



UNIVERSIDAD NACIONAL
AUTÓNOMA DE
MÉXICO

UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO

POSGRADO EN CIENCIA E INGENIERÍA DE LA COMPUTACIÓN

"ANÁLISIS DE LA ACTIVIDAD DE
CUATRO ESTUDIANTES Y UN PROFESOR
EN EL AULA DEL FUTURO UTILIZANDO
TÉCNICAS DE VISIÓN
COMPUTACIONAL"

T E S I S

QUE PARA OBTENER EL GRADO DE:

MAESTRO EN INGENIERÍA
(COMPUTACIÓN)

P R E S E N T A:

ING. ADALBERTO BLANCO MEJÍA

DIRECTOR DE
TESIS: DR. FERNANDO GAMBOA RODRÍGUEZ
CO-DIRECTOR: DR. YANN FRAUEL

MÉXICO DF.

2009



Universidad Nacional
Autónoma de México

Dirección General de Bibliotecas de la UNAM

Biblioteca Central



UNAM – Dirección General de Bibliotecas
Tesis Digitales
Restricciones de uso

DERECHOS RESERVADOS ©
PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL

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

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

Agradecimientos

Quisiera agradecer a mis asesores de tesis, **Dr. Fernando Gamboa Rodríguez y Dr. Yann Frauel**, por el apoyo y confianza que depositaron en mí durante el desarrollo de este proyecto. También quiero agradecer el apoyo de mis sinodales, por las observaciones que hicieron a este trabajo, pues me ayudaron a comprender como poder expresar claramente mis ideas.

A mis amigos quienes me alentaron a seguir adelante. Especialmente quiero agradecer el apoyo incondicional de mi compañera de vida "Adriana Hernández", por alentarme día con día para que este trabajo fuera concluido.

Agradecimientos

Quisiera agradecer a mis asesores de tesis, **Dr. Fernando Gamboa Rodríguez y Dr. Yann Frauel**, por el apoyo y confianza que depositaron en mí durante el desarrollo de este proyecto. También quiero agradecer el apoyo de mis sinodales, por las observaciones que hicieron a este trabajo, pues me ayudaron a comprender como poder expresar claramente mis ideas.

A mis amigos quienes me alentaron a seguir adelante. Especialmente quiero agradecer el apoyo incondicional de mi compañera de vida "Adriana Hernández", por alentarme día con día para que este trabajo fuera concluido.

Resumen general

En la actualidad, los avances tecnológicos nos han alcanzado y forman parte de nuestras vidas, permitiéndonos realizar tareas complejas en escenarios y contextos que hace apenas unos años eran impensables. En particular, en este trabajo nos centraremos en el uso de tecnologías como apoyo al problema educativo, y es que si bien ha habido muchas y muy diversos esfuerzos por introducir las Tecnologías de la Información y la Comunicación (TIC) como apoyo a la educación, no siempre han sido exitosos, debido entre otros factores, a la falta de metodologías que contemplen la complejidad y las características propias del proceso educativo.

Percibir el proceso de enseñanza-aprendizaje como un fenómeno complejo ha alentado a grupos de trabajo interdisciplinarios a estudiar de manera conjunta y articulado el problema, surgiendo proyectos como el Aula del Futuro, en el que se busca estudiar desde una perspectiva interdisciplinaria las dinámicas educativas y sociales que surgen en el aula, para entonces desarrollar y evaluar las tecnologías que mejor se adapten a los objetivos educativos que se persiguen. Así, el Aula del Futuro es un proyecto en el que un espacio educativo con capacidad para cuatro estudiantes y un profesor es enriquecido con técnicas de cómputo ubicuo y realidad aumentada, buscando promover las tareas de participación, cooperación y colaboración entre los estudiantes.

Un segundo objetivo es poder brindar al profesor las condiciones necesarias para que pueda llevar a cabo su labor, sin ser distraído o agobiado por la presencia de la nueva tecnología. En efecto, el proyecto no busca dar nuevas tareas y obligaciones a los profesores, sino por el contrario facilitar en lo posible su labor. ¿Cómo alcanzar este ideal si el profesor se ve obligado a activar, desactivar y calibrar dispositivos cada quince minutos?

Con el propósito de liberar a los profesores de estas tareas, en este trabajo se propone analizar la actividad dentro del aula de manera automática, utilizando para ello técnicas de

visión por computadora, permitiendo detectar los diferentes escenarios en los que alumnos y maestros llevan a cabo actividades educativas. La detección de los escenarios dentro del aula permitirá determinar que dispositivos deben ser utilizados en ese momento.

Desde hace aproximadamente treinta años el área de visión por computadora ha formado parte de un sistema multidisciplinario en el que intervienen disciplinas como: inteligencia artificial, robótica, procesamiento de señales, reconocimiento de patrones, teoría de control, psicología y otras más. En ese sentido, en este trabajo se busca acercar las técnicas de visión por computadora al área educativa, con la premisa de que la detección de escenarios puede ser un punto de partida para la adaptabilidad de los espacios educativos.

La tesis está organizada de la siguiente forma: el *primer capítulo* presenta las características de las aulas y su evolución, los problemas encontrados en la incorporación de tecnología al proceso de enseñanza-aprendizaje, así como las características del proyecto el Aula del Futuro y la metodología a seguir para la detección de escenarios. En el *segundo capítulo*, se describen las técnicas de visión por computadora usadas para la detección de escenarios, a partir de la detección de marcas y la detección de rostros. Por lo cual, en el *tercer capítulo* presenta la arquitectura del sistema y la integración de módulos que permiten la detección de escenarios. El *cuarto capítulo* describe los casos de uso y las pruebas realizadas para analizar la funcionalidad del sistema en la detección de escenarios. Finalmente, en el *quinto capítulo* se presentan las conclusiones obtenidas del desarrollo de esta tesis.

Índice general

Agradecimientos	i
Resumen general	ii
Índice de figuras	vi
Capítulo 1 El aula	1
1.1 Introducción	1
1.2 Antecedentes	2
1.2.1 La tecnología en el aula	5
1.3 El proyecto: el Aula del Futuro	7
1.3.1 Características del Aula del Futuro	8
1.3.2 Distribución de dispositivos	9
1.3.3 Límites del proyecto y modificaciones necesarias	10
1.4 Problemática	11
1.5 Solución	12
1.6 Metodología	12
1.7 Contribución y relevancia	14
Resumen del capítulo	14
Capítulo 2 Visión por computadora	15
2.1 Antecedentes	15
2.1.1 Antecedentes de la detección de rostros	16
2.1.1.1 Métodos basados en las características faciales.	17
2.1.1.2 Métodos basados en las características invariantes.	17
2.1.1.3 Métodos basados en la correspondencia de plantillas.	18
2.1.1.4 Métodos basados en la apariencia.	18
2.1.2 Antecedentes de la detección de marcas artificiales	18
2.2 Técnicas de visión por computadora para la detección de escenarios	19
2.2.1 Espacio de color HSV	20
2.2.2 Segmentación	21
2.2.2.1 Segmentación basada en píxel	22
2.2.2.2 Segmentación basada en área	23
2.2.2.3 Segmentación basada en bordes	24
2.2.2.4 Segmentación basada en modelos	26
2.2.3 Extracción de contornos	26
2.2.4 Morfología matemática	26
2.2.4.1 Erosión y dilatación.	27
2.2.4.2 Apertura y cierre.	28

2.2.5	Contornos activos -----	29
2.2.6	Detección de rostros -----	33
2.2.6.1	Extracción de características para detección de rostros -----	33
2.1.1.1	Entrenamiento de un clasificador para detección de rostros -----	36
	Resumen del capítulo -----	37
Capítulo 3 Arquitectura-----		38
3.1	Módulos del sistema-----	38
3.1.1	Módulo Detector -----	41
3.1.1.1	Módulo de captura-----	41
3.1.1.1.1	Características de las cámaras-----	42
3.1.1.2	Módulo para la detección de marcas-----	43
3.1.1.2.1	Características de las marcas -----	44
3.1.1.2.2	Detección de marcas -----	45
3.1.1.3	Módulo para la detección de rostros -----	49
3.1.2	Módulo de votación-----	51
3.1.3	Módulo de configuración-----	51
	Resumen del capítulo -----	52
Capítulo 4 Casos de estudio -----		53
4.1	Detección de marcas -----	53
4.2	Detección de rostros-----	55
4.3	Detección de escenarios-----	56
4.3.1	Primera etapa: Detección de escenarios sin obstrucciones -----	57
4.3.1.1	Caso 1: Detección del escenario colaborativo sin obstrucciones -----	57
4.3.1.2	Caso 2: Detección de escenarios proyectivos sin obstrucciones-----	58
4.3.2	Segunda etapa: Detección de escenarios con obstrucciones por objetos-----	59
4.3.2.1	Caso 3: Detección del escenario colaborativo con obstrucciones -----	59
4.3.2.2	Caso 4: Detección de escenarios proyectivos con obstrucciones-----	60
4.3.3	Tercera etapa: Detección de escenarios con obstrucciones por objetos y por alumnos -----	61
4.3.3.1	Caso 5: Detección del escenario colaborativo con obstrucciones y alumnos -----	61
4.3.3.2	Caso 6: Detección del escenario pizarra con obstrucciones y alumnos--	62
4.3.3.3	Caso 7: Detección del escenario de proyección estereoscópica y video conferencia con obstrucciones y alumnos-----	63
4.3.4	Análisis de resultados-----	63
	Resumen del capítulo -----	64
Conclusiones -----		65
5.1	Trabajo Futuro -----	66
Bibliografía -----		68

Índice de figuras

Figura 1. Representación de una organización espacial tipo “tradicional”	4
Figura 2. Representación de una organización espacial tipo “activa”	4
Figura 3. Distribución de dispositivos dentro del aula	9
Figura 4. Áreas de trabajo dentro del Aula del Futuro.....	10
Figura 5. Definición de escenarios a partir de las áreas de trabajo.....	13
Figura 6. Representación del espacio de color HSV.....	21
Figura 7. Histograma para la selección del umbral	23
Figura 8. Segmentación de una imagen por crecimiento de regiones	23
Figura 9. Segmentación de una imagen por separación-fusión de regiones	24
Figura 10. Mascaras de convolución para la detección de bordes.....	25
Figura 11. Operadores más comunes para la detección de bordes	25
Figura 12. Representación de las operaciones de erosión y dilatación	28
Figura 13. Representación de la operación de erosión.....	28
Figura 14. Representación de la operación de dilatación.....	28
Figura 15. Representación de la operación de esqueletización.	29
Figura 16. Contorno activo representado por un conjunto de puntos.....	30
Figura 17. Energía elástica del contorno.....	31
Figura 18. Energía de la curvatura.....	31
Figura 19. Energía de la imagen.	32
Figura 20. Representación de tipos de características.....	34
Figura 21. Clasificación de características presentes en un rostro.	34
Figura 22. Cálculo de la imagen integral mediante puntos de referencia.	36
Figura 23. Representación de la arquitectura por módulos	39
Figura 24. Traducción de los módulos de la arquitectura a diagramas de clase	40
Figura 25. Diagrama de flujo para la determinación de escenarios.....	41
Figura 26. Representación del proceso de captura de la imagen	42
Figura 27. Distribución de las cámaras dentro del aula	42
Figura 28. Diagrama de flujo para la detección de marcas y determinación de escenarios.	44
Figura 29. Representación de marcas sobre las mesas de trabajo.	45
Figura 30. Imágenes capturadas por las cámaras de video	46
Figura 31. Segmentación de imágenes obtenidas por las cámaras de video.	47
Figura 32. Cierre de las imágenes a partir de la segmentación del color.....	48
Figura 33. Ajuste del contorno a las marcas.....	49
Figura 34. Formato de almacenamiento de parámetros.	52
Figura 35. Marcas utilizadas para la detección de escenarios.	53
Figura 36. Detección de marcas realizada por cada cámara.....	54
Figura 37. Detección de rostros de frente a las cámaras.	56
Figura 38. Pruebas para la detección del “escenario colaborativo” sin obstrucciones	58
Figura 39. Pruebas para detección de “escenarios proyectivos” sin obstrucciones.....	59

Figura 40. Pruebas para la detección del “escenario colaborativo” ante la presencia de obstrucciones	60
Figura 41. Pruebas para la detección de “escenarios proyectivos” ante la presencia de obstrucciones	61
Figura 42. Pruebas para la detección del “escenario colaborativo” ante la presencia de alumnos y objetos sobre las mesas de trabajo	62
Figura 43. Pruebas de detección del “escenario pizarra” con alumnos y objetos sobre las mesas de trabajo.....	62
Figura 44. Pruebas de detección del “escenario video conferencia” con alumnos y objetos sobre las mesas de trabajo	63

Índice de Tablas

Tabla 1. Descripción de la distribución de dispositivos.....	9
Tabla 2. Características de las cámaras de video.....	43
Tabla 3. Resultados de la detección de las marcas artificiales.	55
Tabla 4. Resultados de la detección de rostros.....	56
Tabla 5. Criterios de evaluación durante las pruebas aplicadas al sistema.....	64

Resumen general

En la actualidad, los avances tecnológicos nos han alcanzado y forman parte de nuestras vidas, permitiéndonos realizar tareas complejas en escenarios y contextos que hace apenas unos años eran impensables. En particular, en este trabajo nos centraremos en el uso de tecnologías como apoyo al problema educativo, y es que si bien ha habido muchas y muy diversos esfuerzos por introducir las Tecnologías de la Información y la Comunicación (TIC) como apoyo a la educación, no siempre han sido exitosos, debido entre otros factores, a la falta de metodologías que contemplen la complejidad y las características propias del proceso educativo.

Percibir el proceso de enseñanza-aprendizaje como un fenómeno complejo ha alentado a grupos de trabajo interdisciplinarios a estudiar de manera conjunta y articulado el problema, surgiendo proyectos como el Aula del Futuro, en el que se busca estudiar desde una perspectiva interdisciplinaria las dinámicas educativas y sociales que surgen en el aula, para entonces desarrollar y evaluar las tecnologías que mejor se adapten a los objetivos educativos que se persiguen. Así, el Aula del Futuro es un proyecto en el que un espacio educativo con capacidad para cuatro estudiantes y un profesor es enriquecido con técnicas de cómputo ubicuo y realidad aumentada, buscando promover las tareas de participación, cooperación y colaboración entre los estudiantes.

Un segundo objetivo es poder brindar al profesor las condiciones necesarias para que pueda llevar a cabo su labor, sin ser distraído o agobiado por la presencia de la nueva tecnología. En efecto, el proyecto no busca dar nuevas tareas y obligaciones a los profesores, sino por el contrario facilitar en lo posible su labor. ¿Cómo alcanzar este ideal si el profesor se ve obligado a activar, desactivar y calibrar dispositivos cada quince minutos?

Con el propósito de liberar a los profesores de estas tareas, en este trabajo se propone analizar la actividad dentro del aula de manera automática, utilizando para ello técnicas de visión por computadora, permitiendo detectar los diferentes escenarios en los que alumnos y maestros llevan a cabo actividades educativas. La detección de los escenarios dentro del aula permitirá determinar que dispositivos deben ser utilizados en ese momento.

Desde hace aproximadamente treinta años el área de visión por computadora ha formado parte de un sistema multidisciplinario en el que intervienen disciplinas como: inteligencia artificial, robótica, procesamiento de señales, reconocimiento de patrones, teoría de control, psicología y otras más. En ese sentido, en este trabajo se busca acercar las técnicas de visión por computadora al área educativa, con la premisa de que la detección de escenarios puede ser un punto de partida para la adaptabilidad de los espacios educativos.

La tesis está organizada de la siguiente forma: el *primer capítulo* presenta las características de las aulas y su evolución, los problemas encontrados en la incorporación de tecnología al proceso de enseñanza-aprendizaje, así como las características del proyecto el Aula del Futuro y la metodología a seguir para la detección de escenarios. En el *segundo capítulo*, se describen las técnicas de visión por computadora usadas para la detección de escenarios, a partir de la detección de marcas y la detección de rostros. Por lo cual, en el *tercer capítulo* presenta la arquitectura del sistema y la integración de módulos que permiten la detección de escenarios. El *cuarto capítulo* describe los casos de uso y las pruebas realizadas para analizar la funcionalidad del sistema en la detección de escenarios. Finalmente, en el *quinto capítulo* se presentan las conclusiones obtenidas del desarrollo de esta tesis.

Capítulo 1

El aula

En este capítulo presentaremos de manera sucinta el problema de introducir tecnología al aula, tal como este problema es analizado hoy día. Asimismo presentaremos el proyecto del Aula del Futuro, marco de referencia para el trabajo presentado en esta tesis. Finalmente, presentamos los límites de dicho proyecto, mismos que dan origen a la problemática que aquí se trata de resolver.

1.1 Introducción

Vivimos en una sociedad que enfrenta cambios considerables día con día, en donde el uso de tecnologías se hace cada vez más popular, permitiendo comunicarnos de diversas formas, buscar, intercambiar información e incluso como fuente de diversión. Las Tecnologías de la información y la comunicación (TIC) son omnipresentes no sólo en las diferentes áreas de conocimiento, sino también en las actividades de nuestra vida diaria; por ejemplo, los teléfonos celulares, videojuegos, hornos de microondas, computadoras, etc.

A pesar de toda esta revolución tecnológica, diversos autores coinciden en que el área educativa se ha quedado rezagada [GH08,ES04,AC08,Sal97], debido básicamente a que, si bien los dispositivos se multiplican, las prácticas docentes y las dinámicas escolares permanecen inamovibles: todos en el mismo lugar, haciendo las mismas cosas, al mismo tiempo.

Con el propósito de estudiar este fenómeno surge en 2006 el proyecto del Aula del Futuro, auspiciado por el Macroproyecto de Tecnologías para la Universidad de la Información y la Computación (MTUIC) en la Universidad Nacional Autónoma de México. El equipo interdisciplinario detrás de este esfuerzo busca desarrollar y evaluar tecnologías innovadoras, que impacten de manera significativa el aprendizaje de los estudiantes y las

dinámicas dentro del aula, llevándola a un esquema de trabajo de modalidad mixta (tanto presencial como a distancia), con el objetivo de proporcionar un planteamiento educativo más participativo y colaborativo.

El proyecto reúne a más de una docena de académicos e investigadores de diferentes perfiles y formaciones, creando una red interdisciplinaria abocada a estudiar los fenómenos educativos mediados por tecnología. Así, el Aula del Futuro es un aula-laboratorio que permite a los investigadores generar y evaluar propuestas tecnológicas-pedagógicas innovadoras (realidad aumentada, sistemas colaborativos, proyección estereoscópica, interfaces adaptivas, etc.). Estas propuestas involucran la utilización de una gran cantidad de dispositivos, que son calibrados, activados y desactivados desde una cabina por uno de los integrantes del equipo de investigadores.

1.2 Antecedentes

Históricamente, las escuelas se mantenían encerradas en sí mismas desconectándose de la realidad externa. En los últimos años, se han creado vínculos constantes y planificados con el exterior que facilitan el flujo de estímulos e información en su entorno [DV97]. Esta apertura ha llevado a un replanteamiento de los espacios educativos, en donde, si bien se mantiene la estructura institucional en los edificios escolares, se dispersa cada vez más en un conjunto de espacios y escenarios educativos con objetivos complementarios, yuxtapuestos, paralelos o solapados con la institución escolar. El espacio educativo se define como el lugar donde la comunidad educativa reflexiona, con el objetivo de conseguir un crecimiento intelectual, personal y humano, identificando los conocimientos relevantes y significativos, que les permita integrarse con éxito en la sociedad actual [DV97].

