



APROBACIÓN DEL ASESOR

Quien suscribe:

Dr. Julio Cesar Huanca Marín por la presente:

CERTIFICA,

Que, el Bachiller en **Ingeniería de Sistemas**, HERBERTH VARGAS GUTIERREZ ha culminado satisfactoriamente el informe final de tesis intitulado: "SISTEMA BASADO EN LA BINARIZACIÓN DE IMÁGENES CON ALGORITMO ADAPTATIVO PARA LA CONSERVACIÓN DE LAS PARTIDAS DE MATRIMONIO DE LA PARROQUIA ANDAHUAYLAS" para optar el Título Profesional de Ingeniero de Sistemas.

Andahuaylas, 14 de agosto de 2020

Dr. Julio Cesar Huanca Marín
Asesor

Bach. Herberth Vargas Gutierrez
Tesista

ANEXO 33



DECLARACIÓN JURADA DE AUTENTICIDAD

Yo, HERBERTH VARGAS GUTIERREZ, **identificado (a) con DNI N° 43439201** de la **Escuela Profesional de Ingeniería de Sistemas Declaro bajo juramento que el Proyecto de Tesis Titulado:** “SISTEMA BASADO EN LA BINARIZACION DE IMÁGENES CON ALGORITMO ADAPTATIVO PARA LA CONSERVACIÓN DE LAS PARTIDAS DE MATRIMONIO DE LA PARROQUIA ANDAHUAYLAS”. Es auténtico y no vulnera los derechos de autor. Además, su contenido es de entera responsabilidad del autor (es) del proyecto, quedando la UNAJMA exenta de toda responsabilidad en caso de atentar contra la Ley de propiedad intelectual y derechos de autor.

Andahuaylas, 14 de agosto de 2020

Una firma manuscrita en tinta negra, que parece ser la del autor, sobre una línea punteada.

Firma

N° DNI: 43439201

E-mail: hvg.ingsis@gmail.com

N° Celular: 917873844

**UNIVERSIDAD NACIONAL JOSÉ MARÍA ARGUEDAS
FACULTAD DE INGENIERÍA
ESCUELA PROFESIONAL DE INGENIERÍA DE SISTEMAS**



**SISTEMA BASADO EN LA BINARIZACION DE IMÁGENES CON
ALGORITMO ADAPTATIVO PARA LA CONSERVACIÓN DE LAS
PARTIDAS DE MATRIMONIO DE LA PARROQUIA ANDAHUAYLAS**

Presentado por

HERBERTH VARGAS GUTIERREZ

**TESIS PARA OPTAR EL TÍTULO PROFESIONAL DE
INGENIERO DE SISTEMAS**

ANDAHUAYLAS – APURÍMAC – PERÚ

2020

**UNIVERSIDAD NACIONAL JOSÉ MARÍA ARGUEDAS
FACULTAD DE INGENIERÍA
ESCUELA PROFESIONAL DE INGENIERÍA DE SISTEMAS**



Presentado por

HERBERTH VARGAS GUTIERREZ

**SISTEMA BASADO EN LA BINARIZACIÓN DE IMÁGENES CON
ALGORITMO ADAPTATIVO PARA LA CONSERVACIÓN DE LAS
PARTIDAS DE MATRIMONIO DE LA PARROQUIA ANDAHUAYLAS**

Asesor:

Dr. Julio Cesar Huanca Marín

ANDAHUAYLAS – APURÍMAC – PERÚ

2020

Dedicatoria

Quiero dedicar el presente informe de tesis en especial para mi padre, madre y mis hermanos (a); con quienes estoy muy agradecido porque siempre tuvieron un apoyo incondicional y que, gracias a ellos, y la ayuda de Dios pude lograr los objetivos trazados.

Agradecimiento

La gratitud es un producto de la cultura; no es fácil hallarla entre la gente basta. Por ello agradezco a Dios ya que es el autor de nuestra vida y sabiduría; y estar siempre conmigo en cada paso que doy, por hacer cada vez más fuerte mis sentimientos e iluminar mi mente y crecer espiritualmente y formarme como una nueva persona de servicio a la sociedad.

También agradecer a la casa superior de estudios “Universidad Nacional José María Arguedas” especialmente a los docentes de la Escuela profesional de Ingeniería de Sistemas quienes me impartieron sus conocimientos en el transcurso de mi formación profesional para ser una persona de bien y cooperar con la sociedad de manera justa y oportuna.

Índice

Dedicatoria	III
Agradecimiento	IV
Índice.....	V
Índice de tablas.....	VII
Índice de figuras.....	VIII
RESUMEN.....	IX
ABSTRACT.....	X
INTRODUCCIÓN	1
CAPÍTULO 1. PROBLEMA DE INVESTIGACIÓN	2
1.2. DESCRIPCIÓN DEL PROBLEMA.....	2
1.2.1. Árbol de problemas.....	5
1.3. Formulación del problema.....	6
1.4. Justificación.....	6
1.5. Objetivos.....	7
1.5.1. General	7
1.5.2. Específicos	7
1.6. Delimitación de la Investigación	7
1.6.1. Viabilidad técnica.....	7
1.6.2. Viabilidad Operativa.....	7
1.6.3. Limitaciones.....	7
CAPÍTULO 2. ANTECEDENTES	8
2.1. Antecedentes o trabajos previos	8
CAPÍTULO 3. MARCO TEÓRICO	10
3.2. Definición conceptual.....	10
3.2.1. Adquisición de la imagen	10
3.2.2. Digitalización	10
3.2.3. Modelo de color de escala de grises.....	11
3.2.4. Sin compresión	12
3.2.5. Resolución.....	14
3.2.6. Dimensiones De La Imagen	15
3.2.7. Profundidad De Bits O Resolución Cromática.....	16
3.2.8. Fundamentos de procesamiento de imágenes digitales	17
3.2.9. Color.....	19
3.2.10. Mapa de color	20

3.2.11. Histograma de una imagen	20
3.2.12. Ruido en imágenes	20
3.2.13. Compresión de imágenes.....	21
3.2.14. Compresión y redundancia de datos	21
3.2.15. Redundancia de código.....	22
3.2.16. Redundancia entre pixeles	22
3.2.17. Umbralización	22
3.2.18. Binarización de una imagen	23
3.2.19. Mapa de Bits.....	23
3.2.20. Ruido en imágenes	24
3.2.21. Segmentación.....	24
3.3. Entornos de desarrollo.....	25
3.3.1. Python	25
3.3.2. Pillow.....	25
3.3.3. Anaconda Python.....	26
3.3.4. Star uml.....	26
3.3.5. Visual Studio Code.	26
3.3.6. Método de Otsu	27
CAPÍTULO 4. METODOLOGÍA DE LA INVESTIGACIÓN	28
4.2. Propuesta de solución	28
4.2.1. Herramientas para el desarrollo	28
4.2.2. Metodología RUP de desarrollo de la aplicación	28
4.2.3. Diseño con posprueba únicamente y grupo de control	29
4.3. DESARROLLO	30
CAPÍTULO 5. RESULTADOS	32
5.2. Resultados con imágenes de actas de diferentes tipos.....	42
5.3. Resultados de la aplicación	50
5.4. Almacenamiento.....	51
CONCLUSIONES	52
RECOMENDACIONES	53
REFERENCIAS BIBLIOGRÁFICAS.....	54
ANEXOS.....	56
1. Código Fuente.....	56
2. Resultados de pruebas con imágenes de diarios.....	63
3. Partidas Originales	66

Índice de tablas

Tabla 1: Cálculo de píxeles.....	15
Tabla 2: Cálculo de tonos de color.....	17
Tabla 3: Colores RGB	19
Tabla 4: Etapas del procesamiento digital de imágenes.	29
Tabla 5: Cálculo de tonos de color.....	38
Tabla 6: Comparación de umbrales calculados	42
Tabla 7: Espacio de almacenamiento	51

Índice de figuras

Figura: 1 Imagen Original y su representación binaria.....	11
Figura: 2 Modelo de color escale de gris.....	12
Figura: 3 Ilustración a resolución de 100, 200 y 300 ppp.....	15
Figura: 4 Imágenes bitonal y escala de grises.....	16
Figura: 5 Imagen original y su representación binaria	17
Figura: 6 Espacio de colores RGB.....	19
Figura: 7 Diferentes tipos de ruido afectando a una imagen.	21
Figura: 8 imágenes, sus histogramas y sus diagramas de correlación.....	22
Figura: 9 Umbral global en histogramas bimodal.....	23
Figura: 10 Metodología RUP de desarrollo de la aplicación.....	29
Figura: 11 Caso 01 monedas.bmp.....	33
Figura: 12 Caso 02 - habitacion.jpg	34
Figura: 13 Caso 03 - paleta.jpg.....	35
Figura: 14 Caso 04 - lena.jpg	36
Figura: 15 Caso 05 - logoepris.jpg	37
Figura: 16 Imagen escaneada de una página de diario	39
Figura: 17 Imagen procesada con el método Otsu	40
Figura: 18 Imagen procesada con el método a APT.....	41
Figura: 19 Región con zona oscura del acta01 con el método Otsu.....	43
Figura: 20 Región con zona oscura del acta01 con el método APT	43
Figura: 21 Región del texto del acta10 con el método Otsu	44
Figura: 22 Región del texto del acta10 con el método APT	44
Figura: 23 Histograma de la imagen acta01	45
Figura: 24 Histograma de la imagen acta02.....	45
Figura: 25 Histograma de la imagen acta03.....	46
Figura: 26 Histograma de la imagen acta04.....	46
Figura: 27 Histograma de la imagen acta05.....	47
Figura: 28 Histograma de la imagen acta06.....	47
Figura: 29 Histograma de la imagen acta07	48
Figura: 30 Histograma de la imagen acta08.....	48
Figura: 31 Histograma de la imagen acta09.....	49
Figura: 32 Histograma de la imagen acta10.....	49
Figura: 33 Ventana principal del prototipo	50

RESUMEN

La conservación de documentos usando medios digitales es una necesidad, sobre todo en lugares donde sufren deterioro afectando a su legibilidad y a la salud de quienes los manipulan. Las parroquias conservan actas o partidas de los sacramentos que administran, entre ellos el matrimonio. Esta es información valiosa que requiere ser conservada de forma eficiente y ahorrando espacio de almacenamiento usando medios digitales.

El presente estudio propone un método de binarización de imágenes con un algoritmo adaptativo para la conservación de documentos de partidas de matrimonio de la parroquia de Andahuaylas. Se evalúa el algoritmo de Otsu y en base a este se propone e implementa un algoritmo y un método de umbralización adaptativo progresivo. La aplicación del mismo se hace mediante un prototipo desarrollado en el lenguaje de programación Python.

El método propuesto permite obtener imágenes binarizadas en blanco y negro que se ajustan al tipo de imagen propio de una digitalización de documentos manuscritos y a la reducción de espacio de almacenamiento de hasta 12% conservando la legibilidad de la información registrada documento original.

ABSTRACT

The preservation of documents using digital media is a necessity, especially in places where they suffer deterioration affecting their legibility and the health of those who handle them. The parishes keep records or records of the sacraments they administer, including marriage. This is valuable information that needs to be conserved efficiently and saving storage space using digital media.

The present study proposes an image binarization method with an adaptive algorithm for the preservation of marriage certificate documents from the Andahuaylas parish. The Otsu algorithm is evaluated and based on this, an algorithm and a progressive adaptive thresholding method are proposed and implemented. Its application is done by means of a prototype developed in the Python programming language.

The proposed method allows to obtain binarized black and white images that conform to the type of image typical of a digitization of handwritten documents and to a reduction of storage space of up to 12% while preserving the legibility of the information registered in the original document.

INTRODUCCIÓN

La digitalización de documentos es una necesidad creciente en los tiempos de hoy. Esto puede ser tan simple como escanearlos y guardarlos en un medio de almacenamiento como un disco rígido.

Es deseable sin embargo hacerles un procesamiento previo con el fin de reducir el peso de los archivos resultantes ya que la imagen escaneada contiene información irrelevante sobre el color del soporte o papel, así como las marcas de deterioro o suciedad que no es deseable almacenar.

En el presente trabajo de investigación se propone un método de binarización adaptativo y progresivo APT y se contrasta con un método de binarización conocido el cual es el método Otsu.

Para esto se diseña el procedimiento y se programa para poder aplicarlo en diferentes situaciones y con diferentes tipos de imágenes. En una primera etapa se hacen pruebas con imágenes generales con diferentes características para evaluar su efectividad en comparación con un método conocido. En una segunda etapa se comprueba con un dataset de imágenes escaneadas de documentos de texto para comparar los valores obtenidos. En una tercera etapa se aplica a documentos de actas o partidas de diferentes fuentes y tipos de archivo, para hacer una inspección visual y del histograma de los resultados.

CAPÍTULO 1. PROBLEMA DE INVESTIGACIÓN

1.2. DESCRIPCIÓN DEL PROBLEMA

El término “conservación”, en el ámbito de los archivos y bibliotecas, hace referencia a todas aquellas medidas destinadas a proteger adecuadamente los documentos, con el fin de prolongar su utilización en condiciones óptimas durante el mayor tiempo posible (Carpallo Bautista, 2000). A lo largo del tiempo la documentación presenta deterioros debido a múltiples factores (físicos, químicos, biológicos, físico–mecánicos y circunstanciales), que están destruyendo la información contenida en ellos y así mismo, desarrolla riesgos para la integridad física y salud de los archivistas (MONSALVE, 2006).

Por tanto (Rodríguez Mora, 2007) indica que, la preservación involucra a todas aquellas acciones preventivas que contribuyan al buen estado del acervo documental. Quiere esto decir, que la preservación se anticipa al daño que puede sufrir el documento por factores o mecanismos que propician su alteración o deterioro, cuya mayoría son de carácter extrínseco, o sea que se ciernen en el medio que rodea al documento. En consecuencia, la preservación busca eliminar el daño ocasionable, no el ocasionado.

En la actualidad las parroquias cuentan con almacenes de acervo documentario en físico, en la que se encuentran documentos muy antiguos y modernos, referentes a las actividades eclesíásticas realizadas en un total abandono así como indica (Rodríguez Mora, 2007) Ya que, en su mayoría, las empresas del grupo no cuentan con los espacios adecuados para realizar el tratamiento técnico a los documentos y poder realizar una efectiva conservación y preservación de sus archivos, se está generando una desorganización y mal manejo de la documentación y, en última instancia descuido y deterioro de la información. Con todo esto, es muy importante tener en cuenta que, en la actualidad, las compañías generan documentos que por sus valores secundarios pueden llegar a formar parte del patrimonio documental de grupo e incluso del país, teniendo claro que para lograr transmitir la información vital que contienen los mismos se debe contar con los mecanismos y programas apropiados para preservar y conservar los documentos, sin importar su formato.

Cada uno de estos documentos está organizado de manera física en estantes. Para la preservación no se cuenta con un plan sistemático de limpieza, mantenimiento y digitalización menos una sistematización, el personal a cargo del manejo y preservación de los documentos de archivo realiza su trabajo de manera empírica, sin el apoyo de un manual de organización archivística que guíe y oriente las acciones que corresponden para una organización, conservación y acceso adecuado a los documentos.

Los métodos preventivos afectan al entorno del objeto. Procuran para él un hábitat idóneo que le ponga, en la medida de lo posible, a salvo de incidencias

degradantes, manteniendo su integridad química y física: su permanencia y durabilidad. La eficacia de una política preventiva está en razón directa del conocimiento de las características del objeto, de las causas que producen su patología y en el empleo de materiales y medios técnicos que eviten esta última (Hendriks & UNESCO, 1984).

La diócesis de Abancay cuenta en su totalidad con 21 parroquias distribuidas en cuatro provincias: Abancay, Chincheros, Aymaraes y Andahuaylas. Estas provincias no son ajenas a este problema ya que la administración y gestión de los documentos eclesiales, así como los bienes patrimoniales que conforman el legajo eclesiástico propiamente dicho, según su naturaleza, deben administrarlos los sacerdotes conforme a las normas de las leyes eclesiásticas. El manejo según (Vivas, Ma, & Pérez, 2011) Prácticamente desde sus orígenes, la Iglesia ha intentado conservar los documentos por ella generados. En un primer momento, el objetivo no era otro que el de intemporalizar la doctrina. Han sido posteriormente los archiveros los encargados de organizarla y describirla. En otro sentido, al tratarse la Iglesia católica de una institución privada, el acceso a sus archivos es restringido, ocasionando dificultades a los investigadores. En contraposición, en estos archivos se custodia documentación valiosa y antigua dado que como es sabido la Iglesia es una de las instituciones más vetustas que han existido. Sin embargo, no podemos obviar diversas vicisitudes acaecidas a lo largo de los siglos, las cuales ocasionaron pérdidas de incalculable valor a su patrimonio. Aún a pesar de ello nuestras parroquias, catedrales, monasterios, conventos, seminarios, etc. conservan un legado documental de infinito valor histórico. En el contexto de la sociedad actual de las parroquias de nuestra

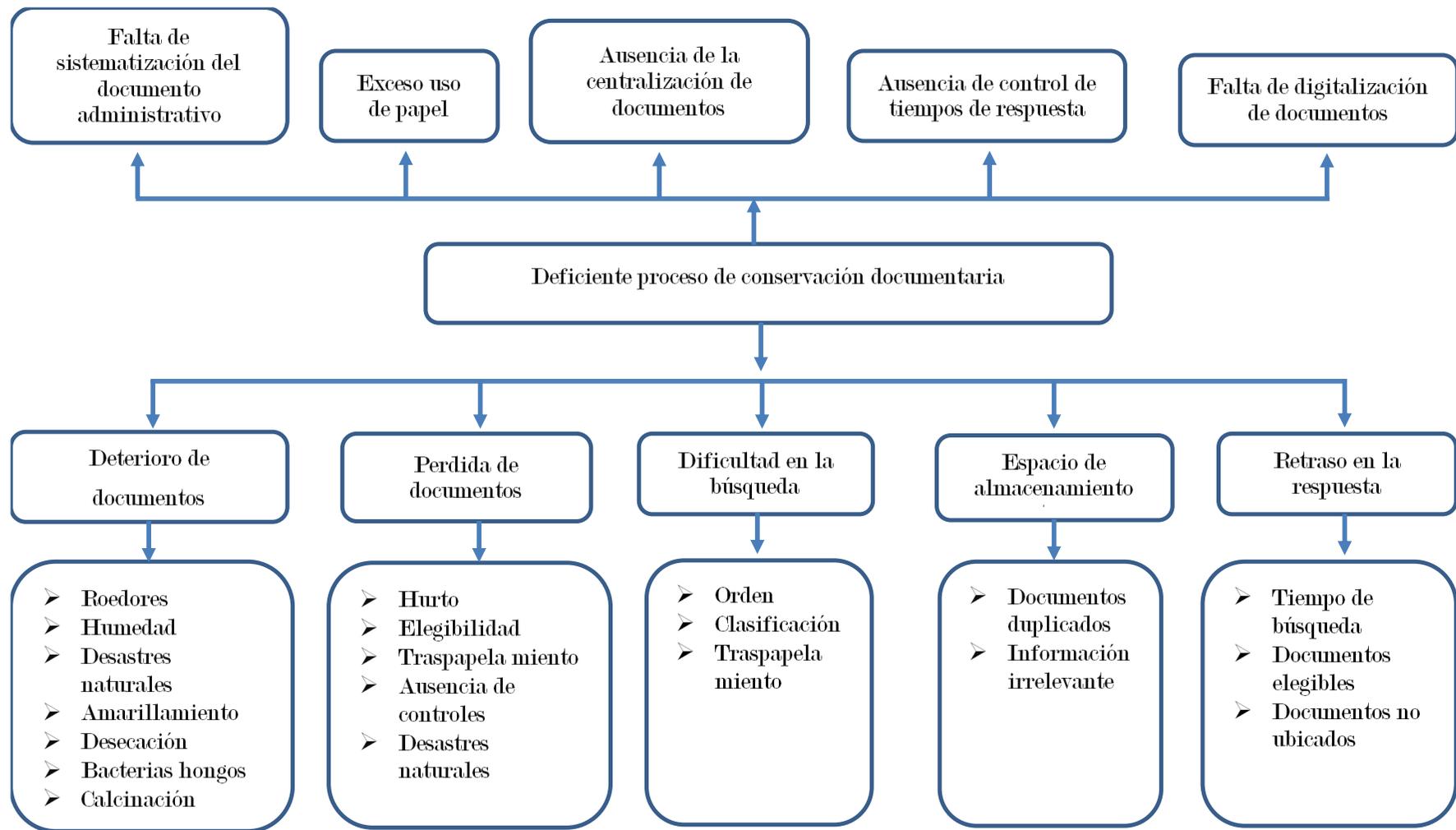
provincia son aun incapaces de desprenderse de sus elaboraciones manuales y mentales preconcebidas la que refleja la resistencia al cambio en los procesos de elaboración de documentos en papel.

Los documentos contienen información que constituye un recurso valioso y un activo importante de la organización. La adopción de un criterio sistemático para la gestión de documentos resulta esencial para las organizaciones y la sociedad, a fin de proteger y conservar los documentos como evidencia de sus actos. Un sistema de gestión de documentos se convierte en una fuente de información sobre las actividades de la organización que puede servir de apoyo a actividades posteriores y a la toma de decisiones, a la vez que garantiza la rendición de cuentas frente a las partes interesadas presentes y futuras. Los documentos permiten a las organizaciones (Sastre Natividad, Perez, & Loyola Vargas, 2015).

Una de las tareas principales en la digitalización es el tema de la preservación y conservación de los documentos. Hoy en día, la era digital aporta las nuevas tecnologías y tendencias para la mejor gestión de los documentos y sus contenidos. Por ello según (Carlos, Cobos, & Javeriana, 2007).Es necesario el desarrollo de estudios e investigaciones que se puedan aplicar a procesos de digitalización de imágenes, en especial para soportes que están más expuestos al deterioro por causas del tiempo, como son los documentos de valor histórico en soporte papel: “Las fuentes varían en cuanto tamaño, formato, medios y estado de conservación; pueden o no basarse en textos; pueden o no contener ilustraciones cuyas características pueden ser muy variadas o las fuentes pueden o no poseer color. Por eso, el sistema de gestión de documentos tiene que garantizar la autenticidad, fiabilidad, integridad y disponibilidad de los documentos, identificándolos en el contexto de las actividades de la organización (Alonso, García, & Lloveras, 2007).

Si se trata de la conversión digital de materiales que están en peligro o dañados, el objetivo es, en primer lugar, la creación de reproducciones exactas de estos originales en un soporte de larga duración y no la selección de materiales de acuerdo con la demanda. Estas reproducciones deben satisfacer tanto a los usuarios actuales como a los potenciales usuarios del futuro y, por tanto, deben ser de una gran calidad y poseer una gran estabilidad física de modo que puedan mantenerse a lo largo del tiempo (IFLA e ICA, 2002).

