

SAVASA: Metodología de Búsqueda de Archivos y Análisis Basada en Estándares

C. Hilario⁽¹⁾, M. Nieto⁽¹⁾, A. Rodríguez⁽²⁾, N. Morocho⁽²⁾, P. Sanchez⁽²⁾, K. Villarroel⁽³⁾, R. Giménez⁽⁴⁾, S. Little⁽⁵⁾, H. Wang⁽⁶⁾, C. Prados⁽⁷⁾, T. Kourtis⁽⁸⁾, A. Arnes⁽⁹⁾, G. Montefiore⁽¹⁰⁾, F. Gaudino⁽¹¹⁾

⁽¹⁾ Vicomtech-ik4, {chilario, mnieto}@vicomtech.org, San Sebastián, España

⁽²⁾ IKUSI – Ángel Iglesias S.A, {aitor.rodriguez, naiara.morocho, pedro.sanchez}@ikus.com, San Sebastián, España

⁽³⁾ Renfe - Dirección de Sistemas de Información, kvillarroel@renfe.es Madrid, España

⁽⁴⁾ HI-Iberia Ingeniería y Proyectos, rgimenez@hi-iberia.es, Madrid, España

⁽⁵⁾ Dublin City University, suzanne.little@dcu.ie, Dublin, Irlanda

⁽⁶⁾ University of Ulster, h.wang@ulster.ac.uk, Ulster, Reino Unido

⁽⁷⁾ INECO - Ingeniería y Economía del Transporte S.A., celso.prados@ineco.es, Madrid, España

⁽⁸⁾ NCSR - National Center for Scientific Research "Demokritos", kourtis@iit.demokritos.gr, Athens, Grecia

⁽⁹⁾ DGT, aarnes@dgt.es, Madrid, España

⁽¹⁰⁾ SINTEL Italia S.p.A, g.montefiore@sintelitalia.com, Pomezia, Italia

⁽¹¹⁾ Studio Professionale Associato a Baker & McKenzie, francesca.gaudino@bakermckenzie.com, Milán, Italia

Abstract- The SAVASA project is a European funded project under the Seventh Framework Programme Theme which faces a standard based approach to video archive search and analysis. The SAVASA project proposes the creation of a video archive search platform that allows authorized users to perform semantic queries over various remote and non-interoperable video archives. The project will exploit the current trends in computer vision, video retrieval and semantic video analysis. It is also a goal of the project to ensure that its results can be deployed in distributed systems and as software services. The project is currently ongoing and preparing the first system prototype in which several of the proposed functionalities will be deployed and evaluated in real scenarios.

I. INTRODUCCIÓN

Dentro de la Unión Europea, la explotación de los sistemas de videovigilancia se ha convertido en una prioridad. Sin embargo, para poder cumplir con las recomendaciones del Foro Europeo para la Seguridad e Innovación ESRIF [1] y las estrategias sobre seguridad interior de la UE [2], la búsqueda eficaz de archivos debe hacer frente a una serie de limitaciones. Las distintas tecnologías de videovigilancia utilizadas por los operadores de infraestructuras (aeropuertos, ferrocarriles, metros, etc.) han dado lugar a la instalación y operación de sistemas de archivo de vídeos no interoperables y diversificados. Como consecuencia, la explotación de la información obtenida a través de dichos sistemas por parte del usuario final resulta desde un punto de vista técnico y operativo complicado. Este problema se ve incrementado debido a las distintas tecnologías, en muchos casos propietarias, utilizadas para la compresión, indexación, almacenamiento y acceso a la información obtenida a partir de los sistemas de videovigilancia. De hecho, las infraestructuras de circuito cerrado de televisión (CCTV) actuales no están siendo del todo explotadas (por ejemplo, para ayudar a reducir o prevenir actos criminales) debido a la imposibilidad de tratar el conjunto masivo de vídeos grabados por múltiples cámaras [3]. Como ejemplo, se muestran las imágenes de la Fig. 1,

correspondientes a grabaciones de Scotland Yard relativas a incidentes ocurridos en Londres en 2011.