Si bien es cierto que el espacio educativo no es determinante en la configuración de las relaciones educativas, también lo es el hecho de que facilita o dificulta algunos procesos, por lo que se cree conveniente analizar los elementos que conforman el espacio educativo (distribución, ubicación, tecnologías) con el propósito de observar si dichos elementos

coadyuvan a la creación de un proceso educativo que permita la renovación y mejora en la calidad de los procesos de enseñanza.

En la actualidad, los espacios educativos se han convertido en instrumentos de convivencia y educación social, conocidos como “Aulas”. En ellas se da lugar gran parte la actividad de los estudiantes; se trata del lugar donde se intenta dar respuesta a un conjunto de demandas con características comunes, derivadas del proceso educativo de los alumnos.

Con respecto a la creación de los espacios físicos de interacción (las aulas), en la investigación realizada por María Isabel Cano, [CB90] se plantean algunos principios como hipótesis de trabajo, que merecen ser retomados aquí.

Principio No. 1: El ambiente de la clase ha de posibilitar el conocimiento de todas las personas del grupo y el acercamiento de unos hacia otros. Progresivamente ha de hacer factible la construcción de un grupo humano cohesionado con los objetivos y metas comunes.

Principio No. 2: El entorno escolar ha de facilitar a todos el contacto con materiales y actividades diversas que permitan abarcar un amplio abanico de aprendizajes cognitivos, afectivos y sociales.

Principio No. 3: El medio ambiente escolar ha de ser diverso, debiendo trascender la idea de que todo aprendizaje se desarrolla entre las cuatro paredes del aula. Deberán ofrecerse escenarios distintos ya sean contruidos o naturales dependiendo de las tareas emprendidas y de los objetivos perseguidos.

Principio No. 4: El entorno escolar ha de ofrecer distintos sub-escenarios de tal forma que las personas del grupo puedan sentirse acogidas, según distintas expectativas e intereses.

Principio No. 5: El entorno ha de ser construido activamente por todos los miembros del grupo al que acoge, viéndose en él reflejadas sus peculiaridades, su propia identidad.

Con base en estos principios se muestran algunos ejemplos sobre la creación y distribución de espacios educativos:

A) Ejemplo de organización espacial tipo “Tradicional”

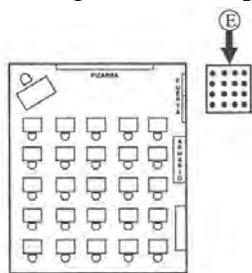


Figura 1. Representación de una organización espacial tipo “tradicional”, [CB90]

- **Estructura de comunicación en clase:**
 - Unidireccional
 - Grupal
 - Informativa académica formal
- **Características de las actividades:**
 - Individuales
 - Competitivas
 - La misma actividad para todos y al mismo tiempo

B) Ejemplo de organización espacial tipo “Activa”

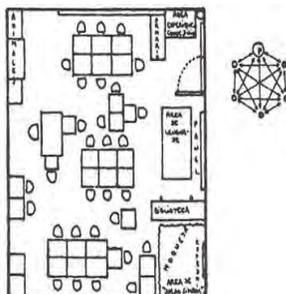


Figura 2. Representación de una organización espacial tipo “activa”, [CB90]

- **Estructura de comunicación en clase:**
 - Bidireccional (todos son emisores y receptores)
 - Grupal e individual
 - Integradora de contenidos “formales” e “informales”
- **Características de las actividades:**
 - Opcionalidad del alumno
 - Grupales e individuales
 - Cooperativas
 - Posibilidad de actividades distintas y simultáneas

Como puede observarse, en estos enfoques modernos el aula se considera no sólo como un punto de encuentro, sino como el lugar en donde el clima, el ambiente y los equipamientos, faciliten el conocimiento del grupo y el acercamiento de los alumnos, permitiéndoles la incorporación en un amplio abanico de aprendizajes cognitivos, afectivos y sociales [DV97]. Debe ser un espacio acogedor donde los alumnos, aún con distintos estados de ánimo, estén dispuestos a responder ante el desarrollo de las actividades que el profesor plantea en el proceso de enseñanza.

1.2.1 La tecnología en el aula

Los avances vertiginosos de las TIC ofrecen a los estudiantes nuevas opciones que, de manera imprecisa, se han denominado “virtuales” [DV97], pues les permite el acceso a una multiplicidad de recursos y personas de manera ubicua, viniendo a romper aún más el concepto convencional del espacio educativo.

Así, es innegable que las TIC han impactado la educación, desdibujando en buena medida los roles tradicionales del profesor y los alumnos, convirtiendo a ambos en sujetos activos en los procesos de enseñanza-aprendizaje, intercambiando roles y conceptos, métodos y estrategias para educar y ser educados. De hecho, en muchos ambientes la fusión de las TIC y la educación ha pasado a ser un argumento reconocido, analizado y estudiado a favor de los cambios en la formación y actualización docente, generando una obstinación por mejorar la enseñanza a través de la tecnología [Lop08].

Diversos estudios sobre el uso de las TIC [ES04,AC08,Sal97,Gom08,Gam07] destacan la necesidad de incorporarlas en los diferentes niveles educativos (desde la educación básica hasta la superior), con un enfoque que supere el manejo básico de las herramientas y que trascienda la visión tradicional de la educación [Gom08]. Estas investigaciones enfatizan que la introducción de las TIC al aula no debe hacerse como un elemento accesorio, sino como un elemento clave en la reformulación del aprendizaje con nuevas herramientas tecnológicas.

Sin embargo, diversos estudios han mostrado la imposibilidad de que las TIC funjan por sí mismas como un elemento suficiente para el desarrollo educativo de la sociedad. En efecto, por un lado se presupone que su sola incorporación a la educación es un paso importante para el bienestar social, dejando de lado todo el problema de producción y administración de contenidos. Por otro lado, si bien hay evidencia de que la incorporación de las TIC al proceso de enseñanza-aprendizaje puede mejorar los procesos de construcción y

socialización del conocimiento, la no actualización de los procesos y las estrategias educativas en los profesores hacen que la educación se mantenga anclada y no se obtenga el éxito esperado.

Hawkins [Haw02] reporta cómo los avances tecnológicos se puede observar en casi todas las actividades humanas, pero no en las aulas. El autor señala: "...mientras la sociedad goza de avances en todo gracias a la tecnología, las aulas siguen desde hace muchos años con los procesos tradicionales de enseñanza: estudiantes en filas, con lápiz y papel en las manos; un profesor escribiendo en el pizarrón los datos más importantes, mientras los estudiantes escriben afanosamente lo que se dice, esperando poder repetirlo en un examen".

Un ejemplo claro de las consecuencias de pretender resolver el problema educativo desde una perspectiva únicamente tecnológica, se observa en el video "Tecnología o Metodología" [Cub07], en donde a una profesora enseña a sus alumnos las tablas de multiplicar. En una primera toma se observa a la profesora señalando las multiplicaciones anotadas sobre una lámina y dirigiendo al coro de voces que recitan (junto con ella), las multiplicaciones. En una segunda toma se ve que el salón de la profesora ha sido completamente remodelado para integrar lo último en tecnología (pizarrones interactivos, dispositivos móviles, una computadora por alumno, etc.). Sin embargo, el uso que la profesora hace de todos estos dispositivos, es proyectar al grupo la misma lámina con las multiplicaciones, esta vez digital, y pedirles que reciten igual que antes las multiplicaciones. Es inevitable: la introducción de tecnologías al aula debe acompañarse de cambios en los procesos educativos.

En ese sentido, hoy día es ampliamente reconocido que las TIC, por sí solas, no pueden alcanzar el objetivo de transformar los procesos por los que los alumnos adquieren conocimientos y habilidades. Es necesario plantear un cambio en los procesos y modelos educativos que permitan dejar de lado las estrategias en las que los alumnos se restringen a asimilar conocimiento mediante la práctica y la repetición; para pasar a un esquema que los haga participes de la construcción de su propio conocimiento [ES04].

Observando esta falta de metodología al hacer uso de tecnologías en el proceso de enseñanza-aprendizaje, se ofrecen nuevas propuestas sobre los modelos metodológicos para el desarrollo educativo, en donde los fundamentos del aula tradicional “todos al mismo tiempo, todos en el mismo lugar haciendo las mismas actividades” trasciendan [Gam07], permitiendo a los estudiantes sentirse responsables de la adquisición y formación de sus propios conocimientos. En este sentido, en el Aula del Futuro se proponen espacios de trabajo colaborativos, con el propósito de hacer partícipes a los alumnos y profesores en la creación de conocimientos, y puedan aprovecharse al máximo los recursos tecnológicos dentro del aula, permitiendo desarrollar nuevas propuestas tecnológicas.

1.3 El proyecto: el Aula del Futuro

El Aula del Futuro es un espacio de investigación en el que un grupo interdisciplinario de académicos busca proponer y evaluar modelos educativos innovadores mediados por tecnología; escenarios de enseñanza-aprendizaje en los que tecnologías de nueva generación (más naturales en su uso y menos invasivas) promuevan la participación y creatividad de los alumnos, faciliten el trabajo colaborativo y coadyuven a la flexibilización de las modalidades presenciales y a distancia, permitiendo esquemas mixtos.

Así, el Aula del Futuro es un aula laboratorio en donde el espacio de evaluación está definido por dos áreas: 1) cabina de control, que es el espacio en donde se encuentran los equipos de cómputo y de control que permiten activar o desactivar los componentes tecnológicos que el profesor utiliza durante una sesión de evaluación. 2) aula experimental, que emula un aula universitaria, diseñada con características especiales que permiten a los investigadores realizar tareas de evaluación de tecnologías, así como diseñar nuevas propuestas tecnológicas como apoyo para la enseñanza.

1.3.1 Características del Aula del Futuro

El Aula del Futuro, a diferencia de las aulas convencionales tiene las siguientes características:

- Un espacio de trabajo flexible: los alumnos trabajan en mesas de trabajo móviles, que se adaptan a los distintos tipos de actividades
- Proyectors y pantallas de proyección, utilizados para mostrar el material que el profesor y el equipo de investigadores han preparado
- Una computadora para el profesor en donde puede presentar la información de una sesión de clases a los alumnos
- Cuatro tablet-pc utilizadas por los alumnos para realizar actividades de aprendizaje que el profesor y el equipo de investigadores han preparado para la sesión
- Un equipo de video-conferencia que permite establecer contacto (al profesor y los alumnos) con otros grupos de estudiantes con el propósito interactuar en espacios presenciales y a distancia
- Un pizarrón interactivo en donde el profesor refuerza su exposición con anotaciones
- Un sistema de proyección estereoscópica que trasciende los conceptos descritos en sistemas visualmente planos (proyectors o pantallas) a un entorno en tercera dimensión
- Un arreglo de cinco cámaras video, distribuidas en las esquinas y en el centro del aula que se conectan a una tarjeta de captura de video y que permiten observar desde la cabina de control lo que ocurre durante una sesión de evaluación.

En conjunto, estos dispositivos tecnológicos permiten desarrollar plataformas y materiales didácticos que integren tecnologías de visualización estereoscópica, ambientes colaborativos, sistemas adaptables y sistemas móviles, de modo a evaluar su impacto en el proceso de enseñanza aprendizaje, y en particular, en las habilidades relacionadas con la autogestión del aprendizaje, el trabajo colaborativo y la creación y gestión de comunidades de aprendizaje. De acuerdo con estas propuestas es necesario tener presente los cambios y adaptaciones que serán demandados por las nuevas tecnologías.

1.3.2 Distribución de dispositivos

El espacio de evaluación del Aula del Futuro es de 7.5 x 6.5 m. La puerta principal comunica al espacio de evaluación del aula y la cabina de control, cada una con puertas independientes que sirven para impedir la entrada y salida de ruido que interrumpa las actividades de evaluación. Un pared con ventanas de cristal oscuro que divide la cabina de control y el área de evaluación que permite observar lo que los involucrados en las evaluaciones realizan; la distribución de dispositivos se observa en la Figura 3 y su descripción en la Tabla 1.

0a	Puerta principal	4	Espejo de proyección hacia las mesas de trabajo
0b	Puerta de acceso a la cabina de control	5	Pizarra interactiva
0c	Puerta de acceso al aula	6	2 Projectores para proyección estereoscópica
0d	Ventanas de cristal oscuro	7	Pantalla para proyección estereoscópica y video conferencia
1	Mesas de trabajo para alumnos	8	Cámaras de video
1a	Mesa de trabajo del profesor	9	Micrófonos ambientales
2	Tablets Pc donde los alumnos realizan tareas	10	Consola de audio
2a	Computadora del profesor	11	Equipo de video conferencia
3	Proyector	12	Red inalámbrica

Tabla 1. Descripción de la distribución de dispositivos

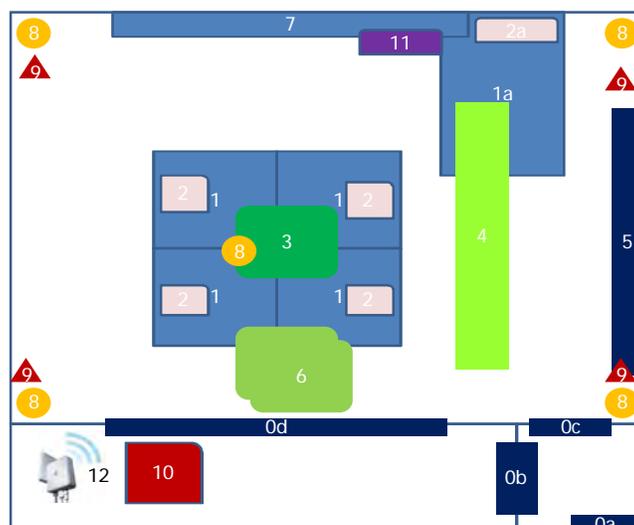


Figura 3. Distribución de dispositivos dentro del aula

Esta distribución dentro del aula permite identificar tres áreas de trabajo (que identificamos como escenarios y se muestran en la Figura 4), en donde los alumnos y profesores participan en el desarrollo de actividades que el grupo de investigadores elaboran y que les permite evaluar propuestas tecnológicas de apoyo para el proceso educativo. Estos escenarios serán identificados como a) escenario colaborativo, b) escenario de proyección en pizarra y c) escenario de proyección estereoscópica.

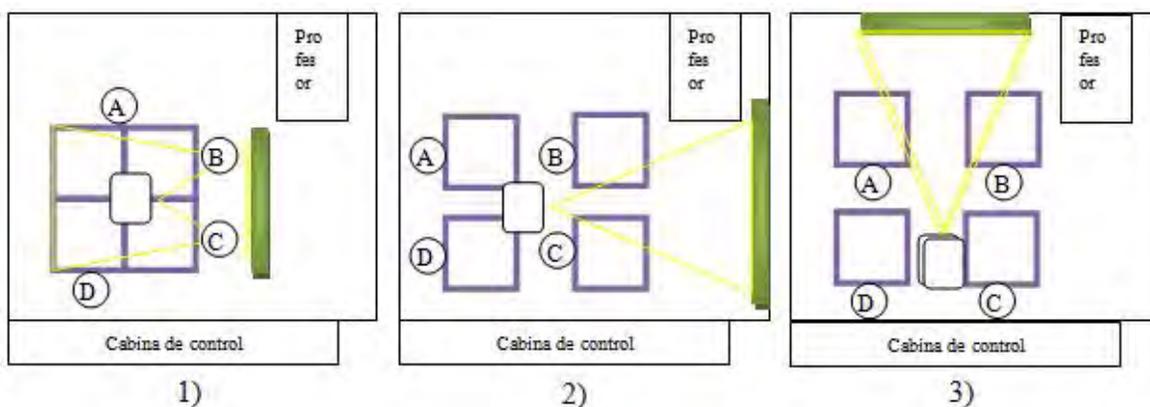


Figura 4. Áreas de trabajo dentro del Aula del Futuro. 1) Área de trabajo colaborativo. Las mesas de trabajo se agrupan para que los alumnos trabajen en equipo; el agrupamiento permite que la información sea proyectada sobre las mesas de trabajo. 2) Área de trabajo con proyección en pizarra digital. Las mesas de trabajo se separan y la información es proyectada sobre una pantalla digital. 3) Área de trabajo con proyección estereoscópica y sistema de videoconferencia. La información es proyectada por dos cañones de proyección generando un sistema estereoscópico; también es utilizado como sistema de videoconferencia.

1.3.3 Límites del proyecto y modificaciones necesarias

Como se ha mencionado, el objetivo del proyecto es generar un espacio que potencie las relaciones y las dinámicas que surjan entre los alumnos y los profesores. Para este objetivo es importante que el profesor pueda concentrarse en su curso, más que en la tecnología que va a utilizar.

Gracias a la distribución y características de los dispositivos con los que se cuenta dentro del aula, llevar a cabo el análisis de la actividad mediante la detección de escenarios a través de técnicas de visión por computadora, no requiere adaptaciones significativas en cuanto al espacio de evaluación, sino mas bien, realizar adaptaciones al proceso de

activación y calibración de los dispositivos para que sea posible dotar de autonomía al aula, y por ese medio, que responda automáticamente a diferentes eventos durante las sesiones.

1.4 Problemática

Como se ha expuesto anteriormente, se ha tratado de introducir tecnología al proceso educativo que permita a los alumnos participar activamente en la gestión de su propio conocimiento, y los profesores puedan apoyarse de estos recursos para llevar a cabo satisfactoriamente la labor de instruir. Sin embargo, aún queda mucho trabajo por hacer. En particular, desde el punto de vista del profesor, puesto que todavía es necesario activar, desactivar y calibrar los diferentes dispositivos tecnológicos en periodos de tiempo muy cortos (de acuerdo con las actividades que se han preparado y los dispositivos requeridos), lo que desvía al profesor de su principal objetivo.

Con el propósito de generar tecnologías que permitan al profesor deslindarse de llevar a cabo estas actividades, en el Aula del Futuro, se analizan, desarrollan y evalúan propuestas y metodologías que permitan hacer el proceso de calibrar, activar y desactivar los dispositivos de manera automática el aula misma, permitiendo así al profesor olvidarse de estas tareas que interrumpen sus actividades docentes.

Una manera de lograr que el aula administre automáticamente los recursos tecnológicos, consiste en la detección de las diferentes actividades educativas que se llevan a cabo, a través de determinar los escenarios asociados a cada una de ellas (arreglo de muebles, punto de atención de los estudiantes, etc.). Que el aula identifique en cuál de los diferentes escenarios se encuentran trabajando los alumnos y el profesor, permitirá analizar y generar estrategias que permitan determinar que dispositivos son requeridos en una sesión, y por consecuencia que el aula se encargue de su administración y control.

1.5 Solución

Gracias a las características del Aula del Futuro, la detección de escenarios se puede realizar a través de la detección de elementos dentro del aula, a partir de la aplicación de técnicas de procesamiento de imágenes y visión por computadora a imágenes obtenidas de las cámaras de video con las que ya se cuenta en el aula. En este sentido, los elementos que al ser detectados permiten identificar en que escenario se llevan a cabo las actividades educativas son las mesas de trabajo y los rostros de los alumnos, debido a que proporcionan información sobre los diferentes escenarios.