1.2.1. Árbol de problemas



1.3. Formulación del problema

Deficiente conservación de la partida de matrimonio en la parroquia de Andahuaylas

1.4. Justificación

Hoy en día el hecho de manejar de manera óptima el legajo y preservar los documentos no solo es tener organizado ya que esto involucra factores como son: la naturaleza del papel, condiciones del ambiente, la manipulación, desastres naturales, factor humano, factor biológico y factores químicos que conllevan el deterioro de los documentos visto estos aspectos lo prioritario debe ser garantizar que los archivos con el paso del tiempo se mantengan en un buen estado y para ello la digitalización de documentos ha llegado a ser una de las grandes soluciones para las empresas que precisan el uso de gran cantidad de éstos, el tratamiento de datos a llegado a un punto muy importante ya que han surgido saltos tecnológicos de relevancia que coadyuvan y facilitan el manejo de la información de manera más práctica y fácil.

El documento digitalizado mejora el tamaño y nitidez de los documentos originales, de forma que los usuarios pueden descifrar detalles de los documentos que son muy difíciles de ver en un original. La digitalización es beneficiosa también en la preservación de la información contenida dentro del material de archivo. como por ejemplo a la recuperación de archivos que piden los usuarios. Sin el archivo documental digitalizado el personal tiene que atender al usuario, buscar el documento que este le pide y entregarlo, posteriormente cuando el usuario finaliza su tarea el personal debe devolver el documento a su carpeta original. Con el archivo documental digitalizado el propio investigador accede a los documentos por medio de un sistema de gestión documental que muestra los archivos y documentos digitalizados.

Y mediante compresión de las imágenes digitales reducimos el tamaño de los archivos de imagen que hoy en día es cada vez mayor. La compresión de una fotografía es la reducción de los datos digitales que no resultan necesarios e importantes. Esta compresión permitirá almacenar mayor número de imágenes al conseguir que los archivos resultantes no ocupen mucho espacio. Y por consiguiente la administración de la parroquia se optimizará mediante el buen manejo de la información, ¿cómo se puede almacenar información de todas las actividades, procesos realizados por una Iglesia y el seguimiento a cada uno de los miembros, de tal forma que dicha información este siempre disponible y actualizada para mejorar la organización y toma de decisiones?

Es importante que las parroquias tomen conciencia de la necesidad de llevar un buen manejo de su información actualizada, ordenada, sistematizado y que perdure en el tiempo, ya que le permitirá conservar sus documentos de manera virtual y que tengan un acceso en tiempo real y que se pueda cruzar información con las otras sedes parroquiales del ámbito nacional.

1.5. Objetivos

1.5.1. General

Desarrollar un sistema basado en la binarización de imágenes con algoritmo adaptativo para la conservación de las partidas de matrimonio de la parroquia Andahuaylas.

1.5.2. Específicos

- Utilizar métodos adecuados de tratamiento de imágenes
- Implementar un algoritmo adaptativo.
- Almacenar las imágenes procesadas.

1.6. Delimitación de la Investigación

1.6.1. Viabilidad técnica

Es viable técnicamente ya que se dispone de las herramientas necesarias para el desarrollo del sistema integrado, las herramientas de desarrollo para la implementación están basados en software libre y disponibles para su descarga gratuita en sus páginas oficiales respectivamente, también se cuenta con los equipos computacionales adecuados, servicios de internet, energía eléctrica y otros.

1.6.2. Viabilidad Operativa

Es viable operativamente ya que se cuenta con el personal adecuado para la binarización de las imágenes, modelamiento de la base de datos e implementación del sistema de manera óptima y eficiente.

1.6.3. Limitaciones

El presente proyecto abarca desde la recolección de archivos(documento), escaneo del documento, tratamiento de la imagen, binarización, almacenamiento y recuperación del mismo.

CAPÍTULO 2. ANTECEDENTES

2.1. Antecedentes o trabajos previos

La conservación del patrimonio documental que atesoran los diferentes archivos, bibliotecas y museos de Cuba constituyen un factor de interés y preocupación en la política cultural; la pérdida de estos documentos traería como consecuencia que las generaciones futuras permanezcan ignorantes a muchos de los elementos que han formado el acervo cultural. La conservación es uno de los cuatro pilares que abarcan las tareas de un archivero. Objetivo: Profundizar en los conocimientos sobre la conservación preventiva en los Centros de Información de Salud. Conclusiones: Una conservación preventiva eficiente y permanente evita pérdidas del material cultural de archivos y bibliotecas. La conservación del patrimonio documental nacional depende en gran medida de la concientización de los recursos humanos más que de los medios existentes. (Yasminda Rodríguez Hernández, 2017)

Según (Ruiz, 2017) Uno de los fines de los centros que custodian el patrimonio bibliográfico y documental es el de su conservación a perpetuidad. Sin embargo, en lo que se refiere a los soportes convencionales, su degradación y las exigencias de difusión plantean dudas acerca de que la conservación deba realizarse exclusivamente en los soportes originales. En este sentido se plantean los medios digitales como una solución que puede ser duradera si se enmarca dentro de un plan de preservación integral como el que proporciona el modelo OAIS de la norma UNE-ISO 14721:2015. La gestión de documentos, resulta una oportunidad de alto valor para cualquier organización, ante la necesidad de ejercer la toma de decisiones y acometer procesos de la administración, basados en la evidencia documental, con la garantía de una mayor transparencia y credibilidad. (Laurencio, 2018)

Un documento histórico es un testimonio material de un hecho o acto realizado en funciones por instituciones o personas, registrado en una unidad de información en cualquier tipo de soporte sea papel, cintas, discos magnéticos, fotografías, etc. en lengua natural o convencional. También se puede decir que es el testimonio de una actividad humana fijada en un soporte, dando lugar a una fuente archivística (Vázquez Ávila, 2018).

La gestión documental, como área de gestión responsable de un control eficaz y sistemático de la creación, organización, conservación, y cuidado de todo documento

para el uso y su adecuada difusión. También se puede definir como el conjunto de normas técnicas y prácticas usadas para administrar el flujo de documentos de todo tipo en una organización, permitir la recuperación de información desde ellos, determinar el tiempo que los documentos deben guardarse, eliminar los que ya no sirven y asegurar la conservación indefinida de los documentos más valiosos, aplicando principios de racionalización y economía (Vázquez Ávila, 2018).

La gestión de documentos de forma sistematizada es capaz de personalizar soluciones adaptadas a la medida de las necesidades que una organización demande. Por lo tanto, no se trata de una solución específica sino adaptable, permitiendo parametrizar cualquier necesidad a la hora de estructurar documentos, gestionar contenidos o automatizar procesos con ilimitados requerimientos. La importancia de los documentos y el volumen de información es lo que motiva a buscar soluciones innovadoras que ofrezcan ventajas y valor añadido sobre los sistemas tradicionales de archivo y almacenamiento (Gomez Ruiz, 2017)

CAPÍTULO 3. MARCO TEÓRICO

3.2. Definición conceptual

Un computador solo es capaz de manejar formatos digitales de información. La información puede estar codificada originalmente (códigos ASCII, EBCDIC, etc.), o ser de tipo analógica (documento original), que deberemos convertir en una representación digital que permita ser procesada por un Sistema de Gestión Electrónica de Documentos (SGED).

Las imágenes digitales son fotos electrónicas tomadas de una escena o escaneadas de documentos - manuscritos, textos impresos, ilustraciones, fotografías, etc. -. Se realiza una muestra de la imagen y se confecciona un mapa en forma de cuadrícula de puntos elementales (píxeles). A cada píxel se le asigna un valor tonal (negro, blanco, gris, color) el cual

está representado por un código binario (ceros y unos). Estos dígitos binarios ("bits") para cada píxel, son almacenados en una secuencia y con frecuencia reducidos a una representación matemática (compresión). Luego el computador interpreta esta secuencia de bits para reproducir una versión analógica para su visualización o impresión.

Este pequeño manual trata de describir los parámetros a considerar de una imagen digital, de las compresiones y los formatos de archivo más utilizados, y de los periféricos de captura (con detalle de tipos, interfaz, parámetros y funcionalidades) que permiten conseguir la digitalización de imágenes en los SGED

3.2.1. Adquisición de la imagen

Es el proceso por el cual la imagen es obtenida, mediante cámara, impresión de pantalla, escaneo de documentos, foto tomada por cámara web, etc.

Un aspecto que es importante considerar, es que de acuerdo a la calidad de imagen que se adquiera, corresponderá el buen o mal reconocimiento de las letras.

3.2.2. Digitalización

La digitalización es la conversión de una imagen a números binarios con la ayuda de un ordenador, se proporciona a la computadora una descripción detallada de los objetos que comprenden la escena, junto con las especificaciones de la cámara o el escáner.

Las imágenes digitales son fotos electrónicas tomadas de una escena o escaneadas de documentos - manuscritos, textos impresos, ilustraciones, fotografías, etc. Se realiza una muestra de la imagen y se confecciona un mapa en forma de cuadrícula de puntos elementales (píxeles). A cada píxel se le asigna un valor tonal (negro, blanco, gris, color) el cual está representado por un código binario (ceros y unos). Estos dígitos binarios ("bits") para cada píxel, son almacenados en una secuencia y con frecuencia reducidos a una representación matemática (compresión). Luego el computador interpreta esta secuencia de bits para reproducir una versión analógica para su visualización o impresión.

Parámetros de la imagen digital

Una imagen digital puede ser representada mediante una matriz bidimensional de números. En la Figura 1 se muestra una imagen original y su representación en una matriz de puntos bitonales (píxeles), en los que el negro se representa como "0" y el blanco se representa como "1".

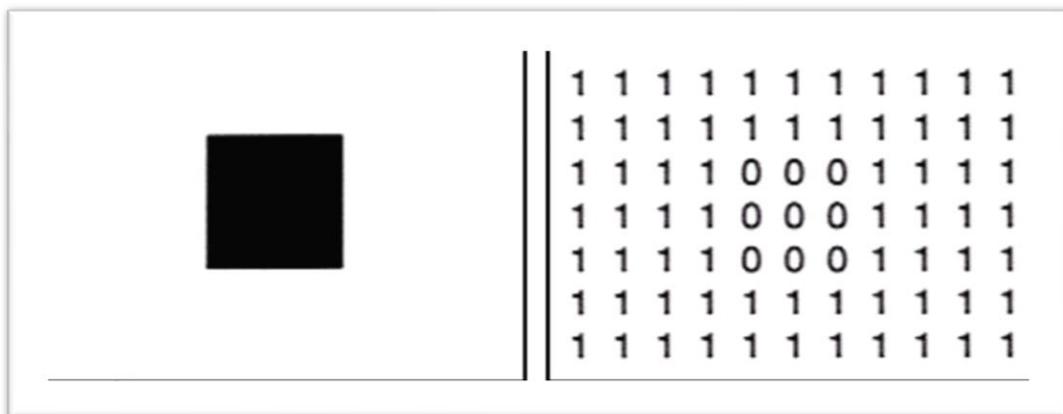


Figura: 1 Imagen Original y su representación binaria

3.2.3. Modelo de color de escala de grises

El modelo de color de escala de grises define el color en un único canal, la luminosidad, que se mide con valores comprendidos en un rango definido similar al de la Fig.3.5 (0-255, 0-1). Cada uno de los colores de la escala de grises tiene un valor equivalente de los componentes rojo, verde y azul del modelo de color RGB. Al cambiar el color de una

fotografía a escala de grises se crea una fotografía comúnmente conocida como en blanco y negro



Figura: 2 Modelo de color escala de gris

Formatos de imagen de mapa de bits

En función de la estructura de la información almacenada, la imagen se puede almacenar en diferentes formatos. Los más extendidos son:

3.2.4. Sin compresión

3.2.4.1. BMP

Windows Bitmap o Bit Mapped Picture (BMP). Se caracterizan por ser muy poco eficientes en su uso de espacio en disco, pero pueden mostrar un buen nivel de calidad. No son utilizables en páginas Web debido a su gran tamaño en relación a su resolución. Dependiendo de la profundidad de color que tenga la imagen cada píxel puede ocupar uno o varios bytes. Generalmente estos archivos se suelen transformar en otros formatos, tales como JPEG, GIF o PNG, que utilizan algoritmos de compresión para obtener un menor tamaño del archivo.

Con compresión

3.2.4.2. PNG

Portable Network Graphic Format (PNG) es un formato gráfico basado en un algoritmo de compresión sin pérdida para bitmaps no sujeto a patentes. Usan la extensión .png y se ha establecido como un tipo Multipurpose Internet Mail Extensions (MIME). Sus principales características son:

- Admite imágenes indexadas con transparencia de 1 bit
- Generalmente PNG realiza una buena compresión, aunque en algunas ocasiones se generan ficheros de mayor tamaño. Esta característica tiene

relación con la posibilidad de generar imágenes en este formato con distintos niveles de compresión, que permite diferentes tamaños de fichero según las necesidades, a expensas de una posible pérdida de velocidad en su representación como consecuencia del proceso de descompresión. En cualquiera de sus niveles, la calidad de la imagen es idéntica.

- El PNG admite formatos con una profundidad de color de millones de colores y canal alfa, lo que proporciona rangos de color muy ricos y precisos y una disposición de valores de transparencia intermedios.
- No soporta animación

3.2.4.3. JPG/JPEG

El formato de archivos Join Photographic Experts Group (JPEG) se abrevia a JPG debido a que algunos sistemas sólo aceptan tres letras de extensión. Es un algoritmo diseñado para comprimir imágenes con 24 bits de profundidad o en escala de grises. Se trata de un algoritmo de compresión con pérdida. Esto significa que al descomprimir la imagen no obtenemos exactamente la misma imagen que teníamos antes de la compresión. Una de las características que hacen muy flexible el JPEG es el poder ajustar el grado de compresión. JPEG es también el formato de fichero que utiliza este algoritmo para comprimir imágenes. El algoritmo de compresión JPEG se basa en dos defectos visuales del ojo humano. Uno es el hecho de que es mucho más sensible al cambio de luminancia que en la crominancia, es decir, notamos mejor los cambios de brillo que de color. El otro es que notamos con más facilidad pequeños cambios de brillo en zonas homogéneas que en zonas donde la variación es grande, como en los bordes de los cuerpos.

3.2.4.4. GIF

Graphics Interchange Format (GIF). Es un formato gráfico utilizado ampliamente en la World WideWeb, tanto para imágenes como para animaciones. GIF consiguió ser muy popular porque podía usar el algoritmo de compresión LZW (Lempel Ziv Welch) para realizar la compresión de imagen. Sus principales características son:

- Profundidad de color: 8 bits máximo
- Permite transparencia de 1 bit. – Permiten hacer animaciones simples.
- Permite utilizar entrelazado en imágenes.

3.2.4.5. TIFF

Tagged Image File Format (TIFF). Su nombre se debe a que los ficheros contienen, además de los datos de la imagen, una serie de etiquetas en las que se archiva información sobre las características de la imagen que sirven para su posterior tratamiento. Estas etiquetas describen el formato que puede ser de distinta naturaleza:

- Binarias (blanco y negro). Adecuadas para texto, entre otras.
- Niveles de gris. Utilizadas para representación de tonos continuos como imágenes en blanco y negro.
- Paleta de colores. Útil para el almacenamiento de diseños gráficos con un número limitado de colores
- Color real. Adecuadas para imágenes de tono continuo, como fotos en color.

Las etiquetas también nos indican el tipo de compresión aplicado. Este tipo de compresión puede utilizar distintos métodos como PackBits, Huffman modificado, LZW o JPEG. También puede no presentar ningún tipo de compresión

3.2.5. Resolución

La resolución es la frecuencia espacial a la cual se realiza la muestra de una imagen digital. La medida de la resolución se determina por el número de píxeles leídos en una distancia lineal de una pulgada (2,54 cm) en el documento digitalizado.

Las unidades más utilizadas son “puntos por pulgada” - ppp - (dots per inch - dpi -), o “píxeles por pulgada” - ppp - (pixels per inch - ppi -). Las resoluciones de digitalización de documentos están normalmente en el rango de 50 a 600 ppp.

Las resoluciones de escaneado más frecuentemente utilizadas en documentos blanco y negro en SGED son 200, 300 y 400 ppp. En la digitalización de imágenes color de alta calidad, son típicas resoluciones de 1.200 a 2.400 ppp.

En la Figura 2 se muestra una ilustración a resoluciones de 100, 200 y 300 ppp. Las mejoras en las imágenes de mayor resolución son apreciables.

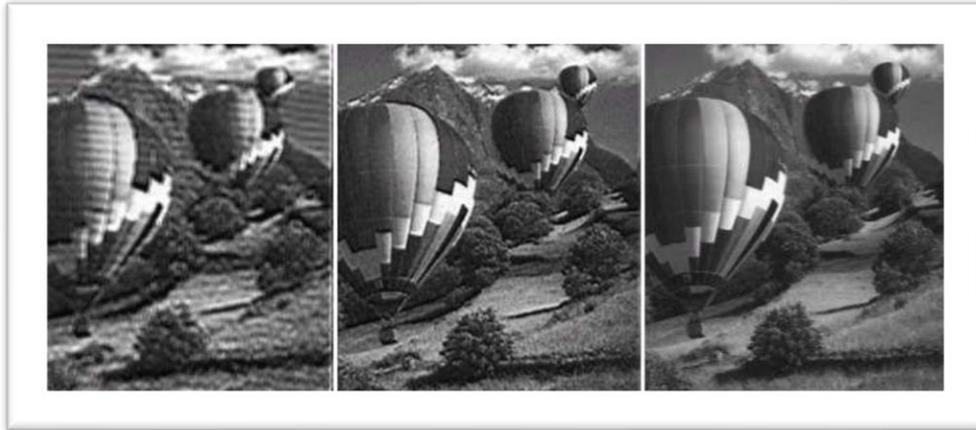


Figura: 3 Ilustración a resolución de 100, 200 y 300 ppp.

3.2.6. Dimensiones De La Imagen

Las dimensiones de la imagen son las medidas horizontales y verticales de la misma, expresadas en píxeles. Pueden determinarse multiplicando tanto el ancho como la altura de la imagen (medido en pulgadas) por la resolución en ppi.

La tabla 01 muestra, por ejemplo, una página de tamaño DIN A4 en blanco y negro, digitalizada a resolución de 600 ppi, tendrá una dimensión total de:

Cálculo de dimensiones en píxeles	
Medidas	Equivalencias
Ancho: 21,0	cm = 8,27"
8,27" x 300 ppi	2.500 píxeles
Alto: 29,7 cm	11,69"
11,69" x 300 ppi	3.500 píxeles
La dimensión total del documento DIN A4 será	
2.500 x 3.500	8.750.000 píxeles

Tabla 1: Cálculo de píxeles

3.2.7. Profundidad De Bits O Resolución Cromática

La profundidad de bits especifica la cantidad de información de color que está disponible para cada píxel de una imagen. Cuanto mayor sea la profundidad de bits, mayor será la cantidad de tonos que puedan ser representados.

Las imágenes digitales se pueden digitalizar en blanco y negro (binario), a escala de grises o a color:

Binario, utilizando un bit por píxel para representar el blanco o el negro utilizando los valores 0 para el negro y 1 para el blanco.

En escala de grises, utilizando múltiples bits por píxel para representar las escalas de grises. El mejor nivel de escala de grises es de 8 bits por píxel, y en este nivel la imagen que se muestra puede seleccionar hasta 256 diferentes niveles de grises.

En Figura 3, distinguimos dos imágenes digitales: bitonal (1 bit) y escala de grises (8 bits).



Figura: 4 Imágenes bitonal y escala de grises

En la tabla 02 se muestra Estos son los cálculos binarios para la cantidad de tonos representados por profundidades de bits comunes:

Tabla de Tonos	
1 bit (2^1)	2 tonos
2 bits (2^2)	4 tonos
3 bits (2^3)	8 tonos
4 bits (2^4)	16 tonos
8 bits (2^8)	256 tonos
10 bits (2^{10})	1024 tonos
12 bits (2^{12})	4096 tonos
14 bits (2^{14})	16384 tonos
16 bits (2^{16})	65536 tonos
24 bits (2^{24})	16,7 millones de tonos

Tabla 2: Cálculo de tonos de color

3.2.8. Fundamentos de procesamiento de imágenes digitales

Píxel: Elemento básico de una imagen (picture element).

Imagen: Arreglo bidimensional de píxeles con diferente intensidad luminosa (escala de gris).

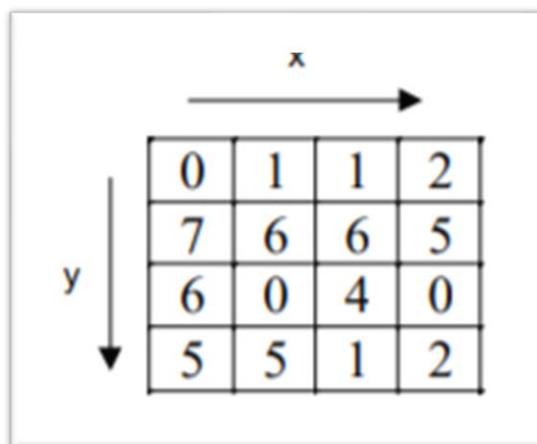


Figura: 5 Imagen original y su representación binaria

Si la intensidad luminosa de cada píxel se representa por n bits, entonces existirán 2^n escalas de gris diferentes.

Matemáticamente, una imagen se representa por $r = f(x, y)$, donde r es la intensidad luminosa del píxel cuyas coordenadas son (x, y) . Matemáticamente, un sistema para procesar imágenes se representa como $g(x, y) = T[f(x, y)]$.

Color: El color se forma mediante la combinación de los tres colores básicos rojo, azul y verde (en inglés RGB). A continuación, se presentan algunas definiciones básicas para comprender los espacios de color:

Brillo: Indica si un área está más o menos iluminada.

Tono: Indica si un área parece similar al rojo, amarillo, verde o azul o a una proporción de ellos.

Luminosidad: Brillo de una zona respecto a otra zona blanca en la imagen.

Croma: Indica la coloración de un área respecto al brillo de un blanco de referencia.

Para obtener una imagen a color deben transformarse primero los parámetros cromáticos en eléctricos y representar los colores, lo cual puede realizarse de diferentes maneras, dando lugar a diferentes espacios de colores o mapas de color.

Espacio RGB: se basa en la combinación de tres señales de luminancia cromática distinta: rojo, verde, azul (Red, Green, Blue). La forma más sencilla de obtener un color específico es determinar la cantidad de color rojo, verde y azul que se requiere combinar para obtener el color deseado, ver la figura 2; para lo cual se realiza la suma aritmética de las componentes: $X = R + G + B$, gráficamente representada por un cubo.