Además de estas dificultades, existen otras relacionadas con el acceso remoto a los archivos de vídeos y su explotación, y la necesidad del uso de redes seguras para permitir su integración en la plataforma. Por un lado, habría que autorizar el acceso de los usuarios finales al sistema, y por otro, sería necesario autenticar de un modo adecuado cada uno de los componentes de la CCTV y la infraestructura de los archivos de vídeo.



Fig. 1. Ejemplos de caras capturadas por sistemas CCTV.

Fuente: Scotland Yard

<http://singularityhub.com/2011/08/20/scotland-yard-using-facial-recognition-to-find-rioters-%E2%80%93-sadly-it-wont-work/>

Por último, con el fin de garantizar la protección de datos de los ciudadanos y para evitar un uso indebido de los mismos, se debe asegurar que los usuarios del sistema hagan uso del mismo cumpliendo con el marco legal, ético y de privacidad establecido. En la Fig. 2 se muestra un ejemplo de la tecnología existente para efectuar ofuscamiento de imágenes que contengan información potencialmente privada.

En la actualidad, existen algunos proyectos ya terminados o en fase de desarrollo con algunos puntos en común con el SAVASA. Dentro del proyecto SAMURAI [4], se han implementado herramientas para la detección de personas, vehículos y equipajes, realizando un seguimiento de los mismos a través de una red de cámaras. El proyecto SUBITO [5] va a estar enfocado a la detección en tiempo real de equipajes u otros objetos abandonados y la identificación del

individuo que los dejó, así como la trayectoria que siguió. El PAPYRUS [6] trabaja en la gestión del conocimiento y el análisis semántico de las consultas. En el proyecto IMP [7] se genera una infraestructura capaz de almacenar metadatos en una base de datos distribuida. Finalmente, en el COAST [8], el usuario especificará qué contenido necesita y el sistema le devolverá la información deseada de un modo eficiente y en un entorno amigable.

A pesar de las contribuciones de todos estos proyectos, por el momento, no existe un sistema de búsqueda de archivos de vídeo automatizado que supere las limitaciones técnicas, legales y de interoperabilidad antes mencionadas.



Fig. 2. Ejemplo de ofuscación de imágenes para la protección de privacidad. Fuente: University of Toronto, <http://www.dsp.utoronto.ca/projects/surveillance/>.

A diferencia de otros proyectos, el objetivo del SAVASA es la creación de una plataforma de búsqueda de archivos de vídeo que permita a aquellos usuarios autorizados realizar búsquedas semánticas a través de diversos archivos remotos y no interoperables. El proyecto va a hacer uso de las últimas tecnologías de visión por computador, de recuperación de vídeos y de análisis semántico de los mismos. Además de esto, los resultados obtenidos por la plataforma deberán poder ser usados por sistemas distribuidos así como ofrecer las distintas soluciones como servicios de software.

El SAVASA forma parte de un proyecto Europeo, y debido a los objetivos tan ambiciosos que persigue, se ha creado un consorcio formado por expertos en ética y leyes, así como por los usuarios finales. Este artículo se ha organizado del siguiente modo; en la sección II se discute brevemente la arquitectura SAVASA. Como ejemplo de uno de los módulos que componen la plataforma SAVASA, en la sección III se exponen las técnicas de detección y seguimiento de caras más habituales y se muestran algunos resultados del detector de caras implementado dentro del módulo de análisis de vídeo. La sección IV presenta varios casos de test y estrategias.

II. ARQUITECTURA

La plataforma SAVASA va a permitir a los usuarios autorizados realizar consultas semánticas sobre diferentes archivos de vídeo de un modo amigable. La participación de los usuarios finales del proyecto (RENFE y DGT principalmente) ha permitido definir la arquitectura del sistema preliminar que se muestra en la Fig. 3. Como se puede apreciar en la figura, los posibles usuarios de la plataforma SAVASA se pueden clasificar en dos tipos. Por un lado, aquellos que producen y gestionan los archivos de vídeo (i.e *Producers*), como por ejemplo RENFE o la DGT. Por otro lado, aquellos que realizan las consultas (i.e *Consumers*). Dentro de este grupo podrían estar los agentes de seguridad del estado o las autoridades judiciales. Los usuarios se conectan a la plataforma de modo seguro, bien a través de una conexión SSL/TLS o VPN.

