

**SISTEMA PROTOTIPO DE DETECCIÓN DE MOVIMIENTO Y
TRACKING DE PERSONAS CON CONTROL SOBRE EL PAN-
TILT DE LA CÁMARA, COMO APOYO A SISTEMAS DE
VIGILANCIA**



**UNIVERSIDAD DISTRITAL
FRANCISCO JOSÉ DE CALDAS**

Autor
Jorge Andrés Pinzón

**UNIVERSIDAD DISTRITAL FRANCISCO JOSÉ DE CALDAS
Facultad Tecnológica
Ingeniería en Control**

Bogotá D.C., 2021

SISTEMA PROTOTIPO DE DETECCIÓN DE MOVIMIENTO Y TRACKING DE PERSONAS CON CONTROL SOBRE EL PANTIL DE LA CÁMARA, COMO APOYO A SISTEMAS DE VIGILANCIA



**UNIVERSIDAD DISTRITAL
FRANCISCO JOSÉ DE CALDAS**

**Jorge Andrés Pinzón
Código: 20191383032**

Email: jpinzon9201@gmail.com

Monografía

Presentado para optar al título de: Ingeniero en Control y Automatización

Director

Ing. Msc. Miguel Ricardo Pérez Pereira

**UNIVERSIDAD DISTRITAL FRANCISCO JOSÉ DE CALDAS
Facultad Tecnológica
Ingeniería en Control**

Bogotá D.C., 2021

Índice

Índice.....	i
Índice de Figuras.....	ii
Índice de Tablas.....	v
Índice de Anexos.....	vi
Lista de Abreviaturas y Siglas.....	vii
Resumen.....	viii
1. Introducción.....	1
2. Justificación.....	4
3. Objetivos.....	5
3.1. Objetivo General.....	5
3.2. Objetivos específicos.....	5
4. Marco de referencia.....	6
4.1. Antecedentes.....	6
4.2. Marco teórico.....	13
4.3. Marco Legal.....	32
5. Metodología.....	35
6. Resultados.....	53
6.1. Propuesta definitiva.....	49
6.2. Restricciones.....	52
6.3. Validación de la solución.....	52
6.4. Evaluación del impacto de la solución propuesta.....	54
7. Conclusiones y Recomendaciones.....	66
8. Referencias.....	67
9. Anexos.....	71

Índice de Figuras

Figura 1.: Imagen digital y su representación numérica	13
Figura 2.: Estructura básica de OpenCV	18
Figura 3.: Representación esquemática de los wavelets	20
Figura 4.: Diagrama de flujo del método de sustracción de fondo	23
Figura 5.: Posición relativa de dos circunferencias cuando son exteriores	24
Figura 6.: Posición relativa de dos circunferencias cuando son tangentes exteriores	25
Figura 7.: Posición relativa de dos circunferencias cuando son secantes	25
Figura 8.: Posición relativa de dos circunferencias cuando son tangentes interiores	25
Figura 9.: Posición relativa de dos circunferencias cuando son interiores	26
Figura 10.: Posición relativa de dos circunferencias cuando son concéntricas	26
Figura 11.: Diagrama de cableado para RS232	27
Figura 12.: Representación de un sistema de control por lazo cerrado	30
Figura 13.: Diagrama de bloques para el ESP8266	31
Figura 14.: Módulo NodeMCU WI-FI certificado FCC, con antena PCB	32
Figura 15.: Diagrama de bloques simplificado del sistema	35
Figura 16.: Primeros resultados algoritmo de detección de personas y sustracción de fondo por media	37
Figura 17.: Representación de movimiento (grados), para el soporte del prototipo. Fuente (imagen propia)	39
Figura 18.: Explorador de variables plataforma Python	40
Figura 19.: Script sencillo de envío y recepción con el uso de la Librería Pyserial	41
Figura 20.: Algoritmo pre entrenado Haar y sustracción de fondo por media	42

Figura 21.: Método analítico para encontrar el centroide de un rectángulo	43
Figura 22.: Integración del algoritmo Haar y sustracción de fondo por media con el principio de posición relativa de dos circunferencias	44
Figura 23.: Color y probabilidad asociados a dos circunferencias con posición relativa al exterior	45
Figura 24.: Color y probabilidad asociados a dos circunferencias con posición relativa tangente exterior	45
Figura 25.: Color y probabilidad asociados a dos circunferencias con posición relativa secante	46
Figura 26.: Color y probabilidad asociados a dos circunferencias con posición relativa tangentes exteriores	46
Figura 27.: Color y probabilidad asociados a dos circunferencias con posición relativa interiores	47
Figura 28.: Color y probabilidad asociados a dos circunferencias con posición relativa concéntrica	47
Figura 29.: Mapeo del frame principal, conforme a la distribución de las coordenadas por escena captada	48
Figura 30.: Tiempo calculado para la reconstrucción de las señales, movimiento pan y tilt	50
Figura 31.: Tiempo calculado para la reconstrucción de las señales, movimientos aleatorios de la cámara	50
Figura 32.: Eliminación de perturbaciones en la integración de los algoritmos para la detección de personas	51
Figura 33.: Simulación primaria del sistema con eje pan de la cámara y el monitor serie del IDE asociado al microcontrolador escogido	52
Figura 34.: Diagrama de flujo sistema prototipo detección y tracking de persona	53
Figura 35.: Servomotor SG90 y dimensiones	55
Figura 36.: Soporte para el servomotor SG90	55
Figura 37.: Sistema prototipo montado	56

Figura 38.: Detección y tracking de personas por el prototipo del sistema	58
Figura 39.: Instrucción para obtener una fuente de video por el protocolo RTSP	61
Figura 40.: Representación de los datos por simulación del puerto serie servomotor panorámico	63
Figura 41.: Representación de los datos por simulación del puerto serie servomotor de inclinación	63
Figura 42.: División teórica por cuadrantes del frame principal acorde a la señal de error (dx,dy)	64
Figura 43.: División teórica por cuadrantes del frame principal acorde a la señal de error	65

Índice de Tablas

Tabla 1: Tabla de protocolos serial.....	26
--	----

Índice de Anexos

Anexo 1.....	62
--------------	----

Lista de Abreviaturas y Siglas

Sigla/Abreviatura	Significado
FOV	Field Of View
SVM	Support Vector Machines
ML	Machine Learning
ROI	Region Of Interest
HOG	Histogram of Oriented Gradients
PID	Proporcional Integral Derivativo
LQR	Linear-Quadratic Regulator
DSL	Digital Subscriber Line
LBP	Local Binary Patterns
USB	Universal Serial Bus
PWM	Pulse Width Modulation
IDE	Integrated Development Enviroment

Resumen

El presente documento expone y destaca la importancia de las técnicas de procesamiento de imágenes, desarrollo e implementación de controladores, instalación y puesta en marcha de proyectos orientados a sistemas de videovigilancia o seguridad, cuyo valor radica en garantizar la seguridad de la persona a detectar y en otras ocasiones la integridad de los bienes e instalaciones, adicionalmente se usan con fines de investigación, la asistencia sanitaria o el control de la prestación laboral por parte de los trabajadores. Y es debido al cambio constante en el hardware, técnicas y manejo de los datos en la detección/tracking de objetivos, en cuánto van a la par de las cámaras fotográficas y de video disponibles en el mercado, que hacen de este proyecto idóneo en referencia a la tecnología actual y de los aportes que se puedan realizar a los sistemas implementados en la nación.

Un conjunto significativo de aplicaciones de videovigilancia plantean sistemas de cámara estática, lo que traduce en errores de detección en el cambio de fondo (entorno por el que se movilizan los objetivos), movimiento fuera del rango que captura la cámara, o por cambios abruptos en el sistema que ocasionan errores en el seguimiento de uno o varios objetivos. Razón por la cual lo conveniente en este tipo de sistemas es modificar la ubicación de cada cámara a zonas de interés, o donde exista un mayor número de personas u objetivos, este último enfoque se establece en metodologías que van dirigidas a vehículos (tripulados o no), robots, parte de maquinaria industrial entre otros sistemas o subsistemas de seguridad.

Luego es posible dar un cambio a las implementaciones por cámara estática, utilizando una ampliación del campo de visión (FOV), enfatizado en el proyecto con un movimiento en el eje panorámico y de inclinación de la cámara, por supuesto contemplando perturbaciones por iluminación, cambios de escena (resultado del mismo movimiento de la cámara), comunicación, entre otros, además del enfoque que permita determinar un valor de referencia o de "algo" que lo compense, entonces lo desarrollado aplica al mantenimiento del eje óptico de la cámara en el frame principal (cuadro donde se ve la transmisión de lo captado por el sensor de la imagen). El mapeo de dicho frame sirve para hallar información relevante de comportamiento del sistema por control de realimentación, y por supuesto la señal de error asociada al punto consigna que varía de acuerdo a la fuente de video.

El resultado se traduce en el seguimiento de una(s) persona(s) que se encuentre en el FOV del frame principal, con el algoritmo de detección y ajuste que realiza el sistema con la estrategia de control desarrollada, en segunda medida con las indicaciones que hacen pertinente trazar una ruta del objetivo por medio de herramientas de simulación; luego del cómo puede distar de un funcionamiento relativamente aceptable, dado por la señal de error establecida en el algoritmo principal y por último de la validación de la operatividad por medio de un prototipo del sistema.

Palabras claves: Campo de visión, detección/tracking, eje óptico, estrategia de control, procesamiento de imágenes, seguridad, videovigilancia

1. Introducción

La gama inmensa de formas, figuras, colores, texturas, movimiento y tranquilidad ha acompañado los seres vivos desde el inicio, brindando a una parte de ellos la capacidad de percibir, adquirir, integrar e interpretar la abundante información visual que les rodea. Ahora bien para los seres humanos particularmente, se ha vuelto un reto impartir estas capacidades a una máquina, para que esta analice los datos incrustados en imágenes fijas, gráficos, videos o imágenes en movimiento del mundo sensorial; por ello es importante comprender las técnicas de almacenamiento, procesamiento, transmisión, reconocimiento y por ultimo interpretación de dichas escenas visuales.

Uno de los aspectos fundamentales del procesamiento de imágenes implica la compresión y codificación de la información visual y es debido a la creciente demanda de aplicaciones de esta tecnología, además de los requisitos en almacenamiento de las imágenes digitales que hacen de esta ultima una opción creciente y excepcional en el mercado, junto con la transmisión a través del ancho de banda de comunicación, así logra ser un área de desarrollo activa en la actualidad; luego dichas técnicas de compresión de imágenes ayudan a reducir las redundancias en los datos de imágenes sin procesar y con esto reducir el ancho de banda de almacenamiento y comunicación.

Existen muchas aplicaciones al procesamiento de imágenes en el amplio espectro de actividades humanas, que van desde la interpretación de escenas por teledetección hasta la interpretación de imágenes biomédicas, algunas de ellas pueden contener aplicaciones o técnicas de otras actividades con el fin de mejorar u optimizar lo desarrollado; para la alternativa propuesta existen diversos principios uno de ellos es la detección-tracking de objetivos que consiste en el seguimiento de objetos en movimiento, midiendo los parámetros del mismo y obtener un registro visual del objeto moviéndose, dicho esto de forma general existen dos enfoques diferentes para el seguimiento de objetos.

El tracking basado en el reconocimiento y el que está basado en movimiento; un sistema para tracking de objetos rápidos (por ejemplo, un avión militar, un misil, etc.) está desarrollado en técnicas de predicción basadas en el procesamiento de objetos en movimiento, como el filtrado de Kalman, Kalman ampliado, filtrado de partículas entre otros. Aquellos sistemas de tracking de objetos con lógica de procesamiento automatizado de imágenes están sujetos a que una vez marcado por el campo de visión del sensor, estos objetivos se adquieran sin la necesidad de intervención humana. Por su parte en el tracking basado en reconocimiento, utilizado en el proyecto, presenta el patrón del objeto que se identifica en los fotogramas sucesivos de cada imagen y el tracking se lleva a cabo utilizando su información posicional.

En las siguientes secciones se establecerán las pautas necesarias para comprender lo que concierne al antes, durante y después (trabajos que puedan surgir en consecuencia a la investigación) de la ejecución del proyecto, entonces, primero se abordara el planteamiento la justificación y los objetivos que hicieron pertinente la realización del prototipo, luego del marco conceptual, sección donde se fundamentan los aspectos más destacables en torno a las técnicas y

metodologías desarrolladas, se procede a exponer esta última como una serie de etapas: la observación sistémica, la medición, experimentación, formulación, análisis de resultados con modificaciones y/o mejoras donde sean posibles aplicarse, para así dar una conclusión al proyecto con recomendaciones varias a la hora de abordar o implementar el sistema prototipo.

Planteamiento del problema

Uno de los primeros pasos para la detección y tracking de personas, es determinar si existió un movimiento en base a una serie de imágenes, lo cual ha sido tema de estudio por años. Existen varias alternativas para calcular dichos cambios entre cuadros consecutivos, algunos son más complejos que otros a nivel computacional y eficacia en los resultados, sin embargo pueden surgir factores a considerar cuando se instalan cámaras exteriores, como lo son el momento del día, la nubosidad y época del año. En consecuencia se hace necesario utilizar ventanas de movimiento cuya finalidad consiste en aplicar un conjunto de operaciones lógicas, para procesar solo los movimientos dentro que se detecten dentro de sus dimensiones y descartar aquellos que estén por fuera.

Este último aspecto es crucial cuando se hace necesario implementar un algoritmo, que a diferencia de la mayoría ha sido diseñado para configuraciones con cámara fija, ya que las técnicas de sustracción de fondo o los algoritmos de diferenciación temporal no pueden utilizarse directamente para cámaras activas (movimiento en sus ejes *pan-tilt*).

Acorde al paso de los años han existido varios intentos para desarrollar sistemas de videovigilancia que sean capaces de operar con cámaras en movimiento, muchos de estos proyectos han concluido en algoritmos que con ciertas restricciones han dado solución al problema, sin embargo gran parte de ellos son inconclusos en los métodos y simplificación de los resultados en referencia a dichas restricciones; junto con todos aquellos análisis y obtención de los datos involucrados que caractericen por sencillez e interpretación, que en consecuencia funcionan como incentivo en la investigación de los sistemas de videovigilancia y/o seguridad.

2. Justificación

En la sociedad y jurisdicción colombiana (cámaras de vigilancia en las calles) los dispositivos que se suelen instalar son de emplazamiento fijo, lo que en la mayoría de los casos implica la ubicación estratégica en los puntos de interés, dejando de lado los objetivos que se mueven fuera del marco de la imagen, dicho esto, las tecnologías que implican tracking aún están en fase de investigación y pruebas en el contexto de desarrollo y puesta en marcha, no obstante existen muchos dispositivos disponibles en el mercado extranjero y que posteriormente han sido asequibles para la mayoría de las aplicaciones que se suelen implementar en empresas, locales comerciales, eventos públicos, conjuntos residenciales entre otros.

Esto hace que sea de carácter imperativo la investigación para proyectos donde se desee mitigar los errores por cambio de fondo en la imagen, factores de instalación o del medio donde esta se realiza, que a menudo es de modo dinámico, adicionalmente de las técnicas distintas al tracking dadas por ejemplo en Machine Learning, SVM (Support Vector Machines), Big Data, (junto con el hardware, conectividad y software requerido), junto con las bases para un uso adecuado y coherente de las tecnologías disponibles, así como el aporte a proyectos más específicos con respecto a la codificación y transmisión de los datos adquiridos y/u obtenidos.

3. Objetivos

3.1. Objetivo General

Diseñar e implementar un sistema prototipo de detección y tracking de una persona utilizando procesamiento de imagen y estrategias de control sobre el pan-tilt de la cámara, como apoyo a sistemas de videovigilancia o seguridad.

3.2. Objetivos específicos

- Definir e implementar el algoritmo de detección y tracking de una persona más acorde con el sistema a desarrollar
- Implementar la estrategia de control sobre el pan-tilt de la cámara, que permita realizar el tracking de una persona
- Realizar la integración del sistema
- Validar el funcionamiento mediante un prototipo del sistema, haciendo pruebas en espacios reales

4. Marco de referencia

4.1. Antecedentes

En cuanto a los referentes nacionales que involucran buena parte del proyecto en lo que respecta a detección esta lo desarrollado por Santos D., Dallos L, A. Paulo, Gaona Garcia [1], que expone un análisis de los algoritmos basados en técnicas de visión por computador y el machine learning, como el papel que desempeña para identificar, rastrear y clasificar diferentes elementos y patrones que se encuentran en los archivos de video, no obstante su importancia recae en la metodología utilizada; es del tipo descriptivo/experimental, esto ayuda particularmente a caracterizar las variables de entrada para el desarrollo en presencia del conjunto de imágenes del objeto a seguir, para posteriormente utilizar la etapa de aprendizaje no supervisado donde se encuentran las características relevantes con k-means y posteriormente con los diferentes frames del video se pueda realizar la clasificación.

Dicho esto y siendo aún más exhaustiva la labor de desarrollo, los autores plantean las fases de la siguiente manera: 1. La fase de exploración que consiste en la recolección de los datos necesarios para el entrenamiento y pruebas del modelo basado en red neuronal, 2. Fase de procesamiento de los datos; aquí se lleva un análisis para la limpieza y entereza de los datos con el fin de eliminar el ruido y demás datos atípicos, 3. Fase de definición del modelo inicial se plantea un modelo de red neuronal, realizando un conjunto de pruebas y ajustes con el fin de determinar el idóneo para el sistema, 4. Fase de pruebas que como su nombre lo indica consiste en modificar la configuración del modelo inicial, para hallar aquellos que se ajustan mejor en lo que corresponde a las predicciones y por último, 5. La fase de comparación de resultados; aquí se llevó a cabo la correspondencia entre los trabajos realizados de forma inicial y final, ya que es un hecho que para diferenciar las características de un objeto de otro, su descripción debe tener tal detalle y particularidad que garantice inequívocamente su diferenciación, ahora esta última parte es clave en la visión computacional, luego otra manera de alcanzar dicha exhaustividad es con un dataset (conjunto de datos referido a variables) con una función en particular, como lo son los tipo Dashcam utilizados por Victoria Mosquera C. y Narvaez Dussan G [2], cuya aplicación son la detección y seguimiento de vehículos e implementarlos en tiempo real, aplicando criterios de selección tales como mayor cantidad de imágenes del objeto que se quiere detectar, eficacia con el entrenamiento de la red siendo este último factor decisivo para aplicar la red de convoluciones multicapa que ayudan con la precisión, velocidad y entrenamiento computacional. No obstante tal y como radica este costo lo que se acostumbra a hacer es acudir inclusive a múltiples redes pre entrenadas con el fin de determinar cuál es la mejor en cada caso o experimento, para así concluir con un recurso de web (Darknet) que permite comparar su desempeño con respecto a otras metodologías y datasets cada vez más dedicados en cuanto a la precisión y uso del hardware utilizado (tarjeta de gráficos de alta gama Nvidia Gforce Pascal Titan X).

Ahora ¿qué sucede cuando el objeto a detectar se encuentra en ambientes de alta oclusión?, y lo que es más, que se centre en aquellos que priman en la seguridad de las personas, pues bien este es el caso de Oviedo Espinosa J. [3], que al enfocarse en motocicletas está mitigando la creciente problemática de los accidentes de tránsito incluso en países emergentes. Así los operadores de los sistemas de vigilancia de tráfico urbano podrían mejorar dicho monitoreo y los índices de

accidentalidad. A nivel operativo implementa en primera medida un clasificador para el objeto deseado en escenarios urbanos utilizando un modelo de red neuronal convolucional de pre entrenamiento, que como en el caso anterior funciona como un extractor de características, con la diferencia de que ahora estos datos extraídos de la red CNN serán evaluados por una máquina de soporte vectorial (SVM), así la estrategia permite evaluar el conjunto de datos del tráfico urbano, para llegar a precisiones cercanas al 99.4% con un dataset de 3 y 5 clases respectivamente. Luego esto permite que el objetivo de dicho proyecto esté en mejorar y comparar métodos más robustos de detección con conjuntos de video y aplicar la parte de tracking con las herramientas más usadas que son el filtro de Kalman y Lucas Kanade.

Resumiendo, Kalman en primera instancia identifica una función de distancia entre los centros de gravedad de las regiones detectadas en diferentes fotogramas, así dicha función fall aplica otras dos funciones alternativas: por comparación de información fotométrica y una estimación del movimiento basada en el filtro de Kalman. El método implementa también una estrategia de "propagación de pistas", diseñada para hacer frente a la oclusión y a los errores esporádicos, en la que cada pista identificada como desunida se incluye en un conjunto de objetos de rastreo no conectados; las pistas se propagan con fin de comprobar si algunas de las pistas inconexas realizan el seguimiento del objeto. Con respecto al método de Lucas Kanade este implementa un marco de aprendizaje de patrones de movimiento multinivel (MLMP) para el análisis del comportamiento de la trayectoria en cámaras de videovigilancia, pese a que el número de trayectorias utilizadas para representar un solo objeto podría fallar en escenarios congestionados o utilizando un ángulo de cámara no ortogonal hacia los objetos de interés.

Otras alternativas en el tracking de objetos residen en la predicción y actualización del estado futuro a partir de una secuencia de observaciones, dicho enfoque fue realizado por Flórez Pinzón C. [4], donde el método pila es el filtro de partículas que se desenvuelve por medio de la evolución temporal de las mismas, e implementa secuencialmente modelos aleatorios con ellos logra aproximar las funciones de densidad de probabilidad y su evolución en el tiempo, directamente como un número finito de muestras discretas llamadas partículas. El enfoque es acercar distribuciones relevantes con diferentes mediciones aleatorias compuestas por partículas aleatorias, cada una por supuesto con su respectiva posición tamaño y forma. Siendo un poco más detallado con respecto al cómo trabaja el algoritmo, la autora recalca las fases o etapas que lo componen siendo la inicialización la primera de ellas, que se encarga de crear una población M de partículas usando una función de densidad de probabilidad conocida o aleatoria, en la segunda fase el re-muestreo se elige el rango de la ubicación de las partículas y su tamaño, por último se encuentra, la fase de estimación dinámica que se puede representar como el movimiento del objeto de un cuadro a otro, es decir, en este paso es donde se predice la posición del objeto, el siguiente cuadro es con respecto a la posición del mismo referido al cuadro actual.

Cuando la labor es implementar el sistema de seguimiento como tal, ya sea por medio de un dispositivo programable ajeno al hardware computacional o bien por control remoto como sucede en [4], lo importante es fijar un punto de referencia que se pueda "llamar/utilizar" con el fin de aplicar estrategias que sirvan para optimizar el proceso desde su captura hasta su ejecución, tal es así que a nivel nacional Imbachi- Buhelly F, Montenegro-Bravo D, Fierro-Pencue L y D. Ferrin-Bolaños [5], destacan el uso de técnicas y en general un sistema de control cuya referencia este

en el centro de la imagen (eje óptico de la cámara), ya que en perspectiva afecta entre otras cosas a la distancia focal que corresponde a la distancia entre la película (plano) y el centro óptico de la lente que apunta al objetivo cuando el mismo pretende dirigirse hacia el "infinito". Ahora bien de dicha longitud focal depende el ángulo de visibilidad de cada objetivo, permitiendo abarcar en mayor o menor parte la escena a capturar. Luego para seguir el objeto en tiempo real se detecta en cada fotograma su posición actual calculando el centro de masa y el centro de la imagen, estas diferencias son enviadas al algoritmo de control como señales de error, para que los actuadores (motores) hagan la corrección adecuada en cada dirección.

Debido a que en gran parte de los proyectos el hardware es decisivo a la hora de tomar decisiones, tratándose del tracking de un objetivo cuando sucede fuera del cuadro de la imagen, es corroborado por trabajos como el de Valencia Ortiz N. y Londoño Vargas L. [6], cuyo sistema está basado en la cabeza de un robot imoov para la detección y seguimiento. Para la implementación del sistema los autores trazan una ruta de objetivos generalizada que consisten en: realizar el modelo cinemático (aspectos del movimiento con la geometría involucrada) de la cabeza del robot, selección e implementación de la instrumentación que requiere el movimiento de la cabeza, que al igual que como sucede con las tarjetas electrónicas deben operar en conjunto con los actuadores y las etapas de adquisición de datos, para seleccionar el algoritmo basado en la visión computacional, que detecte cada vector de movimiento bidimensional en el campo de visión del robot. Por último diseñan e implementan un algoritmo de interacción entre aquel que detecta el vector de movimiento y el que se encarga de mover la cabeza del autómatas.

A nivel internacional y en referencia a lo citado en el presente documento, se ve una constante en la detección y tracking, siendo esta última junto con el hardware materia de énfasis en las fuentes que más se aproximan a la metodología desarrollada en el proyecto, tal y como sucede con Eapen Mathews A. , Jacob R., S. Reshma Biju y Ushus Joseph M. [7], cuyo modelo simplificado se basa en un subsistema de captura de imágenes con un módulo y plataforma PixyCam que opera con un acelerómetro, que posiciona al microcontrolador en una etapa que concuerda a una salida, en consecuencia logra poner en marcha los servomotores y la cámara. Así gracias a la plataforma que utiliza un algoritmo de código de color inteligente; que rastrea y guarda muchas formas de color con la configuración que requiere cada usuario, simplifica bastante el costo computacional que implican proyectos de este calibre (eficacia y bajo coste).

Es en estos casos particulares donde la detección/tracking de objetivos se hacen con otro tipo de señales y enfoques que permiten establecer un punto de partida para el sistema prototipo tratado, tal como el proyecto de Sáenz Esqueda, Sáenz López, Reyes Espino y Morales Salazar [8], establece un servomotor de giro continuo que recibe señales de velocidad con el fin de lograr el seguimiento de un objeto junto con su posición en cada instante empleando librerías Aforge.NET, para que la computadora en dos fases (segmentación e interpretación), realice el filtro de la información en la imagen que no es requerida y aplique la que si lo es, entre otras cosas para realizar cálculos como el área y ubicación del centroide. El sistema se acopla a una interfaz gráfica de usuario con los elementos que permiten al usuario modificar los resultados, que corrigen la captura y posición con el hardware implementado.

