Instituto Politécnico Nacional

Centro de Investigación en Computación

Reconocimiento automático de glifos mayas

T E S I S

Que para obtener el grado de:

Maestría en Ciencias de la Computación.

PRESENTA:

Ing. Steffy Mónica Robles Ruiz

Directores de Tesis:

Dr. Francisco Hiram Calvo Castro
Dr. Alexander Gelbukh

Ciudad de México, México. Junio 2017

SIP-14 bis

INSTITUTO POLITÉCNICO NACIONAL

SECRETARÍA DE INVESTIGACIÓN Y POSGRADO

ACTA DE REVISIÓN DE TESIS

En la Ciudad de México siendo las 17:00 horas del día 16 del mes de junio de 2017 se reunieron los miembros de la Comisión Revisora de la Tesis, designada por el Colegio de Profesores de Estudios de Posgrado e Investigación del:

Centro de Investigación en Computación

para examinar la tesis titulada:

"Reconocimiento automático de glifos mayas"

Presentada por el alumno:

ROBLES	RUÍZ	STEFFY MÓNICA
Apellido paterno	Apellido materno	Nombre(s)

Con registro:

A	1	5	0	1	8	7

aspirante de: **MAESTRÍA EN CIENCIAS DE LA COMPUTACIÓN**

Después de intercambiar opiniones los miembros de la Comisión manifestaron ***APROBAR LA TESIS***, en virtud de que satisface los requisitos señalados por las disposiciones reglamentarias vigentes.

LA COMISIÓN REVISORA

Directores de Tesis

Dr. Francisco Hiram Calvo Castro

Dr. Alexander Gelbukh

Dr. Sergio Suárez Guerra

Dr. Grigori Sidorov

Dra. Obdulia Pichardo Lagunas

Dr. Ildar Batyrshin

PRESIDENTE DEL COLEGIO DE PROFESORES

Dr. Marco Antonio Ramírez Salinas

INSTITUTO POLITÉCNICO NACIONAL
CENTRO DE INVESTIGACIÓN EN COMPUTACIÓN
DIRECCIÓN

INSTITUTO POLITÉCNICO NACIONAL

SECRETARÍA DE INVESTIGACIÓN Y POSGRADO

CARTA CESIÓN DE DERECHOS

En la Ciudad de México el día 21 del mes Junio del año 2017, el (la) que suscribe Steffy Mónica Robles Ruiz alumno (a) del Programa de Maestría en Ciencias de la Computación con número de registro A150187, adscrito al Centro de Investigación en Computación, manifiesta que es autor (a) intelectual del presente trabajo de Tesis bajo la dirección de Dr. Francisco Hiram Calvo Castro y Dr. Alexander Gelbukh y cede los derechos del trabajo intitulado Reconocimiento automático de glifos mayas, al Instituto Politécnico Nacional para su difusión, con fines académicos y de investigación.

Los usuarios de la información no deben reproducir el contenido textual, gráficas o datos del trabajo sin el permiso expreso del autor y/o director del trabajo. Este puede ser obtenido escribiendo a la siguiente dirección steffy451@gmail.com. Si el permiso se otorga, el usuario deberá dar el agradecimiento correspondiente y citar la fuente del mismo.

Steffy Mónica Robles Ruiz

Nombre y firma

RESUMEN

Uno de los elementos que distinguen a los seres humanos del resto de los seres vivos es la capacidad para comunicarse haciendo uso del lenguaje. Gracias a la invención del lenguaje escrito, ha sido posible transmitir ideas complejas a través del tiempo. Por tal motivo, el estudio de materiales históricos invaluables, tales como los sistemas de escritura antiguos es un área significativa y de gran interés social; nos ofrece la oportunidad de enriquecer nuestra identidad como seres humanos, permite determinar las áreas en las que las primeras civilizaciones tenían amplios conocimientos y conocer las aportaciones que brindaron a la humanidad.

Por otra parte, desde la invención de la computadora se han desarrollado técnicas computacionales que han demostrado ser muy útiles en el desarrollo científico, un ejemplo claro es el caso de historiadores, arqueólogos, antropólogos y lingüistas quienes a menudo necesitan consultar, anotar y catalogar documentos pictográficos manualmente. La integración de herramientas computacionales puede facilitar significativamente el trabajo diario de dichos investigadores, por ejemplo: la generación y manejo de bases de datos, análisis estadísticos, aplicación de técnicas de clasificación y el uso de algoritmos de visión por computadora.

Sin embargo, no todas las aplicaciones han logrado resolver las tareas por completo, como es el caso de la clasificación automática de glifos mayas, la cual, a pesar de algunos intentos, es una tarea que aún se encuentra sin ser resuelta satisfactoriamente. Gracias al desarrollo de diferentes algoritmos de análisis digital de imágenes y reconocimiento de patrones, en los últimos años ha sido posible analizar y clasificar información de casi cualquier tipo de imágenes, por tal motivo, en este trabajo se propone una metodología para su reconocimiento automático haciendo uso de redes neuronales convolucionales las cuales últimamente han demostrado ser una poderosa clase de modelos para reconocer y clasificar imágenes.

Los trabajos previos que se han realizado en esta área han intentado en su mayoría trabajar directamente con los rasgos que caracterizan a cada glifo, lo cual resulta en una tarea compleja debido a las diferencias de estilo entre un escriba y otro, y además, la naturaleza de la escritura permite que puedan existir variaciones entre glifos de una misma clase. La finalidad de este trabajo es probar que es posible obtener una buena tasa en el reconocimiento automático, enfocándose principalmente en el algoritmo de clasificación y permitiendo que los rasgos característicos sean extraídos por dicho algoritmo.