La detección de las mesas de trabajo y alumnos dentro del aula permite determinar a través de un análisis de votación el escenario que cada cámara detecta, proporcionando información del escenario presente en el aula.

1.6 Metodología

El poder reconocer el tipo de actividad que se está realizando en un cuarto cerrado es un tema de investigación abierto, estudiado desde diferentes perspectivas en laboratorios de todo el mundo. En nuestro caso, buscamos en primer lugar simplificar el problema, ejecutando una serie de actividades que permitan generar una solución eficiente y original al problema planteado.

Estas actividades son numeradas a continuación:

- 1 ***Definición de escenarios.*** A partir de las áreas de trabajo identificadas, se definen arreglos “rígidos” de los muebles y los alumnos dentro del aula, a los cuales llamaremos “escenarios”. En efecto, el análisis de la actividad que nuestro sistema realiza se basa en la definición de tres escenarios posibles como se muestra en la Figura 5, en donde cada escenario representa un tipo específico de acomodo de los alumnos dentro del aula.

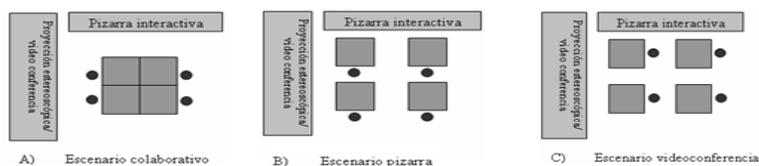


Figura 5. Definición de escenarios a partir de las áreas de trabajo dentro del Aula del Futuro. A) Escenario colaborativo en donde los alumnos realizan actividades en equipo. B) Escenario de proyección en pizarra interactiva y en donde los alumnos observan lo que el profesor expone en una pizarra interactiva. C) Escenario de proyección estereoscópica y videoconferencia en donde los alumnos analizan una proyección estereoscópica o participan en una videoconferencia.

A partir de esta definición de escenarios, es posible intentar la identificación del escenario actual durante una sesión, procurando la detección y análisis, de dos elementos presentes en los escenarios: *las mesas* y *los alumnos*. Así, el algoritmo principal busca detectar la distribución de las mesas de trabajo y la ubicación de los alumnos, para a partir de ello tratar de identificar el escenario.

- 2 **Detección de mesas y alumnos.** Un segundo paso en la solución del problema consiste en determinar, de entre diversas posibilidades, aquellos algoritmos que permiten detectar de manera eficiente y robusta estos dos elementos (alumnos y mesas) dentro de imágenes capturadas por cámaras de video.
- 3 **Reconocimiento del escenario.** Una vez detectados los alumnos y las mesas se establece un mecanismo que permita decidir de entre cuatro cámaras de video cuál o cuáles detectan las mesas, y al mismo tiempo debe decidir cuál o cuáles detectan a los alumnos. Combinando la información de qué cámaras han detectado las mesas y qué cámaras han detectado a los alumnos, el sistema determinara a través de un proceso de votación a que escenario corresponde la distribución de elementos detectada.

Cada uno de estos problemas resulta, en sí mismo, un reto que debe ser superado a través de análisis, desarrollo, pruebas y refinamiento. En todos los casos, siempre se busca que la respuesta del sistema sea lo más robusta y eficientemente posible.

1.7 Contribución y relevancia

Dados los avances en el área de visión computacional, es posible combinar técnicas de detección y reconocimiento de objetos, trayendo nuevas soluciones a problemáticas que antes requerían ser realizadas por un humano. En nuestro caso, el sistema de detección permitirá, además de ofrecer una primera aproximación a la solución del problema de adaptabilidad ya descrito, sentar las bases para futuros desarrollos que requieran de mayor información sobre lo que ocurre en el aula; por ejemplo, llevar a cabo un seguimiento automatizado de las acciones que los alumnos y los profesores realizan: como se organizan durante las diferentes actividades y como se conformaron los grupos de trabajo analizando por ejemplo que actividad tuvo un alumno en particular, etc.

Resumen del capítulo

En este capítulo se analizó de manera breve los espacios educativos y los principios que se deben tomar en cuenta para mejorar el aprendizaje de los alumnos. Se describe también el proyecto el Aula del Futuro y las características del espacio en donde se llevan a cabo actividades de investigación para mejorar el proceso educativo mediado por tecnología. Analizando las características del aula, se describieron los escenarios detectados a partir de las áreas de trabajo, mismos que describen la problemática que aquí se trata de resolver a partir de la aplicación de técnicas de visión por computadora. Estas técnicas serán descritas en el siguiente capítulo.

Capítulo 2

Visión por computadora

En este capítulo se muestran de manera breve los antecedentes de visión por computadora, particularmente una descripción rápida sobre las técnicas para la detección de rostros y marcas artificiales, así mismo, se analizan las técnicas de visión por computadora y procesamiento de imágenes utilizadas para la detección de marcas artificiales y rostros, elementos que permiten la detección de los diferentes escenarios dentro del Aula del Futuro.

2.1 Antecedentes

El objetivo de la visión por computadora es el reconocimiento de los diferentes objetos que se encuentran en una imagen digitalizada, así como la descripción de las relaciones y de su ambiente [PM95]. En un sistema de visión, la tarea del reconocimiento de objetos y su descripción es desarrollada en una serie de tareas consecutivas como: adquisición de imágenes, procesamiento, reconocimiento, medición, decisión, etc., que contribuyen a este proceso. La visión por computadora es un área multidisciplinaria que busca modelar matemáticamente los procesos de la percepción visual de los seres vivos mediante el análisis, procesamiento e interpretación de imágenes provenientes desde distintos tipos de sensores (fundamentalmente cámaras) y generar programas que permitan simular estos procesos por computadora [San05]. El desarrollo de sistemas de visión requiere una combinación de etapas (*e.g.* eliminación de ruido, ajuste de contraste, reconocimiento de patrones, interpretación de imágenes) para reconocer los elementos presentes en una escena.

Estos sistemas implementan diferentes técnicas (por mencionar algunas: extracción de puntos característicos, detección de líneas y círculos, reconocimiento de gestos, detección y reconocimiento de objetos, visión estéreo, entre otras) para el procesamiento de imágenes

en cada una de las etapas. La aplicación de estas técnicas difiere de un sistema a otro, debido a la variación del espacio donde son capturadas las imágenes, o las características particulares del problema a resolver, que llevan a la implementación de técnicas específicas. En este sentido, de acuerdo a las características de los escenarios presentados en el capítulo anterior, se realiza la detección de marcas artificiales que permitan identificar la distribución de las mesas y al mismo tiempo la detección de rostros para conocer la posición de los alumnos.

2.1.1 Antecedentes de la detección de rostros

La detección de rostros inició en la década de los setenta con algoritmos basados en técnicas heurísticas y antropométricas [GHL71],[Kan77]. Los resultados no eran del todo buenos, pues presentaban gran cantidad de errores y eran muy sensibles a cambios, es decir, detectaban un rostro en una fotografía con fondo blanco, pero si se quería reconocer un rostro con gafas se tenía que modificar todo el algoritmo. No fue sino hasta la década de los noventa, cuando las computadoras permitieron realizar tareas de procesamiento con mayor rapidez y contaron con mayor capacidad de almacenamiento, que estas técnicas fueron retomadas [MKN02].

Gracias al avance que estas técnicas alcanzaron, hoy en día se puede realizar la detección de rostros mediante diferentes enfoques o técnicas como lo muestra Ming-Hsuan *et.al* [MKN02], en su clasificación de las técnicas usadas para la detección de rostros, basada en cuatro categorías:

- Métodos basados en las características faciales.
- Métodos basados en características invariantes.
- Métodos basados en la correspondencia de plantillas.
- Métodos basados en la apariencia.

Estas categorías se describen a continuación:

2.1.1.1 Métodos basados en las características faciales.

Los métodos basados en las características faciales [MKN02] utilizan reglas derivadas del conocimiento de los rostros humanos, es decir, una regla para definir un rostro en una imagen es una suposición de que existen dos ojos que se encuentran simétricamente distanciados, una nariz y una boca. La relación entre estas características ayuda a determinar la presencia de un rostro en una imagen mediante sus distancias y posiciones relativas. La determinación de la existencia de un rostro se realiza mediante la extracción de las características a partir de la imagen y la evaluación de candidatos basados en las reglas descritas.

2.1.1.2 Métodos basados en las características invariantes.

Se basan en la hipótesis de que los humanos pueden detectar rostros y objetos en diferentes posiciones y condiciones de iluminación. De acuerdo a esta idea se han propuesto varios métodos que permitan localizar las características faciales y así deducir la presencia de un rostro en una imagen. Estos métodos toman en consideración características faciales, textura, color de piel y un conjunto de múltiples características que proporcionan información para el proceso de detección [MKN02].

Dentro de los métodos propuestos se encuentran el trabajo de Sirohey [SB+93], quien propone un método de localización de rostros, utilizando detectores de bordes y técnicas heurísticas para remover y agrupar los bordes, donde solo los bordes que pertenecen al contorno de un rostro son preservados. El trabajo de Chetverikov [CL93] se basan en un modelo de manchas y líneas que define las características de un rostro como un conjunto de dos manchas oscuras y tres manchas claras, que representan los ojos, pómulos y nariz ya que utiliza líneas para representar la división del contorno del rostro, los ojos y los labios, y por utilizar dos configuraciones triangulares para codificar la relación espacial entre las manchas.

Hay algunos otros trabajos dentro de esta categoría [MKN02], que utilizan diferentes métodos para extraer y representar las características faciales y cuyo objetivo es proporcionar métodos robustos para la detección de rostros logrando así el mayor porcentaje de rostros detectados.

2.1.1.3 Métodos basados en la correspondencia de plantillas.

En estos métodos se define un rostro como patrón estándar predefinido manualmente (frontal), o parametrizado por una función, en donde a partir de una imagen los valores de correlación junto con el patrón estándar son calculados para el contorno del rostro, los ojos, la nariz, y la boca independientemente. A partir de aquí la existencia de un rostro en la imagen es determinada en base a los valores de correlación. Sin embargo se ha notado que estos métodos son inadecuados para la detección de rostros debido al bajo rendimiento con variaciones en escala, pose y forma [MKN02].

2.1.1.4 Métodos basados en la apariencia.

Estos métodos se basan en un entrenamiento a partir de un conjunto de imágenes [MKN02]. El entrenamiento se puede realizar utilizando diferentes modelos estadísticos para encontrar características relevantes de las imágenes que son rostros y las que no lo son. Estos modelos estadísticos pueden ser modelos bayesianos [Dem68], eigenfaces [TP91], redes neuronales [AS93].

2.1.2 Antecedentes de la detección de marcas artificiales

La detección de marcas artificiales es una técnica que se utiliza en sistemas de visión por computadora que permiten detectar y reconocer objetos *e.g.* en los sistemas de navegación para robots [BS+00]. Una marca artificial es un patrón formado por un conjunto de figuras geométricas y colores que permiten la identificación de las figuras y que puede ser reconocido aplicando algoritmos de procesamiento de imágenes. El uso de marcas artificiales en los sistemas de visión por computadora permite obtener información del

entorno en donde operan dichos sistemas, pues las marcas artificiales son específicamente diseñadas para facilitar la detección bajo varios cambios de condiciones en el ambiente *e.g.* iluminación o ruido

Los sistemas de navegación no son el único campo de estudio donde la detección de marcas artificiales puede implementarse, ya que también se aplica la detección de marcas en sistemas como en el análisis de escenas, indexación de imágenes y adquisición de imágenes desde bases de datos, etc., basándose en la detección de marcas artificiales y en ocasiones marcas obtenidas del ambiente.

La aplicación de marcas artificiales para la detección y reconocimiento de objetos es el punto de partida mediante el cual los sistemas de visión obtienen información para determinar la ubicación y orientación de los objetos de interés *e.g.* la posición del robot y obstáculos dentro del entorno. El éxito de estos sistemas se basa en el diseño de las características de las marcas, tomando en consideración *e.g.* color, tamaño, forma, etc., que permiten su fácil detección y reconocimiento [LSM06].

Aunque no existe un método único para la detección y reconocimiento de las marcas, se han realizado diferentes trabajos en los cuales la detección de marcas es la pieza clave del buen desempeño de los sistemas de visión que las implementan [KI01,Sua05,TT05,BS+00,LSM06]. A continuación se describen las técnicas que permiten llevar a cabo la detección de marcas y detección de rostros.

2.2 Técnicas de visión por computadora para la detección de escenarios

Las técnicas para la detección de marcas contemplan la detección del color de las marcas registrando los píxeles que corresponden al color de la marca e ignorando todos los demás colores de la imagen; los píxeles registrados en la detección de color se analizan con el

propósito de encontrar los píxeles que correspondan al contorno de las marcas; el contorno es ajustado lo más cercano posible a las marcas colocadas sobre las mesas; de esta manera los contornos ajustados proporcionan la información de la distribución de las mesas de trabajo.

La detección de rostros se realiza a través de la extracción de características que los conforman. Estas características son extraídas a partir del entrenamiento de un clasificador que determina la presencia de rostros.

Por tal motivo a continuación se describen las técnicas utilizadas para llevar a cabo la detección de marcas y rostros en la detección de escenarios.

2.2.1 Espacio de color HSV

El espacio de color HSV (por sus siglas en inglés Hue, Saturation, Value – Tonalidad, Saturación, Valor), define un modelo de color en términos de sus componentes constituyentes que pueden ser representados mediante un cono [CAP02]. La Figura 1 muestra la relación que existe entre las componentes del espacio para la definición del color.

La tonalidad representa el tipo de color como un grado de ángulo cuyos valores van de 0° a 360° (algunas aplicaciones normalizan su representación de 0 a 100%) donde cada valor representa un color. La saturación representa la pureza del color, es decir, la proporción de luz pura de la longitud de onda dominante y la luz blanca necesaria para definir un color. Un color 100% puro tiene una saturación de 100% y no contiene luz blanca, así los colores que se alejan del color puro contienen una mezcla de luz blanca y del color puro. Un color que tiene una saturación de 0% es un blanco puro. El valor del color representa el brillo del color, dependiendo de la saturación podría ser un gris o un color más o menos oscuro; los valores que puede tomar están entre 0 y 100% [CAP02].

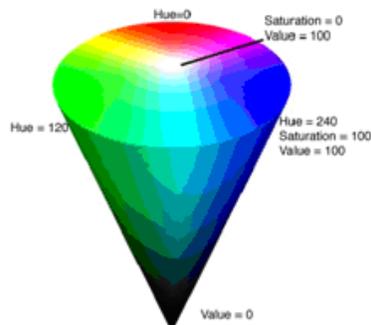


Figura 1. Representación del espacio de color HSV. [CAP02]

El espacio HSV facilita el trabajo en tareas que requieren hacer distinción entre diferentes colores y permite separar la componente de brillo que depende fuertemente de la iluminación. Debido a esta característica es posible hacer una comparación entre un color y otro, permitiendo seleccionar el color de interés. Esta característica es la que nos facilita realizar tareas como la segmentación de imágenes a partir de sus componentes de color.

2.2.2 Segmentación

Una habilidad que el ser humano posee para el reconocimiento de formas y objetos, se debe a que reconoce objetos en el ambiente mediante la segmentación de color [LP07], de esta misma manera en el área de visión por computadora se han realizado investigaciones sobre técnicas de segmentación, de formas y objetos en color dentro de una imagen, tratando de simular el proceso en que éste realiza dicha actividad.

La segmentación de imágenes describe el proceso de dividir una imagen en regiones similares y homogéneas [KA08]. Es uno de los procesos más importantes de un sistema automatizado de visión ya que permite extraer los objetos de la imagen para su posterior descripción y reconocimiento. Este proceso puede apoyarse del conocimiento previo de las características [KA08] como son el color, la forma, el tamaño, etc., de los objetos que se desean separar o reconocer facilitando así el proceso. La segmentación juega un papel importante en diferentes aplicaciones de procesamiento de imágenes y visión computacional [LP07] *e.g.* reconocimiento de objetos y seguimiento, recuperación de

imágenes en una base de datos, etc., pues los resultados dependen de la calidad de segmentación que se realice.

El resultado de la segmentación, constituye la formación de una imagen binaria, que es una representación de la imagen original. La imagen binaria es representada mediante dos valores posibles: 0 para los píxeles que no cumplen con el criterio de segmentación y 1 para los píxeles que pertenecen a los objetos segmentados. De acuerdo a Lukav et. al. y Jähne [LP07,Jah05], existen diferentes enfoques que se pueden aplicar al proceso de segmentación, y estos se clasifican en cuatro categorías en base a: píxeles, área, bordes y características físicas del modelo de color en el cual se representan las imágenes.

2.2.2.1 Segmentación basada en píxel

El proceso de segmentación basado en píxeles es realizado mediante dos enfoques principales [KA08], el primero comprende las técnicas que utilizan la información del histograma, observando la aparición de tonalidades o niveles de gris en la imagen y clasifica los píxeles de acuerdo a este valor, seleccionando un umbral T a partir del cual los píxeles que pertenecen a éste son considerados como parte de los objetos de interés. Una manera de seleccionar el umbral es a partir de los máximos locales ubicados en el histograma como se observa en la Figura 2 en donde el proceso para obtener una imagen umbralizada U a partir de una imagen I es definido por la ecuación 1.

$$U(x, y) = \begin{cases} 1 & \text{si } I(x, y) > T \\ 0 & \text{si } I(x, y) \leq T \end{cases} \quad (1)$$

En donde la imagen umbralizada U se obtiene a través de la evaluación de cada píxel en la imagen con respecto al umbral T .

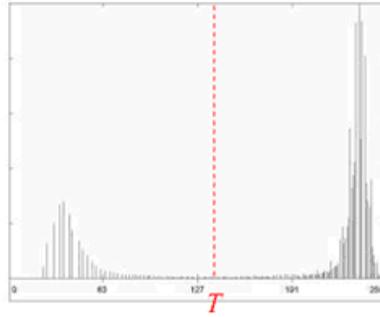


Figura 2. Histograma para la selección del umbral a partir de los valores máximos y mínimos [GW02]

El segundo enfoque está basado en el agrupamiento de píxeles en el espacio de color y se realiza mediante la selección de un número de clases $Z_1 \dots Z_n$ dentro del espacio, con un vector de probabilidad m-dimensional $p_{i1} \dots p_{im}$, donde el valor probabilístico para cada píxel p_i se determina a partir de una imagen de color C . La componente p_{ik} determina la probabilidad de que un píxel corresponda a la clase Z_k con $k = 1, \dots, m$ [KA08].

2.2.2.2 Segmentación basada en área

La segmentación basada en el área localiza zonas uniformes en la imagen; existen dos formas de llevar a cabo la segmentación dentro de esta categoría. La primera se enfoca en el crecimiento de regiones, cuya idea consiste esencialmente en crecer una región a través de un punto de partida llamado semilla y mediante el cual la región comienza a crecer hasta agrupar los píxeles; un ejemplo de segmentación mediante el crecimiento de regiones se muestra en la Figura 3.