Entrando en materia de detección de personas los algoritmos cambian notablemente, es decir, en trabajos como el de Matilla Garcés A., Peña Guerrero F., Cálvelo Garcés A., Silva Hernández E., y Belett Matilla M.,[9], las personas son detectadas por un método similar al de Dalal y Triggs (HOG+ Máquina de Soporte Vectorial SVM), limitándose a aquellas regiones de interés (ROIs) que contengan al menos una persona, con el objetivo de mejorar la mayor eficiencia en el sistema. El detector utiliza una ventana dedicada a la localización de deslizamiento en la imagen acorde a la ventana del detector, posteriormente se calcula un respectivo descriptor HOG, que a su vez es procesado por la SVM entrenada previamente para clasificar al objetivo como una persona o no. Como se mencionó anteriormente en esta y otras propuestas se realiza una adaptación de la imagen que permite encontrar el contorno correctamente y también lograr mayor eficiencia. Esta fase puede contemplar varios aspectos como lo son la transformación de la imagen de RGB a escala de grises, para mejorar el rendimiento y simplificar los cálculos sucesivos, redimensionar la imagen y en general la ecualización del histograma de la imagen, lo cual maximiza el contraste de los píxeles que se tienen que computar.

Tal y como sucede con los objetos, en lo que compete a la detección de personas existen alternativas como el hardware específico, que facilite y por supuesto permita dicha labor, en el prototipo desarrollado por Mazher Iqbal y K. Iswarya Lakshmi [10], básicamente, define la capacidad que tiene una cámara térmica para traducir el calor en luz visible para analizar una escena u objeto en concreto, destacando su funcionalidad incluso en lugares oscuros, durante la nubosidad o en cualquier condición anormal. A este tipo de imágenes producidas se le conoce como termograma, de manera que el sistema trabaja con una cámara térmica (testo 870-1), que al contrario de la vista o captura convencional que distingue objetos iluminados por el sol u otra forma de luz, funciona viendo el calor en vez de la luz reflejada, razón por la cual tiene un aspecto diferente al de las imágenes visibles de la cámara o del ojo, no obstante puede utilizar las metodologías y parte de las herramientas que maneja cualquier otro sistema y/o prototipo que contenga el procesamiento de señales/imágenes, pasando en este caso por una fase de computación, colección de fotogramas, lectura de fotogramas, detección de objetos (en primer plano), asignación de las detecciones de movimiento, tracking (filtro de Kalman) con su respectivo resultado de seguimiento, para volver a la fase de lectura en naturaleza de un sistema realimentado.

Independientemente de que se use una cámara térmica o una de captura convencional, los puntos clave en el tracking del proyecto siguen una lógica similar a la planteada por Mir-Nasiri N. [11], que deja un apartado completo a dicha estrategia; consiste en el cálculo de un centroide acorde a la aplicación de un cuadrante de la imagen (división como el plano bidimensional cartesiano), que al igual a como lo hace el ser humano sigue objetos en movimiento cambiando la orientación de los ojos hacia el objeto, comprobando constantemente los errores de localización del centroide respecto al centro del plano de la imagen y toma medidas correctoras para compensarlo. Las cuatro posibles combinaciones de signos de error para dx y dy (diferencia entre punto medido y centroide), determinan de forma única el cuadrante en el que se mueve el objeto y los controladores del sistema eligen adecuadamente una de las cuatro posibles combinaciones de rotación para dos motores instalados en la cámara, cuyo mecanismo opera a dos grados de libertad siendo el movimiento del motor panorámico el que compensa a dx (Pan), y el motor de inclinación por separado el que corresponde a dy (Tilt). Como resultado el objetivo permanece en

el centro del plano de la imagen mientras que la cámara se oriente hacia el objeto. Sin embargo si el objeto se mueve a una velocidad superior a la que el mecanismo de giro e inclinación pueda soportar, saldrá de la vista de la cámara antes de que esta pueda realizar el ajuste de orientación; ahora si el sistema identifica más de un objetivo de interés o no detecta nada, el autor plantea detener la etapa de movimiento hasta que llegue el siguiente fotograma disponible tal y como lo hacen algunos de los algoritmos de sustracción de fondo.

Otros proyectos llevan más allá el desarrollo de los mecanismos que replican las instrucciones que permiten detectar y seguir objetos con diversos fines, uno de estos mecanismos es el pan-tilt que como sucede con Olmez H. [12], se implementaron para mecanismos de giro e inclinación ya que el movimiento del cuello humano puede imitarse mediante la combinación de estos ejes, inclusive agregando otro mecanismo puede surgir otro movimiento de inclinación (el balanceo), así cada carcasa de motor es independiente posee rodamientos que protegen los ejes de los mismos y se ajustan simultáneamente al mecanismo de inclinación con un panel de rodamientos, lo que permite acoplar fácilmente la cámara a la estructura.

En ocasiones integrar los mecanismos y algoritmos, se convierte en una labor bastante dispendiosa para aquellos usuarios que deseen interpretar los resultados de cada proyecto realizado, o bien operar el prototipo observando cambios en el objeto a detectar, por esa razón y similar a lo realizado en [8], se tiene el proyecto de Chaitra J. M., [13], cuyo enfoque opera el sistema en diversos modos (manual y automático), puesto en marcha y visualizado en la pantalla de un ordenador para que se distinga tanto el modo de operación, como el conjunto de datos relevantes que lo componen, por ejemplo el modo manual se activa con un NumPad (el comando numérico 2), así la configuración se mueve con un cursor como lo indique el usuario junto con el movimiento del cursor corresponde a la posición de la cámara, en el modo de seguimiento, la cámara toma el valor del color y la cámara actúa como los algoritmos le ordenan. Ahora bien los resultados de esta propuesta parecen, van encaminados más para el proceso de detección y seguimiento de rostros que de personas u objetos.

La cuestión ahora es determinar específicamente que factores, afectan o mitigan la problemática de tracking sin depender de algoritmos complicados y de alto coste computacional, para ello y como objetivo principal del sistema realizado, están las estrategias de control que en el caso de Miroslaw Pajor, Marek Grudzinski y Lukasz Marchewka [14], lo establecen como un sistema de control de servoaccionamiento que como se planteó anteriormente funciona en bucle cerrado de realimentación, esta vez, utilizando un controlador PID de valor constante en paralelo, debido a la variación esencial entre la dinámica de los objetos que poseen magnitud y dirección (pares, transmisores, pesos y rangos de movimiento), aplicando diferentes parámetros PID para cada eje. Los límites de velocidad se calculan en base a la saturación integral inherente a este tipo de tecnología (error vs tiempo). Luego con respecto a la señal de entrada para el algoritmo, es proporcionada por lo que el autor denomina "marcador de seguimiento en la imagen capturada por la cámara", la desviación producto de la posición en relación con el centro de la imagen en este caso ya no compete a una coordenada como se cita en trabajos anteriores, sino a una velocidad propiamente dicha, que al ser convertida en unidades que los servomotores puedan interpretar deja de lado la parte de experimentación (ensayo y error), para pasar a una completa caracterización y análisis de la planta donde se desea desarrollar esta metodología.

Esta caracterización implica por supuesto la parte mecánica también, ya que como lo expresa el autor no hay demasiada permisividad de falla en una maquina real (grúa de carga), entonces se vale de los modos de funcionamiento para establecer algo que denomina "modo rueda", lo que traduce en un número infinito de rotaciones, ya que esto implica la forma de construcción de la planta por engranajes de reducción en los ejes del cardan (estructura que permite los ejes de movimiento panorámico y de inclinación operar de forma fluida), esto imposibilita montar codificadores integrados sobre esta parte de la estructura, lo que solo permite controlar la velocidad del motor sin operación fija mediante herramientas y plataformas de software convencionales o en carácter del tipo de comunicación utilizada. Adicionalmente las velocidades serán desde luego las salidas del algoritmo PID aplicado y determinaran la posición real de los ejes del cardan para su posterior calibración o retorno a posiciones iniciales.

Sin embargo como lo citan algunos proyectos y su documentación en repositorios, revistas, tesis de grado entre otros, en ocasiones cuando el hardware o los requerimientos no lo ameritan, utilizar el controlador PID (parte Proporcional, Integral y Derivativa), se vuelve innecesario cuando sus variantes o componentes presentan resultados que no distan mucho de esta primera alternativa (P, PD, PI...), ejemplo de uno de ellos se encuentra el prototipo de Maab Luca F. [15], el cual por la construcción y hardware utilizado (similar a un número significativo de propuestas), consigue con respecto al controlador velocidades entorno a los 33ms, luego bajo varios experimentos y gráficos de los datos encontró, en el controlador tipo P un funcionamiento aceptable en términos de velocidad y sobreimpulso, por su parte el controlador PI muestra un comportamiento en "Zig Zag" a largo plazo y logra en un punto alcanzar el ajuste, no obstante el tiempo de estabilización es demasiado largo para el tiempo en que el móvil se desplaza por el marco de la imagen. Cuando opera el sistema en respuesta a una señal tipo paso para el PD y PID respectivamente, la señal pasa a ser demasiado aleatoria al punto de ajuste deseado. La salida de los controladores en ocasiones supera los valores nominales a los que el sistema puede operar y eso traduce en cambios peligrosos para el hardware; luego bajar la componente derivativa no muestra ningún efecto perceptible en el comportamiento.

Ahora bien cuando las técnicas frecuentes en este tipo de proyectos convergen por llamarlo de alguna manera, con la lógica de los sistemas de control es cuando se encuentran las bases sólidas para una posterior ejecución de los objetivos trazados, ejemplo de dicha convergencia se encuentra en trabajos que se centran en el manejo exhaustivo de un solo algoritmo como pasa con Hamsa D., Resham Verma y Shivaram K. R.[16], con un robot que realiza tracking de objetos usando el filtro de Kalman luego lo destacable esta entre otras cosas es el detalle de dicho algoritmo en un aspecto técnico, que permite asociar de alguna manera dichas convergencias textualmente consigna: Sea el estado inicial X_{init} y el valor de la estimación del estado predicho X_p inicialmente sea $X_p(0)$. Este valor es igual al valor inicial y viene dado por $X_p(0) = X_{init}$, ahora la estimación de la medición predicha se calcula a partir del estado predicho. La estimación de la medición predicha queda expresada como $Y_p(k) = g [X_p(k)]$, luego la variable de innovación es el error de estimación que está definido como la diferencia entre la medida y la medida predicha, viene dada por la expresión $e(k) = y(k) - y_p(k)$, a continuación, añadiendo el termino corrector al estado predicho, se calcula la estimación del estado corregido que viene dado por $X_e(k) + ke(k)$, luego se calcula la estimación del estado predicho para el siguiente paso con $X(k+1) = f [X_e(k), U(k)]$, así el estado corregido es la estimación del estado corregido $X_c(k)$.

Finalmente cuando parece ser que con respecto a los algoritmos y sistemas de control lo que resta es optimizar lo desarrollado, surgen alternativas que complementan algunos de los factores no profundizados en la documentación previamente citada; es así como Niramon Ruangpayoogsak [17], describe un apartado en el cual aplica el control LQR para el posicionamiento de un robot móvil que utiliza un sistema de tracking de bajo coste, en él se encuentra el modelo cinemático que bajo la notación de Denavit-Hartenberg establecen los elementos para la matriz de transformación, luego se plantea el modelo de espacio de estados con sus respectivas matrices asociadas a la entrada, para posteriormente utilizar el controlador LQR y lo que ello implique, pasando primero por una ley de control de realimentación, una función de minimización y la ganancia referida al sistema k , así al igual que en otros trabajos la diferencia entre el objeto y el centro del cuadro de la imagen será el error de pixel en la imagen y de nuevo se utilizara como un punto de ajuste, esta vez para un controlador LQR, de tal manera que cuando el objeto permanezca en el centro de la imagen el punto de ajuste corresponderá a un “cero” y esto por supuesto apoyara la estimación de la trayectoria y una consigna de la primera tanda de resultados.

En el repositorio de la facultad (revista visión electrónica) donde se desarrolló la propuesta, se encuentra documentación que va dirigido a las alternativas mencionadas, una de ellas es la descrita por Duarte B. y Borrás P. quienes plantean las estrategias de control como un medio para modelar robots paralelos, pese a que parece distar de la propuesta del sistema prototipo, los autores destacan su importancia en sistemas dinámicos no lineales, que es precisamente el comportamiento que se espera encontrar en este tipo de aplicaciones (visión artificial), no obstante destacan la liberalización por realimentación como una manera sólida de cambiar esta característica, de forma tal, que se puedan aplicar otras técnicas de control convencional como por ejemplo el control por par calculado (Computed Torque Control), cuya aplicabilidad lo hace el segundo más utilizado a nivel industrial después del PID; ya que de manera general toma un modelo linealizado por realimentación e implementa sobre este un control proporcional-derivativo aplicando una ecuación específica. [18]

Cuando no es posible o impráctico realizar la realimentación de un sistema, se acuden a otro tipos de estrategias de control, algunas de ellas aplicadas al control modular de la estructura en cuestión, tal y como sucede con lo desarrollado por Cruz Carbonell V. y Estepa Castillo R. A., que plantean las CPGs (Central Pattern Generators), por las ventajas que representa para con ese proyecto, tales como la ejecución de movimientos complicados con pocos parámetros de control y la "suave y/o armoniosa" respuesta con respecto a los métodos tradicionales. Debido a que representan un fuerte costo computacional, los CFG han dado paso a alternativas como generadores sinusoidales que al ser versiones simplificadas de las GPC se han utilizado en estructuras mecánicas semejantes a las de una serpiente y oruga, proporcionando buenos resultados con una cantidad relativamente menor de cálculos a realizar; por tanto para la propuesta no se descartó el prescindir de ciertas técnicas y estrategias de control para que desde este punto de vista se vuelva ineficiente e inviable computacionalmente.[19]

Otro enfoque utilizado en proyectos de esta naturaleza, inclusive poco antes de aplicar técnicas complejas de control, se utiliza el reconocimiento eficiente de los objetos por visión de máquina, que facilite una eficiencia en la ejecución de tareas como la navegación, el seguimiento entre otros. Dicha tarea de reconocimiento fue abordada por Jiménez Moreno R, Avilés O y Ovalle

D.M., como un algoritmo de aprendizaje a máquina que acompañe a los de procesamiento de imágenes, luego son las redes RNC quienes muestran una mayor robustez que las técnicas de procesamiento de imagen tradicionales, esto puede radicar por su forma general de funcionamiento, la cual resume en tomar directamente la imagen y predecir el objeto presentado por capas de filtrado y convolución, esto las hace idóneas para robots inteligentes, aplicaciones en el campo de la medicina como la de imágenes diagnosticas e incluso en extracción de características para la posterior determinación de patrones de precisión bajo medidas antropométricas (medidas del cuerpo humano). [20]

4.2. Marco teórico

Procesamiento digital de imágenes

Las imágenes son producidas por una variedad de dispositivos físicos como cámaras fotográficas, de video, escáneres, dispositivos de rayos X, microscopios electrónicos, radares y ultrasonidos, y se utilizan para diversos fines como el entretenimiento, la medicina, los negocios, la industria, armada militar, seguridad civil, ciencia experimental entre otras. El interés por las imágenes digitales radica a la mejora de la información pictórica de la interpretación humana y el procesamiento de los datos de la escena para la percepción autónoma de las maquinas. Luego según el diccionario Webster define a la imagen digital como "Una representación numérica de un objeto", con esta definición es posible asimilar el análisis de las imágenes digitales como un proceso que convierte una imagen digital, como un conjunto de datos de medición o una decisión; la digitalización es un proceso que convierte una forma pictórica en datos numéricos; la imagen digital queda definida como $f(x,y)$ que ha sido discretizada tanto en coordenadas espaciales, como en un brillo (intensidad). La imagen se divide en pequeñas regiones denominadas elementos de imagen o pixeles.

Entonces la digitalización incluye el muestreo de la imagen en coordenadas espaciales (x,y), y su posterior cuantificación del nivel de gris, en matrices rectangulares de números enteros donde el tamaño y niveles de gris suelen ser potencias de 2 tal y como se evidencia en la siguiente figura.[21]

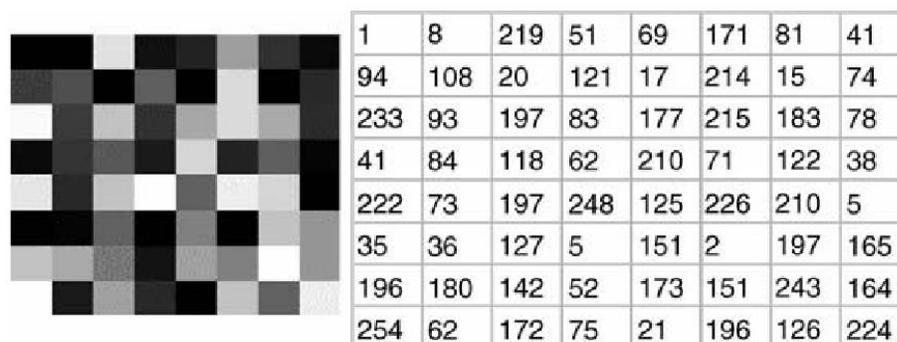


Figura 1. Imagen digital y su representación numérica. (Frank Y.Shih, 2010)

La calidad de una imagen depende en gran parte del número de muestras y de los niveles de gris; cuantos más sean estos dos, mejor será la calidad de una imagen. Sin embargo esto supone una gran cantidad de espacio de almacenamiento que a su vez es el producto de las dimensiones de la imagen y el número de bits necesarios para almacenar los niveles de gris; con una resolución baja, una imagen puede tener un efecto que se asemeja a un tablero de ajedrez o granulado, entonces cuando por ejemplo una imagen de 1024X1024 se reduce a 512X512, puede mostrar mucho deterioro, pero cuando se reduce a 256x256 y luego se vuelve a escalar a 1024x1024 por duplicación, podría mostrar un granulado discernible, dicho esto, es usual que según la aplicación se manejen diferentes calidades de imagen.

Ahora lo elementos de un sistema de procesamiento de imágenes incluyen

Adquisición de imágenes: Un dispositivo físico que es sensible a una banda del espectro de energía electromagnética puede producir una salida de señal eléctrica. Luego se utiliza un digitalizador para convertir la salida de señal eléctrica a una de forma digital; dichas imágenes digitales pueden obtenerse mediante la conversión de las imágenes analógicas (impresiones de 35mm, diapositivas transparencias o arte reflectivo) en imágenes digitales con un escáner, o bien capturando directamente el objeto o la escena en forma digital mediante una cámara digital o un dispositivo de captura de video.

Almacenamiento: Que denota 3 clasificaciones

- Almacenamiento a corto plazo para su uso durante los procesos. Uno de los medios para proporcionar este tipo de almacenamiento es la memoria del ordenador y otro una placa especializada llamada frame buffer
- Almacenamiento en línea para una recuperación relativamente rápida
- Almacenamiento de archivo caracterizado por acceso poco frecuente. El término "calidad de archivo" es usado para denotar materiales que son permanentes y duraderos por lo tanto, pueden utilizarse con seguridad para su conservación. El objetivo del almacenamiento de archivos es la protección contra manipulación, el borrado, los virus y corrupción de los datos.

Procesamiento: Gran parte de las funciones de procesamiento de imágenes pueden implementarse en el software, que se ejecuta en un computador host.

Comunicación: Los servicios de internet por módem de cable estiman en promedio, mayores niveles de ancho de banda que los servicios de internet por DSL(Línea de Abonado Digital, conectividad por espectros de frecuencia), y este ancho de banda traduce en mejor velocidad de comunicación esto debido a que la mayoría de las formas DSL no llegan a los 10Mbps.

Visualización: Los dispositivo de visualización más frecuentes pueden adoptar la forma de una lámpara de imagen luminosa que proporcionan medios por los cuales las imágenes pueden ser iluminadas por una fuente de luz sobre una base seleccionable y removible, algunas variantes de monitores monocromáticos y de color, además de otros medios de visualización como los tubos de rayos catódicos (CTR) de acceso aleatorio y dispositivos de impresión.

Sin embargo en la actualidad los sistemas han tomado enfoques distintos que ayudan entre otras cosas a agilizar y/u optimizar la labor de algunas de las etapas citadas con anterioridad, tal es el caso de aplicaciones como la detección de personas y rostros que por ejemplo pasan por una etapa de adquisición, donde el objeto de interés podría capturarse mediante una cámara digital fija de alta resolución y según la aplicación, comprimirse en un archivo de imagen, una etapa de preprocesamiento donde la imagen adquirida puede mejorarse mediante el cambio de contraste, nitidez, color entre otros, la de segmentación que involucra un posible recorte por áreas de interés (ROI), esto por supuesto para pasar a la etapa de representación y descripción donde las áreas segmentadas se pueden caracterizar por datos estadísticos como componentes principales, texturas, relaciones de aspecto, color entre otros, con el objetivo de que en la etapa de reconocimiento e interpretación de la correspondencia las características derivadas de la segmentación, puedan emparejar áreas segregadas individualmente para aplicar a algoritmos de reconocimiento específicos, para denotar en ocasiones clasificación, descripción del conjunto, vector características entre otros, finalmente la fase de conocimientos toma el vector características que puede utilizarse en una base de conocimientos de todos los objetos conocidos para asociarlos a una base de datos, para retornar, a menudo en números de referencia o puntuación de confianza de dicha lectura.

Introducción al Tracking de objetos

Una vez que se ha realizado un modelado de fondo adecuado con su respectiva detección y eliminación de sombras, la siguiente tarea es el seguimiento del objeto en medio de un "desorden". Se han documentado y reportado muchos enfoques interesantes sobre diferentes algoritmos de seguimiento, además todos en menor o mayor medida han aplicado algunas de las siguientes consideraciones de diseño, que desde luego pueden incorporarse a un sistema de seguimiento (*tracking*) de personas.

Fondo estacionario: Cuando la escena contiene varios objetos, el fondo es inmóvil mientras que parte o todos los objetos en el primer plano pueden encontrarse en movimiento.

Variación en el tamaño del objetivo: El tamaño del objetivo se reduce a medida que el objetivo se aleja de la cámara, razón por la cual es necesario incorporar un mecanismo de escalado durante el proceso de seguimiento.

Modelo del Objetivo: Es necesario incluir el modelo del objetivo al cual se desea aplicar un seguimiento, por ejemplo cuando se trata de una persona, una figura humana puede ser modelada como un conjunto de varias elipses, donde cada una de ellas representa las partes individuales del cuerpo como la cabeza, el torso, las manos y las piernas entre otros. El color, forma, la intensidad y demás atributos del objeto pueden variar mientras este se encuentre en movimiento, sin embargo el rastreador (*tracker*) debe estar en la capacidad de seguirlo correctamente.

Detección automática de objetivos: El rastreador (*tracker*), debe ser capaz de detectar todos los nuevos objetivos de forma automática y empezar a rastrearlos.

Tiempo real: El algoritmo de seguimiento debe ser lo computacionalmente sencillo y óptimo, para que al realizarse el seguimiento se contemple su implementación en tiempo real.

Trayectoria del objetivo: El objetivo puede o no seguir una trayectoria determinada, así mismo pueden haber cambios bruscos en la trayectoria del objetivo.

Velocidad del objetivo: Dicha velocidad también puede cambiar bruscamente como puede ser constante, creciente o decreciente según el caso.

Un sistema de seguimiento (*tracking*) puede modelarse como una máquina secuencial de tres estados, es decir, bloqueo, seguimiento y recuperación. Las funciones de cada estado son las siguientes:

Estado de bloqueo: Inicialmente el sistema se encuentra en estado de bloqueo, y en búsqueda cuando está buscando objetivos. Durante este estado, el procesamiento se lleva a cabo en todo el marco de la imagen. El sistema dividirá el cuadro de la imagen capturado por la cámara en un número de objetos en movimiento, así el historial de estos objetos se extrae comprobando la trayectoria que lleva el objeto u objetos en cuestión; la confirmación de donde se encuentre se lleva en modo automático, una vez confirmado el objetivo del sistema, este se transfiere al estado de seguimiento.

*Estado de seguimiento (*tracking*):* Esta etapa debe utilizar técnicas de bajo coste computacional como se mencionó anteriormente. La ubicación actual extraída por el estado de bloqueo se utiliza para el procesamiento, identificándose en su momento la posición del objetivo, y dicha información posicional se almacenara en la base de datos del historial, entonces si el objetivo no existe en el área de la ventana prevista, el control del sistema se transfiere al estado recuperación.

Estado de recuperación: Es frecuente que el objeto de interés que a su vez se encuentra en movimiento pueda perderse temporal o permanentemente. En estado, si el objetivo se pierde, el sistema intentara recuperar el objetivo a partir de una imagen de baja resolución. Si el objetivo se recupera en pocos fotogramas, el sistema transferirá el control al estado de seguimiento; de lo contrario permanecerá en estado de recuperación hasta que expire el tiempo predefinido, una vez transcurrido este tiempo se transfiere al estado de bloqueo. [22]

La estimación dinámica y el tracking de objetos en una escena a partir de una secuencia de imágenes tienen una amplia gama de aplicaciones, como las de videovigilancia, sistemas de defensa, modelado biomédico y meteorológico etc. De hecho en algunas aplicaciones es común encontrar todo un bloque de operaciones aparte y con respecto a lo expuesto se le denote al mismo como un rastreador o tracker, entonces los métodos descritos van enfocados al mundo real, debido entre otras cosas por el modelado adaptativo del fondo, cuya importancia corresponde a los frecuentes cambios de iluminación de la escena observada. Un modelo de fondo robusto es esencial para eliminar los cambios no deseados en las escenas y su correcta actualización, mitiga las estimaciones erróneas del objeto en movimiento.

En esencia una vez que los píxeles se han clasificado como primer plano y fondo, se debe aplicar un umbral para filtrar el ruido y los puntos aislados, que se detectan falsamente. El etiquetado de componentes conectados produce un conjunto de manchas suaves que corresponden a cada objeto en movimiento. Ahora si solo hay un objeto de interés en un fondo claro y no variable el problema es relativamente sencillo, pero si por el contrario se trata de múltiples objetos con un fondo variable en tiempo real/aproximado al real se convierte en una área de investigación muy desafiante, cuyo desarrollo de un modelo de fondo adaptable y robusto complementara la eficacia del tracking de objetos. Así la investigación sobre la gestión eficaz de la detección de sombras y el problema de la oclusión de objetos puede conducir a la construcción de una solución de seguimiento robusta. [22]

OpenCV

Es una biblioteca de visión computacional de código abierta escrita en lenguaje C y C++ funciona en Linux, Windows y Mac OS X, con un soporte activo para plataformas como Python, Ruby, Matlab entre otros lenguajes y está particularmente diseñado para la eficiencia en aplicaciones de tiempo real. La biblioteca contiene más de 500 funciones que abarcan más que solo la visión como por ejemplo la inspección de productos en fábricas, imágenes médicas, seguridad, interfaces graficas de usuario, calibración de cámaras, visión estereoscópica y robótica. Como es usual que la visión por computador y el aprendizaje automático vayan de la mano OpenCV también contiene biblioteca de aprendizaje automático (MLL) completa y de uso generalizado, que se centra en el reconocimiento de patrones estadísticos y conjuntos de datos; así la MLL es útil en tareas de visión que involucren casi cualquier problema de aprendizaje automático.