ABSTRACT

One of the elements that distinguish human beings from the rest of living beings is the ability to communicate by making use of language. Because of the invention of written language, it has been possible to communicate complex ideas through time. Therefore, the study of historical materials, such as ancient writing systems is a significant area with great social interest; it gives us the opportunity to enrich our identity as human beings, allows us to determine in which areas the first civilizations had extensive knowledge and recognize the contributions they provided to humanity.

On the other hand, since the invention of computers there have been developed different computational techniques that have proven to be very useful in scientific development, a clear example is the case of historians, archaeologists, anthropologists and linguists who often need to consult, annotate and catalog pictographic documents manually. The integration of computational tools can facilitate the daily work of researchers, for example: generation and management of databases, statistical analysis, the use of classification techniques and computer vision algorithms.

However, not all applications have been able to solve tasks completely, as is the case of automatic classification of Maya glyphs, which, despite some attempts, it hasn't been solved satisfactorily. Thanks to the increase of development on different algorithms for digital image analysis and pattern recognition, it has been possible to analyze and classify information from almost any type of image, for this reason, a methodology for automatic recognition is proposed, making use of convolutional neural networks which have recently proved to be powerful models for recognizing and classifying images.

Many of the previous works from this area have tried to work directly with the features that characterize each glyph, which results in a complex task due to the style differences between one scribe and another, besides the Mayan writing nature which allows some variations for glyphs from the same class.

The intention of this work is to prove that it is possible to obtain a good rate of precision in recognition of Mayan glyphs, letting it to extract automatically the features and focusing mainly on the classification algorithm.

AGRADECIMIENTOS

A mi familia, en especial a mis padres y mis hermanos, por ser los principales promotores de mis sueños. En mis triunfos y en mis tropiezos, siempre han estado presentes, apoyándome.

A mis directores: Dr. Hiram Calvo y Dr. Alexander Gelbukh, por su apoyo y sus valiosos consejos brindados para la realización de este proyecto.

A mis compañeros: Jesús Pérez Romero, por ser quien siempre me alentaba en los malos momentos, y José Eduardo Valdez Rodríguez, por su ayuda para hacerme entender las redes neuronales.

A Dr. Joan Andreu, por brindarme la oportunidad de realizar mi estancia en Valencia.

A Sergio por su enorme paciencia y su amor incondicional.

A mis amigos, los nuevos y los eternos, por llenarme de risas y momentos inolvidables.

Al Instituto Politécnico Nacional y al Centro de Investigación en Computación por brindar una educación de calidad.

Al Consejo Nacional de Ciencia y Tecnología, por el apoyo económico brindado durante la realización de este proyecto.

"Los abuelos dijeron que uno de ustedes vendría y entendería su escritura.
Esto sería un signo del regreso del antiguo poder Maya…"

Sueño del camino maya
Richard Luxton (1993)

CONTENIDO

INTRODUCCIÓN **6**

1.1 DESCRIPCIÓN DEL PROBLEMA 7

1.2 MOTIVACIÓN 8

1.3 PROPUESTA 8

1.4 OBJETIVO GENERAL 9

1.5 OBJETIVOS ESPECÍFICOS 9

1.6 ESQUEMA GENERAL 10

MARCO TEÓRICO **12**

2.1 LA ESCRITURA MAYA 12

2.1.1 El sistema de escritura 13

2.2 PROCESAMIENTO DIGITAL DE IMÁGENES 16

2.2.1 Representación digital de una imagen 16

2.2.2 Modelo de degradación de imagen 18

2.2.3 Operaciones morfológicas 20

2.2.4 Modelos de ruido 22

2.2.5 Suavizado de imágenes 24

2.2.6 Métodos para transformaciones geométricas 24

2.3 CLASIFICACIÓN DE IMÁGENES 25

2.3.1 Clasificación por vecinos más cercanos 27

2.3.2 Validación cruzada 28

2.3.3 Redes neuronales artificiales 29

2.3.1 Redes neuronales convolucionales 35

TRABAJO RELACIONADO **38**

3.1 CLASIFICACIÓN DE CARACTERES 38

3.1.1 Reconocimiento de caracteres manuscritos 39

3.1.2 Clasificación de símbolos antiguos 39

3.2 CATÁLOGOS DE GLIFOS MAYAS 41

3.2.1 Catálogo de Thompson *41*
3.2.2 Diccionario de Montgomery *42*
3.3 LA COMPUTACIÓN EN EL ÁREA DE LOS GLIFOS MAYAS 42
DESARROLLO 47
4.1 CORPUS 47
4.1.1 Descripción del corpus *48*
4.1.2 Etiquetado del corpus *49*
4.2 AUMENTO DEL CONJUNTO DE DATOS 50
4.2.1 Operaciones morfológicas *51*
4.2.2 Transformaciones geométricas *52*
4.2.3 Modelos de ruido *54*
4.2.4 Filtros *55*
4.2.5 Degradaciones combinadas *56*
4.3 CLASIFICACIÓN DE GLIFOS MAYAS 57
4.3.1 Implementación de red neuronal convolucional *59*
EXPERIMENTOS Y RESULTADOS 62
5.1 DISEÑO DE EXPERIMENTOS 62
5.1.1 Clasificación de datos sintéticos *63*
5.2 CLASIFICACIÓN DE DATOS REALES 63
5.3 ANÁLISIS DE RESULTADOS 65
CONCLUSIONES 71
6.1 TRABAJO FUTURO 73
ANEXO 74