Debido a restricciones técnicas y legales, los vídeos no pueden ser utilizados por terceros, por lo que el sistema SAVASA se ha dividido en dos partes: el nodo SAVASA local y el cloud SAVASA. Dentro de cada productor se instalará un nodo encargado de adquirir videos a través de los sistemas de CCTV y realizar un procesamiento de las imágenes, para extraer rasgos visuales de bajo nivel, así como realizar la detección y seguimiento de objetos de interés tales como personas. Dentro del cloud SAVASA, van a existir cuatro módulos principales que van a permitir que el usuario realice consultas y obtenga un resultado en tiempo real. Son los siguientes:

- **Módulo de interfaz web y gestión de acceso (“Web interface / access gateway”):** Los usuarios van a acceder al sistema a través de una interfaz web donde podrán identificarse, realizar búsquedas y recibir informes. Se va a prestar especial cuidado a la capa de acceso, ya que en función de los permisos del usuario, éste tendrá acceso a determinados datos. Por ese motivo, se van a desarrollar algoritmos de procesamiento de imágenes que en el nodo local realicen un ofuscamiento de los objetos de interés, garantizando así la protección de datos.
- **Módulo de análisis semántico de vídeo (“Semantic video analysis”):** Este módulo analiza los rasgos de bajo nivel previamente extraídos de los vídeos y lleva a cabo el reconocimiento de objetos y escenarios. Como resultado obtiene información de alto nivel codificada como metadatos semánticos.
- **Módulo de actualización de índices de búsqueda:** Identifica los conceptos semánticos y los indexa, con el

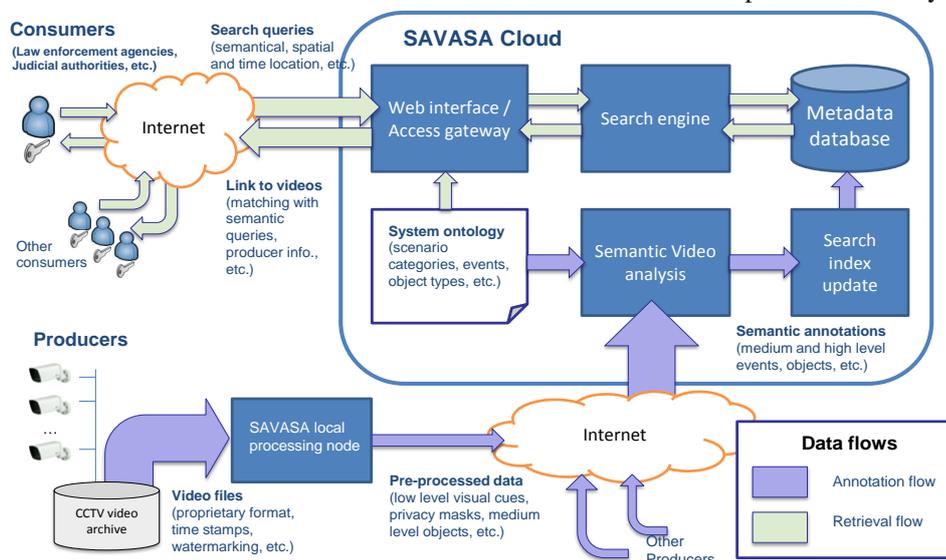


Fig. 3 Arquitectura preliminar del sistema SAVASA

fin de facilitar la recuperación de los datos una vez realizada la consulta por parte del usuario. Actualiza la base de datos de metadatos a medida que llegan nuevas anotaciones.

- **Motor de búsqueda (“Search engine”):** Convierte las consultas realizadas por el usuario en índices semánticos. A través de estos índices se accede a la base de datos de metadatos y encuentra la anotación que mejor se ajusta a la petición del usuario. La base de datos de metadatos se ha generado previamente como resultado del análisis semántico y las anotaciones realizadas a partir de los archivos de vídeos almacenados en los archivos de los productores (por ejemplo, RENFE y DGT).