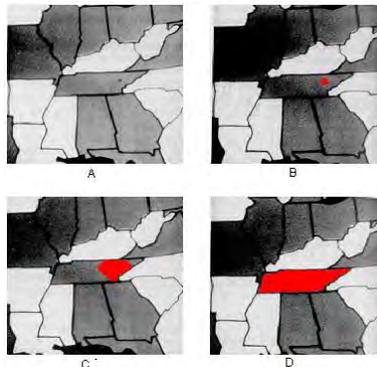


Figura 3. Segmentación de una imagen por crecimiento de regiones a partir de una semilla. Inicialización de la semilla (A). Crecimiento de la semilla (B) y (C). Segmentación del área por semilla (D). [GW87]

La segunda se enfoca en la separación-fusión de áreas, iniciando con áreas no uniformes que pueden dividirse en sub-regiones, las cuales luego se fusionan hasta alcanzar un criterio de uniformidad como se muestra en la Figura 4.

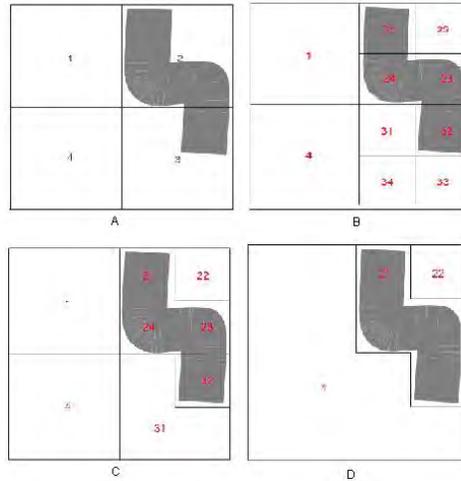


Figura 4. Segmentación de una imagen por separación-fusión de regiones. Separación de regiones (A) y (B). Fusión de regiones (C) y (D). Imagen segmentada (D) [GW02]

2.2.2.3 Segmentación basada en bordes

Un borde es el límite entre dos regiones con propiedades relativamente distintas en el nivel de gris que presenta cambios bruscos en la tonalidad [GW02]. La segmentación basada en bordes usa medidas no uniformes o discontinuidades dentro de una vecindad de los píxeles para determinar si un píxel pertenece a un borde, empleando operadores locales en distintas aproximaciones de la primera y segunda derivada de los niveles de gris de la imagen [GW02]. El método más común en aplicaciones de procesamiento de imágenes para la obtención de las derivadas es el gradiente ∇ . El operador gradiente se basa en diferencias entre los niveles de gris de la imagen. En este caso la derivada es aproximada por la diferencia de píxeles adyacentes mediante las ecuaciones 6 y 7 que forman las mascararas de convolución H_F y H_C respectivamente Figura 5.

$$G_x(i, j) = \frac{\partial f(x, y)}{\partial x} \approx \nabla_x f(x, y) = f(x, y) - f(x - 1, y) \quad (6)$$

$$G_y(i, j) = \frac{\partial f(x, y)}{\partial y} \approx \nabla_y f(x, y) = f(x, y) - f(x, y - 1) \quad (7)$$

Donde G_x es el gradiente en la dirección x y G_y el gradiente en la dirección y de la imagen. El gradiente es obtenido aplicando una convolución de la imagen con las mascararas H_F y H_C Figura 5. La Figura 6 muestra los operadores más utilizados para la detección de bordes.

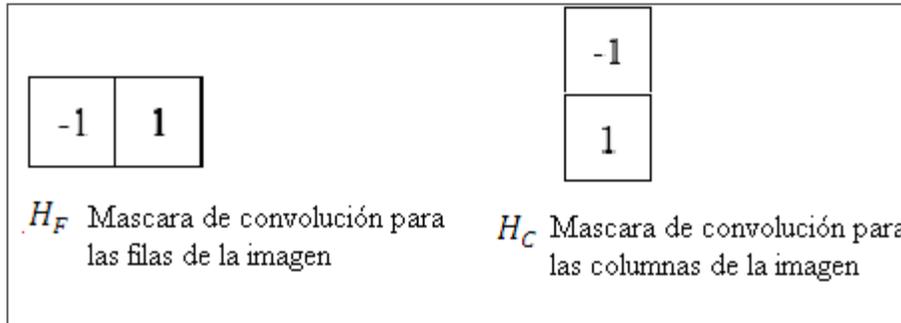


Figura 5. Mascaras de convolución para la detección de bordes

Operador	Gradiente de renglón	Gradiente de columna
Diferencia de pixeles	$\begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 & -1 \\ 0 & 0 & 0 \end{bmatrix}$	$\begin{bmatrix} 0 & -1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}$
Diferencia de pixeles separada	$\begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 1 & 0 & -1 \\ 0 & 0 & 0 \end{bmatrix}$	$\begin{bmatrix} 0 & -1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}$
Roberts	$\begin{bmatrix} 0 & 0 & -1 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}$	$\begin{bmatrix} -1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}$
Prewit	$\frac{1}{3} \begin{bmatrix} 1 & 0 & -1 \\ 1 & 0 & -1 \\ 1 & 0 & -1 \end{bmatrix}$	$\frac{1}{3} \begin{bmatrix} -1 & -1 & -1 \\ 0 & 0 & 0 \\ 1 & 1 & 1 \end{bmatrix}$
Sobel	$\frac{1}{4} \begin{bmatrix} 1 & 0 & -1 \\ 2 & 0 & -2 \\ 1 & 0 & -1 \end{bmatrix}$	$\frac{1}{4} \begin{bmatrix} -1 & -2 & -1 \\ 0 & 0 & 0 \\ 1 & 2 & 1 \end{bmatrix}$
Frei-Chen	$\frac{1}{2 + \sqrt{2}} \begin{bmatrix} -1 & 0 & -1 \\ \sqrt{2} & 0 & \sqrt{2} \\ 1 & 0 & -1 \end{bmatrix}$	$\frac{1}{2 + \sqrt{2}} \begin{bmatrix} -1 & \sqrt{2} & -1 \\ 0 & 0 & 0 \\ 1 & \sqrt{2} & 1 \end{bmatrix}$

Figura 6. Operadores más comunes para la detección de bordes [Jan05].

2.2.2.4 Segmentación basada en modelos

La mayoría de las técnicas de segmentación implementan criterios uniformes para la definición de regiones, sin embargo los resultados no siempre son buenos, lo que nos lleva a buscar diferentes formas de realizar esta tarea. Una solución a este problema consiste en la aplicación de estrategias que contemplan el conocimiento previo de las características de los objetos que se desean separar *e.g.* las propiedades geométricas como líneas y círculos. En este sentido, realizar la segmentación consiste en el análisis de modelos matemáticos que definen a los objetos como es el caso de la transformada de hough [Jan05,Bal81,DH72] para detectar líneas, círculos, elipses, etc.

2.2.3 Extracción de contornos

Un contorno es un conjunto de puntos que representan una curva en una imagen. Esta técnica es ampliamente usada en aplicaciones como el reconocimiento, análisis topológico, conteo de objetos y compresión de imágenes [HAA92]. Para la extracción de contornos se determina la conectividad de los píxeles alrededor de una vecindad. Un píxel p , es 8-conectivo si los ocho píxeles que lo rodean comparten un borde o una esquina con p , y son etiquetados en orden contrario a las manecillas del reloj con números de 0 a 7, iniciando a la derecha de p . Ahora bien un píxel es 4-adyacente, si comparte un borde común, aunque 4-adyacencia implica 8-adyacencia, 8-adyacencia no siempre implica 4-adyacencia. Se puede decir entonces que dos píxeles son 4 u 8 conectados si tienen el mismo color y son 4 u 8 adyacentes [HAA92]. Esta conectividad de píxeles dan lugar a la formación de contornos, de tal manera que un contorno de longitud n , es una secuencia de píxeles conectados $P = (p_1, p_2, \dots, p_n)$, donde p_i está conectado con p_{i+1} para $1 \leq i < n$.

2.2.4 Morfología matemática

La morfología matemática para el procesamiento de imágenes se basa en la “teoría de conjuntos” y sus métodos son usados en la clasificación, reconocimiento y análisis de

objetos. Estos métodos se han vuelto muy populares dado su utilidad y sus rigurosas descripciones matemáticas [Haw94].

Por lo general, la práctica de morfología se aplican una vez que se ha obtenido la imagen segmentada, que puede ser vista como una máscara de la imagen original y en donde los objetos de interés pueden ser detectados, reconocidos y analizados. Esto se debe a que, por muy cuidadoso que se haya realizado el proceso de segmentación de la imagen, los resultados obtenidos presentan algunas deficiencias *e.g.* píxeles aislados que forman parte del fondo de la imagen y fueron clasificados como objetos de interés.

Esto se debe entre otros factores, a que los niveles de gris de estos píxeles se encuentran dentro del criterio utilizado en la segmentación, o porque la cercanía entre los objetos de interés, es muy pequeña, o más aun, éstos no son totalmente reconocidos por la existencia de condiciones irregulares en la iluminación de la imagen que dan lugar a que un objeto sea particionado y sea considerado como un conjunto de objetos durante la detección y análisis de los objetos [Per03].

Estas y otras deficiencias de los algoritmos de segmentación, pueden ser depuradas mediante la aplicación de métodos de morfología matemática, que permiten obtener una máscara adecuada para realizar la detección, reconocimiento y análisis de los objetos de interés. Las operaciones morfológicas más comunes se describen a continuación:

2.2.4.1 Erosión y dilatación.

Se denomina erosión y dilatación a dos operaciones complementarias que se realizan sobre las imágenes, que consisten en la eliminación o incremento de píxeles alrededor de un objeto. La erosión consiste en la búsqueda de los píxeles marginales a un objeto (*e.g.* píxeles con valor lógico 1 que tengan un vecino con valor lógico 0) y es producida cuando los píxeles marginales cambian su valor de 0 a 1. La dilatación aplica un criterio similar,

buscando los píxeles marginales con valor lógico 0 cambiando su valor a 1 como se muestra en la Figura 7.

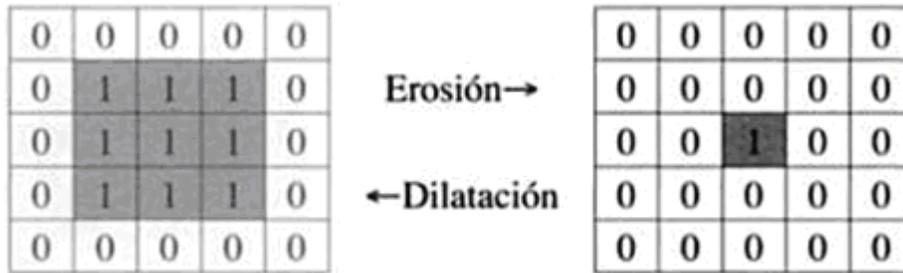


Figura 7. Representación de las operaciones de erosión y dilatación [Per03]

Como se puede apreciar en la Figura 7 anterior la erosión produce una disminución del área del objeto y la dilatación produce un incremento de la misma. La erosión y dilatación puede realizarse de acuerdo a diferentes criterios en la conectividad de los píxeles (*e.g.* erosionar o dilatar un píxel marginal que tenga un vecino con valor cero inmediatamente a la derecha pero no a la izquierda, erosionar la esquina de un objeto si se tienen tres píxeles consecutivos con valor cero, etc.). Ejemplos de estas operaciones se muestran en la Figura 8 y Figura 9.

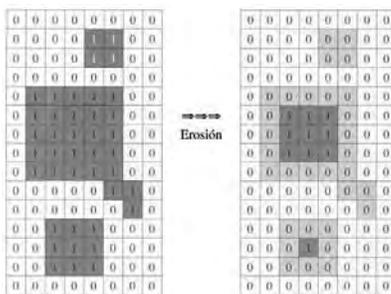


Figura 8. Representación de la operación de erosión. Izquierda: representación de imagen binaria. Derecha: resultado de erosionar la imagen binaria. [Per03]

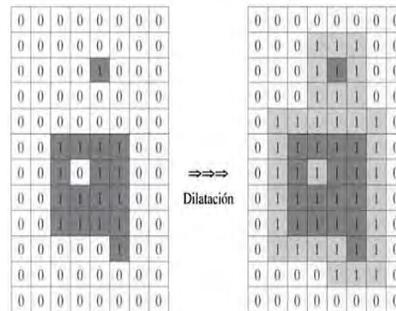


Figura 9. Representación de la operación de dilatación. Izquierda: representación de imagen binaria. Derecha: resultado de dilatar la imagen binaria. [Per03]

2.2.4.2 Apertura y cierre.

La apertura y cierre de la imagen es realizada mediante las operaciones de erosión y dilatación *e.g.* la apertura de una imagen consiste de la aplicación n erosiones seguidas de n

dilataciones. Usando la misma analogía, el cierre de la imagen se realiza dilatando n veces la imagen y erosionando n veces.

Relleno de agujeros. Al igual que las operaciones de apertura y cierre, esta actividad es realizada mediante una dilatación condicionada en donde solo se dilatan los píxeles que delimitan las regiones internas de los objetos, dando como resultado una imagen en donde los píxeles que formaban parte de un objeto con orificios son convertidos para formar parte del objeto.

Esqueletización. Puede entenderse como una erosión que concluye cuando se alcanza una línea final de píxeles de tal manera que la ésta es el eje del objeto constituida por un píxel. Un ejemplo de esta aplicación se muestra en la Figura 10.

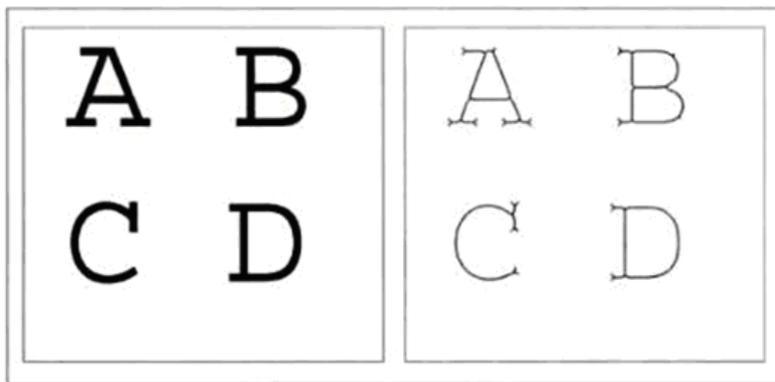


Figura 10. Representación de la operación de esqueletización. Izquierda: Imagen original; Derecha: Aplicación del operador de esqueletización. [Per03]

2.2.5 Contornos activos

Los contornos activos [BI98], también conocidos como *Snakes*, son uno de los modelos utilizados en la segmentación de imágenes que permiten ajustar de manera directa curvas suaves y cerradas asociadas a las componentes de interés dentro de la imagen. Una manera sencilla de describir un contorno puede ser la colocación de una banda elástica sometida a una atracción de un objeto donde la banda se ajusta a los bordes [KWT88]. Basándonos en esta idea, se considera solo el caso de contornos discretos definidos por una cadena de

puntos p_1, \dots, p_N , donde $p_i = (x_i, y_i)$ lo conforman como se observa en la Figura 11. Cada elemento tiene asociada una energía funcional, de tal forma que la energía del contorno sea mínima. Esta energía funcional está compuesta por tres energías que corresponden a la fuerza de atracción del contorno, la energía elástica, la energía de la curvatura, y la energía de la imagen, definida por:

$$E = \alpha E^{elastica} + \beta E^{curvatura} + \gamma E^{imagen} \quad (8)$$

Donde α, β y γ determinan la importancia de los diferentes términos y pueden variar con respecto a cada uno de los puntos.

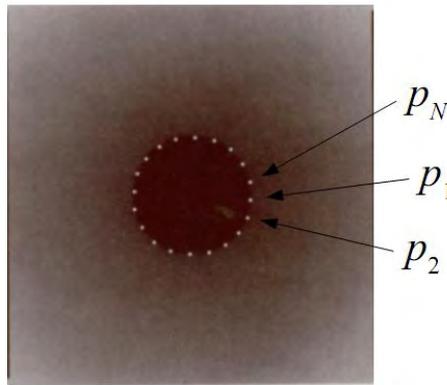


Figura 11. Contorno activo representado por un conjunto de puntos. [TV98]

Cada término de la energía funcional tiene diferente propósito, de tal manera que la energía elástica y la energía de la curvatura ayudan a mantener la continuidad y suavidad del contorno y es considerada como una forma de energía interna. Mientras que la energía de la imagen proporciona la fuerza de atracción al contorno haciendo que éste se ajuste lo más cerca al borde del objeto; la energía de la imagen es considerada como una forma de energía externa. Es necesario que los puntos que forman el contorno se encuentren regularmente espaciados manteniendo una separación uniforme. Sea d la distancia promedio entre puntos consecutivos, la energía elástica mide la desviación a esta distancia promedio mediante la ecuación 9 equivalente a conectar los puntos consecutivos con resortes de longitud d como se muestra en la Figura 12 en donde la distribución equidistante de los puntos presenta una energía menor.

$$E^{elastica} = (\|p_i - p_{i-1}\| - d)^2 \quad (9)$$

Donde p_i es el punto en el cual se calcula la energía.

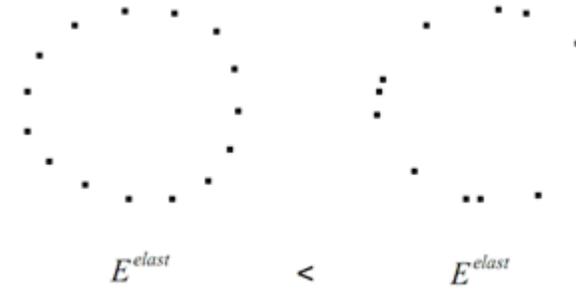


Figura 12. Energía elástica del contorno. Los puntos uniformemente espaciados tienen menor energía que una distribución variante.

Por otro lado el contorno debe tener una forma regular sin muchos huecos ni protuberancias. Para esto la energía de la curvatura penaliza los cambios de dirección repentinos haciendo una aproximación a la segunda derivada del contorno mediante la ecuación 10 que mide el grado de flexión del contorno como se muestra en la Figura 13. De acuerdo con el ejemplo de la Figura 11, la Figura 13a presenta menor energía que la Figura 13b debido a que se ajusta cercanamente al contorno:

$$E^{curvatura} = \|p_{i+1} - 2p_i + p_{i-1}\|^2 \quad (10)$$

Donde p_i es el punto a partir del cual se calcula la energía de la curvatura a partir de la segunda derivada del contorno.

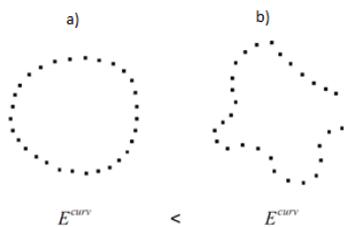


Figura 13. Energía de la curvatura. La energía de la curvatura es menor cuando se ajusta el contorno

De igual forma la energía de la imagen (energía externa) proporciona la fuerza necesaria para mantener el contorno lo más ajustado posible a la región a la cual se desea ajustar el

contorno. La energía de la imagen es la representación espacial del gradiente mediante la ecuación 11 siendo menor sobre los bordes de la región como se muestra en la Figura 14.

$$E^{imagen} = \|\nabla I\| \quad (11)$$

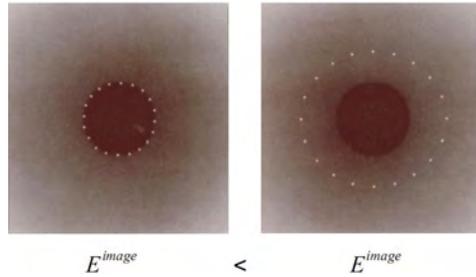


Figura 14. Energía de la imagen. Los puntos ajustados a los bordes del objeto tienen menor energía que los puntos más distanciados. [TV98]

Una vez definidos los términos de la energía que ajustara el contorno es necesario mencionar que al aplicar las ecuaciones 9 y 10 se debe considerar una vecindad (*e.g.* una vecindad de 5 x 5 píxeles) alrededor de p_i lo cual permite desplazar a p_i dentro de la vecindad con menor energía. Antes de aplicar el algoritmo es necesario normalizar términos. Para normalizar la energía elástica así como la energía de la curvatura solo se requiere dividir por el mayor valor de energía correspondiente en la vecindad considerada, mientras que la normalización aplicada a la energía de la imagen está dada por:

$$\frac{\|\nabla I\| - m}{M - m} \quad (12)$$

Donde M y m son el máximo y el mínimo en la vecindad respectivamente. Para el ajuste de un contorno activo se sigue el siguiente algoritmo.

