

# Visión Computacional aplicado en la Seguridad y Comunicaciones

Medina, Jahir., Pastor, Christian., Salinas, Josep y Sifuentes, Víctor.  
{jahirmedina, cmpastors, jhsalinas, vsifuentes}@unitru.edu.pe  
Universidad Nacional de Trujillo

**Resumen**—La visión computacional como técnica de detección o análisis de patrones, en la actualidad, es una norma. Desde la vigilancia mediante cámaras de circuito cerrado hasta la eliminación de ruidos en señales de audio.

La visión computacional ha avanzando hasta un punto donde es capaz de analizar contextos, pudiendo identificar si se comete un delito o si existe una persona con actitud sospechosa.

Es por esto que en el presente artículo, se hará un recuento de las aplicaciones modernas de la visión computacional. Mas concretamente su aplicación en el capo de la seguridad, vigilancia y telecomunicaciones.

**Índice de Términos**—Visión Computacional, Seguridad, Seguridad Informática, Vigilancia, Telecomunicaciones, Procesamiento de Señales, DeepFake.

## I. INTRODUCCIÓN

Escribir Luego

## II. LA VIGILANCIA AUTOMATIZADA

### A. En los 90's

Desde que se empezó a usar de forma comercial tecnologías de procesamiento de imágenes para detectar movimiento en grabaciones tomadas por Cámaras de Circuito Cerrado (*CCTV*, por sus siglas en ingles) en los años 80's, se veía el potencial pero también su mal rendimiento, especialmente por la alta tasa de falsos negativos en la detección de intrusos (Sage and Young, 1998).

Sin embargo, una solución que se considero y trabajo por mucho tiempo fue la de recopilar mas información para así poder garantizar la disminución de falsos negativos (Esto Basado en una cuestión estadística, mas información, mejor predicción).

Sin embargo, arrojar hardware a un problema de software es una solución, que a la larga aumenta los costos de cualquier sistema. Ante esta problemática, se comenzó a plantear modelos estocásticos para no solo detectar variaciones en la escena filmada, sino para intentar también, trazar una ruta y aproximar este comportamiento a uno próximo de un humano (Sage and Young, 1998).

Gracias a las mejoras en las técnicas de análisis y la mejor calidad en vídeo, en los últimos años de los 90's, se empezaron a plantear sistemas de detección en escenarios dinámicos, siendo un caso particular, las carreteras (Manendez et al., 1999). La motivación, tal como menciona el paper *Vigilancia de Autopistas mediante visión computacional stereo* (Manendez et al., 1999, Abstract) se origina por el aumento

de la demanda de automatización, la ubicuidad de cámaras y la mayor necesidad de automatización y abaratamiento de costos.

En todo este escenario de crecimiento tecnológico, no solo de hardware, sino también de software y sus respectivos algoritmos, es que comienza a surgir la idea de extender estas aplicaciones a campos mas delicados: Detección de crímenes y Verificación Biométrica.

### B. Hardware y Matemática

Desde comienzos de los 2000's ya se comenzaba a visualizar el verdadero efecto de la Internet: Niveles inauditos de información, organizada o desorganizada, pero información. No solo aumentaba la información disponible, sino que el hardware especializado se comenzaba a hacer mas accesible: Tarjetas para procesamiento grafico, Tarjetas de calculos para fisicas simuladas y microprocesadores con un conjunto de instrucciones mas amplio y eficiente.

Este abaratamiento (no solo abaratamiento sino tambien un aumento en la calidad ofrecida en un determinado rango de precios (Chaki et al., 2010)) y masificación del hardware sumado con una revolución del software como servicio fue fundamental para implementaciones "artesanales" de sistemas de vigilancia (S et al., 2020).

Aun mas importante y critico para esta etapa de transición es la revolucion de las Redes Neuronales, no su descubrimiento pero si su uso extensivo. Evolucionando desde simples clasificadores (perceptron clasico) hasta sistemas capaces de analizar contextos y componer una escena con multiples videos facilitando el análisis forense en el caso de ataques terroristas (Schindler et al., 2020).

