Anexo 1. Formato de protocolo de investigación

1. **Datos generales de la propuesta**

   - **Nombre de la persona proponente:**
     Edgar Gonzalez Fernandez

   - **Título de la propuesta:**
     Análisis forense de imágenes y videos

   - **Línea de Generación y Aplicación del Conocimiento (LGAC) de INFOTEC en la que incide la propuesta:**
     Analítica de datos e información

   - **Periodo de ejecución** Del 01/01/2022 al 31/12/2023**:**

2. **Descripción de la propuesta:**

   - **Resumen** (ejecutivo)**:**
     La existencia de dispositivos móviles con cámaras de alto rendimiento y potentes aplicaciones de procesamiento de imágenes facilita la alteración de las imágenes digitales con fines maliciosos. El objetivo principal del trabajo de tesis es identificar las modificaciones comunes y novedosas utilizadas para alterar las imágenes digitales y el vídeo y proponer nuevos enfoques para detectarlas. En general, las técnicas de detección hacen uso de características extraídas de bloques de imágenes, para después ejecutar pruebas estadísticas que ayuden a determinar los bloques que han sido modificados. Todos los enfoques aplicados se probarán en conjuntos de datos de acceso público para medir su precisión y eficiencia.

   - **Antecedentes:**
     Una clasificación sencilla de las técnicas de detección que es habitual en la literatura, considera dos tipos de análisis:
     - Tecnicas activas: Este enfoque verifica las marcas de agua o firmas dejadas por un dispositivo durante el proceso de generación de una imagen digital.

La autenticidad se verifica si se puede probar la integridad de la marca de agua, generalmente con el uso de técnicas criptográficas o esteganográficas. El mayor inconveniente de este enfoque reside en las limitadas capacidades de muchos dispositivos de imagen, por lo que su alcance es limitado.

- Técnicas pasivas: Este enfoque examina el contenido y las características de la imagen digital sin necesidad de conocimientos adicionales, como el dispositivo de origen, el uso de marcas de agua, o firmas digitales.

En el presente proyecto, centraremos nuestros esfuerzos en las técnicas pasivas.

Para poder abordar los problemas asociados a las categorías anteriormente descritas, se pueden considerar las siguientes técnicas:

- Análisis de metadatos (EXIF). Los metadatos refieren a información descriptiva de un archivo, como puede ser el tamaño de una imagen, longitud de un video, y geolocalización, entre mushos otros. En [2, 3, 4, 5], se estudia, por ejemplo, la información que deja software de edición, cambios en la estructura original de archivos, incoherencias en la información de los metadatos (velocidad de los fotogramas, tamaño, duración, etc.), o las diferencias en miniaturas (thumbnails) frente a las imágenes a tamaño completo.
- Ruido del sensor (PRNU). Todos los dispositivos digitales que capturan información del entorno físico, incorporan errores por distintos motivos, uno de ellos debido a imperfecciones desde el proceso de manufactura. Estas imperfecciones son consideradas como una ”huella digital” del dispositivo, lo cual es aprovechado en diversos trabajos, como [6, 7, 8, 9, 10, 11], donde se analizan diversas características del ruido del sensor para detectar manipulaciones, identificar dispositivos de origen, o agrupar de acuerdo a la huella digital.
- Artefactos de pre y post-procesamiento. Diversos algoritmos utilizados para generar imágenes de color, corregir iluminación, compresión de archivos, entre otros, dejan también rastros de su aplicación, lo que es aprovechado en diversos para identificar zonas de una imagen que han sufrido modificaciones, como en los siguientes trabajos: [12, 13, 14, 15].

- **Justificación:**
  Las modificaciones a imágenes obtenidas por cámaras fotográficas han existido desde hace mucho tiempo, y el poder que una imagen tiene para transmitir

información puede observarse en casi cualquier ámbito de la vida cotidiana: noticias, anuncios, redes sociales, periodismo, etc. Sin embargo, para el común de la gente resulta en muchos casos casi imposible distinguir una fotografía auténtica de una modificada. Aunado a esto, se observa un gran avance en el desarrollo de técnicas de edición de imágenes, audio y video, así como una gran oferta de herramientas que facilitan su empleo y dispositivos cada vez más poderosos y compactos, lo que permite su uso en casi cualquier momento. La información adquirida por estos dispositivos, puede ser utilizada para fines maliciosos, como ataques políticos, ejercer influencia negativa en la opinión pública, o incluso influir en procesos legales, lo que cobra especial relevancia teniendo en cuenta la gran disponibilidad de evidencia generada por dispositivos móviles, cámaras de seguridad, drones, entre otros,

No obstante, aún conociendo los posibles riesgos y perjuicios que presenta la disponibilidad de este tipo de software, menos conocidas son las técnicas y herramientas existentes dedicadas al análisis forense multimedia. Por tal motivo, resulta indispensable contar con recursos informáticos para descubrir manipulaciones maliciosas con el fin de proveer de procesos confiables para la validación de material digital.

