



TEC2014-53176-R HAVideo (2015-2017)

High Availability Video Analysis for People Behaviour Understanding

D1.1 v2

System infrastructure

Video Processing and Understanding Lab

Escuela Politécnica Superior

Universidad Autónoma de Madrid

AUTHORS LIST

Juan Carlos SanMiguel

Juancarlos.sanmiguel@uam.es

HISTORY

Version	Date	Editor	Description
0.9	19 December 2015	Juan Carlos San Miguel	Final Working Draft
1.0	20 December 2015	José M. Martínez	Editorial checking
1.9	22 June 2017	Juan Carlos San Miguel	Final Working Draft v2
2.0	26 June 2017	José M. Martínez	Editorial checking

CONTENTS:

1. INTRODUCCIÓN	1
1.1. ESTRUCTURA DEL DOCUMENTO	1
2. INFRAESTRUCTURA DE CAPTURA	2
2.1. CÁMARAS FIJAS	2
2.2. CÁMARAS PTZ (PAN-TILT-ZOOM)	3
2.3. CÁMARAS DE VISIÓN PORTÁTILES	5
2.4. CÁMARA DE PROFUNDIDAD	5
2.5. SISTEMA DE CAPTURA CON LUZ ESTRUCTURADA LÁSER INFRARROJA	6
<i>Emisión de luz estructurada infrarroja</i>	6
<i>Captura de luz estructurada infrarroja</i>	7
2.6. CÁMARAS DE VISIÓN PORTÁTILES CON BATERÍA	7
3. INFRAESTRUCTURA DE PROCESO	10
3.1. EQUIPAMIENTO	10
3.2. SUBSISTEMAS SOFTWARE DE BASE	11
3.3. ARQUITECTURA DE PROCESO DÍVA	11
3.3.1. <i>Módulos</i>	12
3.4. ARQUITECTURA DE PROCESO SMCS	14
3.4.1. <i>Módulos</i>	15
3.4.2. <i>Comunicación</i>	18
3.4.3. <i>API para la funcionalidad Android</i>	19

1. Introducción

En este documento se describe la infraestructura de la que dispone actualmente el VPULab para llevar a cabo los desarrollos propuestos en el marco del Proyecto HAVideo.

1.1. Estructura del documento

El documento está estructurado de la siguiente manera:

- Capítulo 1, “Introducción”: describe el objetivo del presente documento
- Capítulo 2, “Infraestructura de captura”.
- Capítulo 3, “Infraestructura de proceso”.

2. Infraestructura de captura

Esta sección describe los dispositivos con que cuenta el Grupo para capturar secuencias de vídeo, tanto en lo referente a cámaras de diversos tipos como a escenarios y sistemas de iluminación.

2.1. Cámaras fijas

La infraestructura actual de cámaras fijas data del año 2005, en que se instaló un sistema de tres cámaras de visión artificial en dos pasillos de acceso de la Escuela Politécnica Superior (ver Fig. 1). Uno de los accesos tiene dos cámaras, para considerar aplicaciones de estéreo-visión; el otro tiene una única cámara con distancia focal variable, lo que permite variar la región de interés. Ambos consideran escenarios con iluminación natural, no artificial, escenarios especialmente adecuados para el conteo de personas, ya que su uso habitual es como pasillos que comunican dos edificios de la Escuela.

Se trata de cámaras digitales, modelo DFW-X710 de SONY, dos de ellas con óptica fija y la tercera con óptica variable. Las cámaras envían secuencias de imágenes sin comprimir (1024x768, color, 15 fps.) a través de un bus IEEE1394, extendido mediante un enlace de fibra óptica (GOF), a un bastidor de PCs. Cada PC recibe la señal de hasta dos cámaras. El bastidor interconecta los PCs de recepción con un servidor de disco a través de una red Gigabit Ethernet, conectada a Internet y protegida por un Firewall.

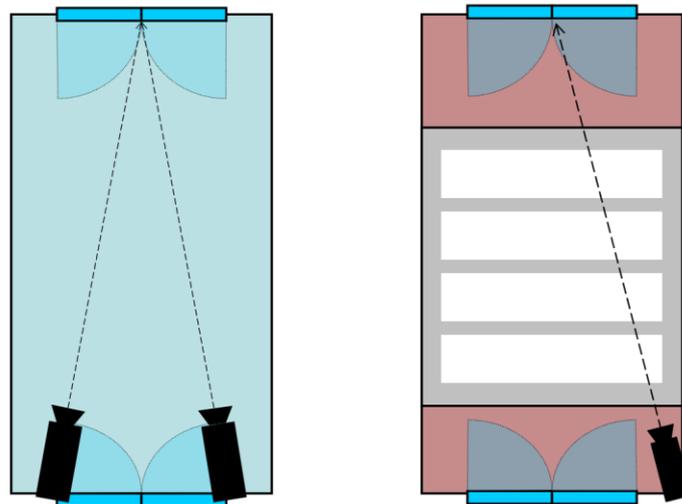


Figura 1. Esquemas de colocación de las cámaras fijas en planta 3 (izda.) y planta baja (dcha.) en la EPS



Figura 2. Situación de las cámaras fijas de la planta 3 (izda.) y detalle de una de las cámaras montadas (dcha.)

2.2. Cámaras PTZ (Pan-Tilt-Zoom)

La infraestructura actual data del año 2006, en que se instaló un sistema de tres cámaras PTZ en el vestíbulo de la Escuela Politécnica Superior (ver esquema en Fig. 3, aspecto del área cubierta en Fig. 4, y detalle de una cámara en Fig. 5). El escenario que considera esta instalación es múltiple. Por una parte, un escenario en el que se graba una zona amplia y no diáfana (por motivo de las cuatro columnas centrales) con alta densidad de personas circulando por ella y con iluminación parcialmente natural, ya que el frontal del vestíbulo está acristalado; por otra, teniendo en cuenta que las cámaras laterales pueden orientarse hacia el extremo contrario al vestíbulo, que presenta dos pasillos de acceso a las aulas, se contempla un escenario adecuado para el conteo en situaciones de densidad media y con iluminación eminentemente artificial. En ambos casos la alta reflectividad del suelo plantea complicaciones adicionales al análisis.

Se trata en este caso de cámaras IP digitales PTZ, modelo SNC-RZ50P de SONY. Presentan varios modos de funcionamiento y su posición y distancia focal es controlable por software.