Algoritmo: Ajuste de un contorno activo

La entrada al algoritmo es una imagen I que contiene un contorno cerrado para el cual se quiere ajustar el contorno, y un conjunto de puntos p_1, \dots, p_N que definen la posición inicial

de los puntos que forman el contorno que se desea ajustar. N Representa el número total de puntos del contorno.

- 1 Calcular la distancia promedio entre los puntos vecinos
- 2 Para cada $i = 1, \dots, N$
 - a. Encontrar la posición dentro de la vecindad p_i donde la energía E es mínima
 - b. Desplazar p_i a la posición encontrada
- 3 Para cada $i = 1, \dots, N$
- 4 Estimar la curvatura k_i del contorno en p_i con $k = \|p_{i+1} - 2p_i + p_{i-1}\|^2$
- 5 Repetir 1-4 mientras los puntos se muevan

La salida del algoritmo es el conjunto de puntos ajustados al contorno de la región de interés de la imagen.

2.2.6 Detección de rostros

Como se mencionó anteriormente, existen diferentes enfoques para llevar a cabo la detección de rostros. En este trabajo, se aplica la detección de rostros como una de las técnicas de visión por computadora para la detección de escenarios. Para la detección de rostros nos basamos en el trabajo de Viola. *et. al* [VJ04] a partir de la extracción de características y el entrenamiento de un clasificador para la detección de rostros.

2.2.6.1 Extracción de características para detección de rostros

Una característica es definida por un conjunto de píxeles claros y oscuros que representan áreas de la imagen en donde existen cambios de iluminación [VJ01]. Existen tres tipos de características utilizadas para la detección de rostros: características formadas por dos rectángulos, características formadas por tres áreas rectangulares y formadas por cuatro rectángulos como se muestra en la Figura 15.

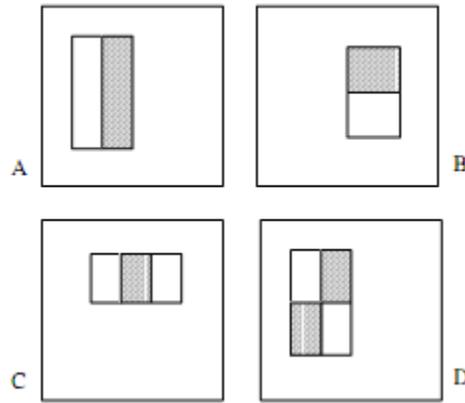


Figura 15. Representación de tipos de características. A y B representación de características formadas por dos rectángulos. C representación de características formadas por tres rectángulos. D representación de características formadas por cuatro rectángulos. [VJ01]

El valor de una característica formada por dos rectángulos es la diferencia entre la suma de los píxeles dentro de las regiones rectangulares (Figura 17), el valor de una característica formada por tres rectángulos es la diferencia de la suma de las áreas exteriores y el área central, el valor de una característica formada por cuatro regiones rectangulares es la diferencia entre la suma de las regiones diagonales. Estas características pueden ser encontradas en los rostros como se muestra en la Figura 16, donde el área de los ojos es más oscura que el área de las mejillas y es clasificada como una característica formada por dos rectángulos, de igual forma el área de la nariz es más clara que el área ocupada por cada uno de los ojos y por tanto es clasificada como una característica formada por tres rectángulos.

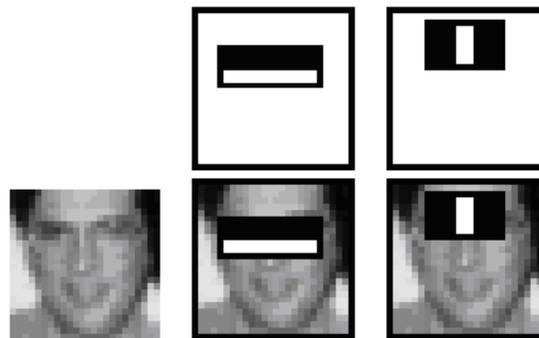


Figura 16. Clasificación de características presentes en un rostro. [VJ01]

Las características rectangulares son fácilmente calculadas usando una representación intermedia llamada imagen integral [VJ01] que permite calcular las características de una manera rápida y eficiente a diferentes escalas realizando un número reducido de operaciones por píxel. La imagen integral en la posición (x, y) está formada por la suma de los píxeles ubicados en la parte superior izquierda de (x, y) denotada por:

$$ii(x, y) = \sum_{x' \leq x, y' \leq y} i(x', y') \quad (13)$$

Donde $ii(x, y)$ es la imagen integral y $i(x, y)$ es la imagen original. A partir de la ecuación 3.15 se obtiene entonces un par de recurrencias que permiten calcular la imagen integral realizando un solo cálculo a partir de la imagen original.

$$s(x, y) = s(x, y - 1) + i(x, y) \quad (14)$$

$$ii(x, y) = ii(x - 1, y) + s(x, y) \quad (15)$$

Donde $s(x, y)$ es la suma acumulada por renglón.

Usando la imagen integral la suma de cualquier área rectangular puede ser calculada mediante un conjunto de cuatro puntos de referencias como se muestra en la Figura 17, claramente se puede observar también que la diferencia entre la suma de dos áreas rectangulares puede ser calculada a partir de ocho puntos de referencia. Estos puntos de referencia representan las coordenadas (x, y) que demarcan la región a la cual se calcula la imagen integral para la extracción de características.

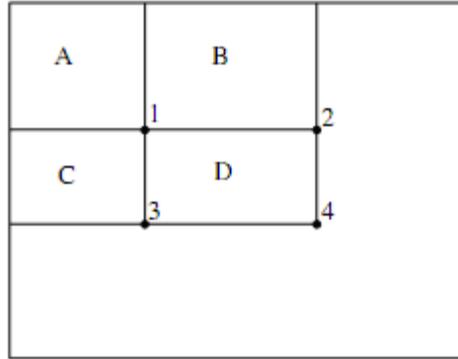


Figura 17. Cálculo de la imagen integral mediante puntos de referencia. [VJ01]

2.1.1.1 Entrenamiento de un clasificador para detección de rostros

El entrenamiento consiste en la búsqueda de los mejores parámetros para la detección de rostros. Los métodos para la detección de rostros basados en la extracción de características realizan la detección a partir de métodos de entrenamiento como: *clasificadores Bayesianos, Boosting, árboles aleatorios, el k-vecino más cercano, redes neuronales* [A+01]. Para el entrenamiento es necesario un conjunto de imágenes de objetos que se desean detectar (para este caso en particular rostros), que reciben el nombre de “positivos”, y un conjunto de imágenes (de cualquier tipo) que no contengan objetos de interés (rostros) llamado “negativos”.

Basado en el trabajo de Viola, et. al [VJ01], el entrenamiento se realiza aplicando una modificación del algoritmo AdaBoost [FS97] para la selección del conjunto de características para realizar el entrenamiento del clasificador. En su forma original, el algoritmo de entrenamiento AdaBoost es usado para impulsar el rendimiento de un algoritmo simple (*e.g.* puede ser usado para impulsar el rendimiento de un perceptron simple) combinando una colección de clasificadores débiles que dan como resultado un clasificador fuerte [TV00]. Un perceptron es una red neuronal simple que permite llevar a cabo el entrenamiento del clasificador.

El entrenamiento es llamado débil debido a que no se espera que ningún perceptron pueda clasificar de manera adecuada los datos de entrenamiento (probablemente el perceptron pueda clasificar los datos de entrenamiento en 51% [VJ01]); el entrenamiento se realiza iterativamente hasta obtener el porcentaje de error más bajo (*e.g.* 0.001 de error). Después de cada ronda de entrenamiento, el conjunto de datos de entrenamiento es re-ponderado con el propósito de destacar los datos que fueron incorrectamente clasificados en la etapa anterior por el clasificador [FA91].

Resumen del capítulo

En este capítulo se presentaron los antecedentes de visión por computadora, así como las técnicas para el procesamiento de imágenes que permiten realizar la detección de marcas artificiales y la detección de rostros. En el siguiente capítulo se analizará la arquitectura del sistema así como la descripción del proceso para la detección de marcas y rostros.

Capítulo 3

Arquitectura

En este capítulo se describen los componentes que conforman la arquitectura del sistema, así como su integración. La arquitectura es organizada en módulos; cada módulo realiza tareas específicas como: capturar imágenes de las cámaras de video, detectar marcas de color sobre las mesas de trabajo, detectar rostros, administrar la configuración del sistema y determinar el escenario actual durante una sesión dentro del aula.

1.1 Módulos del sistema

El análisis de la actividad en el aula se realiza mediante la integración de módulos. La Figura 1 muestra la organización y distribución de los módulos. El módulo principal del sistema es el módulo para la determinación de escenarios. Este módulo, está integrado a su vez por cuatro módulos detectores (correspondientes al número de cámaras conectadas como se muestra en el módulo de captura), que proporcionan información sobre el escenario que cada una de ellas está analizando. Esta información es retomada por un módulo de votación, en el que cada detector vota por el escenario que ha detectado; esto es, debido a la existencia de objetos o personas en la imagen, el escenario puede no ser identificado por algunos de los módulos detectores. En este caso, el módulo en cuestión informa al módulo de votación que no ha sido capaz de identificar el escenario; mientras que los detectores que sí lograron determinar un posible escenario, envían un identificador al módulo de votación, en donde se realiza un conteo del escenario identificado por la mayoría de los detectores y el sistema da a conocer el escenario con mayor número de votos.

Para que cada uno de los módulos detectores pueda realizar adecuadamente el proceso de detección, es necesario ajustar los parámetros (*e.g. valores que definen el color de la marca, eliminación de ruido en las imágenes, ajuste de un contorno sobre las marcas y valores que permiten detectar rostros en las imágenes, etc*) que cada módulo necesita para realizar su función. Así pues, este ajuste de parámetros es llevado a cabo por un módulo de configuración que se comunica con los módulos detectores y proporciona los parámetros que cada módulo requiere.

En la Figura 2 se muestra el diagrama de clases como una traducción de la arquitectura a elementos que representan los módulos mediante el cual se ha diseñado el sistema. Esta representación permite analizar desde el punto de vista de programación la interacción de los módulos.

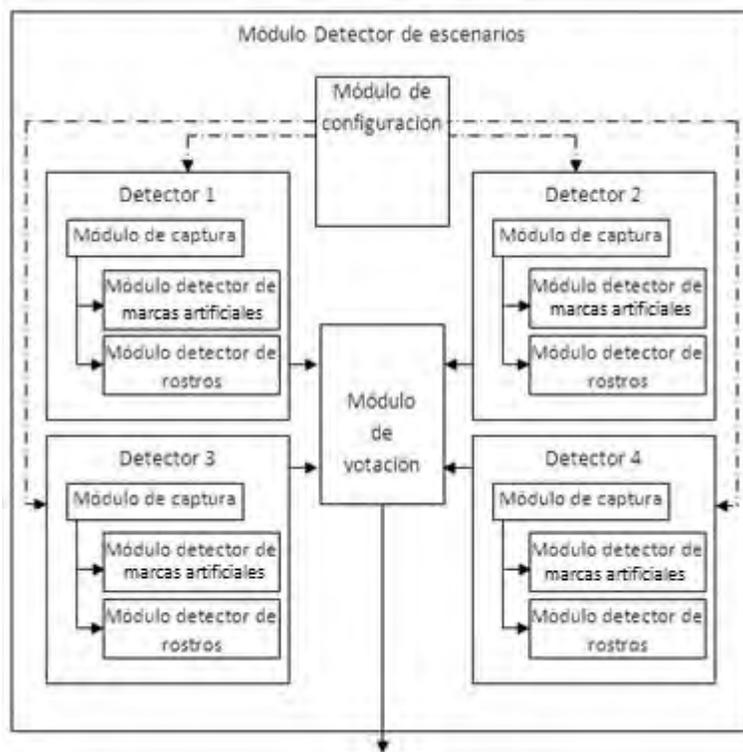


Figura 1. Representación de la arquitectura por módulos

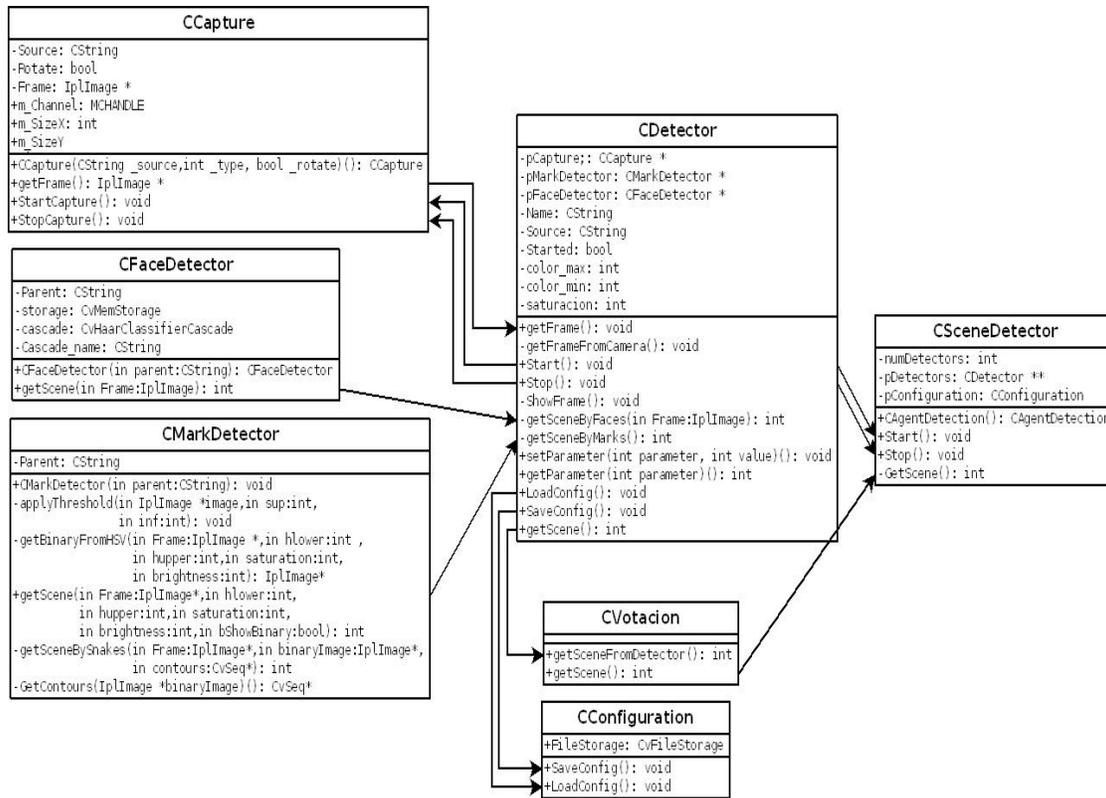


Figura 2. Traducción de los módulos de la arquitectura a diagramas de clase

Cada módulo detector está a su vez integrado por un módulo de captura cuya función es la adquisición de imágenes desde una cámara de video, un módulo para la detección de marcas que procesa las imágenes para detectar la distribución de las marcas sobre las mesas de trabajo y un módulo para la detección de rostros que ayuda a reducir la ambigüedad generada cuando las mesas de trabajo se encuentran separadas. El diagrama de flujo de la Figura 3 muestra de manera grafica las tareas que los módulos una vez integrados realizan para la determinación de los escenarios que a continuación se describen:

El proceso inicia con la captura de imágenes de las cámaras de video. En estas imágenes se detectan las marcas para conocer la distribución de las mesas; si la distribución de las mesas corresponde al escenario colaborativo la determinación del escenario es realizada por el módulo de votación, de lo contrario, si la distribución de las mesas corresponde a un escenario proyectivo (escenario pizarra ó escenario videoconferencia), se realiza la detección de rostros para que el sistema determine si se trata de una actividad con pizarra

digital ó de una actividad con proyección estéreo o videoconferencia. Una vez detectados los rostros, el escenario puede ser determinado por el módulo de votación quien informa al sistema el escenario detectado.

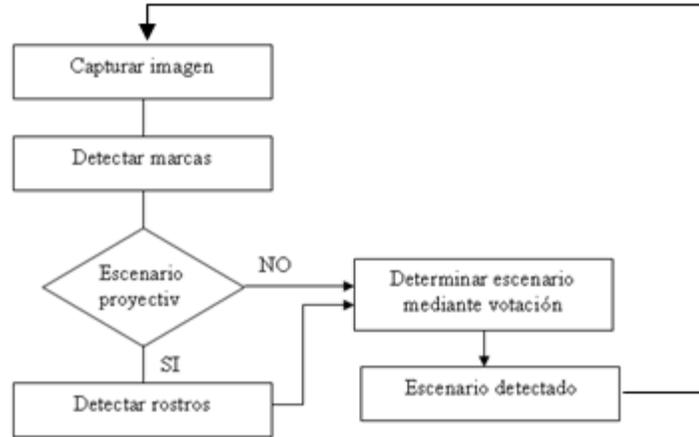


Figura 3. Diagrama de flujo para la determinación de escenarios.

1.1.1 Módulo Detector

Este módulo es el encargado de la detección de escenarios. Está integrado por un módulo para la captura de imágenes, un módulo para la detección de marcas y un módulo para la detección de rostros.

1.1.1.1 Módulo de captura

Este módulo está encargado de capturar las imágenes a través de cuatro cámaras de video, las cuales están conectadas a una tarjeta de captura con cuatro entradas. Las imágenes son capturadas en formato RGB, utilizando 24 bits para representar los colores (8 bits para cada canal). Cada una de las imágenes capturadas por este módulo es enviada a los módulos de detección de marcas y rostros para su procesamiento. La Figura 4 representa el proceso mediante el cual se envían las imágenes para su procesamiento a los módulos para la detección de marcas y detección de rostros. La captura de imágenes se realiza mediante la

librería MultiCam [Eur], que contiene funciones para controlar la captura de imágenes desde la tarjeta y está contenida en el controlador de la tarjeta de captura.

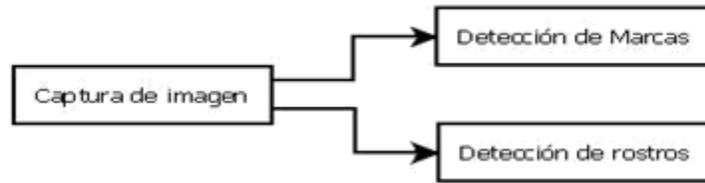


Figura 4. Representación del proceso de captura de la imagen. Cada imagen capturada es enviada a los módulos de detección de marcas y rostros.