### C. Actualizaciones Necesarias

Si bien los sistemas de circuito cerrado han existido por muchísimo tiempo entre nosotros, es comun que un punto de falla se encuentre en el componente humano. El tedio de repetir la misma tarea, por horas, días, hasta semanas de supervisar el contenido en vídeo mostrado en los monitores genera fatiga, y por tanto se reduce la atención al detalle; finalizando en la falla mas obvia: la omisión de contenido critico o de relevancia para la seguridad, monitoreo o control.

La visión computacional se puede usar para automatizar tareas como: detección de movimiento, cambios en la condición del video (Si existe señal o no), segmentación de información en tiempo real y mas aplicaciones (Potgieter and

Niekerk, 2012), una situación interesante que se origina de esta automatización es que, de darse una implementación unilateral.

El abaratamiento del hardware, y la computación como servicio, como se mencionaba; facilita la creación de sistemas artesanales de vigilancia, estos sistemas hacen uso de la internet para usar servicios de terceros y así solo preocuparse por la instalación del hardware requerido (Othman and Aydin, 2017), de forma que aun que el usuario no tenga conocimientos avanzados puede hacer uso de sistemas que se encuentra a la par de los usados por bancos; esto es posible por que estos terceros proveedores no solo ofrecen servicios a usuarios finales sino también a grandes corporaciones.

En el área de la detección de movimiento se puede observar actualizaciones tan avanzadas como el empleo de algoritmos cuánticos (Yu et al., 2019) para optimizar la paralelización de los mismos, si bien estos métodos requieren de un hardware especializado, la existencia de propuestas con esta ambición hacen que no solo la criptografía cuántica represente un punto de inflexión para la seguridad en el futuro, sino también para la vigilancia.

En el área de la segmentación y/o extracción de información se tiene una actualización importante en lo que respecta la vigilancia de carreteras: mayor confiabilidad en la detección de placas e identificación de vehículos.

Si bien la extracción de información de una placa de tránsito puede parecer sencillo, en la vida real implica un 80 % pre-procesamiento y un 20 % procesamiento, siendo el primer 80 % la parte más complicada puesto que la realidad no solo se genera ruido en las imágenes, sino también pérdida de información. Tener algoritmos unificados, capaces de realizar ambos en una sola ejecución es esencial para aumentar no solo la precisión sino también la eficiencia (Quiros et al., 2015). Si para una placa de tránsito el trabajo puede resultar complicado, para la identificación de vehículos dadas ciertas características puede resultar en una tarea inacabable (sino se automatiza, se necesita de operarios humanos), sin embargo si extendemos el uso las redes neuronales profundas y su capacidad para detectar patrones, esta aplicación solo quedaría limitada por la cantidad y calidad de información disponible (Qian et al., 2019).

Por otra parte en esta área se avanza a niveles nunca esperados en el poco tiempo que se han ido dado estos cambios, en abril del presente año se presenta un analizador forense para videos de diferentes ángulos y perspectivas con el objetivo de investigar escenarios posteriores a ataques terroristas (Schindler et al., 2020). Una situación a destacar es que para analizar conjuntos de videos es necesario considerar que se debe reconstruir escenarios tridimensionales y los objetos en ellos, un problema que ya había sido abordado con anterioridad (Sumi et al., 2002).

Algo importante a notar es que si bien la unificación de los avances en las áreas de detección de movimiento, analizadores semánticos (de contexto), reconstrucción de escenarios y segmentación pueden resultar muy útiles para la vigilancia con miras a evitar delitos o ubicar a los culpables de uno, también puede servir para evitar accidentes (Yoshimoto et al., 2004), ya que estas cámaras que son usadas para vigilancia no solo

se encuentran encendidas durante la noche, sino durante el día por igual. Es por esto que de acoplar un sistema con capacidad de detectar los riesgos y peligros en el espacio de trabajo (o espacio donde se encuentren las cámaras), se lograría aumentar la seguridad de los individuos que usan dichas instalaciones (Zubal et al., 2016).