Las cámaras envían secuencias de imágenes comprimidas (JPEG) o directamente un stream de vídeo (MPEG4 o H264) a través de una red Ethernet, a un router Gigabit Ethernet, conectada a Internet y protegido por un Firewall.

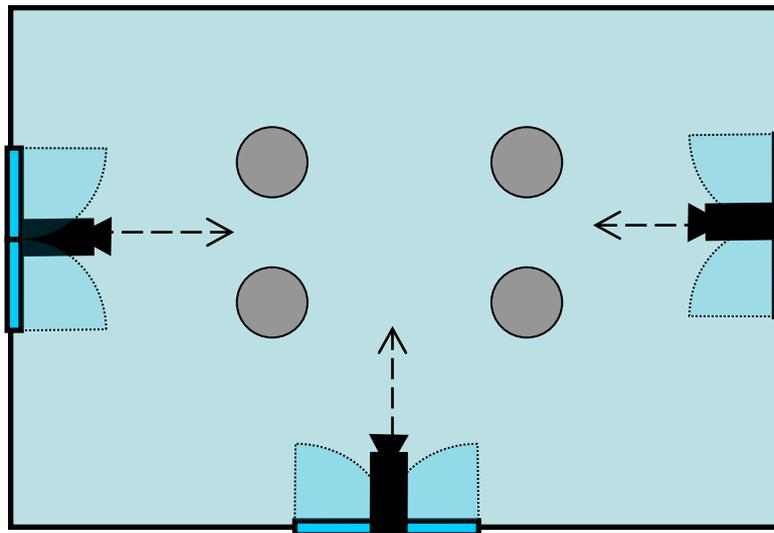


Figura 3. Esquema de colocación de las cámaras PTZ en el hall de la EPS

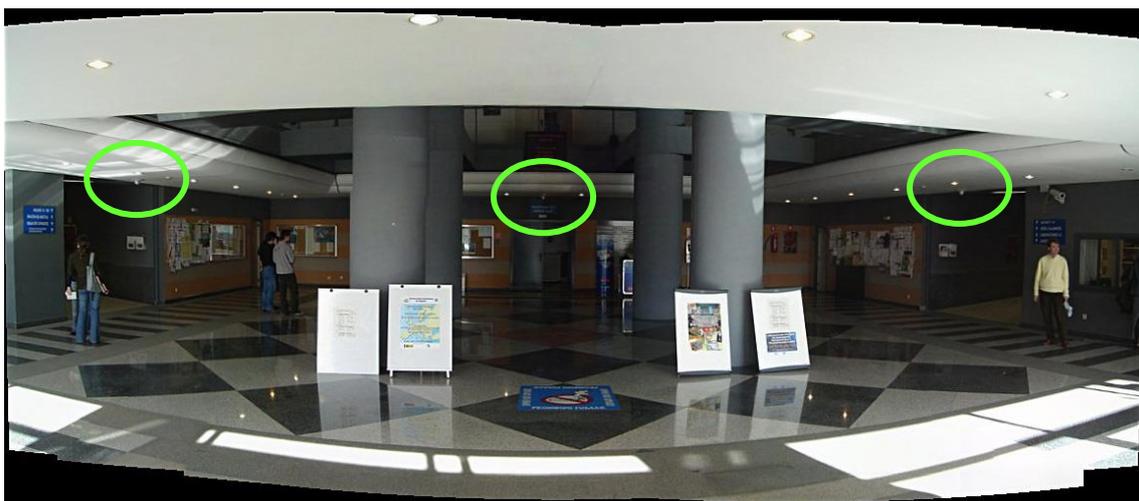


Figura 4. Situación de las cámaras PTZ (izda.) y detalle de una de las cámaras montadas (dcha.)



Figura 5. Detalle de una de las cámaras PTZ situada en el techo del vestíbulo de la Escuela.

2.3. Cámaras de visión portátiles

La infraestructura actual de cámaras de visión portátiles data del año 2007. Consta de dos parejas de cámaras de alta resolución (1600x1200 y 1900x1080) y calidad (captura sin compresión), así como juegos de lentes y de soportes para configurar un equipo de captura multicámara portátil. El esquema está especialmente indicado para escenarios reducidos, del tipo de salas inteligentes, sistemas de visión estéreo, etc.

Se trata de cámaras digitales, modelos Pioneer piA 1600-35gm/gc y piA 1900-32gm/gc de BASLER (ver detalle en Fig. 6). Presentan varios modos de funcionamiento en cuanto a resolución y tasa de cuadro.

Las cámaras envían secuencias de imágenes no comprimidas a través de una red Gigabit Ethernet, normalmente directamente a un equipo portátil.



Figura 6. Detalle de una de las cámaras de visión de alta resolución de BASLER.

2.4. Cámara de profundidad

Aparte de los dispositivos ya descritos, el Grupo dispone de dispositivos de captura especiales que pueden ayudar a resolver problemas de visión en situaciones o contextos complejos.

Uno de ellos es el dispositivo Kinect para Xbox 360 de Microsoft © (ver detalles en Fig. 7). Desde el punto de vista de visión, el dispositivo permite obtener imágenes RGB de una escena, incluyendo para cada píxel la información de la profundidad a la que se encuentra del dispositivo de captura.



Figura 7. Detalle de una de cámara Kinect y de su colocación.

2.5. Sistema de captura con luz estructurada láser infrarrojo

El Grupo dispone de un sistema de emisión de luz estructurada no visible (basada en el uso de un láser infrarrojo y de rejillas de difracción), y una cámara para capturar dicha luz (cámara convencional sin filtro de infrarrojos y con un filtro paso-banda centrado en la banda del láser).

Emisión de luz estructurada infrarrojo

La fuente de luz es un láser infrarrojo. Se trata del modelo ZM-18H de la empresa Z-Laser (ver figura 8). Trabaja a una longitud de onda de 808 nanómetros, correspondiente al rango de frecuencias infrarrojas, que son invisibles al ojo humano, lo que evita que interfiera en la escena sobre la que actúa. Su potencia nominal de 100 mW es elevada, lo que exige tener en cuenta los valores de seguridad de la normativa referente a los dispositivos láser. Se ha desarrollado un dispositivo que permite regular la potencia que emite el láser.



Figura 8. Detalle de Z-Laser infrarrojo .

Para la obtención de un patrón de luz estructurada se coloca al láser una lente que idealmente tiene la banda de paso en su misma longitud de onda (para evitar pérdidas). La Fig. 9 muestra ejemplos de los patrones de que se dispone.