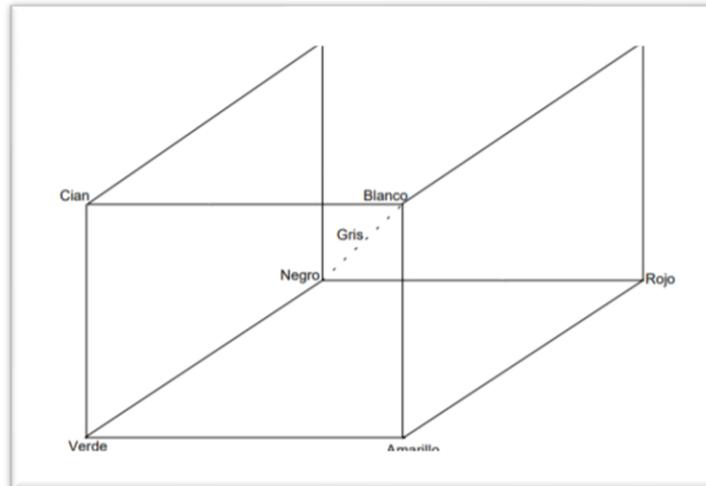


Figura: 6 Espacio de colores RGB

En la recta que une el origen con el valor máximo se encuentran ubicados los grises (escala de gris) debido a que sus tres componentes son iguales. Cuando una cámara adquiere una imagen a color, para cada píxel en color se tienen en realidad 3 componentes, una para cada uno de los colores básicos (rojo, verde y azul); la ganancia máxima para cada componente corresponde a la longitud de onda de los tres colores básicos.

3.2.9. Color

Un color puede definirse como la combinación de tres colores básicos: rojo, verde y azul, y expresarse mediante una tripleta de valores de 0 a 1 (R, G, B), donde R, G y B representan las intensidades de cada uno de los tres colores básicos rojo, verde y azul, respectivamente. En la tabla I se presentan ejemplos de colores definidos mediante estas tripletas.

Color	R	G	B
Blanco	1	1	1
Rojo	1	0	0
Amarillo	1	1	0
Verde	0	1	0
Turquesa	0	1	1
Gris	0.5	0.5	0.5
Rojo Oscuro	0.5	0	0
Azul	0	0	1
Aguamarina	0.5	1	0.83
Negro	0	0	0

Tabla 3: Colores RGB

3.2.10. Mapa de color

El mapa de color es una matriz de $n \times 3$, donde cada renglón es una tripleta de colores. El primer renglón corresponde al valor mínimo del eje de color y el último renglón al máximo. Al definir diferentes distribuciones de intensidad de los tres colores básicos, se crean diferentes mapas de color.

3.2.11. Histograma de una imagen

El histograma de una imagen es una representación del número de píxeles de cierto nivel de gris en función de los niveles de gris.

3.2.12. Ruido en imágenes

Todas las imágenes tienen cierta cantidad de ruido, la cual se puede deber a la cámara o al medio de transmisión de la señal. Generalmente el ruido se manifiesta como píxeles aislados que toman un nivel de gris diferente al de sus vecinos. Los algoritmos de filtrado que se verán más adelante permiten eliminar o disminuir este ruido. El ruido puede clasificarse en los siguientes tipos:

Gaussiano: produce pequeñas variaciones en la imagen; generalmente se debe a diferentes ganancias en la cámara, ruido en los digitalizadores, perturbaciones en la transmisión, etc. Se considera que el valor final del píxel sería el ideal más una cantidad correspondiente al error que puede describirse como una variable aleatoria gaussiana.

Impulsional (sal y pimienta): el valor que toma el píxel no tiene relación con el valor ideal, sino con el valor del ruido que toma valores muy altos o bajos (puntos blancos y/o negros) causados por una saturación del sensor o por un valor mínimo captado, si se ha perdido la señal en ese punto. Se encuentran también al trabajar con objetos a altas temperaturas, ya que las cámaras tienen una ganancia en el infrarrojo que no es detectable por el ojo humano; por ello las partes más calientes de un objeto pueden llegar a saturar un píxel.

Multiplicativo: La imagen obtenida es el resultado de la multiplicación de dos señales.

En la figura 6 se muestran los diferentes ruidos afectando a una imagen.



Figura: 7 Diferentes tipos de ruido afectando a una imagen.

3.2.13. Compresión de imágenes

En nuestros días, una enorme cantidad de información se maneja por medios digitales. La compresión es necesaria en aplicaciones de almacenamiento, transmisión, tiempo real.

3.2.14. Compresión y redundancia de datos

Se requiere distinguir entre los términos datos e información. Para nosotros, los datos son el medio por el que se transporta la información. Esto quiere decir que se pueden utilizar distintas cantidades de datos para representar la misma cantidad de información.

El término compresión de datos se refiere al proceso de reducir la cantidad de datos requeridos para representar una cantidad dada de información.

Cuando un conjunto de datos contiene más del mínimo necesario para transmitir la información, decimos que existe redundancia de datos.

3.2.15. Redundancia de código

Utilizaremos el ya familiar concepto de histograma de niveles de gris de una imagen para mostrar cómo éste puede ayudar en gran medida en la construcción de códigos que reducen la cantidad de datos utilizados para representar la imagen.

3.2.16. Redundancia entre píxeles

Consideremos las imágenes de la figura 6. Sus histogramas son prácticamente idénticos, con 3 rangos de nivel de gris dominantes. Se puede usar un código del tipo que acabamos de ver para reducir la redundancia de datos. Sin embargo, este proceso de codificación no alteraría el nivel de correlación entre los píxeles de las imágenes.

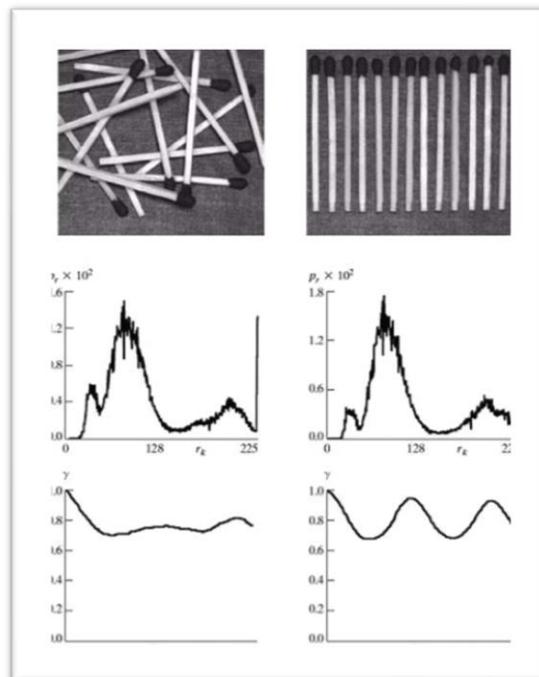


Figura: 8 imágenes, sus histogramas y sus diagramas de correlación

3.2.17. Umbralización

De acuerdo con (Solomon y Breckon, 2011), la umbralización de imágenes es una transformación puntual que produce imágenes en blanco y negro $B(x, y)$ asignando valores de 1 y 0 a los píxeles de una imagen $I(x, y)$, dependiendo si están por debajo o sobre un valor de umbral T .

$$B(x, y) = \begin{cases} 0 & \text{si } I(x, y) > T \\ 1 & \text{si } I(x, y) < T \end{cases}$$

3.2.17.1. Umbralización global

Los métodos de umbralización globales emplean un solo valor umbral que afecta a todo el histograma de la imagen. A partir del valor obtenido, cada píxel es asignado al color blanco o negro dependiendo de la comparación de su valor de brillo con el valor umbral. En la Figura 1, se muestra el umbral en un histograma bimodal de una imagen en escala de grises.

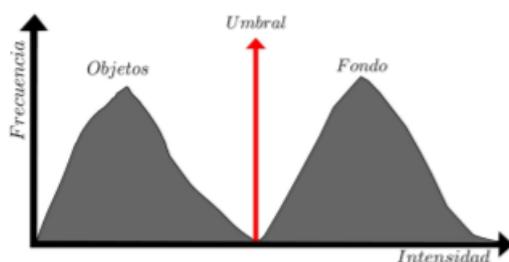


Figura: 9 Umbral global en histogramas bimodal

3.2.17.2. Umbralización local

Los métodos de umbrales locales o adaptativos trabajan seleccionando valores de umbrales en vecindarios de píxeles. Estos métodos parten del hecho de que los píxeles están altamente correlacionados entre sí en cada vecindario. Cada valor de umbral es utilizado para binarizar el vecindario de píxeles correspondiente y a partir de ellos obtener una imagen final binarizada.

3.2.18. Binarización de una imagen

La binarización de una imagen consiste en comparar los niveles de gris presentes en la imagen con un valor (umbral) predeterminado. Si el nivel de gris de la imagen es menor que el umbral predeterminado, se le asigna al píxel de la imagen binarizada el valor 0 (negro), y si es mayor, se le asigna un 1 (blanco). De esta forma se obtiene una imagen en blanco y negro. Generalmente se utiliza un umbral de 128 si se trabaja con 255 niveles de gris, pero en algunas aplicaciones se requiere de otro umbral. En la figura 30 se muestra un ejemplo de imagen binarizada.

3.2.19. Mapa de Bits

La imagen raster o mapa de bits es un archivo o estructura que representa generalmente una rejilla rectangular o matriz de píxeles en un monitor, papel u otro medio. Un mapa de bits se caracteriza por el ancho, el alto de la imagen en píxeles y el número de

canales por píxel, el cual determina el número de colores que puede representar. El color de cada píxel está definido individualmente y la información albergada corresponde a atributos como color, luminosidad o transparencia entre otros. Este conjunto de atributos utilizados para la representación de colores queda definido en los denominados modelos de color. Dichos modelos se explicarán en detalle más adelante. Este tipo de imagen tiene una resolución finita, pudiéndose aumentar gracias a la interpolación.

Esta técnica consigue aumentar tanto resolución como tamaño con una consiguiente pérdida de calidad y nitidez de la imagen, pero no siempre es una solución válida. Por este motivo las imágenes en mapas de bits necesitan un estudio de la resolución necesaria de antemano, para poder preparar la imagen acorde a esas necesidades.

A su vez, una reducción de sus dimensiones presenta pérdidas de información, quedando descartada la posibilidad de un regreso a las dimensiones iniciales sin tener un empeoramiento de la imagen. Este tipo de imagen, debido a sus atributos, es el más extendido en la captura de imágenes digitales y representación realista, puesto que permite un alto grado de detalle.

3.2.20. Ruido en imágenes

Las imágenes digitales comúnmente son contaminadas con errores indeseables. Estos errores son conocidos como ruido. El ruido en una imagen puede ser descrito mediante sus características probabilísticas. El ruido en las imágenes puede ser ocasionado durante la propia formación de la imagen o su transmisión. Además, el ruido puede ser en ciertos casos dependiente del contenido de las imágenes.

3.2.21. Segmentación

Una vez se ha realizado un procesamiento previo de la imagen capturada, el cual ha mejorado la calidad de ésta y/o de los elementos presentes, se busca segmentar la foto para facilitar el análisis digital o reconocimiento automático.

Este fin se logra dividiendo la imagen en zonas homogéneas referentes a una sola característica, como pueden ser el color o el brillo. Este proceso no es algo estandarizado y aplicable a cualquier imagen, depende del problema y del objetivo

que se persiga, siendo así necesario o no aplicar este proceso.

Como se ha mencionado aplicar este proceso no es imprescindible. Por ejemplo, si se aplica segmentación a una fotografía paisajística se crearán una multitud de regiones homogéneas, las cuales no transmitirán ninguna información relevante sobre la imagen.

3.3. Entornos de desarrollo

3.3.1. Python

Lenguaje de programación de propósito general, orientado a objetos, que también puede utilizarse para el desarrollo web.

Python es un lenguaje de scripting independiente de plataforma y orientado a objetos, preparado para realizar cualquier tipo de programa, desde aplicaciones Windows a servidores de red o incluso, páginas web. Es un lenguaje interpretado, lo que significa que no se necesita compilar el código fuente para poder ejecutarlo, lo que ofrece ventajas como la rapidez de desarrollo e inconvenientes como una menor velocidad.

En los últimos años el lenguaje se ha hecho muy popular, gracias a varias razones como:

- La cantidad de librerías que contiene, tipos de datos y funciones incorporadas en el propio lenguaje, que ayudan a realizar muchas tareas habituales sin necesidad de tener que programarlas desde cero.
- La sencillez y velocidad con la que se crean los programas. Un programa en Python puede tener de 3 a 5 líneas de código menos que su equivalente en Java o C.
- La cantidad de plataformas en las que podemos desarrollar, como Unix, Windows, OS/2, Mac, Amiga y otros.
- Además, Python es gratuito, incluso para propósitos empresariales.

3.3.2. Pillow

Python Imaging Library (abreviado como PIL) (en versiones más recientes conocidas como Pillow) es una biblioteca gratuita para el lenguaje de programación Python que agrega soporte para abrir, manipular y guardar muchos formatos de archivos de imágenes diferentes. Está disponible para Windows, Mac OS X y Linux. La última versión de PIL es 1.1.7, fue lanzada en septiembre de 2009 y es compatible con Python 1.5.2-2.7, con soporte de Python 3 para ser lanzado "más tarde".

Pillow ofrece varios procedimientos estándar para la manipulación de imágenes. Éstas incluyen:

- manipulaciones por pixel,
- enmascaramiento y manejo de la transparencia
- filtrado de imágenes, como borrosidad, contornos, suavizado o búsqueda de bordes,
- mejora de la imagen, como nitidez, ajuste de brillo, contraste o color,
- agregar texto a las imágenes y mucho más.

3.3.3. Anaconda Python

Anaconda es una suite de código abierto que abarca una serie de aplicaciones, librerías y conceptos diseñados para el desarrollo de la ciencia de datos con Python. Se trata de una distribución de Python que básicamente funciona como un gestor de entorno, de paquetes y que posee una colección de más de 720 cuya característica primordial es que son de código abierto.

- Anaconda Distribution se agrupa en cuatro sectores o soluciones tecnológicas:
- Anaconda Navigator: Interfaz gráfica de Anaconda Python
- Anaconda Project
- Librerías de Ciencia de Datos
- Conda: Gestor de código del Anaconda Python.

Anaconda spider

Es un entorno de desarrollo integrado y multiplataforma de código abierto (IDE) para programación científica en el lenguaje Python. Este IDE se liberó bajo la licencia de MIT. Spyder es extensible con complementos. Incluye soporte de herramientas interactivas para la inspección de datos e incorpora controles de calidad específicos de Python e instrumentos como Pyflakes, Pylint y Rope.

3.3.4. Star uml

Es una herramienta para el modelamiento de software basado en los estándares UML (Unified Modeling Language) y MDA (Model Driven Architecture), que en un principio era un producto comercial y que hace cerca de un año pasó de ser un proyecto comercial (anteriormente llamado plastic) a uno de licencia abierta GNU/GPL.

3.3.5. Visual Studio Code.

Es un editor de código fuente desarrollado por Microsoft para Windows, Linux y macOS. Incluye soporte para depuración, control de Git integrado, resaltado de sintaxis,

finalización de código inteligente, fragmentos de código y refactorización de código. También es personalizable, de modo que los usuarios pueden cambiar el tema del editor, los métodos abreviados de teclado y las preferencias. Es gratuito y de código abierto.

3.3.6. Método de Otsu

El método de Otsu, llamado así en honor a Nobuyuki Otsu que lo inventó en 1979, utiliza técnicas estadísticas, para resolver el problema. En concreto, se utiliza la varianza, que es una medida de la dispersión de valores – en este caso se trata de la dispersión de los niveles de gris.

El método de Otsu calcula el valor umbral de forma que la dispersión dentro de cada segmento sea lo más pequeña posible, pero al mismo tiempo la dispersión sea lo más alta posible entre segmentos diferentes. Para ello se calcula el cociente entre ambas variancias y se busca un valor umbral para el que este cociente sea máximo.

CAPÍTULO 4. METODOLOGÍA DE LA INVESTIGACIÓN

4.2. Propuesta de solución

4.2.1. Herramientas para el desarrollo

- Lenguaje de programación Python
- Librería Pillow python
- STAR UML
- Anaconda
- Spyder Python
- Visual Code

4.2.2. Metodología RUP de desarrollo de la aplicación

La estructura dinámica del RUP permite que el proceso de desarrollo se fundamentalmente iterativo en las cuatro fases.

Fase de inicio: (se define el alcance del proyecto)

Se define el alcance del proyecto con los clientes, se identifican los riesgos asociados al proyecto, se elabora el plan de las fases y el de la iteración posterior, se detalla de manera general la arquitectura del software.

Fase de Elaboración: (Definición, análisis y diseño)

Se diseña la solución preliminar, se selecciona los casos de uso que permiten definir la arquitectura base del sistema y se desarrollara el primer análisis del dominio del problema.

Fase de desarrollo o construcción: (implementación)

La función de esta fase es completar la funcionalidad del sistema, se clarifican los requisitos pendientes, se administran los cambios de acuerdo a las evaluaciones realizadas por los usuarios, y se realizan las mejoras para el proyecto.

Fase de transición: (fin del proyecto y pruebas)

Fase de cierre, el propósito es asegurar que le software esté disponible para los usuarios finales, se ajustan los errores y defectos encontrados en las pruebas de aceptación, se capacitan a los usuarios y se provee el soporte necesario.

Planear las 4 fases incluye: Asignación de tiempo, Hitos Principales, Iteraciones por Fases, Plan de proyecto.

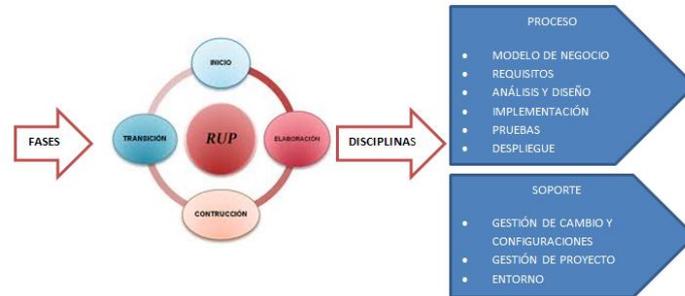


Figura: 10 Metodología RUP de desarrollo de la aplicación

4.2.3. Diseño con posprueba únicamente y grupo de control

Considerando que la manipulación de la variable independiente se realiza en dos grados (ausencia y presencia) se considera el diseño con post prueba únicamente y grupo de control.

Métodos de análisis de datos:

El método usado estará basado en las fases del procesamiento digital de imágenes, descritas a continuación:

Etapa	Descripción
Captura o adquisición de datos.	Mediante un scanner.
Preprocesamiento	Reparar en la imagen los desperfectos producidos, o no eliminados por el hardware.
Segmentación	Dividir una imagen (fotograma) en objetos que sean de nuestro interés de estudio.
Binarización	Generación de los bits
Normalización	Tratamiento de las imágenes obtenidas de la binarización
Almacenamiento	Almacenar los bits en la base de datos
Recuperación	Recuperación de los bits y reconstrucción del documento.
Comparación	Analizar la similitud del documento recuperado y el original.

Tabla 4: Etapas del procesamiento digital de imágenes.

4.3. DESARROLLO

La información obtenida a partir del procesamiento de imágenes tiene utilidad en muchos aspectos hoy en día. Tales como procedimientos médicos y sistemas de inteligencia artificial, y están siendo usadas para obtener análisis variados. El procedimiento propuesto demuestra las capacidades de estas técnicas de procesamiento, la técnica de umbralización adaptativa de Otsu se mejora con la técnica de **umbralización adaptativa progresiva (APT)** para construir imágenes en escala binaria (blanco y negro). Cada una de estas técnicas calcula recursivamente y obtiene un valor de umbral, para lo cual se construye una nueva imagen, en base a los niveles en grises de la imagen original.

La parte más importante de esta umbralización adaptativa puede encontrar un umbral que pueda segmentar la imagen en dos clases, luz y sombra. Primeramente, vamos a ver el método Otsu. Sin embargo, hay un proceso estadístico involucrado que hacer antes de alcanzar este punto. Primero la imagen en color necesita ser convertida a una imagen en escala de grises, luego se usa el histograma para acceder a la imagen. Entonces, para todos los valores de pixeles, lo dividimos en dos clases, el primer método calcula iterativamente el valor de la intensidad y lo multiplica por el número de ocurrencias. El segundo método toma iterativamente el valor del primer método y lo utiliza para sustraerlo del tamaño total de la imagen.

$$q_1(t) = \sum_{i=1}^t P(i)$$

$$q_2(t) = \sum_{i=t+1}^L P(i)$$

Ambos métodos son usados para calcular sus valores individuales para las medias

$$\mu_1(t) = \frac{\sum_{i=1}^t iP(i)}{q_1(t)}$$

$$\mu_2(t) = \frac{\sum_{i=t+1}^L iP(i)}{q_2(t)}$$

Entonces una vez que las medias son calculadas, estas se usan para calcular la varianza interclases para el estado actual

$$\sigma_b^2(t) = q_1(t)q_2(t)[\mu_1(t) - \mu_2(t)]^2$$

La varianza interclase se usa para encontrar el valor más grande para el umbral, el umbral obtiene el valor de 'i' cuando una nueva varianza mejora a la anterior. 'varmax' representa el umbral más grande en el momento actual, si este valor es mejorado, este se reemplaza por la varianza actual al cuadrado.

Cuando el proceso termina, hemos obtenido el valor del umbral para la imagen actual. Entonces iteramos a través de la imagen original, accediendo a sus píxeles. Si el valor de los píxeles actuales de la imagen el más grande que el valor del umbral, se le asigna el valor máximo para el blanco 255. De lo contrario será 0 para negro.

$$g(x, y) = \begin{cases} 1, & f(x, y) \geq T \\ 0, & x < T \end{cases}$$

Esto es todo lo que requiere el método de Otsu. Ahora para aplicar el método APT, debemos normalizar el histograma con respecto al valor del umbral, esto es conocido como cuantización. Entonces, para la imagen cuantizada, se deben repetir los pasos anteriores y calcular el nuevo valor de umbral, este debe ser significativamente menor que el valor de umbral calculado por el método Otsu. Este nuevo valor de umbral se usará en el resto del proceso para obtener la imagen APT.

Se hará la comprobación visual con 5 imágenes diferentes para evaluar el funcionamiento del método APT frente al método Otsu.

Una vez comprobado el resultado general se hará el procesamiento con documentos similares a las partidas de matrimonio escaneadas con diferentes niveles de exposición, contraste, legibilidad y deterioro.

CAPÍTULO 5. RESULTADOS

Para algunas imágenes de entrada donde el fondo está claramente diferenciado, los métodos Otsu y APT trabajan bien, los objetos son claramente distinguibles, mientras que otras imágenes donde el rango de color está más distribuido el resultado no se ve muy bien. Cuando comparamos las imágenes obtenidas con método Otsu versus el método APT, las imágenes APT siempre serán un poco más luminosas, y tienen un umbral menor, dada la naturaleza de la ecuación APT, donde se limita el monto de colores posibles en un cierto rango a partir del umbral obtenido por el método Otsu.