- **Objetivo general:**
  El objetivo principal del proyecto de investigación consiste en el desarrollo e implementación de **técnicas de análisis forense** para imágenes considerando el estado del arte, así como el desarrollo de propuestas novedosas que permitan la detección de manipulaciones realizadas con herramientas y técnicas actuales.

- **Objetivos específicos:**
  Para lograr este objetivo global, se propone alcanzar los siguientes objetivos específicos:
  1. Reconocimiento de las principales herramientas, técnicas y propósitos utilizados por entidades maliciosas para manipular imágenes digitales.
  2. Identificación de las ventajas y desventajas de los algoritmos actuales frente a las modificaciones comunes.
     a. Identificar las hipótesis restrictivas que dificultan el éxito de los algoritmos existentes.

   b. Reconocer las técnicas novedosas de manipulación (deep fakes, realidad aumentada y disminuida, ediciones profesionales) y analizar en busca de características que permitan un análisis estadístico.
3. Proporcionar nuevas técnicas para las manipulaciones no detectadas o para mejorar la precisión y la tasa de detección.
   a. Identificación de las características significativas que proporcionan información estadística fiable.
   b. Utilización de las nuevas tendencias en el análisis de datos basados en redes neuronales e inteligencia artificial para mejorar los resultados.
4. Desarrollo y aplicación de los algoritmos propuestos para proporcionar un conjunto de herramientas para la investigación forense de imágenes digitales.

- **Metas:**
  1. Obtener una alta eficiencia y éxito en los algoritmos propuestos, poniéndolos a prueba utilizando datasets de acceso público,
  2. Identificar la necesidad y en dado caso generar datasets ad-hoc para lograr un entendimiento adecuado de los artefactos generados por las modificaciones estudiadas.
  3. Crear una línea de investigación en el área de procesamiento de imágenes mediante la inclusión de estudiantes en la resolución de problemas relacionados con el procesamiento de imágenes.

- **Metodología:**
  La revisión del estado del arte pone de manifiesto una metodología general para realizar la autenticación pasiva en imágenes digitales utilizando el ruido del sensor, que se perfecciona posteriormente según el análisis específico. El enfoque inicial se basa en el análisis de los artefactos de interpolación cromática. La metodología que consta de 4 fases:
  1. Extracción de ruido. La imagen inspeccionada I' se modela como una señal ruidosa. Entonces el primer problema a tratar es recuperar la señal original I sin el ruido ε introducido en la formación. Formalmente,
     $$I' = I + \varepsilon$$
     donde I es el original y ε es el ruido.
  2. Cálculo de errores. Se obtiene una matriz de error, resultante de la diferencia entre la imagen original y la estimación.

3. Extracción de características. La matriz de error se segmenta en bloques de un tamaño predefinido para su análisis y clasificación individual mediante métodos estadísticos. El tamaño del bloque debe seleccionarse adecuadamente, ya que el análisis de bloques muy grandes suele ignorar las manipulaciones pequeñas, mientras que los bloques demasiado pequeños dan lugar a resultados estadísticamente inexactos, lo que hace imposible una detección correcta.
4. Segmentación. Finalmente, con las características extraídas se decide si la imagen ha sido modificada. La identificación de las zonas en las que se ha modificado la imagen depende en gran medida de la eficacia de los 3 pasos anteriores.
5. Evaluación de resultados: La eficacia y el éxito de los algoritmos propuestos se probarán utilizando conjuntos de datos disponibles públicamente, aunque para el análisis inicial y la detección de errores, también se pueden generar conjuntos de datos ad hoc.