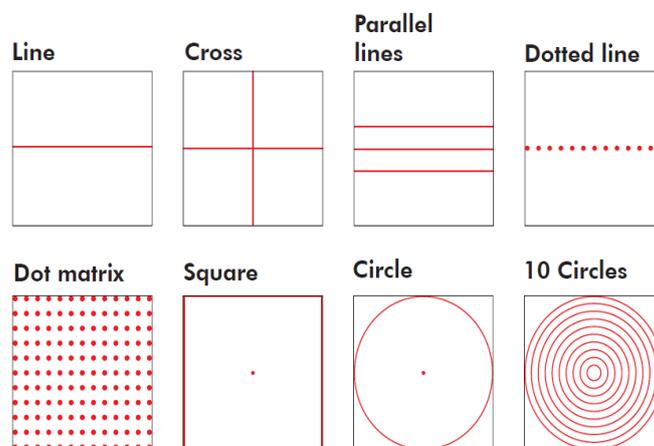


Figura 9. Patrones de luz estructurada generados por las lentes del láser .

Captura de luz estructurada infrarroja

El elemento básico de captura es una cámara IP, modelo CDN-62 con CCD SONY Exview HAD 1/3", escogido por ser especialmente sensible a la luz infrarroja. Esta cámara dispone de un filtro mecánico de IR, que se desactiva al situarla en el modo noche (ver Fig. 10)



Figura 10. Cámara CDN-62 (izda.) y filtros infrarrojos (dcha.)

A fin de obtener máxima sensibilidad al patrón de luz estructurada, frente a la lente de la cámara se incluye un filtro paso-banda de infrarrojos lo más centrado posible en la banda de funcionamiento del láser. En este sentido, se dispone de dos filtros:

- FB810-10: Filtro centrado en 810 nm con una banda de paso de 10 nm, es decir, trabaja entre longitudes de onda de 800 y 820 nanómetros.
- FB800-40: Filtro centrado en 810nm con una banda de paso de 40 nm, es decir, trabaja entre longitudes de onda de 760 y 840 nanómetros.

2.6. Cámaras de visión portátiles con batería

Como parte del material presupuestado, el Proyecto considera la ampliación de los recursos actuales basados en cámaras fijas que no pueden ser fácilmente reubicadas para otros escenarios, ya sea dentro o fuera del campus. Por ello, son necesarias cámaras fácilmente reubicables que no requieran depender de la infraestructura de red del campus, con el objetivo de poder capturar video en distintos entornos de manera flexible, y de poder realizar demostraciones en vivo. Actualmente, existen cámaras de vídeo-seguridad equipadas con conectividad WiFi y alimentadas con baterías de litio, capaces de cumplir con los requisitos de flexibilidad necesarios para su instalación. Adicionalmente, pueden realizar grabación off-line, para aquellos casos en los que se necesite realizar captura desatendida de video durante varias horas.

El dispositivo inicial presupuestado “Cámara de vigilancia portátil WiFi” (<https://www.foscam.es/camara-vigilancia-portatil/>) se encuentra fuera de stock y por ello, se ha buscado un dispositivo que se ajuste a las necesidades del proyecto. Se considera que los actuales teléfonos móviles cumplen estas necesidades (captura, procesamiento y comunicación embebida; batería;...). Se realizó un estudio de los

dispositivos existentes y se seleccionó el modelo Xiaomi Redmi Note 4 3GB/32GB debido a su relación calidad/precio.

Las características del material adquirido son las siguientes (solo se muestran las relevantes para el proyecto):

- **Sistema operativo** Android 6.0 (Marshmallow) MIUI V8
- **Procesador** Qualcomm Snapdragon 625 Octa-core
- **Memoria RAM:** 3GB
- **Memoria Interna:** 32GB
- **Resolución de captura:** Imagen (max 13MP), vídeo (720p)
- **Conectividad** Wifi 802.11ac y Bluetooth 4.2
- **Batería** 4100mAh

Las Figuras 11 y 12 muestran ejemplos del dispositivo adquirido y su montaje para emular un sistema de video-vigilancia portátil. Para el desarrollo de esta infraestructura, se utilizaron trípodes disponibles previamente y se adquirieron adaptadores para el montaje.



Figura 11. Dispositivo Xiaomi Redmi Note 4 3GB/32GB adquirido

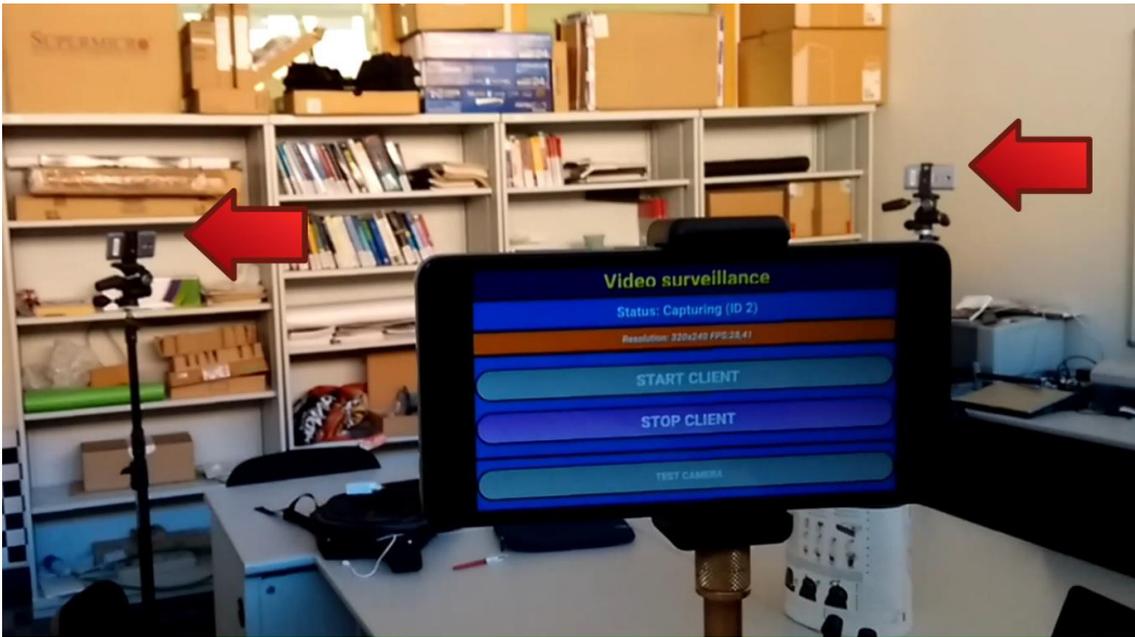


Figura 12. Sistema de videovigilancia usando trípodes y adaptadores con las cámaras móviles

3. Infraestructura de proceso

Esta sección describe los equipos con que cuenta el Grupo para distribuir y procesar las secuencias procedentes de los sistemas de captura. Se trata de un aspecto de especial importancia, ya que un aspecto fundamental en la filosofía de trabajo del Grupo es desarrollar algoritmos que puedan operar en de tiempo real.