Estructura y contenido

OpenCV está estructurado en cinco componentes principales, de los cuales se muestran en la figura 2, contiene los algoritmos básicos de procesamiento de imágenes y de visión por ordenador de alto nivel; *ML* es la biblioteca de aprendizaje, que incluye muchos clasificadores estadísticas y herramientas de agrupación (*clustering*). *HighGUI* contiene rutinas tipo E/S y funciones para almacenar y cargar video e imágenes; *CXCore* contiene estructuras de datos básicas y de contenido. [23]

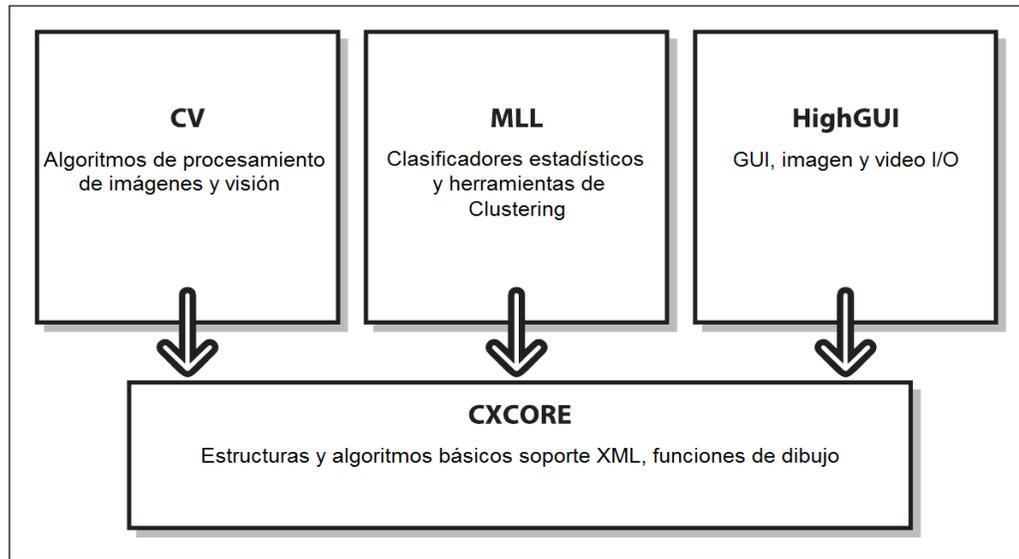


Figura 2. Estructura básica de OpenCV. Modificado de (Bradski y Kaehler, 2008)

La figura no incluye CVAux que contiene errores (reconocimiento facial HMM incrustado), como algoritmos experimentales (segmentación de fondo), que aún no han sido bien documentados razón por la cual tampoco está en el subdirectorio principal de ayuda, no obstante cuenta con lo siguiente

- Objetos Eigen, es en esencia una técnica de reconocimiento computacional eficiente de comparación de plantillas
- Modelos de 1D y 2D Markov ocultos, consiste en una técnica de reconocimiento estadística resuelta mediante programación dinámica
- HMMs incrustados (las observaciones parentales HMM son a su vez HMMs)
- Reconocimiento de gestos del soporte de visión estéreo
- Extensiones a la triangulación de Delaunay, secuencias, etc.
- Visión en estéreo
- Coincidencia de formas acorde a contornos de región
- Descriptores de textura
- Seguimientos de ojos y boca
- Seguimiento 3D
- Esqueletos (líneas centrales) de objetos en una escena
- Deformación de vistas intermedias entre dos vistas de cámara
- Segmentación de fondo-primer plano
- Videovigilancia
- Calibración de cámara por clases C++

DetECCIÓN DE OBJETOS BASADOS EN OPENCV

Aplicados en un número significativo de aplicaciones por visión computacional, la biblioteca proporciona herramientas para el aprendizaje discriminativo y generativo, que se pueden combinar para determinar por ejemplo, si una imagen contiene un objeto en particular y de ser posible su localización en un espacio de píxeles, es por esta y otras razones que es válido aplicar las técnicas de aprendizaje automático (*machine learning*) para evidenciar una mejora o cambio en los resultados.

El uso convencional de las herramientas por aparte suele referenciarse a métodos de bajo nivel, cuando se aplican técnicas específicas aluden a algo mucho más complejo así mismo surgen los métodos de aprendizaje, cuyo fin es detectar objetos de interés en las imágenes. Actualmente existen dos detectores de este tipo en OpenCV el primero es el clasificador en cascada (*cascade classifier*), que generaliza aquel implementado por Viola y Jones para la detección de rostros, y el segundo es el de cascada suave (*soft cascade*), una evolución de ese algoritmo que utiliza un nuevo enfoque, para brindar lo que en la mayoría de los casos representa una clasificación más robusta que aquel que posee el clasificador de cascada; ambos se han utilizado con éxito en muchas clases de objetos además de los rostros, de tal manera, que generalmente aquellos objetos que caractericen por una estructura excelente y rígida suelen responder bien a estos métodos.

Estos métodos no solo comprenden los componentes de aprendizaje automático en los que se basan, sino que también implican otras etapas que condicionan la entrada para el aprendizaje o post-proceso en la salida del algoritmo de aprendizaje. Luego no es de extrañar que estos algoritmos de detección de objetos no tengan una interfaz tan uniforme como aquellos algoritmos de núcleo de aprendizaje (*core machine learning*), por esta razón suscita la cantidad y variedad natural de las necesidades y resultados en la metodología de alto nivel, luego en la práctica estos algoritmos son soportados por los mismos usuarios, quienes proporcionan y poseen interfaces más parecidas a las implementaciones originales. [24]

Clasificadores en cascada

Los componentes técnicas basadas en árboles conocidos con el nombre de clasificador en cascada y se basa en el concepto de la cascada de rechazo potenciada (*boosted rejection cascade*). Tiene un formato diferente al de la mayor parte de la librería ML (*Machine Learning*) en OpenCV, por que fue inicialmente desarrollada como una aplicación completa de detección de rostros, para que después de una pequeña modernización se haga un poco más general que la implementación original.

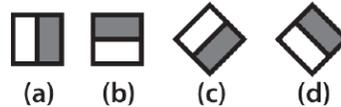
Sin embargo, la detección de caras es una necesidad tan común que valía la pena tener una técnica de referencia que funcionara bien, como dicha técnica se basaba en el conocido campo del boosting estadístico, en realidad tenía una gama más amplia de aplicaciones; desde entonces varias organizaciones han diseñado el detector de "caras" en OpenCV para detectar objetos "mayormente rígidos" (coches, bicicletas, partes del cuerpo, personas...), entrenando nuevos detectores en miles de conjuntos de imágenes de entrenamiento cuidadosamente seleccionados

para cada vista del objeto u objetos de interés. Esta técnica se ha utilizado para crear detectores de última generación, aunque con un detector diferente en cada caso, por lo tanto es una valiosa herramienta a destacar en tareas de reconocimiento en general, actualmente la biblioteca recibe soporte a niveles que le permiten abarcar aún más implementaciones a nivel mundial.

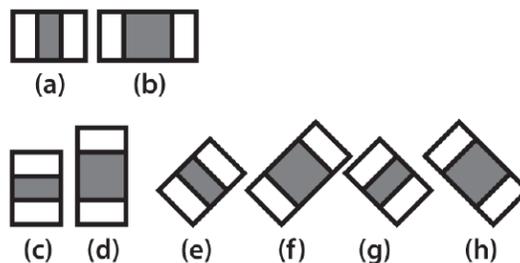
Como se mencionó anteriormente el clasificador en cascada implementa una versión de la técnica de detección de rostros desarrollada por Paul Viola y Michael Jones, originalmente esta técnica y su implementación solo soportaba un conjunto particular de características, las wavelets Haar. Esta técnica fue ampliada posteriormente por Rainer Lienhart y Jochen Maydt, para utilizar lo que se conoce como características diagonales, que fueron incorporadas a OpenCV así este conjunto ampliado se conoce comúnmente como "características tipo Haar"; en versiones 3.x de OpenCV las cascadas se amplían para trabajar en conjunto con patrones binarios locales o LBP.

Tal y como sucede con el algoritmo principal el detector de OpenCV, este funciona en dos capas la primera es el detector de características que encapsula y modula el computo de dichos datos, y la segunda capa corresponde a la cascada real reforzada, que utiliza sumas y diferencias sobre regiones rectangulares de las características computadas. En la siguiente figura se hace mención a las características que trae por defecto el clasificador, de varias maneras estas constituyen la "materia prima" que utilizaran los clasificadores potenciados, y todas ellas tienen esa propiedad que les permite calcularse rápidamente a partir de una imagen integral tomada de la imagen original en escala de grises.

1. Características de los bordes



2. Características de línea



3. Características del centro-fondo

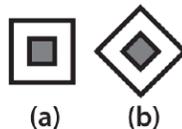


Figura 3. En esta representación esquemática de los wavelets, la región clara se interpreta como "añadir esa zona" y la región oscura como "restar esa zona"(Bradski y Kaehler, 2017)

Actualmente hay dos conjuntos de características admitidas, esto por supuesto incluyen las "originales" de los wavelets Haar (incluidas las características diagonales) y la alternativa LBP, y no se descarta que en un futuro otros tipos de propiedades se puedan escribir o reutilizar de otras características (ya que la interfaz en cascada es totalmente extensible), siguiendo la forma en la que se añadió LBP.

Principio de funcionamiento del clasificador Haar

El clasificador se entrena para el reconocimiento de muestras, tanto positivas como negativas y así básicamente reconoce la forma del objeto que se desea. Tal y como se mencionó este ha de pasar por diferentes etapas que son aplicadas secuencialmente con un proceso de optimización. La primera versión donde se utilizó este proceso o técnica tomo el nombre de "*Haar wavelets*" propuestas por el profesor Papageorgiou para la detección de caras y peatones, sin embargo al principio no fue muy usado. Entonces Viola y Jones adaptaron la idea desarrollando lo que vendría a ser las denominadas características Haar (Haar-like Features), donde básicamente se identifican aquellos rasgos que definen el objeto en base a la estructura de los niveles de intensidad que presentan sus píxeles en una imagen. Esta información será extraída aplicando sobre la imagen del objeto de interés una serie de funciones ligadas a las denominadas características Haar.

En resumen el clasificador puede verse como la ventana de píxeles de tamaño y orientación variable, dividida en regiones rectangulares, pudiendo ser cada una de estas regiones de dos tipos diferentes (positiva o negativa), y las características Haar en particular se definen sobre la misma en un escalado de grises; razón por la cual estas se encuentran formadas por un número finito de rectángulos y su valor escalar consistirá en la suma de los píxeles de cada rectángulo aplicando un factor de peso. Luego la siguiente fórmula suele ser utilizada como el valor de una característica al menos en la base del algoritmo para la detección de rostros:

$$\sum_{In} I(x,y) - \sum_{Ib} I(x,y) \quad (1)$$

Donde el resultado es la suma de intensidad de píxeles situadas en la ventana de color negro menos la suma de las intensidades de píxeles de la ventana de color blanco. Esto aporta un rasgo importante del método Haar, debido a que lo define el tamaño, la orientación o la distribución de las regiones positivas y negativas, pudiéndose construir infinidad de patrones. Entonces estos parámetros van a depender de las características del objeto a detectar, y más en concreto de la distribución de intensidades de los píxeles que componen dicho rasgo. En consecuencia, el objetivo a la hora de construir una característica Haar es buscar que su estructura se asemeje a la del elemento a detectar tal y como sucede con los bordes, líneas o contornos. [25]

Entrenamiento y detectores pre entrenados

OpenCv incluye un conjunto de archivos de reconocimiento de objetos pre entrenados, sin embargo también existe código abierto que permite entrenar y almacenar nuevos modelos de objetos para el detector, así mismo existe una aplicación en el paquete traincascade que puede utilizarse para entrenar detectores de casi cualquier objeto, aunque su fuerte varié sustancialmente de un objeto a otro. Los objetos pre entrenados que vienen con OpenCV se encuentran en los directorios `../opencv/data/haarcascades` y `../opencv/data/lbpcascades`; en ocasiones las vistas laterales son más difíciles de detectar con precisión, luego estos modelos han tenido que ser mejorados; entre otras cosas gracias al trabajo realizado durante el Google Summer of Code 2012.

Una de las razones por las cuales las vistas de perfil son complicadas para el clasificador en cascada Haar, es porque este utiliza características en bloque, por lo tanto se ve obligado a intentar aprenderla variabilidad del fondo que "asoma" a través del borde informativo del perfil de la vista lateral de los objetos. Este problema ocurre de forma generalizada y afecta tanto las características Haar como en menor medida a las LBP, entonces en la mayoría de los casos lo mejor es entrenar al detector para que encuentre una región cuadrada que sea un subconjunto del objeto que se busca, y es de esta manera como las cascadas pre entrenadas tratan las formas redondas; buscan aquella región cuadrada que está circunscrita por una cara, por ejemplo, que suelen ser las cascadas Haar mayormente implementadas por su mejorado funcionamiento. [24]

La media como algoritmo de sustracción de fondo

La sustracción de fondo es un método que se utiliza para detectar un objeto en el primer plano de la imagen, y de los usos que comúnmente se le dan incluyen: el seguimiento de objetos, el seguimiento de objetivos, análisis de tráfico y demás aplicaciones de video; a menudo la detección no solo comprueba la existencia de un objeto en el video sino que también lo localiza, debido a que durante dicha secuencia de video los cambios espaciales y temporales incluyen otro tipo de información como lo es el tamaño, forma, color, entre otros.

Su principal aplicación radica en implementaciones de seguridad, debido a su sencillez, precisión y tiempo computacional, sin embargo al igual que los otros métodos presenta sus retos en cuanto a las limitaciones del sistema donde se desarrolla y de los cambios ambientales como iluminación, similitud de colores, sombras etc. Así que básicamente el algoritmo compara dos fotogramas diferentes, opera la diferencia y crea una matriz de distancia, a su vez realiza lo mismo con el valor del umbral que no está predefinido, es el usuario quien lo establece, para que después el escenario principal compare de la siguiente manera (figura 4), "entre mayor sea un umbral traducirá, como un objeto en movimiento de lo contrario, este se tomara como una imagen de fondo o como parte de ella". [26]

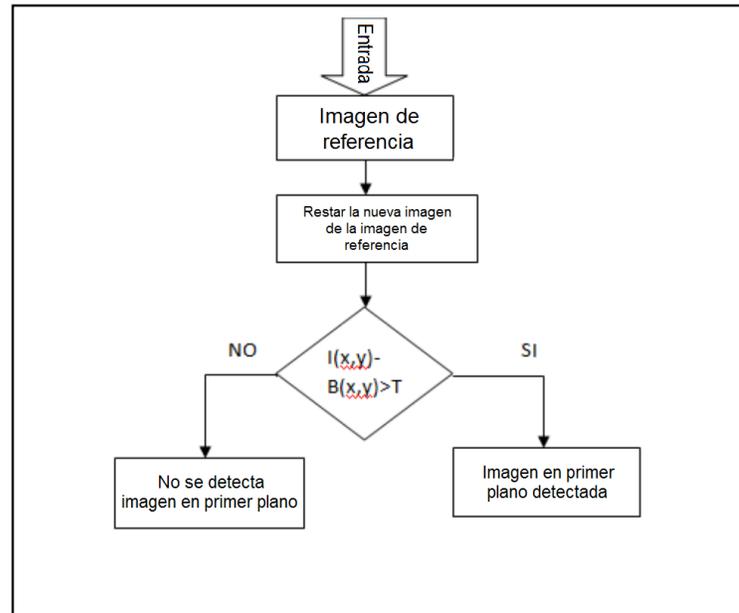


Figura 4. Diagrama de flujo del método de sustracción de fondo. Modificado de (Desai y Gandhi, 2014)

Existen muchos métodos de sustracción de fondo, como la diferencia de fotogramas, el modelo de mezcla gaussiana, la estimación de la densidad del núcleo..., todos dan una precisión diferente acorde a la aplicación; luego en este apartado se especificara un poco más aquel desarrollado en la presente propuesta. La mayoría de las técnicas BS (sustracción de fondo <<US>>) comparten un denominador común: hacen la suposición de que la secuencia I se compone de un fondo estático B delante del cual se observan objetos en movimiento, con la suposición claro esa de que cada objeto tiene un color o la distribución del mismo diferente del observado en B , numerosos métodos permiten resumirse con la siguiente formula.

$$X_t(s) = 1 \text{ si } d(I_{s,t}, B_s) > \tau, 0 \text{ En otro caso} \quad (2)$$

Donde T es un umbral, $X_{(t)}$ es el campo de etiquetas de movimiento en el momento t (llamado también máscara de movimiento), d es la distancia $I_{s,t}$, el color en el momento t y el pixel s , B_s , el modelo de fondo es la principal diferencia entre los distintos métodos de BS como es aquel modelo que modela B y que métrica de distancia d utilicen. La forma más sencilla de detección básica de movimiento en el fondo B , es mediante una única imagen en escala de grises/color sin objetivos en movimiento, dicha imagen puede ser una fotografía tomada en ausencia de movimiento o estimada mediante un filtro de mediana temporal. Para manejar los cambios de brillo y la modificación del fondo se utiliza la siguiente expresión:

$$B_{st+1} = (1 - \alpha) B_{st} + \alpha \cdot I_{st} \quad (3)$$

Donde α es una constante cuyo valor oscila entre 0 y 1, así con este sencillo modelo de fondo los pixeles que corresponden a objetos en movimiento de primer plano, pueden detectarse mediante el umbral de cualquiera de las funciones de distancia:

$$\begin{aligned}
 d_0 &= |I_{st} - B_{st}| \\
 d_1 &= |I_{st}^R - B_{st}^R| + |I_{st}^G - B_{st}^G| + |I_{st}^B - B_{st}^B| \\
 d_2 &= (I_{st}^R - B_{st}^R)^2 + (I_{st}^G - B_{st}^G)^2 + (I_{st}^B - B_{st}^B)^2 \\
 d_\infty &= \max\{|I_{st}^R - B_{st}^R|, |I_{st}^G - B_{st}^G|, |I_{st}^B - B_{st}^B|\}
 \end{aligned}
 \tag{4}$$

Donde R , G , y B son los canales de color en la imagen (rojo, verde y azul) y $d0$ es una medida que opera en imágenes en escala de grises, cabe destacar que es posible utilizar el fotograma anterior I_{t-1} como imagen de fondo, sin embargo con esta configuración, la detección de movimiento se convierte en un proceso de detección de cambios entre fotogramas que se vuelve robusto en términos de cambios de iluminación pero sufre un grave problema de apertura, ya que solo se detectan partes de los objetos en movimiento. [26]

Posición relativa de dos circunferencias como integración de los algoritmos de detección de persona y sustracción de fondo por media

Para comprender de una manera más ágil la interpretación de la presente subsección, es necesario aproximar los conceptos, no obstante, la recomendación es siempre acudir al apartado de metodología para complementar lo aquí consignado:

Sean dos circunferencias que tengan como incógnita su posición de una con respecto a la otra, en el plano del espacio vectorial R^2 , es posible exponer 6 posiciones relativas generalizando las características que puedan tomar para cada caso: [27]

Exteriores: Si no tienen ningún punto en común (coordenadas x,y en R^2 por ejemplo) y la distancia entre sus centros Q_1 y Q_2 en la figura 5 es mayor que la suma de sus radios.

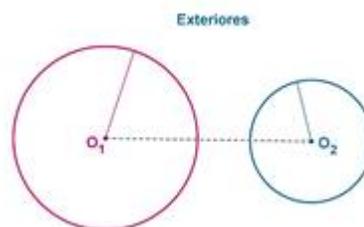


Figura 5. Posición relativa de dos circunferencias cuando son exteriores (Schulz, UNAM)

Tangentes exteriores: Tienen un punto en común y la distancia entre sus centros es igual a la suma de sus radios

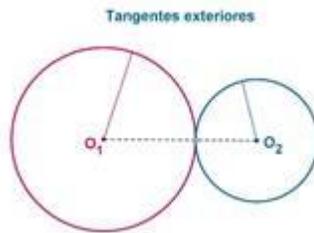


Figura 6. Posición relativa de dos circunferencias cuando son tangentes exteriores (Schulz, UNAM)

Secantes: Tienen dos puntos en común y la distancia entre sus centros es menor que la suma de sus radios y a su vez es mayor que su diferencia

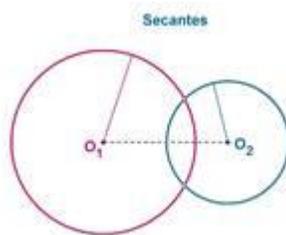


Figura 7. Posición relativa de dos circunferencias cuando son secantes (Schulz, UNAM)

Tangentes interiores: Tienen un punto en común y la distancia entre sus centros es igual que la diferencia de sus radios.

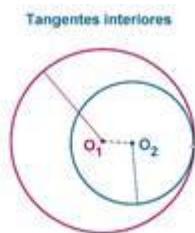


Figura 8. Posición relativa de dos circunferencias cuando son tangentes interiores (Schulz, UNAM)

Interiores: No tienen ningún punto en común y la distancia entre sus centros es menor que la diferencia de sus radios.

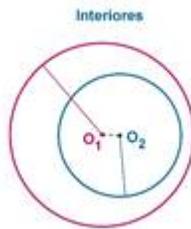


Figura 9. Posición relativa de dos circunferencias cuando son interiores (Schulz, UNAM)

Interiores concéntricas: No tienen puntos en común y la distancia entre sus centros es cero (coinciden)

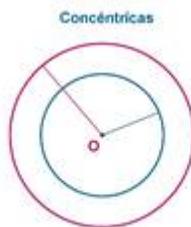


Figura 10. Posición relativa de dos circunferencias cuando son concéntricas (Schulz, UNAM)

Conceptos generales de la comunicación serial

La comunicación entre dispositivos electrónicos consiste en interconectar circuitos (procesadores u otros circuitos integrados), para crear un sistema simbiótico así para que esos circuitos individuales intercambien información, deben a su vez compartir un protocolo de comunicación estándar en común, entre los variados protocolos utilizados habitualmente se encuentran RS232, RS485, RS422, USB y Ethernet. Pero dado que USB y Ethernet requieren sistemas de interfaces bastante potentes y complejas, muchos dispositivos eficientes aun utilizan RS232, RS485, RS422, ahora para interiorizar un poco en las características generales de los protocolos serie, la siguiente tabla es referente a esa información.

Tabla de comparación de protocolos

	RS-232	RS-422	RS-485
Cable	Un solo extremo	Multipunto Simple	Multidireccional
Numero de dispositivos	1 transmisor 1 receptor	1 transmisor 10 receptores	32 transmisores 32 receptores
Modo de Comunicación	Full duplex	Full duplex, Half duplex	Full duplex, Half duplex
Distancia máxima	50 pies a 19.2 kbps	4000 pies a 100 kbps	4000 pies a 100 kbps
Velocidad de datos máxima 50ft	1 mbps	10 mbps	10 mbps

Tabla 1. Tabla de protocolos serial. Fuente: modificada de (Sonnenberg J.)

Cabe aclarar que los protocolos serie RS232, RS485, RS422 solo se refieren a la interfaz de hardware, no a los protocolos de software utilizados para hacer que los dispositivos se comuniquen. Hay muchos protocolos disponibles en el mercado, por lo que no se puede asumir la interoperabilidad entre diferentes fabricantes de puertos "RS232". El bus serie RS232 consta de sólo dos hilos de comunicación: uno es utilizado para enviar datos y el otro para recibirlos, es posible inferir como tal que debe contener dos pines, un receptor RX y el transmisor TX, producto de esto los datos de comunicación se basan en la tensión de los cables en relación con el nivel de tierra, entonces debe realizarse una conexión a tierra precisamente entre los dos dispositivos que se comunican con RS232 tal y como se evidencia en el siguiente diagrama.

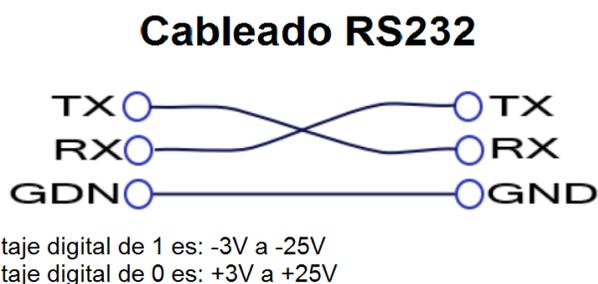


Figura 11. Diagrama de cableado para RS232. Modificado de (Sonnenberg J.)

Los datos de RS232 se envían en serie, esto también quiere decir que cada bit viaja uno tras otro porque sólo hay una línea de datos en cada dirección. Este modo de transmisión de datos también requiere que el receptor sepa cuando llegan los bits de datos reales para que pueda sincronizarse con los datos entrantes, para dicho fin se envía un 0 lógico como bit de inicio de sincronización. [28]

Protocolo RTSP (Real Time Streaming Protocol)

Es un protocolo de flujo de datos en tiempo real no orientado a conexión utilizado para definir cómo se hará el envío de información entre el cliente y el servidor. Trabaja a nivel de aplicación, pues el tipo de contenido con el que trabaja para realizar el streaming es muy sensible a la sincronía temporal (o la falta de ella). Así pues se podría considerar que el RTSP actúa como si de una especie de mando a distancia de red para servidores multimedia se tratase. [29]

Las peticiones que maneja RTSP son las siguientes:

- **OPTIONS:** Esta petición puede ser enviada en cualquier instante y no influye en el estado de la sesión. Es utilizada en el inicio, para establecer la sesión con los parámetros necesarios.
- **DESCRIBE:** Se encarga de inicializar la sesión. Obteniendo la descripción de un objeto requerido en base el flujo necesario para la correcta reproducción

- SETUP: Especificará los parámetros de transporte tales como el protocolo a utilizar, el puerto de recepción o el puerto para los paquetes RTCP.
- PLAY: Indica al servidor cuando debe comenzar el envío de datos, con los parámetros especificados en el SETUP.
- PAUSE: Detiene temporalmente el flujo de datos.
- TEARDOWN: Petición para detener completamente el flujo de datos y liberar los recursos utilizados
- GET_PARAMETER: recupera el valor de algún parámetro de la presentación en curso.
- SET_PARAMETER: Análogamente a la anterior petición, sirve para establecer el valor de un parámetro de la presentación en curso.
- REDIRECT: Petición que informa al cliente de que debe conectar a otra dirección de servidor.

