



UNIVERSIDAD POLITÉCNICA DE PUEBLA



Programa Académico de Posgrado

**Sistema de reconocimiento
automático de logotipos comerciales
en imágenes**

TESIS
QUE PARA OBTENER EL GRADO DE
**MAESTRÍA EN INGENIERÍA EN SISTEMAS Y
CÓMPUTO INTELIGENTE**

PRESENTA:
Miguel Ángel Medina Nieto

DIRECTOR
Dr. Jorge de la Calleja Mora



Juan C. Bonilla, Puebla, México, Octubre 2016.

El presente trabajo fue realizado gracias a las facilidades otorgadas por el Centro de Ingeniería del Instituto Nacional de Astrofísica, Óptica y Electrónica, ubicado en Calle Luis Enrique Erro No. 1, Santa María Tonantzintla, San Andrés Cholula, Puebla CP. 72840.

La presente tesis titulada “**SISTEMA DE RECONOCIMIENTO AUTOMÁTICO DE LOGOTIPOS COMERCIALES EN IMAGENES**” y realizada por el Lic. Miguel Ángel Medina Nieto, ha sido revisada y aprobada por el Jurado para obtener el Título de:

MAESTRO EN INGENIERÍA EN SISTEMAS Y CÓMPUTO INTELIGENTE
UNIVERSIDAD POLITÉCNICA DE PUEBLA
PNPC-CONACYT

Jurado integrado por:

Profesor

Firma

Director: Dr. Jorge de la Calleja Mora

Revisor: Dr.

Revisor: M.

Revisor: Dr.

Juan C. Bonilla, Puebla, México, Octubre 2016.

gradecimientos

Tabla de contenido

TABLA DE CONTENIDO	V
ÍNDICE DE TABLAS	VII
ÍNDICE DE FIGURAS	IX
NOMENCLATURA	X
Resumen	xi
Abstract	xi
CAPÍTULO 1 PLANTEAMIENTO DEL PROBLEMA DE INVESTIGACIÓN..	1
1.1 Introducción	1
1.2 Objetivos	3
1.2.1 Objetivo general	3
1.2.2 Objetivos específicos.....	3
1.3 Justificación	4
CAPÍTULO 2 MARCO TEÓRICO	5
2.1 Visión por computadora	6
2.1.1 Sistemas de visión por computadora	7
2.1.2 Procesamiento de imágenes	7
2.2 Reconocimiento de objetos en imágenes	8
2.2.1 Clasificación de objetos.....	8
2.2.2 Adaboost.....	9
2.2.3 Local Binary Patterns	11
2.3 Publicidad móvil	12
2.3.1 Realidad aumentada aplicada a la publicidad.....	13
2.4 Herramientas de desarrollo	14
2.4.1 OpenCV	14
2.4.2 Unity.....	15
2.5 Trabajos relacionados	16
CAPÍTULO 3 METODOLOGÍA	22
3.1 Metodología empleada	22
3.1.1 Descripción del sistema.....	22
3.1.2 Variables consideradas	23

3.1.3 Instrumentos de recabado de datos y análisis	23
3.2 Evaluación de las técnicas de visión por computadora	23
3.3 Generación del reconocedor	24
3.3.1 Logotipos que se identifican	25
3.3.2 Generación de los clasificadores.....	26
3.3.2.1 Entrenamiento de un clasificador.....	27
3.3.2.2 Prueba de un clasificador.....	29
3.3.2.3 Experimentos de prueba	31
3.4 Escenario de pruebas	32
3.4.1 Integración y generación de contenido 3D	33
3.4.2 Integración de imágenes del conjunto de pruebas.....	33
3.4.3 Implementación del recorrido inmersivo.....	33
3.4.4 Implementación del reconocedor	33
3.4.5 Evaluación del sistema.....	33
3.5 Implementación en un dispositivo móvil.....	35
CAPÍTULO 4 RESULTADOS.....	36
4.1 Evaluación de las técnicas de visión por computadora	36
4.2 Generación del reconocedor	37
4.2.1 Logotipos que se identifican	37
4.2.2 Generación de los clasificadores.....	40
4.3 Escenario de pruebas	51
4.3.1 Integración y generación de contenido 3D	51
4.3.2 Integración de imágenes del conjunto de pruebas.....	52
4.3.3 Implementación del recorrido inmersivo.....	52
4.3.4 Implementación del reconocedor	53
4.3.5 Evaluación del sistema en el escenario de pruebas	54
4.4 Implementación en un dispositivo móvil.....	56
4.4.1 Evaluación del sistema en el dispositivo móvil.....	58
CAPÍTULO 5 CONCLUSIONES Y PERSPECTIVAS.....	60
5.1 Conclusiones	60
5.1.1 Publicaciones.....	61
5.2 Perspectivas	61
REFERENCIAS BIBLIOGRÁFICAS	63
ANEXO	69
A.1 Resultados de los experimentos de detección de los clasificadores	69
A.2 Ejecución de los experimentos de detección de logotipos	80

A.3 Código de implementación del reconocedor	80
--	----

Índice de Tablas

Tabla 2.5.1. Análisis de los datos obtenidos de los trabajos relacionados. Se presenta el número de clases y la exactitud lograda (en porcentaje)	20
Tabla 2.5.1. Variables consideradas	23
Tabla 3.3.1.1. Parámetros para la creación de ejemplos positivos	27
Tabla 3.3.2.5. Parámetros para el entrenamiento de cada clasificador.....	28
Tabla 3.3.2.2.1. Matriz de confusión	29
Tabla 3.4.5.1. Heurísticas de Nielsen (Nielsen, 1995)	34
Tabla 3.4.5.2. Heurísticas adicionales para aplicaciones de realidad aumentada (de Paiva Guimaraes & Farinazzo Martins, 2014).....	34
Tabla 3.4.5.3. Escala de Likert empleada en la evaluación heurística del sistema... 34	
Tabla 4.1.1. Comparación de resultados obtenidos para determinar las técnicas de visión por computadora a emplear en este trabajo. Exactitud (en porcentaje) y tiempo de detección (en milisegundos).....	36
Tabla 3.3.1.1. Lista de los centros comerciales más valorados de México de acuerdo con (Hispanic Searching, 2015)	37
Tabla 4.2.1.1. Tiendas de los centros comerciales populares en México y su número de aparición.....	38
Tabla 4.2.2.1. Resultados de los experimentos de detección del clasificador Ferrioni basado en la técnica <i>Haar</i> . Exactitud (en porcentaje) y tiempo de detección (en milisegundos)	41
Tabla 4.2.2.2. Resultados de los experimentos de detección del clasificador Ferrioni basado en la técnica <i>LBP</i> . Exactitud (en porcentaje) y tiempo de detección (en milisegundos).....	41
Tabla 4.2.2.3. Comparación de la exactitud obtenida de los clasificadores entre las técnicas <i>Haar</i> y <i>LBP</i> . Exactitud (en porcentaje) y tiempo de detección (en milisegundos).....	42
Tabla 4.2.2.4. Comparación de la exactitud lograda (en porcentaje) en este trabajo respecto a otros trabajos.....	43
Tabla 4.2.2.5. Comparación de la exactitud obtenida con peso sobre el tiempo de detección entre las técnicas <i>Haar</i> y <i>LBP</i> a partir de los resultados de los experimentos. Exactitud (en porcentaje) y tiempo de detección (en milisegundos)	45
Tabla 4.2.2.6. Comparación de los logotipos identifica cada técnica	46
Tabla 4.2.2.7. Comparación de las técnicas <i>Haar</i> y <i>LBP</i> sobre el conjunto de pruebas para cada logotipo.....	46
Tabla 4.3.5.1. Lista de validación utilizada	54
Tabla 4.3.5.2. Evaluación de los expertos en el uso del sistema en el escenario de pruebas	55
Tabla 4.4.1.1. Evaluación de los expertos en el uso del sistema en el dispositivo móvil.....	58
Tabla A.1.1. Resultados del clasificador Adidas basado en <i>Haar</i>	69
Tabla A.1.2. Resultados del clasificador Adidas basado en <i>LBP</i>	69

Tabla A.1.3. Resultados del clasificador Burger King basado en <i>Haar</i> .	70
Tabla A.1.4. Resultados del clasificador Burger King basado en <i>LBP</i> .	70
Tabla A.1.5. Resultados del clasificador Cinépolis basado en <i>Haar</i> .	70
Tabla A.1.6. Resultados del clasificador Cinépolis basado en <i>LBP</i> .	70
Tabla A.1.7. Resultados del clasificador Devlyn basado en <i>Haar</i> .	71
Tabla A.1.8. Resultados del clasificador Devlyn basado en <i>LBP</i> .	71
Tabla A.1.9. Resultados del clasificador Ferrioni basado en <i>Haar</i> .	71
Tabla A.1.10. Resultados del clasificador Ferrioni basado en <i>LBP</i> .	71
Tabla A.1.11. Resultados del clasificador HSBC basado en <i>Haar</i> .	72
Tabla A.1.12. Resultados del clasificador HSBC basado en <i>LBP</i> .	72
Tabla A.1.13. Resultados del clasificador Liverpool basado en <i>Haar</i> .	72
Tabla A.1.14. Resultados del clasificador Liverpool basado en <i>LBP</i> .	73
Tabla A.1.15. Resultados del clasificador Martí basado en <i>Haar</i> .	73
Tabla A.1.16. Resultados del clasificador Martí basado en <i>LBP</i> .	73
Tabla A.1.17. Resultados del clasificador Nutrisa basado en <i>Haar</i> .	73
Tabla A.1.18. Resultados del clasificador Nutrisa basado en <i>LBP</i> .	74
Tabla A.1.19. Resultados del clasificador Puma basado en <i>Haar</i> .	74
Tabla A.1.20. Resultados del clasificador Puma basado en <i>LBP</i> .	74
Tabla A.1.21. Resultados del clasificador Radioshack basado en <i>Haar</i> .	74
Tabla A.1.22. Resultados del clasificador Radioshack basado en <i>LBP</i> .	75
Tabla A.1.23. Resultados del clasificador Santander basado en <i>Haar</i> .	75
Tabla A.1.24. Resultados del clasificador Santander basado en <i>LBP</i> .	75
Tabla A.1.25. Resultados del clasificador Starbucks basado en <i>Haar</i> .	76
Tabla A.1.26. Resultados del clasificador Starbucks basado en <i>LBP</i> .	76
Tabla A.1.27. Resultados del clasificador Steren basado en <i>Haar</i> .	76
Tabla A.1.28. Resultados del clasificador Steren basado en <i>LBP</i> .	76
Tabla A.1.29. Resultados del clasificador Subway basado en <i>Haar</i> .	77
Tabla A.1.30. Resultados del clasificador Subway basado en <i>LBP</i> .	77
Tabla A.1.31. Resultados del clasificador Taco Inn basado en <i>Haar</i> .	77
Tabla A.1.32. Resultados del clasificador Taco Inn basado en <i>LBP</i> .	77
Tabla A.1.33. Resultados del clasificador Telcel basado en <i>Haar</i> .	78
Tabla A.1.34. Resultados del clasificador Telcel basado en <i>LBP</i> .	78
Tabla A.1.35. Resultados del clasificador Todomoda basado en <i>Haar</i> .	78
Tabla A.1.36. Resultados del clasificador Todomoda basado en <i>LBP</i> .	78
Tabla A.1.37. Resultados del clasificador Tous basado en <i>Haar</i> .	79
Tabla A.1.38. Resultados del clasificador Tous basado en <i>LBP</i> .	79
Tabla A.1.39. Resultados del clasificador Vans basado en <i>Haar</i> .	79
Tabla A.1.40. Resultados del clasificador Vans basado en <i>LBP</i> .	79

Índice de Figuras

Figura 2.1. Tópicos que se comprenden en este trabajo.	5
Figura 2.2.2.1. <i>Haar-like Features</i>	10
Figura 2.2.2.2. Reconocimiento de rostros con <i>Haar-like Features</i> . Recuperado de (Ikeuchi, 2014).	11
Figura 2.2.3.1. Cálculo de un valor decimal para un código <i>LBP</i>	12
Figura 2.2.3.2. Ejemplo de aplicación de <i>LBP</i> . Imagen original (a), imagen modificada con operador <i>LBP</i> (b). Recuperado de (Huang, Shan, Ardabilian, Wang, & Liming, 2011).	12
Figura 2.3.1.1. <i>Smart glasses</i> de diferentes compañías.	14
Figura 2.4.2.1. Recorrido en un escenario inmersivo. Vista del usuario (a), sensor Oculus Rift (b).	16
Figura 3.3.1. Diseño de implementación del reconocedor.	25
Figura 3.3.1.1. Diagrama de bloques de selección de logotipos.	26
Figura 3.3.2.1.1. Algunas imágenes del conjunto de ejemplos negativos para entrenamiento.	27
Figura 3.3.2.2. Izquierda, logotipo original. Derecha, algunos ejemplos positivos del logotipo HSBC para entrenamiento.	28
Figura 3.3.2.3. Varias imágenes para el logotipo de Devlyn.	28
Figura 3.3.2.3. Algunas imágenes del conjunto de ejemplos negativos para prueba.	30
Figura 3.3.2.4. Ejemplos positivos del conjunto de prueba para logotipos Ferrioni, Adidas, Puma y Tous.	31
Figura 3.4.1. Pasos para generar el escenario de pruebas.	32
Figura 4.2.1.1. Logotipos seleccionados para ser identificados por el reconocedor. Adidas (a), Burger King (b), Cinépolis (c), Devlyn (d), Ferrioni (e), HSBC (f), Liverpool (g), Martí (h), Nutrisa (i), Puma (j), Radioshack (k), Santander (l), Starbucks (m), Steren (n), Subway (o), Taco Inn (p), Telcel (q), Todomoda (r), Tous (s), Vans (t).	40
Figura 4.3.1.1. Integración y generación de contenido 3D. Modelo original (a), modelo iluminado (b), generación de plano y límites de colisión, líneas en color verde (c).	52
Figura 4.3.2.1. Centro comercial virtual con las tiendas de los logotipos que el reconocedor puede identificar.	52
Figura 4.3.3.1. Recorrido inmersivo con Oculus Rift DK1 en el cual se identifica el logotipo Ferrioni.	53
Figura 4.3.4.1. Ajuste de valores de variables y técnicas de detección para cada logotipo en el ambiente de desarrollo Unity.	54
Figura 4.4.1. Diseño de la aplicación del sistema final.	57
Figura 4.4.2. Casco de realidad virtual del sistema implementado en un dispositivo móvil.	57
Figura 4.4.3. Usuario utilizando el sistema mientras recorre un centro comercial.	58
Figura A.2.1. Sección de código para medir la exactitud de detección del logotipo Adidas.	80
Figura A.3.1. Código del método principal del reconocedor.	80

Nomenclatura

Resumen

El objetivo de esta tesis es desarrollar una aplicación de realidad aumentada que integra algoritmos de visión por computadora y aprendizaje automático para reconocer logotipos comerciales a partir de una secuencia de imágenes y proveer una forma innovadora de hacer publicidad móvil.

En la primera parte, se muestran los resultados de aplicar *Haar-like features* y *Local Binary Patterns*, dos técnicas conocidas en la comunidad de visión por computadora en el reconocimiento de logotipos comerciales. La tesis identifica la mejor técnica para cada logotipo y las reúne para producir un reconocedor. En la sección central se detalla el análisis realizado con base en experimentos de detección. En una siguiente etapa, se muestran dos implementaciones del reconocedor en dos aplicaciones de realidad aumentada. La primera de ellas sobre un mundo virtual y la segunda sobre el mundo real. Como resultado de la implementación del reconocedor se pueden mostrar mensajes publicitarios.

Los resultados experimentales de reconocimiento sobre un conjunto de imágenes del mundo real muestran una exactitud promedio del 92%. La tesis establece que los resultados obtenidos son similares a los de otros autores. Se expone una prueba de usabilidad para comprobar la funcionalidad de las aplicaciones de realidad aumentada y se aprueba por un grupo de evaluadores. Por último, se establece el trabajo futuro.

Abstract

The aim of this thesis is to develop an augmented reality application that integrates computer vision and machine learning algorithms to recognize trademark logos over an image sequence and to provide an innovative way to do mobile marketing.

In the first part, results of applying Haar-like features and Local Binary Patterns, two well-known techniques in the computer vision community, for trademark logo recognition are presented. Then the best technique for each logo is selected, and all of them are joined to produce a single recognizer. In a detail central section the thesis provides the analysis of the detection experiments. In a following stage two implementations of the recognizer in augmented reality applications are shown. The first one over a virtual world that represents a mall and the second one, over a real world by using a mobile device. As a result of the recognizer implementation advertising messages can be displayed.

The experimental results on several real-world images show a recognition average accuracy of 92%. The thesis state that these results are similar to other author's results. It is also exposed a usability test to check the functionality of the augmented reality applications and it is approved by a group of evaluators. Finally, future work is set.

Capítulo 1 Planteamiento del problema de investigación

1.1 Introducción

Las empresas son entidades que están en continua evolución. Aquellas que enfocan la promoción de sus productos o servicios mediante la presencia de tiendas en centros comerciales están en la constante búsqueda de nuevas maneras de publicitarlos para hacerlos llegar a sus consumidores. Por otro lado, los consumidores son cada vez más demandantes en cuanto a la calidad del producto o servicio ofertado. De aquí que la publicidad se adapte continuamente a las expectativas y exigencias de los clientes.

En este sentido, es común ver que estas empresas utilizan elementos tecnológicos para difundir y penetrar en todos los ámbitos posibles de la sociedad con el fin de hacer presente su imagen en la mayor cantidad posible de espacios.

Es conveniente observar la forma en la que plasman en su marca su imagen. De hecho, los consumidores llegan a estar dispuestos a hacer esfuerzos notables por conseguir los productos y servicios que portan los símbolos o logotipos pegados en prendas y objetos de consumo (Martín Requero & Alvarado López, 2007). De esta manera, se puede observar que los consumidores asocian a las marcas con los logotipos.

En estos días, los anuncios publicitarios llegan desde todos los medios: prensa escrita y digital, radio, televisión, internet, dispositivos móviles. Se comprende que las empresas usen estos recursos para ofrecer sus productos. Más aún, las empresas revisan continuamente las tecnologías emergentes para valorar su potencial, para

insertarlas en los medios de difusión publicitaria. En efecto, la publicidad se basa de cualquier herramienta para promocionar la venta de un artículo o servicio. Una de ellas incluye el uso de tecnologías de visión por computadora para el procesamiento de imágenes.

En este sentido, es posible observar que la realidad aumentada y la realidad virtual evolucionan la manera en la que se puede hacer publicidad. Es por ello que las empresas aprovechan estas tecnologías.

Las aplicaciones de realidad aumentada se basan en la identificación de símbolos. De tal manera que cuando se identifica uno de ellos, se presenta información relevante a él. Si a estos símbolos se visualizan como logotipos comerciales, se encuentra un área de oportunidad para las empresas. Esto es, hacer uso de su propio logotipo para difusión de mensajes publicitarios.

No cabe duda que los dispositivos móviles son cada vez más inteligentes y sirven como plataformas de estrategias de mercado (Ruiz Davis & Polo Serrano, 2012). Desde luego, sus capacidades de adquisición de imágenes, almacenamiento y procesamiento aumentan constantemente (Escalera, Baró, Pujol, Vitriá, & Radeva, 2011). Por tanto, es común que se desarrollen aplicaciones que aprovechen estas características.

El propósito de este trabajo consiste en someter a consideración a las empresas de una nueva tecnología que les permita promocionar sus ofertas y utilizar sus propios logotipos como símbolos de identificación de una aplicación de realidad aumentada.

Dado que actualmente es posible desarrollar aplicaciones para dispositivos móviles que realicen reconocimiento de imágenes, se considera viable el desarrollo de una aplicación que identifique logotipos de marcas comerciales que, cuando reconozca a un logotipo sea capaz de proporcionar información publicitaria de su marca propietaria.

