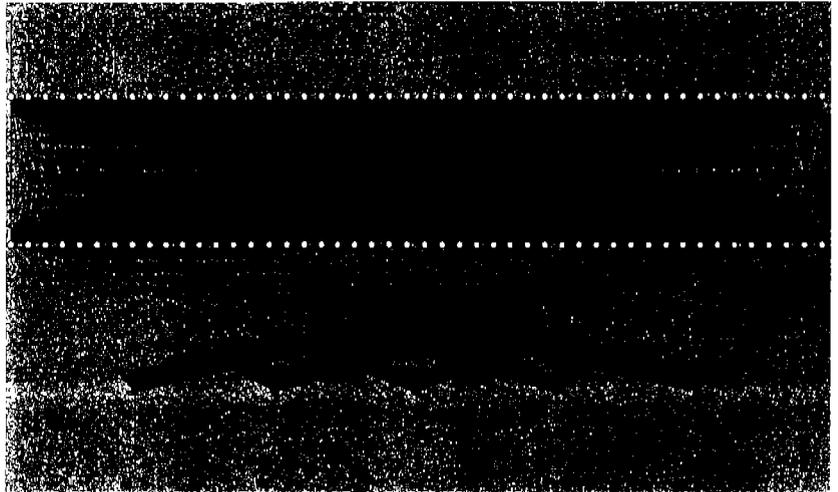


Figura U. Arquitectura de la Red para los momentos de Zemike

Resultados

Algoritmo 1. Imágenes Sintéticas varias vistas - Momentos Geométricos

Los resultados del algoritmo 1, se muestran en la Tabla 4, esta contiene la matriz de confusión la cual para la clase 1 reconoció 62 de 72, la clase 2 reconoció 57 de 72 ejemplos, para la clase 3 reconoció 71 de 72 ejemplos, para la clase 4 reconoció 68 de 72 ejemplos y para la clase 5 reconoció 67 de 72 ejemplos.

Datos

Ejemplos	Clases	Atributos	RMS					CC	EP
360	5	46	0.241	0.2449	0.1635	0.0579	0.1500	90.2778%	0.0972

RMS: Error Raíz Cuadrática Media por nodo de salida.

CC: Porcentaje de Clasificación Correcta.

EP: Error de Predicción.

Red Neuronal

NC	Objetivo	Épocas	FT
3	0.05	8000	Tansig

FT: Función de Transferencia

NC: Numero de Capas

	1	2	3	4	5
1	62	4	4	0	2
2	10	57	5	0	0
3	1	0	71	0	0
4	0	0	2	68	2
5	3	2	0	0	67

Tabla 4. Matriz de confusión con momentos geométricos para imágenes sintéticas.

El error de predicción de la red fue calculado se la siguiente manera:

$$Error = 1 - \left(\frac{Num_Clasificados_Correctamente}{Total_Ejemplos} \right) = 1 - \left(\frac{325}{360} \right) = 0.0972$$

Algoritmo 2. Imágenes Infrarrojas de una sola vista - Momentos Geométricos

Los resultados de este algoritmo se muestran en la Tabla 5, en este se observa la **matriz de confusión** para el algoritmo 2, para la clase 1 reconoció 0 ejemplos de 12, para la clase 2 reconoció 9 ejemplos de 10, en la clase 3 reconoció 5 de 9, en la clase 4 reconoció 5 de 8, y en la clase 5 reconoció 10 de 13.

	1	2	3	4	5
1	0	0	5	4	3
2	0	9	0	1	0
3	0	0	5	4	0
4	0	0	0	5	3
5	0	0	0	3	10

Tabla 5. Matriz de confusión con momentos geométricos para imágenes Infrarrojas.

Clasificación correcta = 55.7692%

El error fue calculado se la siguiente manera:

$$Error = 1 - \left(\frac{Num_Clasificados_Correctamente}{Total_Ejemplos} \right) = 1 - \left(\frac{29}{52} \right) = 0.442307692307692$$

Algoritmo 3. Imágenes Infrarrojas de una sola vista - Momentos de Zernike

En la Tabla 6, se muestra la matriz de confusión para esta prueba, para la clase 1 reconoció 12 ejemplos de 13, para la clase 2 reconoció 10 ejemplos de 10, en la clase 3 reconoció 4 de 9, en la clase 4 reconoció 6 de 9, y en la clase 5 reconoció 11 de 14.