1.1.1.1 Características de las cámaras

Se cuenta con cinco cámaras que se conectan en las entradas de la tarjeta de captura. Dado que la tarjeta de captura solo cuenta con cuatro entradas, dos de ellas son conectadas a una mezcladora de video, que permite combinar las imágenes que se obtienen de dos cámaras. Cuatro de las cinco cámaras se encuentran ubicadas en las esquinas del aula y una en el centro del techo como se muestra en la Figura 5.

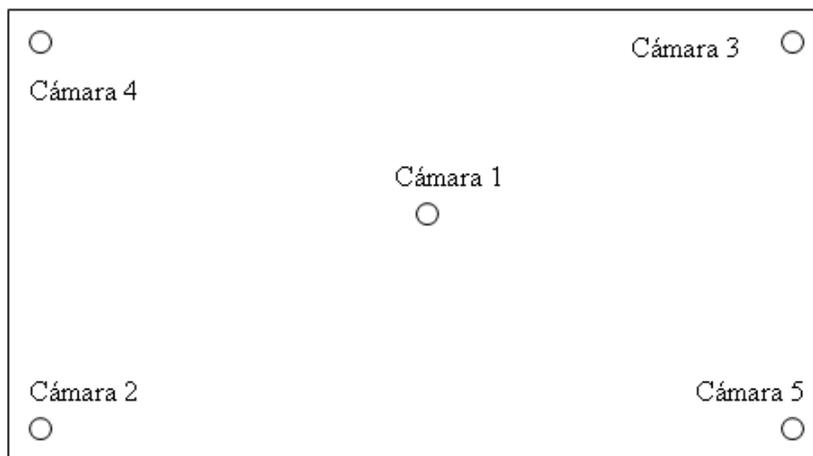


Figura 5. Distribución de las cámaras dentro del aula

Para identificar las cámaras, se han numerado del uno al cinco. A continuación se describen sus características Tabla 1:

Tabla 1. Características de las cámaras de video

Cámara	Detalles
1	<p>Marca: Sony Modelo: SKD106M290 Descripción: Mini Cámara Domo / 3X zoom digital / 470 TVL</p> <p>Características: 470 líneas de resolución. Lente de 3.8mm. Zoom Digital 3X. Control remoto por controlador RS-485.</p>
2 y 3	<p>Marca: Steren Modelo: CCTV-153 Descripción: Mini cámara a color para circuito cerrado de televisión con LEDs infrarrojos para visión nocturna.</p> <p>Características: Compatible con el formato NTSC 720 x 480 píxeles Salida de video en BNC Alimentación: 12 Vcc 500 mA Resolución horizontal de 420 líneas de TV Dimensiones: 7,5 cm de frente x 6,5 cm de alto x 6 cm de espesor</p>
4 y 5	<p>Marca: Sony Modelo: EVI-D100</p> <p>Características: Cabezal con amplia capacidad y velocidad de pan/tilt Completamente controlable a distancia mediante un cable RS-232C Zoom de 40X(zoom óptico de 10x y zoom digital de 4x)</p>

1.1.1.2 Módulo para la detección de marcas

En este módulo se realiza la detección de marcas de color sobre las mesas de trabajo. La detección de marcas proporciona información sobre la distribución de las mesas (una de las

técnicas que se está utilizando para poder detectar los posibles escenarios en una sesión de clases). El diagrama para la detección de marcas se muestra en la Figura 6.

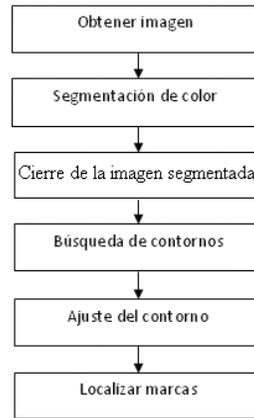


Figura 6. Diagrama de flujo para la detección de marcas y determinación de escenarios.

1.1.1.2.1 Características de las marcas

El diseño de las marcas es una pieza clave que permite al sistema detectar los escenarios dentro del Aula del Futuro. La Figura 7 muestra la distribución de las marcas sobre las mesas de trabajo. Debido al papel que desempeña la detección de marcas en la detección de escenarios, las marcas deben tener características diferentes a los objetos dentro del aula que permitan su detección:

Color: Se utilizó el color verde debido a que no forma parte del ambiente dentro del Aula del Futuro, además de ser fácil de detectar mediante la segmentación de color.

Dimensiones: La marca debe ser grande de tal forma que se pueda realizar la segmentación del color adecuadamente. Las marcas son ajustadas a las dimensiones de las mesas de trabajo de 60 x 60 cm.

Ancho: Esta característica facilita la detección de las marcas; marcas muy delgadas no son detectadas correctamente por la presencia de ruido en la imagen. El ancho de la marca utilizada es de 5 cm.

Forma: La forma de la marca permite la detección aun con la presencia de objetos que puedan obstruir en parte las marcas.



Figura 7. Representación de marcas sobre las mesas de trabajo.

1.1.1.2 Detección de marcas

El procesamiento de la imagen se realiza mediante la librería OpenCV (Open Source Computer Vision Library) [Int]de distribución en código abierto, que contiene funciones de programación enfocadas al procesamiento de imágenes y a la visión por computadora en tiempo real; ejemplos de las aplicaciones que se desarrollan con esta librería son el reconocimiento de objetos (*e.g. la detección de rostros*), sistemas de visión para robots,

segmentación de imágenes, reconocimiento de patrones, análisis de movimiento y seguimiento de objetos, reconstrucción 3D, entre otros más.

La detección de marcas se realiza en una serie de procesamientos que se aplican a la imagen. Estas técnicas fueron descritas en el Capítulo anterior y a continuación se mencionan las técnicas utilizadas para la detección de marcas:

- 1 **Obtención de imagen.** Las imágenes son capturadas por el módulo de captura como se muestra en la Figura 7. Este módulo es el encargado de enviar las imágenes al módulo de detección de marcas en donde se aplican el procesamiento necesario para detectar las marcas de color sobre las mesas de trabajo.



Figura 8. Imágenes capturadas por las cámaras de video

- 2 **Segmentación del color.** La segmentación consiste principalmente en la obtención de las regiones de la imagen que contienen el color de interés (verde de las marcas) formando una imagen binaria con dos colores: blanco y negro (Figura 8), donde el color blanco representa los píxeles que contienen el color de interés (marcas de color), y el negro todos los demás colores presentes en la imagen, pero que no proporcionan mayor información para la detección de las marcas.

Una imagen en color RGB representada por 24 bits puede contener 16777216 colores diferentes; por tal motivo realizar la segmentación en este espacio

representa un trabajo laborioso y un tanto complicado. Por lo que el espacio de color HSV facilita este proceso, pues representa el color en términos de sus componentes: tonalidad, saturación y brillo. Como se mencionó en el capítulo 3, la tonalidad representa el color como el “color puro”, la saturación indica el nivel de oscuridad o claridad que tiene el color y el brillo representa que tan brillante es un color. De aquí entonces el proceso de segmentación se lleva a cabo a partir de la conversión de la imagen RGB al espacio de color HSV separando la imagen en sus componentes H, S y V.

El proceso de segmentación del color se logra estableciendo un límite inferior L_{inf} y superior L_{sup} dentro del cual se puede encontrar el color de la marca, y un valor de saturación s así como un valor de brillo v que nos permiten precisar el color de interés. Estos valores son establecidos a partir de los resultados visuales y son ajustados hasta que se logra segmentar las marcas de acuerdo a la apreciación visual. Una vez establecidos estos valores que nos definirán el color de las marcas, se utiliza una variante de la ecuación 3.1 que tome en cuenta el límite inferior y superior así como el valor de saturación y brillo definida por:

$$I_{seg} = \begin{cases} 1 & \text{si } H(x, y) \geq L_{inf} \wedge H(x, y) \leq L_{sup} \wedge S(x, y) \leq s \wedge V(x, y) \leq v \\ 0 & \text{Cualquier otro caso} \end{cases} \quad (4.1)$$

Donde I_{seg} es la imagen segmentada, H, S, V representan la imagen HSV distribuida en sus diferentes componentes.

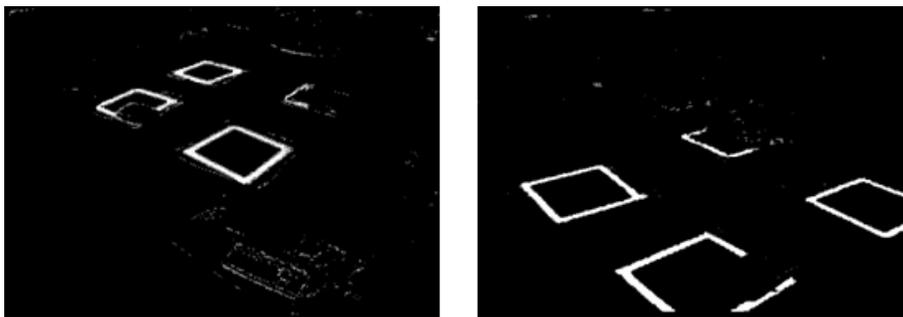


Figura 9. Segmentación de imágenes obtenidas por las cámaras de video.

- Cierre de la imagen segmentada.** La operación de cierre de la imagen segmentada consiste aplicar el proceso de dilatación y erosión a la imagen segmentada (Figura 10). Esto debido a que durante la segmentación, la presencia de objetos que puede obstruir las marcas generando regiones desconectadas que corresponden a una misma marca o que no correspondan a ninguna. Este proceso permite unir las regiones desconectadas y desaparece regiones que no pertenecen a las marcas, lo que permite realizar la detección de una forma eficiente.



Figura 10. Cierre de las imágenes a partir de la segmentación del color.

- Búsqueda de contornos.** La búsqueda de contornos consiste en la localización de píxeles blancos conectados dentro de la imagen. Esta operación se realiza mediante el análisis de conectividad entre los píxeles.
- Ajuste del contorno.** Una vez que se ha realizado la búsqueda de los contornos, se ajusta un contorno deformable (contorno activo ó *snake*) a partir del cual se puede conocer la distribución y ubicación de las mesas como se muestra en la Figura 11.



Figura 11. Ajuste del contorno a las marcas

- 6 **Localización de marcas.** Los contornos ajustados proporcionan la información correspondiente a la distribución, es decir, dependiendo del número de contornos ajustados se determina si las mesas están juntas o separadas.

La determinación del escenario se basa en la distribución de las mesas de tal forma que si las mesas se encuentran juntas se tiene la correspondencia con el escenario colaborativo, mientras que la distribución separada de las mesas corresponde con los escenarios pizarra y videoconferencia. De acuerdo con los escenarios propuestos la separación de las mesas genera ambigüedad para la determinación de escenarios debido a que no se puede determinar con exactitud si los alumnos están observando lo que el profesor explica en la pizarra interactiva o si están trabajando con una proyección estereoscópica. Para solucionar el problema de ambigüedad, se realiza la detección de rostros, que permite identificar el lugar hacia el cual los alumnos están trabajando.

1.1.1.3 Módulo para la detección de rostros

En este módulo se realiza la detección de rostros presentes en las imágenes obtenidas; el proceso de detección consiste en determinar la presencia ubicación y tamaño de estos. Este proceso se basa en la implementación del algoritmo propuesto por Viola et. al. [VJ04], en donde se detectan rostros con características frontales basado en un clasificador entrenado

que contiene las descripciones faciales (entrenado para detectar rostros con orientación frontal) para la detección de estos en las imágenes obtenidas de cámaras de video.

De acuerdo con la distribución de las cámaras, la detección de rostros debe realizarse tomando en cuenta que éstos tendrán diferentes orientaciones con respecto a las cámaras. Por tal motivo, el algoritmo implementado requiere de adaptaciones que permitan detectarlos en diferentes orientaciones (*e.g.* frontal, perfil, inclinado). Estas adaptaciones son realizadas en base a dos enfoques:

- a) **Búsqueda de parámetros.** Ajustar parámetros como el factor de escala a partir del cual la ventana de búsqueda es escalada entre las características almacenadas en el clasificador y la las características encontradas en la imagen, así como el tamaño mínimo de ventana de correspondencia con el clasificador entrenado y el número mínimo de rectángulos (características) vecinos que forman un rostro.

- b) **Entrenamiento de un clasificador.** Entrenar un clasificador para llevar a cabo la detección de rostros a partir variaciones de iluminación, tamaño y orientación. Esta actividad se lleva se realizó con la librería OpenCV [Int], pues cuenta con un conjunto de algoritmos que permiten entrenar un clasificador a partir de un conjunto de imágenes mediante la extracción de características (como se indicó en el capítulo 2).

El entrenamiento del clasificador analiza las características extraídas a partir de una colección imágenes de rostros con orientaciones y condiciones de iluminación diferentes (llamados positivos) que conforman los elementos a detectar y una colección de imágenes de cualquier tipo (llamados negativos), que junto con los positivos permiten obtener las características faciales para el clasificador. Los negativos no deben contener elementos para los cuales se realiza el entrenamiento; para este caso particular, dentro del conjunto de negativos no deben existir imágenes que contengan rostros. Así, el entrenamiento se llevó a cabo, a partir de la recopilación de 6000 imágenes de rostros y 14800 imágenes que no

contienen rostros. Los rostros contenidos en estas imágenes presentan diferentes orientaciones (frontal, perfil), así como cambios de iluminación en las imágenes. Para realizar el entrenamiento, se delimitan las regiones de los rostros (a partir de esta delimitación, el algoritmo de entrenamiento extraen las características).

El buen desempeño del clasificador depende de la selección minuciosa de imágenes que conforman la colección de positivos y negativos, por lo que una selección deficiente de imágenes no asegura que el clasificador pueda realizar una buena detección.

1.1.2 Módulo de votación

Este módulo es el encargado de informar al sistema que escenario ha sido detectado por la mayoría de módulos detectores, quienes envían a este módulo un identificador que corresponde al escenario detectado. Cada identificador recibido es sometido a votación, esto es, se realiza un conteo de los identificadores y se decide sobre el identificador con mayor número de votos. El proceso de votación es llevado a cabo de la siguiente manera: cada módulo detector envía el identificador del escenario que ha detectado. Con base a los escenarios propuestos, se verifica a que escenario corresponde cada identificador recibido, generando una suma de votos para cada escenario. Por último, se analiza el escenario que obtuvo mayor número de votos. El escenario con mayor número de votos es el escenario que da a conocer el sistema a través de este módulo.

1.1.3 Módulo de configuración

Este módulo es el encargado de realizar las operaciones de lectura, escritura y actualización de los parámetros necesarios para la detección de marcas y rostros. Estas operaciones son controladas por el módulo detector que sirve de intermediario entre el módulo de

configuración y los módulos en cuestión. Los parámetros son almacenados en un diagrama de bloques dentro de un archivo con formato XML Figura 12.

```
- <DETECTOR_1>
- <IMAGEN>
  <ROTAR>0</ROTAR>
</IMAGEN>
- <DETECCION_MARCAS>
- <SEGMENTACION>
  <COLOR_MAX>75</COLOR_MAX>
  <COLOR_MIN>31</COLOR_MIN>
  <SATURACION>0</SATURACION>
  <BRILLO>0</BRILLO>
  <MOSTRAR_IMAGEN_BINARIA>0</MOSTRAR_IMAGEN_BINARIA>
  <ITERACIONES>4</ITERACIONES>
  <EROSION>3</EROSION>
  <DILATACION>7</DILATACION>
</SEGMENTACION>
- <CONTORNO_ACTIVOS>
  <CONTORNO_ALFA>0.5000000000000000</CONTORNO_ALFA>
  <CONTORNO_BETA>0.200000029802322</CONTORNO_BETA>
  <CONTORNO_GAMA>0.100000014901161</CONTORNO_GAMA>
  <CONTORNO_ITERACIONES>100</CONTORNO_ITERACIONES>
  <CONTORNO_VENTANA>7</CONTORNO_VENTANA>
  <MOSTRAR_CONTORNOS_ACTIVOS>0</MOSTRAR_CONTORNOS_ACTIVOS>
</CONTORNO_ACTIVOS>
</DETECCION_MARCAS>
- <DETECCION_ROSTROS>
  <DETECTAR_ROSTROS>0</DETECTAR_ROSTROS>
  <MOSTRAR_ROSTROS>0</MOSTRAR_ROSTROS>
  <ESCALA>1.299999523162842</ESCALA>
  <ESCALA_DETECCION>1.200000476837158</ESCALA_DETECCION>
  <NOMBRE_CLASIFICADOR>"/.Clasificador/haarcascade_profileface1_0.xml"</NOMBRE_CLASIFICADOR>
  <NUMERO_VECINOS>3</NUMERO_VECINOS>
- <TAMANO_VENTANA>
  <WIDTH>20</WIDTH>
  <HEIGHT>20</HEIGHT>
</TAMANO_VENTANA>
</DETECCION_ROSTROS>
</DETECTOR_1>
```

Figura 12. Formato de almacenamiento de parámetros.

Resumen del capítulo

En este capítulo se ha presentado la arquitectura del sistema, describiendo los módulos que la conforman, así como su integración, que permiten realizar la detección de escenarios. Se describió la función que cada módulo realiza, así como el proceso en que se lleva a cabo la detección de marcas artificiales y rostros. En el siguiente capítulo se analizará el funcionamiento del sistema a partir de casos de uso.

Capítulo 4

Casos de estudio

En este capítulo se muestran los casos de uso, mediante los cuales se determina la funcionalidad del sistema en la detección de escenarios. Así mismo se describen los resultados obtenidos de la detección de marcas artificiales y rostros.

4.1 Detección de marcas

Se realizaron pruebas para analizar la detección de marcas. Estas pruebas se aplicaron con tres tipos de marcas (Figura 1); estas marcas se fueron modificando, pues no eran detectadas adecuadamente debido a factores como:

Ubicación: las marcas pueden ser obstruidas por objetos o por alumnos.

Color: el color de las marcas puede ser mezclado con algunos colores del ambiente.

Tamaño: el tamaño de las marcas no permitía realizar la detección del color adecuadamente debido a la distancia de las marcas con respecto a las cámaras.

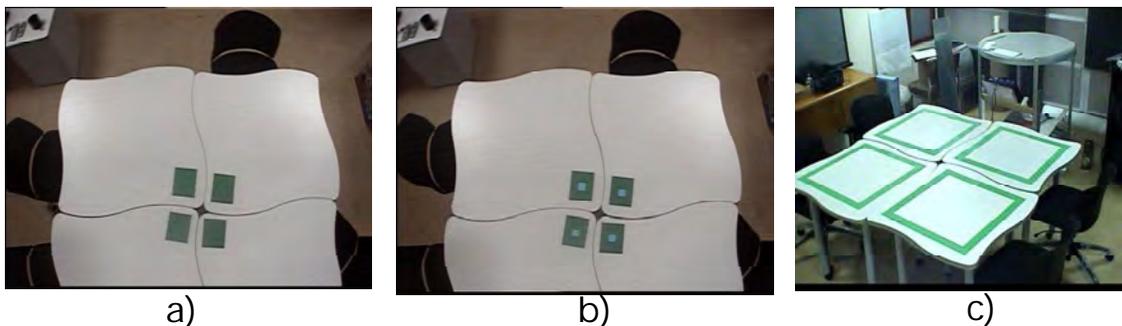


Figura 1. Marcas utilizadas para la detección de escenarios.