LISTA DE TABLAS

Tabla 1. Funciones de activación más comunes *33*
Tabla 2. Resumen de resultados de trabajos relacionados *46*
Tabla 3. Conjuntos de datos *48*
Tabla 4. Resumen de transformaciones realizadas *57*
Tabla 5. Arquitecturas CNN propuesta *61*
Tabla 6. Distribución de glifos (1) *63*
Tabla 7. Distribución de glifos (2) *64*
Tabla 8. Resultados clasificación de datos sintéticos *65*
Tabla 9. Resultados de clasificación de datos reales *66*
Tabla 10. Significancia estadística experimentos datos sintéticos *69*
Tabla 11. Significancia estadística experimentos datos reales *69*
Tabla 12. Clasificación sin transformaciones *70*
Tabla 13. Comparación entre métodos. *71*

LISTA DE FIGURAS

Figura 1. Orden de lectura para un texto maya. Tomada de Coe (2001) 14
Figura 2. Elementos de un glifo compuesto 14
Figura 3. Orden de lectura para glifos compuestos. Tomada de Coe (2001) 15
Figura 4. Composición glífica. Tomada de Kettunen y Helmke (2008) 15
Figura 5. Ejemplo de representación de una imagen digital 17
Figura 6. Representación binaria de una imagen 18
Figura 7. Modelo del proceso de degradación/restauración de imágenes. Basado en González y Woods (2006) 20
Figura 8. Ejemplo de dilatación. 21
Figura 9. Ejemplo de erosión. Basada en González y Woods (2006) 22
Figura 10. Ejemplo de ruido Gaussiano. 23
Figura 11. Ejemplo de ruido sal y pimienta. Basada en González y Woods (2006) 24
Figura 12. Ilustración de un modelo de clasificación KNN.* 27
Figura 13. Esquema de validación cruzada de 5 iteraciones. 29
Figura 14. Esquema de una neurona artificial. Basado en Demuth, Beale, Jess, y Hagan (2014) 30
Figura 15. Ejemplo de red neuronal con 2 capas ocultas 34
Figura 16. Arquitectura de una red neuronal convolucional 37
Figura 17. Equivalencias de símbolos en los conjuntos de datos 49
Figura 18. Ejemplos de pares de símbolos sin equivalencia 50
Figura 19. Ejemplo de dilatación del glifo T001 52
Figura 20.Ejemplo de erosión 52
Figura 21. Ejemplo de rotación 53
Figura 22. Ejemplo de permutación de pixeles 53
Figura 23. Ejemplo de arremolinado 54
Figura 24. Ejemplo de ruido gaussiano 54
Figura 25. Ejemplo de ruido sal y pimienta 55
Figura 26. Ejemplo de desenfoque 55

Figura 27. Ejemplo de suavizado gaussiano *56*
Figura 28. Combinación de erosión y dilatación *57*
Figura 29. Diseño de experimentos. (a) Datos sintéticos. (b) Datos reales *62*
Figura 30. Elementos de estructura para dilatación. Tamaño 3x3 ..... *74*
Figura 31. Elementos de estructura para dilatación. Tamaño 2x2 ..... *74*
Figura 32. Elementos de estructura para erosión. Tamaño 2x2 ..... *75*
Figura 33. Arremolinado de una imagen *75*

1

Introducción

La civilización Maya es considerada una de las más importantes del mundo antiguo, equiparable con la egipcia y la china. Son reconocidos por su arte, arquitectura, sus conocimientos en matemáticas y astronomía, además de ser los únicos del continente americano que, previo a la conquista española, desarrollaron un sistema de escritura completo con el cual pudieron transmitir su visión y conocimiento del mundo, sus ideas religiosas y su historia. Las señales de su pasado son muy abundantes y los trabajos arqueológicos siguen obteniendo descubrimientos relacionados con ellos.

Se asentaron en un área muy extensa del sureste de México y en los territorios de América Central: Belice, Guatemala, Honduras y El Salvador. Con una historia de aproximadamente 3000 años, la historia de la civilización maya se divide en tres períodos principales: el preclásico (2000 a. C. – 250 d. C.), clásico (250 – 950) y posclásico (1000 - 1519).

La mayoría de textos mayas de los que se tiene conocimiento fueron escritos durante el periodo clásico; algunos fueron tallados o pintados en monumentos de piedra, elementos arquitectónicos como columnas, dinteles, pinturas en murales, así como en artículos personales como piedras preciosas, vasos de cerámica, huesos de animales, entre otros. Existe también otro tipo de textos mayas conocidos como códices, libros en forma de biombo hechos de papel amate, utilizados para registrar principalmente asuntos calendáricos

y rituales. Tras la llegada de los españoles, se perdió todo rastro de la escritura maya y sólo se conservan tres códices: Madrid, Dresde y París, nombrados así por la ciudad en la que se encuentran actualmente.

1.1 Descripción del problema

El análisis de la escritura maya ha sido durante mucho tiempo una tarea muy compleja, ya que no existe algún tipo de texto traducido a alguna lengua conocida que acompañe a los textos, como fue el caso de los jeroglíficos egipcios y la piedra de Rosetta. Sin embargo, los grandes esfuerzos de arqueólogos, epigrafistas e historiadores han llevado a la obtención de importantes avances en el desciframiento de la escritura maya y, tras casi 200 años de investigación, más de 1,000 signos han sido clasificados en diferentes catálogos.