Contras

Los principales problemas están basados en la necesidad de descargar el contenido entero a la memoria cache del cliente

- La cantidad de memoria que se tiene que reservar en el cliente para el búfer tiene que ser muy parecida o igual a la del contenido a reproducir
- Para asegurar una reproducción sin cortes o ausencias de la señal se necesita un tiempo de predescarga que será directamente proporcional al contenido a reproducir
- No se puede adelantar la transmisión hasta un punto de interés hasta que no se haya descargado al buffer, todo el contenido anterior a ese punto.

Pros

- Utilizar esta tecnología no implica ningún soporte especial del servidor, como único requisito es que el cliente tenga instalado el plug-in adecuado en su navegador web y/o plataformas de desarrollo
- Las "descargas progresivas" se transmiten utilizando el protocolo HTTP que garantiza el correcto envío de todos los paquetes y por tanto, asegura la retransmisión del video entero sean cuales sean las condiciones de la red
- Por usar trafico web estándar se "salta" el firewall (si existiese), lo que ocasionaría un bloqueo en el streaming
- A mayor o menor velocidad la descarga tiene que llegar a su término, luego es prácticamente seguro que todos los clientes conectados disfrutaran de la misma calidad de reproducción independientemente de la velocidad de su conexión a internet, esto además permite utilizar algoritmos de compresión con mayor calidad, como sucede con los que se podrían implementar en un dispositivo de grabación digital DVR (Digital Video Recorder).

Tipos de estrategias en control y su rol en los procesos

Una pregunta que surge de manera natural cuando se abordan los procesos industriales o no es: ¿cuándo es necesario aplicar el control o sus estrategias?, para responder dicha pregunta existen

dos grandes razones. La primera razón para implementar, ejercer una labor de control, es mantener la variable proceso en su valor deseado aunque se produzcan perturbaciones en el sistema. Algunas perturbaciones típicas en un tanque de uso industrial pueden ser: el caudal y temperatura del fluido de entrada, la temperatura y presión del fluido de calentamiento (previo a sistemas completos de válvulas), así es posible determinar cómo debe ajustarse cada una (cierre o apertura según el proceso), como en repuesta a un cambio de cada una de las variables que responden a la perturbación.

La segunda razón para el control es responder a los cambios en el valor deseado, que a su vez se basan en un análisis exhaustivo del funcionamiento y de los objetivos de la planta a desarrollar, así los problemas que enfrentan los sistemas de control pueden contemplar las siguientes categorías: [30]

- Seguridad
- Protección del medio ambiente
- Protección de los equipos
- Buen funcionamiento de la planta y ritmo de producción
- Calidad del producto
- Optimización del profit
- Monitorización y diagnóstico

Esto traduce en valores de variables, que hay que controlar y que requieren de un conocimiento exhaustivo de la planta y proceso a implementar. Ahora bien una vez establecidas unas pautas básicas para la aplicación de un control, así mismo se originan unas estrategias que participan activa y naturalmente en cada proceso, sin embargo para llegar a ese punto primero debe definirse formalmente una estrategia de control; *"Una estrategia de control es un conjunto planificado de controles, derivado de la comprensión actual del producto y proceso, que asegura el rendimiento del proceso y calidad del producto"* FDA de EE. UU., ICH Q10, entre las que la normativa considera, es una estrategia fundamental de control la siguiente: [31]

La esencia del control de realimentación

Tal y como lo expresan (Pérez Mario .A, Hidalgo Pérez A. y Berenguer Pérez E., 2007) la realimentación *"es una propiedad de los sistemas que permiten que la salida del sistema o cualquier variable del mismo sea comparada con la entrada al sistema o con cualquier componente del sistema, de tal manera que pueda establecerse la acción de control apropiada entre la entrada y la salida."* Luego las componentes clave de estos sistemas son: la variable de proceso (PV) que ha de desviarse a su valor objetivo, la consigna (SP), para que se produzca cualquier acción de control.

- La diferencia entre PV y SP se denomina error (E)
- El signo de error depende de la dirección en la que se tenga que mover algún elemento cuando los niveles "superen" o "no alcancen" el valor establecido `+

$$E = PV - SP \quad \text{ó} \quad E = SP - PV \quad (5)$$

La variable controlada (CV), en algunos casos corresponde a la posición de los actuadores como en el caso de una válvula donde es directamente proporcional al error.

Entonces básicamente se dice que la realimentación existe en un sistema cuando hay una secuencia de relaciones causa-efecto entre las variables de un sistema, que se vuelve fácilmente identificable por el afán de controlar su existencia y su función. Existen dos tipos de realimentación, dada la forma en cómo se comparan las variables.

Realimentación positiva: cuando ambas variables comparadas son de igual signo

Realimentación negativa: cuando ambas variables comparadas son de signo contrario

Ahora bien dos términos que tienden a confundirse en los sistemas de control son el control de realimentación y el sistema de control realimentado entonces: el control de realimentación corresponde a la operación, que en presencia de perturbaciones reduce las diferencias entre la salida y la entrada del sistema, esto lo realiza sobre la base de dicha diferencia (la señal de error). Por su parte el sistema de control realimentado es aquel que mantiene la relación preestablecida entre la salida y la entrada de referencia, comparando ambas y utilizando la diferencia como variable de control.

Dicho esto no es complejo inferir que lo desarrollado corresponde a un sistema de control de lazo cerrado, por tanto la salida o la señal controlada debe ser realimentada y comparada con la entrada de referencia, para posteriormente enviar una señal actuante o acción de control proporcional a la diferencia entre la entrada y salida a través del sistema, para disminuir el error y corregir la salida. En otras palabras el termino lazo cerrado implica el uso de realimentación negativa para reducir el error del sistema, luego una representación típica de la relación entrada y salida de un sistema de control de lazo cerrado puede verse en la siguiente figura. [32]

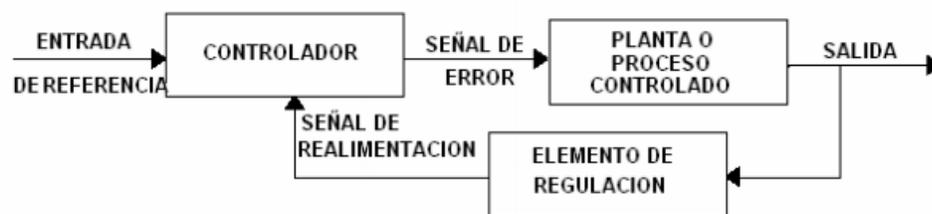


Figura 12. Representación de un sistema de control por lazo cerrado. Tomado de (Pérez Mario A., Hidalgo Pérez A. y Berenguer Pérez E., 2007)

Los rasgos más importantes en la implementación de este tipo de sistemas se encuentran:

- Aumento en la exactitud
- Reducción de los efectos de no linealidad y distorsión
- Aumenta el ancho de banda del sistema

- Disminuye la ganancia del sistema
- El sistema tiene a ser menos estable

El microcontrolador ESP8266

El ESP8266 es un microcontrolador RISC de 32 bits de bajo coste con conectividad Wi-Fi, capaz de funcionar a 80 MHz o 160 MHz, dispone de 64Kb de RAM de instrucciones y 96 Kb de RAM de datos. Para el almacenamiento de firmware y otro tipo de datos, se le conecta una flash QSPI externa cuyo tamaño puede variar entre 512 KB y 4MB el chip expone varios pines de entrada/salida de propósito general (GPIO), pero algunos de ellos se utilizan para la conexión de la flash QSPI. El resto de los pines son capaces de realizar la interfaz periférica SPI, I2C, I2S, recepción/transmisión asíncrona universal (UART) y un convertidor Analógico-Digital (ADC) de 10 bits. En la siguiente figura se encuentra el diagrama de bloques para el ESP8266. [33]

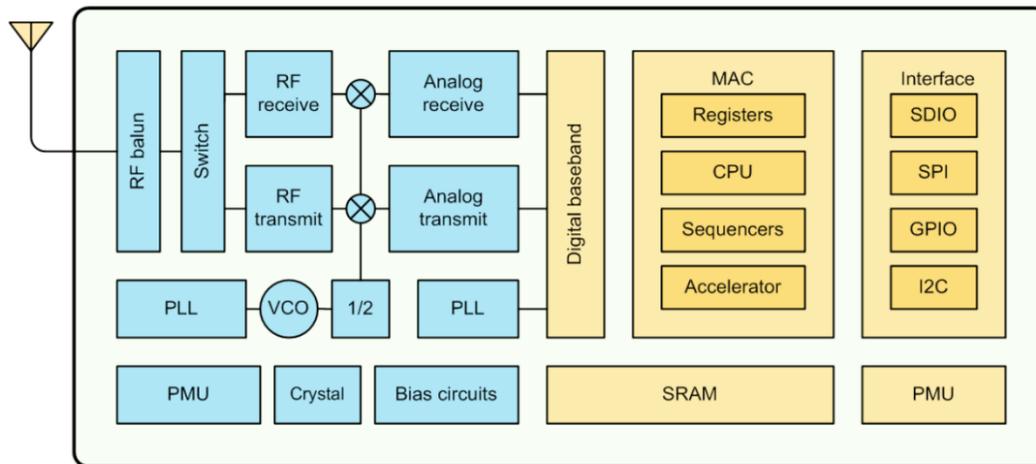


Figura 13. Diagrama de bloques para el ESP8266. Tomado de (Espressif Systems ESP8266 datasheet)

La capacidad Wi-Fi está acorde a la norma IEEE 80.11 b/g/n y proporciona autenticación WPA/WPA2 y WEP, pero también puede conectarse a redes abiertas. La forma o escogencia de la placa depende de los alcances o limitaciones de cada proyecto, en el sistema se utilizó el kit de desarrollo NodeMCU como el que se ve en la figura 14 que es similar a la Olimex, pero también tiene un convertidor USB a serie integrado, así como su propia fuente de alimentación. Es fácil de usar y puede programarse con los SDKs oficiales de Espressif que contienen los SDKs NonOS y FreeRTOS escribiendo código en C/C++, pero otras empresas y personas añaden otros lenguajes de programación, como Lua, Javascript o Micropython, aunque es habitual utilizar el SDK de NonOS y como IDE de desarrollo el Arduino IDE. [33][35]



Figura 14. Módulo NodeMCU WI-FI certificado FCC, con antena PCB. Sugerido en (Schwartz M., 2016)

Los principales campos de aplicaciones que tiene el ESP8266 son:

- Electrodomésticos
- Automatización del hogar
- Enchufe inteligente y luminarias
- Redes de malla
- Control inalámbrico industrial
- Monitores para bebés
- Cámaras IP
- Redes de sensores
- Electrónica cableada
- Dispositivos con reconocimiento de ubicación WiFi
- Etiquetas de identificación de seguridad
- Balizas del sistema de posición WiFi

4.3. Marco Legal

Normativa y conceptos en el tratamiento de imágenes y videovigilancia en general

Según la superintendencia Colombiana de industria y comercio el tratamiento de datos de personas se define como "cualquier operación o conjunto de operaciones sobre datos personales, tales como la recolección almacenamiento, uso, circulación o supresión". En el caso de las imágenes de personas determinadas o determinables, operaciones como la captación, grabación, transmisión, almacenamiento, conservación o reproducción en tiempo real o posterior entre otras, son consideradas como tratamiento de datos personales, y en consecuencia se encuentran sujetas al Régimen General de Protección de Datos Personales.

Dado el artículo 17 emitido por la superintendencia el responsable del tratamiento, es una persona natural o jurídica que decide sobre la base de datos y/o tratamiento de datos, razón por la cual la ley ha dispuesto una serie de obligaciones a su cargo, en particular, para garantizar la protección de los derechos de los titulares de información respecto de la recolección, almacenamiento, uso y disposición de sus datos personales, en este caso de su imagen. [36]

Luego el encargado del tratamiento es un tercero (persona natural o jurídica ajena al responsable que trata los datos personales por propia cuenta). En sistemas de videovigilancia por ejemplo se realiza por medio de empresas de vigilancia y seguridad privada que hacen uso, entre otros, de medios tecnológicos para la prestación de su servicio, estos terceros están en calidad de tratamiento-cliente en casos como copropiedades, establecimientos comerciales, empresas, instituciones educativas etc. Que serán obviamente responsables del mismo, sin que el responsable del tratamiento se desligue de sus obligaciones al garantizar el manejo de datos personales.

Dicho esto los sistemas de videovigilancia (SV) son considerados como intrusivos de la privacidad al involucrar herramientas como el monitoreo y la observación de las actividades que realizan las personas a lo largo del día. Por estas razones, antes de tomar la decisión de implementar SV, se debe tener en cuenta la necesidad de utilizarlos y así determinar si suple con la implementación o si existen otros mecanismos que se puedan utilizar además generen el menor impacto en la privacidad de las personas.

Concepto No 13-102526 del 2013

Recapitulando aparece el principio de finalidad, que rige por supuesto el tratamiento de datos personales, el cual impone que el fin perseguido con dicho tratamiento sea legítimo en relación con la constitución y la presente ley, expreso con objetivos específicos, explícitos e informados, así la normativa vigente permite los desarrollos jurisprudenciales y los estándares internacionales en la materia, entonces el principio de finalidad implicaría: [36]

- Un ámbito temporal: es decir que el periodo de conservación de los datos personales no exceda del necesario para alcanzar la finalidad para la cual se ha recolectado
- Un ámbito material: que exige que los datos recaudados sean solo los necesarios para cumplir las finalidades perseguidas, lo que implica que los mismos se limitan a los que resulten adecuados, pertinentes y acordes con las finalidades para las cuales fueron recolectados

Ley 1581 del 2012 y principio de confidencialidad

Consigna que el SV y la base de datos conformada deben contar con las medidas técnicas, humanas y administrativas necesarias para garantizar la seguridad de los datos personales, evitando su adulteración, pérdida, deterioro, consulta, uso o acceso no autorizado o fraudulento, mantener la integridad de la información y garantizar los derechos de los titulares de los datos.

Algunas de las medidas preventivas y de seguridad a implementar pueden ser limitar el acceso a la información, cifrarla y realizar auditorías periódicas a las medidas adoptadas. En los casos en

que el tratamiento se limite a la reproducción o emisión de imágenes en tiempo real, la visualización debe limitarse solamente al personal autorizado para ello, así las imágenes pueden ser grabadas y posteriormente visualizadas en un área de acceso restringido que garantice la seguridad de las mismas.

De igual forma, se deben establecer medidas sobre la confidencialidad y reserva de datos personales y exigir su cumplimiento a todas aquellas personas naturales o jurídicas que tengan acceso y hagan tratamiento de las imágenes recolectadas y dicha obligación debe atenderse incluso después de finalizada la relación laboral o contractual de la que se derivó el tratamiento de datos personales.

Cámaras con acceso a vía pública

Según lo cita la superintendencia de industria y comercio una de las finalidades más comunes de los SV es garantizar la seguridad de bienes o personas en entornos públicos y privados, en ocasiones puede implicar la toma de imágenes necesaria para alcanzar la finalidad establecida, por ejemplo garantizar la seguridad de bienes y personas en entornos privados. De ser así dicha recolección debe limitarse a aquella que no pueda evitarse para alcanzar el fin legítimo perseguido. El caso más común es la toma de imágenes en puntos de acceso, como porterías y entradas a propiedades privadas, la cual se debe restringir a las áreas más próximas al espacio objeto de vigilancia. [36]

Normativa internacional específico a la proporcionalidad con uso de cámaras en sistemas de seguridad

El RGPD recoge en su artículo 5, en virtud en cual los datos personales serán recogidos con fines determinados, explícitos y legítimos sin tratarlos ulteriormente de manera incompatible con dichos fines, de manera que los datos que sean objeto de tratamiento a través de la videovigilancia motive su instalación velado por la seguridad de las personas, bienes e instalaciones.

Como regla general la agencia española de protección de datos, expresa la captación de imágenes con fines de seguridad de la vía pública debe realizarse por las fuerzas y cuerpos de seguridad. Ya que les corresponde la prevención de hechos delictivos y la garantía de la seguridad en la vía pública en conformidad por la ley orgánica 4/1997 de 4 de agosto y su respectivo reglamento de desarrollo, sin embargo sobre la normativa es posible aplicar excepciones cuando: [37]

- En algunas ocasiones para la protección de espacios privados, donde se hayan instalado cámaras en fachadas o en el interior, puede ser necesario para garantizar la finalidad de seguridad la grabación de una porción de la vía pública, es decir, las cámaras y videocámaras instaladas con fines de seguridad no podrán obtener imágenes de la vía pública salvo que resulte imprescindible para dicho fin, o resulte imposible evitarlo por la razón de la ubicación de aquellas. Por lo tanto las cámaras podrían captar la mínima parte necesaria para la finalidad de seguridad que se pretende.

- Sera posible la captación de la vía publica en una extensión superior cuando fuese necesario para garantizar la seguridad de bienes, instalaciones estratégicas o de infraestructuras vinculadas al transporte.

- Gran parte de la actividad de los ciudadanos se desarrolla en espacios que admiten el acceso al público en general, como centros comerciales, restaurantes, lugares de ocio o aparcamientos. Referido a lugares a los que los ciudadanos pueden tener libre acceso aunque sean de propiedad privada, en los cuales los titulares utilizan los sistemas de videovigilancia para garantizar la integridad de las personas e instalaciones

5. Metodología

El siguiente apartado consigna los métodos en conjunto con el análisis, estándares, técnicas y demás, que llevaron al sistema a su implementación, pruebas y puesta en marcha. Para ello se expondrá en forma detallada el diseño y descripción del sistema en conjunto con las actividades que van en línea con un cronograma preestablecido.

Diseño y descripción del sistema

Una forma general de describir el sistema como un conjunto de actividades y prácticas teórico/experimentales, puede evidenciarse en el siguiente diagrama de bloques, donde cada uno de ellos requirió un proceso de diseño e implementación, que será explicado etapa a etapa junto a los resultados más representativos en el marco de cómo se desarrolló en las actividades propuestas.

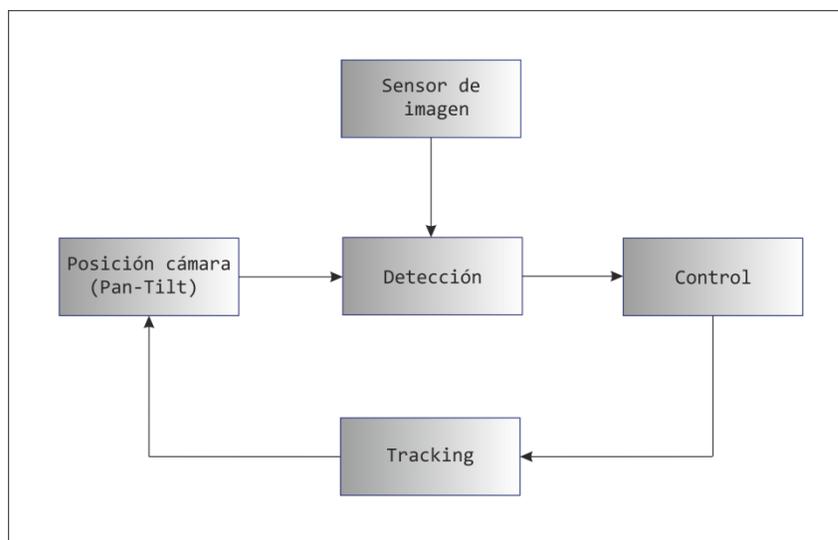


Figura 15. Diagrama de bloques simplificado del sistema. Fuente (imagen propia)

Sensor de imagen: Este bloque representa la entrada del sistema, por tanto se le atribuye todo lo comprendido a la escena captada por la cámara, ahora bien, ¿Cuál y específicamente por que se selecciona este dispositivo para con el sistema?; para responder este tipo de cuestionamientos es necesario verlo desde varios puntos de vista.

Operativo

- La cámara o dispositivo en cuestión, debe tener cierta compatibilidad con la plataforma (software IDE + librerías) dispuesta para el procesamiento y visualización de las imágenes (video), en conjunto con todos los parámetros requeridos para llegar hasta este punto.
- Retomando el ítem anterior es obvio que el dispositivo debe tener al menos un puerto de comunicaciones asociado al ordenador para completar las etapas posteriores a esta.
- Debe cumplir unos requisitos mínimos en cuanto a la resolución y tiempo de adquisición de la imagen mostrada en la pantalla del ordenador, ya que de ello se hará un primer dictamen del costo computacional del hardware a elegir.

Estructural

- El dispositivo debe ser liviano puesto que independientemente de que se trabaje con el soporte para el servomotor o no, la misma va a estar sometida a cambios en su eje panorámico y de inclinación, lo que eventualmente requerirá otros elementos en su construcción (perturbaciones por bajo equilibrio y balanceo), probablemente añadirían más peso a los servomotores que harían inviable algunos de los conceptos descritos en el apartado de anexos.
- Lo primordial en el diseño con respecto a esta etapa es buscar dispositivos en el mercado que se asemejen a los elementos descritos en la sección operativa, esto porque es lo que referencia a un estándar o familiaridad en las personas que conocen o han trabajado con sistemas similares. Luego el concepto de domo transparente es utilizado por cámaras que están en la intemperie, ya que se les atribuye una protección frente a los elementos (lluvias, vientos moderados,...), razón por la cual, el tamaño de la cámara no debería extralimitar esta parte de la estructura.

Estético

- Si es utilizado un soporte prefabricado o construido específicamente para una referencia de servomotor, lo que se esperaría es que en conjunto con la cámara, las dimensiones no fuesen superiores a 4 ó 5 veces las dimensiones del microcontrolador (5X2.6X1.6 cm) como para que las mismas superen a las estructuras destinadas al resguardo de los circuitos análogos y/o digitales, esto por el hecho de mantener al mínimo el espacio sobrante y que por supuesto no se mitigue la funcionalidad del prototipo
- El concepto de domo transparente es la elección, pero como dato a tener en cuenta también limita el tamaño de la cámara y soporte, hecho que repercute directamente en el desarrollo, ya que el domo debe estar en la capacidad de albergar dichos elementos,

manteniendo la libertad de movimiento que se le otorgaran a los servomotores con las señales PMW (pan-tilt)

Detección: Esta parte del sistema tiene un largo conjunto de tareas que contemplan inicialmente el preprocesamiento y manejo en general de las imágenes que se muestran en la pantalla del ordenador, que posteriormente traducen en un video que muestra tanto los resultados del algoritmo (frame principal), así como la escena que captura el sensor de la imagen o el archivo alojado en alguno de los directorios del ordenador que contiene y ejecuta el algoritmo.

La detección contempla dos algoritmos uno que concierne a la clasificación de objetos en la escena (catalogados o no, como personas), mediante un algoritmo de pre entrenamiento Haarscade, y otro que compete a la concordancia de objetos que son dictaminados como fondo respecto a un umbral y al uso de una constante de aprendizaje (α en el algoritmo utilizado, sustracción de fondo por media). En la siguiente imagen se pueden evidenciar algunos de los primeros resultados con respecto a la detección de los dos algoritmos por separado (1 y 2 respectivamente), lo que necesariamente trae implementaciones de otro tipo para con los objetivos del proyecto.

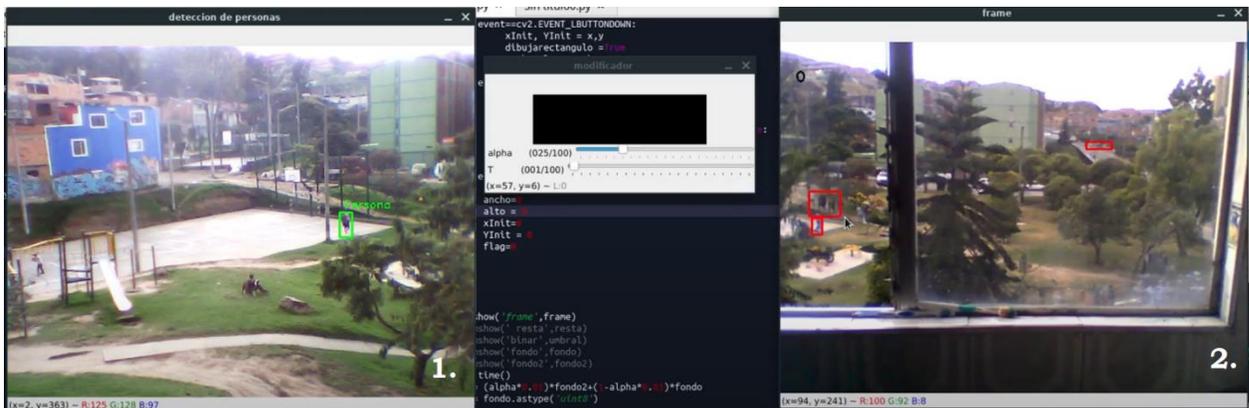


Figura 16. Primeros resultados algoritmo de detección de personas y sustracción de fondo por media. Fuente (imagen propia)

Observaciones

- Como se puede ver en el algoritmo de detección 1, no se contempla elementos que modifiquen significativamente la escena, para que las personas sean clasificadas y denotadas con el recuadro color verde que se muestra en dicho instante.
- Así mismo ocurren errores en la clasificación, que limitan el algoritmo a ciertas distancias y proporciones con respecto a la herramienta utilizada.
- En el algoritmo de detección 2, en contraste, se detectan algunos objetos que se mueven por la escena (automóviles, personas...), lo cual no facilita la posibilidad de clasificar las que corresponden o no a personas, esto pese a un umbral y alpha ajustables que solo añaden más objetos encerrados con rojo en el video o del tiempo que perdura el recuadro,

hasta que el algoritmo lo identifique como fondo y este sencillamente desaparezca del frame

- Los elementos que modifican el video (T, alpha), se tienen que ajustar cada vez debido a perturbaciones en la escena o movimiento de la cámara, lo que eventualmente traería más variables a considerar en las etapas previas a la integración del sistema prototipo.