En general, la detección y reconocimiento automático de objetos en imágenes son temas con grandes retos en las áreas de visión por computadora y aprendizaje computacional. De hecho, uno de los principales problemas consiste en tratar con la variabilidad de imágenes que se pueden obtener del mundo real (Escalera, Baró, Pujol, Vitriá, & Radeva, 2011). La identificación de objetos en imágenes del mundo real se enfrenta con diferentes dificultades en la detección, como son, diferentes perspectivas de vista, cambio de iluminación, fondos confusos, oclusiones, limitaciones de las cámaras y otras distorsiones (Brahnam, C. Jain, Nanni, & Lumini, 2014).

Para ello, esta tesis propone la integración de algoritmos de visión por computadora y aprendizaje computacional para obtener como resultado un sistema de reconocimiento de imágenes, que permita reconocer logotipos de marcas comerciales a partir de una secuencia de imágenes.

Este proyecto pretende, además de los propósitos científicos y tecnológicos, innovar en la publicidad móvil al aplicar visión por computadora mediante la creación

de un reconocedor automático de logotipos comerciales en una aplicación de realidad aumentada. Entiéndase por reconocedor el proceso que recibe una imagen e indica como salida si en la imagen existe algún objeto y determina su identidad. En este caso, busca si en la imagen existe un logotipo y determina cuál encontró, si es que encontró alguno.

A partir de la integración de dos técnicas de identificación de objetos, se desarrolla el reconocedor automático de logotipos comerciales en imágenes que funciona en una aplicación para dispositivos móviles. Cuando identifique a un logotipo, se le presenta al usuario un mensaje publicitario que contiene la información de la marca del logotipo. Se precisa además que los usuarios podrán utilizar la aplicación mientras recorren un centro comercial. Cabe señalar que solo se reconocen ciertos logotipos.

Este trabajo de tesis se organiza de la siguiente manera: en el presente Capítulo, además del planteamiento expuesto, se incluye el objetivo general, los objetivos específicos y la justificación.

Posteriormente, en el Capítulo 2 se describen las técnicas y algoritmos relacionados con el reconocimiento de imágenes. En lo que concierne al Capítulo 3, en él se describe el diseño del reconocedor y se presenta el diseño de un mundo virtual para su implementación. Después, en el Capítulo 4 se muestran los resultados obtenidos de los experimentos realizados y se incluye la evaluación de la aplicación de realidad aumentada. Por último, en el Capítulo 5 se destacan las principales conclusiones, y se establece el trabajo futuro que surge a partir de los resultados obtenidos.

1.2 Objetivos

1.2.1 Objetivo general

Desarrollar una aplicación de reconocimiento automático de imágenes utilizando visión por computadora para identificar logotipos de marcas y mostrar al usuario mensajes publicitarios.

1.2.2 Objetivos específicos

1. Evaluar dos técnicas de visión por computadora para identificar logotipos que traten con la variabilidad de imágenes.
2. Desarrollar un reconocedor automático de logotipos de marcas comerciales con base en las técnicas evaluadas.
3. Construir un escenario de pruebas por medio de herramientas 3D para mostrar y evaluar la funcionalidad de la aplicación.

1.3 Justificación

Actualmente la competencia de mercado presiona a todas las empresas a buscar nuevas formas de publicidad. Aunado a esto, existe un incremento en capacidades de adquisición, almacenamiento y procesamiento de dispositivos móviles. Esto permite a las empresas aprovechar un nuevo mercado para publicitar sus productos y servicios.

El desarrollo de sistemas automáticos de identificación de imágenes en dispositivos móviles es un problema que está abordándose de forma global, Por ejemplo, la empresa NEC con su servicio Gaziru (Fukuzawa, 2014). Más aún, dispositivos como Google Glass forman alianzas con empresas como Catchoom (Catchoom, 2015) para implementar reconocimiento de imágenes en aplicaciones de compras. Asimismo, la empresa Microsoft libera su dispositivo Hololens cuya tecnología involucra el reconocimiento de imágenes y realidad aumentada (Microsoft, 2015).

Estos hechos establecen tendencias en el desarrollo de aplicaciones para dispositivos portátiles que se enfoquen en el reconocimiento de imágenes y que hacen uso de algoritmos de aprendizaje computacional. Esto se soporta al observar los resultados actuales tanto en investigación de visión por computadora como en herramientas de código abierto desarrolladas para este fin.

El sistema de reconocimiento automático de logotipos comerciales en imágenes pretende seguir estas tendencias y enfocar su aplicación en centros comerciales. Con base en la idea de (Ferrell & Hartline, 2012) que establece que la publicidad se basa en la idea o noción de que, mientras más llegada al público tenga un producto, más conocido se hará y por lo tanto, tendrá más posibilidades de ser consumido.

En el capítulo siguiente se presentan las bases teóricas inmersas en la elaboración de este trabajo de investigación.

Capítulo 2 Marco teórico

En este capítulo se presentan las bases teóricas que dan lugar a este trabajo así como el conjunto de trabajos relacionados que son el marco de referencia de los resultados obtenidos.

Para ubicar al lector en el contexto de este trabajo, obsérvese la Figura 2.1. En ella se presentan los principales tópicos envueltos en este trabajo. Por un lado, el reconocimiento de objetos que se fundamenta en la visión por computadora y el aprendizaje automático. Por otro, la realidad aumentada que representa el campo de aplicación y, por último, la publicidad móvil, que es hacia donde se destina el uso de la aplicación.



Figura 2.1. Tópicos que se comprenden en este trabajo.

Para comprender cada uno de estos tópicos se presentan las secciones siguientes.

2.1 Visión por computadora

La visión por computadora ha visto sus inicios décadas atrás. Es una rama de la Inteligencia Artificial que ha experimentado un mayor crecimiento en los últimos años. Aun cuando líneas de investigación son muy diversas, la de reconocimiento de objetos en imágenes ha sido una de las más estudiadas y ha evolucionado bastante en la última década al introducir nuevas técnicas y algoritmos para este fin.

Existen diferentes definiciones de visión por computadora. En este trabajo se basa en la definición que expone (Sucar & Gómez, 2008): “Visión es un proceso que produce a partir de las imágenes del mundo exterior una descripción que es útil para el observador y que no tiene información irrelevante”.

En concreto, visión por computadora consiste en el estudio de los procesos de reconocimiento y localización de objetos mediante el procesamiento digital de imágenes. Es así como, su principal propósito se enfoca en la construcción de sistemas artificiales que obtienen información a partir del análisis de imágenes.

El procesamiento digital de imágenes se refiere al proceso que toma una imagen como entrada y produce como salida una versión modificada de ésta. Por su parte, el análisis de imágenes recibe una imagen como entrada y obtiene una medición, interpretación o decisión a partir de su contenido (Mery, 2004).

De acuerdo con (Y. Shih, 2010), la visión por computadora comprende los siguientes procesos:

- **Adquisición de imágenes.** En este proceso se obtienen las imágenes con las que se va a trabajar, puede ser a través de archivos de imágenes, cámaras digitales u otros dispositivos. Cabe señalar que las imágenes que se obtienen dependen de la aplicación que se pretenda realizar, por ejemplo, fotografías, radiografías, termografías, etcétera.
- **Pre-procesamiento.** Su objetivo es el de mejorar las imágenes obtenidas para resaltar determinadas características de interés o bien para eliminar aquello que se considere irrelevante. Esto se realiza con el uso de filtros.
- **Extracción de características o medición.** Este proceso determina en cada imagen las principales características que la definen y realiza una medición objetiva de atributos como color, tamaño o posición, entre otros.
- **Segmentación.** Separa en diferentes conjuntos aquellos grupos de píxeles u objetos con características comunes con el propósito de contar con una representación significativa y fácil de analizar, que la

imagen original. Esto permite identificar el o los objetos de estudio en la imagen.

- Transformaciones morfológicas. A través de la teoría de conjuntos, modifican la estructura o forma de los objetos en la imagen para su análisis individual.
- Interpretación, reconocimiento o clasificación. A partir de los valores obtenidos por las mediciones se lleva a cabo la interpretación de los objetos: su clasificación o bien su categoría.

2.1.1 Sistemas de visión por computadora

Un sistema de visión por computadora se conforma de los siguientes elementos: un dispositivo para capturar imágenes (cámara), una o más computadoras y software que realice la manipulación de las imágenes adquiridas de tal forma que se puedan reconocer características en la imagen. En algunos casos se considera como parte de un sistema de visión por computadora a un sistema de iluminación que permite asegurar la calidad de las imágenes.

La cámara es el elemento básico sensible del sistema. Existen diferentes cámaras de acuerdo con el tipo de imágenes que pueden captar, como visibles, térmicas o radiográficas (Nixon & Aguado, 2008).

2.1.2 Procesamiento de imágenes

Las aplicaciones de visión por computadora requieren de un gran cuidado en la etapa del procesamiento de imágenes para alcanzar resultados aceptables (Szeliski, 2010).

El procesamiento de imágenes se lleva a cabo por medio de un conjunto de operaciones o transformaciones que permiten mejorar las imágenes para resaltar determinadas características de interés o bien para eliminar aquello que se considere irrelevante. Algunos ejemplos de estas operaciones son la corrección de exposición, balance de color, reducción de ruido y aumento de nitidez.

Un operador de procesamiento de imágenes es una función que toma como entrada una o más imágenes y produce a partir de ellas una imagen de salida. Las Ecuaciones 1 y 2 muestran dos definiciones de esta función en un dominio continuo donde x está en el dominio D-dimensional (en el caso de imágenes, $D = 2$) y, las funciones f , g y h operan sobre un rango escalar o vectorial. El dominio consiste de un número finito de ubicaciones de píxeles.

$$g(x) = h(f(x)) \qquad \text{Ec. (1)}$$

$$g(x) = h(f_0(x), \dots, f_n(x)) \quad \text{Ec. (2)}$$

El procesamiento de imágenes se realiza en varios niveles, el nivel más básico es el procesamiento de punto, cuya tarea consiste en mapear los valores originales de los píxeles a nuevos valores. Esto produce una nueva imagen de salida. Algunos ejemplos de estos operadores corresponden a la corrección de brillo y contraste. La Ecuación (3) presenta la definición conjunta de los operadores de brillo y contraste a través de la multiplicación y la suma de una constante. Los parámetros a y b se conocen también como ganancia y tendencia.

$$g(x) = af(x) + b \quad \text{Ec. (3)}$$

2.2 Reconocimiento de objetos en imágenes

Es posible decir que, los objetos en general pueden reconocerse por sus características como color, forma y textura. De manera que, la detección de objetos en imágenes consiste en encontrar si una imagen contiene algún objeto de interés y, de ser así, indicar su posición en la imagen. Por consiguiente, algunos ejemplos de detección de objetos pueden ser: detección de rostros, movimientos de las manos, detección de autos y señales de tránsito. Cuando se necesita identificar a una persona en específico o bien a una señal de tránsito en específico, el problema se conoce como reconocimiento (Boguslaw, 2013).

2.2.1 Clasificación de objetos

De acuerdo con (Escudero, 2009), un clasificador es un método para determinar la posible clase de un objeto desconocido dentro de un conjunto.

Generalmente, para clasificar objetos, el primer paso a seguir es la extracción de características. Cuyo objetivo es que cada objeto del conjunto de entrenamiento sea expresado por un vector de valores que lo caracterizan. El espacio que abarca a todas las combinaciones posibles de características se le denomina espacio de características. Si se considera a los píxeles de una imagen como características de un objeto, el problema de clasificación deberá trabajar con cada pixel, lo que aumenta la complejidad de la clasificación. Por ello comúnmente se recurre a una reducción de características. De entre los métodos más comunes se encuentran *Principal Component Analysis (PCA)*, *Independent Component Analysis (ICA)*, *Linear Discriminant Analysis (LDA)*, *Local Binary Patterns (LBP)* y *Haar-like Features*.

Una vez que se han extraído las características de los objetos, se procede con su clasificación; ésta puede ser supervisada o no supervisada. Algunos clasificadores supervisados son: discriminante lineal de *Fisher*, vecinos más

cercanos (*KNN*), máquinas de soporte vectorial (*SVM*) y *Adaboost*. Por otra parte, los clasificadores que distinguen entre dos tipos de objetos o eventos se conocen como clasificadores binarios. Por consiguiente, para distinguir más de dos clases se utilizan los clasificadores multiclase.

2.2.2 Adaboost

El reconocimiento de objetos puede verse como un problema de clasificación. Cuando se requiere localizar un objeto en una imagen, una técnica utilizada consiste en recorrer toda la imagen y aplicar múltiples ventanas detectoras de diferentes tamaños. Para cada una de estas ventanas se aplica un clasificador que indicará si existe o no el objeto buscado. Se debe tomar en cuenta que el desplazamiento de la ventana a través de la imagen es un factor determinante en el desempeño de la clasificación. Si el desplazamiento es muy pequeño, el tiempo de clasificación es más grande. Por otra parte, si el desplazamiento es muy grande podrían no encontrarse objetos dentro de la ventana.

Adaboost es un método de agregación de muchos clasificadores débiles para crear un clasificador fuerte para detectar objetos en imágenes. Este clasificador determina si diferentes ventanas en una imagen contengan a un objeto de interés. (Viola & Jones, 2001), demuestran que *Adaboost* puede ser usado tanto para reducción de características como para entrenamiento del clasificador. La principal desventaja de este algoritmo es el alto consumo de tiempo de entrenamiento.

(Uddin & Akhi, 2016), determinan cuatro pasos para construir un sistema de detección de objetos con *Adaboost*.

1. *Seleccionar un conjunto de datos.* Deben generarse ejemplos para construir el conjunto de entrenamiento. Los ejemplos pueden ser positivos y negativos. Los ejemplos positivos consisten de cada imagen que contenga al objeto que se desea identificar. Los objetos de interés en las imágenes de los ejemplos positivos pueden aparecer en diferentes perspectivas, colores y escalas. Los ejemplos negativos corresponden a imágenes en donde no se encuentra el objeto a identificar; pueden ser imágenes arbitrarias.
2. *Entrenar los valores de umbral para cada característica.* Con base en los ejemplos positivos y negativos se realiza el entrenamiento del clasificador.
3. *Seleccionar y entrenar un subconjunto de clasificadores.*
4. *Entrenar la cascada de clasificadores.*

Una vez que se ha construido el clasificador, se escanean imágenes de forma exhaustiva donde los objetos a localizar estén en diferentes ubicaciones y escalas.

Cada característica se representa por un modelo, su coordenada relativa al origen de la ventana de búsqueda y su escala. Cada característica se compone de dos o tres rectángulos blancos y negros juntos. Estos rectángulos se pueden rotar a 45 grados. El valor de la característica se calcula como una suma ponderada de dos componentes: el nivel de gris del pixel más el rectángulo negro más toda el área que comprende la característica (todas las áreas blancas y negras). Un ejemplo de estas características se presenta en la Figura 2.2.2.1.

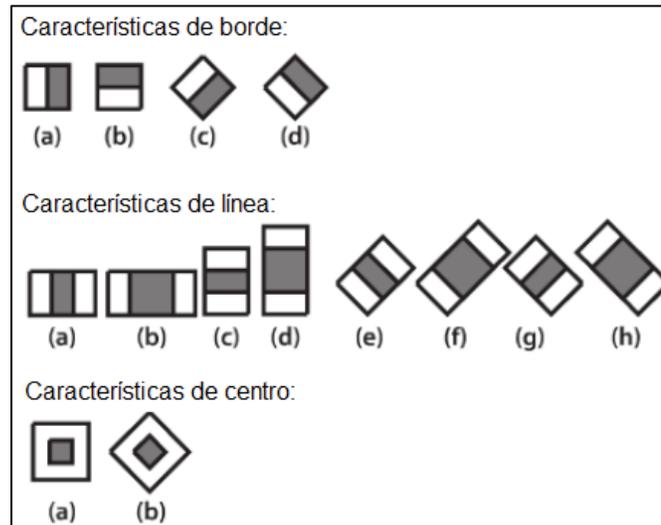


Figura 2.2.2.1. *Haar-like Features*.

Adaboost encuentra los pesos adecuados para sus clasificadores débiles. Su algoritmo de *boosting* maximiza el margen entre un conjunto de ejemplos positivos de uno negativo. Cada ejemplo se convierte a una imagen en escala de grises y debe escalarse a la resolución del detector.

La Figura 2.2.2.2 muestra ejemplos de reconocimiento de rostros con *Haar-like Features*.



Figura 2.2.2.2. Reconocimiento de rostros con *Haar-like Features*. Recuerdo de (Ikeuchi, 2014).

2.2.3 Local Binary Patterns

La clasificación de texturas es un problema básico en el procesamiento de imágenes y visión por computadora. Los métodos de clasificación de texturas deben ser robustos a la traslación, rotación y escala (Zhao, 2012). El método *Local Binary Patterns (LBP)* es uno de ellos, su característica más importante es su tolerancia a cambios de iluminación y su simplicidad computacional.

En un principio *LBP* fue usado como un clasificador de texturas, probado ser invariante a la rotación en imágenes en escala de grises. Sin embargo, recientemente ha sido utilizado como técnica de extracción de características, principalmente en la detección de rostros y análisis de expresiones faciales. *LBP* caracteriza la estructura espacial de la textura local de una imagen. La Ecuación 8 describe cómo se aplica a una región de una imagen.

$$LBP_{P,R} = \sum_{p=0}^{P-1} s(g_p - g_c) 2^p, \quad s(x) = \begin{cases} 1, & x \geq 0 \\ 0, & x < 0 \end{cases} \quad \text{Ec. (8)}$$

Donde g_c representa al pixel central, $g_p (P=0, \dots, P-1)$ denota a la vecindad de g_c con un radio R . P es el número total de vecinos. En una vecindad de 3×3 , cada pixel se compara con sus ocho vecinos y se resta el valor del pixel central. Para cada pixel se obtiene un valor binario a partir de la concatenación de los valores del vecindario y se realiza un recorrido en dirección de las manecillas del reloj. Este valor se utiliza para etiquetar al pixel. Los valores resultantes se conocen

como códigos *LBP* o bien como *LBP*s (Huang, Shan, Ardabilian, Wang, & Liming, 2011). En la Figura 2.2.3.1 se presenta un ejemplo de este cálculo.

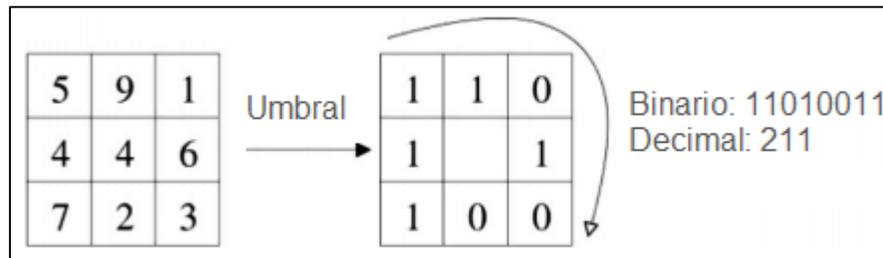


Figura 2.2.3.1. Cálculo de un valor decimal para un código *LBP*.

Un ejemplo de aplicar *LBP* a una imagen se observa en la Figura 2.2.3.2.