Knórozov (1999) explica que para las tareas de transcripción y traducción de cualquier texto desconocido es necesario tener una aguda capacidad para reconocer cada uno de los grafemas y los errores en el reconocimiento de ellos son muy perjudiciales debido a que su corrección posterior resulta sumamente difícil. Por esta razón, el proceso para la traducción de los textos mayas resulta muy laborioso ya que requiere que los epigrafistas inviertan una gran cantidad de tiempo buscando en los catálogos existentes para identificar los glifos individuales de cada uno de los bloques que conforman un escrito.

La clasificación de cada glifo es el paso inicial para obtener la transcripción, posteriormente generar la transliteración y finalmente obtener una traducción. Sin embargo, la complejidad de la tarea de identificación aumenta

si se toma en consideración la gran cantidad de variaciones que existen para algunos signos y que una de las principales características de la escritura mayas es que los glifos eran ordenados de forma arbitraria, con variaciones de posición, tamaño y orientación.

En la actualidad, a pesar del desarrollo tecnológico con el que se cuenta, los epigrafistas y expertos mayistas continúan realizando su trabajo de forma manual debido a que no existe una herramienta computacional que pueda ser de ayuda para realizar sus investigaciones.

1.2 Motivación

La motivación de este trabajo recae en que, a pesar del desarrollo tecnológico con el que se cuenta en la actualidad, los epigrafistas y expertos mayistas tienen que realizar su trabajo de forma manual debido a que no existe una herramienta computacional que pueda ser de ayuda para realizar sus investigaciones.

1.3 Propuesta

Gracias al desarrollo de diferentes algoritmos dentro de las áreas de análisis digital de imágenes y reconocimiento de patrones en los últimos años, ha sido posible analizar y clasificar información de casi cualquier tipo de imágenes.

Por esta razón, este trabajo propone una metodología para automatizar el proceso del reconocimiento de ciertos glifos individuales y que, además, permita agilizar y facilitar la investigación destinada al reconocimiento de glifos mayas.

No obstante, aún existen diversos obstáculos que deben ser tomados en consideración para el análisis automático de glifos mayas. En primer lugar, la cantidad de datos de glifos mayas digitalizados y de libre acceso es muy limitada. En segundo lugar, a pesar de que es posible segmentar digitalmente glifos de las diferentes fuentes físicas, los textos han perdido su calidad visual con el paso del tiempo o algunos otros se han borrado parcialmente.

Por último, es importante añadir que, para propósito de este trabajo de tesis, los temas referentes al modelo del lenguaje y traducción quedan fuera de los alcances de la misma, debido a que éstos requieren de conocimientos más específicos acerca del entorno y contexto de los textos.

1.4 Objetivo general

El objetivo principal de esta tesis es desarrollar una metodología para el reconocimiento de glifos mayas simples con la ayuda de técnicas de procesamiento de imágenes.

1.5 Objetivos específicos

Para lograr el objetivo principal de esta tesis es necesario realizar una serie de pasos intermedios. En primer lugar, se debe recopilar un conjunto de datos de glifos digitalizados. Además, es imprescindible abordar el problema de escasez de datos con la generación de variaciones sintéticas. Posteriormente, desarrollar y entrenar un clasificador adecuado para la tarea del reconocimiento de glifos mayas. Finalmente, se deberá evaluar el desempeño del clasificador desarrollado. Para conseguir una evaluación más acertada del

desempeño clasificador, será necesario que, además de realizar experimentos con el conjunto de datos sintéticos, se realicen experimentos con glifos reales, extraídos de códices y que con ellos se pueda brindar una evaluación con una confiabilidad mayor que con la evaluación obtenida al clasificar únicamente datos sintéticos.

1.6 Esquema general

Este trabajo de tesis consta de 6 capítulos explicados a continuación:

El capítulo 2 contiene todos los conceptos que serán utilizados para la propuesta de solución del objetivo principal. En primer lugar, se presenta una breve introducción al sistema de escritura maya, lo cual brinda una idea de la magnitud del problema. Después, se habla brevemente de las técnicas que son utilizadas para el análisis digital de imágenes. Finalmente, se describen los algoritmos de clasificación de imágenes que serán utilizados.

En el capítulo 3 se presentan los trabajos que se encuentran relacionados con esta tesis. Primero se analizan algunos trabajos relevantes acerca del reconocimiento de caracteres. Posteriormente se habla de la catalogación de glifos mayas que es útil para la transcripción y traducción de los símbolos. Finalmente se detallan los trabajos en los que la computación ha sido partícipe en la tarea específica de clasificación de glifos mayas.

El capítulo 4 explica cómo fueron combinados los conceptos presentados en el marco teórico. Primero se describe cómo se logró solucionar el problema de falta de datos para entrenar un algoritmo de aprendizaje supervisado. El desarrollo de las técnicas de clasificación para este problema y las arquitecturas utilizadas.

En el capítulo 5 se exponen los resultados obtenidos en los experimentos de cada una de las técnicas descritas en el capítulo 4 y los resultados de los algoritmos de clasificación. Se presentan también las comparaciones pertinentes con los trabajos relacionados expuestos en el capítulo 3.

El capítulo 6 presenta las conclusiones obtenidas, además se habla sobre las contribuciones de este trabajo de tesis. Finalmente, se exponen algunos puntos importantes para trabajo futuro.