### III. METADATOS HUMANOS

#### A. Firma Biométrica

Es indiscutible que el aumento del hardware multimedia ha creado un abanico de información que puede usarse para identificar a una persona, esto no solo implica imágenes del rostro de una persona, sino también de su huella dactilar, talla, peso, firma y patrones de voz (Goldenfein, 2019).

Es importante recordar que los escaners de huella dactilar son un tipo de cámara, capaz de tomar una foto con alta precisión de la huella dactilar. Es por esto que cualquier procesamiento asociado a la detección de huellas dactilares será con imágenes (Spinoulas et al., 2020). Si bien se pueden acoplar más sensores, como uno de presión o temperatura para reforzar la data asociada a una captura de huella digital, esto no es viable en muchos escenarios, pues la idea detrás de este sistema de identificación es la conveniencia y su reducido costo de implementación en la actualidad.

En el presente año se ha presentado múltiples trabajos de investigación que buscan mejorar la calidad de detección y a su vez reducir la posibilidad de ataques hacia los lectores de huella digital, un enfoque es añadir redundancias modificando el ángulo de la cámara y luz dentro del lector (Spinoulas et al., 2020) o el empleo de métodos multitarea en una red neuronal profunda para detectar huellas dactilares modificadas, ubicando las zonas de interés (Giudice et al., 2020).

Ahora, analicemos que sucede con una firma biométrica directa: una fotografía de nuestro rostro, esta información de carácter delicado es actualmente de fácil acceso; sea por las redes sociales o por que cualquier persona con celular también es dueña de una cámara, obtener una fotografía de nosotros nunca fue tan fácil.

La detección de rostros, que forma parte la visión computacional (Pezoa and Dominguez, 2017) hay llegado a un punto de automatización que gobiernos (como el chino o el norteamericano) usan para el seguimiento de personas de interés. Incluso nos encontramos en un punto donde se puede monitorear la asistencia de alumnos a las clases (Harikrishnan et al., 2019), si bien esto puede resultar ilegal en cierta medida (Cote and Albu, 2017), no es algo que una carta de consentimiento no pueda corregir.

La facilidad de contratar un servicio de terceros para implementar sistemas de visión computacional nos ha llevado a un punto donde no solo podemos implementar sistemas de seguridad que detectan movimiento o peligros, sino también rostros (Aydin and Othman, 2017); si bien esto puede usarse como una extensión del caso previo y marcar la asistencia al trabajo del personal, puede usarse también para identificar intrusos y obtener datos del individuo en cuestión en tiempo real.

En el caso que un atacante comprometa la integridad del sistema encargado de la obtención de imágenes, también

existen técnicas capaces de detectar videos alterados para no revelar el rostro, si bien esto puede usarse como una medida antagónica a los famosos *deep fake*, en el caso particular del software *VideoForensicsHQ* (Fox et al., 2020), se usa para el escrutinio forense. Un ataque similar es alterar completamente la grabación para incluir rostros creados o alterados de tal manera que, aun que reconocibles, no sean de la persona original; técnicas para su detección y corrección han sido propuestas en los últimos años, sin embargo el uso de redes neuronales con memoria de corto plazo con capacidad de generalización fue propuesto recién en el presente año (Aneja and Nießner, 2020).

Aquí es importante aclarar que métodos que buscan generalizar sistemas de detección se enfrentan a un tipo de problema bastante crítico: la similaridad visual, esto causa que por mas preciso que sea el método de detección, corrección o segmentación, se empiece a entregar falsos positivos. Para situaciones cotidianas esto es aceptable, pero cuando se intenta buscar el culpable de algún crimen o identificar sospechosos, uno no se puede permitir esto; por consiguiente se debe agregar un componente crucial: El contexto. El contexto, puede ser desde la fecha hasta, como explora el paper *Detecting Suspicious Behavior: How to Deal with Visual Similarity through Neural Networks*, las conductas sospechas de las personas (Martínez-Mascorro et al., 2020).

Concluyendo este tópico, debemos mencionar a las firmas escritas, tan antiguas como la tinta, si bien su uso actualmente se ha reducido gracias a digitalización de los servicios (de casi todos los rubros, incluso salud) y un comercio *cash-less*, no se puede bajar la guardia, es por esto que los sistemas de validación / autenticación de firmas escritas son necesarios, si estos son emparejados con un cifrado asimétrico o firma digital, se tiene un sistema de autenticación robusto (Alam, 2016).