Control: Este comprende básicamente a la estrategia de control, que permite dictaminar de manera aproximada como debe realizarse un seguimiento posterior a la etapa de detección, dicha aproximación va con respecto a unos principios básicos de tracking por información posicional (píxeles en coordenadas (x,y) en el frame), comparte una similitud con la que el ojo humano capta y sigue los objetos (eje óptico de las cámaras en el mercado), y la información en general que provee el hardware utilizado.

Luego de establecer los parámetros que permitirían mantener el sistema en un punto de operación (actividad 4), fue elemental trabajar con un protocolo de comunicaciones que sirviera como un punto de partida en futuras prácticas de tracking y de unos primeros vestigios de integración del sistema, como lo son por ejemplo elección del campo de visión a abarcar por el detector y tracker (como grados en los servomotores), distancia máxima/mínima en las cuales es posible operar el prototipo para que se muestren unos resultados acordes al bloque control, tiempos de proceso entre otros valores hallados experimentalmente.

Observaciones en la implementación de la estrategia de control vs operación convencional

- La estrategia de control permitió entre otras cosas conocer en profundidad el sistema con tiempos de restablecimiento de las señales para el posterior tracking, establecer una señal de error para futuras investigaciones y determinar el trasfondo que existe entre el punto consigna y la variable proceso.
- Mitigar la presencia de perturbaciones exista o no una detección de personas debido a la naturaleza misma de la configuración, aplicando entre otras cosas acciones correctivas desde este punto.
- Reducción en la no linealidad del sistema, ya que entre otras cosas es claramente impredecible cuantos o cuales son los objetos/personas que se involucran en la escena en conjunto con otras variables que se originan en el diseño y desarrollo del proyecto.
- La operación convencional comprende únicamente el tomar un valor posicional de la(s) persona(s) u objetivo(s) en el frame principal, y asignarles un dato en específico que posteriormente interpretara el microcontrolador, además del uso del protocolo de comunicaciones (serial), para operar el eje panorámico o de inclinación de la cámara en la orientación donde podría encontrarse una persona.
- El modo convencional incrementa así los niveles de perturbación, luego no es posible establecer tiempos de restablecimiento de la señal y el sistema opera de forma aleatoria lo que a menudo representa una salida del FOV preestablecido, haciendo que el tracker eventualmente rote sobre su eje y fuerce el cableado hasta el punto de desconexión o falla.

Tracking: Este bloque comprende a la acción que ejercen los servomotores sobre el soporte y cámara acoplados (tracker), para operarlos acorde a unos parámetros configurados en el bloque

de control (estrategia y comunicación serial), para así ajustar el valor presente en la escena al valor deseado, que en este caso sería la posición actual de la persona detectada y un cambio en la escena capturada. Esto es posible gracias a los grados de inclinación pan-tilt programados (grabados también en el microcontrolador), que permiten utilizar el protocolo serial como vía de comunicación entre el IDE de Python y dicho dispositivo, cuyo funcionamiento resume en la utilización de una librería que “escriba” una señal PMW en el pin del servomotor asociado, en conjunto con un tiempo de restablecimiento de la señal (actividad 5), que cambia la posición del tracker y por ende de la entrada de captura.

Observaciones con respecto a los resultados de esta etapa

- El tracking solo se pudo implementar cuando la detección por sustracción de fondo por media y clasificador Haar se incorporaron en uno solo para conformar lo que más adelante se llamaría algoritmo principal (actividad 3).
- Cuando el detector que alude al conjunto de los dos algoritmos del ítem anterior se prueba con diferentes archivos de video, es posible acoplar el protocolo de comunicación para que las etapas posteriores apliquen de manera adecuada, no obstante se hace necesario reajustar el umbral en más de una ocasión (con el video aun mostrándose en el frame principal), cuando por ejemplo los objetivos cambian su proximidad o hay cambios de luz o perturbaciones en la escena.

Posición de la cámara pan-tilt: Este bloque comprende a la ubicación actual del tracker, que puede ser producto de un principio de operación de todo el sistema, o bien de un conjunto de tareas que llevan a que la instrucción se ejecute por la detección o movimiento de la persona en el FOV, y pueda verificarse en tiempo real con el sonido y posterior movimiento (tracking). En la siguiente figura se puede evidenciar una representación de los grados de movimiento establecidos en cuanto al eje panorámico y de inclinación que compete a esta parte del sistema prototipo

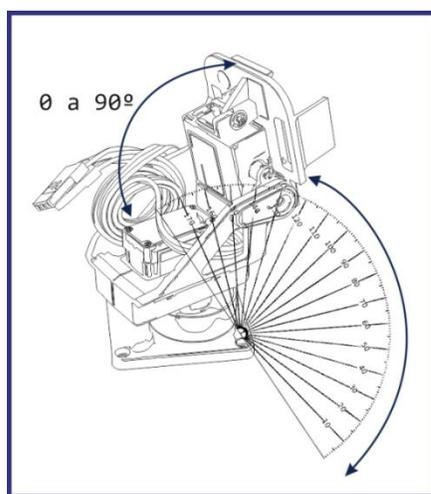


Figura 17. Representación de movimiento (grados), para el soporte del prototipo. Fuente (imagen propia)

Ahora bien realizar el tracking de una persona con una cámara convencional, requiere de ciertas plataformas de hardware/software más un conjunto de conocimientos básicos en procesamiento de señales (imágenes), electrónica además de sistemas y estrategias de control, todo esto enfocado a los sistemas de videovigilancia o seguridad, razón por la cual se debe seguir un orden determinado para la ejecución e integración del prototipo; por ejemplo en pro de los objetivos del proyecto no se puede realizar un movimiento del pan-tilt sin la existencia de una persona en el marco de la imagen, no es posible poner el sistema en marcha aludiendo a videovigilancia o seguridad sin que exista un procesamiento de imágenes que garantice pruebas en entornos reales, a su vez la naturaleza del procesamiento implica el desarrollo del algoritmo de detección, el algoritmo de tracking y unos parámetros iniciales de configuración para el controlador, en ese orden de ideas las actividades desarrolladas se describen a continuación.

Actividad 1

Consiste en realizar la investigación de la idea y documentación de los sistemas disponibles en el mercado, con el fin de consultar los estándares, normativas y procedimientos generales al implementar un algoritmo de detección, inicialmente, esto implica por supuesto una reevaluación del procesamiento de imágenes y a su vez la escogencia de las plataformas disponibles para dicho fin.

Por su parte Python provee de una librería de visión artificial llamada OpenCV, que entre su amplia gama de aplicaciones se encuentra por supuesto una parte clave del proyecto presentado, además cuenta con las librerías para la comunicación, extracción y manipulación en general de los datos que se operan desde que se realiza la captura, hasta que el ordenador la "muestra" en el frame principal tal como se evidencia en la siguiente figura.

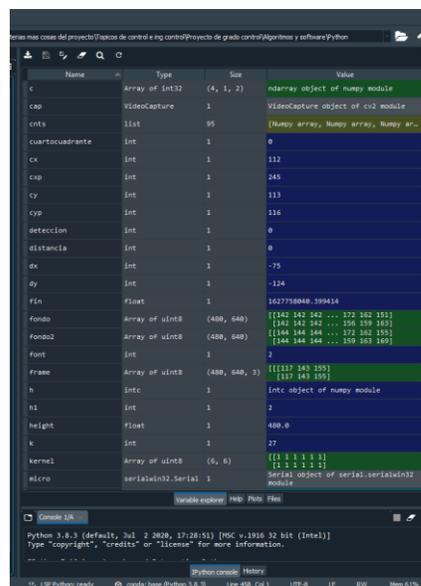
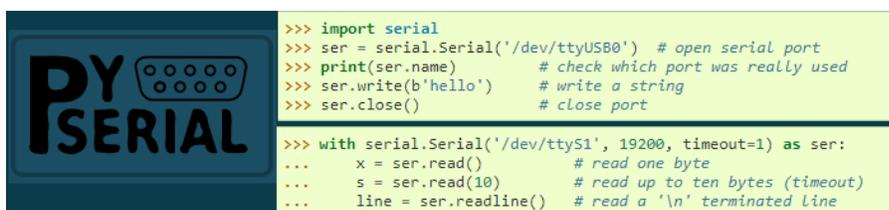


Figura 18. Explorador de variables plataforma Python. Tomado del entorno de desarrollo Spyder(variables propias)

Luego de realizar la búsqueda y lectura de los protocolos de comunicación disponibles para los proyectos de investigación afines al propio, gran parte de ellos opta por una comunicación serial, que no les presente mayor contratiempo cuando se desee integrarlo al hardware que amplíe el rango de visión y captura de las ROI (regiones de interés), con el movimiento sobre el eje panorámico y de inclinación del emplazamiento de la cámara, replicando en ocasiones lo que se le conoce en el mercado como cardan.

Una muestra del envío y recepción de datos respectivamente se implementa por medio de la librería Pyserial manejada por Python, está por ser la plataforma base de detección y a su vez represente una reducción de costes computacionales para otras tareas a ejecutar, luego una muestra las operaciones (envío y recepción) con datos por puerto serie puede detallarse en la siguiente figura.



```
>>> import serial
>>> ser = serial.Serial('/dev/ttyUSB0') # open serial port
>>> print(ser.name) # check which port was really used
>>> ser.write(b'hello') # write a string
>>> ser.close() # close port

>>> with serial.Serial('/dev/ttyS1', 19200, timeout=1) as ser:
...     x = ser.read() # read one byte
...     s = ser.read(10) # read up to ten bytes (timeout)
...     line = ser.readline() # read a '\n' terminated line
```

Figura 19. Script sencillo de envío y recepción con el uso de la Librería Pyserial. Modificada de (<https://pythonhosted.org/pyserial/shortintro.html>)

Como resultado de esta primera actividad se cierne una base sólida para el desarrollo de lo que sería en primera instancia el algoritmo de detección de personas, y la búsqueda de todas aquellas herramientas que consolidan la parte de tracking visto desde la perspectiva del software.

Actividad 2

Involucra fuertemente la parte de detección, al implementar uno de los múltiples algoritmos disponibles para una parte importante del objetivo general del proyecto; el algoritmo de sustracción de fondo por media es idóneo para dicha labor (videovigilancia o seguridad), ya que proporciona la "separación" de un objeto que se mueve en una escena estática durante un periodo de interés (Kaehler y Bradski, 2017), así es posible mitigar o ajustar el carácter de detección por efectos de iluminación o cambio absoluto del fondo y de lo que allí, bajo cierto periodo se clasificase como tal.

Luego de aplicar la forma básica, respectiva al algoritmo de movimiento con la formula correspondiente [21], el desarrollo deriva en cómo podría el algoritmo diferenciar de lo que podría ser un objeto o una persona, para ello se utilizó otra de las herramientas de OpenCV, que determina en sus clasificadores pre entrenados basados en (P.Viola y M. Jones), el tipo de objeto que es detectado por medio de una cámara o elemento multimedia (imagen o video), así con el conjunto de características admitidas las llamadas "Características Haar", es posible almacenar información relevante para su manejo en las actividades posteriores al sistema prototipo.

Sin embargo pese a lo factible en lo que se pueda probar el algoritmo hasta este punto, presenta fallas en la integración, ya que como se mencionó antes no es posible por sustracción de fondo, saber si lo que se mueve corresponde a una persona, tampoco es posible determinar si lo que detecta el Haar pre entrenado es una persona en movimiento, son varias personas o si por el contrario se clasifico inadecuadamente tal y como lo muestra la figura 17 donde el recuadro rojo representa el objeto que se "mueve" por el marco de la imagen y el de color verde la o las personas detectadas por el clasificador de cascada; las barras deslizantes representan el T (umbral) y α (alpha) en el algoritmo de media con el que se ajusta la cantidad de objetos detectados (niveles de gris por segmentación de la imagen), y el tiempo en el que permanecen en el frame principal.

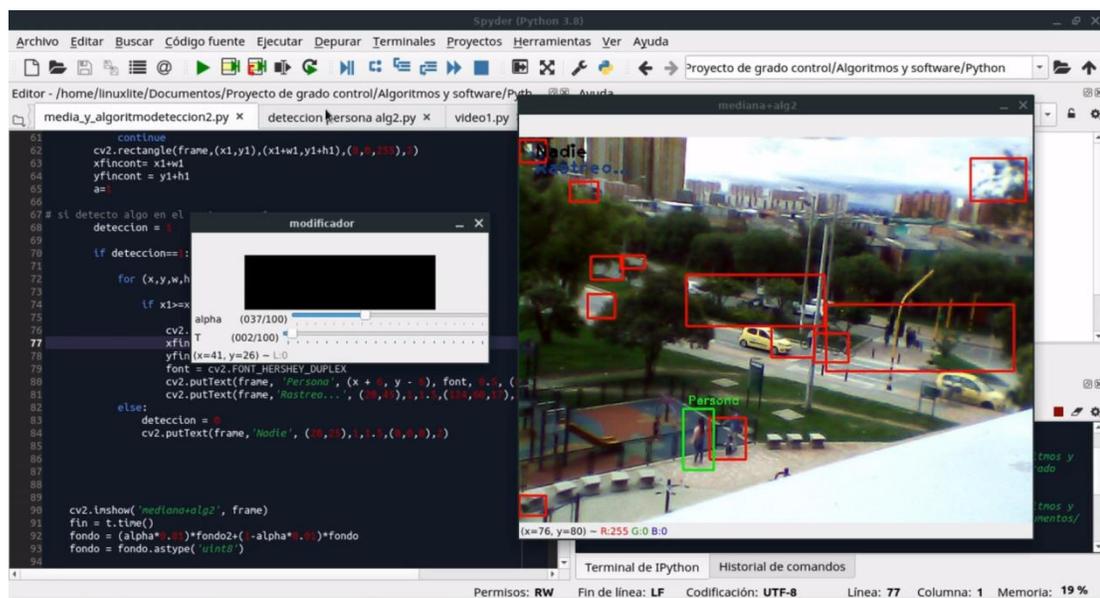


Figura 20. Algoritmo pre entrenado Haar y sustracción de fondo por media. Fuente (imagen propia)

Estos resultados apuntan a la necesidad de realizar una integración de las dos alternativas, para conformar lo que en actividades futuras será llamado como algoritmo de detección de personas.

Actividad 3

Para integrar los algoritmos de sustracción de fondo y clasificador de cascada Haar, es necesario buscar características comunes entre el objeto detectado y la persona detectada conforme estos se desplacen por el marco de la imagen, y así el "tracker" (rastreador) siga a quien corresponde. Para ello existen varios puntos de partida ya que esto es inherente a la imagen o secuencia de imágenes capturadas para este caso, y debido a que juntos poseen dimensiones (x,y), valor de pixel en escala RGB o de grises para donde sea necesario aplicarse, entonces es posible establecer estas mismas características para realizar un dictamen de lo que puede clasificarse como persona o no.

Debido a que el pre entrenamiento tal y como lo expresa la documentación al respecto "*funciona en dos capas la primera es el detector de características que encapsula y modulariza el computo de dichos datos y la segunda capa corresponde a la cascada real reforzada, que utiliza sumas y diferencias sobre regiones rectangulares de las características computadas*"(Bradski y Kaehler, 2017), es posible aprovechar dichos rectángulos para compararlos en conjunto con el algoritmo de sustracción de fondo por media, haciendo uso por supuesto de un criterio que consiste básicamente en como el ojo humano sigue objetos y desde luego se mantiene el eje óptico de la cámara para apuntar a un objetivo cuando parece dirigirse hacia el infinito, cuidando por supuesto ángulo de visibilidad y cambios en la escena, aprovechando su "centro" para compararlo con el de toda la imagen captada lo cual servirá para el controlador y tracker mencionados en su momento.

El centro tanto del clasificador Haar como el de sustracción de fondo se pueden calcular dinámicamente (es decir según como cambie la escena), para este caso en particular se utilizaron dos submétodos para el primero, se halló con el centroide de este rectángulo preestablecido con la formula y método analítico para un rectángulo.

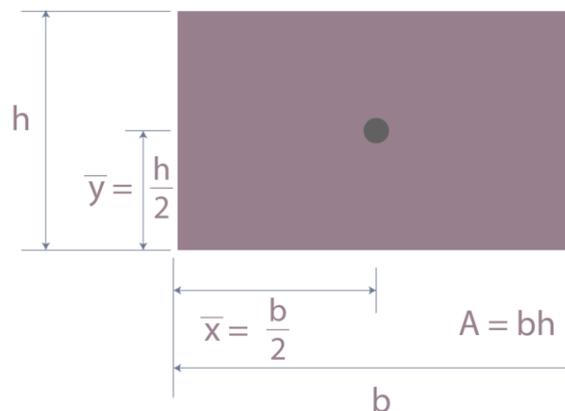


Figura 21. Método analítico para encontrar el centroide de un rectángulo. Tomado de (repositorio Moodle UNAM)

Para el algoritmo de sustracción de fondo se utilizó la herramienta de OpenCV denominada *moments* la cual calcula los centroides de una figura en particular hallando el promedio ponderado de las intensidades de píxeles, la cual a su vez también ayuda a encontrar propiedades como el radio, área entre otras, esto lo hace por supuesto con una imagen binarizada y utilizando la siguiente formula dispuesta por algunas de las características extraídas por la funcionalidad. [38]

$$C_x = \frac{M_{10}}{M_{00}} \tag{6}$$

$$C_y = \frac{M_{01}}{M_{00}} \quad (7)$$

Luego de obtener los centroides y graficarlos en el frame principal, se procede a aplicar una razón de distancia con el fin de determinar cuál es la probabilidad de que el objetivo sea una persona a la cual rastrear mientras permanezca en el rango de visibilidad, sin embargo hacerlo por puntos sobre el plano involucraría un conjunto de operaciones aritméticas de distancia, que en su mayoría retornaría lo que en términos de programación se conoce como datos de punto flotante, es por esta razón que se utiliza los círculos/circunferencias porque son pocos los elementos en comparación, los que se han de computar para obtener resultados favorables, ello al no depender en demasía de ángulos, lados, vértices, dimensiones entre otras terminologías geométricas y matemáticas. Hallar la información entre dos circunferencias teóricas, implica el axioma que involucra la relación de posición entre juntas y así exprese en una probabilidad aceptable, lo que el algoritmo de detección definitivo clasifica como una persona tal y como lo muestra la figura 19.

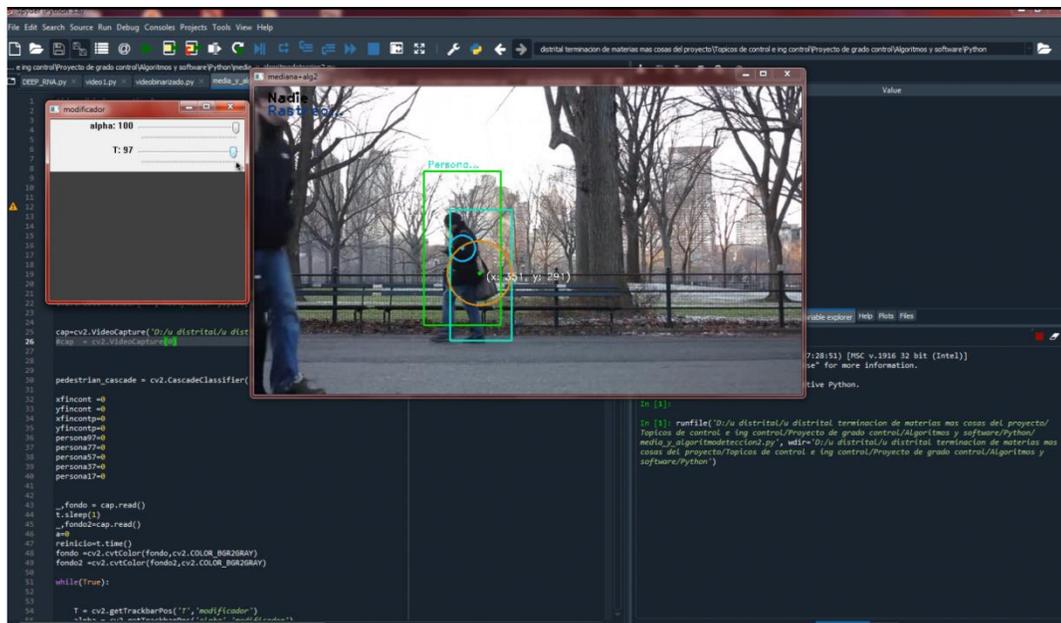


Figura 22. Integración del algoritmo Haar y sustracción de fondo por media con el principio de posición relativa de dos circunferencias. Fuente (imagen propia, video “People walking in central park Youtube”)

Donde el recuadro color verde, correspondiente al clasificador Haar se mantiene y su centroide con circunferencia asociada van en un azul claro, por su parte el algoritmo de sustracción de fondo cambio su trazo rectangular de color rojo a azul claro, porque esto hace referencia al índice de probabilidad dictaminado por su posición relativa con respecto a la circunferencia del clasificador, obviamente se tuvo que tener una circunferencia a asociar conforme se iban

detectando objetos en movimiento, que es desde luego aquella color marrón y centroide color verde claro, cuyas coordenadas se priorizaron al punto de dejarlas plasmadas en el frame principal de todas las pruebas realizadas debido a la dinámica comportamental que tiene el rastreador y la base misma de la sustracción de fondo en una escena no estática.

Como resultado de esta actividad se obtiene la consigna las probabilidades y colores asociados que permite determinar lo que el algoritmo detecta como una persona con un porcentaje y radios dictaminados por el ponente, teniendo en cuenta la distancia y resolución mínima permitida a la cual se va a instalar el prototipo para captar la gente, además de algunas propiedades descritas para determinar las posiciones de la circunferencias por clasificador Haar y por el algoritmo de sustracción de fondo por media, los métodos se evalúan y programaron con la siguiente lógica:

Exteriores: La distancia entre los centros, distancia (en la figura 20 y programa), es mayor que la suma de los radios, razón por la cual las circunferencias no tiene puntos en común. Probabilidad de persona al 17%.

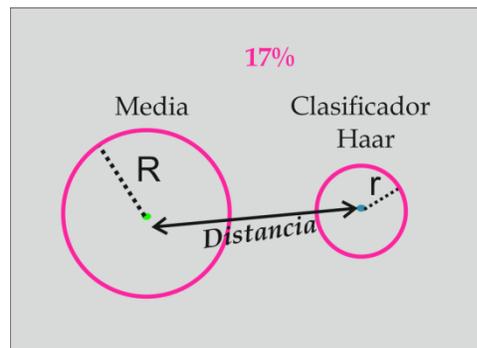


Figura 23. Color y probabilidad asociados a dos circunferencias con posición relativa al exterior.
Fuente (imagen propia)

Tangentes exteriores: La distancia entre los centros es igual a la suma de los radios por eso el centro de cada circunferencia es exterior a la otra y tienen un punto en común (punto de tangencia figura 24). Probabilidad de persona al 37%.

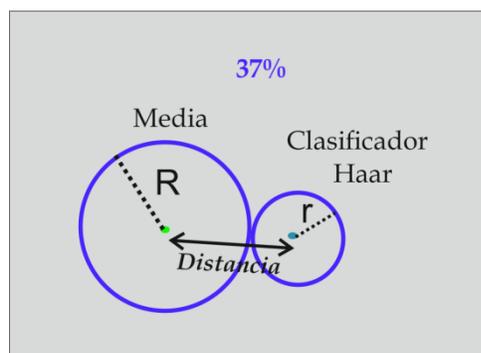


Figura 24. Color y probabilidad asociados a dos circunferencias con posición relativa tangente exterior. Fuente (imagen propia)

Secantes: La distancia entre los centros es menor que la suma de los radios y mayor que su diferencia, tienen dos puntos de cruce en común (figura 25). Probabilidad de persona al 57%.

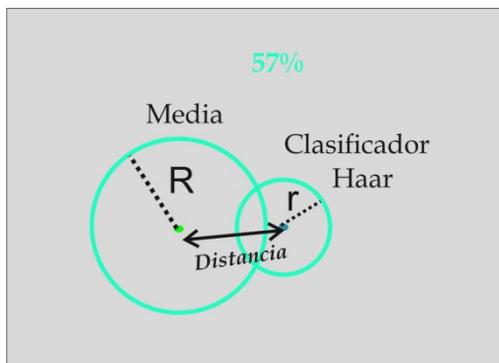


Figura 25. Color y probabilidad asociados a dos circunferencias con posición relativa secante. Fuente (imagen propia)

Tangentes interiores: La distancia entre los centros es igual a la diferencia entre los radios, razón por la cual una de las circunferencias está dentro de la otra (figura 26), poseen un punto en común. Probabilidad de persona al 77%.

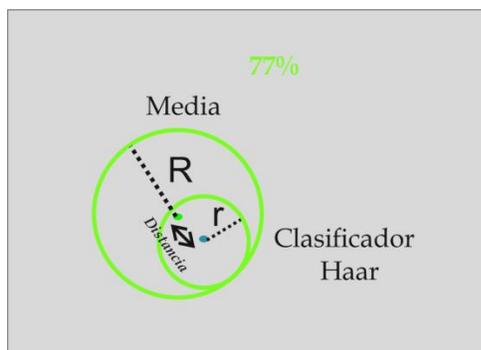


Figura 26. Color y probabilidad asociados a dos circunferencias con posición relativa tangentes exteriores. Fuente (imagen propia)

Interiores: La distancia entre los centros es mayor que cero y menor que la diferencia entre los radios y a diferencia del caso anterior la circunferencia está dentro de la otra pero no existen puntos en común (figura 27). Probabilidad de persona al 97%.

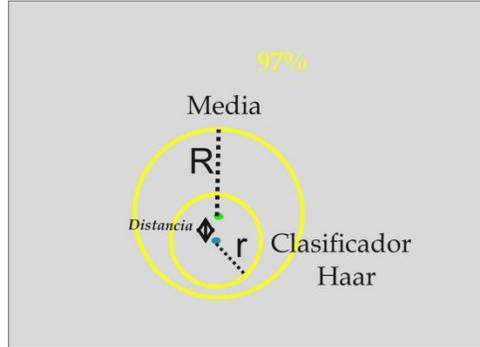


Figura 27. Color y probabilidad asociados a dos circunferencias con posición relativa interiores.
Fuente (imagen propia)

Concéntrico: Tienen el mismo centro lo que quiere decir que la distancia entre sus radios es igual a cero (figura 28) no tienen puntos en común a menos de que se trate de la misma circunferencia.