2

Marco teórico

2.1 La escritura maya

La escritura maya puede considerarse como una de las más complejas del mundo antiguo. En sus inscripciones, los mayas registraron sus conocimientos, sus mitos y la historia de los linajes de sus gobernantes. Los primeros textos mayas identificados datan del siglo III a.C., mientras que los más tardíos se realizaron hasta poco después de la llegada de los conquistadores españoles en el siglo XVI. Es posible que la tradición de escribir con jeroglíficos mayas continuara hasta el siglo XVII en algunas regiones alejadas de la invasión española, como el caso de la ciudad de Tayasal ubicada en el norte del Petén.

Kettunen y Helmke (2008) afirman que existen alrededor de 10,000 textos individuales, incluyendo aquellos que se encuentran en museos o colecciones privadas alrededor del mundo, como los códices Madrid, Dresde y París y los que han sido descubiertos arqueológicamente. La mayor parte de estos textos fueron escritos en vasijas cerámicas y en monumentos de piedra, como estelas y dinteles; algunos otros aparecieron en fachadas estucadas, paredes interiores de edificios y de cuevas, conchas, huesos, objetos de jade, obsidiana y barro, que fueron encontrados en tumbas o en sitios que fueron abandonados o enterrados antes de la llegada de los españoles.

2.1.1 El sistema de escritura

Los mayas crearon un sistema de escritura capaz de registrar todas las manifestaciones orales de su lengua. Está compuesto por jeroglíficos ordenados en líneas horizontales o verticales. Los jeroglíficos son bloques que se forman por uno o varios grafemas. Un grafema se define como la unidad mínima e indivisible de la escritura de una lengua y dependiendo de la cantidad de grafemas se pueden tener monogramas, bigramas, trigramas, tetragramas o incluso pentagramas. (Knórozov et al., 1999).

Lingüísticamente, el sistema de escritura maya se describe como un sistema logo silábico, compuesto por logogramas, signos que representan palabras completas, y silabogramas o signos silábicos que pueden funcionar de forma vocálica o silábica. Se calcula que existen más de mil signos diferentes, contando simples y compuestos; sin embargo, muchos de ellos son alógrafos o variaciones del mismo signo, elementos distintos pero que tienen la misma lectura u homófonos o bien, variantes que fueron empleadas en un cierto periodo o en un área determinada. (Kettunen & Helmke, 2008).

Los textos mayas comúnmente son ordenados dentro de un tipo de tabla imaginaria. La manera correcta de leer un bloque de glifos es de izquierda a derecha, de arriba hacia abajo en pares de columnas.

Cabe mencionar que existen algunas excepciones a esta regla, un caso pueden ser los glifos ordenados en una sola columna, los que tienen forma de “L” o los que se encuentran dispuestos horizontalmente formando una fila. Existen algunos otros ejemplos, pero se reitera que la regla básica de lectura es de izquierda a derecha y de arriba hacia abajo.

Si designáramos una numeración a los elementos de un bloque de glifos, el orden de lectura sería el siguiente:

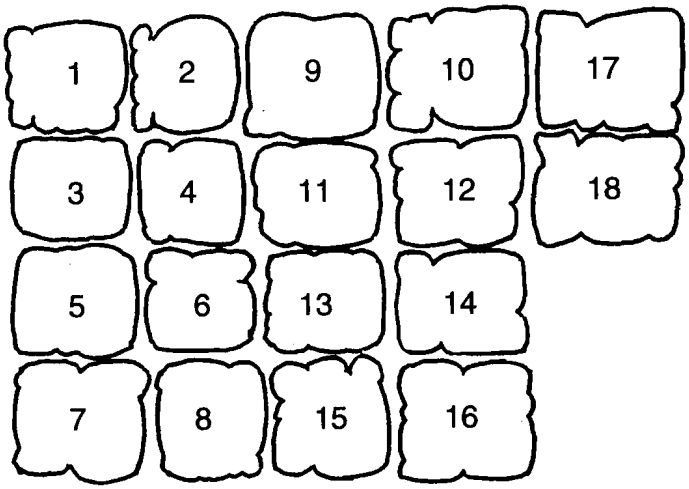

Figura 1. Orden de lectura para un texto maya. Tomada de Coe (2001)

Además del orden de lectura general de un texto, se debe tener en consideración el orden de cada uno de los elementos contenidos en él. Existen dos tipos de elementos: glifos simples o glifos compuestos. Como su nombre lo indica, los glifos simples contienen un solo elemento, mientras que los compuestos contienen más de un elemento significativo (Figura 2).

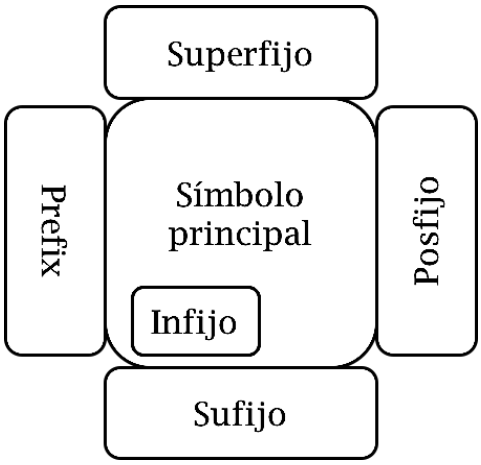

Figura 2. Elementos de un glifo compuesto

En un glifo compuesto, el elemento más grande es llamado símbolo principal y los que lo rodean son llamados afijos. La lectura de un glifo compuesto va de arriba hacia abajo y de izquierda a derecha (Figura 3). Sin embargo, este orden en ocasiones era alterado principalmente por motivos estéticos y algunos otros por motivos culturales, el cual es el caso de los afijos que denotan la palabra *ajaw,* "señor", que debían ser colocados arriba del símbolo principal en lugar de colocarlo debajo de éste.