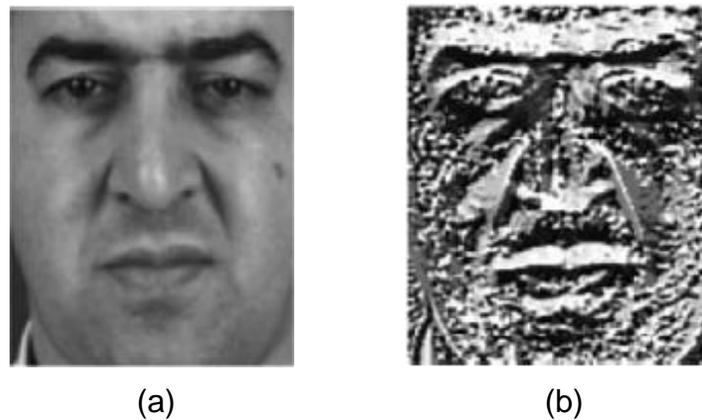


Figura 2.2.3.2. Ejemplo de aplicación de *LBP*. Imagen original (a), imagen modificada con operador *LBP* (b). Recuperado de (Huang, Shan, Ardabilian, Wang, & Liming, 2011).

2.3 Publicidad móvil

Los teléfonos móviles se han convertido en una herramienta de nuestra vida y quehacer diario. Son utilizados por millones de personas gracias a su plataforma de comunicación y a sus características de integración social. La telefonía móvil se ha convertido en la forma de comunicación más extendida en el mundo y permite interacciones entre anunciantes y clientes de forma rápida y fácil.

En (Dickinger, Heinzmann, & Murphy, 2005), se define la publicidad móvil como “la utilización de medios inalámbricos para proporcionar a los clientes información personalizada sensible a la hora y al lugar, que promociona mercancías, servicios, ideas, generando valor para todos los grupos de interés”.

La industria publicitaria aprovecha el nivel de penetración que tiene la telefonía móvil para convertir los teléfonos en plataformas de distribución. Como se señala en (Marín Dueñas, 2013), la integración de la publicidad es tal que puede transformar de forma radical la estructura de la cadena de valor y los modelos de negocio conforme la conectividad y las aplicaciones tecnológicas avanzan; lo que permite ofrecer nuevas formas de promoción.

De acuerdo con (Martínez & Aguado, 2006), la telefonía móvil puede verse como una plataforma mediática caracterizada por la translocalidad (que se define como la posibilidad de estar conectado en un desplazamiento a través de espacios o bien, de trayectorias completas) y la integración de servicios (como Internet y televisión) en un dispositivo de comunicación y acceso. Los desarrollos de estándares tecnológicos, la portabilidad y la identidad individual han hecho del teléfono móvil el dispositivo idóneo para sentar las bases de una nueva integración de opciones y servicios. De esta manera se constituye la base del proceso mediático de la telefonía móvil.

2.3.1 Realidad aumentada aplicada a la publicidad

La popularidad de los teléfonos inteligentes y otros dispositivos móviles ha permitido el desarrollo de aplicaciones de realidad aumentada, lo que abre al mercado un concepto que parecía estar limitado a entornos de desarrollo. La realidad aumentada, mezcla de realidad y virtualidad, requiere de diferentes capacidades tecnológicas, como son, sistemas de visión en gafas, grandes necesidades de procesamiento, sensores, sistemas de localización y conectividad, entre otros.

En la actualidad, las posibilidades de adaptación de la realidad aumentada son muy amplias. Aunque se pueden señalar aplicaciones en sectores como entretenimiento, turismo, educación, salud, militar, manufactura, mantenimiento automovilístico y aeronáutico, entrenamiento de habilidades y destrezas, la mercadotecnia y los procesos de venta son los ámbitos donde más se aplica (Telefónica, 2011). El elemento fundamental de la mercadotecnia es captar la atención. Las empresas ven en la realidad aumentada una forma de diferenciarse con respecto a la competencia, ya que ofrece a los usuarios la posibilidad de acceder a nuevas experiencias visuales. Como lo hace notar (Martí Parreño, 2011), la información proporcionada por una aplicación de realidad virtual podría resultar determinante en el momento de la compra, en especial de aquellos productos con los que el consumidor está menos familiarizado.

La realidad aumentada requiere de una pantalla, que puede ser de una computadora, de un teléfono móvil o de un *Head Mounted Display (HMD)*, en la que la información tomada de la realidad se combina con información específica, como una imagen, una animación o un modelo en 3D.

Dentro de la categoría de *HDM* se encuentran los dispositivos conocidos como *Smart glasses*. En la Figura 2.3.1.1 se muestran algunos de estos con su respectiva marca.



Figura 2.3.1.1. *Smart glasses* de diferentes compañías.

En (Rasiel, 2015), se predice que el mercado de los *Smart glasses* alcanzará el billón de ventas al término de esta década. La idea que propone es que las aplicaciones que se utilizan en *smartphones* poco a poco migrarán para convertirse en aplicaciones para *Smart glasses*.

2.4 Herramientas de desarrollo

En las secciones siguientes se presentan las herramientas que se emplean en el desarrollo del sistema: OpenCV y Unity. Las funciones de OpenCV se utilizaron para la detección y reconocimiento de logotipos. El ambiente de desarrollo Unity permite crear un ambiente para evaluar el reconocedor automático de logotipos.

2.4.1 OpenCV

OpenCV es un conjunto de librerías de código abierto de visión por computadora. Comenzó como un proyecto de investigación de la empresa Intel en 1998 enfocado en proveer las herramientas necesarias para resolver problemas de visión por computadora (Pulli, Baksheev, & Korniyakov Kirill, 2012). Fue diseñado para ser eficiente computacionalmente y con un gran enfoque en aplicaciones de tiempo real. Está escrito en lenguaje C optimizado y puede aprovechar las ventajas de los procesadores de núcleos múltiples.

Uno de los objetivos de OpenCV es proveer una infraestructura de visión por computadora, fácil de usar, que le permita a los desarrolladores, construir aplicaciones sofisticadas de visión rápidamente. Contiene una mezcla de

funciones de procesamiento de imágenes de bajo nivel y algoritmos de alto nivel como detección de rostros, detección de peatones, correspondencia de características y seguimiento automático (traqueo). A partir de 2010 cuenta con un módulo GPU de aceleración por hardware que permite hacer uso de la GPU sin tener que programarla.

2.4.2 Unity

El uso de herramientas para la creación de videojuegos ha establecido un patrón importante en el desarrollo de nuevas y mejores herramientas. Las áreas de oportunidad que surgen de ello son bastas y comprenden aplicaciones como simulación, animación, inmersión, realidad aumentada, interfaces humano computadora, entre otras. Los desarrollos se pueden distribuir no solo en sistemas operativos tradicionales o sistemas móviles como Android, iOS, Windows Phone y Blackberry, sino también en diferentes consolas como Xbox, Play Station, PS Vita por mencionar algunas.

Los motores de videojuegos son las herramientas que están detrás de muchos videojuegos que se encuentran en el mercado. La funcionalidad básica de un motor de juegos es la de proveer de un motor de renderizado para los gráficos 2D y 3D, un motor físico, detector de colisiones, sonidos, codificación, animación, inteligencia artificial, redes, administración de memoria y un escenario gráfico (Goldstone, 2011).

El motor de juegos Unity es uno de los más utilizados y lo pueden usar tanto desarrolladores novatos como profesionales. Las plataformas de desarrollo de este motor incluyen Windows, Mac y Linux. Los juegos producidos se pueden ejecutar en Web (con su plug-in Unity Web Player), PC, iOS, Android, Windows Phone, Blackberry y consolas de videojuegos como Xbox 360, Xbox One, PS3, PS4, Nintendo Wii.

Aunque Unity es una herramienta para el desarrollo de video juegos, tiene aplicación en otras áreas como: simulación, realidad virtual, sistemas inmersivos y realidad aumentada. Los sistemas inmersivos le permiten al usuario sumergirse en un mundo artificial a través de dispositivos sensoriales. De esta manera se simula la realidad y se proporciona una experiencia en primera persona. En los escenarios que se crean, el usuario puede desplazarse de manera virtual. En la Figura 2.4.2.1 se muestra un ejemplo de ejecución de este escenario inmersivo que utiliza un sensor OculusRift. Las dos salidas de video corresponden a la señal que llega a cada ojo del usuario.



(a)



(b)

Figura 2.4.2.1. Recorrido en un escenario inmersivo. Vista del usuario (a), sensor Oculus Rift (b).

2.5 Trabajos relacionados

En este apartado se presentan diferentes trabajos relacionados con la detección y reconocimiento de logotipos. Se resume cómo abordan el problema y qué utilizan para solucionarlo; además se exponen las características de los experimentos que abordaron, el número de clases que utilizaron (cuando se encontró disponible) y el porcentaje de exactitud reportado.

(Doermann, Rivlin, & Weiss, Logo Recognition Using Geometric Invariants , 1993), exploraron el reconocimiento de logotipos en documentos para obtener información semántica y decidir cuándo es conveniente analizar componentes del texto. Implementan la detección y extracción de características con algoritmos *global invariants* y *local invariants* para demostrar su exactitud en un conjunto de nueve logotipos.

(Yunqiong, Zhifang, & Fei, 2007), proponen un método de *template matching* e histogramas de orientación de bordes para reconocer logotipos de vehículos con el propósito de reconocer los tipos de vehículos a través de una cámara de vigilancia. Los experimentos los realizaron para diecisiete clases de logotipos, y obtienen un noventa por ciento de exactitud.

(Doermann & Zhu, Automatic Document Logo Detection, 2007), exploran un nuevo método para el reconocimiento de logotipos en documentos. Utilizan una estrategia de *boosting* sobre imágenes a múltiples escalas. El clasificador que eligieron es Fisher por su característica de que requiere menos ejemplos para su entrenamiento. Las regiones donde se ubican los logotipos se clasifican a diferentes escalas a través de una cascada de clasificadores simples. Los experimentos de la detección de logotipos están basados en el conjunto de datos Tobacco-800. De él se utilizaron cincuenta documentos como conjunto de entrenamiento y el resto, para pruebas. La exactitud que lograron fue del ochenta y cuatro por ciento.

(Xia, Qi, & Qianhao, 2008), presentan un método para el reconocimiento de logotipos en imágenes naturales. Como retos marcan que gran variedad de imágenes hace que los logotipos varíen en forma y ubicación; además, las imágenes cuentan con diferentes tipos de logotipos y, en algunos casos, los logotipos están ocluidos. Para resolver esto, proponen utilizar *SIFT* al conjunto de pruebas para detectar las regiones de interés y extraer las características discriminantes. Posteriormente aplican la estrategia de búsqueda de vecinos más cercanos para la comparación. La exactitud que se obtiene es del noventa y cinco por ciento sobre un conjunto de prueba de mil imágenes. El inconveniente de este método es consume más tiempo en la clasificación que otros métodos.

El trabajo de (Wei, Li, Chau, & Li, 2009) aborda el problema de tratar con un número vasto de imágenes de marcas comerciales en un sistema de registro de marcas comerciales. Hace uso del detector de bordes *Canny*, ejecuta una normalización de forma y por último extrae las características globales y locales. Para medir la similitud entre una imagen de entrada y las imágenes de una base de datos utiliza una estrategia de *feature matching* de doble componente. El conjunto de prueba para este experimento constó de mil tres imágenes distribuidas en catorce clases. Los resultados muestran que su método propuesto se desempeña mejor que algoritmos como *momento invariants*, descriptores de *Fourier*, *Zernike moments* y *CSS*.

(Phan & Dimitrios, 2010), presentan una extensión al algoritmo de *Color Edge Cooccurrence Histogram (CECH)* para detección recuperación de logotipos y marcas comerciales en una base de datos de imágenes sin restricciones. La identificación se realiza en una base de datos de cinco mil cuatrocientas imágenes con cuatro clases de logotipos; cuatrocientas de estas imágenes contienen los logotipos buscados. Los resultados permiten tener un promedio de exactitud del noventa y dos por ciento.

(Wang, 2010), propone un simple y dinámico enfoque para la detección y reconocimiento de logotipos en documentos. Primero ubican la región donde se encuentra un logotipo, aplica *geometric invariants* y realiza el *shape matching*. El algoritmo que utilizan para reconocimiento es *Bayesian Belief Network*; en la última fase de verificación se emplea un clasificador de distancia Euclidiana entrenado por un número de características globales que incluye *Zernike moments*. Los resultados muestran una exactitud del noventa y dos por ciento.

El análisis inteligente de imágenes en documentos propuesto por (Li, Matthias, & Martin, 2010), se basa en una arquitectura libre de segmentación e independiente del diseño para la detección y reconocimiento de logotipos. Para la tarea de reconocimiento se utiliza un descriptor de formas invariante a la rotación y a la escala. Se obtiene la envoltura convexa del logotipo un descriptor de componentes conectados. El emparejamiento se calcula con la distancia Euclidiana. Uno de los experimentos mostrados permite realizar la detección y el reconocimiento de doscientos sesenta y dos logotipos alcanzan un grado de exactitud del noventa y ocho por ciento.

(Pysillos, Christos-Nikolaos, & Eleftherios, 2010), presentan un algoritmo basado en *SIFT* modificado con *feature-matching*. El algoritmo se prueba sobre una base de datos de mil doscientos logotipos distribuidos en diez clases. Alcanzan un grado de exactitud del noventa y un por ciento.

El trabajo de (Hassanzadeh & Pourghassem, 2011), se enfoca en el tratamiento de ruido, oclusiones y diferentes escalas y orientaciones de imágenes en documentos. El trabajo se basa en características estructurales y espaciales. En primer lugar se normalizan las imágenes, se extraen las características basadas en histogramas verticales y horizontales. Por último utilizan el clasificador *KNN* para el reconocimiento. Sus experimentos los realizan en ciento cinco logotipos y logran en promedio un noventa y siete por ciento de exactitud.

(Pourghassem, 2012), explora un algoritmo de detección y reconocimiento de logotipos basado en redes neuronales y *KNN*. El análisis de detección se ejecutó con una base de datos de mil novecientas ochenta imágenes en siete clases. La exactitud obtenida es del noventa y cuatro por ciento.

(Zhang, Zhang, Liang, & Liang, 2012), exponen un trabajo para la detección y la ubicación de logotipos en videos de anuncios. El trabajo permite predecir las regiones donde se ubican los logotipos. Hace uso de ecualización de color *HSV*. La extracción de características se realiza por medio de una fusión multi-característica basada en contenido. El experimento utiliza veinte videos de anuncios con catorce logotipos. Los resultados muestran que la detección y reconocimiento es rápido. Sin embargo la exactitud reportada es del setenta por ciento.

(Lei, Chen, & Lim, 2012), tratan la variación de logotipos de imágenes de mercancías. Su estrategia combina el descriptor *DAISY*, histogramas espaciales y detección de regiones salientes. Se obtiene eficiencia de este método al utilizar descriptores tipo *SIFT* como lo es *DAISY*. La clasificación la logran a través de *SVM* con un *kernel* no lineal. El experimento se realiza con una colección de 560 logotipos correspondientes a siete marcas, ochenta imágenes por marca. Los resultados muestran una exactitud promedio del setenta y cinco por ciento.

(Bagheri & Qigang, 2012), abordan el problema de la clasificación de múltiples clases con el uso de un método de clasificación de vecinos más cercanos (*KNN*) y de clasificadores binarios como *OVA*, *OVO*, y *ECOC*. El experimento se hace con veinte bases de datos con un promedio de dos mil setecientas imágenes. Cada base de datos tiene en promedio seis clases. La exactitud promedio que se obtiene es del ochenta y dos por ciento.

(Yang, Zhai, Liu, & Li, 2013), describen un método para la identificación de vehículos a través del reconocimiento de sus logotipos. El método se basa en el gradiente de imágenes para identificar la región del logotipo. Se aplica *pixelization* y el algoritmo de *Sobel*. Para la clasificación utilizan *SVM* basado en características de baja frecuencia *DCT*. El conjunto de pruebas consta de mil noventa y seis imágenes de doce clases. Los resultados obtenidos cuentan con una exactitud del noventa y siete por ciento en un tiempo de análisis de treinta milisegundos.

El trabajo de (Ku, Jianghua, & Gui, 2013), se orienta al reconocimiento de logotipos en televisión por el métodos de *pixel-wise template matching*. La identificación de la región de los logotipos se realiza en tiempo real. Posteriormente se aplica el algoritmo de *SUSAN (Smallest Univalve Segment Assimilating Nucleus)* para extraer las esquinas de las regiones donde se ubica el logotipo. Por último se realiza el emparejamiento de esquinas para identificar los logotipos. Los resultados experimentales se realizan para ocho logotipos de televisión. El video se toma de fuentes como: grabaciones, internet, capturas, etcétera. Los resultados que obtienen alcanzan el setenta y siete por ciento de exactitud.

Para el reconocimiento de logotipos de vehículos, (Yu, Zhang, Yang, & Liang, 2013), representan a las imágenes como histogramas de palabras visuales. La clasificación basada en *SVM* consta de tres fases: extracción de características con *sense-SIFT*, cuantización a través de *SIFT-assignmnet*, y por último, construcción de histogramas de palabras visuales con información espacial. Su propuesta se evalúa con un conjunto de ochocientos cuarenta logotipos de baja resolución distribuidos en catorce clases. Los resultados ofrecen una exactitud del noventa y siete por ciento con un tiempo de reconocimiento de veintidós milisegundos.

(Pan, Yan, Xu, & Sun, 2013), expone el problema de reconocimiento de logotipos con una red neuronal *CNN (Convolutional Neural Network)* y la compara con *SIFT*. El experimento se realiza con un conjunto de pruebas de cinco mil doscientas imágenes para reconocer veintiséis clases. Para *CNN* se alcanza una exactitud promedio de noventa y nueve por ciento en tanto que para *SIFT* es del noventa por ciento.

(Sun, Lu, & Hu, 2014), aplican la detección y reconocimiento de logotipos en cámaras de vigilancia de vías carreteras. El método que proponen utiliza primeramente *Adaboost* y *LBP (Local Binary Patterns)* para reducir el área de búsqueda del logotipo en la imagen. Posteriormente realizan la clasificación por histogramas orientados a gradientes (*HOG*) y máquinas de soporte vectorial (*SVM*). El conjunto de pruebas consta de cerca de mil cuatrocientas imágenes correspondientes a quince clases. Un promedio de noventa por ciento es la exactitud que obtienen para imágenes en tres condiciones de luz: soleado, oscuro y lluvioso.

(Kalaiyarasi & Karthikeyan, 2014), proponen una mejora al algoritmo de *SIFT* basado en componentes principales (*PCSIFT*). En un experimento donde se toman veinte logotipos para su reconocimiento, la exactitud mostrada de *PCSIFT* es de setenta y cinco por ciento, que supera a la de *SIFT* reportada en sesenta y nueve por ciento.

La detección y reconocimiento de logotipos que describen (Xiao, Xiang, & Yuncai, 2015), propone un algoritmo para reconocer logotipos en tiempo real. Construye una máquina de soporte vectorial multi-clase ponderada basada en las características de histogramas de nitidez. El conjunto de datos utilizado es de ochocientos imágenes correspondientes a ocho clases. Cada clase tiene cien

imágenes. Los resultados promediados muestran una exactitud del ochenta y dos por ciento.

Un análisis de estos trabajos que permita distinguir claramente el número de clases que utilizan así como la exactitud que logran se presenta en la Tabla 2.5.1.

Tabla 2.5.1. Análisis de los datos obtenidos de los trabajos relacionados. Se presenta el número de clases y la exactitud lograda (en porcentaje).

Trabajo	Técnica	Número de clases	Exactitud
(Yunqiong, Zhifang, & Fei, 2007)	TM	17	90
(Phan & Dimitrios, 2010)	CECH	4	92
(Psylos, Christos-Nikolaos, & Eleftherios, 2010)	SIFT	10	91
(Pourghassem, 2012)	MLP	7	94
(Zhang, Zhang, Liang, & Liang, 2012)	CEGCH	14	70
(Lei, Chen, & Lim, 2012)	DAISY	7	75
(Bagheri & Qigang, 2012)	OVA	6	82
(Yang, Zhai, Liu, & Li, 2013)	SVM	12	97
(Ku, Jianghua, & Gui, 2013)	SVM	8	77
(Yu, Zhang, Yang, & Liang, 2013)	SIFT	14	97
(Pan, Yan, Xu, & Sun, 2013)	CNN	26	99
(Sun, Lu, & Hu, 2014)	SVM	15	90
(Kalaiyarasi & Karthikeyan, 2014)	PCSFIT	20	75
(Xiao, Xiang, & Yuncai, 2015)	HOG	8	82

En la Tabla 2.5.1 se incluyen aquellos trabajos relacionados que de manera explícita describen los valores de número de clases y de exactitud alcanzada. No están por tanto, todos los trabajos mencionados.