3.1. Equipamiento

Aparte de dispositivos de propósito general, el Grupo dispone de varios servidores exclusivamente dedicados a la ejecución de algoritmos de análisis de las secuencias de vídeo capturadas.

Las características principales de estos equipos, montados en bastidores situados en salas refrigeradas (ver Fig. 11) son:

- DELL R415: Dual Octa Core @ 2.7GHz, 64 GB RAM, SAS RAID
- DELL R300: Dual Core @ 3GHz, 24 GB RAM
- DELL R300: Quad Core @ 2,83GHz, 24 GB RAM
- 2x IBM x346: 2x Xeon @ 3GHz, 4 GB RAM, U320 SCSI RAID

Como parte del material presupuestado en el Proyecto, se ha adquirido nueva infraestructura para computación de alto rendimiento (servidor). Este equipo permite entrenar modelos complejos (e.g. detección de personas). La parte de test se realiza en equipos estándar manteniendo el objetivo de alta disponibilidad del proyecto. La infraestructura tiene los siguientes detalles:

- SUPERMICRO SUPERSERVER: Dual Octa Core @ 2.4Ghz, 64 GB RAM
- GPU NVIDIA TITAN X 12GB



Figura 13. Detalle de equipos de proceso (izda.) y del punto de acceso y control de parte de las cámaras (dcha.).

3.2. Subsistemas software de base

El Grupo sostiene su infraestructura en un 100% sobre sistemas libres Debian GNU Linux en versión 6 (Squeeze), mediante los cuales se construye una red para trabajar en el procesamiento de vídeo e imágenes de forma segura y multiusuario.

Dado que los equipos están distribuidos en la red de la UAM, es decir, a una red no privada e insegura, el acceso a la información de los dispositivos de captura se realiza a través de equipos terminadores de VPN "site to site" (OpenVPN) que conectan de forma directa con ellos y además nos proporcionan un entorno de trabajo, mediante consola o escritorio remoto, para un primer procesado previo o trabajo más fluido.

Todas estas VPNs se establecen contra un servidor central en nuestro CPD principal. Este equipo proporciona también VPN de tipo "client to site" para facilitar el acceso de los usuarios en itinerancia o cualquier localización remota, controlado por clave y certificado. Adicionalmente actúa como directorio unificado de autenticación (OpenLDAP), que comparten todos los equipos y aplicaciones del entorno, y, entre otras variadas funciones, como DNS local y servidor web del grupo de trabajo.

Una vez garantizada la conectividad y disponibilidad de los datos a procesar, la capacidad de proceso se ofrece a través de una serie de servidores de virtualización formados por varias máquinas corriendo Oracle VirtualBox. De esta forma es posible realizar pruebas o procesamiento en cualquier sistema operativo, así como compatibilizar y monitorizar el consumo de recursos de las máquinas virtuales en ejecución.

Los usuarios almacenan su código en un servidor de repositorios Subversión, en una máquina dedicada con almacenamiento redundado. Así mismo existen "Baúles" de espacio común para el intercambio de datos entre máquinas virtuales, distintos procesos que operan sobre una misma fuente. etc. El almacenamiento de datos se apoya en unidades de red sobre discos NAS de Lacie, modificados para trabajar con Linux en lugar de Windows, ser accesibles por CIFS, NFS y SFTP, y poder así utilizar el directorio común de usuarios y controlar las cuotas de disco de éstos.

Por último, se intenta garantizar alta disponibilidad de todos los servicios monitorizandolos mediante Nagios y Cacti.

3.3. Arquitectura de proceso DiVA

Sobre la infraestructura descrita, el Grupo ha desarrollado una arquitectura o plataforma que da soporte al desarrollo y ejecución de aplicaciones de análisis de secuencias de vídeo. El término con el que se denomina esta plataforma de desarrollo es DiVA (Distributed Video Analysis).

DiVA establece un entorno distribuido para la intercomunicación simultánea de múltiples fuentes de vídeo con algoritmos de proceso, la conexión en cascada de estos algoritmos de proceso, la visualización de resultados parciales, y la inclusión formalizada de información contextual en el proceso de análisis, todo ello posibilitando

que los flujos de datos se procesen en tiempo real. La descripción detallada de esta plataforma puede encontrarse en el D1.2 “DiVA documentation”.

3.3.1. Módulos

Para poder alcanzar este objetivo se ha propuesto un desarrollo modular en distintos subsistemas. Cada subsistema está relacionado con una función específica dentro de la plataforma. Estas funciones contemplan desde la captura de datos desde distintos dispositivos físicos hasta el almacenaje y presentación de los resultados obtenidos por los algoritmos.