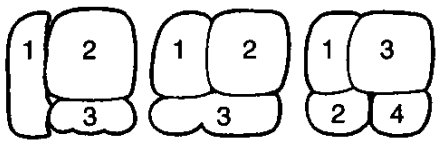

Figura 3. Orden de lectura para glifos compuestos. Tomada de Coe (2001)

La naturaleza logosilábica de la escritura maya permite que ésta sea extremadamente flexible, lo que permite que una misma palabra pueda ser escrita de múltiples formas sin que el significado o la lectura se vean afectados (Figura 4).

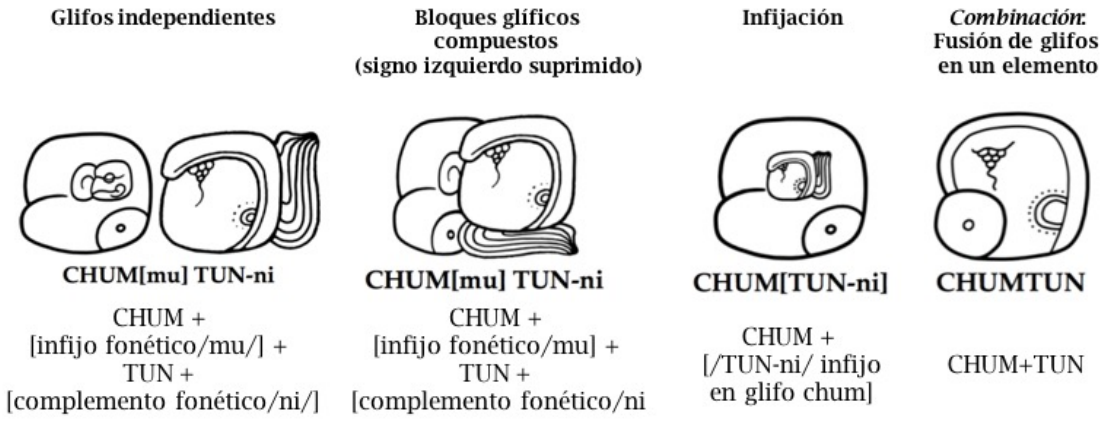

Figura 4. Composición glífica. Tomada de Kettunen y Helmke (2008)

Por motivos tanto de economía gráfica como de estética, cualquiera de estas combinaciones puede llegar a aparecer en un texto, incluso más de una de ellas al mismo tiempo. En algunas ocasiones por falta de espacio y en otras, se buscaba evitar repeticiones. Además de esto, es habitual encontrar dentro de un texto alternancia entre signos con valores fonéticos idénticos, lo que sirve como guía para la lectura de algunos signos.

2.2 Procesamiento digital de imágenes

El procesamiento digital de imágenes es un tema de mucho interés y últimamente se ha extendido para atender necesidades de diferentes disciplinas, desde las ciencias médicas hasta la industria agroalimentaria. Debido a que es un área grande, se puede dividir de la siguiente manera:

Realce y mejora: Son técnicas que se utilizan para mejorar la apariencia visual de las imágenes o para restaurar o recuperar imágenes degradadas.

Segmentación: Se encarga de la división de las imágenes en regiones o áreas significativas.

Extracción de características*:* Se encarga de la detección y localización de particularidades visuales que pueden ser líneas, puntos o regiones, las cuales permiten comparar una imagen con alguna otra.

2.2.1 Representación digital de una imagen

Una imagen puede ser vista como una función $f(x, y)$, donde (x, y) equivalen a las coordenadas espaciales de la imagen y el valor de la función, evaluada en

cualquier punto (x, y), representa la intensidad de luz de la imagen en esa posición.

Adicionalmente, una imagen digital es una imagen o función que ha sido discretizada en sus coordenadas espaciales y en sus valores de intensidad de luz. El resultado es un arreglo numérico tridimensional (para el caso de imágenes a color) donde cada dimensión representa uno de los 3 canales de color: rojo, verde y azul, mejor conocidos como RGB. Cada elemento del arreglo es llamado elemento de imagen o pixel el cual es un número real en un rango que va desde 0, que representa al color negro, hasta 255 que representa al color blanco.