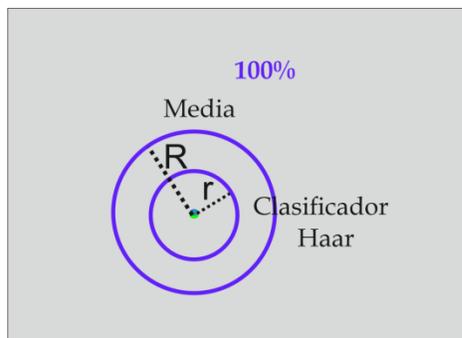


Figura 28. Color y probabilidad asociados a dos circunferencias con posición relativa concéntrica. Fuente (imagen propia)

A pesar de que se incluye, el que se dé el caso concéntrico es prácticamente un atípico que se produce por el ruido de la señal.

Actividad 4

Resume los parámetros de comunicación a tener en cuenta tanto para el desarrollo del circuito, como de la planta tracker (rastreador), entonces la vista preliminar de los hechos permite tener en cuenta lo siguiente:

- Como es un dato a comunicar por el puerto serie necesario para establecer un primer indicio de funcionamiento, el sistema se configurara en realimentación que tal y como lo expresan (Pérez Mario .A, Hidalgo Pérez A. y Berenguer Pérez E.,2007) "*es una propiedad de los sistemas que permiten que la salida del sistema o cualquier variable del mismo sea comparada con la entrada al sistema o con cualquier componente del sistema, de tal manera que pueda establecerse la*

acción de control apropiada entre la entrada y la salida", luego en forma general, se dice que la realimentación existe en un sistema cuando hay una secuencia cerrada de relaciones causa-efecto entre las variables de un sistema.

- Una vez establecida la forma de comunicación y escrita en código con la plataforma donde reside el algoritmo de detección de personas (Python), se procede a determinar pautas de mapeo del frame principal lo que quiere decir, aplicar una de las técnicas básicas de tracking (trayectoria del objetivo), así, este siga o no una trayectoria determinada las subdivisiones aplicadas sobre eje horizontal y vertical de la escena permiten enviar un dato con respecto a los siguiente:

1. La variable error se establece como una diferencia entre el centroide del objeto que se desplaza por la escena (cx, cy), que al estar integrado con el algoritmo pre entrenado Haar corresponderá al de la(s) persona(s) y el valor de pixeles que representa la región de interés tal como se muestra en la figura 29, este valor esta dictaminado para pasos de 10 porque a eso corresponde la distancia aproximada del lente de la cámara utilizada (64 horizontal y 48 vertical), con una resolución de trabajo de 640X480, para que el valor de referencia (SP) se ajuste con la persona "moviéndose" por el frame principal con las coordenadas (cx, cy) visualizadas en todo momento en referencia a las pruebas realizadas (videos), razón por la cual (x,y) corresponden para este caso como la variable proceso.

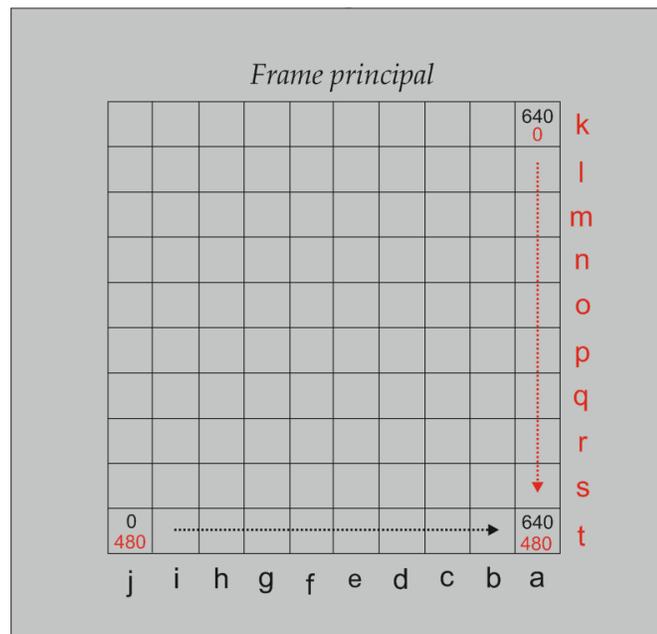


Figura 29. Mapeo del frame principal, conforme a la distribución de las coordenadas por escena captada. Fuente (imagen propia)

2. Dicho esto es posible afirmar que con los parámetros de configuración citados, se está hablando de un sistema de control realimentado ya que cumple con mantener una relación preestablecida entre la salida (cambio en el pan-tilt y escena si de tiempo real se tratase), y

la entrada de referencia (escena donde aparece o se desplaza la persona en cuestión FOV), comparándolas ambas y utilizando la diferencia como variable de control CV (señal que ajusta el eje panorámico o de inclinación de la cámara). [39]

3. Finalmente se ejecuta una operación que en presencia de perturbaciones tiende a reducir diferencias entre la salida y la entrada del sistema (error), así que es usual que cuando se acuda al control de realimentación se considere a las perturbaciones como aquellas que surgen aleatoriamente, y que a su vez pueden incluir una compensación dentro del sistema siempre y cuando puedan ser predichas, que hagan hasta cierto punto innecesaria la labor de control.

- En consecuencia y dados los antecedentes (estado del arte) del presente documento, se optó para que la labor de la variable controlada fuese adoptada por servomotores, que estuviesen en la capacidad tanto de lidiar con el peso del mecanismo, así como de tener una precisión en el marco de lo aceptable por el ojo humano. Entonces la planta, específicamente el tracker fueron desarrollados en base a soportes pan-tilt disponibles en el mercado en conjunto a los servomotores compatibles a ellos, para así emitir los posibles conceptos que puede adoptar el sistema prototipo en investigaciones futuras.

Como resultado de esta actividad y del montaje parcial del tracker (rastreador), fue necesaria la elección de un dispositivo que pudiese interpretar los datos por escritura serial y posteriormente traducirla a una señal tipo PWM que pondría los actuadores en modo de rastreo (tracking), dicho dispositivo es el microcontrolador ESP8266 NodeMCU, que posee aplicaciones interesantes enfocadas al IoT, y una compatibilidad con la plataforma Arduino IDE que en conjunto a las librerías asociadas *servo.h* (para manejar ángulos de movimiento en el servomotor), *PID_v1.h* (una alternativa viable para PID básico) entre otras; podría ser material de estudio para proyectos que tomen el propio como una base de investigación en la videovigilancia o seguridad.

Actividad 5

Conciérne a los ajustes del sistema que hacen de la señal de datos por puerto serial un tanto menos susceptible a perturbaciones, producto del algoritmo de detección de personas, el movimiento del eje vertical u horizontal de la cámara o por simples cambios en la escena para esto se tuvo en cuenta:

- *Eliminación del ruido:* Fue posible su desarrollo aplicando un contador de objetos y personas en el algoritmo, con el fin de que solo se pudiese realizar el rastreo y conteo de personas cuando hubiesen cierta cantidad de objetos que el algoritmo de sustracción de fondo pudiesen señalar por medio de un rectángulo rojo, si fuese el caso de una "no persona"; y bajo esta misma lógica solo enviase los datos por el puerto serial y reiniciase los contadores cuando juntos fuesen en su momento distintos a cero.

- *Tiempo de reconstrucción:* Producto de una serie de pruebas, el algoritmo tarda un lapso de tiempo en omitir ciertos elementos que aparecen en el frame principal que no corresponden a personas ni a objetos a pesar de ser clasificados como tal, en la mayoría de los casos cuando no

era por un cambio abrupto en el threshold (umbral), del modificador por barra deslizante, se originaba por el movimiento del pan y tilt de la cámara, lo que en su momento fue ajustado acorde a un exhaustivo método de captura de los tiempos donde se desarrollaba dicha acción, en conjunto con una variedad y muestra de videos donde se varían los parámetros que se esperan sean una fuente de perturbación tal y como lo demuestran los cálculos de las siguientes figuras.

Tiempo de reconstrucción de las señales, lecturas realizadas

Primer video (alpha al máximo T bajo sin cambios notables ensayo con tilt <<inclinacion>>)

$14912ms - 13869ms = 1043 \text{ ms}$
 $14403ms - 13869ms = 534 \text{ ms}$

Segundo video (alpha al máximo T sin cambios notables ensayo
Con pan <<panoramica>>)

$132685ms - 131572ms = 1113 \text{ ms}$
 $132102ms - 131572ms = 530 \text{ ms}$
 $148988ms - 147074ms = 1914 \text{ ms}$

Tercer video (alpha al máximo T sin cambios notables ensayo con pan <<movimientos muy poco
abruptos en la panorámica>>)

$47367ms - 46031ms = 1361 \text{ ms}$
 $52027ms - 51371ms = 656 \text{ ms}$
 $99707ms - 99056ms = 651 \text{ ms}$
 $122714ms - 122037ms = 677 \text{ ms}$
 $150801ms - 150251ms = 550 \text{ ms}$

Figura 30. Tiempo calculado para la reconstrucción de las señales, movimiento pan y tilt. Fuente (imagen propia)

Tiempo de reconstrucción con videos de ruido considerable

$43025ms - 41586ms = 1439 \text{ ms}$
 $43025ms - 42616ms = 409 \text{ ms}$
 $165741ms - 165516ms = 226 \text{ ms}$
 $165741ms - 165057ms = 684 \text{ ms}$
 $184163ms - 184ms = 163 \text{ ms}$

Figura 31. Tiempo calculado para la reconstrucción de las señales, movimientos aleatorios de la cámara. Fuente (imagen propia)

Resultado de esta actividad trajo en una reducción considerable de ruido tanto en el manejo que se hace en la detección de personas, como en el posterior envío de datos que maneja el

rastreador, constancia de ello está la siguiente figura donde a pesar de que se capturan y muestran las coordenadas de aquellos detectados por el algoritmo como objetos, no se alcanzan a clasificar como tal, lo que impide que el algoritmo pre entrenado por cascada Haar, lo tenga en cuenta para catalogarlo como persona, contarlos y posteriormente emitir una orden a ejecutar por los actuadores del sistema.

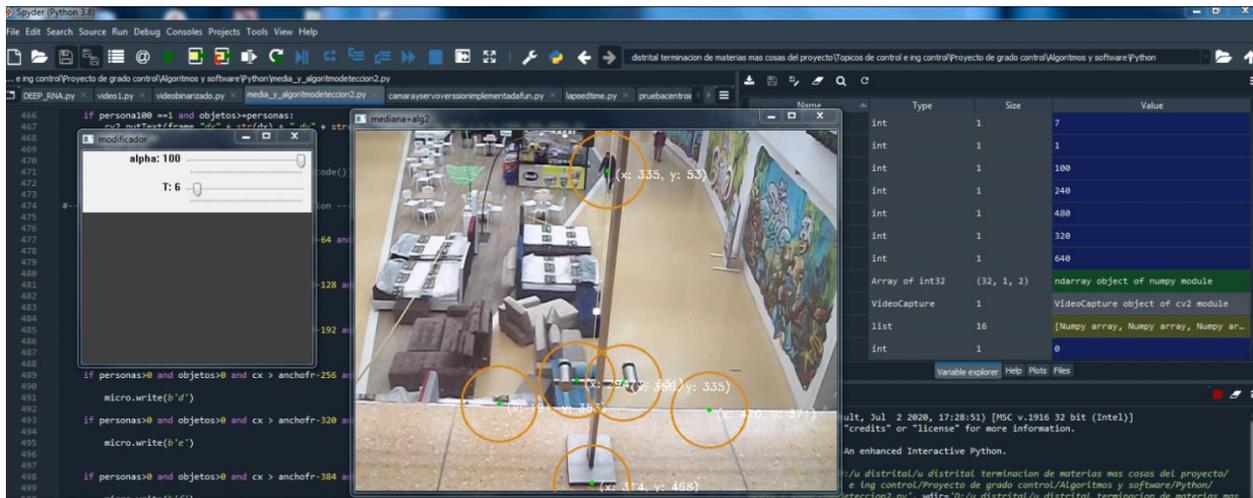


Figura 32. Eliminación de perturbaciones en la integración de los algoritmos para la detección de personas. Fuente (imagen propia)

Dónde:

- El círculo color marrón representa a la circunferencia "central" del objeto en movimiento, puesto que se ha descartado por superar la cantidad de objetos detectados en el frame principal lo que se exployó como perturbaciones en la detección
- El punto en color verde claro representa el centroide con las coordenadas (x,y color blanco), de la posición en píxeles a donde refiere el objeto al cual el algoritmo de sustracción de fondo no representa en un recuadro por las razones ya descritas

Actividad 6

Consolida la parte de detección y tracking de persona, haciendo la validación del funcionamiento, hasta este punto (sistema de control por realimentación), con la simulación (figura 30) y posterior puesta en marcha. En primera instancia con el eje horizontal y vertical por separado, para después probarlo en conjunto con el soporte y circuito del sistema que incluye las condiciones mínimas de funcionamiento en pruebas primarias.

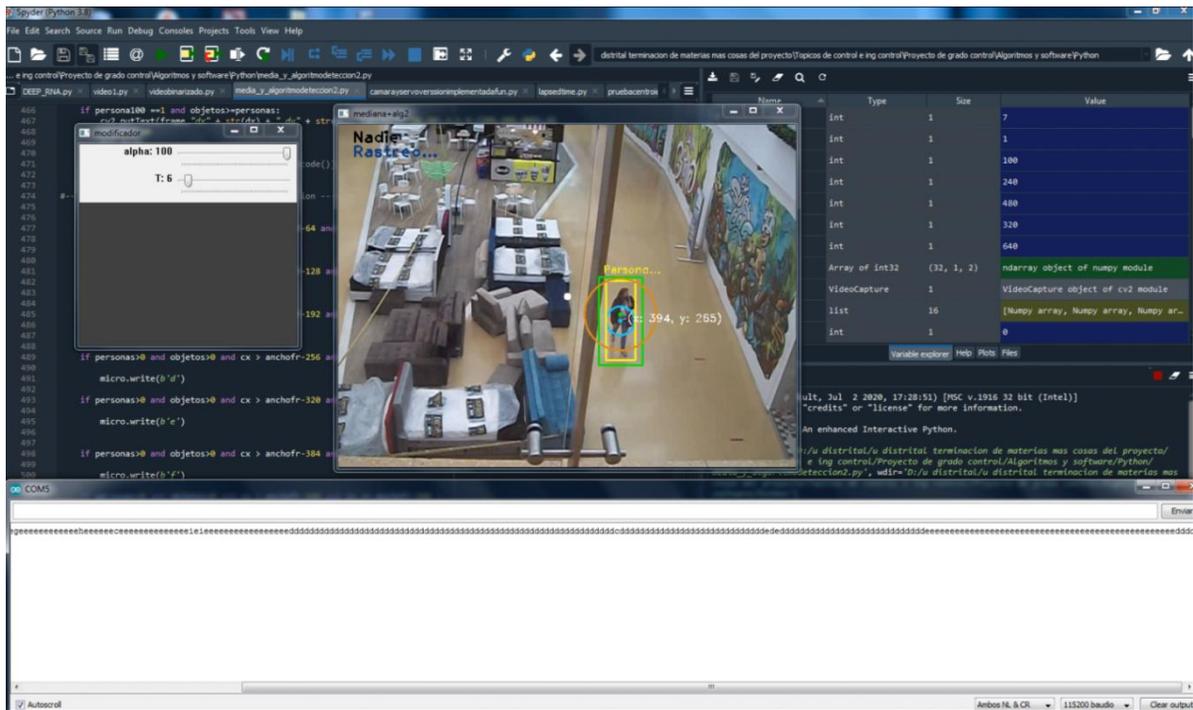


Figura 33. Simulación primaria del sistema con eje pan de la cámara y el monitor serie del IDE asociado al microcontrolador escogido. Fuente (imagen propia)

Donde las herramientas que sirvieron como medio para la funcionalidad de los algoritmos por simulación fueron:

- El monitor serie del Arduino IDE que funciona como una herramienta para depurar y establecer comunicación entre un ordenador y un dispositivo compatible, destaca por su sencillez y en la actualidad recibe soporte en su mayoría para las tarjetas que pertenecen al núcleo del IDE. [39]
- El emulador de puertos serie virtual, cuya función fue la de "reemplazar" el microcontrolador en el COM5 (SO Windows), para determinar el comportamiento del envío y recepción de los datos entre las etapas de software y hardware que caracterizan al sistema prototipo. Esta aplicación entre otras cosas crea una compatibilidad que ayuda al usuario a hacer que el flujo de datos de cualquier fuente como la conexión TCP, USB etc., sea visible para cualquier programa a través de un puerto serie virtual (Eterlogic.com software Helper)

Acto seguido se prueban los datos adquiridos haciendo una integración del sistema y así mismo surge una verificación de que se cumplen a cabalidad todas las metas operativas para con el prototipo. Como resultado de esta actividad surge el diseño y/o implementación de otras técnicas que mejoren o atenúen algunas de las características propias del tracking como lo son el tiempo en que se mueve el pan-tilt a su momento de ajuste, la forma en la que se maneja la señal de error (diferencia entre la entrada y salida del sistema), entre otros factores inherentes a la(s) estrategia(s) de control.

6. Resultados

6.1. Propuesta definitiva

Antes de abordar el diseño de la solución propuesta, se va a exponer de forma general el funcionamiento final del proyecto, con el flujograma de la siguiente figura, además de su descripción.

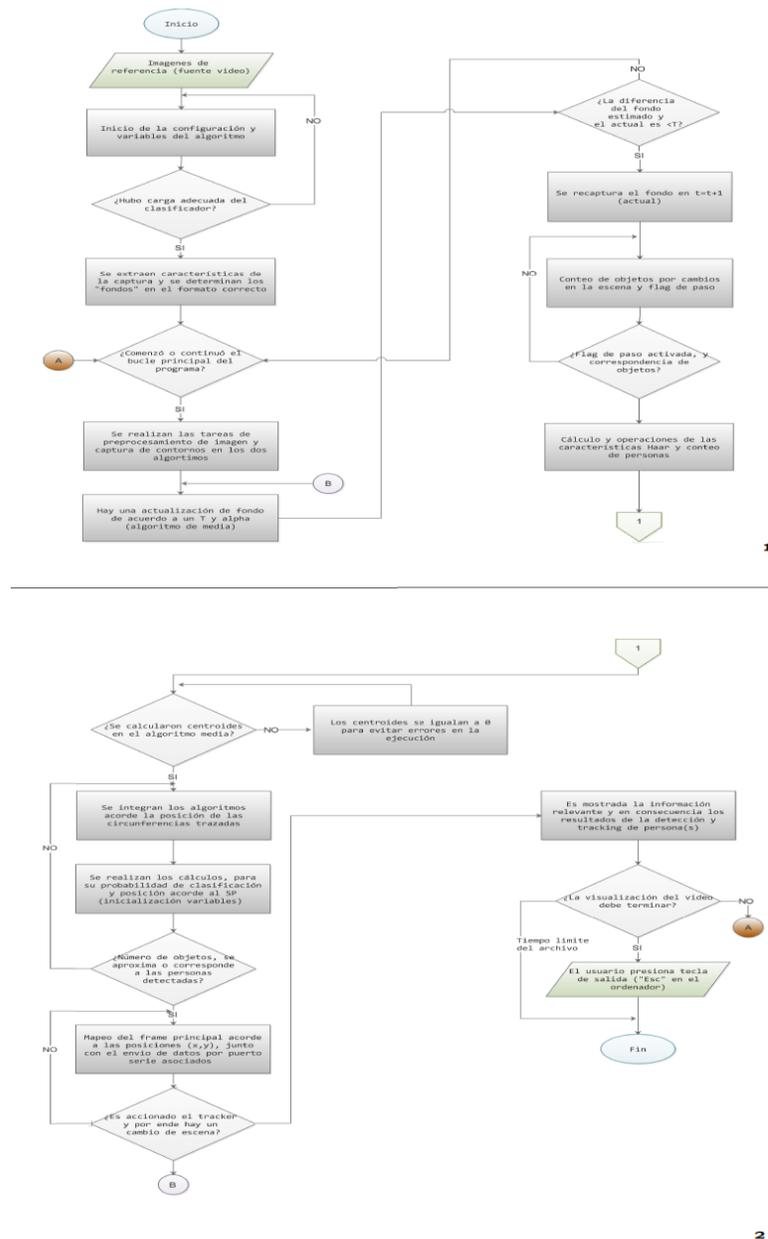


Figura 34. Diagrama de flujo sistema prototipo detección y tracking de persona. Fuente (imagen propia)

El diagrama anterior es una representación de cómo opera y cambia el sistema, acorde a la integración de los algoritmos de sustracción de fondo por media y clasificador de persona tipo Haar, para tal fin se plantean los siguientes pasos (en orden por el flujograma):

1. Inicia el programa con las librerías requeridas (IDE Python) y es cargado desde un archivo en el directorio del ordenador, o en tiempo real por un dispositivo externo. Un video que corresponde a la escena que funciona como entrada al sistema
2. Comienza la configuración del sistema que consta de la creación de las barras deslizantes que corresponden al umbral y aprendizaje (T y alpha respectivamente), luego de las variables que componen al algoritmo principal además del puerto serial que utiliza el microcontrolador y por supuesto lo que será el punto consigna (SP)
3. Es cargado el archivo que corresponde al clasificador preentrenado, para posteriormente extraer características relevantes al algoritmo principal (es el algoritmo de detección y tracking integrados)
4. Luego se realiza la captura de dos fondos, para definir, lo que sería el inicio del algoritmo de sustracción de fondo por media con el preprocesamiento de imágenes asociados a el
5. Comienzan las fases de actualización y segmentación en el algoritmo dentro del bucle principal, que eventualmente traducirán en cambios sobre el frame acorde al T y alpha que ajuste el usuario una vez se habiliten las funciones y puedan visualizarse en la pantalla del ordenador
6. Producto de estas acciones es realizado un cambio sobre el frame principal, que corresponde al trazo de un rectángulo (rojo) sobre el objeto que se desplaza en una escena, lo que eventualmente se convierte en un punto de partida tanto para activar una "flag" (puntero, bandera....) de detección, contar la cantidad de objetos no clasificados como fondo y así mismo ajustar que tan rápido tiene que suceder (aprendizaje)
7. Un concepto similar al paso anterior sucede con lo obtenido en el clasificador (conteo de personas solamente), pero el criterio a este momento solo se determina por la flag y la cantidad de objetos dictaminados por código (10 por defecto); funciona para ajustar posibles perturbaciones, así las personas serán "encerradas" por el rectángulo de contorno y clasificadas como tal (el color es distinto)
8. Son calculados los centroides tanto del rectángulo del algoritmo de sustracción de fondo por media y el que corresponde al clasificador Haar, en ese orden, para realizar tanto la integración de los dos algoritmos, como los cálculos que permiten determinar la probabilidad de que el objeto que se desplaza por la escena del frame principal, corresponda o no a una persona
9. Se realizó un mapeo del frame principal (sección de metodología), con el fin de ajustar al punto consigna lo que compete a la señal de error, como una resultante de la integración de los algoritmos y conforme se desplace el o las personas por la escena, realizar el seguimiento con el eje panorámico y/o de inclinación del tracker (servomotores, cámara y soporte mecánico debidamente acoplados a los demás elementos del sistema)
10. Inmediatamente después retorna al bucle principal (flujograma), halla o no cambiado la escena (video aun mostrado en pantalla), con el fin de repetir indefinidamente todo el proceso desde ese punto (ítem 5) si de tiempo real se tratase, o si por el contrario el archivo de video cargado aún no ha finalizado
11. Dentro de todo el proceso se guardó la información relevante de todos los pasos hasta este punto, para visualizar los más representativos por medio de la pantalla del ordenador (frame

principal llamado con el comando ejecutar en el IDE), independientemente del comportamiento que se ajuste para con el detector o tracker

12. Si la ejecución del programa debe terminar el usuario presionara la tecla “Esc” en el ordenador, o esperar a que la misma “desaparezca” por qué el archivo o fuente de video ha finalizado

En lo que respecta al diseño definitivo de la estructura para la solución propuesta en el proyecto, en referente al uso de “ventanas de movimiento” (cuadros de imagen que traducen en el ojo como un video). Cuya finalidad consiste en aplicar un conjunto de operaciones lógicas, para procesar lo que respecta a movimientos o personas en general que se detecten en las dimensiones y campo de visión establecidas, descartando los objetivos que estén por fuera (personas), se aplica un concepto de pan-tilt desarrollado por medio de dos servomotores SG90 (figura 31) que están instalados en el soporte correspondiente a la referencia como se ve en la figura 32.

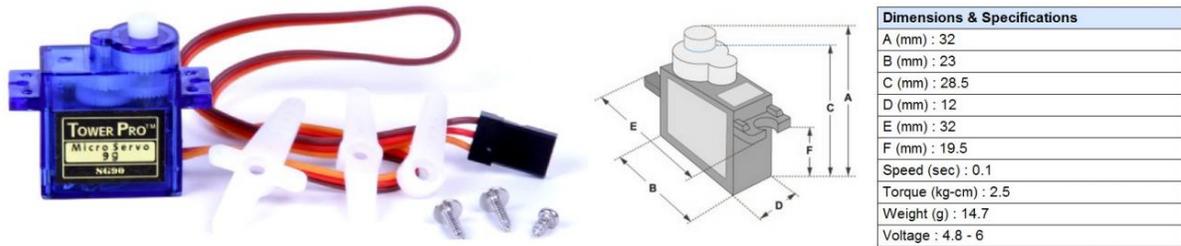


Figura 35. Servomotor SG90 y dimensiones. Modificado de (SG90 datasheet)



Figura 36. Soporte para el servomotor SG90. Tomado de (Mactronica-Mercadolibre.com)

En el anexo 1 se puede ver los planos del sistema para el movimiento del pan-tilt que se implementó, además en la figura 33 se puede ver una foto del sistema total montado.



Figura 37. Sistema prototipo montado. Fuente imagen propia

Para la implementación de un algoritmo que pueda solventar las implicaciones del uso de una cámara estática, se optó por la integración de dos algoritmos usados comúnmente por separado:

- El primero de ellos (pre entrenamiento en cascada Haar), se encarga de encontrar los patrones correspondientes para clasificar lo que en la imagen podría corresponder a una persona
- El segundo por su parte se encarga de establecer bajo ciertas condiciones (umbral y velocidad de aprendizaje), cuáles de los objetos captados en una escena son catalogados como "fondo".

Dicha solución fue planteada con el lenguaje de programación Python y el IDE Spyder, que proveen tanto las librerías de visión artificial, como aquellos algoritmos de pre entrenamiento recomendados para la clasificación de objetos en imágenes o videos, por último se aprovechó otras de las herramientas de la plataforma para añadir la comunicación por puerto serie que permite entre otras cosas el desarrollo operativo del tracker (rastreador) junto con la(s) estrategia(s) de control requerida(s).