Tomando en cuenta estos factores, se diseñó una marca que permitiera su detección ante la presencia de obstrucciones (la Figura 1c muestra el diseño de la marca). Este diseño de marcas permite su detección ante la presencia de elementos que obstruyan las marcas o debido a condiciones de iluminación que no permitan realizar una detección adecuada, pues la ubicación de las cámaras es un factor que favorece la detección, es decir, la marca puede ser obstruida con respecto al campo de visión de una cámara pero no para alguna(s) del resto de estas. En la Figura 2 se muestra el campo de visión de las cámaras.

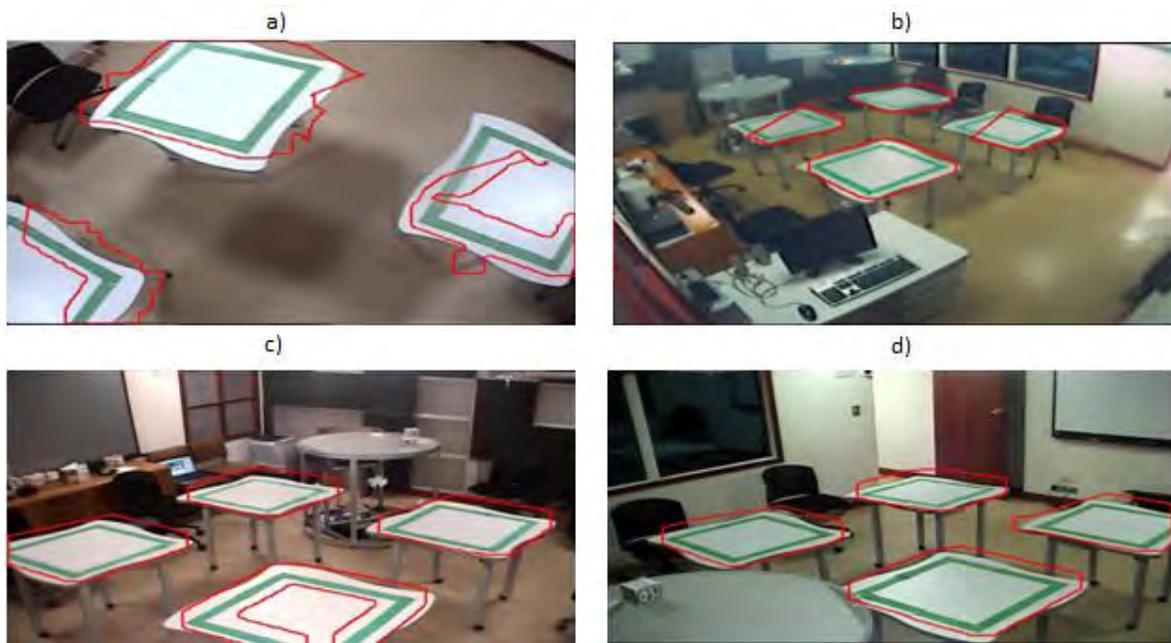


Figura 2. Detección de marcas realizada por cada cámara. a) Visualización incompleta de las marcas. b) Visualización completa de las marcas. c) y d) Visualización de marcas parcialmente.

Como se mencionó anteriormente, se probaron tres tipos de marcas dentro del ambiente del Aula del Futuro, lo cual permitió determinar que marca permite llevar a cabo la detección de escenarios. De acuerdo con la Figura 1, las marcas se enumeran de izquierda a derecha. El tipo de marca se determinó de acuerdo a los resultados que se muestran en la Tabla 1. Estos resultados se obtuvieron a partir del análisis de segmentos de video en donde se colocaron los diferentes tipos de marcas. El porcentaje de detección se determinó analizando cuadro por cuadro si la marca es detectada.

Tabla 1. Resultados de la detección de las marcas artificiales.

Marca	Tiempo de prueba	Fps	% detección correcta de la marca
a)	60 seg	28 fps	15%
b)	60 seg	28 fps	5%
c)	60 seg	28 fps	90%

Como se muestra en la tabla anterior, la marca de la Figura 1c presenta mayor porcentaje durante la detección. Debido a estos resultados se utilizó la marca de la Figura 1c para la detección de escenarios.

4.2 Detección de rostros

Estas pruebas se realizaron a partir del algoritmo para la detección de rostros implementado en la librería [Int] que implementa un clasificador entrenado. Las pruebas para la detección de rostros se realizaron utilizando tres clasificadores entrenados bajo diferentes condiciones; el primer clasificador utilizado fue el clasificador contenido en la librería OpenCV; para el segundo clasificador se realizó el entrenamiento a partir de la recolección de 6000 imágenes de rostros (positivos) y 14800 imágenes que no contienen rostros (negativos); el tercer clasificador fue descargado de páginas de visión por computadora [Seo].

El ajuste de parámetros como: a) la escala a la cual se pueden encontrar rostros en la imagen, respecto al tamaño de ventana utilizado para la extracción de características durante el entrenamiento del clasificador, y b) el tamaño mínimo de rostros en la imagen, permitieron realizar la detección de rostros. La Figura 3 muestra la detección de rostros con respecto a lo que cada cámara está visualizando al momento de la detección.



Figura 3. Detección de rostros de frente a las cámaras.

En la Tabla 2 se muestran los resultados de la detección de rostros utilizando tres clasificadores para la detección. Estos resultados se obtuvieron a partir del análisis de segmentos de video con alumnos. El porcentaje de detección se determinó analizando cuadro por cuadro la detección de rostros.

Tabla 2. Resultados de la detección de rostros

Clasificador	Tiempo de prueba	Fps	% detección correcta de rostros
1	60 seg	28 fps	55%
2	60 seg	28 fps	15%
3	60 seg	28 fps	92%

Como se muestra en la tabla, el clasificador que mejor detecta rostros es el clasificador 3, por lo que la detección de rostros se realiza utilizando el clasificador con mayor porcentaje.

4.3 Detección de escenarios

Para evaluar el desempeño de la aplicación, se realizaron pruebas que permitieron observar si los escenarios propuestos eran detectados. Las pruebas se realizaron en tres etapas: a) la primera consistió en observar si los escenarios podían ser detectados sin la presencia de elementos (objetos y alumnos) que obstruyeran el campo visible de las cámaras; b) la segunda etapa consistió en la detección de escenarios colocando objetos sobre las mesas de

trabajo para analizar si los escenarios podían ser detectados ante la presencia de oclusiones generadas por la presencia de objetos; c) la tercera etapa consistió en la colocación de objetos y la presencia de alumnos para determinar si los escenarios podían ser detectados con los elementos (objetos y alumnos) que se encuentran en una sesión de evaluación.

4.3.1 Primera etapa: Detección de escenarios sin obstrucciones

En esta etapa se aplicaron pruebas al sistema, con el propósito de analizar la detección de los escenarios en un ambiente donde solamente la presencia de marcas, fuera el único factor en la detección de escenarios.

4.3.1.1 Caso 1: Detección del escenario colaborativo sin obstrucciones

Las pruebas aplicadas para la detección del escenario colaborativo sin obstrucciones (Figura 4) reflejan que la detección de marcas permite la identificación este escenario por lo cual es posible ajustar un contorno que determine la ubicación de las mesas de trabajo. El ajuste del contorno depende del campo de visión de las cámaras y de la calidad en el proceso de segmentación del color. En la Figura 4b se observa que la presencia de elementos idénticos al color de las marcas, pueden ser detectados como marcas. En este sentido, se observa que la detección del escenarios, se puede realizar a partir de la detección que cada cámara realiza y es determinada por el módulo de votación (descrito en el capítulo anterior).

En estas pruebas se observa que el ajuste del contorno corresponde con la ubicación de las marcas que permite proporcionar información sobre la distribución de las mesas, aun cuando el contorno no sea ajustado adecuadamente por algunas cámaras.

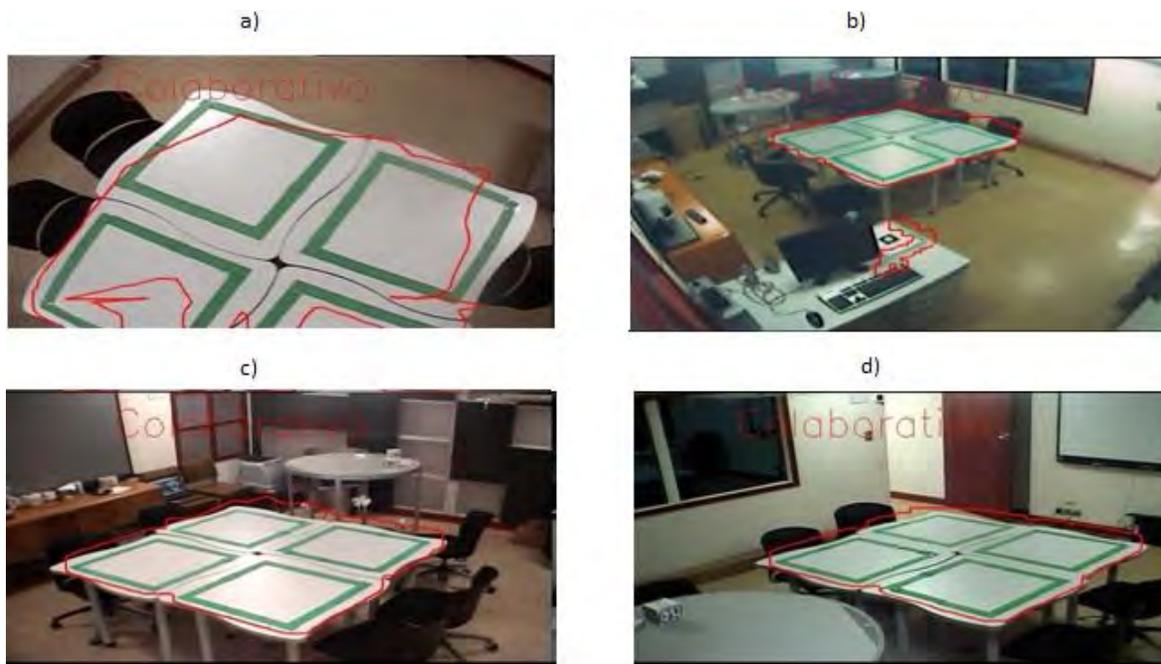


Figura 4. Pruebas para la detección del “escenario colaborativo” sin obstrucciones

4.3.1.2 Caso 2: Detección de escenarios proyectivos sin obstrucciones

Las pruebas para la detección de escenarios proyectivos (escenario de proyección en pizarra y escenario de proyección estereoscópica) sin obstrucciones permiten identificar la distribución de las mesas de trabajo pero no proporcionan información sobre qué escenario se encuentra presente; por lo que la detección de este escenario es ambigua (Figura 5). En este sentido se observa necesaria la presencia de alumnos dentro del aula. La detección del escenario se considera ambigua debido a que el sistema no puede relacionar el escenario con los escenarios planteados. La ambigüedad es solucionada con la presencia de alumnos a través de la detección de rostros.

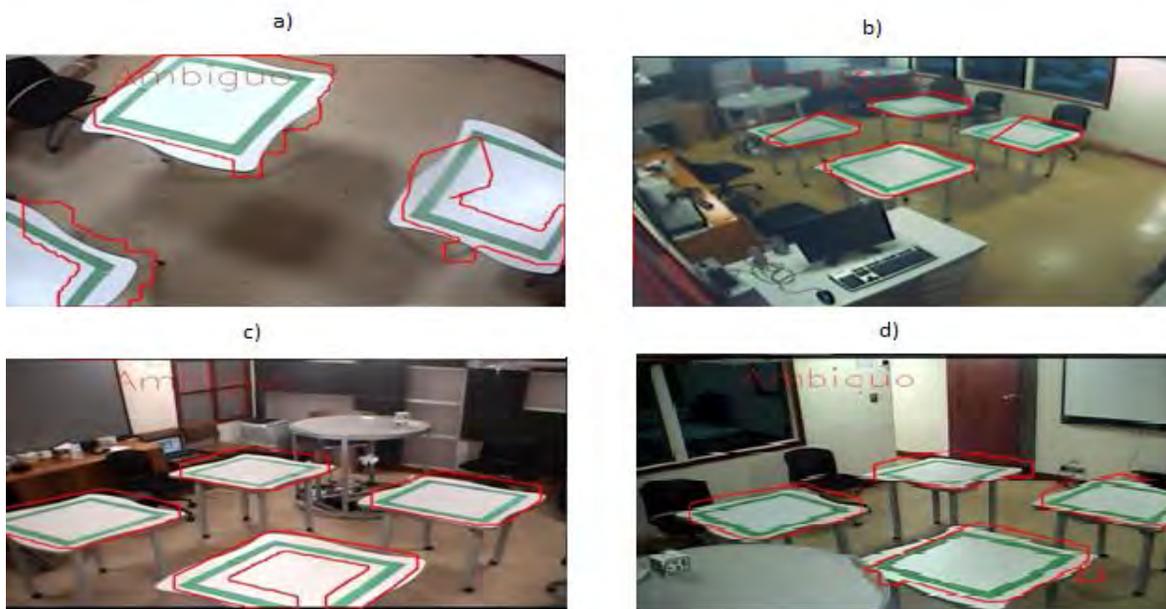


Figura 5. Pruebas para detección de “escenarios proyectivos” sin obstrucciones

4.3.2 Segunda etapa: Detección de escenarios con obstrucciones por objetos

Durante las pruebas realizadas en la etapa anterior se observó que el escenario colaborativo fue detectado sin elementos que obstruyeran las marcas, también se observó que la distribución de las mesas de trabajo fue detectada en las pruebas para la detección de escenarios proyectivos, pero el escenario no pudo ser detectado correctamente y es mostrado como ambiguo.

En esta etapa, el propósito de las pruebas fue determinar si los resultados obtenidos en la etapa anterior mostraban cambios ante la presencia de elementos que obstruyeran las marcas.

4.3.2.1 Caso 3: Detección del escenario colaborativo con obstrucciones

Las pruebas aplicadas para la detección del escenario colaborativo ante la presencia de elementos que obstruyen parcialmente las marcas, permiten determinar la distribución de

las mesas (Figura 6). Debido a que las marcas son obstruidas parcialmente por los objetos, es posible conocer la distribución de las mesas de trabajo, pues las marcas no son obstruidas completamente.

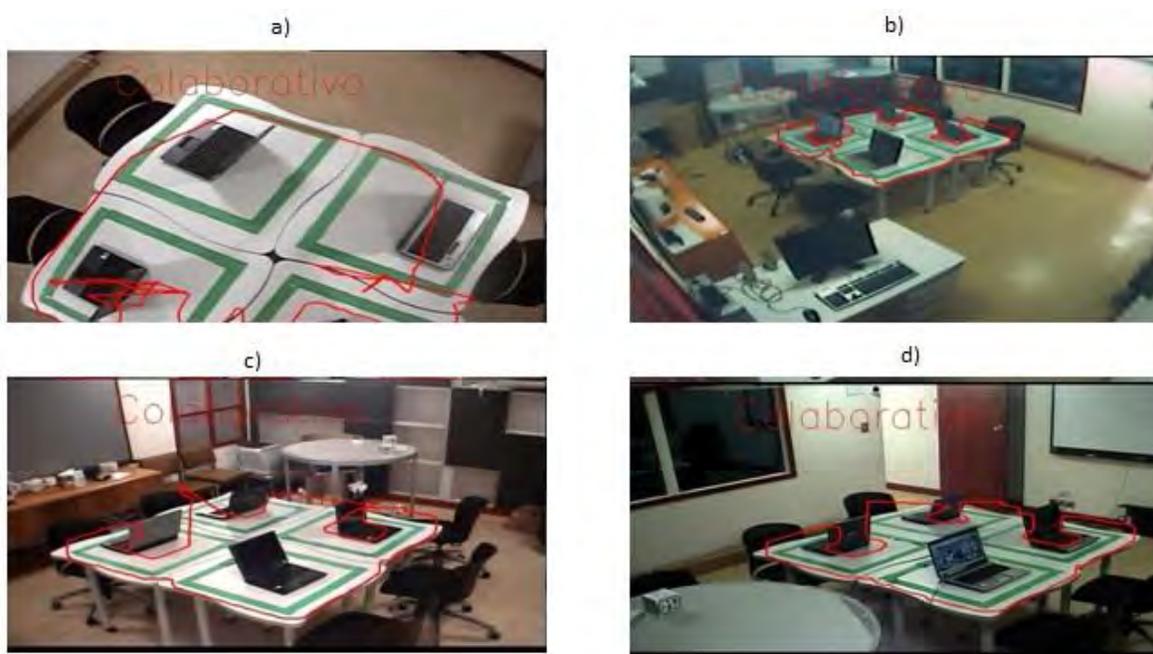


Figura 6. Pruebas para la detección del “escenario colaborativo” ante la presencia de obstrucciones

4.3.2.2 Caso 4: Detección de escenarios proyectivos con obstrucciones

En las pruebas para la detección de escenarios proyectivos sin obstrucciones, se puede identificar la distribución de las mesas, pero no es posible definir a que escenario corresponde, esto se debe a la ambigüedad generada, por lo que durante la aplicación de pruebas para la detección de escenarios proyectivos con obstrucciones, no se espera conocer si el escenario se detecta o no, por el contrario, los resultados esperados deben permitir conocer la distribución de las mesas de trabajo ante la presencia de elementos que obstruyan parcialmente las marcas (Figura 7).

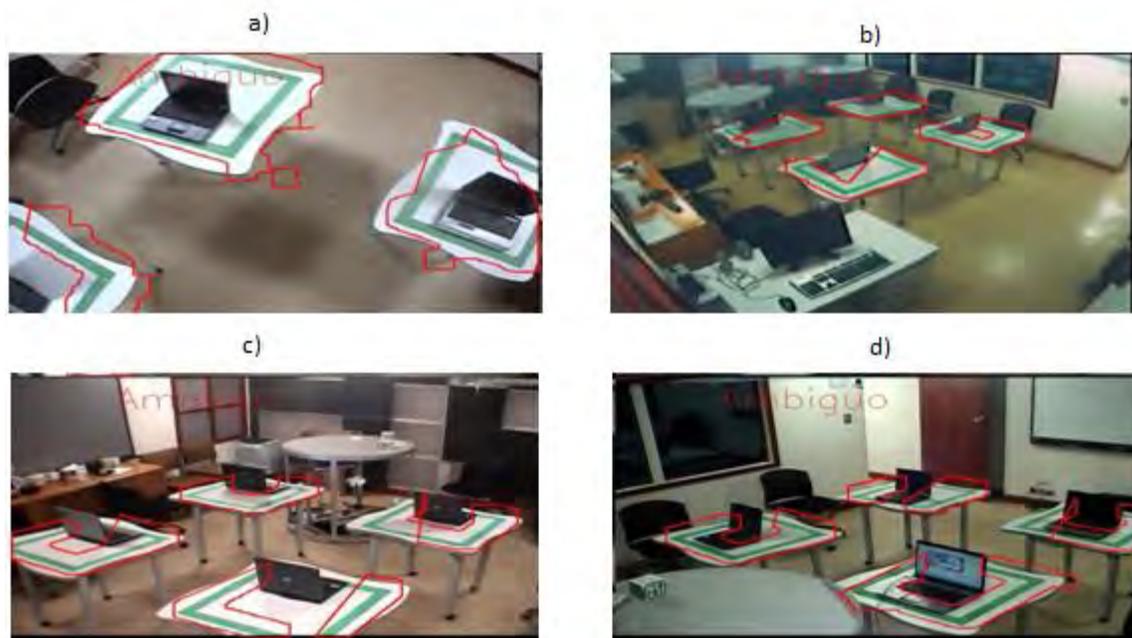


Figura 7. Pruebas para la detección de “escenarios proyectivos” ante la presencia de obstrucciones

4.3.3 Tercera etapa: Detección de escenarios con obstrucciones por objetos y por alumnos

Las pruebas realizadas en las etapas anteriores indican que es posible detectar el escenario colaborativo ante la presencia de elementos sobre las mesas de trabajo, pero existe ambigüedad en la detección de escenarios proyectivos. Por lo que en esta etapa se aplicaron pruebas con alumnos.