Primeramente, se hizo la comparación del método APT con el método Otsu con imágenes representativas diferentes a las partidas de matrimonio del estudio.

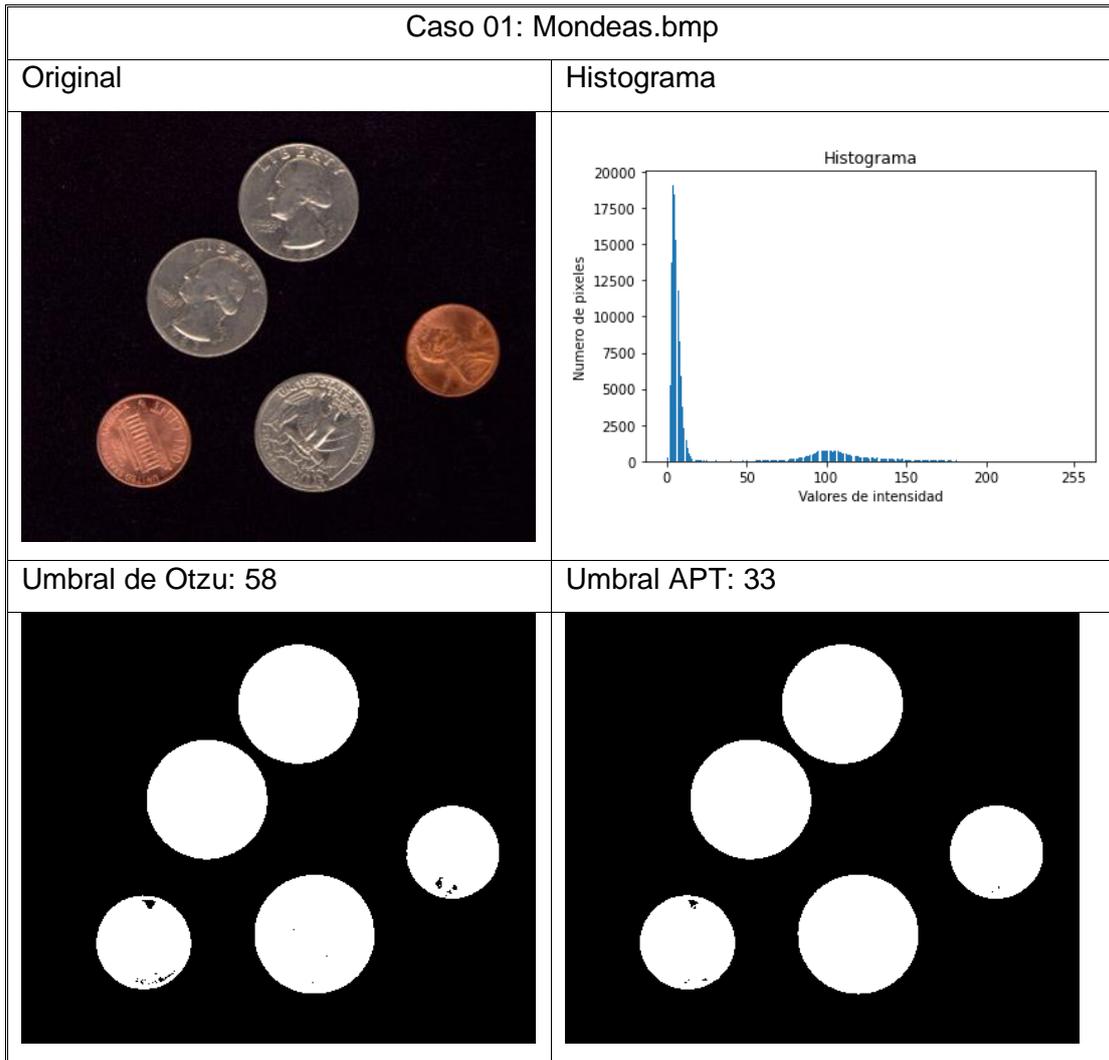


Figura: 11 Caso 01 monedas.bmp

En esta imagen los objetos son claramente distinguibles del fondo, el algoritmo adaptativo ha tenido éxito en la umbralización de las imágenes. Cada moneda se ha hecho blanca, independientemente de sus valores cuando se compara con el fondo oscuro. Debido a que el umbral para la imagen APT es significativamente menor que el umbral Otsu, ha sido capaz de remover algunos bits oscurecidos, por lo que los círculos son prácticamente blancos.

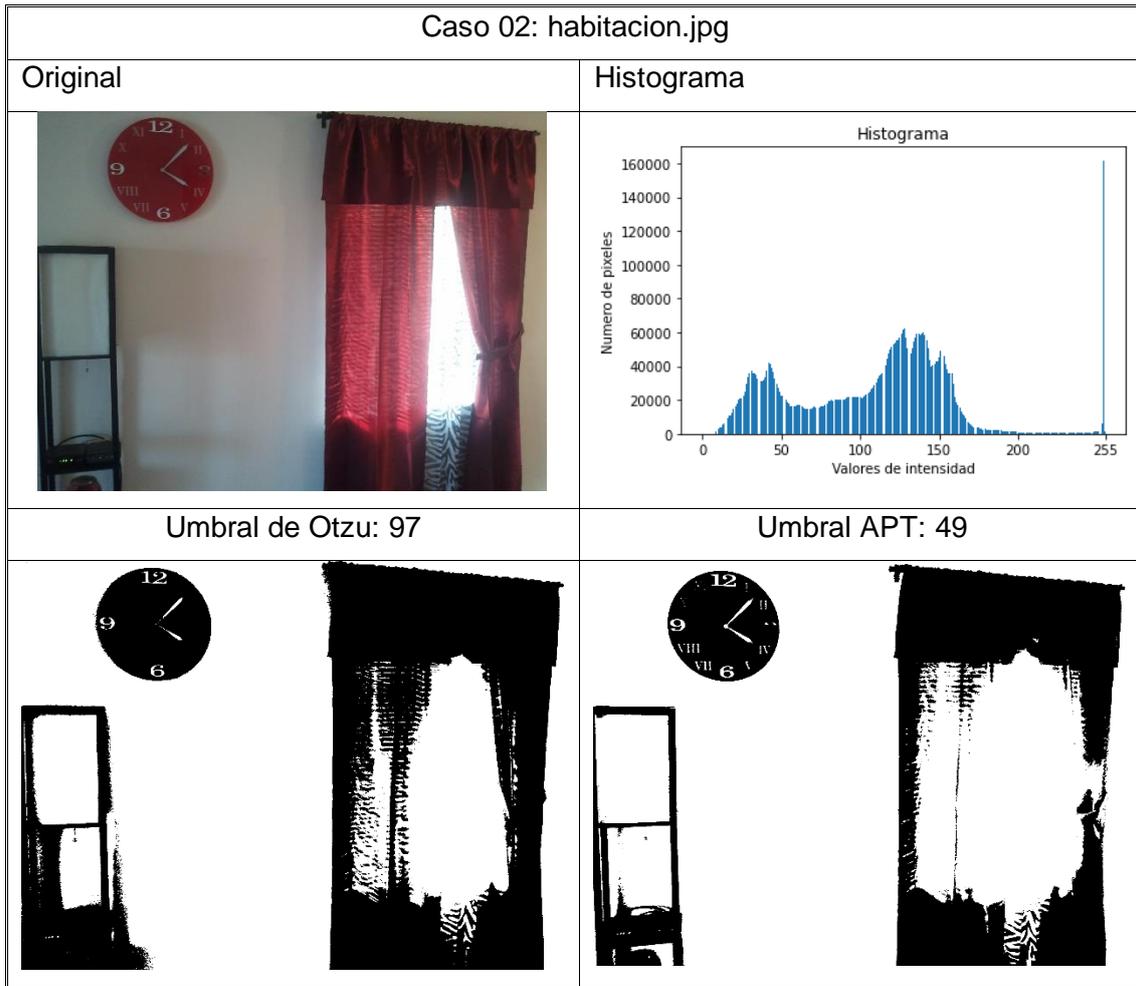


Figura: 12 Caso 02 - habitacion.jpg

Esta imagen tiene muchos objetos distinguibles, el reloj, cortinas y un mueble. Para la umbralización de Otsu, se nota que el umbral es muy grande, porque hace notorios los degradados creados a través de la cortina cuando es atravesada por la luz. Algunos números son visibles en el reloj, mientras que otros han sido sombreados. Se ve también que algunos pliegues de la cortina son visibles los que está cercanos al umbral de Otsu. Al observar la imagen con umbralización APT, el umbral es mas bajo, esto hace que los objetos sean más luminosos, y es posible ver los números del reloj. Se ve también que los pliegues de la cortina han sido convertidos en blanco.

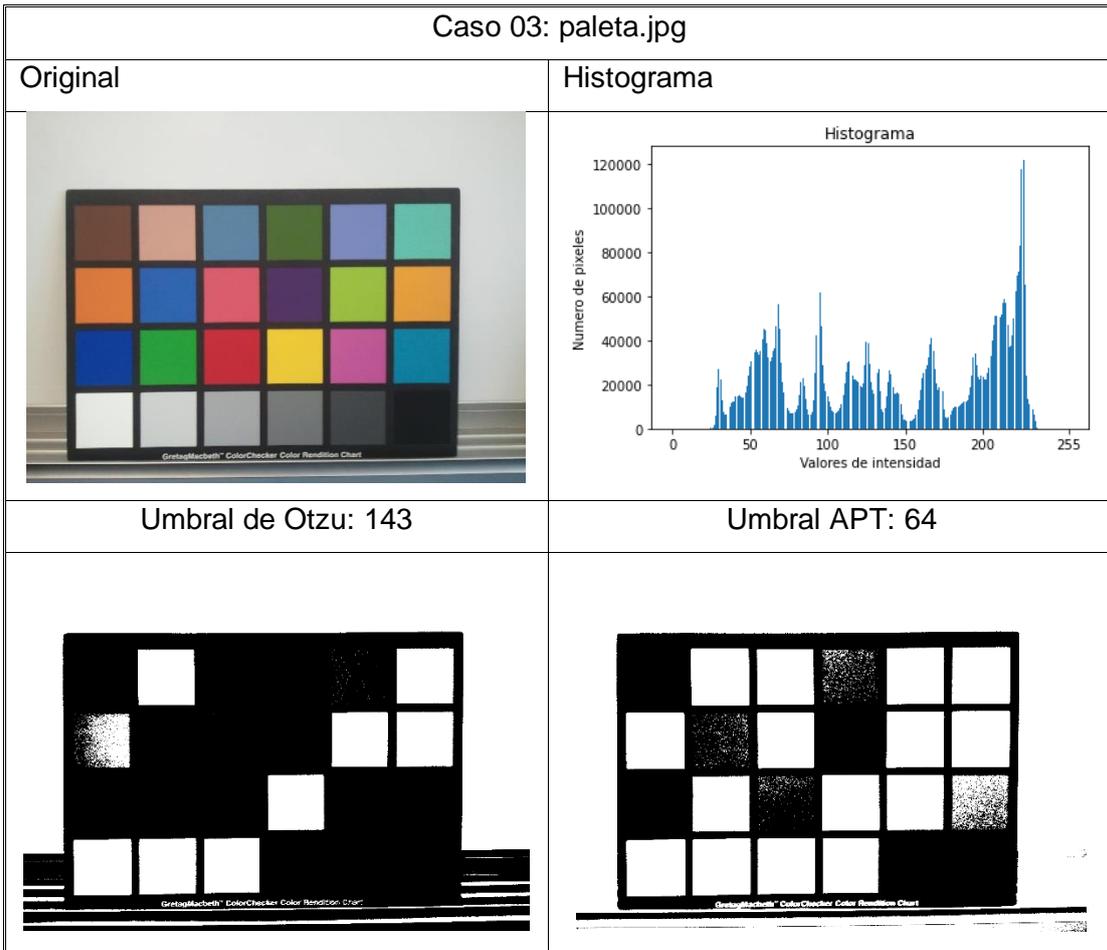


Figura: 13 Caso 03 - paleta.jpg

En esta imagen, tenemos una variedad de colores diferentes. Es un poco más difícil analizar esta imagen, porque el principal factor que afecta en esta imagen es el método usado para generar la imagen en escala de grises. Sin embargo, una cosa que es fácil de notar en la imagen coloreada en el cuadrado naranja en el lado izquierdo, tiene una parte blanca y otra negra con el método de Otsu. Esto es por las condiciones de iluminación que tiene impacto en los valores de pixeles, lo que significa que los valores en gris de los pixeles en el área naranja son muy cercanos al umbral de 142. Para la imagen APT, muchos de los colores han sido convertidos a blanco, se nota que en los dos cuadrados que se han mantenido negros corresponden al azul oscuro y el negro.

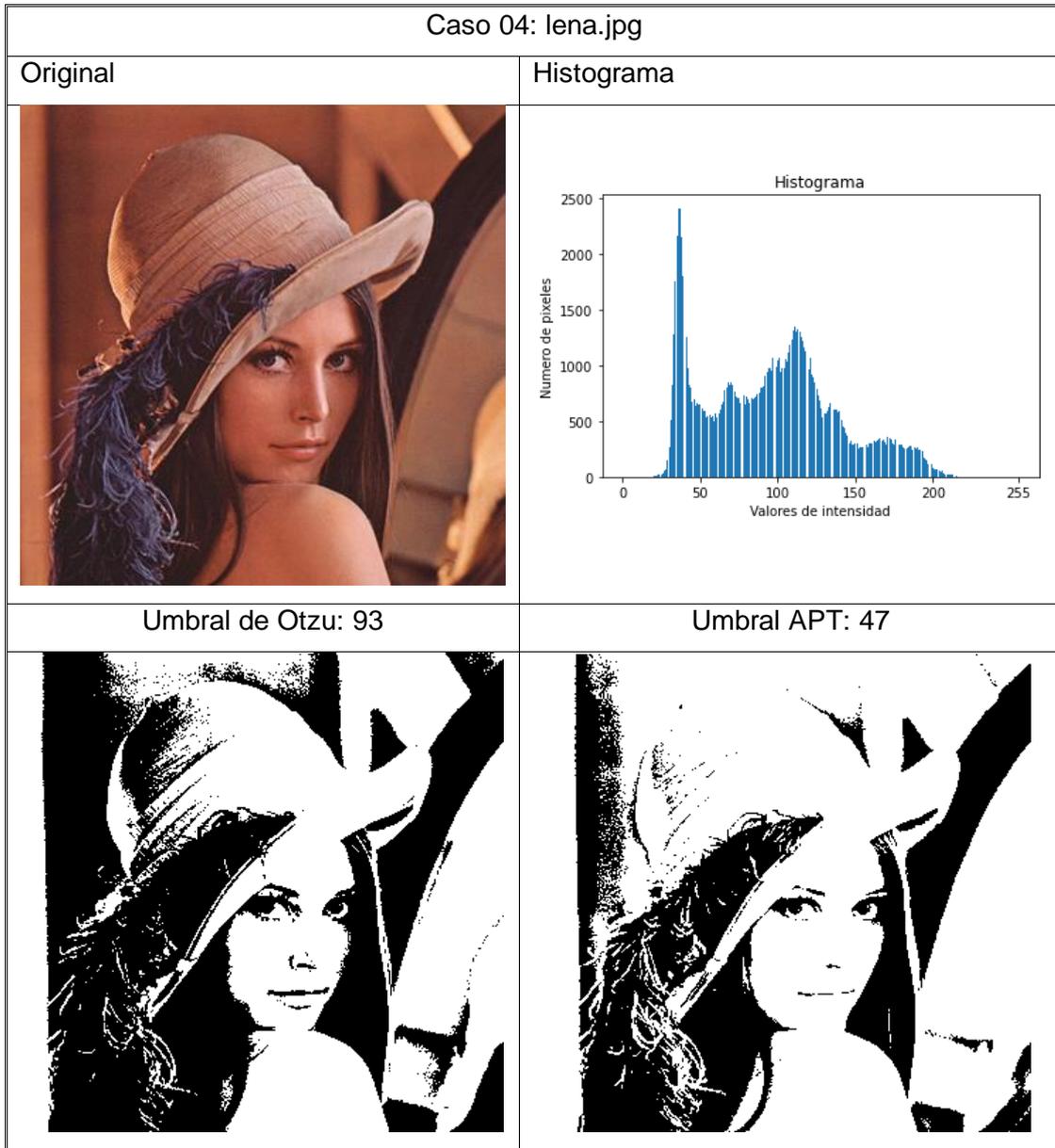


Figura: 14 Caso 04 - lena.jpg

Para la imagen específica, el algoritmo de umbralización adaptativo no es especialmente útil, específicamente el método APT. Hay una variedad de colores diferentes en esta imagen, dejando una amplia gama de píxeles blancos y negros; estos algoritmos son buenos cuando la diferencias entre colores son notables al ojo humano. Los píxeles que están bajo el umbral son predominantemente partes luminosas de la imagen, el sombrero, la piel y el espejo.

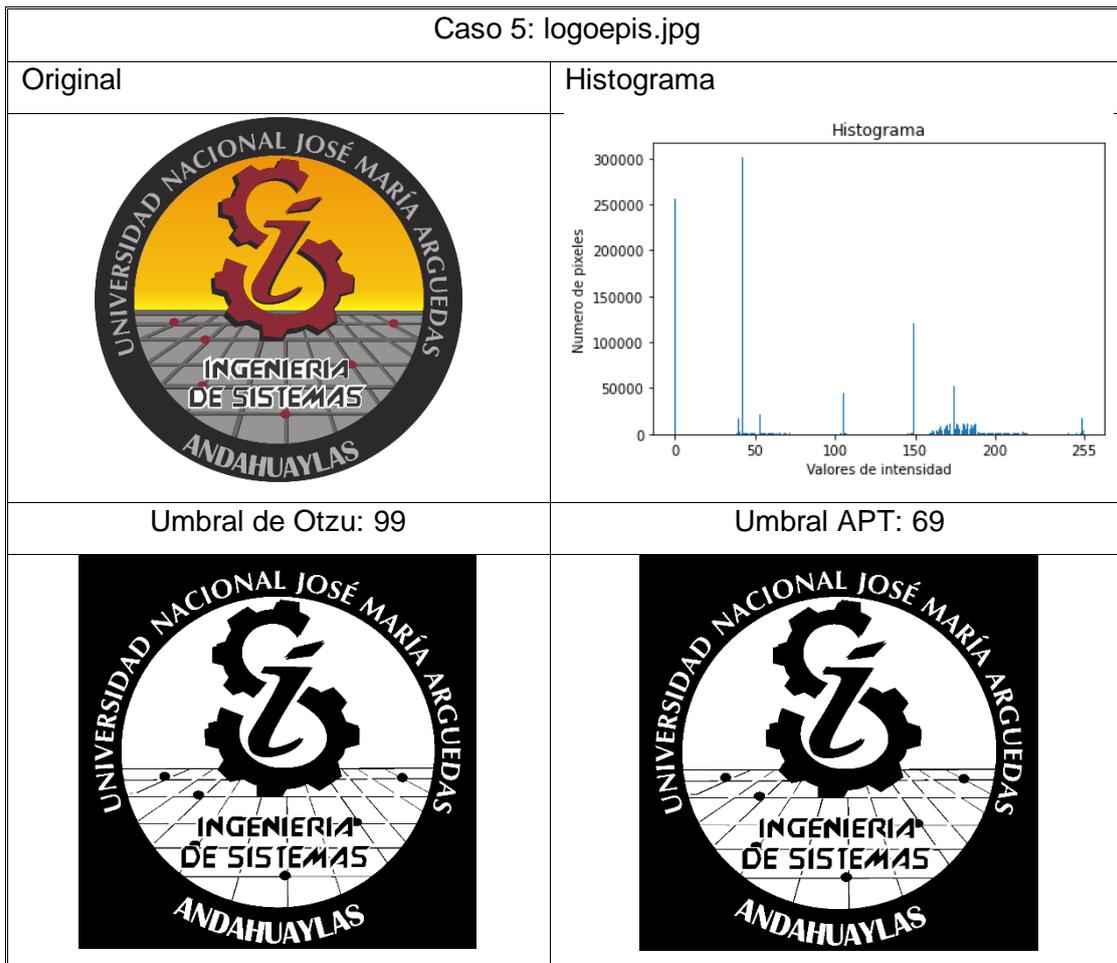


Figura: 15 Caso 05 - logoeepis.jpg

En esta imagen en particular, a pesar de los diferentes valores de umbral obtenidos para el método Otsu y el método APT. Los resultados son iguales, se debe a que la imagen es de colores sólidos sin degradados. Esta es una situación ideal que no se ve en imágenes tomadas con un scanner como en los documentos de partidas que se van a binarizar.

Todos los casos analizados corresponden se han realizado con la finalidad de validar el método APT propuesto se aprecia en general que el comportamiento de los resultados es similar al del método Otsu aunque los resultados en general son imágenes con menores niveles de negro que en el método Otsu, es decir que son más luminosas.

Para comprobar que esta claridad es algo deseable para documentos manuscritos y tal vez deteriorados obtenidos mediante un scanner, se realizaron las mismas pruebas en una base de datos abierta de documentos escaneados. Por tratarse de una base de datos extensa se presentan a continuación solo el resumen de los valores del umbral obtenidos para ambos métodos Otsu y APT.

IMAGEN	ORIGINAL Tamaño en Bytes	OTSU Tamaño en Bytes	APT Tamaño en Bytes	Tasa Compresión	OTSU Umbral	APT Umbral
1	582'618	284'059	284'049	0.49	166	149
2	506'339	257'657	257'642	0.51	163	146
3	1'757'539	195'091	194'852	0.11	168	151
4	993'791	180'421	180'439	0.18	167	150
5	5'723'709	290'755	339'706	0.06	185	166
6	844'836	164'927	164'239	0.19	170	153
7	6'379'980	492'270	331'533	0.05	195	175
8	10'789'114	474'464	636'899	0.06	187	168
9	4'038'065	237'061	318'902	0.08	206	185
10	6'554'915	528'491	407'840	0.06	197	177
11	3'925'276	229'871	299'156	0.08	205	184
12	6'392'328	450'053	306'632	0.05	188	169
13	3'883'356	237'717	300'565	0.08	207	186
14	774'799	109'107	109'773	0.14	159	143
15	2'876'725	99'637	100'329	0.03	163	146
16	1'150'265	48'207	48'173	0.04	165	148
17	292'969	383'409	397'889	1.36	203	182
18	689'491	96'329	97'173	0.14	159	143
19	2'200'713	451'675	451'061	0.20	170	153
20	3'811'484	472'708	425'880	0.11	181	162
21	1'503'513	385'728	387'524	0.26	170	153
22	6'010'861	632'139	572'004	0.10	201	180
23	3'797'472	395'403	452'651	0.12	204	183
24	2'432'693	417'758	420'275	0.17	198	178
25	2'649'857	425'447	426'074	0.16	198	178
26	3'868'164	432'949	394'506	0.10	177	159
27	1'254'429	99'036	98'855	0.08	156	140
28	3'207'930	142'766	142'471	0.04	159	143
29	1'802'332	415'969	427'578	0.24	202	181

Tabla 5: Cálculo de tonos de color

De los valores de umbrales calculados se aprecia que el valor de umbral APT siempre es menor que el valor del umbral Otsu. También se puede apreciar que el peso en bytes de las imágenes procesadas con el método APT y guardadas en formato PNG, es ligeramente menor que el peso de aquellas imágenes tratadas con el método Otsu. Esto

puede ser consecuencia de que el formato PNG es un formato sin pérdida que utiliza un algoritmo de compresión por zonas, al haber zonas más blancas en las imágenes procesadas con el método APT tendrán zonas blancas más amplias.