Los subsistemas que componen la plataforma se muestran en la siguiente figura

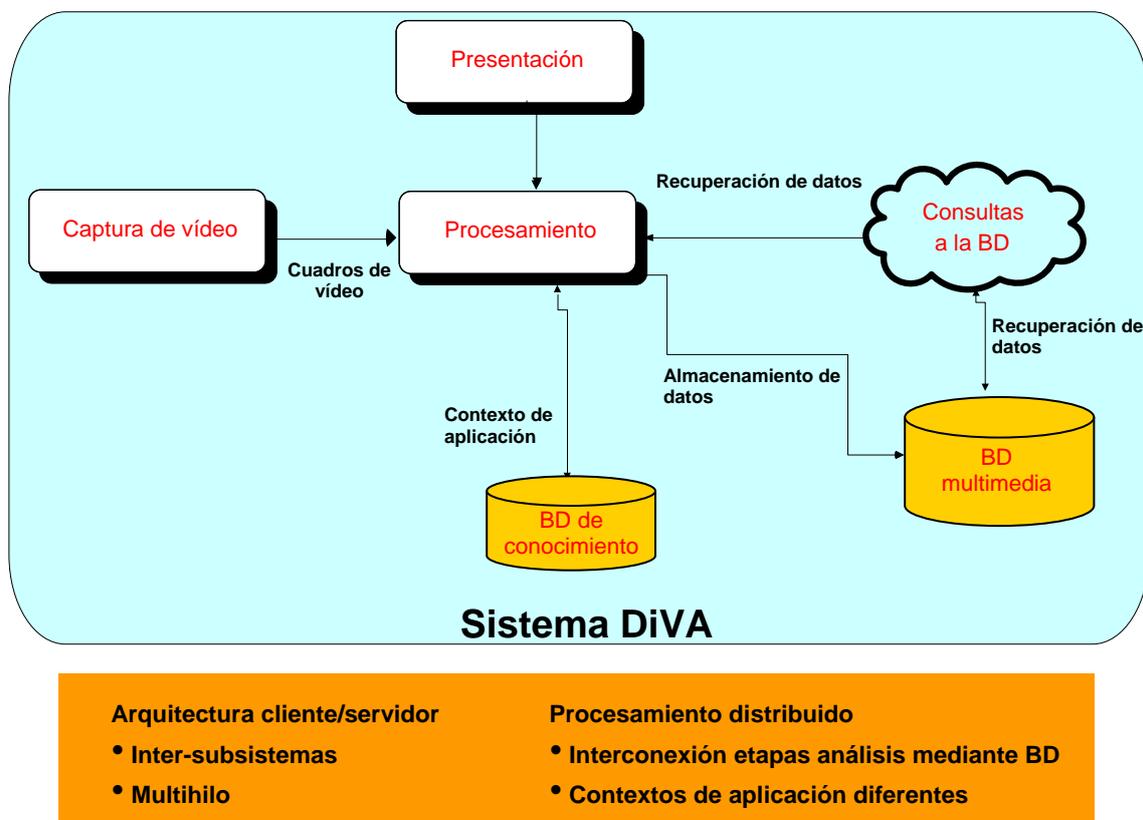


Figura 14. Arquitectura Global de la plataforma DiVA

El esquema de funcionamiento de cada subsistema es el siguiente.

- El *subsistema de captura de datos* reserva los recursos necesarios para su funcionamiento y comienza a capturar datos de las fuentes de vídeo que han sido seleccionadas para capturar señal de vídeo. Dicho sistema de captura dispone de un buffer para almacenar cuadros completos de la señal de vídeo que se transmitirán a los algoritmos de análisis (cuando así lo soliciten).
- El *subsistema de bases de datos* tiene dos objetivos: proporcionar un contexto de aplicación a los algoritmos de análisis y almacenar los resultados del procesado de otros módulos de análisis. Para ello dicho subsistema está compuesto a su vez de otros dos subsistemas que proporcionan las funcionalidades anteriormente descritas.

- El *subsistema de procesamiento* tiene dos objetivos: uno es realizar el propio procesado de la señal de vídeo y otro proporcionar una interfaz de trabajo para cualquier algoritmo de procesado de vídeo (interfaz que gestiona la obtención de un contexto, el procesamiento de la imagen y el posterior almacenamiento de resultados finales o parciales).
- El *subsistema de presentación de datos* tiene por objetivo la presentación en pantalla de los datos de los análisis resultantes del subsistema de procesamiento. Este sistema permite por ejemplo visualizar en una sola ventana un grupo de imágenes.

Para compartir archivos y recursos la plataforma esta desarrollada bajo un modelo cliente/servidor en el que se separa la parte servidora de contenido (en este caso cuadros de la señal de vídeo) de la parte que lo consume (los clientes).

Los subsistemas que funcionan como servidores de contenido son los subsistemas de captura y almacenamiento de datos mientras que los subsistemas de procesamiento y presentación de datos funcionan como clientes.

En la siguiente figura se muestra un ejemplo del funcionamiento de dicha arquitectura cliente/servidor.

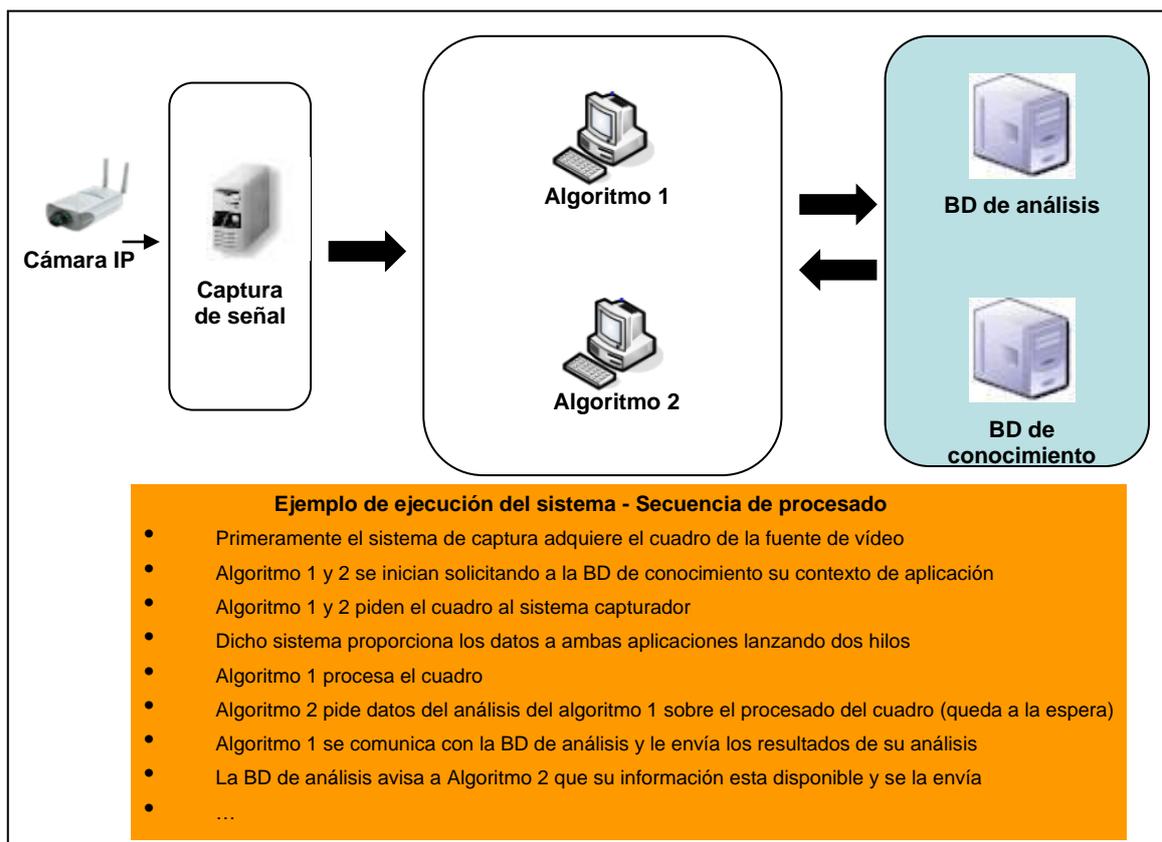


Figura 15 – Ejemplo de ejecución del sistema DiVA mostrando el modelo cliente/servidor y el procesamiento distribuido

Para poder implementar los módulos software anteriormente descritos, se utiliza una red de distribución de datos Ethernet. La red de distribución se compone a su vez de dos subredes: un núcleo central de alta velocidad y una red exterior de velocidad media.