En el siguiente apartado se muestran algunos de los resultados preliminares obtenidos en el módulo de análisis de vídeo, concretamente en las aportaciones en el campo de la detección de caras para identificación de personas.

III. DETECCIÓN Y SEGUIMIENTO DE CARAS

En los entornos de videovigilancia, el rasgo que mejor se presta a la hora de detectar a una persona es su cara. En las dos últimas décadas se han desarrollado un gran número de métodos para detectar caras. Estos se pueden clasificar en 4 tipos: basados en conocimiento, basados en rasgos invariantes, basados en la correspondencia de plantillas (“*Template matching*”) y basados en la apariencia[9]. Los métodos basados en el conocimiento usan reglas predefinidas basadas en el conocimiento humano para determinar qué es una cara; las técnicas basadas en rasgos invariantes pretenden encontrar rasgos faciales que sean robustos a cambios de postura e iluminación; los métodos basados en la correspondencia de plantillas emplean máscaras faciales previamente almacenadas para evaluar si una imagen contiene o no una cara; los métodos basados en la apariencia aprenden los modelos faciales a partir de un conjunto representativo de imágenes de entrenamiento [10].

En los últimos años, la detección de caras ha sufrido un progreso importante, en parte, debido al trabajo de Viola y Jones [11]. La simplicidad de su idea consistente en un clasificador *Adaboost* en cascada hizo posible la detección de caras en aplicaciones del mundo real y ha servido de inspiración a muchos otros autores.

Una vez llevada a cabo la detección, el siguiente paso sería el reconocimiento de la cara, entendiéndose como tal la tarea de identificación del rostro presente en la imagen o vídeo. Esta tarea es particularmente compleja en los sistemas de CCTV, debido a la baja resolución de las imágenes, a las caras no frontales y a las condiciones de iluminación cambiantes. Para poder realizar este proceso, los pasos de detección y seguimiento del rostro son de gran importancia, ya que el reconocimiento se realiza en función de la información proporcionada por ambos.

El seguimiento de una cara se puede realizar en función del color de la piel u otros rasgos faciales, así como el movimiento. Otra alternativa consiste en emplear modelos que contengan información facial, como por ejemplo la forma o apariencia de un rostro. Dentro de este enfoque se engloban los modelos de la forma activa (ASM, “*Active Shape Model*”) [12], los modelos de apariencia activa (AAM, “*Active Appearance Model*”) [12] y los modelos de apariencia online (OAM, “*Online Appearance Models*”) [14], ampliamente usados tanto en el seguimiento como en el reconocimiento de caras. Estos métodos tratan de ajustar un modelo facial a la imagen y comparan el resultado obtenido

con las imágenes de una base de datos para llevar a cabo la identificación. Este es el enfoque utilizado en Vicomtech.

A. Resultados preliminares del detector de caras

En el proyecto SAVASA se está trabajando en un detector de caras basado en la apariencia que, a partir de imágenes tomadas con una cámara crea un modelo que se ajusta a la forma de la cara. En la Fig. 4 se muestra varios ejemplos de su funcionamiento. El sistema es capaz de detectar y seguir el rostro de una persona en tiempo real, siendo robusto a cambios de expresión facial, identificando zonas de interés, tales como los ojos, la boca o la nariz, todo lo cual permite incrementar la precisión de la medición de los rasgos de la persona y por tanto incrementar la posibilidad de reconocer la persona a través de un sistema automático.

La contribución científica principal de esta parte reside en la utilización combinada de técnicas de seguimiento en tiempo real con OAM y una inicialización robusta basada en modelos de apariencia AAM.