### B. DeepFake

Con el avance de la capacidad de hardware, se volvió mas fácil implementar arquitecturas (de redes neuronales) mas 'extremas'. Estas arquitecturas capaces de procesar *batches* (bloques de datos) de mas 10gb empezaron a ser prometedoras en el ambito de la creación de información artificial, análisis de patrones mas complejos y detección de características jamás pensadas. Sin embargo, se ha convertido no solo en un tira y afloja entre que tan bien podemos falsear imágenes o videos y que tan bien podemos identificar un video falso. Esto a convertido al *deep fake* en una amenaza a la seguridad y libertad de expresión (se puede destruir la imagen de alguien haciendo parecer dio una opinión o hizo algo malo) (Lyu, 2020). Sin embargo, nuestro tema es enfocarlo desde como evitar que un *deep fake* amenace nuestra seguridad y no como crearlo, por lo que debemos saber primero, que tipo de mecanismos se usan en la actualidad para mejorar la calidad de un *deep fake*.

Cuando se construye un *deep fake*, se crea información, usando una imagen o video original se busca construir o simular acciones que no se encuentran en los datos originales, esto genera artefactos en la imagen, además del ruido de toda la

vida, es por esto que se construyen redes neuronales profundas encargadas solo de eliminar estos defectos haciendolo mas indetectable (Huang et al., 2020).

Sin embargo por el lado de la detección de los *deep fake*, se tiene métodos que usan redes neuronales recurrentes capaces de mejorar la detección mientras veces analice los mismos datos (Güera and Delp, 2018), métodos mas demandantes pero mas robustos que emplean en análisis de datos en su espacio original (espacial y temporal) y no en el de frecuencias mediante la aplicación de redes convencionales (de Lima et al., 2020) o métodos mas ortodoxos basados en técnicas de aprendizaje de máquina (Maksutov et al., 2020).

## IV. TELECOMUNICACION Y ESPIONAJE

Esta sección sera una muy difusa, pues aun que ciertas aplicaciones puede señarse solo al ambito de las comunicaciones, tambien pueden ser extendidas al área de la seguridad informática, es por esto que se ira detallando los alcances estas aplicaciones y los resultados que ofrecen.

### A. Comunicación inter-personal

La comunicación entre individuos de una misma especie es crucial, si dicha especie es una del tipo social, sabiendo que los seres humanos una de estas especies, tener miembros de nuestra sociedad con impedimentos de habla o escucha nos genera una barrera de comunicación, y es aqui donde la vision computacional nos ayuda. Crear un sistema de traducción entre personas con dichos problemas que empleen el lenguaje de señas (Bohra et al., 2019) es una aplicación que ido progresando mas y mas conforme va pasando los años, se ha llegado a conseguir no solo un sistema de traducción, sino un sistema de comunicación en ambos sentidos en tiempo real (Bohra et al., 2019).

A esta situación de impedimentos físico se suma un impedimento social: la falta de comunicación con nuestros iguales, las causas de esta evasión de la comunicación con otras personas, reduciendolo a una comunicación con ciertas personas se suele notar cuando alguien llega a un establecimiento público y es incapaz de pedir indicaciones, solo atina a quedarse en silencio y seguir al grupo. Si la tecnología nos empieza a crear un distanciamiento no voluntario, debemos usarla tambien para notar estos cambios y evitarlos (Bohra et al., 2019).

### B. Telecomunicaciones

La comprensión de archivos multimedia es algo fundamental a la hora de transmitir información, si bien en muchos casos esto implica comprimir el video omitiendo fotogramas estáticos, tambien puede implicar la comprensión sin pérdida buscando que la gran mayoría de datos lleguen tal y como fueron capturados (Porat, 2010), aun que claro; esta situación es algo que parece ser mas concerniente al área de telecomunicaciones que de vision computacional. Sin embargo debemos recordar que las imágenes son una forma de información, y como tal pueden ser convertidos al espacio de frecuencias (Phonsri et al., 2015) y manejarse como una señal de toda la vida.