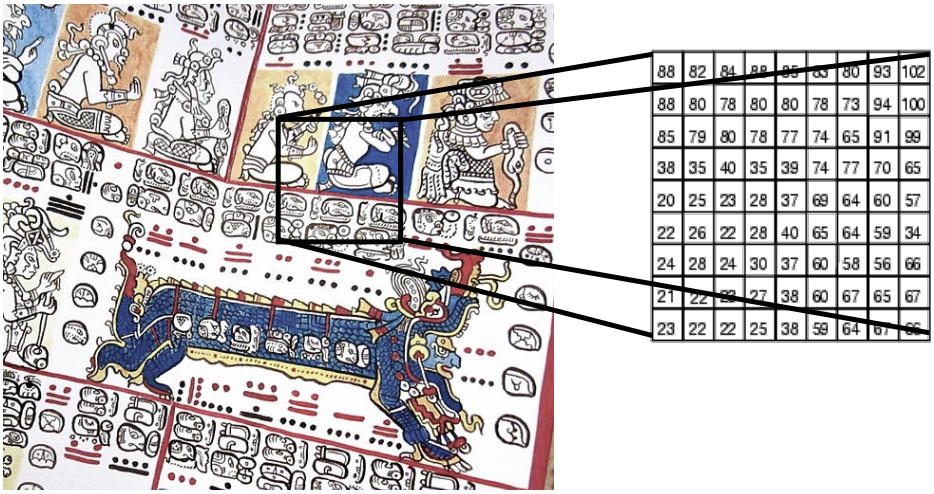

Figura 5. Ejemplo de representación de una imagen digital

Representación binaria de una imagen

Una imagen binaria es una imagen digital que únicamente tiene dos valores posibles para cada pixel. La binarización de una imagen consiste en un proceso de reducción de la información de la misma, en la que sólo persisten dos valores: encendido y apagado. En una imagen digital, estos valores de encendido y apagado, pueden representarse por los valores 0 y 1 o, más

frecuentemente, por los colores negro, valor de gris 0, y blanco, valor de gris 255.

0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0
0	0	1	1	1	1	0	0	0	0
0	0	1	1	1	1	1	0	0	0
0	0	1	1	1	1	1	0	0	0
0	0	1	1	1	1	1	0	0	0
0	0	0	1	1	1	1	1	1	0
0	0	0	0	0	0	1	1	1	0
0	0	0	0	0	0	1	1	1	0
0	0	0	0	0	0	0	0	0	0

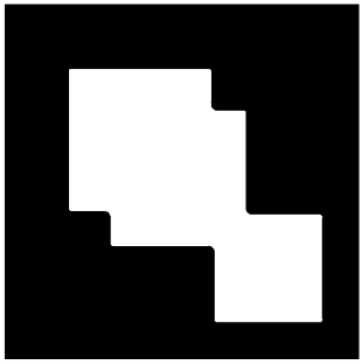

Figura 6. Representación binaria de una imagen

En el proceso y análisis de imagen, la binarización se emplea para separar las regiones u objetos de interés en una imagen del resto. Las imágenes binarias se usan en operaciones booleanas o lógicas para identificar individualmente objetos de interés o para crear máscaras sobre regiones.

2.2.2 Modelo de degradación de imagen

Existe una gran cantidad de transformaciones u operaciones que se pueden realizar sobre las imágenes con el propósito de mejorar su apariencia. Debido a que es un proceso subjetivo, realizado generalmente de una forma interactiva, la selección de los métodos apropiados y de los parámetros adecuados dependen de la calidad de la imagen original y de la aplicación final. Hay varios criterios para clasificar las técnicas de realce, pero por lo general se dividen en cuatro categorías:

Transformaciones puntuales: El resultado de una operación puntual depende sólo del nivel de gris de entrada de un punto. Las operaciones típicas puntuales incluyen la manipulación individual de los pixeles, por ejemplo, la binarización, el umbralado, entre otras.

Transformaciones locales: Aquí, los valores de entrada de varios pixeles vecinos influyen en el resultado del pixel de salida. Muchas operaciones son locales, por ejemplo, suavizado utilizando la media o el promedio y extracción bordes.

Transformaciones globales: El total de datos de la imagen como valor de entrada contribuye al resultado de salida. Las operaciones globales se realizan a menudo en el dominio de la frecuencia. Un ejemplo es la compresión de imágenes que tomando el total de una imagen entrada obtiene una imagen comprimida de salida.

Transformaciones geométricas: El resultado depende de las diferentes posiciones de los niveles de gris en la imagen de entrada. Ejemplos típicos son rotación, traslación, cambios de escala y rectificación.

Aunque las técnicas anteriores en su mayoría son empleadas con el objetivo de recuperar la calidad de imágenes también pueden ser utilizadas para modelar la degradación, aplicando el proceso en orden inverso. La Figura 7 muestra el modelo del proceso de degradación representado como una función H que, junto con la adición de ruido $\eta(x, y)$, opera sobre una imagen de entrada $f(x, y)$ para producir una imagen degradada denotada como $g(x, y)$.

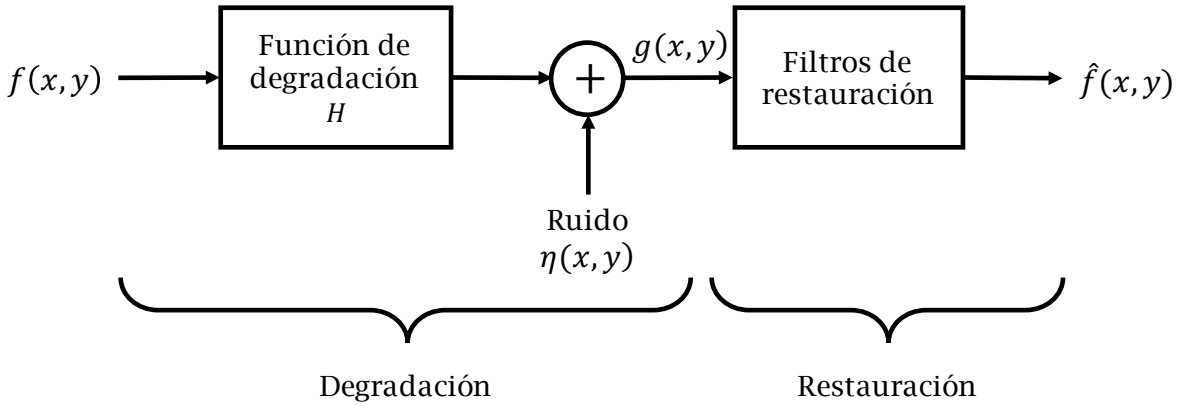

Figura 7. Modelo del proceso de degradación/restauración de imágenes. Basado en González y Woods (2006)

El objetivo del modelo es obtener una aproximación $\hat{f}(x,y)$ lo más semejante posible a la imagen de entrada original $f(x,y)$. Es por esto que mientras mejor se conozca la función de degradación H y el ruido $\eta(x,y)$, la función $\hat{f}(x,y)$ *más* se asemejará a $f(x,y)$.