	1	2	3	4	5
1	12	0	0	0	1
2	0	10	0	0	0
3	0	0	4	5	0
4	0	0	3	6	0
5	0	0	2	1	11

Tabla 6. Matriz de confusión con momentos de Zernike para imágenes Infrarrojas de una sola vista.

Clasificación correcta = 78.18%

El error fue calculado se la siguiente manera:

$$Error = 1 - \left(\frac{Num_Clasificados_Correctamente}{Total_Ejemplos} \right) = 1 - \left(\frac{43}{55} \right) = 0.2181818181818181$$

Algoritmo 4. Imágenes Infrarrojas de varias vistas - Momentos de Zernike

Este algoritmo consta de la fase de entrenamiento y de prueba. Los resultados obtenidos en cada una de estas son mostrados a continuación.

Entrenamiento

Datos

Ejemplos	Clases	Atributos	RMS				CC	EP	
455	5	48	0.066	0.117	0.2066	0.0752	0.2379	93.1868%	0.0681

RMS: Error, Raíz Cuadrática Medio por cada nodo de salida.

CC: Porcentaje de Clasificación Correcta.

EP: Error de Predicción.

Red Neuronal

NC	Objetivo	Épocas	FT
3	0.05	10000	tansig

FT: Función de Transferencia

NC: Numero de Capas

En la Tabla 7, se muestra la matriz de confusión para la fase de entrenamiento, en el cual para la clase 1 reconoció 87 ejemplos de 87, para la clase 2 reconoció 137 ejemplos de 140, en la clase 3 reconoció 94 de 99, en la clase 4 reconoció 60 de 63, y en la clase 5 reconoció 46 de 66.

	1	2	3	4	5
1	87	0	0	0	0
2	0	137	0	0	3
3	0	0	94	0	5
4	1	2	0	60	0
5	1	0	18	1	46

Tabla 7. Matriz de confusión con momentos de Zernike para imágenes infrarrojas de varias vistas.

El error de predicción de la red fue calculado se la siguiente manera:

$$Error = 1 - \left(\frac{Num_Clasificados_Correctamente}{Total_Ejemplos} \right) = 1 - \left(\frac{424}{455} \right) = 0.0681$$

Prueba

Datos

Ejemplos	Clases	Atributos	CC
148	5	46	93.9189%

RMS: Error Raíz Cuadrática Media por nodo de salida.

CC: Porcentaje de Clasificación Correcta.

EP: Error de Predicción.

Red Neuronal

NC	Objetivo	Épocas	FT
3	0.05	1000	tansig

FT: Función de Transferencia

NC: Numero de Capas

En la Tabla 8, se muestra la **matriz de confusión** para la fase de prueba, en la clase 1 reconoció 44 ejemplos de 47, para la clase 2 reconoció 42 ejemplos de 42, en la clase 3 reconoció 19 de 23, en la clase 4 reconoció 21 de 22, y en la clase 5 reconoció 13 de 14.

	1	2	3	4	5
1	44	2	1	0	0
2	0	42	0	0	0
3	0	0	19	0	4
4	0	1	0	21	0
5	0	0	1	0	13

Tabla 8. Matriz de confusión con momentos de Zernike para la fase de prueba.

El error de predicción de la red fue calculado de la siguiente manera:

$$Error = 1 - \left(\frac{Num_Clasificados_Correctamente}{Total_Ejemplos} \right) = 1 - \left(\frac{139}{148} \right) = 0.0608$$

Conclusiones

El análisis y procesamiento de imágenes infrarrojas es una tarea no trivial debido principalmente a las siguientes características: el ruido asociado que existe en la imagen, contraste de los objetos detectados y su morfología, ya que ésta depende en algunas ocasiones a la cantidad de calor que el objeto emite.