Figura: 16 Imagen escaneada de una página de diario

La figura muestra una imagen escaneada original, se aprecia que el texto no es completamente negro sino gris antes de calcular el umbral y binarizar. El problema de las zonas grises es que se perderán, el proceso de binarización usando el método APT debe resaltar las letras porque estas tienen pocos degradados.

Science in the Prince Charles Mountains — the 1991-92 summer expedition

Louise Crossley, Field Leader, 1991-92 Prince Charles Mountains Expedition

Friday 13 December 1991 was a fateful day for the 1991-92 Prince Charles Mountains expedition. In the early morning, *Aurora Australis* reached Mawson ice edge. The three helicopters assigned to the fly-in and then to the expedition were soon at the station, carrying mail, fresh vegetables and the usual sense of excitement that a ship and its consignment of new faces brings at the beginning of the season. But by 2000 hours that evening the expedition was in jeopardy as one of the helicopters lay smashed beyond repair beside the Casey Range, 40 km from Mawson. No one was injured in the accident, but having only two helicopters put some doubt on the logistics and safety of the expedition.

Good luck and hard work, however, transformed disaster into a highly successful PCM season. By chance, the Chinese ship *Ji Di* was due in Hobart bound for Davis, and CHINARE authorities quickly agreed to transport a replacement helicopter. The engineers dismantled VTI-HRF where it lay, and the sad remains were flown back to the ship for transport to Australia. With overtime worked on ship to shore operations and a miraculous burst of brilliant weather, the helicopter pilots managed to ferry the entire expedition of 23 members out to Dovers in four



Moving fuel for the PCM field program. Photo A Drinnan

sorties over two days—a record with only two-thirds normal capacity!

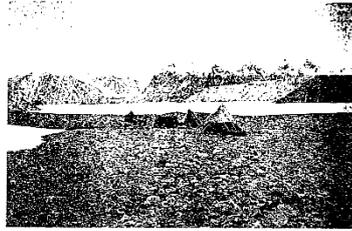
Three days after Christmas, all field parties had been deployed to their initial campsites and remaining cargo collected from Mawson, and a small party flew to Davis to collect the third aircraft and pilot. By happy coincidence we arrived at Davis, after a flight via Corry Rocks, Drygalski IV and the Larsmann Hills, just in time to join in the station's New Year's Eve festivities.

On our return to Dovers, the program forged ahead in continuing brilliant weather. Using the three helicopters, field parties moved freely from campsites to campsite while the Dovers base party began the arduous task of cleaning up the old ANARE base at Moore Pyramid. Three hard days' work and 22 sling loads later, all that remained to mark the site of six years occupation in the early 1970s was a section of the radio mast and three deeply buried containers. The 20-year-old food, especially the flour and orange juice, proved a welcome addition to the Dovers larder, while all remaining provisions, 200 l fuel drums and three cage pallets of rubbish, were loaded on traverse sleds for Mawson and Australia.

Dovers itself was paradise compared to previous seasons. The exceptionally warm spring had melted the snow on the saddle between Mt Jacklyn and Farley Massif upwind of the camp, so that drift, which is normally a persistent problem, was virtually non-existent. In addition, the spring traverse crew had put all the huts up on sleds, so that far from having constantly to dig access to the doors, we had to make steps to reach them.

No paradise lasts forever. By mid-January the first of the Prydz Bay lows brought cloudy, unpredictable weather.

frequent blizzards and whiteout—weather more typical of the region. Dovers was clearly in the wrong place. Day after day, while Dovers was shrouded in cloud or lashed by katabatics, field parties in the south and east of the northern PCMs, around Beaver Lake and Fisher Massif, reported perfect flying weather. The windows for flying became shorter and less predictable; once the helicopters had to wait at Radok Lake for several hours



Campsite at *Glossopteris Cully*, with Radok Lake and Dragons Teeth in background. Photo A Drinnan

before the whiteout cleared enough for a return to Dovers.

Nonetheless the scientific program proceeded well. All the field parties completed at least 85 percent of their primary program. Several found the opportunity to achieve some secondary objectives as well.

The traverse party left for Mawson on 30 January, carrying nearly 1.5 tonnes of rocks amongst its other cargo, greatly reducing the need for helicopter rock runs to Mawson, normally a routine feature of the logistics operation. The scientists were not far behind—the first group returned to Mawson on 2 February, and a second on 5 February.

Another low interrupted flying operations for a few days, and gave the crew remaining at Dovers plenty of time to clean up, perform a stocktake and close down the base. The final two sorties of people and cargo were flown out on 8 February. There will be no PCM program in the 1992-93 season, but hopefully all will be intact when this tiny principality is once again visited by an ANARE expedition. □

Figura: 17 Imagen procesada con el método Otsu

Science in the Prince Charles Mountains — the 1991-92 summer expedition

Louise Crossley, Field Leader, 1991-92 Prince Charles Mountains Expedition

Friday 13 December 1991 was a fateful day for the 1991-92 Prince Charles Mountains expedition. In the early morning, *Aurora Australis* reached Mawson ice edge. The three helicopters assigned to the fly in and then to the expedition were soon at the station, carrying mail, fresh vegetables and the usual sense of excitement that a ship and its consignment of new faces brings at the beginning of the season. But by 2000 hours that evening the expedition was in jeopardy as one of the helicopters lay smashed beyond repair beside the Casey Range, 40 km from Mawson. No one was injured in the accident, but having only two helicopters put some doubt on the logistics and safety of the expedition.

Good luck and hard work, however, transformed disaster into a highly successful PCM season. By chance, the Chinese ship *Ji Di* was due in Hobart bound for Davis, and CHINARE authorities quickly agreed to transport a replacement helicopter. The engineers dismantled VH-HRF where it lay, and the sad remains were flown back to the ship for transport to Australia. With overtime worked on ship to shore operations and a miraculous burst of brilliant weather, the helicopter pilots managed to ferry the entire expedition of 23 members out to Dovers in four



Moving fuel for the PCM field program. Photo A Drinnan

sorties over two days—a record with only two-thirds normal capacity!

Three days after Christmas, all field parties had been deployed to their initial campsites and remaining cargo collected from Mawson, and a small party flew to Davis to collect the third aircraft and pilot. By happy coincidence we arrived at Davis, after a flight via Cory Rocks, Dryshinaya IV and the

Larsenmann Hills, just in time to join in the station's New Year's Eve festivities.

On our return to Dovers, the program forged ahead in continuing brilliant weather. Using the three helicopters, field parties moved freely from campsite to campsite while the Dovers base party began the arduous task of

cleaning up the old ANARE base at Moore Pyramid. Three hard days' work and 22 sling loads later, all that remained to mark the site of six years' occupation in the early 1970s was a section of the radio mast and three deeply buried containers. The 20-year-old food, especially the flour and orange juice, proved a welcome addition to the Dovers larder, while all remaining provisions, 200 L fuel drums and three cage pallets of rubbish, were loaded on traverse sleds for Mawson and Australia.

Dovers itself was paradise compared to previous seasons. The exceptionally warm spring had melted the snow on the saddle between Mt Jacklyn and Farley Massif upwind of the camp, so that drift, which is normally a persistent problem, was virtually non-existent. In addition, the spring traverse crew had put all the huts up on sleds, so that far from having constantly to dig access to the doors, we had to make steps to reach them.

No paradise lasts forever. By mid-January the first of the Prydz Bay lows brought cloudy, unpredictable weather,

frequent blizzards and whiteout—weather more typical of the region. Dovers was clearly in the wrong place. Day after day, while Dovers was shrouded in cloud or lashed by katabatics, field parties in the south and east of the northern PCMs, around Beaver Lake and Fisher Massif, reported perfect flying weather. The windows for flying became shorter and less predictable, once the helicopters had to wait at Radok Lake for several hours



Campsite at *Glossopteris Cully*, with Radok Lake and Dragons Teeth in background. Photo A Drinnan

before the whiteout cleared enough for a return to Dovers.

Nonetheless the scientific program proceeded well. All the field parties completed at least 85 percent of their primary program. Several found the opportunity to achieve some secondary objectives as well.

The traverse party left for Mawson on 30 January, carrying nearly 1.5 tonnes of rocks amongst its other cargo, greatly reducing the need for helicopter rock runs to Mawson, normally a routine feature of the logistics operation. The scientists were not far behind—the first group returned to Mawson on 2 February, and a second on 5 February.

Another low interrupted flying operations for a few days, and gave the crew remaining at Dovers plenty of time to clean up, perform a stocktake and close down the base. The final two sorties of people and cargo were flown out on 8 February. There will be no PCM program in the 1992-93 season, but hopefully all will be intact when this tiny principality is once again visited by an ANARE expedition. □

ANARE News Spring/Summer 1992-93 —

Figura: 18 Imagen procesada con el método a APT

El resultado no es satisfactorio del todo debido a que en la imagen original las partes oscuras no tenían buen contraste, es algo que se puede resolver haciendo ajustes previos a las imágenes de contraste y brillo. El método APT no abarca estos procedimientos.

5.2. Resultados con imágenes de actas de diferentes tipos

Las imágenes digitalizadas de actas, tienen características que las hacen particulares, tales como el trazo y color de la tinta por ser manuscritas. Contienen sellos y timbres muy notorios que en algunos casos están sobre el texto. Pueden contener también filigranas o fondos tenues además de logotipos.

El procesamiento de las actas utilizando el método APT, en base a las pruebas realizadas previamente se ajusta a lo esperado, existen algunos documentos que por sus características particulares escapan a la regla, pero en general el resultado es satisfactorio.

Se hizo el procesamiento de 10 actas de diferentes tipos y orígenes, digitalizadas en formatos diferentes JPG y PNG. Se hizo la comparación visual de las partes representativas de estas: los textos de los datos y los sellos y timbres principalmente, que constituyen las partes que guardan información relevante.

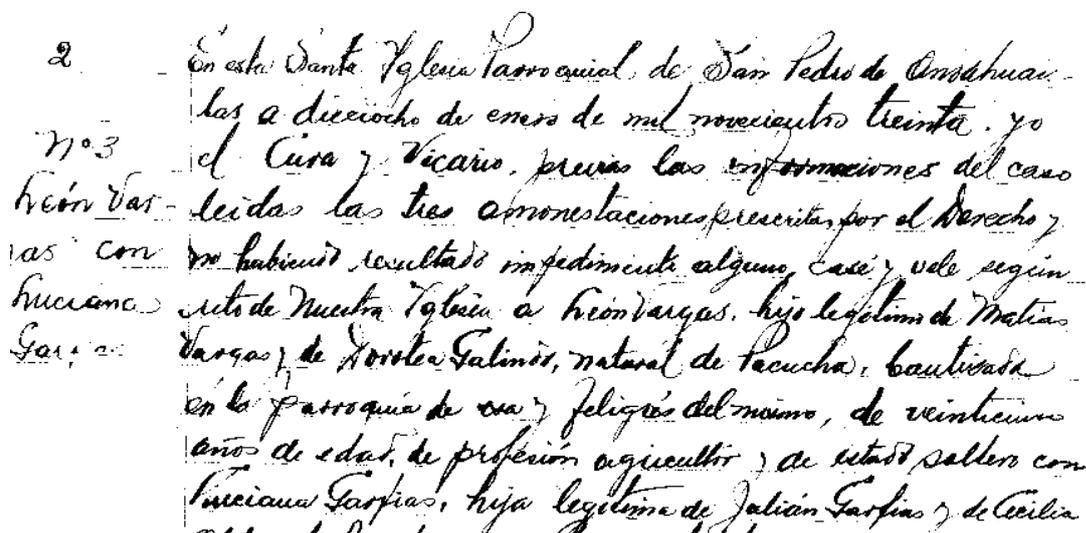
Se muestran los resultados de los umbrales calculados para los umbrales. Debido a que las imágenes originales son de formatos diferentes de compresión, no se comparan los tamaños o pesos.

Documento	OTSU	APT
Acta01	179	161
Acta02	180	162
Acta03	191	171
Acta04	195	175
Acta05	187	168
Acta06	176	158
Acta07	201	180
Acta08	189	170
Acta09	183	164
Acta10	175	157

Tabla 6: Comparación de umbrales calculados

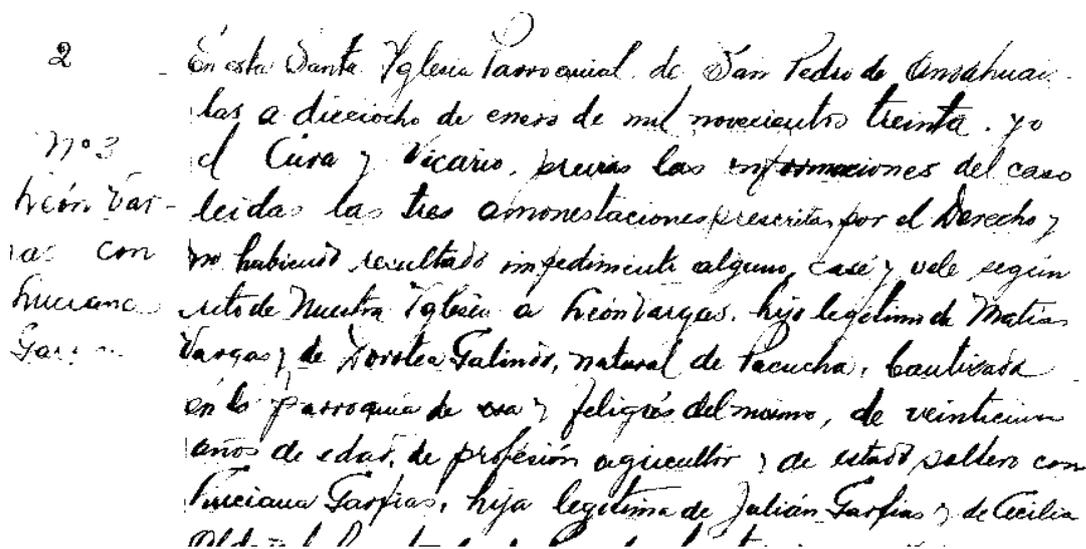
Se aprecia que al igual que con las imágenes de la base de datos de prueba el umbral calculado con el método APT es siempre menor, con lo que las imágenes resultantes

tienen más píxeles blancos. Esta característica puede ser deseable o no dependiendo de la imagen procesada y las partes del documento que se quieren resaltar.



2. En esta Santa Iglesia Parroquial de San Pedro de Amahuacu-
las a dieciocho de enero de mil novecientos treinta. yo
Nº 3 el Cura y Vicario, previas las informaciones del caso
Heón Vargas - leídas las tres amonestaciones prescritas por el Derecho y
ras con no habiénd resultado impedimento alguno, case y vele según
buena costumbre a Leon Vargas, hijo legítimo de Matías
García de Norberto Galindo, natural de Tacucha, bautizado
en la parroquia de esta feligresía del mismo, de veintinueve
años de edad, de profesión agricultor y de estado soltero con
Inocencia García, hija legítima de Julián García y de Cecilia

Figura: 19 Región con zona oscura del acta01 con el método Otsu



2. En esta Santa Iglesia Parroquial de San Pedro de Amahuacu-
las a dieciocho de enero de mil novecientos treinta. yo
Nº 3 el Cura y Vicario, previas las informaciones del caso
Heón Vargas - leídas las tres amonestaciones prescritas por el Derecho y
ras con no habiénd resultado impedimento alguno, case y vele según
buena costumbre a Leon Vargas, hijo legítimo de Matías
García de Norberto Galindo, natural de Tacucha, bautizado
en la parroquia de esta feligresía del mismo, de veintinueve
años de edad, de profesión agricultor y de estado soltero con
Inocencia García, hija legítima de Julián García y de Cecilia

Figura: 20 Región con zona oscura del acta01 con el método APT

La imagen procesada con el método Otsu, tiende a oscurecer las zonas con logotipos, sombras de otros folios o detalles que en original estaban degradados. La imagen procesada con el método APT rescata mejor el contenido en desmedro de las sombras de profundidad. En el texto no hay diferencia apreciable.

51 En esta Santa Iglesia parroquial de San Pedro de Andahuailas, a los veintinueve
Cinco dias del mes de octubre de mil novecientos treinta y seis, yo el infrascripto párroco
Datis premito las informaciones canónicas e leídas las amonestaciones forenses
con las por el Derecho i no habiendo resultado ningun impedimento cano
Ceferna por palabras de presente i cede según el rito de nuestra Santa Madre Iglesia
Viguera a Cirilo Datis hijo legítimo de José Datis i de Concepción Altamirano soltero
de veintisiete años de edad natural de San Jerónimo i residente en Estora
bautizado i feligres de San Jerónimo de profesión agricultor con Ceferna
Viguera hija legítima de Manuel Viguera i de Juana María Velasco soltera
de treinta años de edad natural de San Jerónimo i residente en Estora
bautizado i feligres de San Jerónimo de profesión hilandera casada

Figura: 21 Región del texto del acta 10 con el método Otsu

51 En esta Santa Iglesia parroquial de San Pedro de Andahuailas, a los veintinueve
Cinco dias del mes de octubre de mil novecientos treinta y seis, yo el infrascripto párroco
Datis premito las informaciones canónicas e leídas las amonestaciones forenses
con las por el Derecho i no habiendo resultado ningun impedimento cano
Ceferna por palabras de presente i cede según el rito de nuestra Santa Madre Iglesia
Viguera a Cirilo Datis hijo legítimo de José Datis i de Concepción Altamirano soltero
de veintisiete años de edad natural de San Jerónimo i residente en Estora
bautizado i feligres de San Jerónimo de profesión agricultor con Ceferna
Viguera hija legítima de Manuel Viguera i de Juana María Velasco soltera
de treinta años de edad natural de San Jerónimo i residente en Estora
bautizado i feligres de San Jerónimo de profesión hilandera casada

Figura: 22 Región del texto del acta 10 con el método APT

La imagen tratada con el método Otsu, ha resaltado algunos defectos del papel como los del borde izquierdo y las líneas del rayado de la página. La imagen tratada con el método APT en cambio se ve más limpia. El texto sin embargo es casi igual en ambos métodos.

La inspección visual nos permite verificar las imágenes procesadas, pero para tener una apreciación más general se analizan los histogramas de las imágenes, para comparar sus características y similitudes en función de los tonos de grises que contienen las imágenes originales.