El núcleo central incluye el conjunto de equipos que debido al diseño del sistema necesitan un mayor uso de recursos de red (como son los subsistemas de captura de datos y de almacenamiento en bases de datos). Este sistema consta de varios equipos conectados entre si mediante tecnología Gigabit Ethernet. Además, se incluyen todas las conexiones necesarias para la captura de vídeo desde las distintas fuentes de vídeo de la plataforma. Este núcleo de red se conecta con la red exterior mediante un router Gigabit.

La red exterior incluye todos los algoritmos de procesado. Los equipos que ejecutan algoritmos requieren un menor ancho de banda (debido principalmente a que no tienen que realizar múltiples conexiones) y este hecho posibilita su uso en redes de peor calidad. Actualmente la tecnología de dicha subred es de Ethernet a 100Mbps.

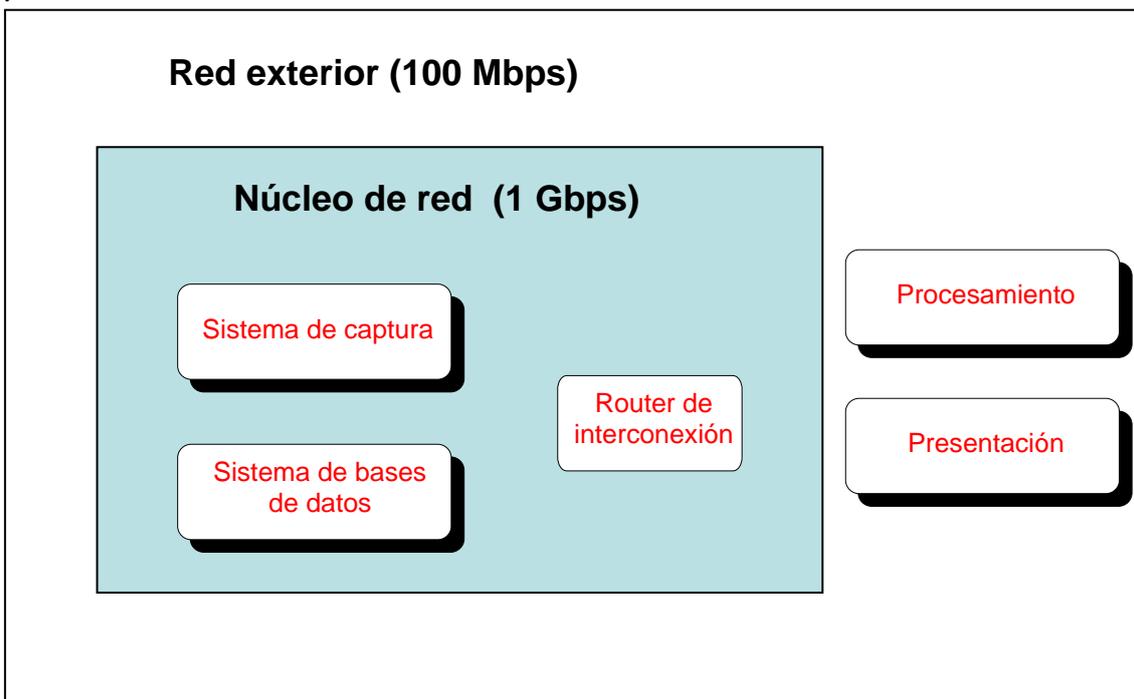


Figura 16 – Arquitectura de red de la plataforma DiVA

3.4. Arquitectura de proceso SMCS

Sobre el hardware adquirido en el punto “2.6 Cámaras de visión portátiles con batería”, el Grupo ha desarrollado una arquitectura o plataforma que da soporte al desarrollo y ejecución de aplicaciones de análisis de secuencias de vídeo. El término con el que se denomina esta plataforma de desarrollo es SMCS (*Smartphone MultiCamera System*).

SMCS establece un entorno distribuido para la intercomunicación simultánea de múltiples fuentes de vídeo (i.e. smartphones) con algoritmos de proceso ejecutados en un servidor, y la inclusión formalizada de información contextual en el proceso de análisis, todo ello posibilitando que los flujos de datos se procesen en tiempo real.

3.4.1. Módulos

En una primera fase, el software desarrollado considera el uso de móviles como unidades de sensado (i.e., cámaras). El sistema de videovigilancia está compuesto de tres módulos (ver Figura 17): cámaras, servidor y aplicación cliente.

Las cámaras serán los móviles Android encargados de conectarse al servidor y permanecer a la escucha de peticiones que vengan del mismo para realizar acciones como configurar parámetros de la grabación, controlar las conexiones wifi o bluetooth, enviar información sobre el estado actual, comenzar la captura de imágenes y enviar dichas imágenes al servidor. El servidor es el módulo intermedio entre la aplicación cliente y las cámaras. La aplicación cliente es la encargada de enviar comandos al servidor, que será el encargado de adaptar dichos comandos a peticiones a las diferentes cámaras, obtener sus respuestas y mandar dicha información a la aplicación cliente. La aplicación cliente es la encargada de enviar los comandos a las cámaras a través del servidor. El hecho de tener un servidor intermedio hace que sea mucho más sencilla de implementar la comunicación con múltiples cámaras de manera simultánea.