4.3.3.1 Caso 5: Detección del escenario colaborativo con obstrucciones y alumnos

Las pruebas aplicadas en esta etapa, indican que el escenario colaborativo puede ser detectado, aun con la presencia de elementos sobre las mesas de trabajo y alumnos (Figura 8).

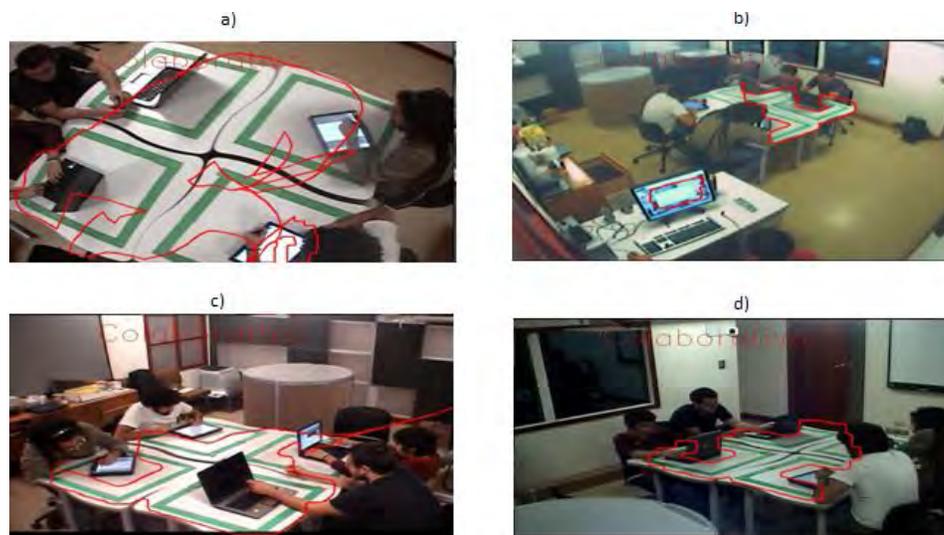


Figura 8. Pruebas para la detección del “escenario colaborativo” ante la presencia de alumnos y objetos sobre las mesas de trabajo

4.3.3.2 Caso 6: Detección del escenario pizarra con obstrucciones y alumnos

Las pruebas de las etapas anteriores no detectaron este escenario correctamente, debido a la ambigüedad causada por la falta de alumnos dentro del aula. La presencia de alumnos es necesaria para la detección de este escenario, ya que se requiere determinar hacia donde están observando (*i.e.* hacia donde se encuentra el centro de atención) los alumnos. El escenario fue detectado, pues se logró identificar que el centro de atención de los alumnos corresponde a este escenario (Figura 9).



Figura 9. Pruebas de detección del “escenario pizarra” con alumnos y objetos sobre las mesas de trabajo

4.3.3.3 *Caso 7: Detección del escenario de proyección estereoscópica y video conferencia con obstrucciones y alumnos*

Durante las pruebas realizadas para la detección de este escenario, se observa que la ambigüedad sobre escenarios proyectivos desaparece con la presencia de alumnos, los cuales permiten detectar la ubicación y orientación (

Figura 10). En este sentido, el sistema determina que se ha detectado el escenario de proyección estereoscópica y video conferencia, debido a que el centro de atención de los alumnos es asociado con este escenario.

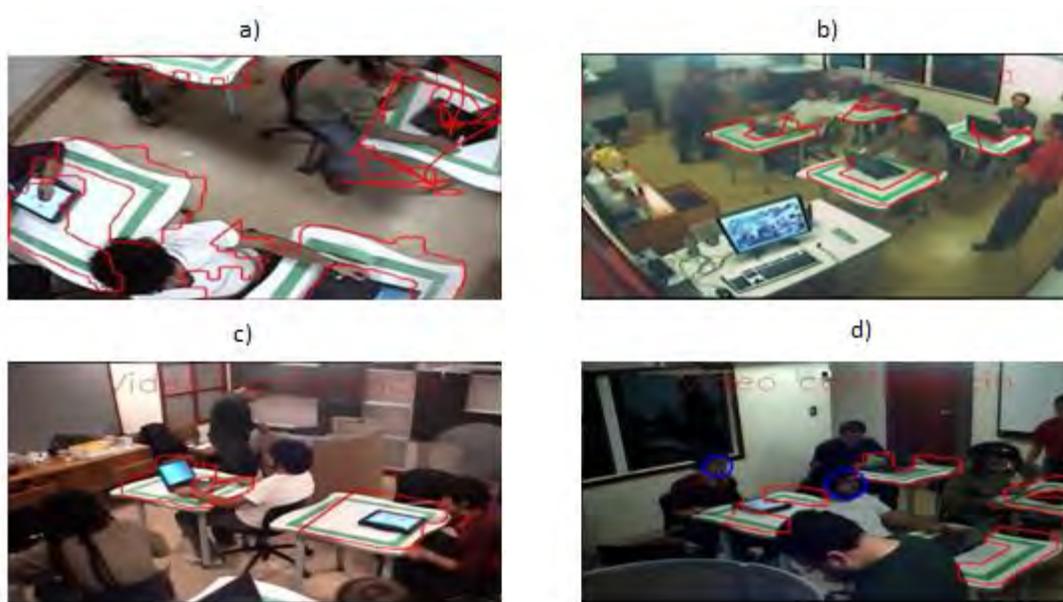


Figura 10. Pruebas de detección del “escenario video conferencia” con alumnos y objetos sobre las mesas de trabajo

4.3.4 **Análisis de resultados**

Los criterios de evaluación durante las pruebas aplicadas al sistema permitieron determinar el porcentaje de fiabilidad que el sistema puede alcanzar en la detección de escenarios durante una sesión de evaluación. Tomando como base la definición de tres tipos de escenarios, las pruebas se realizaron con tiempos de evaluación similares (aproximadamente de 15 minutos) para cada escenario, obteniendo el porcentaje de detección con respecto al porcentaje de marcas detectadas por cada cámara y generalizando el porcentaje de detección de las cuatro cámaras (esto debido a que durante el proceso de detección, el campo de visión de las cámaras varía de acuerdo a la ubicación de objetos y

los alumnos dentro del aula). Asimismo se observa el número de fotogramas por segundo que pueden ser procesados para detectar los escenarios y el tiempo promedio que dura el sistema en dar respuesta ante la presencia de cambios en los escenarios presentes.

En este sentido, los criterios que determinan la funcionalidad del sistema se muestran en la Tabla 3, en donde se muestra los valores obtenidos durante las pruebas aplicadas.

Tabla 3. Criterios de evaluación durante las pruebas aplicadas al sistema

Tipo de escenario	Tiempo de prueba	% de detección del escenario	% de error	% detección de marcas	% detección de rostros	Fps	Tiempo promedio de detección del escenario
Colaborativo	15:00 min	98%	2%	90%	25%	1.45	2.3 seg
Pizarra	14:20 min	90 %	10%	80%	70%	1.3	3.2 seg
Video conferencia	15:40 min	88%	12%	80%	70%	1.3	3.5 seg

Resumen del capítulo

En este capítulo se presentaron los casos de uso mediante los cuales se determinó el desempeño del sistema en la detección de escenario. Así mismo, se presentaron resultados parciales correspondientes a la detección de marcas y rostros, y los resultados obtenidos en la detección de escenarios. En el siguiente capítulo se presentan las conclusiones y trabajo futuro.

Capítulo 5

Conclusiones

En este trabajo hemos presentado la propuesta de un sistema que permite, con base a técnicas de visualización, identificar automáticamente uno de tres escenarios definidos previamente en un aula-laboratorio (el Aula del Futuro); y esto, con un nivel de robustez y fiabilidad determinado en 91%. Este resultado está dentro de los rangos útiles al proyecto, por lo que consideramos que nuestra propuesta representa un uso novedoso y eficiente de dichas técnicas para resolver un problema que, hasta ahora, requería de la intervención humana.

Entre las virtudes de nuestra solución podemos además señalar el que su funcionamiento no precisa de efectuar adaptaciones al aula o su equipamiento. En efecto, la detección de escenarios es realizada a través del análisis de imágenes provenientes de las cámaras de video con que ya cuenta el aula, permitiendo detectar la distribución de las mesas de trabajo a partir de la colocación de marcas sobre éstas. Como ya se explicó, esta detección de marcas permite determinar el tipo de distribución de las mesas de trabajo, uno de los elementos clave para la identificación del escenario.

Otro punto fuerte es que, a la información sobre la distribución de las mesas, el sistema es capaz de añadir, o combinar, la información proveniente de la detección de rostros, independientemente del ángulo con que éstos son percibidos por la cámara. Al combinar la información sobre la distribución de las mesas de trabajo y la detección de rostros el escenario es detectado en un tiempo promedio de 3 segundos, tiempo razonable dentro del contexto de una clase.

Estos dos aspectos nos implicaron revisar, implementar, y ajustar los algoritmos de segmentación, morfología y contornos activos que permitieran lograr la detección de marcas, así como los algoritmos para la detección de rostros, con el objetivo de afinar su desempeño en el contexto particular del problema. Asimismo nos permitió comprender mejor las ventajas de tratar con imágenes en el espacio de color HSV, con respecto de imágenes en un espacio de color complejo como es el espacio RGB, así como sus limitaciones.

Por otra parte, si bien consideramos que los requerimientos iniciales fueron satisfechos, no podemos negar que se trata, indudablemente, de un primer paso en el camino a la detección automática de la actividad dentro de un aula, y que todavía deja muchos aspectos por resolver o mejorar: escenarios menos rígidos (por ejemplo, con sólo algunas de las mesas juntas, y otras separadas, o con alumnos de pie, etc.); mejor ponderación de las tomas que provienen de las cuatro cámaras, seguimiento de un mayor número de mesas y alumnos, etcétera.

5.1 Trabajo Futuro

Como se expuso anteriormente, realizar el análisis de la actividad de los estudiantes dentro del Aula del Futuro, es en principio, el punto de partida para realizar actividades más complejas como: activar, desactivar y calibrar los diferentes dispositivos tecnológicos que se integren al proceso educativo, y probablemente, realizar el seguimiento automatizado de las acciones que los alumnos y los profesores realizan, si las nuevas problemáticas así lo requieren. En ese sentido, determinar la lista de los trabajos futuros no es un problema, por ejemplo determinar los dispositivos necesarios en una sesión de evaluación con base al escenario identificado por el sistema. Consideramos sin embargo mucho más complicado ordenarlos y sistematizarlos, de modo a tener avances consistentes con la problemática a resolver y que, al mismo tiempo, permitan aportar nuevas ideas a las técnicas de

Conclusiones

reconocimiento, por ejemplo, el reconocimiento de la actividad en espacios cerrados como puede ser una oficina o una habitación.

Bibliografía

- [ES04] R. Escontrela Mao and L. Stojanovic Casas, “La integración de las TIC en la educación: Apuntes para un modelo pedagógico pertinente,” *Revista de Pedagogía*, vol. 25, 2004, pp. 481-502.
- [GH08] M. González Sanchez and M.J. Hernández Serrano, “Interpretación de la virtualidad. El conocimiento mediado por espacios de interacción social,” *Revista apertura*, vol. 8, 2008.
- [AC08] J.A. Andrade Castro and M.S. Campo Redondo, “Tecnologías de Información. Inclusión en la educación basada en lo digital,” *Revista Mexicana de Investigación Educativa*, vol. 13, 2008, pp. 223-248.
- [Educ] Educatic, “¿Por qué usar Tecnología para enseñar y aprender?,” *Tecnología para la educación*.
- [Sal97] J. Salinas, “Nuevos ambientes de aprendizaje para una sociedad de la información,” *Revista Pensamiento Educativo*, vol. 20, 1997, pp. 81-104.
- [DV07] J. Domènech and J. Viñas, *La organización del espacio y del tiempo en el centro educativo*, GRAO, 1997.
- [CB90] M.I. Cano Martínez and Á.L. Lledó Becerra, *Espacio, comunicación y aprendizaje*, Diada, 1990.
- [Lop08] M. López Ornelas, “¿Por qué hablar de usos educativos en las nuevas tecnologías?,” vol. 10, 2008.
- [Gom08] G. Gómez Ávalos, “El uso de la tecnología de la información y la comunicación y el diseño curricular,” vol. 32, 2008, pp. 77-97.
- [Gam07] F. Gamboa Rodríguez, “Diseño, uso y evaluación de tecnología de cómputo ubicuo en el aula universitaria, bajo un enfoque centrado en el usuario y su tarea,” 2007.
- [Haw02] R. Hawkins J, “Ten Lessons for ICT and Education in the Developing World: Readiness for the Networked World,” *The Global Information Technology Report 2001-2002*, 2002, pp. 38-43.
- [Cub07] J. Cubero Allende, *¿Tecnología e Metodología?*, 2007.
- [PM95] G. Passariello and F. Mora, *Imágenes Médicas. Adquisición, Análisis, Procesamiento e Interpretación*, Equinoccio, 1995.
- [San05] Á. Sánchez Calle, *Aplicaciones de la visión artificial y la biometría informática*, Librería-Editorial Dykinson, 2005.
- [GHL71] A.J. Goldstein, L.D. Harmon, and A.B. Lesk, “Identification of human faces,” *Proceedings of the IEEE*, vol. 59, 1971, pp. 748-760.
- [Kan77] T. Kanade, “Computer Recognition of Human Faces,” *Interdisciplinary Systems Research*, vol. 47, 1977.
- [MKN02] Y. Ming-Hsuan, D.J. Kriegman, and A. Narendra, “Detecting Faces in Images: A Survey,” *IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE*, vol. 24, 2002, pp. 34-58.
- [SB+93] S.A. Sirohey, M. Begum, I. Sirohey, and Z. Sirohey, “Human face segmentation and identification,” *Technical Report CS-TR3176*, 1993.
- [CL93] D. Chetverikov and A. Lerch, “Multiresolution Face Detection,” *Theoretical Foundations on Computer Vision*, vol. 69, 1993, pp. 131-140.

Bibliografia

- [Dem68] A.P. Dempster, "A Generalization of Bayesian Inference," *Journal of the Royal Statistical Society. Series B (Methodological)*, vol. 30, 1968, pp. 205-247.
- [TP91] M. Turk and A. Pentland, "Eigenfaces for Recognition," *Cognitive Neuroscience*, vol. 3, 1991, pp. 71-86.
- [AS93] M.F. Augusteijn and T.L. Skujca, "Identification of Human Faces through Texture-Based Feature Recognition and Neural Network Technology," *Proc. IEEE Conf. Neural Networks*, 1993, pp. 392-398.
- [BS+00] A.J. Briggs, D. Scharstein, D. Braziunas, C. Dima, and P. Wall, "Mobile Robot Navigation Using Self-similar Landmarks," *Proceedings of the 2000 IEEE International Conference on Robotics & Automation*, vol. 2, 2000, pp. 1428-1434.
- [LSM06] D. Langdon, A. Soto, and D. Mery, "Automatic Selection and Detection of Visual Landmarks Using Multiple Segmentations," *Advances in Image and Video Technology*, vol. 4319/2006, 2006, pp. 601-610.
- [KI01] Y. Kuk-jin and K. In-so, "Artificial Landmark tracking based on the color histogram," *Proceedings of the 2001 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pp. 1918-1923.
- [Sau05] P. Suau, "Robust Artificial Landmark Recognition Using Polar Histograms," *Lecture Notes in Computer Science*, vol. 3808/2005, 2005, pp. 455-461.
- [TT05] E. Todt and C. Torras, "Color-Contrast Landmark Detection and Encoding in Outdoor Images," *Computer Analysis of Images and Patterns*, vol. 3691/2005, 2005, pp. 612-619.
- [CAP02] P. Capilla Perea, J.M. Artigas, and J. Pujol i Ramo, *Fundamentos de colorimetria*, Universitat de València, 2002.
- [LP07] R. Lukac and K. Plataniotis, *Color Image Processing: Methods and applications*, Taylor and Francis Group, 2007.
- [KA08] A. Koschan and M. Abidi, *Digital Color Image Processing*, Wiley, 2008.
- [Jan05] B. Jähne, *Digital Image Processing*, Springer, 2005.
- [GW02] R. Gonzalez and R. Woods, *Digital Image Processing*, Prentice Hall, 2002.
- [GW87] R.C. Gonzalez and P. Wintz, *Digital Image Processing*, Addison Wesley, 1987.
- [DH72] R.O. Duda and P. Hart, "Use of the Hough Transform to Detect Lines and Curves in Pictures," *Communications of Association for Computing Machinery*, vol. 15, 1972, pp. 11-15.
- [Bal81] D.H. Ballard, "Generalizing the Hough Transform to Detect Arbitrary Shapes," *Pattern Recognition*, vol. 13, 1981.
- [HAA92] T.D. Haig, Y. Attikiouze, and M.D. Alder, "Border following : new definition gives improved borders," *IEEE PROCEEDINGS-I*, vol. 139, 1992, pp. 206-211.
- [Haw94] P.W. Hawkes, *Advances in Electronics and Electron Physics, Volumen 89*, Academic Press, 1994.
- [Per03] J.F. Pertusa Grau, *Técnicas de análisis de imagen: aplicaciones en biología*, Universitat de València, 2003.
- [BI98] A. Blake and M. Isard, "Active Contours," *Springer-Verlag*, 1998.
- [KWT88] M. Kass, A. Witkin, and D. Terzopoulos, "Snakes: Active Contour Models," *International journal of Computer Vision*, 1988, pp. 321-331.
- [TV98] E. Trucco and A. Verri, *Introductory Techniques for 3-D computer vision*, Prentice Hall, 1998.
- [VJ04] P. Viola and M. Jones, "Robust Real-Time Face Detection," *International journal of Computer Vision*, vol. 57, 2004.

Bibliografía

- [VJ01] P. Viola and M. Jones, “Robust Real-Time Object Detection,” *International journal of Computer Vision*, 2001.
- [A+01] D. Abrosimov, A. Alekseev, M. Asbach, and S. Boldyrev, “Intel® Open Source Computer Vision Library,” 2001.
- [FS97] Y. Freund and R.E. Schapire, “A Decision-Theoretic Generalization of On-Line Learning and an Application to Boosting,” *journal of computer and system sciences*, vol. 55, 1997, pp. 119-139.
- [TV00] K. Tieu and P. Viola, “Boosting Image Retrieval,” *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2000, pp. 228-235.
- [FA91] W.T. Freeman and E.H. Adelson, “The desing and use of steerable filters,” *IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE*, vol. 13, 1991, pp. 891-906.
- [Eur] Euresys Inc, *Euresys Multicam Driver Library*, Euresys, .
- [Seo] N. Seo, “Tutorial: OpenCV haartraining (Rapid Object Detection With A Cascade of Boosted Classifiers Based on Haar-like Features).”