A fin de realizar la experimentación necesaria que valide las técnicas forenses desarrolladas, se consideran los recursos computacionales y fuentes de información que se describen a continuación:

- Herramientas computacionales: Dado el avanzado estado de desarrollo y la facilidad de uso, se ha seleccionado el lenguaje de programación Python para su aplicación. Las librerías de Python, OpenCV [17] y SciPy [18] se utilizan para leer, almacenar y analizar imágenes y videos. Más experimentación usando redes neuronales se llevará a cabo usando el framework TensorFlow. De ser posible, se espera proporcionar implementaciones que hagan uso de tarjetas gráficas, para mejorar la eficiencia.
- Fuentes de datos: Varios datasets que contienen modificaciones junto con el archivo original y la verdad sobre el terreno (para las imágenes) están disponibles públicamente. Algunos de ellos realizan modificaciones específicas, mientras que otros pueden utilizar varias técnicas para obtener un resultado más realista. A continuación se enumeran algunos de los conjuntos de datos considerados para validar los métodos propuestos:
    - CASIA Dataset[1], contiene ejemplos de empalme en pequeñas imágenes con alta compresión.

[1] https://paperswithcode.com/dataset/casia-v2

- Realistic Tampering Dataset[2] que contiene diversas modificaciones en imágenes de alta calidad (formato crudo).
- Columbia Uncompressed Image Splicing Detection Evaluation Dataset[3]. Contiene modificaciones de copy-move y empalme de imágenes utilizando varias marcas de cámaras.

Una extensa lista de más conjuntos de datos para la manipulación de imágenes se puede encontrar en [1].

- **Beneficios esperados:**
  Se espera fortalecer la LGAC de "Analítica de datos e información" incorporando líneas específicas de investigación en el área de procesamiento de imágenes. Aunado a esto, se espera poder generar lazos de colaboración que incorporen a investigadores de otras instituciones.

- **Resultados esperados:**
  Se espera obtener una producción inicial de proyectos técnicos en la Maestría en Ciencia de Datos e Información, así como producción científica relevante y una participación cada vez mayor de estudiantes.

3. Plan de actividades

- **Descripción de las actividades**

#	Actividad	Descripción de la actividad
1	Revisión del estado del arte	Se identifican en primera instancia los métodos de modificación usuales y las nuevas tendencias en modificación de imágenes. Posteriormente se analizan las técnicas de análisis forense que abordan cada una de estas modificaciones
2	Diseño de pruebas de análisis forense	Una vez identificadas las ventajas y desventajas de los algoritmos actuales frente a modificaciones comunes se propondrán nuevas técnicas para las manipulaciones no detectadas

[2] https://paperswithcode.com/dataset/casia-v2
[3] https://www.ee.columbia.edu/ln/dvmm/downloads/authsplcuncmp/

		o para mejorar la precisión y la tasa de detección.
3	Implementación de técnicas	Se implementarán los algoritmos propuestos para proporcionar un conjunto de herramientas para la investigación forense de imágenes digitales.
4	Publicación de resultados	Redacción de artículo científicos, de divulgación, y proyectos con estudiantes abordando la problemática del análisis forense en imágenes

- **Descripción de las metas**

#	**Actividad**	**Meta**
1	2, 3	Obtener una alta eficiencia y éxito en los algoritmos propuestos, poniéndolos a prueba utilizando datasets de acceso público,
2	1, 2	Identificar la necesidad y en dado caso generar datasets ad-hoc para lograr un entendimiento adecuado de los artefactos generados por las modificaciones estudiadas.
3	4	Crear una línea de investigación en el área de procesamiento de imágenes mediante la inclusión de estudiantes en la resolución de problemas relacionados con el procesamiento de imágenes.

- **Productos (entregables)**

#	**Actividad**	**Producto**
1	1	Documento tipo “survey” con el estado del arte y técnicas relevantes de manipulación y detección.
2	2, 3	Código fuente de las implementaciones de técnicas existentes y propuestas

3	4	Artículos científicos JCR y factor de impacto Q! ó Q2
4	4	Artículos de congresos nacionales o internacionales relevantes
5	4	Proyectos de titulación de estudiantes de INFOTEC

4. Cronograma de actividades

#	**Actividad**	**Productos** (entregables)	**Fecha de inicio**	**Duración** (núm. de semanas)
1	1, 2	Revisión del estado del arte	01/01/2022	16
2	2, 3	Primer ronda de implementaciones	01/03/2022	20
2	4	Publicación de artículo JCR o de congreso	01/03/2022	16
3	4	Proyectos de titulación	01/01/2022	52
4	2,3	Diseño de técnicas forenses	01/07/2022	26
4	4	Proyectos de titulación	01/08/2022	52
5	4	Artículo científico	01/08/2022	20
6	2, 3	Código fuente implementaciones adicionales	01/01/2023	40
7	4	Artículo científico o de congreso	01/04/2023	20

5. Referencias

[1] P. Korus, Digital image integrity – a survey of protection and verification techniques, Digital Signal Processing 71C (2017).