En esta misma tabla se puede observar que al calcular los promedios de cada variable, se obtiene un promedio de doce clases y una exactitud del ochenta y seis por ciento. Con base en estos promedios, se propone que el presente

trabajo de tesis opere con veinte clases y, compare la exactitud lograda con el mostrado en esta tabla.

Capítulo 3 Metodología

En este capítulo se describe la manera en que se desarrollaron las diferentes etapas encaminadas a la consecución de los objetivos planteados. Para ello se emplearon diferentes tipos de metodología y se determinó para cada una de ellas las principales variables que se ven involucradas.

3.1 Metodología empleada

Dada la naturaleza de este trabajo y de acuerdo con los objetivos que se plantearon, para la primera parte, correspondiente al primer objetivo específico, se eligió un tipo de metodología descriptiva. A partir de este estudio se determinaron las técnicas de visión por computadora a utilizar.

Para el segundo objetivo se planteó un enfoque cualitativo en el que se mide la exactitud de la detección de logotipos comerciales en imágenes.

Por último, para el tercer objetivo se aplicó una metodología también cualitativa que evalúa la usabilidad de la aplicación por medio de escalas de Likert sobre un conjunto de criterios definidos por (Nielsen, 1995) y (de Paiva Guimaraes & Farinazzo Martins, 2014).

3.1.1 Descripción del sistema

Dado que no existe sistema experimental para realizar el estudio, en el tercer objetivo se implementa un mundo virtual que representa a un centro comercial y que contiene aquellas tiendas de las marcas cuyos logotipos se pretenden reconocer.

El sistema se puede recorrer por medio de un casco de realidad virtual y cuenta con varios elementos gráficos que generan la impresión de estar en un

mundo real. El usuario puede desplazarse dentro de él. Por tanto, podrá acercarse a los escaparates para efectuar el reconocimiento de logotipos.

3.1.2 Variables consideradas

La Tabla 3.1.2.1 presenta las variables que considera este trabajo así como su tipo y descripción.

Tabla 2.5.1. Variables consideradas.

Nombre de variable	Tipo de variable	Descripción
Número de clases	Independiente	Número de logotipos que se reconocerán
Exactitud	Dependiente	Porcentaje de exactitud de detección
Tiempo de detección	Dependiente	Tiempo que se toma en realizar la detección
Vecinos mínimos	Independiente	Número de vecinos
Factor de escala	Independiente	Factor para realizar la búsqueda de un logotipo en una imagen

3.1.3 Instrumentos de recabado de datos y análisis

El recabado de datos se conforma con los resultados obtenidos a partir de experimentos. Éstos son descritos en la sección 3.3.2.

El recabado consiste de valores de exactitud y tiempo de detección que son registrados en hojas de cálculo para su análisis posterior.

El análisis consiste en seleccionar de cada experimento los parámetros que hayan proporcionado la mayor exactitud. Sin embargo, si en un experimento se encuentran exactitudes similares, el criterio de selección se aplica en elegir los parámetros cuyos resultados se obtengan con el menor tiempo de detección.

3.2 Evaluación de las técnicas de visión por computadora

En la sección 2.5 se revisaron diferentes técnicas que se aplican al reconocimiento de logotipos. El problema que se propone se enfoca en encontrar dos valores fundamentales: la exactitud y el tiempo de detección.

El trabajo de (Sun, Lu, & Hu, 2014), propone la utilización de *Adaboost* y obtiene una exactitud del noventa por ciento sobre un conjunto de mil

cuatrocientas imágenes distribuidas en quince clases. Por lo que se pretendió obtener resultados similares que permitieran evaluar la exactitud de un clasificador basado en *LBP*.

Por otra parte, se planteó analizar la técnica de *Haar-like Features* (*Haar*) para comparar la utilidad de este método en el reconocimiento de logotipos, aun cuando se haya empleado para el reconocimiento de rostros (Medina Nieto, de la Calleja, Medina, & Benitez Ruiz, 2016).

Los resultados que se obtengan tanto de exactitud como de tiempo de detección permitirán corroborar las técnicas de *LBP* y *Haar* como las que se emplearán en este trabajo. Otra cuestión que se tomó en cuenta, fue la existencia de las implementaciones de *LBP* y *Haar* dentro del conjunto de librerías de *OpenCV*.

Es por ello que, para esta sección se realiza un análisis que permite determinar si las técnicas de *LBP* y de *Haar* serán las adecuadas o no. El criterio de aceptación consiste en lograr que los resultados obtenidos sean similares a aquellos mencionados en los trabajos relacionados de la Tabla 2.5.1.

3.3 Generación del reconocedor

El reconocedor establece su función como el proceso que toma como entrada una imagen, la analiza y determina si en ella se encuentra algún logotipo. En caso afirmativo establece la identidad del logotipo.

Su implementación incorpora aprendizaje automático supervisado, por lo que los logotipos que identifica están limitados. Por tanto no puede identificar logotipos que no se encuentran en el conjunto de logotipos seleccionados.

El reconocedor está compuesto de varios clasificadores binarios y, cada uno de ellos es entrenado para reconocer un solo logotipo. El diseño de la implementación del reconocedor se presenta en la Figura 3.3.1.

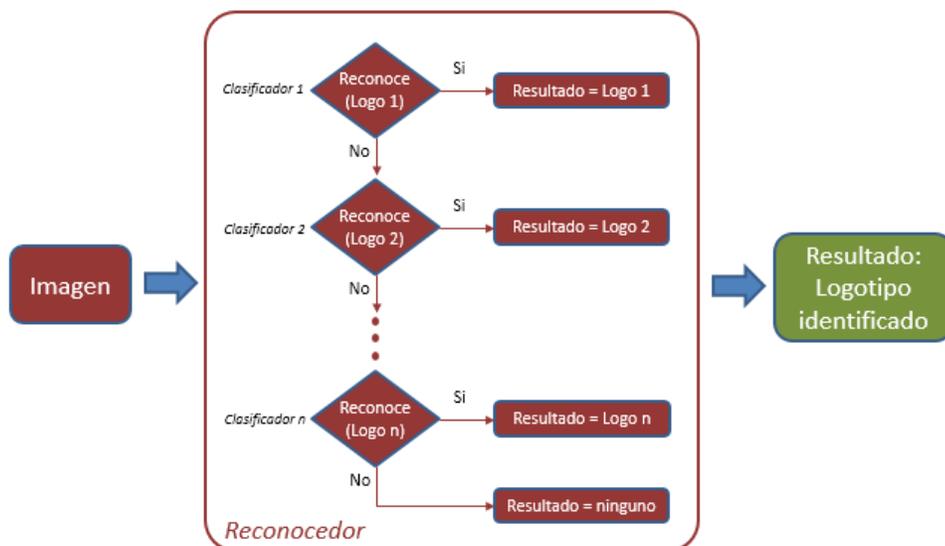


Figura 3.3.1. Diseño de implementación del reconocedor.

Como se muestra en la Figura 3.3.1, el resultado corresponde al logotipo que ha sido identificado por alguno de los clasificadores.

Recuérdese que se determinó que el reconocedor puede identificar a veinte logotipos. Esto quiere decir que el reconocedor se compone de veinte clasificadores; cada uno capaz de reconocer a un solo logotipo. Por tanto, primero se deben establecer los veinte logotipos que se identificarán y después, se generan sus respectivos clasificadores.

3.3.1 Logotipos que se identifican

Los logotipos corresponden a marcas comerciales de tiendas de centros comerciales. Para determinar los logotipos que son identificados por el reconocedor se sigue el proceso mostrado en la Figura 3.3.1.1.

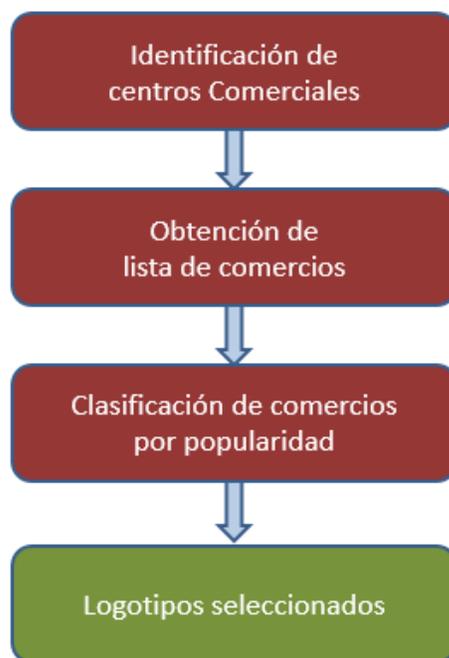


Figura 3.3.1.1. Diagrama de bloques de selección de logotipos.

Como se muestra en la Figura 3.3.1.1, el resultado es el conjunto de veinte logotipos que el reconocedor debe identificar. En otras palabras, corresponde a la elección de las veinte tiendas de centros comerciales que más presencia tienen.

3.3.2 Generación de los clasificadores

En la Figura 3.3.1 se describió que el reconocedor está compuesto por clasificadores. Cada uno se entrena para reconocer un logotipo, por lo que es necesario generar veinte clasificadores. Por otro lado, las técnicas de visión que se evaluaron fueron *Haar* y *LBP*. Aun cuando solo una de ellas se integre en el reconocedor, se deben generar clasificadores para las dos y determinar cuál de ellas tiene el más alto valor de exactitud. Esto hace que existan dos clasificadores para cada logotipo: uno entrenado con base en *Haar* y otro entrenado en *LBP*, lo que da un total de cuarenta clasificadores.

Cada clasificador se genera de la misma manera. Se requiere de dos conjuntos de imágenes, un conjunto de entrenamiento y un conjunto de pruebas con el que se obtiene su exactitud y tiempo de detección.

El conjunto de entrenamiento está compuesto a su vez de dos conjuntos: el conjunto de ejemplos positivos, que consta de imágenes que contienen el objeto a reconocer y, el conjunto de ejemplos negativos, compuesto por imágenes arbitrarias cuyo requisito es que no contengan al objeto a reconocer.

De igual manera, el conjunto de pruebas también está compuesto de ejemplos positivos y ejemplos negativos. El clasificador generado se aplica al conjunto de prueba para medir su exactitud.

3.3.2.1 Entrenamiento de un clasificador

Para propósitos de este trabajo, se utiliza el mismo conjunto de ejemplos negativos para todos los clasificadores. Este conjunto está compuesto por 1000 ejemplos negativos de un tamaño de 160 pixeles de ancho por 200 pixeles de alto. En la Figura 3.3.2.1.1 se presentan algunos elementos.



Figura 3.3.2.1.1. Algunas imágenes del conjunto de ejemplos negativos para entrenamiento.

El conjunto de ejemplos positivos de cada clasificador cuenta con 2000 imágenes de aproximadamente 50 pixeles de ancho por 50 pixeles de alto; aunque estas dimensiones varían según la forma de cada logotipo. Estas imágenes se generan a partir de la imagen de una o varias versiones de un logotipo. Las imágenes resultantes corresponden a variaciones del logotipo original en rotación, enfoque, iluminación y deformación de perspectiva. Estas deformaciones se generan con la herramienta *createsamples* de OpenCV con base en los parámetros mostrados en la Tabla 3.3.2.1.1.

Tabla 3.3.1.1. Parámetros para la creación de ejemplos positivos.

Parámetro	Valor
Angulo de rotación en el eje X	34°
Angulo de rotación en el eje Y	17°
Desviación máxima de intensidad	100
Color de fondo	Negro
Umbral de color de fondo	0

Se especifican también parámetros para la dimensión de los ejemplos positivos: alto y ancho. Sin embargo, estos valores varían para cada logotipo con respecto a su forma. Cabe resaltar que ninguno supera los 50 píxeles, ni en alto, ni en ancho.

Con base en los parámetros anteriores, se puede observar en la Figura 3.3.2.2 algunos ejemplos que se generaron para el logotipo de HSBC.

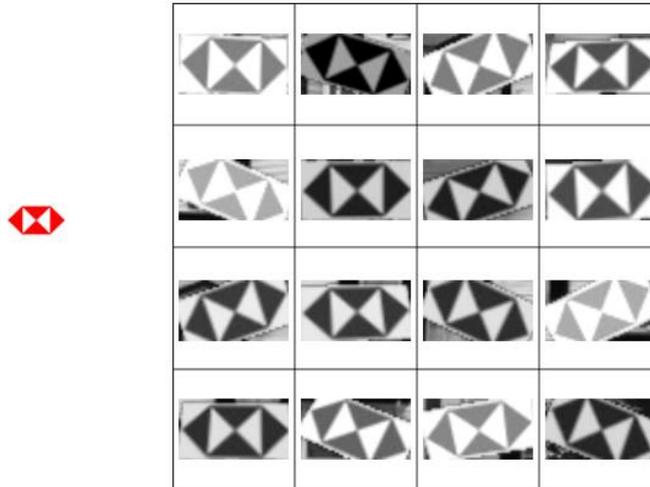


Figura 3.3.2.2. Izquierda, logotipo original. Derecha, algunos ejemplos positivos del logotipo HSBC para entrenamiento.

En algunos casos la generación de los ejemplos positivos se realizó a partir de varias imágenes de un mismo logotipo. En la Figura 3.3.2.3 se presenta un ejemplo de estas imágenes correspondientes al logotipo de Devlyn. Los ejemplos positivos que se generan a partir de estas imágenes se unen en un solo archivo que es procesado durante el entrenamiento. Para realizar esta unión, se utilizó la herramienta *mergevec* de (Wulfe, 2015).



Figura 3.3.2.3. Varias imágenes para el logotipo de Devlyn.

Para entrenar los clasificadores se hace uso de la herramienta *traincascade* de OpenCV. Los clasificadores se entrenaron de acuerdo con los parámetros mostrados en la Tabla 3.3.2.5.

Tabla 3.3.2.5. Parámetros para el entrenamiento de cada clasificador

Parámetro	Clasificador <i>Haar</i>	Clasificador <i>LBP</i>
Ejemplos positivos	1000	2000
Ejemplos negativos	400	1000
Exactitud esperada	0.9995	0.9995
Número de etapas	4	4
Tasa de detección de negativos	0.000001	0.000001

El hecho de que el conjunto de entrenamiento para los clasificadores *Haar* sea menor se debe a dos factores: tiempo y funcionalidad. Ya que, en pruebas iniciales se notó que entrenamientos con 2000 ejemplos positivos y 1000 negativos tardaban de 15 a 20 días por logotipo. Más aún, algunos de estos entrenamientos detenían su ejecución sin alcanzar la exactitud esperada, o bien, algunos detenían su ejecución. Lo que provocaba que el tiempo invertido se desaprovechara por no conseguir el resultado deseado. En cambio, al disminuir el conjunto de entrenamiento, se pudieron generar los clasificadores y la exactitud que se obtuvo se consideró aceptable.

3.3.2.2 Prueba de un clasificador

La exactitud de un clasificador se puede expresar a través de la matriz de confusión de la Tabla 3.3.2.2.1.

Tabla 3.3.2.2.1. Matriz de confusión

	Positivo Predicho	Negativo Predicho
Positivo	TP	FN
Negativo	FP	TN

El resultado de realizar la detección de un objeto en una imagen se contabiliza en una de cuatro condiciones. Si el clasificador reconoce un objeto y en la imagen se encuentra dicho objeto, se obtiene un verdadero positivo (TP). Si la imagen contiene al objeto, pero el clasificador no lo reconoce, se obtendrá entonces un falso negativo (FN). Si por el contrario, la imagen no contiene al objeto, pero el clasificador propone que sí, se obtiene entonces un falso positivo (FP). Por último, si el clasificador analiza una imagen que no contiene al objeto y la detección determina que el objeto no está presente, entonces se tiene un verdadero negativo (TN).

Lo anterior indica que un buen clasificador será aquel que contabiliza en su mayoría valores TP y TN. De manera más específica, la exactitud (*accuracy*) de un clasificador se puede conocer mediante la aplicación de la Ecuación 9 sobre el conjunto de pruebas.

$$Accuracy = \frac{tp + tn}{tp + tn + fp + fn} \quad \text{Ec. (9)}$$

Los valores obtenidos por esta ecuación toman un rango de 0 a 100; un valor cercano al cero indicará que el clasificador no puede reconocer al logotipo que se quiera detectar. Por el contrario, un valor cercano al cien indicará que el clasificador tiene una alta exactitud.

La prueba de un clasificador se realiza sobre el conjunto de pruebas, cuyo conjunto de ejemplos negativos es el mismo para todos los clasificadores. Consta de 250 imágenes del mundo real cada una de 640 por 480 pixeles. Estas imágenes corresponden a fotografías de centros comerciales que no contienen a ninguno de los logotipos seleccionados. En la Figura 3.3.2.3 se presentan algunas de las imágenes que comprenden a este conjunto.



Figura 3.3.2.3. Algunas imágenes del conjunto de ejemplos negativos para prueba.

Los ejemplos positivos del conjunto de pruebas son imágenes del mundo real que contienen uno o más logotipos. Cada clasificador se prueba con 250 ejemplos positivos de dimensiones de 640 por 480 pixeles y corresponden a fotografías de centros comerciales que incluyen el logotipo a identificar. En la

Figura 3.3.2.4 se presentan algunos de los ejemplos positivos para los clasificadores de los logotipos Ferrioni, Adidas, Puma y Tous.



Figura 3.3.2.4. Ejemplos positivos del conjunto de prueba para logotipos Ferrioni, Adidas, Puma y Tous.

Las pruebas de los clasificadores se realizan a través de los siguientes experimentos.

3.3.2.3 Experimentos de prueba

Los resultados de exactitud de un clasificador dependen de los parámetros que se especifiquen al realizar la detección. Estos parámetros corresponden a las variables de factor de escala y de vecinos mínimos. El factor de escala permite reducir o aumentar la escala de la imagen de entrada de tal manera que se puedan reconocer objetos de distinto tamaño. Por otra parte, el valor de vecinos mínimos se relaciona con la calidad de los objetos detectados, un valor alto implica menos detecciones pero de mayor calidad y, un valor pequeño implica más detecciones pero de menor calidad.

Los experimentos se conforman de varias ejecuciones de detección sobre el conjunto de pruebas con diferentes valores de factor de escala y de vecinos mínimos, buscando obtener las exactitudes más cercanas al cien por ciento.

En el Apéndice A se muestran los registros de los experimentos que se realizaron para cada clasificador.

3.4 Escenario de pruebas

Como se mencionó en la sección 3.1.1, no existe sistema experimental para realizar el estudio que propone este trabajo. Es por ello que se implementa un escenario de pruebas, en este caso, un mundo virtual que representa a un centro comercial y que contiene aquellos locales o tiendas que presentan los logotipos que se pretenden reconocer.

El escenario de pruebas consiste en una aplicación donde el usuario pueda hacer un recorrido virtual dentro de un centro comercial. Su generación se lleva a cabo mediante los pasos mostrados en la Figura 3.4.1.



Figura 3.4.1. Pasos para generar el escenario de pruebas.

En las siguientes secciones se describe la función de cada uno de los pasos de la Figura 3.4.1.