Sabiendo lo anterior, es posible mejorar señales de radio o en forma general, señales portadoras de información. Si obtenemos un espectrograma de una señal, al ser este un proceso bidireccional, es posible mejorar la calidad del histograma como imagen y así reducir el ruido en la señal de la cual fue creada (Phonsri et al., 2015).

Si extendemos esta capacidad para limpiar señales, podemos implementar métodos de corrección de errores, de forma que se pueda reconstruir la información faltante o corregir la que existe, para esto, como siempre se debe tratar a las señales originales como si de una imagen se tratase, ya que una imagen es una señal y una señal puede ser una imagen (Tian et al., 2020).

### C. Internet

Si la internet es la red de redes, y una red usa señales para comunicarse, es fácil pensar otras aplicaciones, una de ellas puede ser, analizar el tráfico de red durante un tiempo determinado, construir una imagen que represente de forma única este periodo, así de forma sucesiva hasta construir los datos necesarios para que al analizar estas imágenes, se pueda construir un sistema capaz de identificar cuando está sucediendo algún evento; digamos por ejemplo un ataque de denegación de servicio Tan et al. (2015).

Existen métodos para limitar el tráfico de red, así evitando los ataques *DoS* (denegación de servicio), sin embargo un método ampliamente usado para evitar el ataque mediante robots, es el empleo de un *CAPTCHA*. Este método usado para evitar ataques automatizados pone a prueba la percepción de la máquina, pidiéndole que identifique caracteres en una escena bastante complicada, pudiendo solo ser resuelta por humanos. Una tentativa para automatizar incluso este obstáculo diseñado solo para las máquinas es una segmentación basada en caracteres conectados, de forma que se pueda inferir el texto (de letras y/o números) se encuentra ofuscado en la imagen de prueba (Hussain et al., 2016).

### D. Espionaje y Privacidad

La estenografía es la acción de esconder algo a simple vista, esto puede usarse con fines de seguridad (validación) o con fines de comunicación entre personas de interés (tal vez espías) (Tran, 2002). Sin embargo un caso especial de lo que se podría llamar estenografía es un cifrado homomórfico, que es cuando teniendo un conjunto de datos, este se transforma en otro, con las mismas propiedades, ahora imaginemos que obtenemos una fotografía, si ciframos esta imagen de tal manera que transformado al espacio de frecuencias, luzca idéntica a otra imagen, se podría usar esta segunda imagen como portador de información. Dicha información podría ser desde datos que deseemos adicionar hasta la misma identidad de quien aparece en la imagen (Bian et al., 2020).

Pasemos a hablar de la generación de escenarios en 3D, digamos que un espía intenta conocer el mapa tridimensional de un lugar de interés; podría comenzar por buscar los planos, o tal vez por infiltrarse en la casa y obtener muchas fotografías del mismo. Sin embargo, gracias a los avances de la visión computacional es posible recrear un entorno 3D

usando microfones bidireccionales, microfones que podrían ser nuestros propios oídos o un dispositivo en ellos. La problemática es que al usar un radar, este necesita de un detector direccional, sin embargo con un nuevo enfoque de reconstrucción inteligente es posible que con un emisor y dos receptores se pueda reconstruir un entorno tridimensional, simplemente con cálculos de tiempo e inferencias asistidas por un modelo pre-entrenado (Christensen et al., 2019).

Un enfoque similar, donde se reconstruye información es el empleado por un equipo del MIT, orientado a extraer sonido de algún ambiente inaccesible, usando solo una filmación del lugar. Esto es posible al medir las variaciones en la iluminación y movimiento de algún objeto referencial, digamos una bolsa de papas fritas, al existir un cambio en la presión del aire en el interior del lugar de interés, este cambio moverá ligeramente la superficie de la bolsa, al tener registrado estas variaciones es posible, reconstruir un audio, que aun que con fallas, es información nueva de utilidad para los fines que sean convenientes (Davis et al., 2014).