[2] T. Gloe, A. Fischer, M. Kirchner, Forensic analysis of video file formats, in: Proceedings of the First Annual DFRWS Europe, volume 11, pp. S68–S76.

[3] M. Iuliani, D. Shullani, M. Fontani, S. Meucci, A. Piva, A video forensic framework for the unsupervised analysis of mp4-like file container, IEEE Transactions on Information Forensics and Security 14 (2019) 635–645.

[4] A. L. Sandoval Orozco, C. Quinto Huam án, D. Povedano Álvarez, L. J. García Villalba, A machine learning forensics technique to detect post-processing in digital videos, Future Generation Computer Systems 111 (2020) 199–212.

[5] C. Quinto Huamán, A. L. Sandoval Orozco, L. J. Garc ía Villalba, Authentication and integrity of smartphone videos through multimedia container structure

analysis, Future Generation Computer Systems 108 (2020) 15–33.

[6] P. Korus, J. Huang, Multi-scale analysis strategies in PRNU-based tampering localization, IEEE T Inf Foren Sec 12 (2017) 809–824.

[7] G. Chierchia, G. Poggi, C. Sansone, L. Verdoliva, PRNU-based forgery detection with regularity constraints and global optimization, in: Proceedings of the IEEE 15th International Workshop on Multimedia Signal Processing (MMSP), Santa Margherita di Pula, Sardinia, Italy, pp. 236–241.

[8] M. Chen, J. Fridrich, M. Goljan, J. Lukas, Determining image origin and integrity using sensor noise, IEEE T Inf Foren Sec 3 (2008) 74–90.

[9] H. Ravi, A. V. Subramanyam, G. Gupta, B. A. Kumar, Compression noise based video forgery detection, in: 2014 IEEE International Conference on Image Processing (ICIP), pp. 5352–5356.

[10] R. C. Pandey, S. K. Singh, K. K. Shukla, Passive forensics in image and video using noise features: A review, Digital Investigation 19 (2016) 1 – 28.

[11] E. A. Armas Vega, E. González Fernández, A. L. Sandoval Orozco, L. J. García Villalba, Image tampering detection by estimating interpolation patterns, Future Generation Computer Systems 107 (2020) 229–237.

[12] Y. Chen, C. Hsu, Detecting recompression of jpeg images via periodicity analysis of compression artifacts for tampering detection, IEEE Transactions on Information Forensics and Security 6 (2011) 396–406

[13] E. González Fernández, A. L. Sandoval Orozco, L. J. García Villalba, J. Hernandez-Castro, Digital image tamper detection technique based on spectrum analysis of CFA artifacts, Sensors 18 (2018) 2804.

[14] Y. Guo, X. Cao, W. Zhang, R. Wang, Fake colorized image detection, IEEE Transactions on Information Forensics and Security 13 (2018) 1932–1944.

[15] C. W. Park, Y. H. Moon, I. K. Eom, Image tampering localization using demosaicing patterns and singular value based prediction residue, IEEE Access 9 (2021) 91921–91933.

[16] W. Luo, J. Huang, G. Qiu, Jpeg error analysis and its applications to digital image forensics, IEEE Transactions on Information Forensics and Security 5 (2010) 480–491.

[17] G. Bradski, The OpenCV Library, Dr. Dobb's Journal of Software Tools (2000).

[18] E. Jones, T. Oliphant, P. Peterson, et al., SciPy: Open source scientific tools for Python, 2001.

[19] E. A. Armas Vega, E. González Fernández, A. L. Sandoval Orozco, L. J. García Villalba, Image tampering detection by estimating interpolation patterns, Future Gener. Comput. Syst. 107 (2020) 229–237

ATENTAMENTE

Dr. Edgar Gonzalez Fernandez
Subgerente de Docencia

C.c.p. **Mtro. Carlos Josué Lavandeira Portillo**, Director Adjunto de Innovación y Conocimiento. Presente.
Dr. Juan Antonio Vega Garfias, Subgerente de Innovación Gubernamental. Presente.