3.4.1 Integración y generación de contenido 3D

En este paso se realiza la implementación del mundo virtual sobre una herramienta de creación de videojuegos. El primer paso consiste en incluir el modelo 3D que represente al centro comercial. Posteriormente, se crean los siguientes componentes necesarios para que un jugador se desplace sobre el mundo virtual:

- Creación de un plano de colisión
- Creación de límites de colisión
- Escalamiento

Por último, se crean todos los elementos de iluminación que propicien un recorrido virtual.

3.4.2 Integración de imágenes del conjunto de pruebas

Puesto que el centro comercial virtual debe contar con las tiendas correspondientes a los logotipos que se deben reconocer, se insertan en el modelo 3D las imágenes del conjunto de pruebas.

3.4.3 Implementación del recorrido inmersivo

El mundo virtual se recorre por medio de un casco de realidad virtual de tal manera que un usuario puede desplazarse dentro de él y pueda invocar al reconocedor para que identifique los logotipos.

3.4.4 Implementación del reconocedor

La implementación del reconocedor dentro del mundo virtual debe permitir especificar para cada logotipo la técnica de detección y los valores de las variables factor de escala y vecinos mínimos. Tener el control de estas variables permite ajustar los clasificadores en caso de ser necesario.

3.4.5 Evaluación del sistema

Para evaluar el sistema se aplica una evaluación heurística. Esta evaluación utiliza un pequeño número de evaluadores que realizan una prueba de usabilidad a una interfaz de usuario y permite documentar errores de software basados en principios de diseño (Mohd. Zainuddin, Badioze Zaman, & Ahmad, 2011).

La evaluación permite registrar las validaciones obtenidas para tener una referencia de los puntos que se pueden mejorar en la aplicación. Las cuestiones propuestas basadas en las heurísticas de (Nielsen, 1995), se muestran en la Tabla 3.4.5.1.

Tabla 3.4.5.1. Heurísticas de Nielsen (Nielsen, 1995).

-
1. Visibilidad del estado del sistema
 2. Correspondencia entre el sistema y el mundo real
 3. Control del usuario y libertad
 4. Consistencia y estandarizaciones
 5. Prevención del error
 6. Reconocer en lugar de recordar
 7. Flexibilidad y eficiencia de uso
 8. Diseño estético y minimalista
 9. Ayuda al usuario para reconocer, diagnosticar y recuperarse de errores
 10. Ayuda y documentación
-

Puesto que la aplicación desarrollada se enfoca en la realidad aumentada, se agregan las heurísticas adicionales que propone (de Paiva Guimaraes & Farinazzo Martins, 2014), mismas que se presentan en la Tabla 3.4.5.2.

Tabla 3.4.5.2. Heurísticas adicionales para aplicaciones de realidad aumentada (de Paiva Guimaraes & Farinazzo Martins, 2014).

-
11. Exactitud
 12. Preparación del ambiente
 13. Satisfacción
-

La ponderación para evaluar el sistema con base en las heurísticas se determina de acuerdo con la escala de Likert mostrada en la tabla 3.4.5.3.

Tabla 3.4.5.3. Escala de Likert empleada en la evaluación heurística del sistema.

Puntaje	Significado
1	Totalmente en desacuerdo
2	En desacuerdo
3	Ni de acuerdo ni en desacuerdo
4	De acuerdo

3.5 Implementación en un dispositivo móvil

La aplicación final hará uso del reconocedor en un dispositivo móvil. Lo que aquí se propone reproducir la implementación del reconocedor y la operatividad del escenario en un dispositivo móvil.

Si bien en el escenario de pruebas se utiliza una cámara virtual para la captura de imágenes para el reconocimiento de logotipos, en un dispositivo móvil se utiliza la cámara del dispositivo. En este sentido, las imágenes que se capturen corresponden a imágenes del mundo real.

Capítulo 4 Resultados

En este capítulo se describen los resultados obtenidos a partir de la metodología planteada.

4.1 Evaluación de las técnicas de visión por computadora

El artículo de investigación de (Medina Nieto, de la Calleja, Medina, & Benitez Ruiz, 2016), permitió realizar el análisis de las técnicas *LBP* y *Haar* para el reconocimiento de logotipos. Éste comprendió a cuatro logotipos. Los resultados de exactitud y tiempo de detección que se obtuvieron se presentan en la Tabla 4.1.1.

Tabla 4.1.1. Comparación de resultados obtenidos para determinar las técnicas de visión por computadora a emplear en este trabajo. Exactitud (en porcentaje) y tiempo de detección (en milisegundos).

Logotipo	<i>Haar</i>		<i>LBP</i>	
	Exactitud	Tiempo de detección	Exactitud	Tiempo de detección
Adidas	94	11	71	8
Ferrioni	92	11	73	8
Puma	56	12	50	10

Tous	94	12	74	10
------	----	----	----	----

Como se aprecia en la Tabla 4.1.1, los valores de exactitud obtenidos para *Haar* son superiores a los obtenidos para *LBP*. Sin embargo, *LBP* demostró ser más rápido para realizar la detección. Por otra parte, puede observarse que los valores de exactitud obtenidos con *Haar*, a excepción del logotipo Puma, son semejantes a los trabajos de (Phan & Dimitrios, 2010) y de (Pourghassem, 2012); en tanto que los valores obtenidos con *LBP* son similares a los de (Zhang, Zhang, Liang, & Liang, 2012), (Lei, Chen, & Lim, 2012) y (Kalaiyarasi & Karthikeyan, 2014).

Para el caso del logotipo Puma, se planteó que era posible obtener una exactitud mayor al emplear un número mayor de imágenes para entrenamiento que el descrito en (Medina Nieto, de la Calleja, Medina, & Benitez Ruiz, 2016). Esto se logró y la mejora se presenta en la Tabla 4.2.2.3.

De esta manera y de acuerdo con el trabajo de (Medina Nieto, de la Calleja, Medina, & Benitez Ruiz, 2016) se pudo establecer las técnicas *Haar* y *LBP* como aquellas a emplear en este trabajo y, con esto, se puede afirmar que se cubrió el primer objetivo específico.

4.2 Generación del reconocedor

4.2.1 Logotipos que se identifican

Los centros comerciales que se analizaron son los quince más valorados de México de acuerdo con (Hispanic Searching, 2015), éstos se muestran en la Tabla 3.3.1.1.

Tabla 3.3.1.1. Lista de los centros comerciales más valorados de México de acuerdo con (Hispanic Searching, 2015).

Centro comercial	Posición de popularidad
La Isla Shopping Village	1
Centro Comercial Santa Fé	2
Centro Comercial Andares	3
Multiplaza Lindavista	4
Centro Comercial Villasunción	5
Centro Comercial Plaza Monarca	6
Centro Comercial Galería Cuernavaca	7
Centro Comercial Pabellón Cuauhtémoc	8
Centro Comercial Plaza del Sol	9
Parque Tezontle	10
Centro Comercial Plaza Satélite	11
La Gran Plaza Fashion Mall	12
Centro Comercial Perisur	13

Centro Comercial Perinorte	14
Centro Comercial Galerías Atizapán	15

De cada centro comercial mostrado en la Tabla 3.3.1.1, se obtuvo la lista de las tiendas que lo conforman. Reunidas éstas, se registró el número de veces que aparecen, Lo que permite identificar aquellas con mayor presencia. En la Tabla 4.2.1.1, se muestran todas las tiendas con su número de aparición.

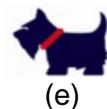
Tabla 4.2.1.1. Tiendas de los centros comerciales populares en México y su número de aparición.

Tiendas	Número de aparición
Telcel	11
Martí	10
GNC	10
Devlyn	10
Zara	10
Radioshack	9
Sunglasshut	9
Aldo Conti	9
Julio	9
Adidas	9
Nutrisa	8
Starbucks	8
Subway	8
Taco Inn	8
Mistertennis	8
LUX	8
Berksha	8
Oysho	8
Pull & Bear	8
Scappinno	8
Shasa	8
Todomoda	8
Sanborns	8
Capa de Ozono	8
Flexi	8
BBVA Bancomer	7
Chillis	7
Crystal Joyas	7
+ Kota	7
Ferrioni	7
LOB	7
Massimo Dutti	7
Liverpool	7
Game Planet	7
CI Banco	6
HSBC	6

Santander	6
Tous	6
Cinépolis	6
Burger King	6
iShop	6
Mobo Shop	6
Steren	6
The Body Shop	6
L'Occitane	6
Campanita	6
Docker's	6
Ivonne	6
Levis Shop	6
Liz Minelli	6
Sfera	6
Studio F	6
Women's Secret	6
Iusacell	6
Dorothy Gaynor	6
Georgie Boy	6
Michel Domit	6
Nine West	6
Nike	6
Puma	6
Vans	6

Cada tienda de la Tabla 4.2.1.1 tiene su propio logotipo y, cada logotipo representa a una clase. El análisis que se realizó consistió en seleccionar los veinte logotipos de las tiendas con mayor número de aparición, en el caso de las tiendas que presentaron el mismo número de aparición, se excluyeron aquellos logotipos que contuvieran solo texto. Así también se excluyeron con base en su forma y combinación de colores.

El resultado de este proceso se conforma de los veinte logotipos que son identificados por el reconocedor; mismos que se muestran en la Figura 4.2.1.1.



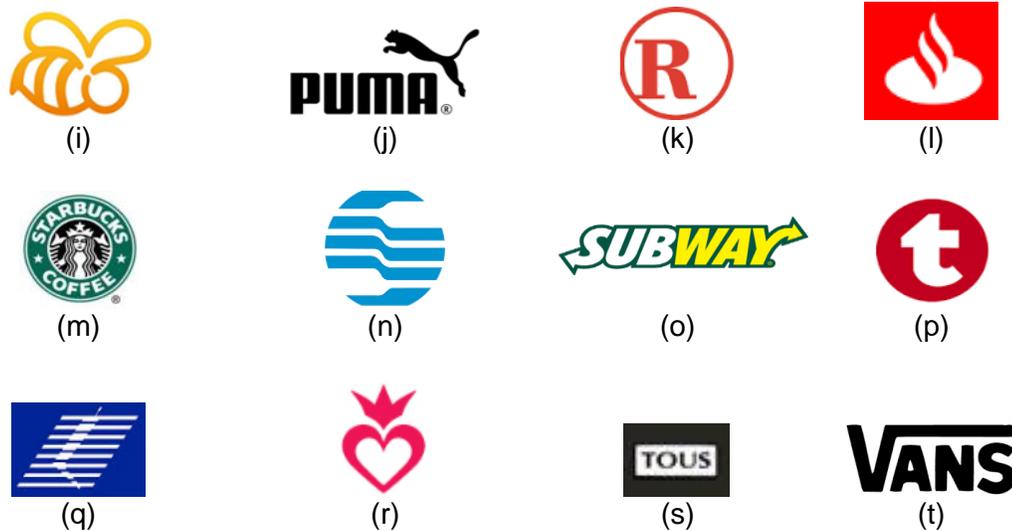


Figura 4.2.1.1. Logotipos seleccionados para ser identificados por el reconecedor. Adidas (a), Burger King (b), Cinépolis (c), Devlyn (d), Ferrioni (e), HSBC (f), Liverpool (g), Martí (h), Nutrisa (i), Puma (j), Radioshack (k), Santander (l), Starbucks (m), Steren (n), Subway (o), Taco Inn (p), Telcel (q), Todomoda (r), Tous (s), Vans (t).

Cae señalar las marcas y logotipos que se presentan en este trabajo han sido utilizados únicamente con fines de investigación. Los logotipos utilizados son marcas registradas de sus propietarios y no se pretende de ninguna manera afectar a las empresas ni tampoco verse beneficiados por su utilización.

Como se muestra en la Figura 4.2.1.1, los logotipos elegidos tienen formas variadas y no forman parte de este conjunto logotipos de tiendas como GNC ó Zara, que aun cuando tienen número de aparición de nueve, de los más altos, su logotipo consiste solo de texto.

Con los logotipos ya definidos, se procedió a generar los clasificadores.

4.2.2 Generación de los clasificadores

A partir del conjunto de entrenamiento descrito en la sección 3.3.2.1, y después de utilizar la herramienta *Traincascade* se entrenaron los clasificadores. Con ello, se obtuvieron cuarenta clasificadores, veinte clasificadores entrenados con la técnica *Haar* y veinte con la técnica *LBP*.

Se observó que el tiempo que tarda el proceso de entrenamiento depende del tamaño y forma de las imágenes del conjunto de entrenamiento y del equipo de cómputo que se utilice para este propósito. En general, durante los entrenamientos iniciales, se observó que la duración con *LBP* radicaba entre

quince minutos y dos días. Con respecto al entrenamiento con *Haar*, los procesos de entrenamiento tardaron entre uno y doce días.

Para disminuir el tiempo de entrenamiento, se utilizaron cinco equipos de cómputo con características diferentes. Como se esperaba, los equipos con mayor capacidad de cómputo entrenan a los clasificadores de una manera más rápida que los equipos de menor capacidad.

Posteriormente se realizaron los experimentos sobre el conjunto de prueba para determinar la exactitud de cada clasificador. Este proceso se llevó a cabo de forma individual para cada logotipo y para cada técnica. La ejecución de los experimentos se llevó a cabo en el lenguaje de programación *Python* integrado con las librerías de *OpenCV*. En el Anexo puede consultar cómo se llevó a cabo esta implementación.

En las Tablas 4.2.2.1 y 4.2.2.2, se presentan los resultados de los experimentos para el clasificador Adidas. Recuérdese que cada experimento se basa en la especificación del factor de escala y de vecinos mínimos y, se registra la exactitud y el tiempo de detección obtenido sobre el conjunto de pruebas.

Tabla 4.2.2.1. Resultados de los experimentos de detección del clasificador Ferrioni basado en la técnica *Haar*. Exactitud (en porcentaje) y tiempo de detección (en milisegundos).

Factor de escala	Vecinos mínimos	TP	FP	FN	TN	Exactitud obtenida	Tiempo de detección
1.2	1	222	28	0	250	94.40	12.15
1.1	1	228	2	5	245	98.54	20.91
1.3	1	174	75	0	250	84.97	7.96
1.15	3	231	19	2	248	95.80	13.42
1.02	1	250	0	43	207	91.40	91.20

Tabla 4.2.2.2. Resultados de los experimentos de detección del clasificador Ferrioni basado en la técnica *LBP*. Exactitud (en porcentaje) y tiempo de detección (en milisegundos).

Factor de escala	Vecinos mínimos	TP	FP	FN	TN	Exactitud obtenida	Tiempo de detección
1.01	5	129	121	40	210	67.80	125.00

1.01	3	162	88	52	198	72.00	122.54
1.01	1	229	21	142	108	67.40	120.62
1.02	1	188	62	85	165	70.45	65.34
1.02	3	114	136	37	213	65.40	62.76

Como puede observarse en las Tablas 4.2.2.1 y 4.2.2.2, para el caso del logotipo Ferrioni, se logra una mayor exactitud con la técnica *Haar* que con la técnica *LBP*, así como también un menor tiempo de detección. De esto se deduce que el reconocedor incorporará, solo para el logotipo Ferrioni, la técnica *Haar* y, los valores de factor de escala y vecinos mínimos de 1.1 y 1 respectivamente.

Puede notarse que la exactitud obtenida con la técnica *Haar* es semejante a la que obtuvieron (Yang, Zhai, Liu, & Li, 2013), (Yu, Zhang, Yang, & Liang, 2013) y (Pan, Yan, Xu, & Sun, 2013).

Cabe señalar, que los experimentos no fueron consistentes para todos los clasificadores. Para ciertos logotipos, se encontró que *LBP* obtuvo una exactitud mayor a la lograda por *Haar*. En otros casos, la exactitud obtenida fue la misma para ambas técnicas, en este caso, se optó por seleccionar el reconocedor que mostró menor tiempo de ejecución.

Las Tablas 4.2.2.1 y 4.2.2.2, presentan los resultados de los experimentos de dos clasificadores. En el Apéndice A se presentan los resultados de los experimentos de detección de los todos los clasificadores.

En la Tabla 4.2.2.3, se reúnen los resultados de los experimentos sobre los cuarenta clasificadores en los que se obtuvo la mayor exactitud. Se presenta la comparación entre ambas técnicas y se distingue en negritas la mejor exactitud.

Tabla 4.2.2.3. Comparación de la exactitud obtenida de los clasificadores entre las técnicas *Haar* y *LBP*. Exactitud (en porcentaje) y tiempo de detección (en milisegundos).

Clasificador	<i>Haar</i>		<i>LBP</i>	
	Exactitud	Tiempo de detección	Exactitud	Tiempo de detección
Adidas	79.40	23.55	95.00	87.74
Burger King	78.60	229.38	96.80	134.89
Cinépolis	87.00	45.79	93.80	141.06
Devlyn	99.80	6.74	100.00	7.50

Ferrioni	98.54	20.91	72.00	122.54
HSBC	77.40	7.67	89.80	63.80
Liverpool	79.20	34.35	95.80	193.37
Martí	95.60	9.93	98.60	105.05
Nutrisa	64.20	104.81	98.80	194.36
Puma	91.00	10.19	97.20	165.04
Radioshack	97.60	11.33	99.00	143.74
Santander	99.80	20.83	81.40	124.07
Starbucks	84.00	52.91	99.00	33.60
Steren	78.80	200.59	85.20	132.69
Subway	94.20	60.53	94.00	100.36
Taco Inn	99.20	10.50	94.80	121.79
Telcel	82.80	14.03	83.40	146.57
Todomoda	74.80	28.24	84.20	29.00
Tous	81.80	53.55	88.20	41.82
Vans	83.20	52.70	90.00	86.49
Promedio	86.35	49.92	91.85	108.77

Como se aprecia en la Tabla 4.2.2.3, para algunos logotipos una técnica es más exacta que la otra. Esto sugiere que no se puede generalizar que una sea mejor que la otra, ni en exactitud, ni en tiempo de detección. Sin embargo se puede notar que *LBP* mostró ser más exacto para más logotipos, puesto que su exactitud es mayor en dieciséis casos, en tanto que *Haar*, solo en cuatro. Sin embargo, recuérdese que el conjunto de entrenamiento para *Haar* es menor, por lo que tampoco puede asegurarse que *LBP* sea más exacto que *Haar*.

En la Tabla 4.2.2.4 se comparan los resultados que se obtuvieron con respecto a los de la Tabla 2.5.1. Para ello se toman en cuenta los valores promedio. La tabla 2.5.1 se presenta ordenada de manera descendente respecto al valor de exactitud.

Tabla 4.2.2.4. Comparación de la exactitud lograda (en porcentaje) en este trabajo respecto a otros trabajos.

Trabajo	Técnica	Número de clases	Exactitud
---------	---------	------------------	-----------

(Pan, Yan, Xu, & Sun, 2013)	CNN	26	99
(Yu, Zhang, Yang, & Liang, 2013)	SIFT	14	97
(Yang, Zhai, Liu, & Li, 2013)	SVM	12	97
(Pourghassem, 2012)	MLP	7	94
(Phan & Dimitrios, 2010)	CECH	4	92
Este trabajo	LBP	20	91
(Psylos, Christos-Nikolaos, & Eleftherios, 2010)	SIFT	10	91
(Yunqiong, Zhifang, & Fei, 2007)	TM	17	90
(Sun, Lu, & Hu, 2014)	SVM	15	90
Este trabajo	Haar	20	86
(Xiao, Xiang, & Yuncai, 2015)	HOG	8	82
(Bagheri & Qigang, 2012)	OVA	6	82
(Ku, Jianghua, & Gui, 2013)	SVM	8	77
(Kalaiyarasi & Karthikeyan, 2014)	PCSIFT	20	75
(Lei, Chen, & Lim, 2012)	DAISY	7	75
(Zhang, Zhang, Liang, & Liang, 2012)	CEGCH	14	70

Como se observa en la Tabla 4.2.2.4, la exactitud lograda con este trabajo la ubica en resultados similares a otros autores. Una parte a destacar es que con los resultados de este trabajo se logra establecer que la técnica *Haar* puede también ser utilizada para el reconocimiento de logotipos.