### REFERENCIAS

- Alam, S. (2016). Suis: An online graphical signature-based user identification system.
- Aneja, S. and Nießner, M. (2020). Generalized zero and few-shot transfer for facial forgery detection.
- Aydin, I. and Othman, N. A. (2017). A new IoT combined face detection of people by using computer vision for security application. In *2017 International Artificial Intelligence and Data Processing Symposium (IDAP)*. IEEE.
- Bian, S., Wang, T., Hiromoto, M., Shi, Y., and Sato, T. (2020). Ensei: Efficient secure inference via frequency-domain homomorphic convolution for privacy-preserving visual recognition.
- Bohra, T., Sompura, S., Parekh, K., and Raut, P. (2019). Real-time two way communication system for speech and hearing impaired using computer vision and deep learning. In *2019 International Conference on Smart Systems and Inventive Technology (ICSSIT)*. IEEE.
- Chaki, A., Prashant, M., and Sen, P. (2010). A comprehensive market analysis on camera and illumination sensors for image processing and machine vision applications. In *2010 International Conference on Computational Intelligence and Communication Networks*. IEEE.
- Christensen, J. H., Hornauer, S., and Yu, S. (2019). Batvision: Learning to see 3d spatial layout with two ears.
- Cote, M. and Albu, A. B. (2017). Teaching computer vision and its societal effects: A look at privacy and security issues from the students' perspective. In *2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*. IEEE.
- Davis, A., Rubinstein, M., Wadhwa, N., Mysore, G., Durand, F., and Freeman, W. T. (2014). The visual microphone: Passive recovery of sound from video. *ACM Transactions on Graphics (Proc. SIGGRAPH)*, 33(4):79:1–79:10.
- de Lima, O., Franklin, S., Basu, S., Karwowski, B., and George, A. (2020). Deepfake detection using spatiotemporal convolutional networks.