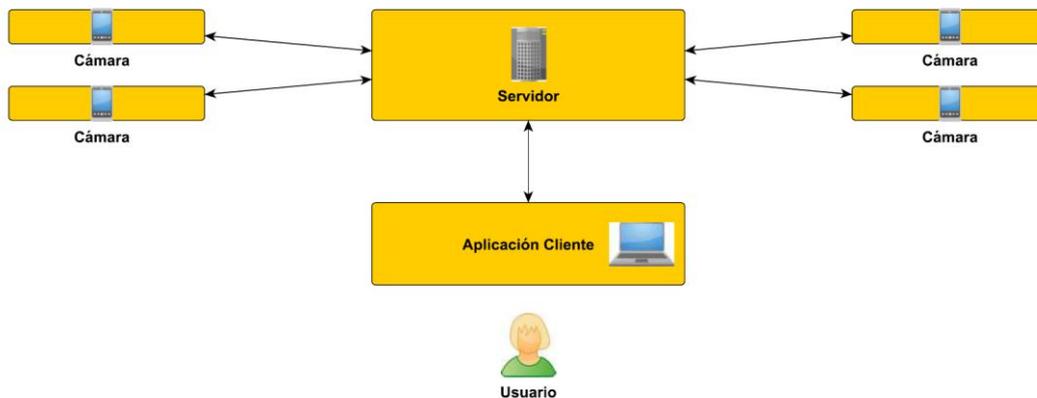


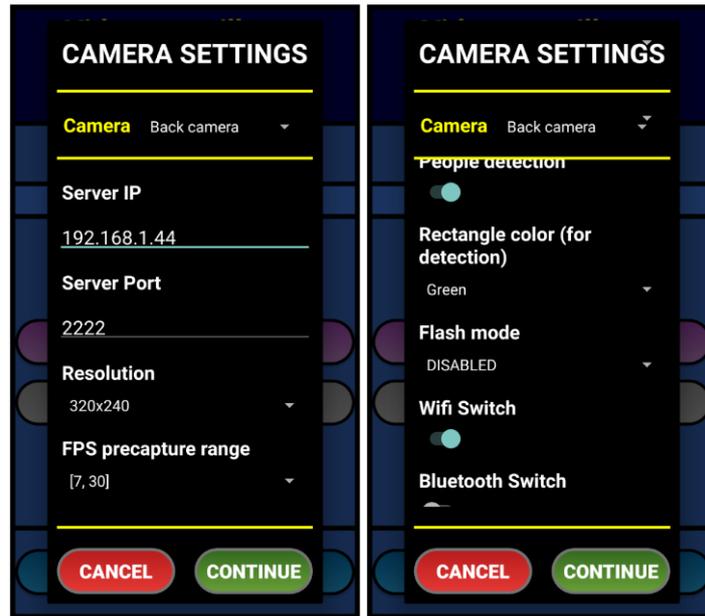
Figura 17 – Arquitectura conceptual de la plataforma SMCS

Cámara Android

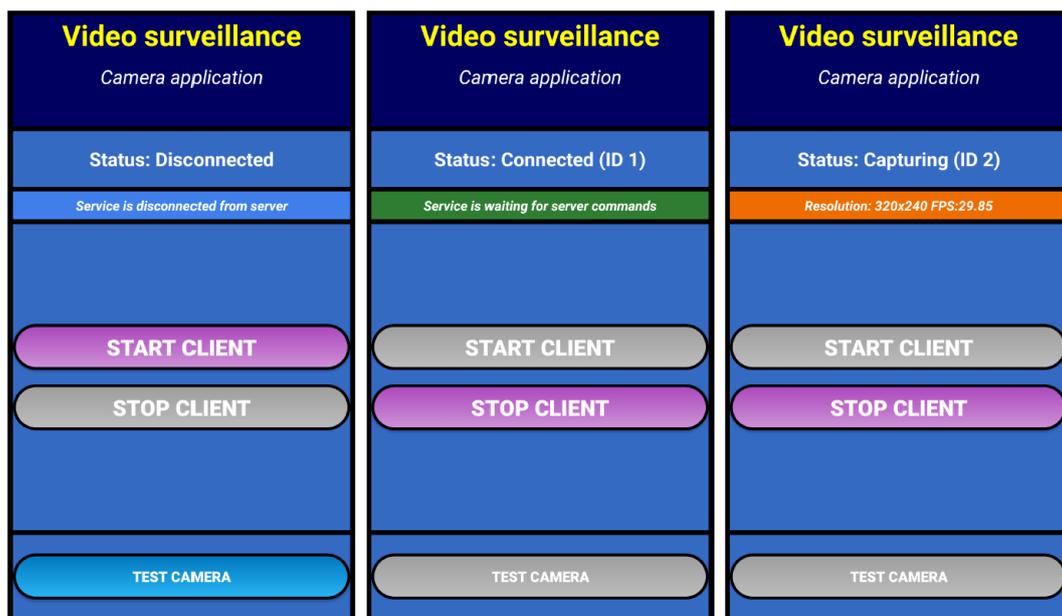
En el desarrollo de la aplicación móvil que permita la captura de imágenes, su procesamiento y envío se ha utilizado el entorno de desarrollo Android Studio para realizar la implementación Java de la app con las librerías de procesamiento de imágenes de OpenCV incluidas en el ejecutable generado como producto final.

Dispone de una interfaz gráfica de usuario que informa en todo momento del estado del dispositivo móvil, en concreto del estado de la cámara y de la conexión con el servidor remoto. Además de funcionalidad visual también dispone de una compleja organización que ofrece desde una API con muchas funcionalidades relativas a la cámara, la conexión y el tratamiento de imágenes hasta una gestión inteligente de los recursos del teléfono para responder rápidamente a los comandos que se reciban de un usuario, a través del servidor, creando dos hilos para su ejecución en paralelo: uno que gestiona todo lo relativo a la conexión con el servidor y otro que gestiona lo relativo a la captura,

conversión y envío de imágenes. La Figura 18 muestra un ejemplo de la aplicación desarrollada.



(a)



(b)

Figura 18 – Aplicación desarrollada para control de cámaras/móviles Android de manera remota
Servidor

El servidor (ver Figura 19) es el corazón del sistema. El servidor es el encargado de realizar la gestión de múltiples cámaras, interactuar con la aplicación cliente y hacer comprobaciones en la integridad de cada comando para ganar fiabilidad y robustez. Es el módulo responsable de la estabilidad de todo el sistema, ya que todas las comunicaciones tienen que pasar por él. Si el servidor cayese, todo el sistema caería.