El planteamiento del problema expone la complejidad de la forma en la que se muestran los resultados, para proyectos que han dado solución en el marco del uso de cámaras en movimiento, con el conjunto de los métodos, análisis y obtención en general de la información asociada al desarrollo. Varias alternativas fueron fruto de la investigación/experimentación con el sistema prototipo y otras destacan por ser sugerencia indirecta de los antecedentes (estado del arte) y de tutorías extracurriculares; entre las opciones se contemplan:

- Repositorio personal de videos donde se pueden evidenciar aspectos generales de funcionamiento del algoritmo de detección, modificando por supuesto el origen del archivo de video a asociar
- Descripción básica del cómo desarrollar simulaciones del pan o tilt (por separado en cuanto a operatividad y orden), así cual aprovechamiento de las características extraídas en cada imagen para la comunicación, el control e interpretación en la acción de tracking

- Lenguaje informal en la muestra de resultados y no en la documentación donde desde luego sea pertinente aplicarse

Integración del sistema

El sistema se consolida gracias a labor generada en cada uno de los bloques de diseño, además de los componentes que le hacen falta a cada uno para no irrumpir significativamente en las etapas siguientes, y así mismo converjan en el objetivo de detectar y seguir personas, por ejemplo, un conjunto de pasos que realiza el sistema prototipo ya integrado en una situación habitual se plantea de la siguiente forma:

Suponga que tiene el prototipo operando con el tracker y cámara apuntando en una dirección de momento el sistema se encuentra “estático” (con una preconfiguración del T y alpha acorde al lugar), esto debido a que no hay personas en la escena, poco tiempo después aparece una persona que se desplaza de izquierda a derecha a una velocidad promedio de unos 5km/h, se mantiene constante a una distancia, que a el dispositivo de captura le facilita la detección, además el sujeto no se detiene a lo largo de su trayecto. Entonces el conjunto de pasos seria:

1. El sensor de la imagen lo capta y visualiza en la escena inicialmente con una parte del cuerpo que es reconocida por el algoritmo de detección como un objeto y lo encierra en un recuadro de color rojo, al cual ya es posible calcularle un centroide que se evidencia con una coordenada (x,y), y un circulo de radio r que lo circunda en el frame principal.
2. Luego de que la persona ha ingresado al FOV del prototipo, el algoritmo clasificador Haar en el detector lo “encierra” en un recuadro color verde claro (frame principal), y le calcula un centroide color azul con su respectivo circulo concéntrico de radio r.
3. Se comparan las distancias entre círculos y sucede la incorporación de los algoritmos en lo que respecta al detector, permiten así asignarle una probabilidad que lo catalogue como una persona acorde al porcentaje preestablecido.
4. Una vez se visualiza en el frame principal la detección realizada, se “encierra” en un recuadro la persona con el color que corresponde a la probabilidad que dictamina el algoritmo, luego según su posición en el eje de coordenadas se le asigna un dato que se envía por puerto serie.
5. Este dato es interpretado en el microcontrolador como un valor PWM que accionara el eje panorámico del tracker, seguirá a la persona hasta que complete todos los valores programados en la estrategia de control.
6. Todo esto por supuesto con todos los elementos que vienen antes, durante y después de desaparecida la persona del FOV del sistema prototipo.

Otras situaciones también fueron evaluadas de tal forma que se siguiera una secuencia similar a la planteada, y se corroborara a nivel no solo teórico sino también a nivel experimental que corresponden a un sistema de detección y tracking de personas integrado.

Conforme el tracker y el algoritmo de detección de personas iban tomando forma, así como la comunicación serie que hace posible el canal de operatividad entre el detector (algoritmo en el ordenador) y la estructura pan-tilt (servomotor con soporte y cámara), se incrementaba la

necesidad de implementar un dispositivo capaz de intermediar entre las partes mencionadas. El microcontrolador NodeMCU ESP8266 es el dispositivo elegido para dicha labor, que junto con el monitor serie y librerías asociadas, puede operar la parte del servomotor y la de una posible implementación PID. Por ultimo para la fase de pruebas el sistema prototipo adapto el mencionado concepto 1 (apartado anexos), que se asemeja a algunas de las cámaras pan-tilt disponibles en el mercado. A continuación en la siguiente figura se muestra el resultado de todo lo recopilado hasta ahora, como una secuencia de movimientos de seguimiento (tracking).



Figura 38. Detección y tracking de personas por el prototipo del sistema. Fuente imagen propia

La figura 38 muestra el seguimiento que hace el sistema de una persona, además muestra que hace el sistema cuando existen más de una persona en la escena. Del fotograma 1 al 2 el sistema detecta el movimiento pero todavía el algoritmo Haar no lo ha detectado como persona, a partir del fotograma 3 al 11 el algoritmo Haar hace su detección la cual tiene una probabilidad del

100% (ver actividad 3 en metodología), también se puede observar que a partir del fotograma 7 entran otras personas a la escena las cuales también son reconocidas por el sistema.

Protocolo de pruebas

En el orden de ideas de lo que el apartado de experimentación contempla, para cada prueba realizada se tuvieron en cuenta los siguientes aspectos técnicos.

En la detección:

- El umbral (T, 0 a 100), se ajusta en el frame principal a niveles que rondan entre los 5 y 22 para archivos de video cargados y valores superiores a 35 para videos en tiempo real, no obstante y debido a que es posible ajustar este valor dinámicamente (mientras se visualiza el video), en ambas situaciones y si la escena lo requiere pueden modificarse bien sea para aumentar la cantidad de objetos que pueden clasificarse o no como una persona, o disminuirla para realizar una mitigación temporal de perturbaciones en la detección
- El alpha (alpha, 0 a 100), se ajusta a niveles altos >88 por que en el algoritmo principal se traduce en el tiempo que perdura el recuadro en el frame principal y por ende los datos enviados por el puerto serial en $t+1$, que dada la situación, de presentarse otro sujeto en la escena seria la causal de un ruido eléctrico significativo que afectarían o sacarían al tracker de su punto de operación en lo que respecta al movimiento (pan-tilt)
- El tiempo de la detección está en el orden de los milisegundos pero depende de los valores mencionados, razón por la cual se realiza la estimación de acuerdo al tiempo de restablecimiento de la señal de 600 a 1000ms aproximadamente

En la estrategia de control y la comunicación:

- Como ya está incluida la estrategia de control en el algoritmo principal en conjunto con el protocolo de comunicaciones, en lo que respecta a las pruebas experimentales se verifico la habilitación del puerto "COM5" del ordenador, cada vez que se conectaba al puerto USB del mismo, lo que permitía hasta cierto punto una simulación de los datos enviados por cada fuente de video, gracias a la utilización del monitor serie del IDE asociado al microcontrolador
- Una vez esta verificada la habilitación del puerto y la recepción de los datos provistos por el algoritmo principal, se puede energizar el prototipo para que los servomotores operen la señal PWM asociado al dato correspondiente a la labor de detección
- No es posible con las herramientas utilizadas, operar el prototipo y ver los datos enviados simultáneamente, razón por la cual el criterio que se utilizó para corroborar el funcionamiento, era el tiempo de restablecimiento de la señal y el seguimiento de la persona en pantalla acorde a la ubicación del prototipo con respecto a la(s) persona(s)

En la operación y apagado/reset del sistema:

- En todo momento el sistema estuvo supervisado, y a la falta de un apagado de emergencia bastaba con una desconexión en los servomotores para prevenir incidentes en su mayoría con el cableado como se mencionó anteriormente
- El “apagado” del sistema se realiza de manera ordenada, esto por razones que van desde la mala ubicación inicial por parte de los servomotores, el “bloqueo” o errores en el puerto del ordenador y posterior cierre inesperado del algoritmo principal, hasta una ausencia parcial o total de movimiento en la estructura
- Dicho esto, el apagado comienza con la finalización del algoritmo principal presionando una tecla en el ordenador (*Esc*), acto seguido se desenergiza el tracker (servomotores y/o cámara) para finalmente desconectar el cable USB que representa la vía de comunicación serial entre el microcontrolador y el computador
- Si lo que se desea es seguir operando el sistema, el *reset* logra asemejarse al proceso de apagado de la siguiente manera: en primera instancia se presiona la tecla escape en el computador y se desenergiza el tracker, luego se presiona el botón reset del microcontrolador y sin desconectar el cable USB asociado al mismo, se verifica que aun el monitor serie (en el IDE del microcontrolador) lo tiene habilitado, inmediatamente después se energiza el tracker y se ejecuta el algoritmo principal, obviamente teniendo en cuenta las condiciones en las que la fuente del video, debe acorde al experimento mantenerse o cambiar

6.2. Restricciones

Detección

Pese a la precisión que posea el algoritmo de detección de personas al ser la integración de dos algoritmos que por separado presentan ciertas condiciones que no se acoplan a cabalidad con los objetivos del proyecto, o en contraste superan las expectativas de cada prueba realizada a manera operativa y general la detección presenta lo siguiente:

- No está en la capacidad de detectar a una sola persona, dicha restricción no fue implementada por diversas razones; las principales fueron corroborar que tan fiable es la comunicación por puerto serie, la medición y operación en las condiciones nominales de los actuadores, que ayudasen a determinar hasta qué punto es necesario o aplicable el control.
- El algoritmo de sustracción de fondo funciona de una manera más eficaz que su "asociado" el algoritmo de pre entrenamiento Haar, razón por la cual en vez de restringir o limitar su funcionamiento para que fuesen a la par, operativamente hablando, se optó por darle prioridad por sobre el algoritmo que clasifica al objeto detectado como persona, lo que derivó en una cantidad desproporcional de perturbaciones cuando el umbral (barra deslizante), está en valores cercanos al 0 lo que necesariamente hace que el operador tenga que realizar un ajuste manual de acuerdo a la escena a capturar o del sitio donde se va a ubicar el sistema.

- En ocasiones el algoritmo de detección captura y clasifica las personas de tal manera que se hacen casi imperceptibles los cambios que se efectúan a medida de que la persona se acerca o aleja de la cámara, lo que incrementa/disminuye notoriamente sus posibilidades de ser clasificado o no como una persona.
- El algoritmo de pre entrenamiento no capta en ocasiones a la persona (desde la cabeza hasta sus pies) debido a su posición o patrón natural de movimiento, lo que en ocasiones provoca que la probabilidad mostrada en el frame principal no corresponda a la esperada por su acercamiento o alejamiento con respecto a la cámara, factor que contribuyó a ubicar la cámara y tracker en el punto de lo posible en forma ortogonal a la base del emplazamiento principal (eje no rotativo de la cámara), lo que implica a su vez que la instalación del prototipo deba tener las mismas características.
- Cuando lo que se busca es una captura en tiempo real del sistema, con los ordenadores disponibles, la librería y la plataforma en general es bastante usual que presente problemas probablemente por el coste computacional e incompatibilidad en general de las funciones y/o hardware implementado, de allí que lo recomendable sea acudir a fuentes de video pregrabadas para validar su funcionamiento o bien de otro tipo de protocolos de comunicación, como el RTSP (disponible en el IDE de Python como otra herramienta proporcionada por OpenCV) y un dispositivo de captura externo (DVR), cuya configuración en la plataforma es rápida y resume en la siguiente figura con el segmento de código que acompaña la instrucción “cv2.VideoCapture(‘rtsp://...’)”.

```
rtsp://<Nombre de usuario>:<Contraseña>@<Dirección IP del equipo>:
<PuertoRTSP>/streaming/channels/<Numero de canal><Numero de flujo>
Nombre de usuario = Nombre de usuario del equipo (DVR/NVR/IPC) "admin"
Contraseña = Contraseña de acceso al equipo(DVR/NVR/IPC) "12345abc"
Direccion IP = Direccion IP del equipo(DVR/NVR/IPC) "192.168.1.64"
Puerto RTSP = Puerto de acceso al RTSP(por default 554) "554"
Streaming = Colocar la palabra Streaming
Channels = Colocar la palabra Channels
Numero de canal = Colocar el numero de canal de la cámara que desea obtener el
flujo de video "12"
Numero de flujo =          01 = Main stream, 02 = Sub stream
Ejemplo
rtsp://admin:12345abc@192.168.1.64:554/streaming/channels/1201 --> Esto obtiene el
main stream (flujo principal) de la cámara 12
```



Figura 39. Instrucción para obtener una fuente de video por el protocolo RTSP. Fuente (Jiménez N. & soporte.syscom.mx, actualizado 2021)

Tracking

Debido a la inestabilidad que presenta la comunicación serial con algunos actuadores, como se mencionó en su momento (estado del arte del presente documento), el proyecto desarrollado no

difiere del hecho, razón por la cual pese a que las simulaciones no mostraron ningún fallo notable o lectura errónea con su integración parcial en el algoritmo de detección, si presento perturbaciones importantes en el tracking, probablemente por la interpretación tardía de los datos suministrados por el puerto asociado. Entre los fallos destacan:

- El algoritmo sobrescribe una cantidad de veces el dato con respecto a la ubicación actual de la persona detectada y en ocasiones el último lugar donde estuvo, lo que ocasiono que la detección abrupta de otro sujeto en el marco de la imagen escribiese otra cadena tan rápidamente, que los 100ms dados por el fabricante en cuestión de velocidad en los servomotores sirviesen para poco o nada, si en términos de movimiento se tratase.
- Poner un tiempo de retardo en la recepción de los datos que realiza el microcontrolador para enviar la respectiva señal PWM a los servomotores pan-tilt, provocaba que por momentos al finalizar la detección y posterior desaparición del frame principal el servomotor siguiera moviéndose como si siguiera detectando, pese a que la fuente principal de video desapareciese junto con la habilitación del puerto (Python)
- Existe una fuente casi imperceptible de ruido cuando se envía el mismo pulso (tipo y forma de dato), lo que rara vez el tracker manifiesta como un "temblor", al igual de querer cambiar de posición y en casos aislados lo ejecutase, como si así lo hubiesen hecho tanto el detector como el microcontrolador utilizados.

Los fallos que se mitigaron fueron abordados aplicando una operación de reducción de ruido en la detección, en primera instancia por conteo de objetos/personas (Python-Spyder IDE), y en segunda instancia aplicando el tiempo de reconstrucción (metodología), tanto en el algoritmo de detección como en el rastreador simultáneamente, con el fin más que de ajustar, calibrando lo que el ojo dictamina como un seguimiento coherente y fluido.

6.3. Validación de la solución

Para la parte de detección y tracking de personas se realiza el siguiente registro de comunicación (control de realimentación) con el monitor serial y el mapeo del frame principal, los videos fueron grabados en el centro comercial (Gran Plaza el Ensueño), y se escogió uno en particular por presentar los patrones que permiten medir y/o resumir de forma generalizada el proceso, esto puede evidenciarse en el video en cuestión y en las siguientes figuras (no hubo cambios de alpha o T), esto con el propósito de trazar "virtualmente" las rutas de los objetivos.

c	c	e	e	g	e	e	e	i	e	d	d	d	d	d	d	d	d	d	d	1	
c	g	e	e	e	e	e	e	e	e	d	d	d	d	d	c	d	d	d	d	2	
c	h	e	e	e	e	h	e	i	d	d	d	d	d	d	d	d	d	d	d	3	
c	h	e	e	e	e	g	e	e	d	d	d	d	d	d	d	d	d	d	d	4	
i	e	e	e	e	e	g	e	e	d	d	d	d	d	c	d	d	d	d	d	5	
f	e	i	e	e	e	e	e	e	d	d	d	d	d	d	d	d	c	d	d	6	
c	g	e	e	e	e	e	e	i	d	d	d	d	d	d	d	d	d	d	d	7	
e	e	e	e	e	e	c	e	e	d	d	d	d	d	d	d	e	d	d	d	8	
i	e	f	e	e	e	e	e	e	d	d	d	d	d	d	d	d	d	d	e	9	
e	e	e	e	e	h	e	e	e	d	d	d	d	d	d	d	d	d	d	d	10	
1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	Orden de muestras

Figura 40. Representación de los datos por simulación del puerto serie servomotor panorámico.
Fuente (Imagen propia)

l	l	l	l	s	k	r	r	m	s	o	l	t	q	r	n	n	o	o	p	q	r	1
l	l	l	l	l	k	l	l	n	s	n	p	o	q	r	n	n	o	o	p	q	r	2
l	l	l	l	p	q	l	s	m	m	k	k	p	q	r	n	n	o	o	q	q	r	3
l	l	l	l	q	k	l	l	n	n	m	p	p	q	s	n	n	o	p	p	q	r	4
l	l	l	k	p	q	r	l	p	n	p	n	p	r	s	n	o	o	p	q	q	r	5
l	l	l	k	m	q	r	l	m	n	o	n	p	q	s	n	o	o	p	p	r	r	6
l	l	l	l	q	q	r	l	n	l	p	k	p	r	t	n	o	o	p	p	q	s	7
l	l	l	l	q	k	l	s	m	o	p	s	p	r	m	n	o	o	p	s	q	r	8
l	l	l	l	k	k	l	m	n	k	n	s	q	r	t	n	n	o	p	q	r	s	9
l	l	l	l	q	q	l	m	n	n	n	s	q	r	t	n	n	o	p	q	r		10
1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	Orden de muestras

Figura 41. Representación de los datos por simulación del puerto serie servomotor de inclinación.
Fuente (Imagen propia)

Ahora bien para el desarrollo e implementación de un controlador PID o sus variantes (junto con las estrategias de control asociadas), es necesario un setpoint (referencia) en el centro del marco de la imagen (eje óptico de la cámara), y una señal de error que sea manejada en naturaleza de la forma del sistema, por supuesto que esta no solo ayude a trazar la posible ruta de las personas rastreadas si no también realizar el ajuste por medio del pan-tilt de la cámara lo que a su vez brinda información acerca del cuadrante (figura 35) en el cual se encuentra la persona y dar un primer dictamen de comportamiento en la integración del todo el sistema (datos relevantes tales como constantes, tiempos, perturbaciones entre otros).

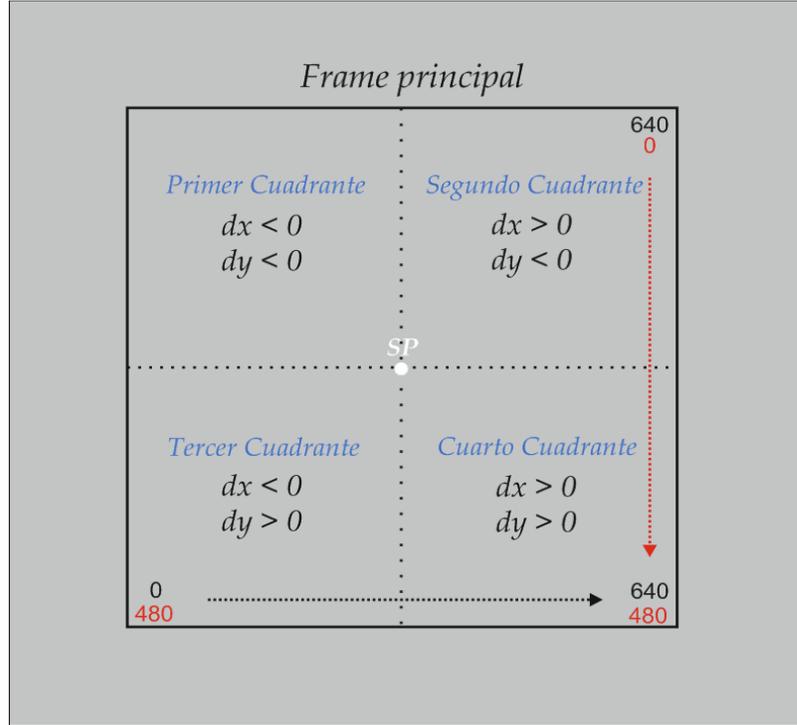


Figura 42. División teórica por cuadrantes del frame principal acorde a la señal de error (dx, dy).
Fuente (Imagen propia)

En el siguiente conjunto de graficas (figura 43), no se ha cambiado la velocidad de entrenamiento (la máxima), pero si se ha realizado con el umbral, para demostrar, que a un coste de disminuir o mejorar la señal de error, así mismo es alterada negativa y recíprocamente la labor de detección en el FOV de la cámara.

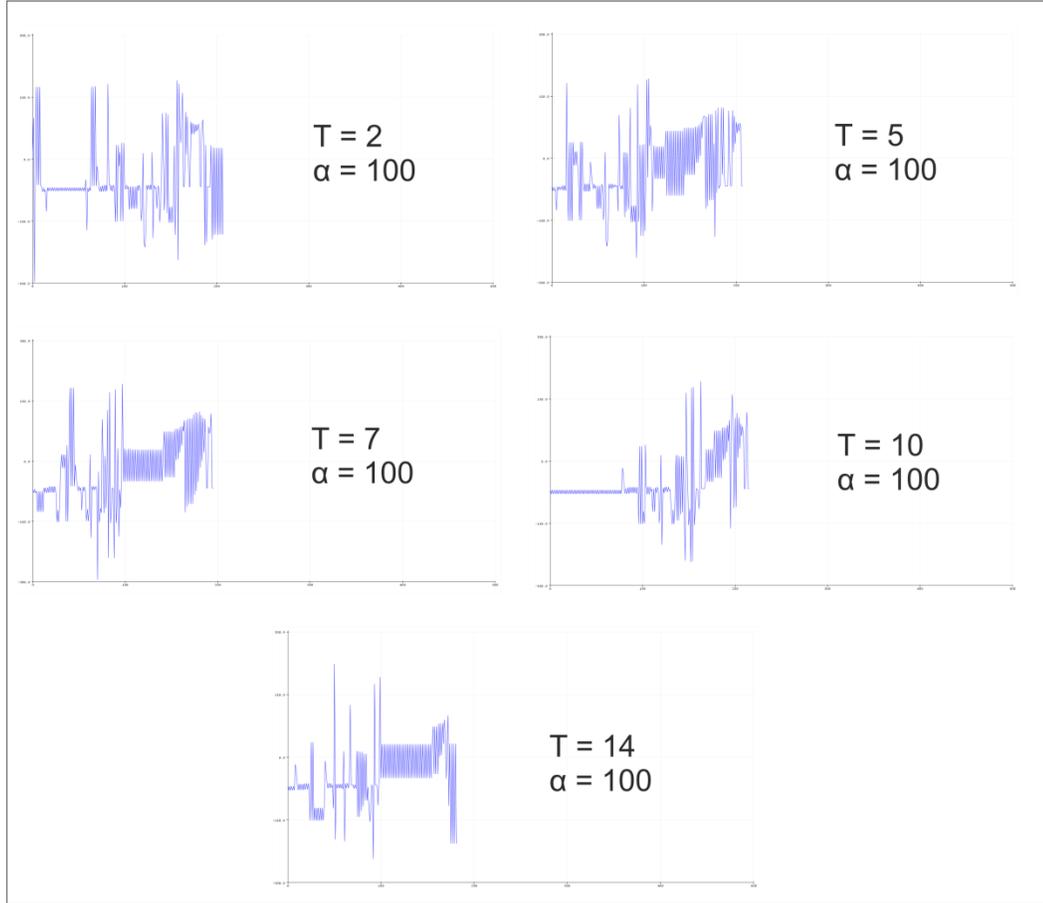


Figura 43. División teórica por cuadrantes del frame principal acorde a la señal de error (dx, dy) .
Fuente (Imagen propia)

Lo anterior es un factor importante si se desea abordar el control proporcional, sin tener en cuenta la acumulación o tasa de cambio de la señal de error.

6.4. Evaluación del impacto de la solución propuesta

Como apoyo a sistemas de videovigilancia o seguridad por medio de una cámara estática, el prototipo muestra una evidente solución a lo que implica el tracking dinámico con un aporte al creciente y demandante mundo de las cámaras que amplían el rango de visión, aunque poniéndolo en términos más técnicos representa claro impacto por los siguientes puntos:

- Es establecida la integración de dos algoritmos para conformar uno solo capaz de bajo ciertas condiciones (tiempo de reconstrucción, análisis trama de datos, ajuste threshold...), puede funcionar adecuadamente en espacios reales (cerrados o abiertos sin muchos cambios en la escena), lo que cierra sobre una base de los conceptos claves en el tracking de un objetivo en particular

- Es posible validar el funcionamiento de todas las etapas descritas con anterioridad gracias al sistema prototipo, con la posibilidad de realizar pruebas en espacios reales o bien cargando un archivo de video diferente cada vez en el IDE Spyder para comparar distintas formas de funcionamiento.

- La señal de error establecida y receptada por el microcontrolador incrementa aún más la posibilidad de desarrollar un controlador que haga más fluido el movimiento lo que traduce en mejores tiempos de comunicación, reducción drástica de perturbaciones entre las etapas y aumento en la probabilidad de clasificación de personas en la escena captada (FOV)

7. Conclusiones y Recomendaciones

Se definió e implemento un algoritmo de detección que involucro la integración de dos enfoques distintos debido a que por aparte no cumplían los requerimientos necesarios, por un lado clasificar a un objeto como persona y por el otro no contar con los elementos necesarios que permitiesen fijar y seguir a las personas de manera coherente.

Es corroborado experimentalmente lo que la labor de tracking en la teoría expresa; es facilitada si se cuenta con las herramientas de detección adecuadas y en el caso particular del sistema prototipo, es implementada una estrategia de control que entre otras cosas propicia una mitigación de las perturbaciones y una no linealidad.

La integración del sistema fue posible por el conjunto de elementos recolectados en el análisis de comportamientos por etapa, que a su vez dieron las pautas para realizar la consigna y calibración del sistema previa a su diseño final, lo que eventualmente formularia una base metodológica para otros sistemas de apoyo a videovigilancia o seguridad.

Cuando se realizó la detección de objetivos que se movían por sobre una escena para clasificarse como personas (prototipo), es propicio medir los alcances y operatividad de todo el sistema en vez de comprobarlo por etapas e ir acoplado sistemas de eliminación de perturbaciones, ya que en ellas como se demostró, también reside una parte importante de la señal captada por el sensor de la imagen.

Como cualquier sistema de visión artificial las perturbaciones medioambientales juegan un papel relevante en el diseño del mismo, por lo tanto, la calibración de estos sistemas deben poderse adaptar a los cambios de iluminación, brillo y sombras del ambiente. Por tal motivo aunque los algoritmos de media y Haar tienen sus puntos de compensación internos propios, el sistema cuenta además con umbrales ajustables por el usuario.