- Fox, G., Liu, W., Kim, H., Seidel, H.-P., Elgharib, M., and Theobalt, C. (2020). VideoforensicsHQ: Detecting high-quality manipulated face videos.
- Giudice, O., Litrico, M., and Battiato, S. (2020). Single architecture and multiple task deep neural network for altered fingerprint analysis.
- Goldenfein, J. (2019). The profiling potential of computer vision and the challenge of computational empiricism. In *Proceedings of the Conference on Fairness, Accountability, and Transparency*, FAT\* '19, page 110–119, New York, NY, USA. Association for Computing Machinery.
- Güera, D. and Delp, E. J. (2018). Deepfake video detection using recurrent neural networks. In *2018 15th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS)*, pages 1–6.
- Harikrishnan, J., Sudarsan, A., Sadashiv, A., and Ajai, R. A. (2019). Vision-face recognition attendance monitoring system for surveillance using deep learning technology and computer vision. In *2019 International Conference on Vision Towards Emerging Trends in Communication and Networking (ViTECoN)*. IEEE.
- Huang, Y., Juefei-Xu, F., Wang, R., Guo, Q., Ma, L., Xie, X., Li, J., Miao, W., Liu, Y., and Pu, G. (2020). Fakepolisher: Making deepfakes more detection-evasive by shallow reconstruction.
- Hussain, R., Gao, H., Shaikh, R. A., and Soomro, S. P. (2016). Recognition based segmentation of connected characters in text based CAPTCHAs. In *2016 8th IEEE International Conference on Communication Software and Networks (ICCSN)*. IEEE.
- Lyu, S. (2020). Deepfake detection: Current challenges and next steps. In *2020 IEEE International Conference on Multimedia & Expo Workshops (ICMEW)*. IEEE.
- Maksutov, A. A., Morozov, V. O., Lavrenov, A. A., and Smirnov, A. S. (2020). Methods of deepfake detection based on machine learning. In *2020 IEEE Conference of Russian Young Researchers in Electrical and Electronic Engineering (EIconRus)*. IEEE.
- Manendez, J., Salgado, L., Rendon, E., and Garcia, N. (1999). Motorway surveillance through stereo computer vision. In *Proceedings IEEE 33rd Annual 1999 International Carnahan Conference on Security Technology (Cat. No.99CH36303)*. IEEE.
- Martínez-Mascorro, G. A., Ortiz-Bayliss, J. C., and Terashima-Marín, H. (2020). Detecting suspicious behavior: How to deal with visual similarity through neural networks.
- Othman, N. A. and Aydin, I. (2017). A new IoT combined body detection of people by using computer vision for security application. In *2017 9th International Conference on Computational Intelligence and Communication Networks (CICN)*. IEEE.
- Pezoa, W. G. and Dominguez, M. J. (2017). Combined approach using artificial vision and neural networks for facial recognition. In *2017 CHILEAN Conference on Electrical, Electronics Engineering, Information and Communication Technologies (CHILECON)*. IEEE.
- Phonsri, S., Mukherjee, S. S., and Sellathurai, M. (2015). Computer vision and bi-directional neural network for extraction of communications signal from noisy spectrogram. In *2015 IEEE Conference on Antenna Measurements & Applications (CAMA)*. IEEE.
- Porat, M. (2010). A computational approach to multimedia communication networks. In *2010 International Conference on Networking and Information Technology*. IEEE.
- Potgieter, M. and Niekerk, J. V. (2012). The use of computer vision technologies to augment human monitoring of secure computing facilities. In *2012 Information Security for South Africa*. IEEE.
- Qian, J., Jiang, W., Luo, H., and Yu, H. (2019). Stripe-based and attribute-aware network: A two-branch deep model for vehicle re-identification.
- Quiros, A. R. F., Abad, A., Bedruz, R. A., Uy, A. C., and Dadios, E. P. (2015). A genetic algorithm and artificial neural network-based approach for the machine vision of plate segmentation and character recognition. In *2015 International Conference on Humanoid, Nanotechnology, Information Technology, Communication and Control, Environment and Management (HNICEM)*. IEEE.
- S, S., Yamuna, S., and George, S. N. (2020). An iot based active building surveillance system using raspberry pi and nodemcu.
- Sage, K. and Young, S. (1998). Computer vision for security applications. In *Proceedings IEEE 32nd Annual 1998 International Carnahan Conference on Security Technology (Cat. No.98CH36209)*. IEEE.
- Schindler, A., Lindley, A., Jalali, A., Boyer, M., Gordea, S., and King, R. (2020). Multi-modal video forensic platform for investigating post-terrorist attack scenarios.
- Spinoulas, L., Mirzaalian, H., Hussein, M., and AbdAlmageed, W. (2020). Multi-modal fingerprint presentation attack detection: Evaluation on a new dataset.
- Sumi, Y., Ishiyama, Y., and Tomita, F. (2002). Hyper frame vision: a real-time vision system for 6-dof object localization. In *Object recognition supported by user interaction for service robots*. IEEE Comput. Soc.
- Tan, Z., Jamdagni, A., He, X., Nanda, P., Liu, R. P., and Hu, J. (2015). Detection of denial-of-service attacks based on computer vision techniques. *IEEE Transactions on Computers*, 64(9):2519–2533.
- Tian, Y., Pan, G., and Alouini, M.-S. (2020). Applying deep-learning-based computer vision to wireless communications: Methodologies, opportunities, and challenges.
- Tran, N. (2002). Hiding functions and computational security of image watermarking systems. In *Proceedings 15th IEEE Computer Security Foundations Workshop. CSFW-15*. IEEE Comput. Soc.
- Yoshimoto, H., Date, N., Arita, D., and Taniguchi, R. (2004). Confidence-driven architecture for real-time vision processing and its application to efficient vision-based human motion sensing. In *Proceedings of the 17th International Conference on Pattern Recognition, 2004. ICPR 2004*. IEEE.
- Yu, C.-H., Gao, F., Liu, C., Huynh, D., Reynolds, M., and Wang, J. (2019). Quantum algorithm for visual tracking. *Physical Review A*, 99(2).
- Zubal, M., Lojka, T., and Zolotova, I. (2016). IoT gateway and

industrial safety with computer vision. In *2016 IEEE 14th International Symposium on Applied Machine Intelligence and Informatics (SAMI)*. IEEE.