Si se observa nuevamente la Tabla 4.2.2.3, puede notarse también que *LBP* obtiene mayores tiempos de detección que *Haar*, lo que lo hace más lento. Dado que el sistema final se ejecuta en un dispositivo móvil, el tiempo de detección es un factor importante a tomar en cuenta. Es por ello que, de los resultados de los experimentos se buscaron otros parámetros que obtuvieran valores aceptables de exactitud, pero con menores tiempos de detección. Como consecuencia, para la implementación del reconocedor se eligieron los clasificadores cuyos valores de factor de escala y de vecinos mínimos disminuyen el tiempo de detección sin afectar la exactitud. Esto da lugar a una nueva comparación entre ambas técnicas mostrada en la Tabla 4.2.2.5.

Tabla 4.2.2.5. Comparación de la exactitud obtenida con peso sobre el tiempo de detección entre las técnicas *Haar* y *LBP* a partir de los resultados de los experimentos. Exactitud (en porcentaje) y tiempo de detección (en milisegundos).

Clasificador	<i>Haar</i>		<i>LBP</i>	
	Exactitud	Tiempo de detección	Exactitud	Tiempo de detección
Adidas	79.40	23.5	92.40	42.82
Burger King	76.40	78.7	94.20	67.9
Cinépolis	87.00	45.79	92.20	70.7
Devlyn	99.80	6.74	100.00	7.5
Ferrioni	95.80	13.42	70.60	65.34
HSBC	76.80	4.36	89.80	63.8
Liverpool	79.20	34.35	93.40	97.93
Martí	94.80	6.99	96.60	52.4
Nutrisa	64.20	104	92.80	90.27
Puma	91.00	10.19	96.60	84.66
Radioshack	97.60	11.33	92.40	71.66
Santander	93.20	6.48	81.40	124.07
Starbucks	84.00	52.91	99.00	33.6
Steren	70.20	58.24	82.40	65.48
Subway	92.20	27.43	93.00	49.97
Taco Inn	99.20	10.5	91.60	61.49
Telcel	82.80	14.03	80.20	72.86
Todomoda	74.40	27.8	84.20	29.00
Tous	81.80	53.55	88.20	41.82
Vans	83.20	52.7	90.00	86.49
Promedio	85.15	32.15	90.05	63.98

Como se observa, los valores de promedio de la Tabla 4.2.2.5, son muy similares a los de la Tabla 4.2.2.3, se puede inferir que no se sacrifica exactitud

y se disminuye considerablemente el tiempo de detección. Elegir los parámetros de factor de escala y vecinos mínimos que producen los resultados marcados en negritas permite que el sistema pueda comportarse manera más eficiente. Por otro lado, a diferencia de lo observado en la Tabla 4.2.2.3 en donde *LBP* demuestra ser más exacto, ahora en la Tabla 4.2.2.5 se presenta cierto equilibrio entre ambas técnicas. Lo que es más, es posible que si el entrenamiento de *Haar* se realiza con un conjunto de entrenamiento más grande, se muestre que *Haar* sea similar a *LBP*.

Puede notarse además en la Tabla 4.2.2.5, que al tomar en cuenta únicamente los valores de exactitud en letras negritas se obtiene en promedio una exactitud, combinada de ambas técnicas, del 92%. Estos valores indican también la técnica que se implementó en el reconocedor. Con ello se definió el diseño del reconocedor mostrado en la Figura 3.3.1. El código de implementación puede consultarse en el Anexo.

A partir de estos últimos resultados se puede observar a los logotipos que son identificados por cada técnica. Estos se presentan en la Tabla 4.2.2.6.

Tabla 4.2.2.6. Comparación de los logotipos identifica cada técnica.

Técnica	Logotipos que identifica
<i>Haar</i>	
<i>LBP</i>	

Como puede notarse en la Tabla 4.2.2.6, los logotipos que identifica la técnica *LBP* tienen ancho y alto similar, son logotipos casi cuadrados. *Haar* en cambio, identifica logotipos de diferentes tamaños. Por otro lado, si se enfoca en las formas de las líneas, curvas, colores, se puede ver que la técnica *Haar* identifica a aquellos que parecer ser simples, en tanto que *LBP* identifica a los más complejos. En cuanto a los logotipos cuyo elemento principal es texto, *Haar* parece identificarlos mejor. Estas observaciones pueden ayudar a decidir la técnica a utilizar cuando se requiera agregar o sustituir un logotipo.

Adicionalmente, en la Tabla 4.2.2.7 se presentan ejemplos de detección sobre el conjunto de pruebas, se puede observar ciertas diferencias en la forma de detectar de ambas técnicas.

Tabla 4.2.2.7. Comparación de las técnicas *Haar* y *LBP* sobre el conjunto de pruebas para cada logotipo.

Logotipo	Clasificador Haar	Clasificador LBP
Adidas		
Burger King		
Cinépolis		
Devlyn		
Ferrioni		

HSBC



Liverpool



Martí



Nutrisa



Puma



Radioshack



Santander



Starbucks



Steren



Subway



Taco Inn



Telcel



Todomoda



Tous



Vans



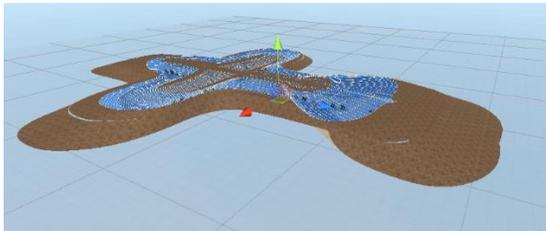
Las imágenes mostradas en la Tabla 4.2.2.7, son similares entre ambas técnicas. Sin embargo, se puede notar que con la técnica *Haar* se detectan un número mayor de falsos positivos que con la técnica *LBP*, es posible que si se incrementa el valor de la variable vecinos mínimos se reduzcan los falsos positivos. Por su parte, la técnica *LBP* muestra ser capaz de identificar más de un logotipo en una sola imagen, cuando éste existe, como es el caso de Burger King, en esa imagen *Haar* solo puede identificar a un solo logotipo. Otra característica que no es muy notable, pero que si se observa con atención se puede distinguir, consiste en que la región de identificación, es decir, el recuadro en color verde, se ajusta más a los logotipos para la técnica *Haar* que para *LBP*.

De esta manera y de acuerdo con los resultados y su análisis presentado en esta sección se puede afirmar que se cubrió el segundo objetivo específico.

4.3 Escenario de pruebas

4.3.1 Integración y generación de contenido 3D

El escenario de pruebas se generó con la herramienta Unity. El modelo de centro comercial que se utilizó se descargó de (Kakasucai, 2015) que es un sitio web que ofrece modelos 3D para uso no lucrativo. Este modelo se escaló para crear una perspectiva real para el usuario. Sobré él se creó el plano y los límites de colisión. Estas acciones se pueden observar en la Figura 4.3.1.1.



(a)



(b)



(c)

Figura 4.3.1.1. Integración y generación de contenido 3D. Modelo original (a), modelo iluminado (b), generación de plano y límites de colisión, líneas en color verde (c).

4.3.2 Integración de imágenes del conjunto de pruebas

Al modelo 3D del centro comercial se le agregaron algunas de las imágenes del conjunto de pruebas. El resultado es un centro comercial virtual con las tiendas que contienen a los logotipos que el reconocedor puede identificar. Esto se puede ver en la Figura 4.3.2.1.



Figura 4.3.2.1. Centro comercial virtual con las tiendas de los logotipos que el reconocedor puede identificar.

4.3.3 Implementación del recorrido inmersivo

Para la implementación del recorrido inmersivo se utilizó un dispositivo Oculus Rift DK1 que reemplaza la cámara del jugador por dos cámaras, una para cada ojo. El control de movimiento del jugador y la invocación al reconocedor se realiza a través de un teclado o bien con el uso de un control de video juegos.

El usuario se desplaza hacia una tienda e invoca al reconocedor. Si éste identifica a un logotipo, le presenta al usuario una imagen con el mensaje publicitario correspondiente al logotipo que identificó. Un ejemplo de identificación del logotipo Ferrioni se observa en la Figura 4.3.3.1.

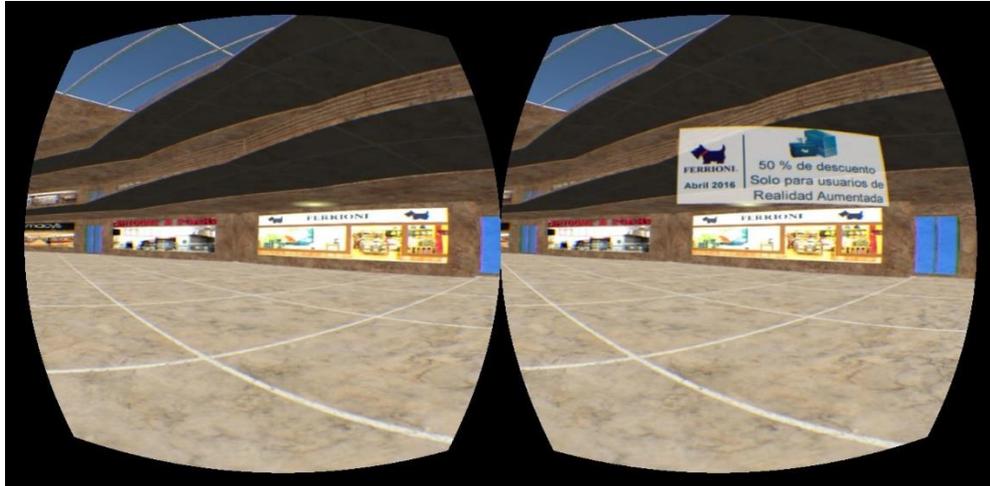


Figura 4.3.3.1. Recorrido inmersivo con Oculus Rift DK1 en el cual se identifica el logotipo Ferrioni.

Para llevar a cabo la identificación, se captura la imagen de una de las cámaras de Oculus Rift. Esta imagen se envía como entrada al reconocedor quien la analiza para determinar si se encuentra algún logotipo o no. En caso afirmativo proporciona como salida la identificación del logotipo. El mundo virtual interpreta la salida y le muestra al usuario una imagen correspondiente al mensaje publicitario del logotipo que se identificó.

4.3.4 Implementación del reconocedor

En el ambiente de desarrollo Unity, es posible especificar las variables de factor de escala y vecinos mínimos para cada logotipo. Por otra parte, la técnica de identificación depende directamente del entrenamiento del clasificador, sin embargo, se puede especificar qué clasificador utilizar. En la Figura 4.3.4.1 se presenta como se ajustan estos valores.

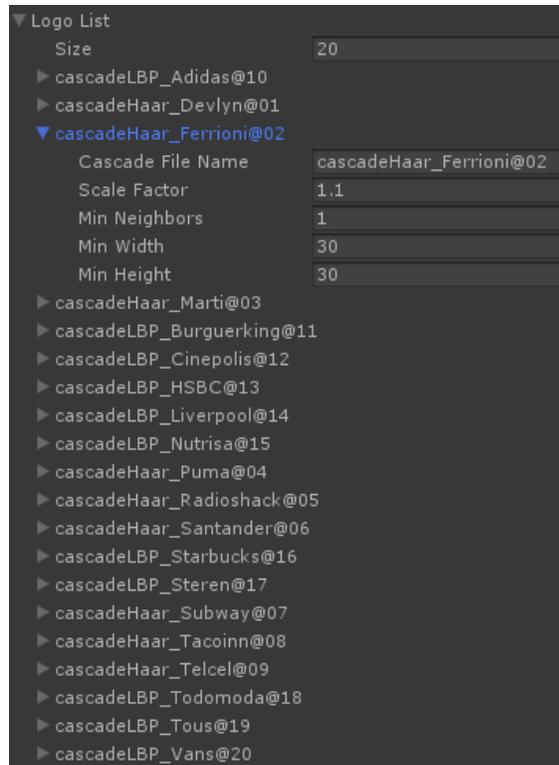


Figura 4.3.4.1. Ajuste de valores de variables y técnicas de detección para cada logotipo en el ambiente de desarrollo Unity.

La implementación del reconocedor dentro del mundo virtual permite especificar para cada logotipo la técnica de detección y los valores de las variables factor de escala y vecinos mínimos. Tener el control de estas variables permite ajustar los clasificadores en caso de ser necesario.

Los llamados a las librerías de *OpenCV* que realiza el reconocedor se implementaron con el complemento *OpenCV ObjectDetector* perteneciente a (Enox, 2015).

4.3.5 Evaluación del sistema en el escenario de pruebas

Para llevar a cabo la evaluación del sistema se estableció el cuestionario de evaluación mostrado en la Tabla 4.3.5.1. Cada cuestión está relacionada con alguna heurística de las Tabla 3.4.5.1 y 3.4.5.2.

Tabla 4.3.5.1. Lista de validación utilizada.

Cuestión	Heurística
1. ¿Se entera de lo que sucede durante el uso de la aplicación?	1

2. Cuando orienta la cámara hacia un logotipo, ¿se logra el objetivo de reconocerlo?	11
3. ¿El tamaño de los mensajes que se presentan es adecuado de acuerdo con la escena?	8
4. ¿Los mensajes son satisfactorios?	10
5. ¿Está satisfecho con el objetivo de la aplicación?	13
6. ¿Está satisfecho con la libertad de moverse y desplazarse mientras usa la aplicación?	13
7. ¿Es correcto el tiempo de presentación y visualización de los mensajes?	1
8. ¿Los mensajes mostrados son coherentes con el mundo real?	2
9. ¿Los mensajes se presentan de manera estandarizada?	4
10. ¿Se presenta algún mensaje al usuario cuando no se pudo reconocer algún logotipo?	5
11. ¿Es fácil recordar las funcionalidades de la aplicación?	6
12. ¿Es simple la curva de aprendizaje de usuarios novatos?	7
13. ¿Pueden los expertos utilizar la aplicación de una manera óptima?	7
14. ¿Se instruye al usuario para que sepa qué hacer en la aplicación?	9
15. ¿Existen requerimientos específicos? (orientación de la cámara, posición del usuario, condiciones de iluminación)	12
16. ¿El sistema es estable?	11
17. Si la aplicación detecta más de un logotipo en la escena, ¿continúa operando de manera correcta?	11

Este cuestionario se aplicó a 5 expertos. A cada uno, se le otorgó un tiempo de 20 minutos de uso de la aplicación.

Los resultados obtenidos del cuestionario se presentan en la Tabla 4.3.5.2. Para cada cuestión se muestra la valoración que dio cada uno de los expertos.

Tabla 4.3.5.2. Evaluación de los expertos en el uso del sistema en el escenario de pruebas.

Experto

Cuestión	1	2	3	4	5
1	5	5	5	5	5
2	4	5	4	4	3
3	5	5	5	4	5
4	5	4	4	5	5
5	5	5	5	4	5
6	5	5	5	5	5
7	5	5	4	5	4
8	5	5	4	4	5
9	4	5	5	5	5
10	3	4	4	3	4
11	5	5	5	5	5
12	5	5	5	5	5
13	4	3	4	3	3
14	5	5	5	5	5
15	4	4	5	4	4
16	4	5	4	4	5
17	5	5	4	5	5

Como puede observarse en la Tabla 4.3.5.2, los resultados obtenidos reflejan cierta satisfacción en el uso del sistema. Un análisis a estos resultados permite obtener suficiente evidencia para determinar oportunidades de mejora de la de usabilidad de la aplicación; un ejemplo se ve en la cuestión 10, en la que se deduce que el sistema debe mejorar en cuanto a informar cuando no se reconoce algún logotipo.

Ahora bien, de acuerdo con los resultados y acciones presentadas en esta sección se puede dar por cubierto el tercer objetivo específico.

4.4 Implementación en un dispositivo móvil

La aplicación final se implementó en un dispositivo móvil con sistema operativo Android. Tomando ventaja de que Unity permite crear aplicaciones para Android y de que el escenario de pruebas se desarrolló en Unity, la implementación del reconocedor se pudo replicar hacia el dispositivo móvil. La

nueva aplicación utiliza la cámara del dispositivo para realizar la detección en imágenes del mundo real.

En la Figura 4.4.1 se presenta el diseño de esta aplicación.

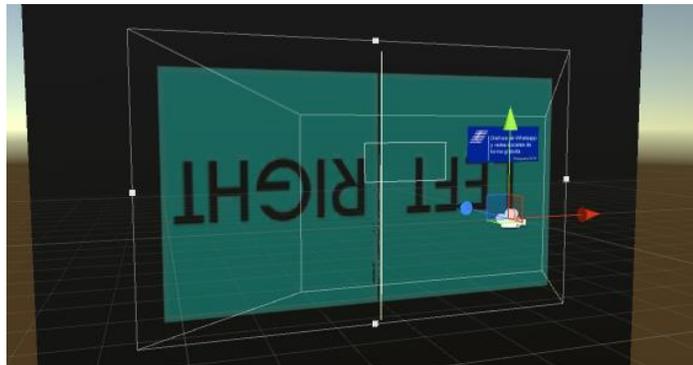


Figura 4.4.1. Diseño de la aplicación del sistema final.

Como puede observarse en la Figura 4.4.1, en esta implementación no está presente el modelo del mundo virtual. Otra diferencia radica en que esta aplicación no funciona con un Oculus Rift, en su caso, se opera a través del casco de realidad virtual mostrado en la Figura 4.4.2, que permite alojar un dispositivo móvil (Google VR, 2016).



Figura 4.4.2. Casco de realidad virtual del sistema implementado en un dispositivo móvil.

En este esquema, la captura de imágenes para identificación de logotipos se realiza por medio de la cámara del dispositivo móvil. Con el sensor Oculus Rift se contaba con dos cámaras, en este caso, la aplicación replica la imagen de video de la cámara del dispositivo para simular las dos cámaras. Con ello, se logra que un usuario pueda hacer uso de la aplicación mientras recorre un centro comercial como se muestra en la Figura 4.4.3.



Figura 4.4.3. Usuario utilizando el sistema mientras recorre un centro comercial.

Por otra parte, dado que esta aplicación también se generó con la herramienta Unity, el reconocedor se implementó de forma idéntica que la mostrada en la sección 4.3.4.

4.4.1 Evaluación del sistema en el dispositivo móvil

Con base en el mismo cuestionario de la Tabla 4.3.5.1, se solicitó a los expertos evaluar al sistema en el dispositivo móvil. Para ello, se preparó un espacio con las imágenes del conjunto de pruebas colocadas en la pared. Los expertos se desplazaban por este espacio mientras utilizaban la aplicación. La ponderación que registraron se presenta en la Tabla 4.4.1.1.

Tabla 4.4.1.1. Evaluación de los expertos en el uso del sistema en el dispositivo móvil.

Cuestión	Experto				
	1	2	3	4	5
1	5	5	5	5	5
2	3	4	4	4	3
3	5	5	5	4	5
4	5	4	4	5	5
5	5	5	5	4	5

6	5	4	5	4	4
7	5	5	4	5	4
8	5	5	4	4	5
9	4	5	5	5	5
10	3	4	4	3	4
11	5	5	5	5	5
12	5	5	5	5	5
13	4	3	4	3	3
14	5	5	5	5	5
15	4	4	5	4	4
16	4	5	4	4	5
17	5	5	4	5	5

Como puede notarse en la Tabla 4.4.1.1, la evaluación abre paso a una serie de actividades enfocadas en la mejora del uso del sistema. Sin embargo se observa que, de manera general, existe una satisfacción por parte de los expertos sobre el uso de la aplicación.

Capítulo 5 Conclusiones y perspectivas

En esta sección se presentan las principales conclusiones y se plantean las perspectivas de este trabajo.