Fig. 4. Detección y modelado 3D de caras, con adaptación del template, ajuste de expresiones y seguimiento ante traslaciones, rotaciones y escalados

IV. ESCENARIOS Y ESTRATEGIAS DE VALIDACIÓN

Con el objetivo de completar integraciones parciales de prototipos y de validar por tanto las distintas funcionalidades del sistema SAVASA se ha trabajado en la descripción de diversos escenarios de test y validación. Estos escenarios se han definido siguiendo una estrategia de validación que permite identificar un vínculo directo entre funcionalidades del sistema y requisitos de usuario. Con esta validación se consigue determinar qué requisitos, y en qué grado, se satisfacen en cada uno de los prototipos incrementales que se han planificado en la fase de integración del proyecto.

Concretamente, el consorcio ha determinado tres potenciales escenarios de test, orientados significativamente a estudiar el posible impacto del proyecto en condiciones de uso realistas. Los escenarios candidatos son:

A. Impacto operacional del proyecto SAVASA

En este escenario se pretende demostrar la capacidad de las herramientas y tecnologías de la plataforma SAVASA para mejorar el tiempo de respuesta en la recogida de material de vídeo, de archivos de vídeo remotos y locales, considerando sistemas CCTV de distinta naturaleza. Igualmente se busca determinar la integridad de la plataforma SAVASA a la hora de proporcionar vídeos como evidencias de suficiente calidad y trazabilidad (mediante verificación de cadena de custodia) para que puedan ser utilizados legal y judicialmente como pruebas materiales.

En este escenario de validación, al contrario que los otros escenarios, más orientados a la parte de investigación y tecnológica del proyecto, se busca medir cuantitativamente el

impacto directo del sistema SAVASA desde una perspectiva operacional para el usuario final. Estos usuarios (RENFE y DGT principalmente, por formar parte del consorcio SAVASA) proporcionarán análisis detallados de la mejora que se obtiene mediante la utilización de un sistema automático como SAVASA frente a los habituales tiempos de respuesta desde que se presenta una orden judicial y se entregan los vídeos correspondientes a las instancias judiciales.

B. Búsqueda semántica jerárquica

En relación con la capacidad de realizar búsquedas semánticas por parte de los usuarios consumidores del SAVASA (agentes y cuerpos de seguridad del estado, policías internacionales, jueces, etc.), se realizarán tests siguiendo la metodología de TrecVID (plataforma de benchmarking desarrollada y mantenida por DCU, socio del proyecto SAVASA). Este escenario proveerá el doble beneficio de evaluar los aspectos más innovadores de las tecnologías de R&D del proyecto a la vez que se evaluará el rendimiento del proyecto con las técnicas de análisis semántico que se utilizan actualmente a nivel mundial.

Los diversos archivos que se integran vía la plataforma SAVASA se presentan al usuario como un único archivo virtual independiente de las distintas tecnologías y estándares en los que se basan, y de si se encuentran almacenados local o remotamente. Este escenario se plantean búsquedas jerárquicas, donde por ejemplo, se trate de identificar en primer lugar vídeos que contengan vehículos pesados en movimiento en diversos archivos, y en segundo lugar, detectar en estos vídeos secuencias de imágenes que contengan camiones articulados viajando en dirección sur.

C. Seguimiento de objetos a través de múltiples archivos de vídeo

La finalidad de este escenario de test es la capacidad de la plataforma SAVASA de determinar la traza de un objeto a través de múltiples bases de datos de vídeos, geográficamente dispersas. La validación de tales capacidades depende de la combinación exitosa de distintos módulos del sistema SAVASA. Por un lado, las herramientas de procesamiento de vídeo conseguirán la detección y seguimiento de determinados objetos en el material de vídeo en función de sus propiedades visuales. La descripción del escenario se podrá utilizar entonces para clasificar objetos como seres humanos, vehículos, etc. De esta forma, las posteriores etapas de identificación y reconocimiento pueden comparar los objetos detectados con las bases de datos disponibles. El resultante metadata se añadirá a la información espacio-temporal de los vídeos para su almacenamiento.