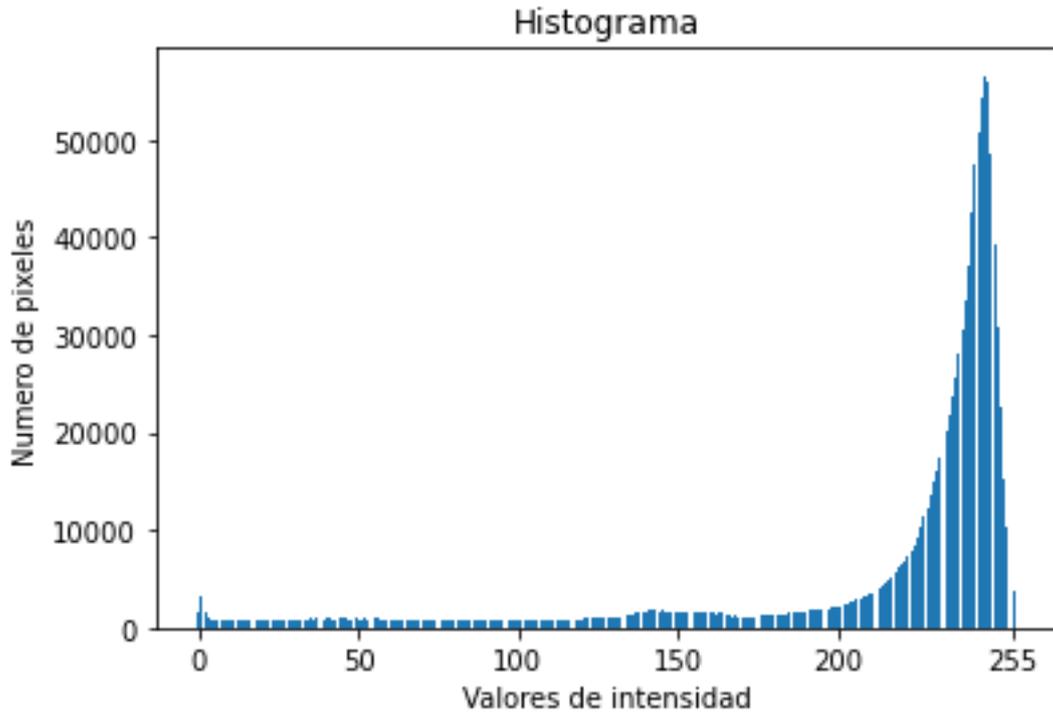


Figura: 23 Histograma de la imagen acta01

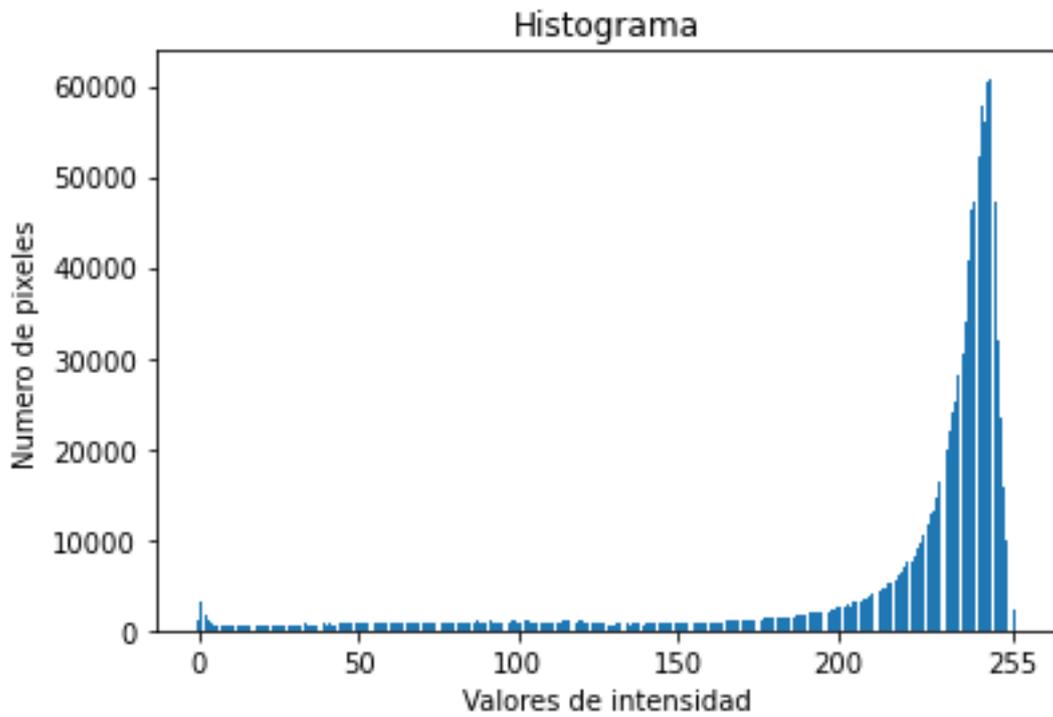


Figura: 24 Histograma de la imagen acta02

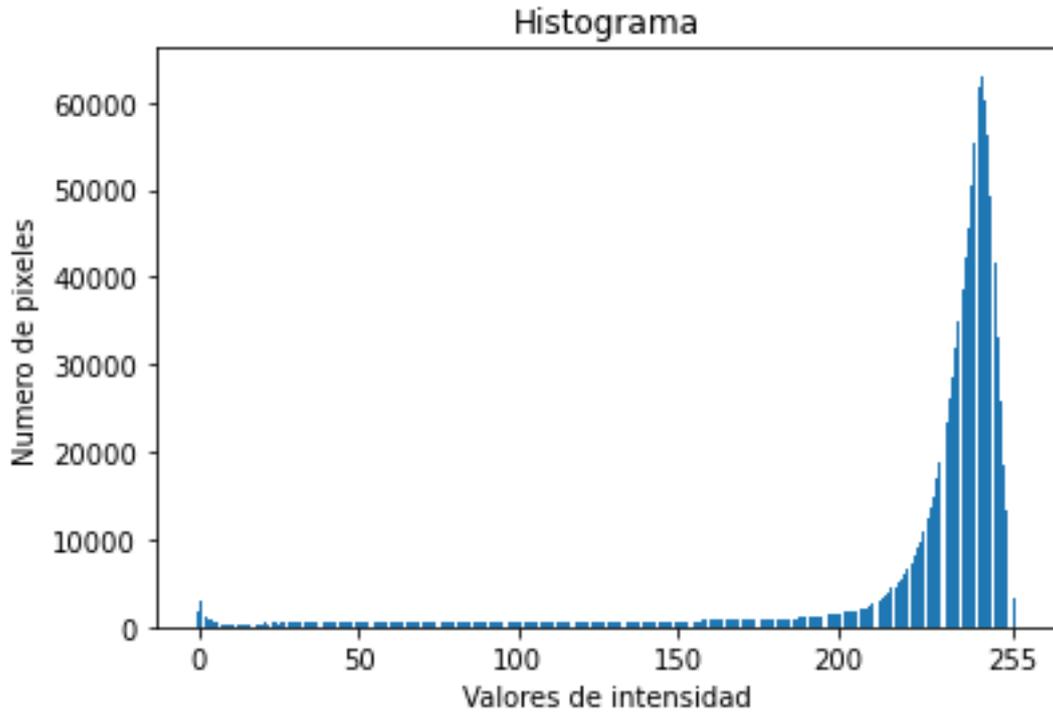


Figura: 25 Histograma de la imagen acta03

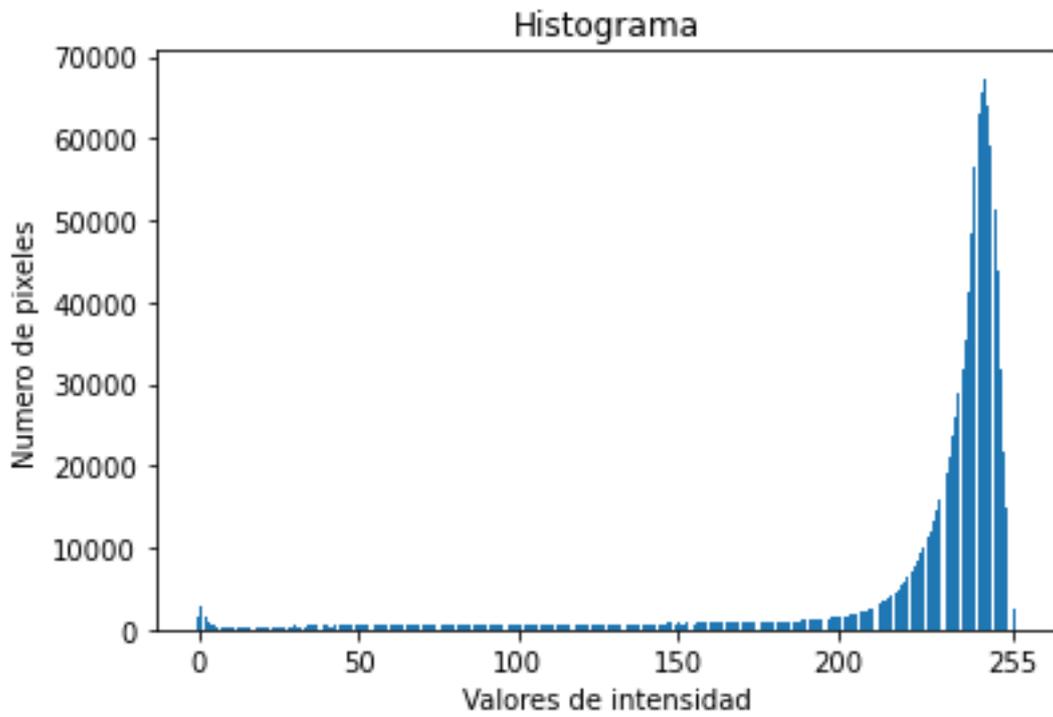


Figura: 26 Histograma de la imagen acta04

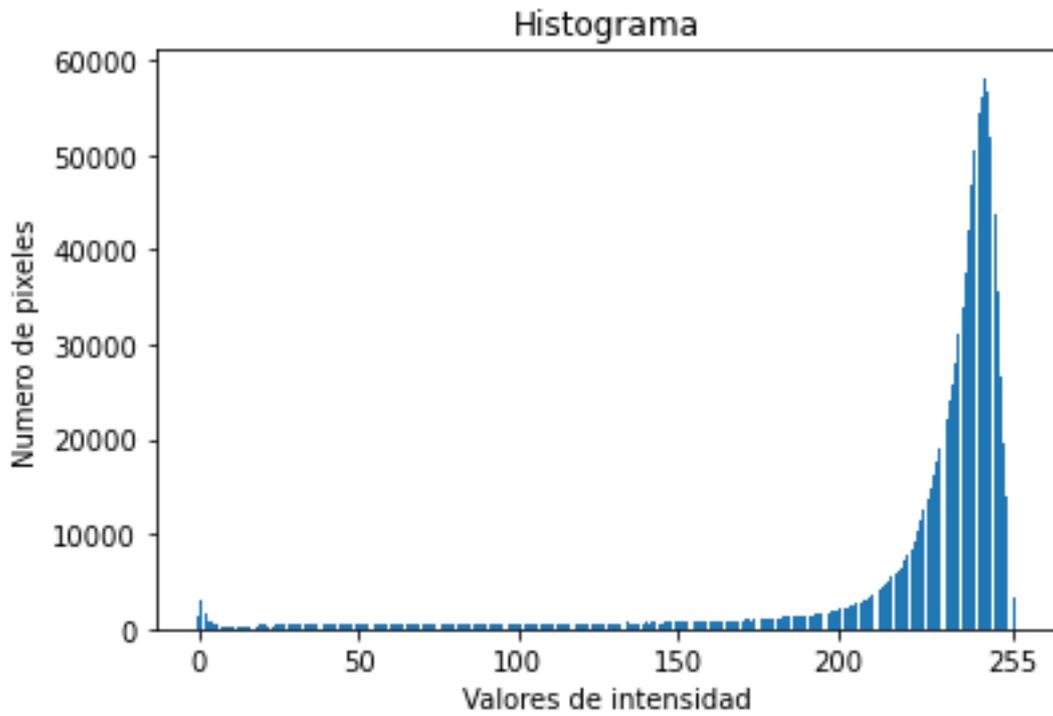


Figura: 27 Histograma de la imagen acta05

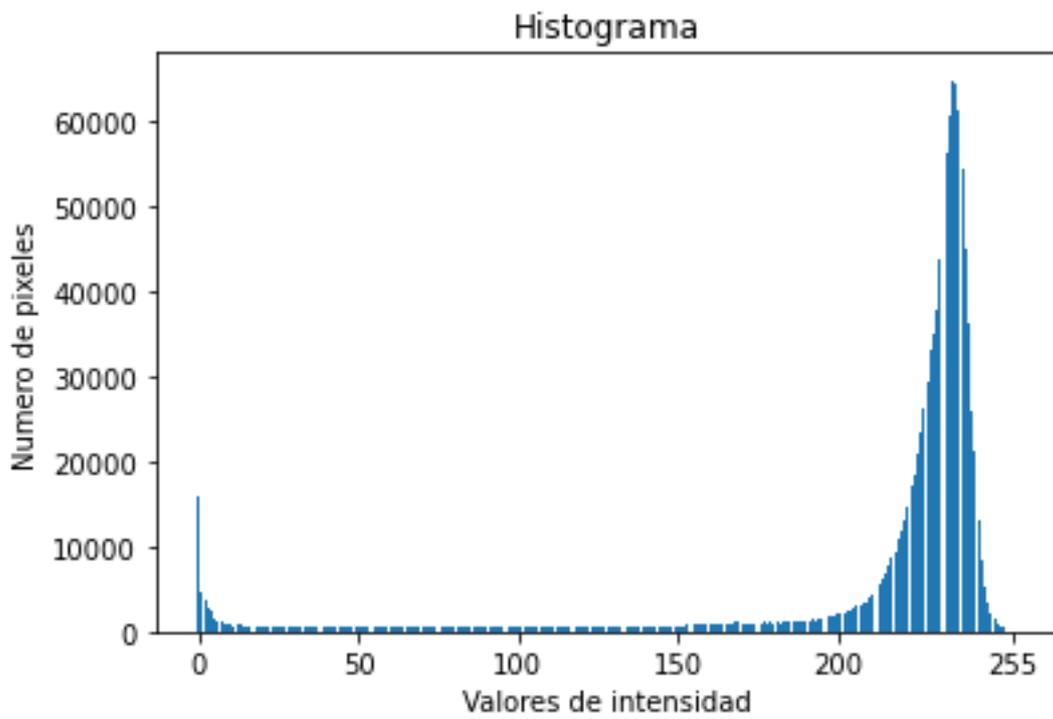


Figura: 28 Histograma de la imagen acta06

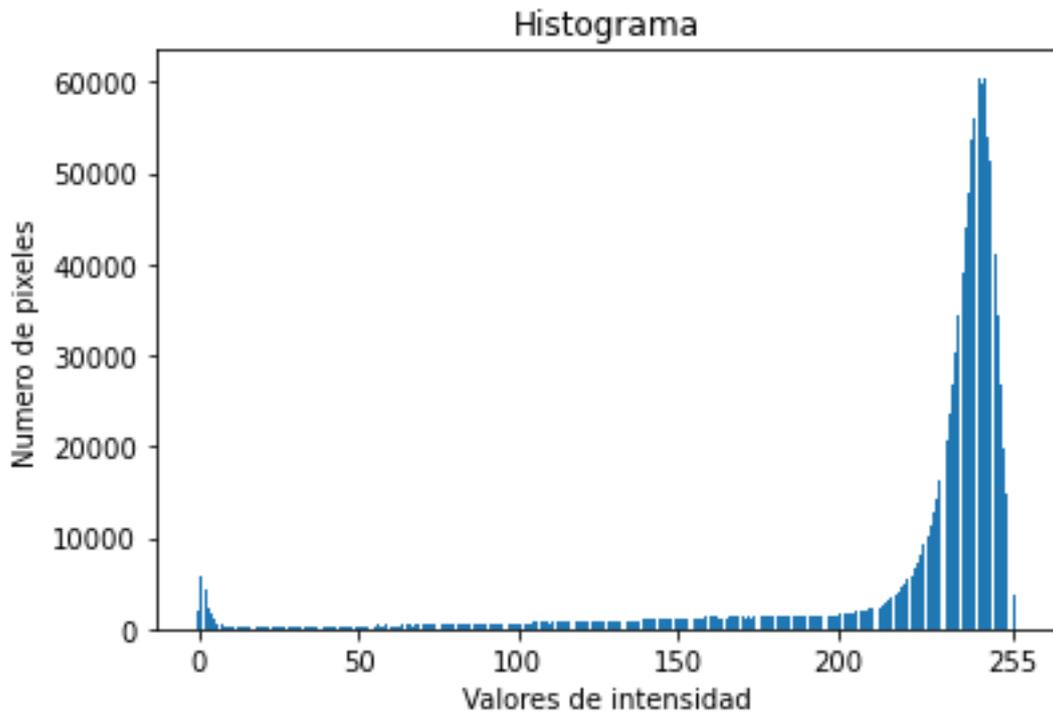


Figura: 29 Histograma de la imagen acta07

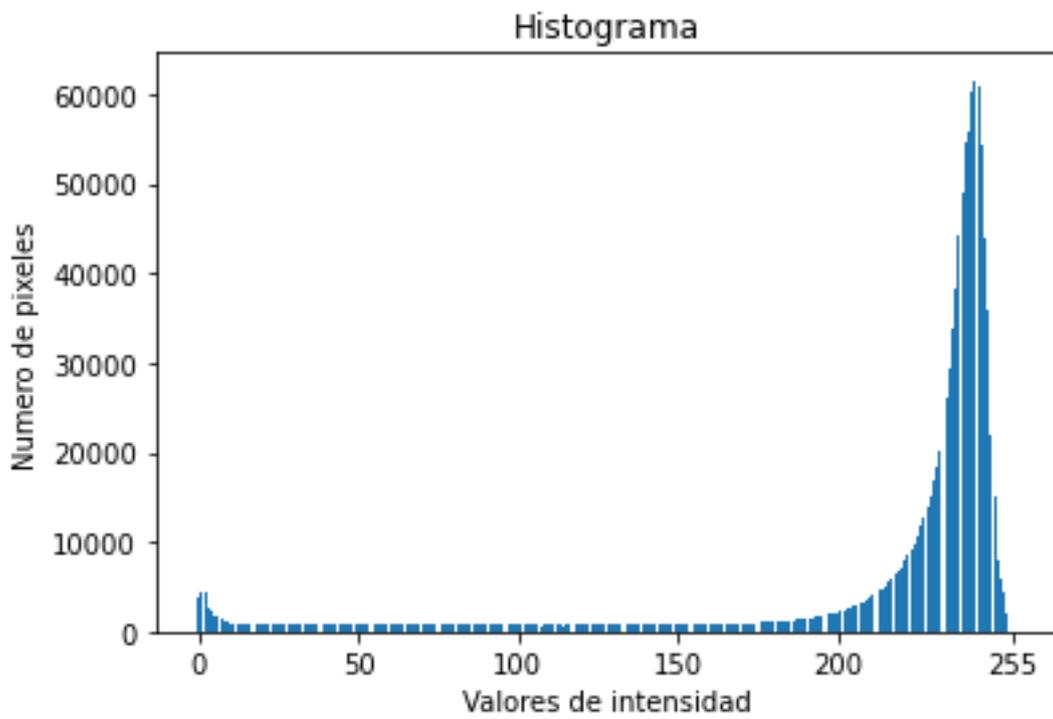


Figura: 30 Histograma de la imagen acta08

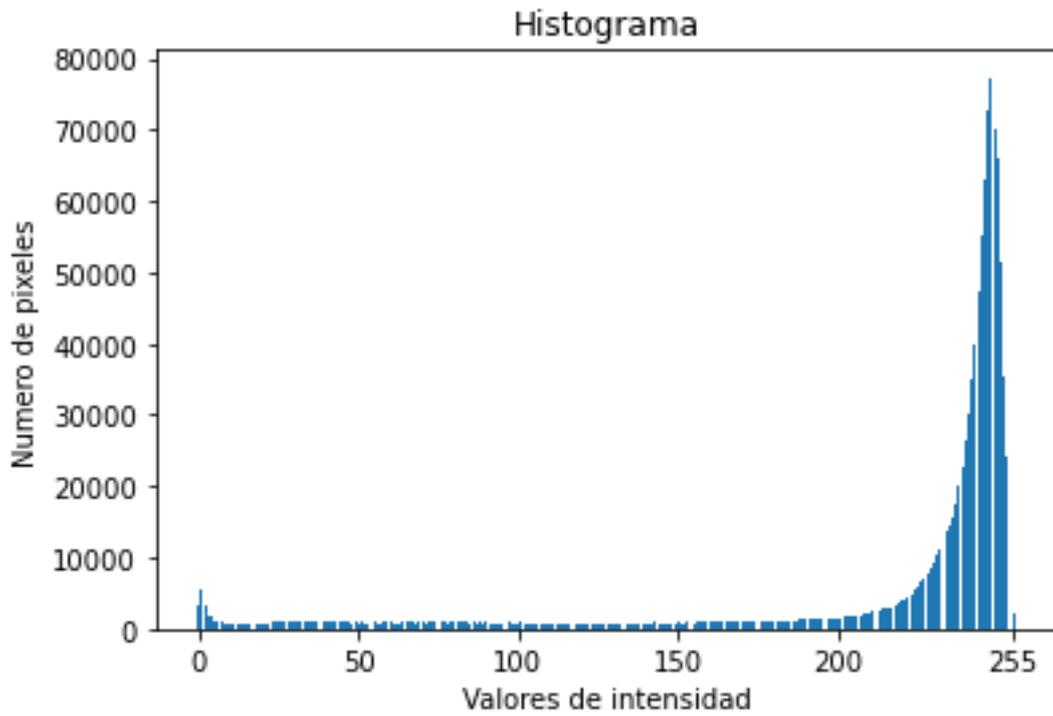


Figura: 31 Histograma de la imagen acta09

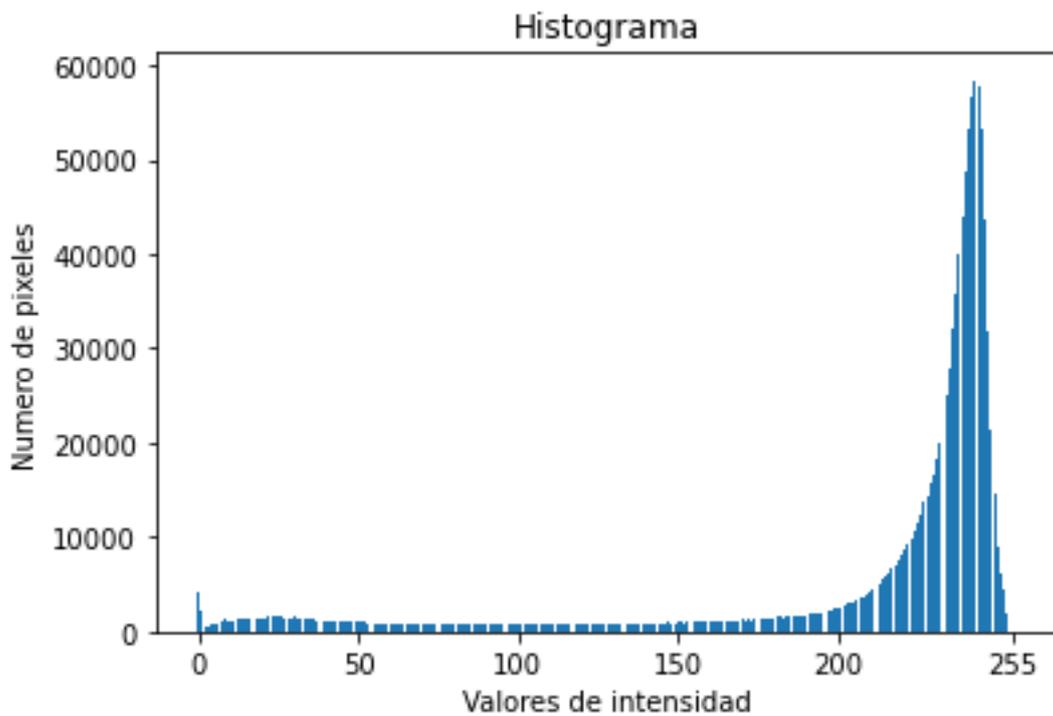


Figura: 32 Histograma de la imagen acta10

Se aprecia que la mayor concentración de valores se encuentra en los valores de 200 y 255, lo que significa que las imágenes originales tienen fuertes contrastes cercanos al

color negro, lo cual permite que el método APT pueda permitirse obtener valores de umbral menores que el método Otsu sin perder demasiada información.

5.3. Resultados de la aplicación

Para el procesamiento de las imágenes se ha desarrollado un prototipo utilizando el lenguaje de programación Python 3.x, programándose los algoritmos y las interfaces mínimas necesarias para procesar imágenes individualmente y por bloques. La aplicación también entrega resultados del histograma de la imagen original en escala de grises a partir de la cual se calcularon los valores de umbral para ambos métodos.