5.1 Conclusiones

Después de desarrollar la aplicación de realidad aumentada, fue posible la identificación de los veinte logotipos seleccionados. Con ello, a los usuarios sometidos a prueba se les presentaron los mensajes publicitarios correspondientes a los logotipos que se identificaron.

A continuación se ahondarán los hallazgos de la investigación al buscar el cumplimiento de los objetivos planteados.

En este proyecto de tesis se desarrolló un método que utiliza visión por computadora y aprendizaje automático para reconocer logotipos comerciales. Se utilizaron las técnicas *Haar-like Features* y *Local Binary Patterns* en la detección de veinte logotipos. Estas técnicas afrontaron retos tales como cambios de iluminación, perspectivas de vista y limitaciones de la cámara del dispositivo. Con ellas se generaron cuarenta clasificadores de los cuales se seleccionaron aquellos que presentaron los más altos valores de exactitud y los menores tiempos de detección. La selección final consistió de nueve clasificadores con *Haar* y once con *LBP*. Los resultados obtenidos respecto al valor de exactitud son similares con respecto a los otros autores. La exactitud lograda es del 92%.

Se tienen dos prototipos completos, uno de ellos se opera con un dispositivo *Oculus Rift DK1* y el otro, con unos lentes de realidad virtual.

El reconocedor se constituyó a partir de los clasificadores seleccionados y fue capaz de determinar la identidad de los logotipos en las imágenes que recibe como entrada.

El escenario de pruebas se constituyó como una aplicación de realidad virtual con un centro comercial que incluye a todos los logotipos que se pueden identificar. En él se implementó el reconocedor y asoció su respuesta con la presentación de mensajes publicitarios. De esta manera se pudo mostrar su funcionalidad. La aplicación resultante fue evaluada de manera positiva por el grupo de expertos.

Por último, con base en la operatividad del escenario de pruebas, se desarrolló una aplicación para un dispositivo móvil. En ella también se implementó el reconocedor. Como consecuencia, es posible reconocer logotipos a partir de imágenes del mundo real y, con el uso de lentes de realidad virtual, los usuarios pueden operar la aplicación mientras recorren un centro comercial. Cuando se realiza la identificación de un logotipo, se despliega un mensaje publicitario al usuario que corresponde al logotipo identificado.

De esta manera, se puede afirmar que con base en la metodología descrita en el Capítulo 3 y los resultados obtenidos en el Capítulo 4 se cubren los objetivos que se plantearon para este trabajo.

5.1.1 Publicaciones

Como trabajo complementario a esta tesis, se elaboraron publicaciones que dieron soporte y fundamento al trabajo realizado.

El primer artículo que se destaca es el de (Benitez Ruíz, Medina Nieto, de la Calleja Mora, & Medina Nieto, 2015) en el cual se realizó un análisis de herramientas 3D para desarrollo de video juegos y con ello se sustentaron las bases para la creación de aplicación de realidad virtual y de realidad aumentada.

El segundo artículo, (Medina Nieto, de la Calleja, Medina, & Benitez Ruiz, 2016), permitió analizar de manera preliminar las técnicas de visión por computadora utilizadas en este trabajo, así también, los resultados que se obtuvieron hicieron posible su continuidad y conclusión.

Con base en los resultados de este trabajo, se está elaborando un tercer artículo que será sometido a una revista especializada.

5.2 Perspectivas

A partir de los resultados y discusiones presentados, las perspectivas de trabajos futuros se orientan en las siguientes direcciones.

En primer lugar está la exactitud obtenida. En concreto, se podría aumentar el número de ejemplos positivos tanto del conjunto de entrenamiento como del conjunto de pruebas, con ello podría aumentarse la exactitud y se disminuiría el número de falsos positivos que llegan a detectarse.

Ahora, con respecto al tiempo de detección, sería interesante realizar un análisis con base en técnicas como *Deep Learning* o *Scale Invariant Feature Transform (SIFT)* y con ello ver si es posible obtener mejores resultados enfocados en mejorar la eficiencia del sistema.

Los mensajes publicitarios que se presentan son fijos. Se podría crear una base de datos que permitiera de forma dinámica actualizar el contenido de los mensajes. La aplicación accedería a esta base de datos para presentar nuevos mensajes y descartar aquellos que han caducado. Más aún, la aplicación también podría utilizar esta misma base de datos para cambiar el conjunto de logotipos que pueda reconocer. Se pueden actualizar los clasificadores y sus parámetros de detección en la base de datos y la aplicación los podría tomar en cuenta.

Algo que no está relacionado con el alcance de este proyecto pero que podría gestionarse, sería por una parte, la publicación de la aplicación en la plataforma Play Store de Google y su desarrollo para iOS y, por otra parte, el registro de la aplicación ante el Instituto Nacional del Derecho de Autor.

Finalmente, sería de interés el implementar esta aplicación para alguno de los *Smart glasses* como los mostrados en la Figura 2.3.1.1 ya que con ellos, el proceso de detección e identificación de logotipos no provocaría ninguna interferencia con la visualización del usuario.

Referencias bibliográficas

- Bagheri, M., & Qigang, G. (2012). Logo Recognition Base on a Novel Pairwise Classification Approach. *2012 16th CSI Interational Symposium on Artificial Intelligence and Signal Processing (AISP)* (págs. 316-321). Shiraz, Fars: IEEE.
- Benitez Ruíz, A., Medina Nieto, M. A., de la Calleja Mora, J., & Medina Nieto, M. (2015). Animación de un modelo 3D del robot Darwin-OP utilizando Kinect. *Pistas Educativas, Instituto Tecnológico de Celaya*.
- Boguslaw, C. (2013). *Object Detection and Recognition in Digital Images: Theory and Practice*. Wiley.
- Bradski, G., & Kaehler, A. (2008). *Learning OpenCV Computer Vision with the OpenCV Library*. O'Reilly Media.
- Brahnam, S., C. Jain, L., Nanni, L., & Lumini, A. (2014). *Local Binary Patterns: New Variants and Applications*. Springer.
- Catchoom. (s.f.). *Image Recognition SDK on Google Glass*. Recuperado el 01 de 2015 de Novimembre, de Catchoom: <http://catchoom.com/documentation/image-recognition-sdk/image-recognition-sdk-on-google-glass/>
- de Paiva Guimaraes, M., & Farinazzo Martins, V. (2014). A checklist to evaluate Augmented Reality Applications. *Symposium on Virtual and Augmented Reality* (págs. 45-52). Piata Salvador: IEEE.
- Dickinger, A., Heinzmann, P., & Murphy, J. (2005). Mobile Environmental Applications. *Proceedings of the 38th Annual Hawaii International Conference System Sciences, 2005. HICSS '05* (pág. 178a). Hawaii: IEEE.

- Doermann, D., & Zhu, G. (2007). Automatic Document Logo Detection. *The 9th International Conference on Document Analysis and Recognition (ICDAR 2007)* (págs. 864-868). Curitiba: LAMP.
- Doermann, D., Rivlin, E., & Weiss, I. (1993). Logo Recognition Using Geometric Invariants . *Proceedings of the Second International Conference on Document Analysis and Recognition, 1993.* (págs. 894-897). Tsukuba: IEEE.
- E.R., D. (2012). *Computer and Machine Vision: Theory, Algorithms, Practicalities, Fourth Edition.* Oxford: Academic Press.
- Escalera, S., Baró, X., Pujol, O., Vitriá, J., & Radeva, P. (2011). *Traffic-Sign Recognition Systems.* Springer.
- Escudero, A. (18 de 02 de 2009). *Reconocimiento de objetos.* Recuperado el 2015 de Noviembre de 2015, de Universidad de Barcelona: <http://www.maia.ub.es/~sergio/linked/alberto09.pdf>
- Ferrell, O., & Hartline, M. (2012). *Estrategia de Marketing* (Quinta ed.). Cengage Learning.
- Fukuzawa, S. (s.f.). *Image Recognition Service "GAZIRU" Available for Smartphone Built-in Cameras.* Recuperado el 01 de Noviembre de 2015, de NEC: <http://www.nec.com/en/global/about/mitatv/12/index.html>
- Goldstone, W. (2011). *Unity 3.x Game Development Essentials.* Packt Publishing.
- Google VR. (15 de 01 de 2016). *Google Cardboard VR.* Obtenido de Google Cardboard VR: <https://vr.google.com/cardboard/>
- Hassanzadeh, S., & Pourghassem, H. (2011). A Fast Logo Recognition Algorithm in Noisy Document Images. *2011 International Conference on Intelligent Computation and Bio-Medical Instrumentation (ICBMI)* (págs. 64-67). Wuhan: IEEE.
- Hispanic Searching, S. (2015). *TN Relaciones.* Recuperado el 10 de 01 de 2016, de [tnrelaciones.com](http://www.tnrelaciones.com) : http://www.tnrelaciones.com/cm/preguntas_y_respuestas/content/391/4681/es/los-centros-y-plazas-comerciales-mas-valorados-de-mexico.html
- Huang, D., Shan, C., Ardabilian, M., Wang, Y., & Liming, C. (2011). Local Binary Patterns and Its Application to Facial Images Analysis: A survey. *IEEE Transactions on Systems, Man, and Cybernetics, Part C: Applications and Reviews*, 41(6), 765-781.
- Ikeuchi, K. (2014). *Computer Vision A Reference Guide.* New York: Springer.
- Kakasucai. (1 de 12 de 2015). *3DModelFree.com.* Obtenido de 3DModelFree.com: <http://www.3dmodelfree.com/>

- Kalaiyarasi, C., & Karthikeyan, S. (2014). Enhancing logo matching and recognition using local features. *2014 International Conference on Information Communication and Embedded Systems (ICICES)* (págs. 1-6). Chennai: IEEE.
- Kettle, R. (2014). *Concise Computer Vision, An Introduction into Theory and Algorithms*. Springer.
- Ku, D., Jianghua, C., & Gui, G. (2013). Translucent-static TV logo recognition by SUSAN corner extracting and matching. *2013 Third International Conference on Innovative Computing Technology (INTECH)* (págs. 44-48). London: IEEE.
- Lei, B., Chen, Y., & Lim, W.-Y. (2012). Logo Classification with Edge-Based DAISY Descriptor. *2012 IEEE International Symposium on Multimedia (ISM)* (págs. 222-228). Irvine: IEEE.
- Li, Z., Matthias, S.-A., & Martin, N. (2010). Fast Logo Detection and Recognition in Document Images. *2010 20th International Conference on Pattern Recognition (ICPR)* (págs. 2716-2719). Istambul: IEEE.
- Marín Dueñas, P. P. (2013). Publicidad Móvil: conocimiento, uso y utilidad para el anunciante. *Congreso Internacional de Comunicación y Sociedad Digital*. España: Dialnet.
- Martí Parreño, J. (2011). Publicidad expandida mediante realidad aumentada. *MK: Marketing + ventas*(267), 30-37.
- Martín Requero, M. I., & Alvarado López, M. C. (2007). *Nuevas tendencias en la publicidad del siglo XXI*. Sevilla: Comunicación Social.
- Martínez, I. J., & Aguado, J. M. (2006). El Desarrollo de la Telefonía Móvil como Plataforma Mediática. *HOLOGRAMÁTICA*, 21-39.
- Medina Nieto, M. A., de la Calleja, J., Medina, M., & Benitez Ruiz, A. (2016). Trademark logo recognition: Preliminary results on a comparative between Haar-like features and Local Binary Patterns. *Visión Politécnica*.
- Mery, D. (17 de 08 de 2004). *Visión por computador*. Recuperado el 01 de 2015 de Noviembre, de Pontificia Universidad Autónoma de Chile: <http://web.ing.puc.cl/~dmery/Prints/Books/2004-ApuntesVision.pdf>
- Microsoft. (s.f.). *Holographic computing is here*. Recuperado el 01 de 2015 de Noviembre, de Microsoft Corporation: <https://www.microsoft.com/microsoft-hololens/en-us>
- Mohd. Zainuddin, N. M., Badioze Zaman, H., & Ahmad, A. (2011). Heuristic Evaluation on Augmented Reality Courseware for the Deaf. *International Conference on User Science and Engineering* (págs. 183-188). Shah Alam, Selangor: IEEE.

- Nielsen, J. (1 de Enero de 1995). *10 Usability Heuristics for User Interface Design*. Obtenido de Nielsen Norman Group: <https://www.nngroup.com/articles/ten-usability-heuristics/>
- Nixon, M., & Aguado, A. (2008). *Feature Extraction & Image Processing for Computer Vision*. Academic Press.
- Pan, C., Yan, Z., Xu, X., & Sun, M. (2013). Vehicle logo recognition based on deep learning architecture in video surveillance for intelligent traffic system. *IET International Conference on Smart and Sustainable City 2013 (ICSSC 2013)* (págs. 123-126). Shangai: IET.
- Phan, R., & Dimitrios, A. (2010). Content-based retrieval of logo and trademarks in unconstrained color image databases using Color Edge Gradient Co-occurrence Histograms. *Computer Vision and Image Understanding*, 66-84.
- Pourghassem, H. (2012). A Hierarchical Logo Detection and Recognition Algorithm Using Two-Stage Segmentation and Multiple Classifiers. *2012 Fourth International Conference on Computational Intelligence and Communication Networks (CICN)* (págs. 227-231). Mathura: IEEE.
- Psyllos, A., Christos-Nikolaos, A., & Eleftherios, K. (2010). Vehicle Logo Recognition Using a SIFT-Based Enhanced Matching Scheme. *IEEE Transactions on Intelligent Transportation Systems*, 11(2), 322-328.
- Pulli, K., Baksheev, A., & Korniyakov Kirill, E. V. (2012). Realtime Computer Vision with OpenCV. *Communications of the ACM*, 61-69.
- Quazzani, I. (29 de 08 de 2012). *Manual de creación de videojuego con Unity 3D*. Recuperado el 01 de Noviembre de 2015, de e-Archivo: <http://e-archivo.uc3m.es/handle/10016/16345>
- Rasiel, E. (27 de 01 de 2015). *Smart Glasses Market Report 2015*. Recuperado el 01 de Noviembre de 2015, de [augmentedreality.org: http://www.augmentedreality.org/#!smart-glasses-report/c88h](http://www.augmentedreality.org/#!smart-glasses-report/c88h)
- Relaciones, T. (2015). *Los Centros y Plazas Comerciales más Valorados de México*. Recuperado el 10 de 01 de 2016, de [tnrelaciones.com: http://www.tnrelaciones.com/cm/preguntas_y_respuestas/content/391/4681/es/los-centros-y-plazas-comerciales-mas-valorados-de-mexico.html](http://www.tnrelaciones.com/cm/preguntas_y_respuestas/content/391/4681/es/los-centros-y-plazas-comerciales-mas-valorados-de-mexico.html)
- Ruiz Davis, S., & Polo Serrano, D. (2012). La realidad aumentada como nuevo concepto de la publicidad online a través de los smartphones. *Primera Revista Electrónica en América Latina Especializada en Comunicación*.
- Sucar, E., & Gómez, G. (2008). *Visión Computacional*. Puebla.
- Sun, Q., Lu, X., & Hu, H. (2014). An Improved Vehicle Logo Recognition Method for Road Surveillance Images. *2014 Seventh International Symposium on*

- Computational Intelligence and Design (ISCID)* (págs. 373-776). Hangzhou: IEEE.
- Szeliski, R. (2010). *Computer Vision: Algorithms and Applications*. Springer.
- Telefónica, F. (01 de 01 de 2011). *Realidad aumentada: una nueva lente para ver el mundo*. Recuperado el 01 de Noviembre de 2015, de realidadaugmentada-fundaciontelefonica.com:
<http://www.realidadaugmentada-fundaciontelefonica.com/realidad-aumentada.pdf>
- Uddin, M. S., & Akhi, Y. (2016). Horse Detection Using Haar Like Features. *International Journal of Computer Theory and Engineering*, 415-418.
- Viola, P., & Jones, M. (2001). Rapid object detection using a boosted cascade of simple features. *Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2001. CVPR 2001*, 1-511-1-518.
- Wang, H. (2010). Document Logo Detection and Recognition Using Bayesian Model. *2010 20th International Conference on Pattern Recognition (ICPR)* (págs. 1961-1964). Istanbul: IEEE.
- Wei, C.-H., Li, Y., Chau, W.-Y., & Li, C.-T. (2009). Trademark image retrieval using synthetic features for describing global shape and interior structure. *Pattern Recognition*, 42(3), 386-394.
- Wulfe, B. (20 de 02 de 2015). *Mergevec*. Obtenido de Github:
<https://github.com/wulfebw/mergevec>
- Xia, L., Qi, F., & Qianhao, Z. (2008). A Learning-based Logo Recognition Algorithm Using SIFT and Efficient Correspondence Matching. *International Conference on Information and Automation, 2008*. (págs. 1767-1772). Changsha: IEEE.
- Xiao, J., Xiang, W., & Yuncai, L. (2015). Vehicle logo recognition by weighted multi-class support vector machine ensembles based on sharpness histogram features. *Image Processing, IET*, 9(7), 527-534.
- Y. Shih, F. (2010). *Image Processing and Pattern Recognition, Fundamentals and Techniques*. New Jersey: Wiley.
- Yang, H., Zhai, L., Liu, Z., & Li, L. (2013). An Efficient Method for Vehicle Model Identification Via Logo Recognition. *2013 Fifth International Conference on Computational and Information Sciences (ICCIS)* (págs. 1080-1083). Shiyang: IEEE.
- Yilmaz, B., & Onay Durdu, P. (2015). Heuristic Evaluation of a Mobile Hand-Writing Learning Application. *Application of Information and Communication Technologies* (págs. 549-552). Rostov on Don: IEEE.

- Yu, S., Zheng, S., Yang, H., & Liang, L. (2013). Vehicle Logo Recognition Base on Bag of Words. *2013 10th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS)* (págs. 353-358). Cracovia: IEEE.
- Yunqiong, W., Zhifang, L., & Fei, X. (2007). A Fast Coarse-to-Fine Vehicle Logo Detection and Recognition Method. *IEEE International Conference on Robotics and Biomimetics 2007* (págs. 691-696). Sanya: IEEE.
- Zhang, Y., Zhang, S., Liang, W., & Liang, J. (2012). A Novel Location and Matching Algorithm for Rapid Logo Recognition in Video Advertisements. *2012 International Conference on Image Analysis and Signal Processing (IASP)* (págs. 1-5). Hangzhou: IEEE.
- Zhao, Y. (2012). Theories and Applications of LBP: A Survey. *Advanced Intelligent Computing Theories and Applications*, 112-121.

Anexo

A.1 Resultados de los experimentos de detección de los clasificadores

Resultados de los experimentos de detección del clasificador Ferrioni basado en la técnica *LBP*. Exactitud (en porcentaje) y tiempo de detección (en milisegundos)

A continuación se presentan los resultados de los experimentos de cada clasificador sobre su conjunto de pruebas correspondiente para las técnicas de *Haar* y *LBP*. Las unidades de exactitud están en porcentaje y el tiempo de detección está en milisegundos.

Tabla A.1.1. Resultados del clasificador Adidas basado en *Haar*.

Factor de escala	Vecinos mínimos	TP	FP	FN	TN	Exactitud	Tiempo de detección
1.2	1	27	223	3	247	54.80	5.78
1.1	1	115	135	6	244	71.80	11.59
1.01	1	233	17	148	102	67.00	106.12
1.05	1	180	70	33	217	79.40	23.55
1.05	3	105	145	1	249	70.80	21.91
1.07	1	139	111	19	231	74.00	16.26

Tabla A.1.2. Resultados del clasificador Adidas basado en *LBP*.

Factor de escala	Vecinos mínimos	TP	FP	FN	TN	Exactitud	Tiempo de detección
1.01	5	225	25	2	248	94.60	87.81

1.01	3	235	15	10	240	95.00	87.74
1.01	1	246	4	29	221	93.40	87.45
1.02	1	231	19	19	231	92.40	42.82
1.02	3	197	53	1	249	89.20	43.14

Tabla A.1.3. Resultados del clasificador Burger King basado en *Haar*.