Esto origina que los objetos de interés no sean identificados con claridad. Además, las condiciones del medio ambiente tales como la temperatura y humedad, deben ser consideradas cuando se desea trabajar con imágenes infrarrojas.

Mediante la técnica de Zernike, la extracción se puede obtener información acerca del objetivo a reconocer, el cual fue posicionado en cuatro diferentes vistas derecho, frente, atrás e izquierdo.

Se demostró que los momentos de Zernike representan formas más exactas que otros momentos y el ruido no es un factor potencial para el perjudicar el reconocimiento, que son invariantes a rotación, traslación y escala en comparación que los momentos Geométricos, como se pudo apreciar en las gráficas anteriores el margen de error entre estos momentos

es considerable, cabe destacar que las características obtenidas por los momentos de Zemike hace que el procesamiento computacional sea más complejo que los momentos Geométricos.

El umbral con el cual se segmentaron los objetivos estáticos fue determinado de forma fija, en base a la observación hecha a diferentes imágenes filtradas.

El procesamiento de imágenes en combinación con los momentos de Zemike proporciona resultados satisfactorios para la extracción de características en imágenes Infrarrojas, por lo tanto la hipótesis es válida.

Trabajo Futuro

En el proceso de desarrollo de los algoritmos y su implementación se descubrieron algunos puntos que no están contemplados en la solución que se propone al problema planteado. Sin embargo esta sección muestra algunas ideas que pueden solucionar estos problemas que se presentan con el sistema.

1. Uno de los problemas importantes que se presentan con un sistema de detección y seguimiento de objetos es la oclusión de los objetos en la escena. Como se menciona antes, la oclusión de un objeto puede ser parcial o total.

El sistema desarrollado no soluciona la oclusión parcial ni total. Una alternativa que puede ayudar con este problema es la restricción convexa, la cual se realiza agrupando segmentos de bordes cuyas tangentes son comunes a los puntos de esquina común en un ángulo menor de 180 grados cuando es vista por dentro del objeto, y después trabajar de manera independiente con cada uno de estos segmentos para mayor información de esta técnica ver la referencia [1].

2. El sistema desarrollado fue diseñado para una cámara sin movimiento, sin embargo, este sistema se puede evolucionar para una cámara con movimiento.

3. El sistema actualmente es capaz de detectar, localizar y reconocer objetivos estáticos, pero si el entorno del objetivo cambia este puede ocasionar que las condiciones del ambiente cambien y a su vez que ocasione errores en el reconocimiento ya que el sistema esta dado para un entorno fijo, por lo tanto como trabajo futuro se considera realizar el reconocimiento en ambientes variables.

4. Aplicar la técnica de Análisis de Componentes Vectoriales de Inteligencia Artificial para la reducción de características de los momentos generados.

Bibliografía

- [1] D. Nair, J.K. Aggarwal, Bayesian recognition of targets by parts in second generation forward looking infrared images, *Image and Vision Computing* 18 (2000), pp 849-864.

- [2] Jorge E. Pérez-Jacome and Vijay K. Madisetti, Target Detection from Coregistered Visual-Thermal-Range Images, center for signal and Image Processing (CSIP), Georgia Institute of Technology –ECE, Atlanta, GA 30332, USA. IEEE 1997, pages 2741-2744.

- [3] Mahmood R. Azimi-Sadjadi, Arta A. Jamshidi and Andrew Nevos, Target identification using Zernike moments and Neural Networks, *Automatic Target Recognition XI*, Firooz A. Sadjadi, Editor Proceedings of SPIE, Vol. 4379 , 2001, pages 136-143.

- [4] D. Maravall, J. Bueno, L. Baumela, J. Ortuño. Reconocimiento Automático de Placas de Matriculación de Vehículos. *Actas del V Simposium Nacional sobre Reconocimiento de Formas y Análisis de Imágenes*. Valencia, 1992 (pp. 222-228).

- [5] Walker, P.F., Smart weapons in naval warfare. *Scientific American*, 248 (May 1983), 53-61.