Como ejemplo, podemos considerar el caso de querer realizar un seguimiento de la ruta realizada por un vehículo o un tipo de vehículos a través de la red de carreteras nacional utilizando su número de matrícula o alguna otra característica que lo identifique, o una persona identificada por su cara a través de diversas redes CCTV, tales como aeropuertos, edificios institucionales, etc. Algunos ejemplos de las búsquedas semánticas que se podría aplicar son:

- Buscar el camino recorrido por un vehículo (con matrícula "###-####") desde el 2011/05/01 al 2011/09/01 en autopistas españolas.
- Buscar la traza de una persona ("Nombre", "Apellido") desde el pasado jueves en París.

Los escenarios de test abordarán la compleja tarea de identificar el camino o traza de un objeto móvil que entra y

sale del ámbito de diversas redes de vigilancia y por tanto de distintas bases de datos de vídeos, utilizando diversas tecnologías e instalaciones geográficamente dispersas. Adicionalmente, este escenario considera los aspectos legales implicados en el traspaso de información en distintas jurisdicciones nacionales e internacionales.

V. CONCLUSIONES

El proyecto SAVASA tiene como objetivo el desarrollo de un sistema de búsqueda de vídeos adaptado a las necesidades de los sistemas de videovigilancia y sus usuarios. Este motor de búsqueda utilizará los resultados obtenidos por las herramientas de análisis de vídeo tales como la detección de objetos o personas, así como el análisis de semántico, entre otros, con el fin de combinar toda esa información para realizar una búsqueda más eficaz. Debido a que los vídeos de videovigilancia pueden ser capturados en distintos sitios y por distintos usuarios, esta plataforma facilitará el acceso a múltiples archivos remotos, mediante la integración en una única plataforma de diferentes tecnologías.

AGRADECIMIENTOS

Este trabajo se engloba dentro de las actividades de diseminación del proyecto SAVASA. Las investigaciones que han dado lugar a estos resultados han sido financiadas por la Unión Europea bajo el 7º Programa Marco (FP7/2007-2013), *Grant Agreement* num. 285621, en el que participan los siguientes socios: IKUSI (España), Dublin City University (Irlanda), University of Ulster (UK), Baker & McKenzie (Italia), Vicomtech-ik4 (España), RENFE (España), DGT (España), Iberia (España), INECO (España), Sintel (Italia), Demokritos (Grecia).

REFERENCIAS

- [1] ESRIF, European Security Research & Innovation Forum – Final Report, Dec. 2009.
- [2] EC, The EU Internal Security Strategy in Action: Five steps towards a more secure Europe, Nov. 2010.
- [3] The Guardian, CCTV boom has failed to slash crime, say police, Tuesday 6 May 2008. <http://www.guardian.co.uk/uk/2008/may/06/ukcrime1>.
- [4] SAMURAI (Suspicious and abnormal behaviour monitoring using a network of cameras & sensors for situation awareness enhancement)
- [5] SUBITO (Surveillance of unattended baggage and the identification and tracking of the owner)
- [6] POPYRUS (Cultural and historical digital libraries dynamically mined from news archives)
- [7] IMP (intelligent metadata-driven processing and distribution of audiovisual media)
- [8] COAST (Content Aware Searching, retrieval and streaming)
- [9] Zhang and Z. Zhang, "A Survey of Recent Advances in Face Detection," Technical Report MSR-TR-2010-66, June 2010.
- [10] M.-H. Yang, D. J. Kriegman, and N. Ahuja, "Detecting faces in images: A survey," IEEE Transactions on PAMI, vol. 24(1), pp. 34-58, 2002.
- [11] P. Viola and M. Jones, "Rapid object detection using a boosted cascade of simple features," in Proc of CVPR, 2001.
- [12] T. F. Cootes, C. J. Taylor, D. H. Cooper, and J. Graham, "Active shape models – their training and application," Computer Vision and Image Understanding, vol. 61, pp. 38-59, 1995.
- [13] T. F. Cootes, G. J. Edwards, and C. J. Taylor, "Active appearance models," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 23, no. 6, pp. 681-684, 2001.
- [14] F. Dornaika, and F. Davoine, "On Appearance Based Face and Facial Action Tracking," IEEE Transactions on Circuits and Systems for Video Technology, vol. 16, no. 9, 2006.