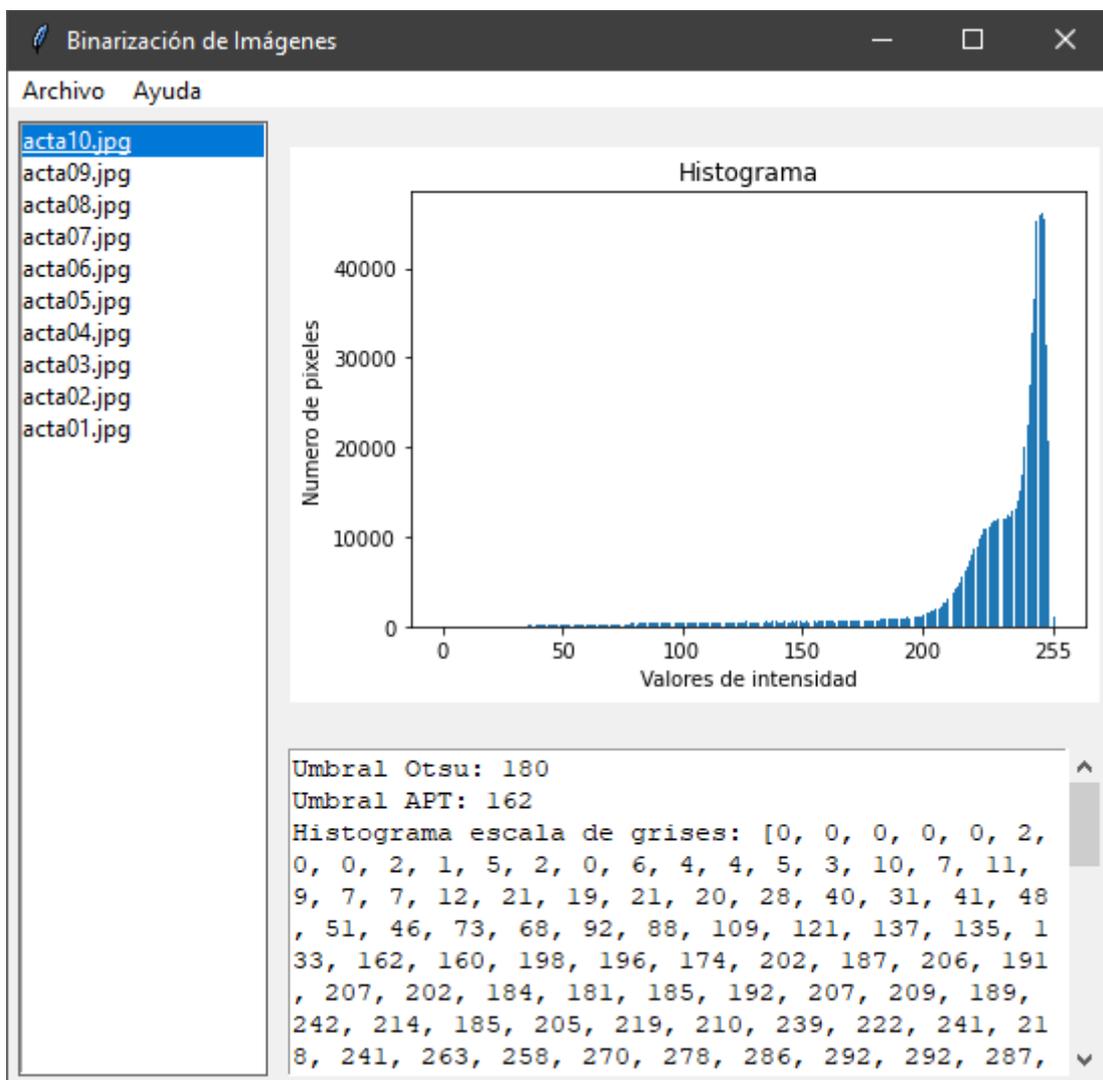


Figura: 33 Ventana principal del prototipo

5.4. Almacenamiento

Como consecuencia de la binarización el tamaño de los archivos a guardar se reduce significativamente. En promedio el archivo generado es 13% del tamaño original.

Archivo	Original en Bytes	Procesado en Bytes	Porcentaje de reducción
Acta01.jpg	544'454	78'752	14%
Acta01.jpg	527'850	72'317	14%
Acta01.jpg	517'726	65'229	13%
Acta01.jpg	516'340	65'987	13%
Acta01.jpg	546'059	68'377	13%
Acta01.jpg	522'294	62'556	12%
Acta01.jpg	549'345	71'386	13%
Acta01.jpg	618'377	72'389	12%
Acta01.jpg	545'959	74'861	14%
			Promedio 13%

Tabla 7: Espacio de almacenamiento

CONCLUSIONES

- Un método adecuado para empezar la binarización es convertir las imágenes a escala de grises, brinda los mejores resultados para cualquier imagen, sobre todo si tienen colores degradados. Otros métodos como la detección de bordes resaltan detalles del soporte de papel que para el caso de las partidas de matrimonio que se desea binarizar y rescatar la información de la escritura, por lo que no son adecuados.
- Se implementó el método propuesto de binarización adaptativa APT utilizando la varianza para corregir el umbral que nos permite obtener imágenes ligeramente más luminosas donde los detalles irrelevantes y tenues son desechados. Resulta adecuado para la binarización de documentos manuscritos como son las partidas de Matrimonio.
- El almacenamiento de las imágenes binarizadas en comparación con las imágenes originales en cualquier formato, es considerable. En promedio 87% menos del tamaño de la imagen original, aunque puede variar dependiendo del tipo de imagen de origen. Esto las hace aptas para ser almacenadas y transmitidas por cualquier medio incluido Internet.

RECOMENDACIONES

- El análisis de los métodos de binarización ha sido por inspección simple, sería mejor tener una métrica que permita evaluar de forma numérica la efectividad de cada método de binarización y tener resultados más concluyentes
- El método de binarización adaptativa ha resultado satisfactorio para el caso de las partidas de matrimonio. Se podría validar su utilidad con otros tipos de documentos o imágenes con degradados apoyándose en técnicas de reconocimiento de patrones.
- Las imágenes se han almacenado utilizando el formato sin pérdida PNG, pero por tratarse de imágenes binarias sería factible utilizar otro formato que permita traducir directamente a un bit cada pixel, facilitando su almacenamiento y transmisión.

REFERENCIAS BIBLIOGRÁFICAS

Alonso, J. A., García, M., & Lloveras, M. R. (2007). La norma ISO 15489: un marco sistemático de buenas prácticas de gestión documental en las organizaciones. *Item: Revista de Biblioteconomía I Documentació*, (47), 41–70. Retrieved from <http://eprints.rclis.org/14604/%5Cnhttp://eprints.rclis.org/handle/10760/12263>

Barzaga Laurencio, Y. (2018). Sistema de gestión documental para la empresa de Servicios Técnicos de Computación, Comunicaciones y Electrónica para la industria del Níquel (SERCONI) (Doctoral dissertation, Departamento de Ciencias de la Información).

Carlos, L., Cobos, C., & Javeriana, P. U. (2007). modelo para digitalización de documentos, aplicado a una obra: perteneciente a la colección libros antiguos y valiosos de la biblioteca de filosofía y teología mario valenzuela., S.J.

Carpallo Bautista, A. (2000). El papel de la “conservación documental” como disciplina al servicio de los profesionales de la Documentación. *Congreso Universitario de Ciencias de La Documentación*, 1, 429–433. Retrieved from <http://www.ucm.es/info/multidoc/multidoc/revista/num10/paginas/pdfs/acarpallo.pdf>

Hendriks, K. B., & UNESCO. (1984). Preservación y restauración de materiales fotográficos en archivos y bibliotecas Un estudio del RAMP con directrices, 87. Retrieved from <http://www.mcu.es/ccbae/es/consulta/registro.cmd?id=61383>

Hernández, Y. R., Fernández, E. M., Martín, N. A., Valdivia, G. G., González, M. M., & Salabarría, S. G. (2017). La conservación preventiva: una acción eficiente para combatir los agresores del patrimonio documental en los centros de información. *Gaceta Médica Espirituana*, 11(3).

IFLA e ICA. (2002). *PROYECTOS DE colecciones y fondos de dominio público*.

MONSALVE, A. (2006). No Title. A lo largo del tiempo la documentación presenta deterioros debido a múltiples factores (físicos, químicos, biológicos, físico–mecánicos y circunstanciales), que están destruyendo la información contenida en ellos y así mismo, desarrolla riesgos para la in.

Rodriguez Mora, H. A. (2007). *Lineamientos Para La Conservación Y Preservación De Los Archivos En Las Empresas Del Grupo Corona*.

Ruiz, G., & Piere, E. (2018). Implementación de un sistema de información bajo plataforma web para la gestión y control documental de la empresa corporación Jujedu EIRL–Talara; 2017.

Sastre Natividad, G., Perez, S. de J., & Loyola Vargas, V. M. (2015). Preservación y conservación de documentos digitales. *ArchivPost*, 15(2), 1–2.

Valentín Ruiz, F. J. (2017). Patrimonio bibliográfico-documental y preservación digital.

Vázquez Ávila, L. A. (2018). Propuesta de un Sistema de Gestión Documental, para la conservación del archivo histórico del Centro de Documentación Regional Juan Bautista Vázquez de la Universidad de Cuenca (Bachelor's thesis, Escuela Superior Politécnica de Chimborazo).

Vivas, A., Ma, D., & Pérez, G. (2011). La información histórica en los archivos eclesiásticos: principales series documentales para la investigación, 34, 441–467.

ANEXOS

1. Código Fuente

```
'''
binariza.py
Herbert Vargas Gutierrez
umbralizacion adaptativa '''

from PIL import ImageTk, Image
import tkinter as tk
from tkinter import scrolledtext as st
from tkinter import filedialog as fd
from tkinter import messagebox as mb
from tkinter import Label as lb
from tkinter import PhotoImage as im
from tkinter import Listbox as li
from tkinter import Toplevel as tl
from tkinter import Scale as sc
from tkinter import Button as bt
from tkinter import DoubleVar as dv
import os
from umbral import binariza
import textos
import time

class Splash(tk.Toplevel):
    def __init__(self, parent):
        tk.Toplevel.__init__(self, parent)
        self.title("Iniciando...")

        self.contenido = lb(self, text=textos.splash, width=50, height=20)
        self.contenido.pack()
        self.lift()
        self.attributes('-topmost', True)
        self.after_idle(self.attributes, '-topmost', False)
        self.update()

class App(tk.Tk):
    def __init__(self):
        tk.Tk.__init__(self)
        self.withdraw()
        splash = Splash(self)

        self.title("Main Window")
        time.sleep(3)

        splash.destroy()

        self.destroy()
```

```

class Aplicacion:

    def __init__(self):

        app = App()

        self.sensibilidad = 0.9
        self.directorio = "."
        self.ventana1 = tk.Tk()

        self.ventana1.title("Binarización de Imágenes")
        self.agregar_menu()
        self.scrolledtext1 = st.ScrolledText(
            self.ventana1, width=40, height=10)
        self.scrolledtext1.grid(
            column=1, row=1, padx=5, pady=5, sticky="nswe")
        self.listbox1 = li(self.ventana1, height=25, width=20)
        self.listbox1.grid (
            column=0, row=0, padx=5, pady=5, rowspan=2, sticky="nswe")
        img = ImageTk.PhotoImage(Image.open("bn.gif"))
        self.listbox1.bind('<Double-1>', self.muestra_info)

        self.labeledtext = lb(
            self.ventana1, image=img, width= 400, height=300)
        self.labeledtext.grid(column=1, row=0, padx=5, pady=5, sticky="nswe"
)

        self.ventana1.lift()
        self.ventana1.attributes('-topmost', True)
        self.ventana1.after_idle(self.ventana1.attributes, '-
topmost', False)
        self.ventana1.mainloop()

    def muestra_info(self, event):
        imselect = self.listbox1.curselection()
        imoriginal = self.listbox1.get(imselect)
        imotsu = self.directorio + \
            "filtrados/" + imoriginal[:-4] + "_otsu.png"
        imapt = self.directorio + \
            "filtrados/" + imoriginal[:-4] + "_APT.png"
        imhisto = self.directorio + \
            "filtrados/" + imoriginal[:-4] + "_hist.png"
        datos = self.directorio + \
            "filtrados/" + imoriginal[:-4] + '_data.txt'
        img = im(file = imhisto).subsample(1)
        self.labeledtext.configure(image = img)
        self.labeledtext.image = img

        f = open(datos, 'r')
        texto = f.read()
        self.scrolledtext1.delete('0.0')

```

```

self.scrolledtext1.insert('0.0',texto)

self.ventana1.update()

time.sleep(2)

imagen = Image.open(imotsu)

imagen.show()
imagen = Image.open(imapt)
imagen.show()

def agregar_menu(self):
    menubar1 = tk.Menu(self.ventana1)

    self.ventana1.config(menu=menubar1)
    opciones1 = tk.Menu(menubar1, tearoff=0)
    opciones1.add_command(
        label="Procesar archivo", command=self.recuperar)
    opciones1.add_command(
        label="Procesar carpeta", command=self.recuperar_carpeta)
    opciones1.add_separator()
    opciones1.add_command(label="Salir", command=self.salir)
    menubar1.add_cascade(label="Archivo", menu=opciones1)

    opciones2 = tk.Menu(menubar1, tearoff=0)
    opciones2.add_command(label="Parámetros", command=self.parametros)
    menubar1.add_cascade(label="Configuración", menu=opciones2)

    opciones3 = tk.Menu(menubar1, tearoff=0)
    opciones3.add_command(label="Acerca de", command=self.about)
    menubar1.add_cascade(label="Ayuda", menu=opciones3)

def salir(self):
    self.ventana1.destroy()

def guardar(self):
    nombreach=fd.asksaveasfilename(
        initialdir = "/",title = "Guardar como",filetypes = ((
            "txt files", "*.txt"),("todos los archivos","*.*")))
    if nombreach!='':
        archi1=open(nombreach, "w", encoding="utf-8")
        archi1.write(self.scrolledtext1.get("1.0", tk.END))
        archi1.close()
        mb.showinfo(
            "Información", "Los datos fueron guardados en el archivo."
        )

def recuperar(self):
    nombreach=fd.askopenfilename(

```

```

        initialdir = "./",title = "Seleccione archivo",filetypes = ((
            "imágenes","*.jpg *.png"),("mapas de bits","*.bmp"),(
            "todos los archivos","*.*"))
    directorio = os.path.split(nombreach)[0]
    nombre = os.path.split(nombreach)[1]
    nuevo_directorio = directorio+"/filtrados/"
    if not os.path.exists(nuevo_directorio):
        os.mkdir(directorio+"/filtrados")
    if nombreach!='':
        print("FILTRANDO")
        binariza(nombreach, directorio+"/filtrados/"+nombre,
            self.sensibilidad)
    self.directorio = directorio + "/"
    self.listbox1.insert(0,nombre)

def recuperar_carpeta(self):
    directorio = fd.askdirectory(title = "Seleccione carpeta")
    lista_archivos = os.listdir(directorio)
    if ("filtrados" in lista_archivos):
        lista_archivos.remove("filtrados")
    # string_archivos = self.seleccionados_to_string(lista_archivos)
    nuevo_directorio = directorio+"/filtrados/"
    # self.scrolledtext1.insert("1.0",string_archivos)
    if not os.path.exists(nuevo_directorio):
        os.mkdir(directorio+"/filtrados")
    for im in lista_archivos:
        self.listbox1.insert(0,im)
        binariza(directorio+"/"+im, nuevo_directorio+im,
            self.sensibilidad)
    self.directorio = directorio + "/"

def seleccionados_to_string(self, lista_archivos):
    return "Archivos seleccionados:\n"+" \n".join(lista_archivos)

def about(self):
    self.acercaDialogo = t1()
    self.acercaDialogo.grab_set()
    self.acercaDialogo.title("Acerca de mi")
    txtacerca = lb(self.acercaDialogo, text=textos.acerca_de)
    txtacerca.pack()
    # self.scrolledtext1.insert("1.0", textos.acerca_de )

def parametros(self):
    self.paramDialogo = t1()
    self.paramDialogo.grab_set()
    self.paramDialogo.title("Parámetros")
    lb(self.paramDialogo, text="Sensibilidad").pack()
    self.var = dv(self.paramDialogo, self.sensibilidad*100)
    scaleSensib = sc(self.paramDialogo, from_=0, to_=100,

```

```

        variable = self.var)
scaleSensib.pack()
bt(self.paramDialogo, text = 'Aceptar',
    command = self.set_sens).pack()

def set_sens(self):
    self.sensibilidad = self.var.get()/100
    self.paramDialogo.destroy()

def leerDatos(ruta):
    ruta = ruta + "/filtrados/"
    lista_archivos = os.listdir(ruta)
    lista_txt = [a for a in lista_archivos if ".txt" in a]
    for arch in lista_txt:
        rutaArchivo = ruta+arch
        f = open(ruta + arch)
        f.seek(13)
        otsu = f.read(3)
        f.seek(30)
        APT = f.read(3)
        print(otsu, APT)

# app.mainloop()
aplicacion1=Aplicacion()

```

```

'''
umbral.py
Herbert Vargas Gutierrez
umbralizacion adaptativa '''

from PIL import Image
import matplotlib.pyplot as plt

umbral = 0 # Global, que utilizaremos para normalizar el APT

def otsu(grayImage):
    global umbral
    pixelEditor = grayImage.load()
    width,height = grayImage.size
    tamImagen = width * height
    # Calcular el histograma de la imagen
    imgHistogram = grayImage.histogram()
    varmax,sum,sumB,q1,q2,mu1,mu2 = 0,0,0,0,0,0,0

    for i in range(0,256):
        sum += i * imgHistogram[i]
    for j in range(0,255):
        q1 += imgHistogram[j] # Sombras
        q2 = tamImagen - q1 # Luces
        if q1 == 0: continue
        if q2 == 0: continue
        sumB += j * imgHistogram[j]
        mu1 = sumB/q1 # Promedio Sombras
        mu2 = (sum-sumB) /q2 # Promedio Luces
        varSquared = q1*q2 * ((mu1 - mu2)*
                               (mu1 - mu2)) # Diferencia de clases

        if varSquared > varmax:
            umbral = j
            varmax = varSquared
    return umbral, pixelEditor

def imageCreator(umbral,pixelEditor,image):
    width,height = image.size
    #Creando imagen binaria
    for a in range(0,width):
        for b in range(0,height):
            grayValue = pixelEditor[a,b]
            if grayValue > umbral:
                pixelEditor[a,b] = 255
            else:
                pixelEditor[a, b] = 0

def APT(grayImage, sensibilidad=0.9):
    global umbral

```

```

pixelEditor = grayImage.load()
width,height = grayImage.size
umbral = int(umbral*sensibilidad)
imageCreator(umbral,pixelEditor,grayImage)

# Corriendo el metodo otso y APT para La imagen
def binariza(imagePath, destino = "./", sensibilidad=0.9):
    global umbral
    imgris = None
    f = open(destino[:-4] + '_data.txt', 'w')
    imgOTSU = Image.open(imagePath).convert('L')
    imgris = imgOTSU.convert('L')
    histograma = imgris.histogram()
    graficar(histograma, destino[:-4] + '_hist.png')
    umbral, pixelEditor = otsu(imgOTSU)
    f.write('Umbral Otsu: ' + str(umbral)+ '\n')
    imageCreator(umbral,pixelEditor,imgOTSU)
    # imgOTSU.show()
    imgOTSU.save(destino[:-4] + '_otsu.png',bits=1,optimize=True)
    imgAPT = Image.open(imagePath).convert('L')
    imgAPT = imgAPT.quantize(umbral)
    #otsu(imgAPT)
    imgAPT = Image.open(imagePath).convert('L')
    APT(imgAPT, sensibilidad)
    f.write('Umbral APT: ' + str(umbral) + '\n')
    imageCreator(umbral,pixelEditor,imgAPT)
    # imgAPT.show()
    imgAPT.save(destino[:-4]+'_APT.png',bits=1,optimize=True)
    umbral = 0
    f.write('Histograma escala de grises: ')
    f.write(str(histograma))
    f.close()

def graficar(datos, nombre_del_archivo):
    plt.figure()
    x=range(len(datos))
    plt.xticks([0, 50, 100, 150, 200, 255],[0, 50, 100, 150, 200, 255])
    plt.bar(x, datos, align='center')
    plt.title('Histograma')
    plt.xlabel('Valores de intensidad')
    plt.ylabel('Numero de pixeles')

    plt.savefig(nombre_del_archivo, bbox_inches='tight')

    return None

```

2. Resultados de pruebas con imágenes de diarios

DIARIOS

Imágenes binarizadas con el método APT

08/08/2020	12:46	284'049	PCM_1996_2_page29_image101_APT.png
08/08/2020	12:47	257'642	PCM_1996_2_page29_image102_APT.png
08/08/2020	12:48	194'852	PCM_1996_2_page29_image118_APT.png
08/08/2020	12:48	180'439	PCM_1996_2_page29_image123_APT.png
08/08/2020	12:49	339'706	PCM_1996_2_page29_image134_APT.png
08/08/2020	12:49	164'239	PCM_1996_2_page29_image147_APT.png
08/08/2020	12:49	331'533	PCM_1996_2_page29_image163_APT.png
08/08/2020	12:50	636'899	PCM_1996_2_page29_image172_APT.png
08/08/2020	12:50	318'902	PCM_1996_2_page29_image179_APT.png
08/08/2020	12:51	407'840	PCM_1996_2_page29_image187_APT.png
08/08/2020	12:51	299'156	PCM_1996_2_page29_image199_APT.png
08/08/2020	12:51	306'632	PCM_1996_2_page29_image213_APT.png
08/08/2020	12:52	300'565	PCM_1996_2_page29_image228_APT.png
08/08/2020	12:52	109'773	PCM_1996_2_page29_image236_APT.png
08/08/2020	12:53	100'329	PCM_1996_2_page29_image245_APT.png
08/08/2020	12:54	48'173	PCM_1996_2_page29_image247_APT.png
08/08/2020	12:53	397'889	PCM_1996_2_page29_image24_APT.png
08/08/2020	12:55	97'173	PCM_1996_2_page29_image255_APT.png
08/08/2020	12:55	451'061	PCM_1996_2_page29_image31_APT.png
08/08/2020	12:56	425'880	PCM_1996_2_page29_image43_APT.png
08/08/2020	12:57	387'524	PCM_1996_2_page29_image51_APT.png
08/08/2020	12:57	572'004	PCM_1996_2_page29_image54_APT.png
08/08/2020	12:58	452'651	PCM_1996_2_page29_image66_APT.png
08/08/2020	12:59	420'275	PCM_1996_2_page29_image75_APT.png
08/08/2020	13:00	426'074	PCM_1996_2_page29_image77_APT.png
08/08/2020	13:00	394'506	PCM_1996_2_page29_image83_APT.png
08/08/2020	13:01	98'855	PCM_1996_2_page29_image87_APT.png
08/08/2020	13:02	142'471	PCM_1996_2_page29_image99_APT.png
08/08/2020	13:01	427'578	PCM_1996_2_page29_image9_APT.png