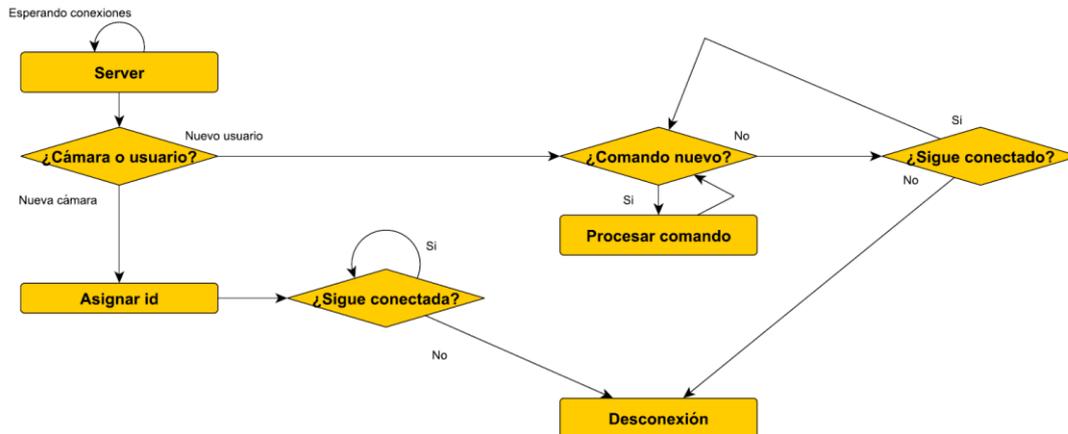


Figura 19 – Diagrama de flujo del servidor

Aplicaciones

Se ha desarrollado una API para implementar aplicaciones en C/C++ mediante OpenCV, cuya funcionalidad se lista en la siguiente Tabla.

	Parámetros	Utilidad
Connect	Cam/User	Conecta una cámara o un usuario al servidor, que le asignará un identificador único
List	Cam/Cmd	El servidor devuelve la lista de cámaras conectadas o la lista de todos los comandos disponibles
Get-current	ID param	Obtiene el valor de un parámetro de una cámara. Puede ser utilizado para devolver todos los parámetros de una cámara o de todas también
Set-current	ID param value	Configura un parámetro concreto para una de las cámaras o todas las conectadas
Start	ID	Comienza a capturar imágenes en una cámara o en todas las que estén inactivas
Stop	ID	Detiene la captura de imágenes en una cámara o en todas las que estén capturando
Get-frame	ID	Obtiene la última imagen disponible en la cámara especificada o en todas
Exit		Termina la conexión con el servidor
Quit		Este comando solo se utiliza en ocasiones que deseamos hacer pruebas y no se accede desde ninguna aplicación. Desasigna el identificador asociado a la conexión pero no cierra la conexión con el servidor. Puede permitir a un usuario dejar de ser usuario y pasar a ser cámara.

Tabla 1 – Descripción de la API

3.4.2. Comunicación

A continuación, se describe brevemente el diseño de las comunicaciones entre los módulos del sistema. Se puede visualizar dichas comunicaciones en la Figura 20.

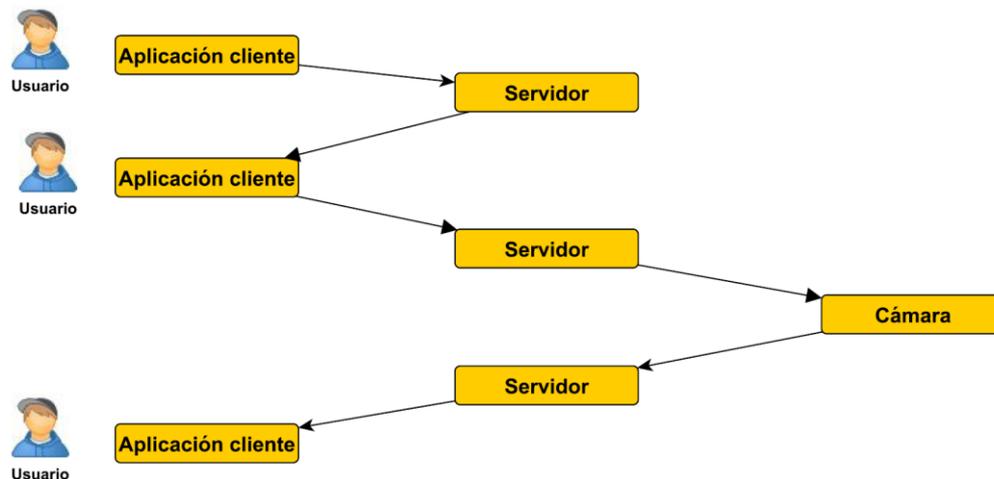


Figura 20 – Flujo de comunicación en SMCS

Comunicación entre cámaras Android y servidor

Los teléfonos móviles Android, es decir, las cámaras, realizan una conexión inicial al servidor en la que se les asigna un identificador único. Después permanecerán inactivas a la espera de recibir comandos. Los comandos que se intercambian entre las cámaras y el servidor son los siguientes:

- *Connect*: Conecta la cámara al servidor, quien le asigna un identificador único.
- *Get current*: Obtiene información de los parámetros de captura, procesador y estado de las conexiones inalámbricas.
- *Set current*: Define un nuevo valor para alguno de los parámetros configurables ya mencionados.
- *Start*: Inicia la captura de imágenes usando los valores configurados previamente.
- *Stop*: Detiene la captura de imágenes iniciada por el comando anterior.
- *Get Frame*: Una vez iniciada la captura de imágenes con el comando start, se puede realizar esta acción, que sirve para obtener la última imagen disponible.

Comunicación entre servidor y clientes

La aplicación cliente es la encargada de enviar las acciones solicitadas por el usuario a las cámaras a través del servidor. De este modo, los comandos de la sección anterior podrán ser enviados al servidor, especificando el identificador de la cámara concreta a la que se quiere enviar.

3.4.3. API para la funcionalidad Android

Tenemos diferentes funcionalidades Android que nos permiten realizar acciones complejas en unas pocas líneas de código. Como estas funcionalidades son muy utilizadas en la aplicación, se ha decidido realizar una API para acceder a esas funcionalidades de una manera eficiente y robusta. Las funcionalidades que se incluyen en esta API son las de:

- Activar/Desactivar la conexión Wifi. Esta opción nos permite activar o desactivar el wifi de forma remota
- Activar/Desactivar la conexión Bluetooth. Esta opción nos permite activar o desactivar el bluetooth de forma remota.
- Configurar los parámetros de captura. Este método nos permite configurar diferentes opciones de captura de imágenes como su resolución, FPS limitado por software (ya que en Android no existe esa posibilidad por hardware), autoenfoco, flash, etc.
- Conversión de imagen en formato YUV a RGB y compresión RGB a JPEG.