Tener en cuenta el costo computacional en todas las etapas del proyecto, se convirtió en un factor clave a la hora de probar el algoritmo en otro tipo de ordenador, razón por la cual la compensación o cambios de perturbaciones se acentúan y evidencian a medida que se muestra el video y las funcionalidades que al frame principal acompañan.

8. Referencias

- [1] Santos, D., Dallos, L., & Gaona Garcia, P. (2020). *Algoritmos de rastreo de movimiento utilizando técnicas de inteligencia artificial y machine learning* [Article pdf] (1st ed., pp. 1-11). Bogotá D.C. - Colombia.
- [2] Mosquera Victoria, C., & Dussan Narvaez, G. (2020). *DETECCIÓN Y SEGUIMIENTO DE MÚLTIPLES OBJETOS EN TIEMPO REAL PARA VEHÍCULOS AUTÓNOMOS*. [tesis pdf] (pp. 35-56). Cali-Colombia.
- [3] Espinosa Oviedo, J. (2019). *Detection And Tracking of Motorcycles in Urban Environments by using Video Sequences with High Level of Oclusion* [tesis pdf] (pp. 41-94). Medellin-Colombia.
- [4] Pinzon Florez, C. (2010). *SEGUIMIENTO DE OBJETOS MÓVILES USANDO UNA CÁMARA PTZ* [tesis pdf] (pp. 11-29). Bogotá D.C- Colombia
- [5] Imbachí, F., Montenegro, D., Fierro, L., & Bolaños, D. (2011). Sistema de seguimiento bidimensional mediante técnicas de control digital y visión computacional. *RCF Revista Colombiana De Física*, (Vol.1), 1-4.
- [6] Ortiz Valencia, N., & Vargas Londoño, L. (2016). *SISTEMA PARA EL SEGUIMIENTO DE UN OBJETO IMPLEMENTADO EN LA CABEZA DEL ROBOT INMOOV* [tesis pdf] (pp. 38-63). Bogotá D.C- Colombia.
- [7] Mathews, A., Jacob, R., Biju, R., & Joseph, U. (2018). *Object Tracking and Image Stabilization using PixyCam and Arduino* [Article pdf] (6th ed., pp. 1-4). Kerala-India.
- [8] Sáenz Esqueda, J., Sáenz López, A., Reyes Espino, V., & Morales Salazar, R. (2014). Seguimiento de Objetos empleando Aforge.Net y Arduino. *Revista De Arquitectura E Ingeniería*, (Vol.8), 1-12.
- [9] Garcés Matilla, A., Guerrero Peña, F., Garcés Calvelo, A., Hernández Silva, E., & Matilla Belett, M. (2017). *Seguimiento de múltiples personas en aplicación de videovigilancia con cámaras de baja resolución* [article pdf] (11th ed., pp. 1-13). Santiago de Cuba-Cuba: Revista Cubana de Ciencias Informáticas.
- [10] Iqbal, D., & Lakshmi, K. (2017). *OBJECT DETECTION AND TRACKING USING THERMAL CAMERA* [Journal pdf] (14th ed., pp. 1-10). Madnapalle- India: International Journal of Pure and Applied Mathematics.
- [11] Nasiri, N. (2006). *Camera-based 3D Object Tracking and Following Mobile Robot* [article pdf] (pp. 1-6). Sarawak Campus-Malaysia.

- [12] Ölmez, H. (2013). *DESIGN AND IMPLEMENTATION OF A HEAD TRACKING CONTROLLED PAN AND TILT VISION SYSTEM* [tesis pdf] (pp. 5-62). Ankara-Turquía.
- [13] Chaitra, J., & Kumar, D. (2018). *Smart Autonomous Camera Tracking System Using myRIO With LabVIEW* [paper pdf] (7th ed., pp. 1-6). Bangalore-India: American Journal of Engineering Research (AJER)
- [14] Pajor, M., Grudziński, M., & Marchewka, Ł. (2018). *Stereovision System for Motion Tracking and Position Error Compensation of Loading Crane* [article pdf] (pp. 1-11). Szczecin-Poland: AIP Conference Proceedings 2029, 020050.
- [15] Maaß, F. *An Analysis on Object Detection and Tracking with a Tilt-Pan Camera on an Embedded Device* [paper pdf] (pp. 1-12). Hamburgo-Alemania.
- [16] Hamsa, D., Verma, R., & Shivaram, K. (2016). *REAL TIME OBJECT TRACKING BY ROBOT USING KALMAN FILTER* [article pdf] (5th ed., pp. 1-7). MSRIT-India: International Journal of Advance Research in Science and Engineering.
- [17] Ruangpayoongsak, N. (2017). *Mobile Robot Positioning by using Low-Cost Visual Tracking System* [article pdf] (pp. 1-5). Bangkok-Thailand.
- [18] Duarte B, K. and Borrás P, C., (2016). Vista de Generalidades de robots paralelos. *Visión electrónica*, vol. 10, no. 1, pp. 102-112. <https://doi.org/10.14483/22484728.11711>
- [19] Cruz-Carbonell, V., & Castillo-Esteba, R. A. (2019). Ensamblaje e implementación de arquitectura cuadrúpeda modular. *Visión electrónica*, 13(2), 280-288. <https://doi.org/10.14483/22484728.15184>
- [20] Jiménez Moreno, R., Avilés, O., & Ovalle, D. M. (2018). Red neuronal convolucional para discriminar herramientas en robótica asistencial. *Visión electrónica*, 12(2), 208-214. <https://doi.org/10.14483/22484728.13996>
- [21] Y. Shih, F. (2010). *IMAGE PROCESSING AND PATTERN RECOGNITION Fundamentals and Techniques* (pp. 3-15). Canada.
- [22] Acharya, T., & Ray, A. (2021). *Image Processing Principles and Applications* (pp. Caps 1 & 6). Canada: Copyright by John Wiley & Sons, Inc.
- [23] Bradski, G. & Kaehler A. (2008). *Learning OpenCV* (1st ed.). California-US: O'Reilly Media, Inc.
- [24] Bradski, G. & Kaehler A (2016 Revision). *Learning OpenCV* (1st ed.). California-US: O'Reilly Media, Inc.

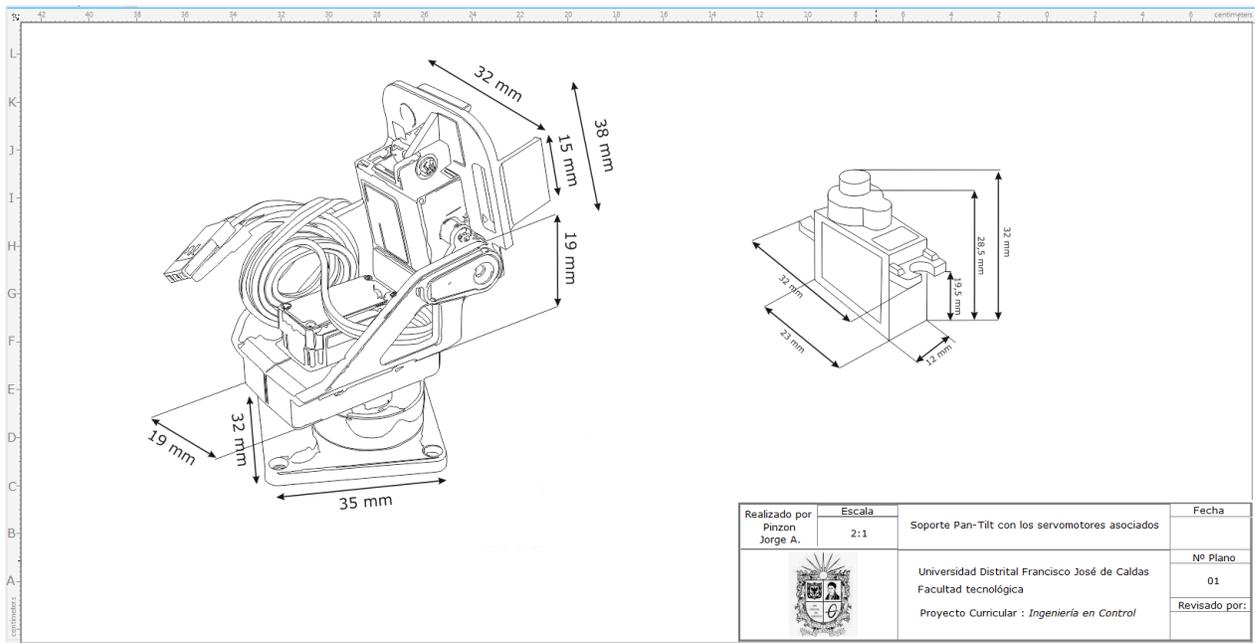
- [25] Sailema Chicaiza, F. (2017). *SISTEMA ELECTRÓNICO DE ALERTA AUTOMÁTICA PARA EL RECONOCIMIENTO DE SEÑALES DE TRÁNSITO REGLAMENTARIAS, PREVENTIVAS E INFORMATIVAS EN LA CIUDAD DE AMBATO* [Tesis pdf] (pp. 35-42). Ambato-Ecuador.
- [26] Desai, M., & Gandhi, M. (2014). *A Survey: Background Subtraction Techniques* [article pdf] (5th ed., pp. 1-3). International Journal of Scientific & Engineering Research.
- [27] UNAM. Posiciones Relativas de Dos Circunferencias. Retrieved 4 August 2021, from http://prepa8.unam.mx/academia/colegios/matematicas/paginacolmate/applets/matematicas_V/Applets_Geogebra/posirel2circ.html#:~:text=Exteriores%3A%20Si%20no%20tienen%20ning%C3%BAn,Tienen%20dos%20puntos%20en%20com%C3%BAn. [Website Schulz]
- [28] Sonnenberg, J. *Serial Communications RS232, RS485, RS422* [Technical Brief] (pp. 1-6). California-US: Raveon Technologies Corp.
- [29] Costilla, D., & Reaño Montoro, S. (2008). Streaming de Audio/Video. Protocolo RTSP [Article pdf] (pp. 15-18). Magdalena Payeras Capellà. Retrieved from http://ibdigital.uib.es/greenstone/sites/localsite/collect/enginy/index/assoc/Enginy_2/008v01p0/15.dir/Enginy_2008v01p015.pdf
- [30] Marlin, T. (2015). *Process Control Designing Processes and Control Systems for Dynamic Performance* (2nd ed., pp. 3-20). McGraw-Hill.
- [31] Reed, D. "The Six" Fundamental Control Strategies Every Process Control Developer Must Know [presentation pdf] (pp. 4-25). ISA Setting the Standard for Automation.
- [32] Perez., M., Hidalgo, A., & Perez Berenguer., E. (2007). *INTRODUCCION A LOS SISTEMAS DE CONTROL Y MODELO MATEMÁTICO PARA SISTEMAS LINEALES INVARIANTES EN EL TIEMPO* [Ebook] (pp. 1-17). San Juan - Argentina.
- [33] Batrinu, C. (2017). *ESP8266 Home Automation Projects* (pp. 30-69). BIRMINGHAM-MUMBAI: Packt Publishing Ltd.
- [34] Espressif Systems. (2015). *ESP8266EX Datasheet* [Datasheet pdf] (pp. 6-15). Espressif Systems IOT Team. Retrieved from https://components101.com/sites/default/files/component_datasheet/ESP8266-NodeMCU-Datasheet.pdf
- [35] NodeMcu Team. (2014). NodeMcu -- An open-source firmware based on ESP8266 wifi-soc. Retrieved 4 August 2021, from https://www.nodemcu.com/index_en.html
- [36] MINCOMERCIO Industria y Turismo. (2016). *Protección de datos personales en sistemas de videovigilancia* [Ebook] (pp. 1-14). Colombia: Superintendencia de industria y comercio.

- [37] AEPD. *Guía sobre el uso de videocámaras para seguridad y otras finalidades* [pdf] (pp. 6-9). España: Agencia española de protección de datos.
- [38] Mallick, S. (2018). Find Center of a Blob (Centroid) Using OpenCV (C++/Python) | LearnOpenCV. Retrieved 4 August 2021, from <https://learnopencv.com/find-center-of-blob-centroid-using-opencv-cpp-python/>
- [39] Arduino. (2021). How to use the Serial Monitor with the Arduino IDE 2.0. Retrieved 4 August 2021, from <https://www.arduino.cc/en/Tutorial/getting-started-with-ide-v2/ide-v2-serial-monitor/>
- [40] Vásquez Chepillós, N. (2019). *RELEVANCIA DEL ÁREA DE GRÁFICA Y DISEÑO DE UNA EMPRESA MANUFACTURERA* [tesis pdf] (pp. 1-41). VALPARAÍSO – CHILE.
- [41] Del Castillo Rodríguez, F. (2015). *NORMAS BÁSICAS PARA DIBUJO TÉCNICO* [pdf] (pp. 1-32). Cuautitlán (Ciudad de México)-México.

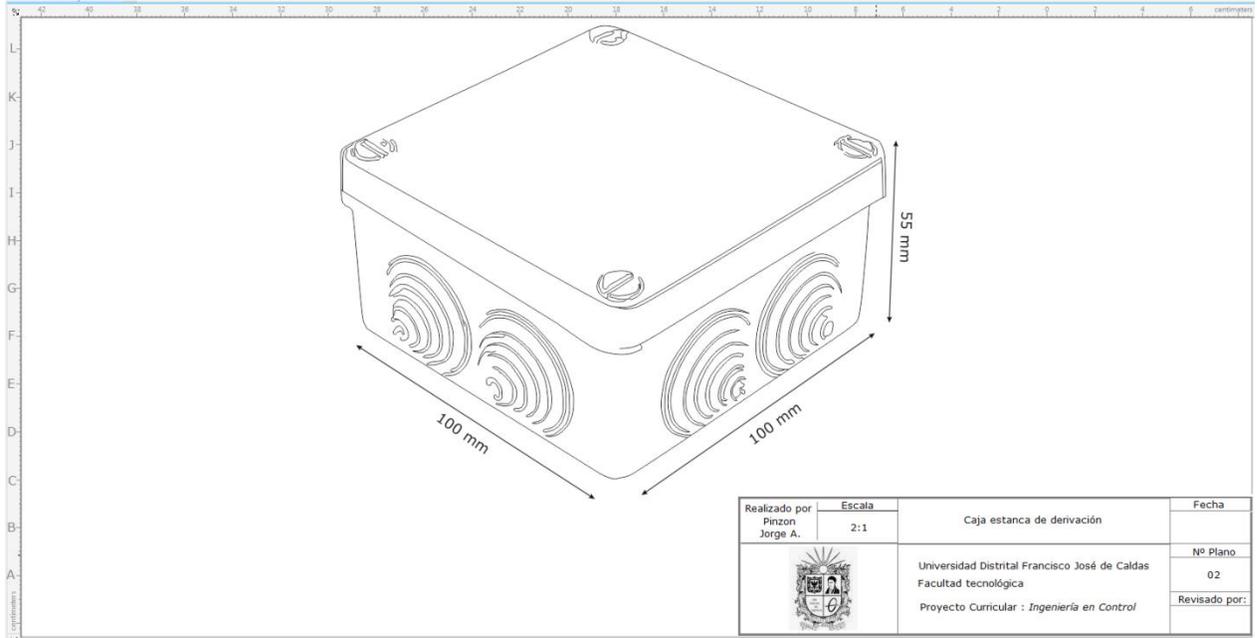
9. Anexos

Anexo 1

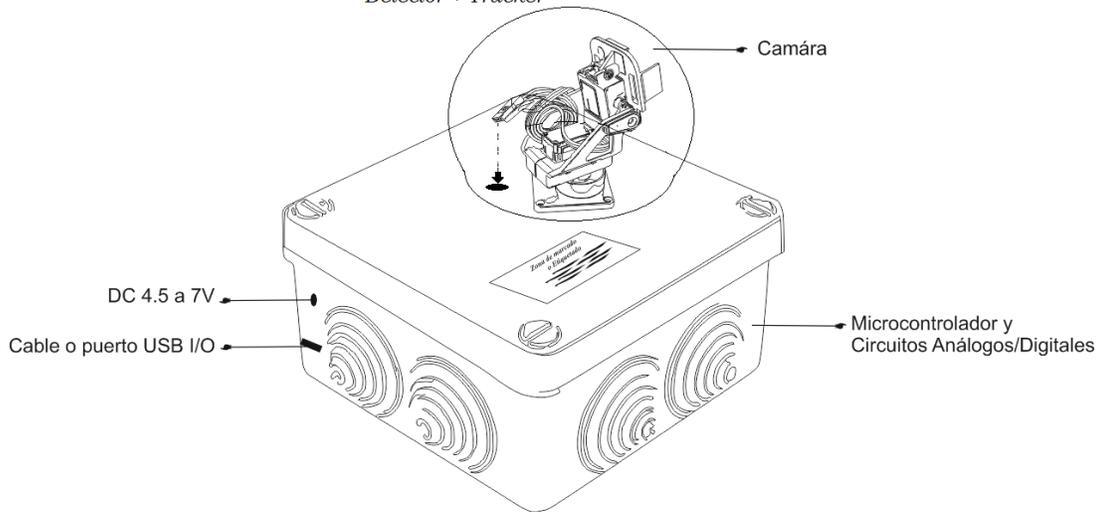
A continuación se consigna el **primer concepto** que podría utilizarse para validar el funcionamiento del prototipo, dichos bosquejos fueron rotulados bajo la normativa IRAM y el tipo de letra corresponde a las recomendadas por la normativa de dibujo técnico. [40] [41]



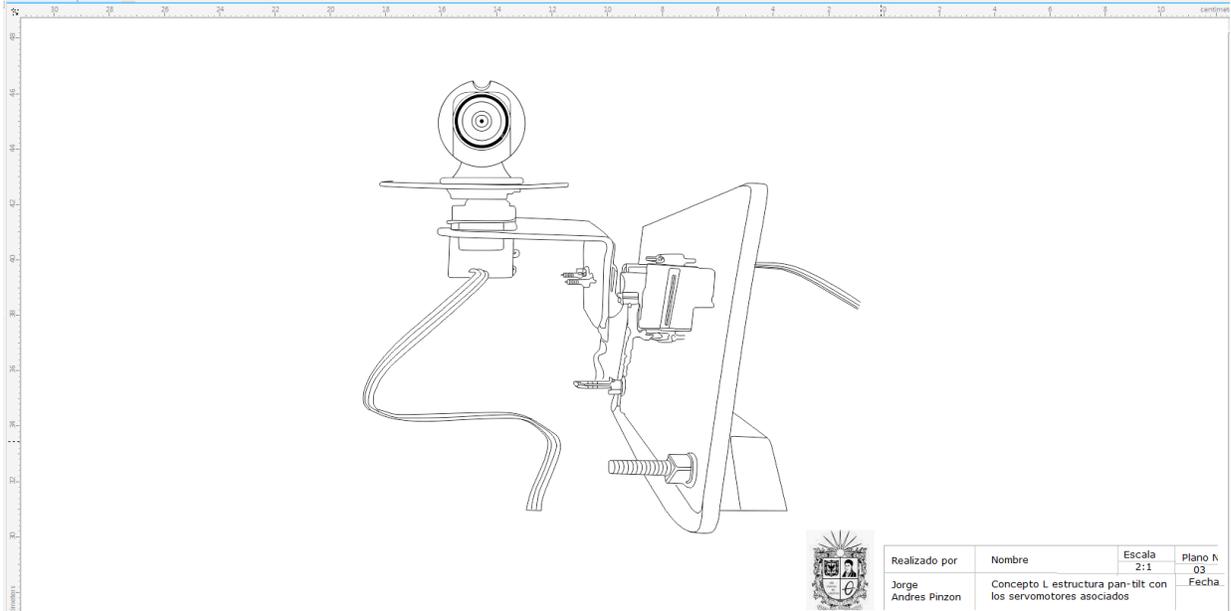
Aquí se puede ver tanto el soporte del servomotor ya construido y el SG90 respectivamente, contienen las dimensiones (escala mm), que obviamente sirvieron tanto para escoger el contenedor de los circuitos y el tipo de protección que se le va a hacer al prototipo.



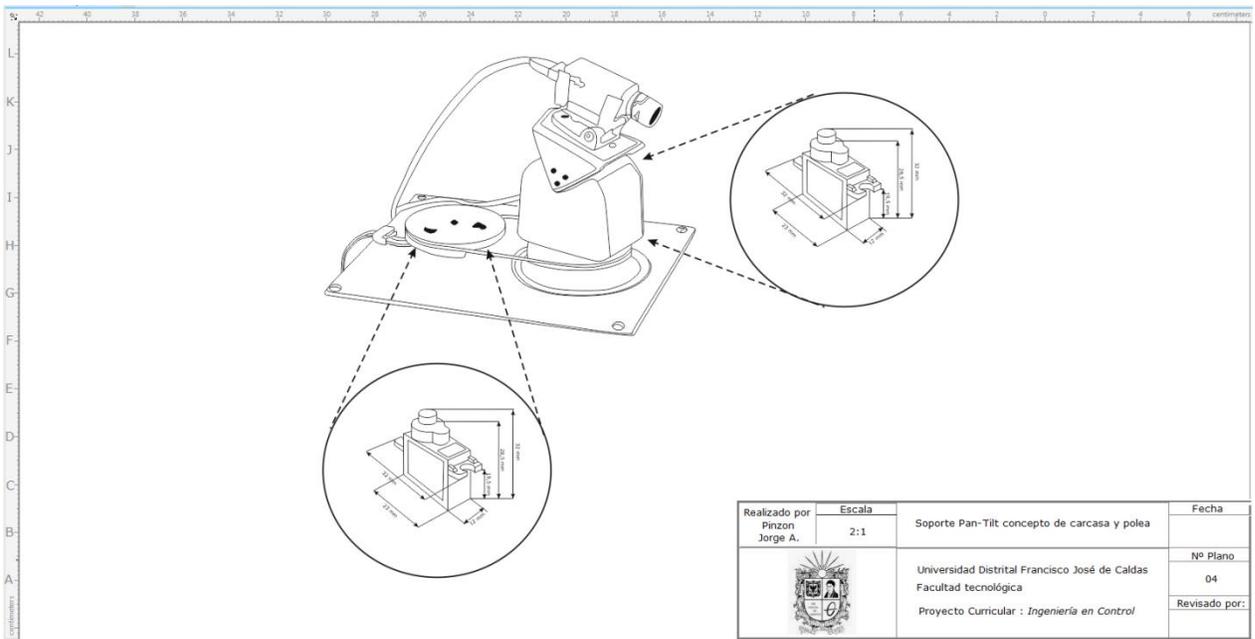
*Concepto final para la primera alternativa
Detector + Tracker*



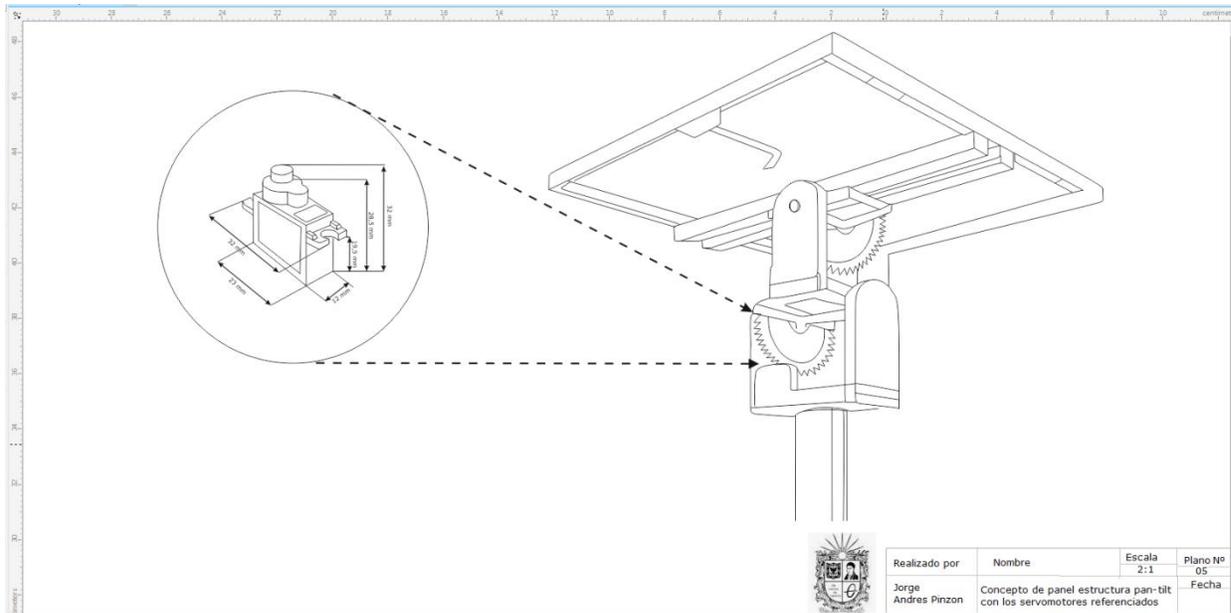
Para el **segundo concepto** se puede realizar una estructura que aproveche el principio de los cardanes más básicos (concepto L) para ubicar la cámara y servomotores de una forma cuasi rustica en una pared usualmente de madera ya que se acostumbra utilizar tornillos.



En el **tercer concepto** aparecen tal y como se muestra en la figura un sistema adaptado de polea que le facilita el movimiento panorámico a la cámara y de inclinación, aquí la cámara queda expuesta pero posee una carcasa que protege los servomotores (internos) tal y como lo muestra el siguiente bosquejo.



En el **último concepto** de estructura sugerida para el proyecto desarrollado, se aprovecha el principio que tienen algunos paneles solares para otorgar más libertad de movimiento, aunque esto comprometa en parte su funcionamiento o seguridad (avería por lectura errática del SG90); tal y como lo muestra el bosquejo los engranajes podrían reemplazarse con un eje acoplado al actuador respectivo y con cierta pericia del diseñador.



Referencias anexo 1

[40] Vásquez Chepillos, N. (2019). *RELEVANCIA DEL ÁREA DE GRÁFICA Y DISEÑO DE UNA EMPRESA MANUFACTURERA* [tesis pdf] (pp. 1-41). VALPARAÍSO – CHILE.

[41] Del Castillo Rodriguez, F. (2015). *NORMAS BÁSICAS PARA DIBUJO TÉCNICO* [pdf] (pp. 1-32). Cuautitlán (Ciudad de México)-México.