Imágenes binarizadas con el método OTSU

08/08/2020	12:46	284'059	PCM_1996_2_page29_image101_otsu.png
08/08/2020	12:47	257'657	PCM_1996_2_page29_image102_otsu.png
08/08/2020	12:47	195'091	PCM_1996_2_page29_image118_otsu.png
08/08/2020	12:48	180'421	PCM_1996_2_page29_image123_otsu.png
08/08/2020	12:48	290'755	PCM_1996_2_page29_image134_otsu.png
08/08/2020	12:49	164'927	PCM_1996_2_page29_image147_otsu.png
08/08/2020	12:49	492'270	PCM_1996_2_page29_image163_otsu.png
08/08/2020	12:50	474'464	PCM_1996_2_page29_image172_otsu.png
08/08/2020	12:50	237'061	PCM_1996_2_page29_image179_otsu.png
08/08/2020	12:50	528'491	PCM_1996_2_page29_image187_otsu.png
08/08/2020	12:51	229'871	PCM_1996_2_page29_image199_otsu.png
08/08/2020	12:51	450'053	PCM_1996_2_page29_image213_otsu.png
08/08/2020	12:52	237'717	PCM_1996_2_page29_image228_otsu.png
08/08/2020	12:52	109'107	PCM_1996_2_page29_image236_otsu.png
08/08/2020	12:53	99'637	PCM_1996_2_page29_image245_otsu.png
08/08/2020	12:53	48'207	PCM_1996_2_page29_image247_otsu.png
08/08/2020	12:52	383'409	PCM_1996_2_page29_image24_otsu.png
08/08/2020	12:54	96'329	PCM_1996_2_page29_image255_otsu.png
08/08/2020	12:55	451'675	PCM_1996_2_page29_image31_otsu.png
08/08/2020	12:55	472'708	PCM_1996_2_page29_image43_otsu.png
08/08/2020	12:56	385'728	PCM_1996_2_page29_image51_otsu.png
08/08/2020	12:57	632'139	PCM_1996_2_page29_image54_otsu.png
08/08/2020	12:58	395'403	PCM_1996_2_page29_image66_otsu.png
08/08/2020	12:59	417'758	PCM_1996_2_page29_image75_otsu.png
08/08/2020	13:00	425'447	PCM_1996_2_page29_image77_otsu.png
08/08/2020	13:00	432'949	PCM_1996_2_page29_image83_otsu.png
08/08/2020	13:00	99'036	PCM_1996_2_page29_image87_otsu.png
08/08/2020	13:01	142'766	PCM_1996_2_page29_image99_otsu.png
08/08/2020	13:01	415'969	PCM_1996_2_page29_image9_otsu.png

Imágenes escaneadas

08/08/2020	00:07	582'618	PCM_1996_2_page29_image101.png
08/08/2020	00:07	506'339	PCM_1996_2_page29_image102.png
08/08/2020	00:08	1'757'539	PCM_1996_2_page29_image118.png
08/08/2020	00:08	993'791	PCM_1996_2_page29_image123.png
08/08/2020	00:08	5'723'709	PCM_1996_2_page29_image134.png
08/08/2020	00:08	844'836	PCM_1996_2_page29_image147.png
08/08/2020	00:08	6'379'980	PCM_1996_2_page29_image163.png
08/08/2020	00:08	10'789'114	PCM_1996_2_page29_image172.png
08/08/2020	00:08	4'038'065	PCM_1996_2_page29_image179.png
08/08/2020	00:08	6'554'915	PCM_1996_2_page29_image187.png
08/08/2020	00:08	3'925'276	PCM_1996_2_page29_image199.png
08/08/2020	00:09	6'392'328	PCM_1996_2_page29_image213.png
08/08/2020	00:09	3'883'356	PCM_1996_2_page29_image228.png
08/08/2020	00:09	774'799	PCM_1996_2_page29_image236.png
08/08/2020	00:06	2'876'725	PCM_1996_2_page29_image24.png
08/08/2020	00:09	1'150'265	PCM_1996_2_page29_image245.png
08/08/2020	00:09	292'969	PCM_1996_2_page29_image247.png
08/08/2020	00:09	689'491	PCM_1996_2_page29_image255.png
08/08/2020	00:06	2'200'713	PCM_1996_2_page29_image31.png
08/08/2020	00:07	3'811'484	PCM_1996_2_page29_image43.png
08/08/2020	00:07	1'503'513	PCM_1996_2_page29_image51.png
08/08/2020	00:07	6'010'861	PCM_1996_2_page29_image54.png
08/08/2020	00:07	3'797'472	PCM_1996_2_page29_image66.png
08/08/2020	00:07	2'432'693	PCM_1996_2_page29_image75.png
08/08/2020	00:07	2'649'857	PCM_1996_2_page29_image77.png
08/08/2020	00:07	3'868'164	PCM_1996_2_page29_image83.png
08/08/2020	00:07	1'254'429	PCM_1996_2_page29_image87.png
08/08/2020	00:06	3'207'930	PCM_1996_2_page29_image9.png
08/08/2020	00:07	1'802'332	PCM_1996_2_page29_image99.png

3. Partidas Originales

2 En esta Santa Iglesia Parroquial de San Pedro de Ondahuailas a dieciocho de enero de mil novecientos treinta, yo
 N^o 3 el Cura y Vicario, preveo las informaciones del caso
 hechas con- leídas las tres amonestaciones preveritas por el derecho y
 ras con no habiéndose resultado impedimento alguno, así y velle según
 bución de Nuestr a Iglesia a Fructo Vargas, hijo legítimo de Matías
 Garfias Vargas y de Dorotea Galindo, natural de Tacucha, bautizada
 en la parroquia de esa y feligres del mismo, de veinticuatro
 años de edad, de profesión agricultor y de estado soltero con
 Francisca Garfias, hija legítima de Julián Garfias y de Cecilia
 Aldasabal natural de Tacucha, bautizada en la parroquia
 de esta y feligres de la misma de veintidos años de edad
 de estado soltero, siendo padrinos don Maximo Altamirano
 y Francisca Altamirano; y testigo Francisco Vargas
 y Pedro Soto; doy fe. *Alfonso San*

N^o 4 En esta Santa Iglesia Parroquial de San Pedro de Ondahuailas a veintinueve de enero de mil novecientos treinta, yo
 Gregorio el Cura y Vicario, hechas las informaciones del caso
 hechas con- leídas las amonestaciones preveritas por el derecho y no habiéndose re-
 sultado impedimento alguno, así y velle según rito de Nuestr
 a Iglesia a Gregorio Huamán, hijo legítimo de Baltasar Huamán
 y de Jacoba Rojas, natural de esta bautizado en la parroquia
 de la misma y feligres de esta de veintian años de edad de prof-
 sión agricultor y de estado soltero con Rosa Pariona hija legítima
 de Altamirano Pariona y de Rosana Quispe, natural de esta bauti-
 zado en la misma y feligres de esta de veintidos años de edad
 de estado soltero, siendo padrinos Francisco Orrego y Bertha Ro-
 ja y testigo Pablo Quispe Francisco Juarez; doy fe. -
Alfonso San

4
n° 7
Moisés
Nehmas
con
Hilda
Isabel
Valdivia

En esta Santa Iglesia Parroquial de San Pedro de Andahuaylas a treinta de Enero de mil novecientos treinta, yo el Cura y Vicario, previas las informaciones del caso, leídas las tres amonestaciones prescritas por el Derecho y no habiendo resultado impedimento alguno, y veli segun rito de nuestra Iglesia a Moises Nehmas, hijo legitimo de don Mayas Nehmas y de Sarah Sillon, natural de Aleppo (Siria) bautizado en la parroquia de esta y feligrés de la misma de veintun años de edad de estado soltero y de profesion Comerciante con Hilda Isabel Valdivia, hija legitima de Alejandro Valdivia y de Maunela Freon, natural de Calavera, bautizada en la parroquia de esa y feligrés de esta de dieciseis años de edad, de estado soltera, siendo padrinos don Saul Nehmas y doña Omelia Valdivia; y testigos Berapio Lavar y Demetrio del Pozo: Doy fe...

N° 8
José J. Gon-
zales
con
Angelica
Menacho.

En esta Santa Iglesia Parroquial de San Pedro de Andahuaylas a los veintidos dias del mes de abril del año mil novecientos treinta, yo el Cura y Vicario previas las informaciones que el derecho exige, leídas las tres proclamas que mandó el Santo Concilio de Trento y no habiendo resultado impedimento alguno, y veli segun el rito de nuestra Santa madre Iglesia a José J. Gonzales de veintidos años de edad, hijo natural de Julian Gonzales y Maria C. Castañeda de profesion sin empleo bautizado en Ahuacay,

N.º 10 En esta Santa Iglesia Parroquial de San Francisco de Andahuaylas a primeros de mayo del año mil novecientos treinta, yo el Cura y vicario previas las informaciones que el derecho exige y leídas las tres proclamas que manda el Santo Concilio de Trento y no habiendo resultado impedimento alguno casé y celebré segun el rito de nuestra Santa Madre Iglesia a Francisco Arvidiano de veintim años de edad hijo legitimo de Manuel Arvidiano y Catalina de las vas de profesion agricultor bautizado en esta Parroquia soltero con Justina Moreno de dieciocho años de edad hija legitima de Andres Moreno ~~y Rosa~~ ^{Justina} Ortiz bautizada en esta Parroquia estado soltera, siendo padrinos Casaro Barrillo y Julia Flores de Carrillo y testigos Baltazar Pando y Mariano Zuniga Duffo.

Aff. sus. *[Signature]*

N.º 11 En esta Santa Iglesia Parroquial de San Santiago de Andahuaylas a primeros de mayo de mil novecientos treinta, yo el Cura y vicario previas las informaciones que el derecho exige y leídas las tres proclamas que manda el Santo

Rivera
con
Carmen
Moreno

veinte años de edad hija legítima de Juan
deancio Valdivia y de Dominga Ortiz de Coracra
bautizada en esta Parroquia de estado sol-
tera siendo padrinos Dr Pedro Casas
ca y Natividad de Salafranca y Testigos
Manuel Barhuaya y Meices Eugman. Day fe.

Alfonso San

Nº 13
En esta Santa Iglesia Parroquial de San Pedro de Ju-
dahuy las a los veinte días del mes de mayo del
año mil novecientos treinta y el cura y vicario
previas las informaciones que el derecho exige
y leídas las tres proclamas que manda el Santo
Concilio de Trento, no habiendo resultado in-
pedimiento alguno case y vele según el rito de
nuestra Santa Madre Iglesia a Manuel Jimenez
de veinte años de edad hijo legítimo de Barto-
lome Jimenez y de Martina Ortega de esta
ciudad de profesión agricultor, bautizado en
esta Parroquia de estado soltero con Alexan-
dra Pastor de veinte años de edad hija
legítima de Manuel Pastor y de
Doña Paula Sifavera bautizada en
esta parroquia de estado soltera. Siem-
do padrinos Lorenzo Romero y Laurita
ma Velasquez y testigos Manuel Peralta
y Pedro Perez Day fe.

Alfonso San

10 de profesión Agricultor, bautizados en esta Parroquia de estado soltero por Emilia Ratonne de veinte y dos años hija legítima de Marcelino Ratonne y Catalina Inocencio bautizados en esta Parroquia de estado soltera siendo padrinos José Adolfo Alarcón y Guimaraes Casapauca de Alarcón y testigos Abatías Chaves y José María Espinosa. Dox fe.

Alfonso Torres

Nº 16 En esta Santa Iglesia Parroquial de San Pedro de Andahuaylas a los diez y seis días del mes de Julio del año mil novecientos treinta, yo el cura y vicario previas las informaciones que el derecho exige y leídas las tres proclamas que manda el Santo Concilio de Trento y no habiendo resultado impedimento alguno, para y vela según el rito de nuestra Santa Madre Iglesia a Justo Fernández de diez y ocho años de edad hijo legítimo de Felisa Fernández y María Orosco de Pampamarca de esta ciudad de profesión Agricultor, bautizados en esta Parroquia de estado soltero por Uta Huaraca de diez y seis años de edad hija legítima de Alberto Huaraca y Getulio Anduadla y bautizados en esta Parroquia de estado soltera siendo padrinos León Quipe y Dolores Huamán y testigos Gerardo Cardenas y Guillermo Orosco. Dox fe.

Alfonso Torres

En esta Santa Iglesia Parroquial de San Pedro de Andahuaylas a los veinticuatro días del mes de Julio del año mil novecientos treinta, yo el cura y vicario previas

No.

En la iglesia de Quilahuayla a quince de marzo de mil novecientos veinte y tres
yo el infrascrito misionero, con licencia del Párroco, previas la información y las proclamas, casé y velé á
Romundo Gato de unos veintiseis años de edad,
soltero natural y vecino de Pollocata
hijo legítimo de Ramón y de Juana Comadilla
con Francisca Cardenas de unos veintinueve años de edad,
soltera natural y vecina de Pollocata
hija legítima de Cornelio y de Isolda Torres
Fueron testigos Gregorio Gato y Mariana Comas
y padrinos Hilberto Espino y Nicolasa Manuco
De que certifico P. Lorenzo Speckert

No.

En la iglesia de Quilahuayla a dieciséis de marzo de mil novecientos veinte y tres
yo el infrascrito misionero, con licencia del Párroco, previas la información y las proclamas, casé y velé á
Bernardo Gullen de unos veintiseis años de edad,
soltero natural y vecino de Pomaspungum
hijo legítimo de Gregorio y de Teodora Quinte
con Leandrea Mancalle de unos veintinueve años de edad,
soltera natural y vecina de Pomaspungum
hija legítima de Gregorio y de Lucinda Huamán
Fueron testigos Bernardo Mancalle y Bernabé Olivera
y padrinos Bernardo Mancalle y Enporaniza Huamán
De que certifico P. Lorenzo Speckert

No.

En la iglesia de Quilahuayla a dieciséis de marzo de mil novecientos veinte y tres
yo el infrascrito misionero, con licencia del Párroco, previas la información y las proclamas, casé y velé a
Pedro Perez de unos veintiseis años de edad,
soltero natural y vecino de Pomaspungum
hijo legítimo de Francisco y de Gregoria Paucari
con Victoria Quispi de unos veintinueve años de edad,
soltera natural y vecina de Comalbat
hija legítima de Marcelo y de Cecilia Allica
Fueron testigos Carpus Quispi y Jaime P. ray
y padrinos Carpus Quispi y María Belarza
De que certifico P. Lorenzo Speckert

Cecilio Chicella e. Hilaria Gauris N°.....
 En la iglesia de Andahuayla a dieciseis de marzo de mil novecientos Treinta y dos
 yo el infrascrito misionero, con licencia del Párroco, previas la información y las proclamas, casé y velé a
 Cecilio Chicella de unos veintifidos años de edad,
 soltero natural y vecino de San Jeronimo
 hijo de Legidimo de Leonor y de Asuncion Rojas
 con Hilaria Gauris de unos veintinueve años de edad,
 soltera natural y vecina de San Jeronimo
 hija de Legidimo de Leonor y de Andreea Chuaco
 Fueron testigos: Juan Cruz Zamora y Estanislao Luján
 y padrinos: Gregorio Perez y Francisca Alcaza
 De que certifico P. Lorenzo Speyer

Elpidio Quispe e. Asunta Centeno N°.....
 En la iglesia de Andahuayla a dieciseis de marzo de mil novecientos Treinta y dos
 yo el infrascrito misionero, con licencia del Párroco, previas la información y las proclamas, casé y velé a
 Elpidio Quispe de unos veintidos años de edad,
 soltero natural y vecino de San Jeronimo
 hijo de Legidimo de Elpidio y de Leguntina Mallque
 con Asunta Centeno de unos veintidos años de edad,
 soltera natural y vecina de Andahuayla
 hija de Legidimo de Asuncion y de Melchora Pichihua
 Fueron testigos: Casimiro Martinez y Prudencio Urbina
 y padrinos: Juan de Dios Centeno y Teresa Quispe
 De que certifico P. Lorenzo Speyer

Gregorio Centeno e. Maria Caza N°.....
 En la iglesia de Andahuayla a dieciseis de marzo de mil novecientos Treinta y dos
 yo el infrascrito misionero, con licencia del Párroco, previas la información y las proclamas, casé y velé a
 Gregorio Centeno de unos veintidós años de edad,
 soltero natural y vecino de Pollocosa
 hijo de Legidimo de Mariano y de Lorenza Cullia
 con Maria Caza de unos veintidós años de edad,
 soltera natural y vecina de Pollocosa
 hija de Legidimo de Damasa y de Juana Lantilla
 Fueron testigos: Marcelino Perales y Prudencio Urbina
 y padrinos: Norberto Lopez y Nicolsa Manuaco
 De que certifico P. Lorenzo Speyer

Mariano Oros con Juana Chiquipa
 En la iglesia de Andahuailas, a de de mil novecientos treinta y seis.
 yo el infrascrito misionero, con licencia del Párroco, previas la información y las proclamas, casé y velé a
 Mariano Oros de unos veintinueve años de edad,
 soltero natural y vecino de Cuncatoca
 hijo legítimo de Lorenzo Oros y de Paula Otáño
 con Juana Chiquipa de unos treinta años de edad,
 soltera natural y vecina de Cuncatoca
 hija legítima de Inocencia Chiquipa y de Bárbara Ruiz
 Fueron testigos Luis Cardenas, Pedro Huaman
 y padrinos Pedro Huaman, Francisca Bello
 De que certifico
 P. Corcuera

Saturnino Oros con Agustina Ruiz N.º 110
 En la iglesia de Andahuailas, a de de mil novecientos treinta y seis.
 yo el infrascrito misionero, con licencia del Párroco, previas la información y las proclamas, casé y velé a
 Saturnino Oros de unos veintiseis años de edad,
 soltero natural y vecino de Cuncatoca
 hijo legítimo de Lorenzo Oros y de Paula Otáño
 con Agustina Ruiz de unos veintiseis años de edad,
 soltera natural y vecina de Cuncatoca
 hija legítima de Victor Ruiz y de Juana Huaman
 Fueron testigos Luis Cardenas, Pedro Huaman
 y padrinos Pedro Huaman, Francisca Bello
 De que certifico
 P. Corcuera

Genaro Arenas con Francisca Guape N.º 111
 En la iglesia de Andahuailas, a de de mil novecientos treinta y seis.
 yo el infrascrito misionero, con licencia del Párroco, previas la información y las proclamas, casé y velé a
 Genaro Arenas de unos veintiocho años de edad,
 soltero natural y vecino de Cuncatoca
 hijo legítimo de Manuel Arenas y de Inocencia Chiquipa
 con Francisca Guape de unos veintiocho años de edad,
 soltera natural y vecina de Cuncatoca
 hija legítima de Pablo Guape y de Branda Guape
 Fueron testigos Pasqual Malpartida, Luis Venegas
 y padrinos Luis Serra, Luisa Martinez
 De que certifico
 P. Corcuera

Benito Contreras

Nº 7

En la iglesia de Ceñuauran a quince de marzo de mil novecientos treinta y seis
 yo el infrascrito misionero, con licencia del Párroco, previas la información y las proclamas, casé y velé á
Benito Contreras de unos 28 años años de edad,
 soltero natural y vecino de l lugar
 hijo legitimo de Demetrio y Contreras y de Rafita Rojas
 con Juliana Huaraca de unos veintaycuatro años de edad,
 soltera natural y vecina de l lugar
 hija legitima de Manuel y Huaraca y de Antonia Castro
 Fueron testigos Mariano Perez y Enrique Soto
 y padrinos Alejandro Arquiya y Rafina Iruvia

De que certifico

Alfonso Maria

Miguel Castro

Nº 8

En la iglesia de Ceñuauran a quince de marzo de mil novecientos treinta y seis
 yo el infrascrito misionero, con licencia del Párroco, previas la información y las proclamas, casé y velé á
Miguel Castro de unos veintay siete años de edad,
 soltero natural y vecino de l lugar
 hijo legitimo de Eustasio Castro y de Cornelia Contre
 con Manuela Aparco de unos veintay seis años de edad,
 soltera natural y vecina de l lugar
 hija legitima de Antonio Aparco y de Gregoria Huaman
 Fueron testigos Lorenzo Huarcaya y Octavio Huaraca
 y padrinos Cyrila Begonia y Cyrila Aribaldo

De que certifico

Alfonso Maria

Pedro Perez

Nº 9

En la iglesia de Ceñuauran a quince de marzo de mil novecientos treinta y seis
 yo el infrascrito misionero, con licencia del Párroco, previas la información y las proclamas, casé y velé a
Pedro Perez de unos treintay cinco años de edad,
 viudo de Feliciana Pora natural y vecino de l lugar
 hijo legitimo de Feliz Perez y de Angela Andrata
 con Simona Quispe de unos treintay cinco años de edad,
 viuda de Alberto Perez natural y vecina de l lugar
 hija legitima de Paulo Quispe y de Isabela Huaraca
 Fueron testigos Octavio Huaraca y Lorenzo Huarcaya
 y padrinos Pedro Andrata y Matias Quispe

De que certifico

Pedro Perez y Simona Quispe
 se casaron con dispensacion del impedimento
 affinitatis in primo gradu lineae collateralis. De que
 certifico

Alfonso Maria

50 En esta Santa Iglesia parroquial de San Pedro de Andahuaila, a los veintinueve
 Francisco días del mes de octubre de mil novecientos treinta y tres, yo el infrascrito párroco
 Valarque vos ferir las informaciones canónicas i leidas las amonestaciones ferir
 con enton por el Derecho i no habiendo resultado ningún impedimento cano
 Gregoria por palabras de presente i vale según el rito de nuestra Santa Madre Iglesia
 Rivera a Francisco Varque hijo legítimo de Jacinto Valarque i de Alejandra
 Gomez coltera de veintinueve años de edad natural de esta i residente
 en Casarucalla bautizado i feligres de esta de profesion agricultor
 con Gregoria Rivera hija legítima de Iniquel Rivera i de Inamela
 Perez coltera de veintidos años de edad natural de esta i residente en
 Casarucalla bautizada i feligres de esta de profesion hilandera. cirri
 ron de testigos Manuel Siches, Paula Romero i Cinotos Espinosa. fue
 con sus padrinos Manuel Siches i Paula Romero En fi de lo cual firmo.

Lesq[ue] de Mendoza Farfán

51 En esta Santa Iglesia parroquial de San Pedro de Andahuaila, a los veintinueve
 Cirilo días del mes de octubre de mil novecientos treinta y tres, yo el infrascrito párroco
 Ortis vos ferir las informaciones canónicas i leidas las amonestaciones ferir
 con tos por el Derecho i no habiendo resultado ningún impedimento cano
 Ceferina por palabras de presente i vale según el rito de nuestra Santa Madre Iglesia
 Figueroa a Cirilo Ortis hijo legítimo de Lusi Ortis i de Concepcion Altamirano coltera
 de veintiseis años de edad natural de San Jeronimo i residente en Estora
 bautizado i feligres de San Jeronimo de profesion agricultor con Ceferina
 Figueroa hija legítima de Manuel Figueroa i de Inamela Velasquez coltera
 de treinta años de edad natural de San Jeronimo i residente en Estora
 bautizado i feligres de San Jeronimo de profesion hilandera cirri
 ron de testigos Saturnino Vivanco, Alfredo Ruizaga i Saturnino Inanico. fue
 con sus padrinos Saturnino Inanico i Pablo Román En fi de lo cual firmo.

Lesq[ue] de Mendoza Farfán