Factor de escala	Vecinos mínimos	TP	FP	FN	TN	Exactitud	Tiempo de detección
1.03	1	145	105	13	237	76.40	78.77
1.01	1	205	45	62	188	78.60	229.38
1.02	3	112	138	7	243	71.00	113.19
1.04	1	130	120	13	237	73.40	59.72
1.03	3	81	169	0	250	66.20	76.98

Tabla A.1.4. Resultados del clasificador Burger King basado en *LBP*.

Factor de escala	Vecinos mínimos	TP	FP	FN	TN	Exactitud	Tiempo de detección
1.01	5	194	56	0	250	88.80	135.91
1.01	3	235	15	1	249	96.80	134.89
1.01	1	250	0	27	223	94.60	138.91
1.02	1	232	18	11	239	94.20	67.90
1.02	3	162	88	0	250	82.40	69.76

Tabla A.1.5. Resultados del clasificador Cinépolis basado en *Haar*.

Factor de escala	Vecinos mínimos	TP	FP	FN	TN	Exactitud	Tiempo de detección
1.05	1	249	1	102	148	79.40	45.50
1.05	3	242	8	57	193	87.00	45.79
1.04	1	250	0	103	147	79.40	57.00
1.04	3	241	9	61	189	86.00	56.44
1.03	1	250	0	114	136	77.20	76.09
1.03	3	250	0	79	171	84.20	72.54

Tabla A.1.6. Resultados del clasificador Cinépolis basado en *LBP*.

Factor de escala	Vecinos mínimos	TP	FP	FN	TN	Exactitud	Tiempo de detección
1.01	5	217	33	1	249	93.20	139.45
1.01	3	230	20	11	239	93.80	141.06

1.01	1	243	7	69	181	84.80	139.77
1.02	1	230	20	19	231	92.20	70.70
1.02	3	197	53	0	250	89.40	71.94

Tabla A.1.7. Resultados del clasificador Devlyn basado en *Haar*.

Factor de escala	Vecinos mínimos	TP	FP	FN	TN	Exactitud	Tiempo de detección
1.2	3	250	0	2	245	99.60	10.83
1.2	5	249	1	0	250	99.80	10.87
1.3	3	250	0	1	249	99.80	6.74
1.3	5	245	5	0	250	99.00	6.82
1.2	1	161	89	28	222	76.60	16.12

Tabla A.1.8. Resultados del clasificador Devlyn basado en *LBP*.

Factor de escala	Vecinos mínimos	TP	FP	FN	TN	Exactitud	Tiempo de detección
1.01	5	250	0	1	249	99.80	108.31
1.01	3	250	0	7	243	98.60	102.36
1.01	1	250	0	43	207	91.40	104.99
1.1	1	249	1	0	250	99.80	12.18
1.2	1	250	0	0	250	100.00	7.50

Tabla A.1.9. Resultados del clasificador Ferrioni basado en *Haar*.

Factor de escala	Vecinos mínimos	TP	FP	FN	TN	Exactitud	Tiempo de detección
1.2	1	222	28	0	250	94.40	12.15
1.1	1	228	2	5	245	98.54	20.91
1.3	1	174	75	0	0	84.97	7.96
1.15	3	231	19	2	248	95.80	13.42
1.02	1	250	0	43	207	91.40	91.20

Tabla A.1.10. Resultados del clasificador Ferrioni basado en *LBP*.

Factor de escala	Vecinos mínimos	TP	FP	FN	TN	Exactitud	Tiempo de detección
1.01	5	129	121	40	210	67.80	125.00
1.01	3	168	88	52	198	72.00	122.54
1.01	1	229	21	142	108	67.40	120.62
1.02	1	188	62	85	165	70.60	65.34
1.02	3	114	136	37	213	65.40	62.76

Tabla A.1.11. Resultados del clasificador HSBC basado en *Haar*.

Factor de escala	Vecinos mínimos	TP	FP	FN	TN	Exactitud	Tiempo de detección
1.1	3	192	58	55	195	77.40	7.67
1.1	1	233	17	115	135	73.60	7.71
1.05	1	245	5	136	114	71.80	14.35
1.2	1	218	32	84	166	76.80	4.36
1.2	3	158	92	32	218	75.20	5.91
1.3	1	158	65	67	183	73.60	3.53

Tabla A.1.12. Resultados del clasificador HSBC basado en *LBP*.

Factor de escala	Vecinos mínimos	TP	FP	FN	TN	Exactitud	Tiempo de detección
1.01	5	126	124	0	250	75.20	64.57
1.01	3	167	83	2	248	83.00	66.66
1.01	1	228	22	29	221	89.80	63.80
1.02	1	192	58	7	243	87.00	31.10
1.02	3	91	159	0	250	68.20	31.48

Tabla A.1.13. Resultados del clasificador Liverpool basado en *Haar*.

Factor de escala	Vecinos mínimos	TP	FP	FN	TN	Exactitud	Tiempo de detección
1.1	3	196	54	50	200	79.20	34.35
1.05	3	243	7	162	88	66.20	68.64
1.05	5	245	35	75	175	78.00	69.27
1.15	3	168	82	29	221	77.80	24.60
1.12	3	158	92	42	208	73.20	28.83

1.05	7	182	68	41	209	78.20	67.21
------	---	-----	----	----	-----	-------	-------

Tabla A.1.14. Resultados del clasificador Liverpool basado en *LBP*.

Factor de escala	Vecinos mínimos	TP	FP	FN	TN	Exactitud	Tiempo de detección
1.01	5	242	8	13	237	95.80	193.37
1.01	3	246	4	34	216	92.40	194.79
1.01	1	250	0	87	163	82.60	192.75
1.02	1	248	2	51	199	89.40	98.20
1.02	3	229	21	12	238	93.40	97.93

Tabla A.1.15. Resultados del clasificador Martí basado en *Haar*.

Factor de escala	Vecinos mínimos	TP	FP	FN	TN	Exactitud	Tiempo de detección
1.2	1	249	1	148	102	70.20	7.49
1.2	3	247	3	31	219	93.20	7.04
1.1	3	250	0	120	130	76.00	16.19
1.1	5	250	0	51	199	89.80	15.75
1.2	5	232	18	8	242	94.80	6.99
1.15	5	241	9	13	237	95.60	9.93

Tabla A.1.16. Resultados del clasificador Martí basado en *LBP*.

Factor de escala	Vecinos mínimos	TP	FP	FN	TN	Exactitud	Tiempo de detección
1.01	5	249	1	6	244	98.60	105.05
1.01	3	250	0	11	239	97.80	103.36
1.01	1	250	0	93	157	81.40	90.39
1.02	1	249	1	39	211	92.00	50.55
1.02	3	239	11	6	244	96.60	52.40

Tabla A.1.17. Resultados del clasificador Nutrisa basado en *Haar*.

Factor de escala	Vecinos mínimos	TP	FP	FN	TN	Exactitud	Tiempo de detección
------------------	-----------------	----	----	----	----	-----------	---------------------

1.01	1	248	2	189	61	61.80	207.29
1.01	3	190	60	127	123	62.60	202.22
1.05	1	115	135	74	179	58.20	43.26
1.02	1	219	31	148	102	64.20	104.81
1.02	3	98	152	67	183	56.20	102.65

Tabla A.1.18. Resultados del clasificador Nutrisa basado en *LBP*.

Factor de escala	Vecinos mínimos	TP	FP	FN	TN	Exactitud	Tiempo de detección
1.01	5	227	28	0	250	94.46	188.47
1.01	3	246	4	2	248	98.80	194.36
1.01	1	243	7	69	181	84.80	167.42
1.02	1	244	6	30	220	92.80	90.27
1.02	3	213	37	0	250	92.60	90.84

Tabla A.1.19. Resultados del clasificador Puma basado en *Haar*.

Factor de escala	Vecinos mínimos	TP	FP	FN	TN	Exactitud	Tiempo de detección
1.2	1	227	23	22	228	91.00	10.19
1.2	3	194	56	7	243	87.40	7.63
1.1	1	225	25	37	213	87.60	14.28
1.05	1	219	31	46	204	84.60	27.71
1.18	1	198	52	27	223	84.20	8.53

Tabla A.1.20. Resultados del clasificador Puma basado en *LBP*.

Factor de escala	Vecinos mínimos	TP	FP	FN	TN	Exactitud	Tiempo de detección
1.01	5	250	0	14	236	97.20	165.04
1.01	3	250	0	51	199	89.80	168.60
1.01	1	250	0	170	80	66.00	139.88
1.02	1	250	0	77	173	84.60	82.64
1.02	3	250	0	17	233	96.60	84.66

Tabla A.1.21. Resultados del clasificador Radioshack basado en *Haar*.

Factor de escala	Vecinos mínimos	TP	FP	FN	TN	Exactitud	Tiempo de detección
1.2	1	247	3	11	239	97.20	14.93
1.1	1	250	0	16	234	96.80	26.41
1.2	3	231	19	0	250	96.20	14.89
1.05	1	250	0	39	211	92.20	51.95
1.3	1	238	12	0	250	97.60	11.33

Tabla A.1.22. Resultados del clasificador Radioshack basado en *LBP*.

Factor de escala	Vecinos mínimos	TP	FP	FN	TN	Exactitud	Tiempo de detección
1.01	5	250	0	5	245	99.00	143.74
1.01	3	250	0	16	234	96.80	144.61
1.01	1	250	0	117	133	76.60	139.79
1.02	1	250	0	5	245	99.00	109.88
1.02	3	250	0	38	212	92.40	71.66

Tabla A.1.23. Resultados del clasificador Santander basado en *Haar*.

Factor de escala	Vecinos mínimos	TP	FP	FN	TN	Exactitud	Tiempo de detección
1.2	1	239	11	0	250	97.80	10.92
1.3	1	216	34	0	250	93.20	6.48
1.15	1	243	7	1	249	98.40	13.80
1.1	1	250	0	1	249	99.80	20.83

Tabla A.1.24. Resultados del clasificador Santander basado en *LBP*.

Factor de escala	Vecinos mínimos	TP	FP	FN	TN	Exactitud	Tiempo de detección
1.01	5	48	202	0	250	59.60	111.41
1.01	3	93	157	1	249	68.40	121.40
1.01	1	187	63	30	220	81.40	124.07
1.02	1	100	150	4	246	69.20	55.62
1.02	3	27	223	0	250	55.40	56.27

Tabla A.1.25. Resultados del clasificador Starbucks basado en *Haar*.

Factor de escala	Vecinos mínimos	TP	FP	FN	TN	Exactitud	Tiempo de detección
1.1	1	222	28	71	179	80.20	26.59
1.1	3	117	133	0	250	73.40	27.13
1.05	1	243	7	131	119	72.40	52.44
1.05	3	212	38	42	208	84.00	52.91
1.02	3	250	0	120	130	76.00	114.47

Tabla A.1.26. Resultados del clasificador Starbucks basado en *LBP*.

Factor de escala	Vecinos mínimos	TP	FP	FN	TN	Exactitud	Tiempo de detección
1.1	1	235	15	0	250	97.00	17.57
1.05	1	248	2	3	247	99.00	33.60
1.01	1	250	0	35	215	93.00	156.17
1.01	1	250	0	19	231	96.20	75.34
1.02	3	250	0	2	248	99.60	74.49

Tabla A.1.27. Resultados del clasificador Steren basado en *Haar*.

Factor de escala	Vecinos mínimos	TP	FP	FN	TN	Exactitud	Tiempo de detección
1.01	1	221	29	77	173	78.80	200.59
1.02	1	169	81	68	182	70.20	58.24

Tabla A.1.28. Resultados del clasificador Steren basado en *LBP*.

Factor de escala	Vecinos mínimos	TP	FP	FN	TN	Exactitud	Tiempo de detección
1.01	5	144	106	0	250	78.80	131.18
1.01	3	171	79	2	248	83.80	132.69
1.01	1	210	40	34	216	82.20	132.69
1.02	1	177	73	15	235	84.40	65.48
1.02	3	141	109	0	250	78.20	67.37

Tabla A.1.29. Resultados del clasificador Subway basado en *Haar*.

Factor de escala	Vecinos mínimos	TP	FP	FN	TN	Exactitud	Tiempo de detección
1.1	1	178	72	0	250	85.60	12.37
1.05	1	212	38	1	249	92.20	27.43
1.01	1	249	1	39	211	92.00	120.73
1.02	1	244	6	23	227	94.20	60.53
1.02	3	221	29	2	248	93.80	63.03

Tabla A.1.30. Resultados del clasificador Subway basado en *LBP*.

Factor de escala	Vecinos mínimos	TP	FP	FN	TN	Exactitud	Tiempo de detección
1.01	5	209	41	0	250	91.80	100.36
1.01	3	230	20	10	240	94.00	100.52
1.01	1	245	5	74	179	84.20	101.98
1.02	1	232	18	17	233	93.00	49.97
1.02	3	191	59	0	250	88.20	52.94

Tabla A.1.31. Resultados del clasificador Taco Inn basado en *Haar*.

Factor de escala	Vecinos mínimos	TP	FP	FN	TN	Exactitud	Tiempo de detección
1.2	1	248	2	2	248	99.20	10.50
1.1	1	250	0	10	240	98.00	20.22
1.3	1	239	11	1	249	97.60	6.48
1.25	1	241	9	0	250	98.20	8.30
1.15	1	250	0	7	243	98.60	13046

Tabla A.1.32. Resultados del clasificador Taco Inn basado en *LBP*.

Factor de escala	Vecinos mínimos	TP	FP	FN	TN	Exactitud	Tiempo de detección
1.01	5	230	20	6	244	94.80	121.79
1.01	3	242	8	14	236	95.60	122.04
1.01	1	244	6	69	181	85.00	119.42
1.02	1	241	9	33	217	91.60	61.49
1.02	3	209	41	3	247	91.20	61.47

Tabla A.1.33. Resultados del clasificador Telcel basado en *Haar*.

Factor de escala	Vecinos mínimos	TP	FP	FN	TN	Exactitud	Tiempo de detección
1.01	1	200	50	159	91	58.20	178.55
1.1	1	181	69	26	224	81.00	22.74
1.2	1	160	90	6	244	80.80	10.76
1.15	1	175	75	11	239	82.80	14.03
1.12	1	178	72	20	230	81.60	17.53

Tabla A.1.34. Resultados del clasificador Telcel basado en *LBP*.

Factor de escala	Vecinos mínimos	TP	FP	FN	TN	Exactitud	Tiempo de detección
1.01	5	132	118	1	249	76.20	135.98
1.01	3	174	76	7	243	83.40	146.57
1.01	1	227	23	82	168	79.00	160.97
1.02	1	181	69	30	220	80.20	72.86
1.02	3	95	155	3	247	68.40	73.35

Tabla A.1.35. Resultados del clasificador Todomoda basado en *Haar*.

Factor de escala	Vecinos mínimos	TP	FP	FN	TN	Exactitud	Tiempo de detección
1.1	1	243	1	210	40	57.29	27.60
1.1	3	198	52	76	174	74.40	27.86
1.1	5	145	105	21	229	74.80	28.24
1.05	3	242	8	162	88	66.00	53.66
1.05	5	221	29	107	143	72.80	54.245

Tabla A.1.36. Resultados del clasificador Todomoda basado en *LBP*.

Factor de escala	Vecinos mínimos	TP	FP	FN	TN	Exactitud	Tiempo de detección
1.01	5	217	33	108	142	71.80	135.86
1.01	3	230	20	137	113	68.60	130.35
1.01	1	245	5	218	32	55.40	135.71
1.05	1	222	28	75	175	79.40	27.25
1.05	3	190	60	19	231	84.20	29

Tabla A.1.37. Resultados del clasificador Tous basado en *Haar*.

Factor de escala	Vecinos mínimos	TP	FP	FN	TN	Exactitud	Tiempo de detección
1.01	1	239	11	89	161	80.00	54.81
1.01	3	209	41	50	200	81.80	53.55
1.02	1	207	43	69	181	77.60	62.91

Tabla A.1.38. Resultados del clasificador Tous basado en *LBP*.

Factor de escala	Vecinos mínimos	TP	FP	FN	TN	Exactitud	Tiempo de detección
1.01	5	88	167	0	250	66.93	41.55
1.01	3	137	113	0	250	77.40	41.58
1.01	1	197	53	6	244	88.20	41.85
1.02	1	155	95	1	249	80.80	20.47
1.02	3	64	186	0	0	62.80	20.45

Tabla A.1.39. Resultados del clasificador Vans basado en *Haar*.

Factor de escala	Vecinos mínimos	TP	FP	FN	TN	Exactitud	Tiempo de detección
1.05	1	161	89	9	241	80.40	22.62
1.03	1	184	66	20	230	82.80	35.39
1.01	1	231	19	75	175	81.20	104.33
1.01	3	205	45	13	237	88.40	115.59
1.02	3	162	88	0	250	82.40	54.69
1.02	1	205	45	39	211	83.20	52.70

Tabla A.1.40. Resultados del clasificador Vans basado en *LBP*.

Factor de escala	Vecinos mínimos	TP	FP	FN	TN	Exactitud	Tiempo de detección
1.01	5	167	83	0	250	83.40	71.35
1.01	3	201	49	1	249	90.00	86.49
1.01	1	230	20	38	212	88.40	87.18
1.02	1	201	49	9	241	88.40	47.70

1.02	3	129	121	1	249	75.60	47.61
------	---	-----	-----	---	-----	-------	-------

A.2 Ejecución de los experimentos de detección de logotipos

La ejecución de los experimentos para medir la exactitud de la detección sobre el conjunto de pruebas se implementó en el lenguaje de programación *Python*. En la Figura A.2.1 se presentan las principales líneas de código de esta implementación. En este caso, sobre el conjunto de pruebas del logotipo Adidas. El código es similar para los demás logotipos.

```

cascade_fn = args.get('--cascade', "cascadeHaarAdidas.xml")
cascade = cv2.CascadeClassifier(cascade_fn)

for file in os.listdir("PositivosAdidas"):
    img = cv2.imread("PositivosAdidas/"+file)
    numImages = numImages + 1
    gray = cv2.cvtColor(img, cv2.COLOR_BGR2GRAY)
    gray = cv2.equalizeHist(gray)
    #Llamada a l
    cascade.detectMultiScale(gray, scaleFactor=1.07, minNeighbors=1, minSize=(40, 40), maxSize=(150,150))
    vis = img.copy()
    draw_rects(vis, rects, (0, 255, 0))
    print 'Procesando: ' + file
    draw_str(vis, (20, 200), 'tiempo: %.1f ms' % (dt*1000))
    cv2.imshow('facedetect', vis)

```

Figura A.2.1. Sección de código para medir la exactitud de detección del logotipo Adidas.

A.3 Código de implementación del reconocedor

En la Figura A.3.1 se presentan las principales líneas de código correspondientes a la implementación del reconocedor.

```

private void Identifier(Texture2D texture){
    foreach (Logo logo in logoList){
        LogoDetected = logo.logoName;
        OpenCVObjectDetector.RemoveAllObjectDetectorParam ();
        IDictionary<string,object> param = new Dictionary<string,object> ();
        param.Add ("filename", logo.fileName);
        param.Add ("scaleFactor", logo.scaleFactor);
        param.Add ("minNeighbors", logo.minNeighbors);
        param.Add ("minwidth", logo.minwidth);
        param.Add ("minHeight", logo.minHeight);
        param.Add ("flipCode", 0);
        OpenCVObjectDetector.AddObjectDetectorParam (Json.Serialize (param));
        OpenCVObjectDetector.DetectAsync (texture, PlaneDetect.name, "LogoDetectCallback");
        byte[] bytes = texture.EncodeToPNG ();
    }
}

```

Figura A.3.1. Código del método principal del reconocedor.