- [6] Barbe, D.F., VHSIC system and technology, *IEEE Computer* (Feb. 1981) 13-22

- [7] Bhanu. B, Evaluation of automatic target recognition algorithms. *Architecture and algorithms for digital image processing*. In *Processing of Spie*, 435 (Aug. 1983), pp. 18-27.

- [8] Proceedings of The IRE, Special segue on "Infrared Technology". (Sept. 1959).
- [9] IEEE Computer, (Special issue on "Computer Architectures for image Processing")
Jan. 1983.
- [10] Gary L. Orlove, "Infrared Imaging and Radiometry", Optical Inspection and Test,
Critical Reviews, Vol. CR46, 1995.
- [11] G.C.M. Meijer and A. W. Hearwaarden, "Thermal Sensors", Institute of Physics
Publishing Bristol and Philadelphia, 1994.
- [12] Astronomía Autodidacta, La luz y el espectro electromagnético, 2001. Carlos Andrés
Carvajal MD., Astrónomo Aficionado ASASAC.
- [13] Arturo de la Escalera, Visión por computador, Fundamentos y Métodos, Prentice
Hall, Cáp. 2.
- [14] Prokop R.J and Reeves A.P., A survey of moment based techniques for unoccluded
object representation, Graphical Models and Image Processing- CVGIP, Vol 54, No. 5,
(1992), pp 438-460.
- [15] Abu Mostafa Y.S, Recognitive aspects of moment invariants, IEEE Trans. On Patt
Anal and Mach Intell, Vol 6, No 6 (1984), pp 698-706.
- [16] Teague M.R.,Image analysis via the general theory of moments, Journal of the
Optical Society of America, Vol. 70, No. 8 (1980), pp 920-930.
- [17] Jiang X. Y, Simple and fast computation of moments, Pattern Recognition, Vol 24, No
8 (1991), pp 801-806.

- [18] Leu J.G, Computing a shape's moments from its boundary, *Pattern Recognition*, Vol 24, No 10 (1991), pp 949-957.
- [19] Li B.C, Fast computation of moment invariants, *Pattern Recognition*, Vol 24, No 8 (1991), pp 807-813.
- [20] The C.H and Chin R.T., On Image analysis by the method of moments, *IEEE Trans. On Patt Anal and Mach Intell*, Vol 10, No 4 (1988), pp 496-513.
- [21] Belkasim S. O, Pattern recognition with moment invariants –A comparative study and new results, *Pattern Recognition*, Vol 24, No 12 (1991), pp 1117-1138.
- [22] C-H. The and R.T. Chin, Image analysis by the methods of moments, *IEEE transactions on Pattern Analysis and Machine Intelligence*, Vol 10, No 4, (1988), pp 496-513.
- [23] Jan Flusser, Fast calculation of Geometric moments of binary images, Institute of Information Theory and Automation, Academy of Sciences of the Czech Republic, e-mail: flusser@utia.cas.cz, 1997.
- [24] Y.S. Kim, W.Y. Kim, Content trademark retrieval system using a visually salient feature, *Image and Vision Computing* 16 (1998) 931-939.
- [25] Yaser S. Abu-Mostafa, Demetri Psaltis, Image Normalization by Complex Moments, *IEEE transactions on Pattern Analysis and Machine Intelligence*, Vol PAMI-7, No 1, (1985), pp 46-55.

- [26] Roth M. W., Survey on neural network technology for automatic target recognition, IEEE Trans. Neural Networks, Vol. 1 (1990), pp. 28-43.
- [27] Bhanu Bir., Automatic target recognition: state of the art survey, IEEE Trans. Aerospace Electr. Syst., Vol 22 (1986) pp. 364 - 379.
- [28] Pau, L.F. and Nahas, M.Y. An Introduction to Infrared Image Acquisition and Classification Systems, England, Research Studies Press LTD(1983).
- [29] J. Flusser and T. Suk Pattern recognition by affine moment invariants, Pattern Recognition, Vol. 26, (1993), pp 167-174.