2.2.3 Operaciones morfológicas

La morfología matemática está basada en operaciones de teoría de conjuntos, dentro del procesamiento digital de imágenes permite simplificar los datos de una imagen, preservar las características esenciales y eliminar aspectos irrelevantes. Algunos usos de la morfología matemática en el procesamiento de imágenes son:

- *Eliminación de ruido*
- *Simplificación de formas*
- *Extracción de esqueleto de un objeto*
- *Detección de objetos*
- *Extracción de bordes*

El conjunto de una imagen corresponde a todos aquellos puntos o pixeles que pertenecen a ella, donde cada elemento del conjunto corresponde a un pixel de la imagen, representado por arreglo de 2 dimensiones o tupla.

Las operaciones fundamentales en el procesamiento morfológico son la dilatación y la erosión. Ambas utilizan un elemento de estructura, o máscara binaria, para procesar la imagen.

Dilatación

En general, la dilatación aumenta el tamaño de un objeto. La cantidad y la forma en que aumenta el tamaño, depende de la elección del elemento de estructura. Dada una imagen A, y un elemento estructural B en un espacio Z^2, la dilatación de A por B se define como:

$$A \oplus B = \{z|\ (\widehat{B})_z \cap A \neq \emptyset\}$$

Esta ecuación consiste en obtener la reflexión de B sobre su origen y trasladar esta reflexión por z. La dilatación de A por B es entonces el conjunto de todos los desplazamientos, z, tal que la reflexión de B y A se solapan por al menos un elemento.

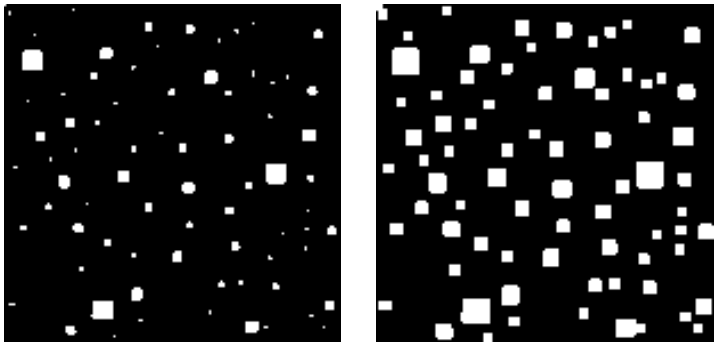

Figura 8. Ejemplo de dilatación.

Erosión

La erosión es la operación opuesta a la dilatación y generalmente, disminuye el tamaño de los objetos. Al igual que en la dilatación, la cantidad y la forma en que se produce esta disminución depende del elemento de estructura elegido. Sean A y B conjuntos en Z2, la erosión de A por B se define como:

$$A \ominus B = \{z|\ (B)_z \subseteq A\}$$

La ecuación de erosión indica que la erosión de A por B, es el conjunto de todos los puntos z tales que B, trasladado por z, está contenido en A. Uno de los usos más simples de la erosión es para la eliminación de detalles irrelevantes, en términos de tamaño, de una imagen binaria.

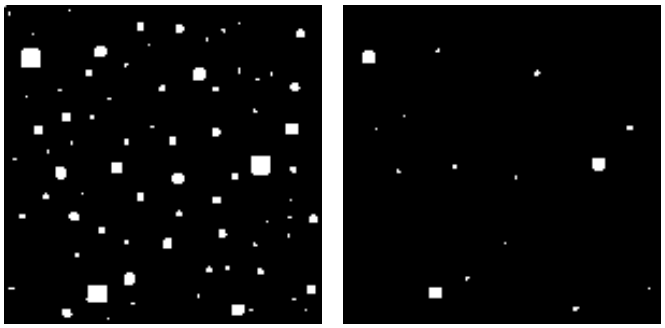

Figura 9. Ejemplo de erosión. Basada en González y Woods (2006)

2.2.4 Modelos de ruido

Las imágenes digitales son vulnerables a una gran cantidad de tipos de ruido. El ruido en una imagen es el resultado de errores en la adquisición o en la transmisión debido a las condiciones del medio entre otras variantes que pueden afectar a los sensores de los aparatos. Esto ocasiona que los valores finales de los pixeles no reflejen la intensidad de color de la escena real de la

imagen. Dependiendo del tipo de perturbación, el ruido puede afectar a una imagen de diferentes maneras.

Ruido Gaussiano

El ruido Gaussiano produce pequeñas variaciones en la imagen debido a las alteraciones en los componentes electrónicos. Es un espectro de energía constante para todas las frecuencias y esto hace que afecte la intensidad de los pixeles de toda la imagen. En el ruido de tipo gaussiano, todos y cada uno de los pixeles que componen la imagen cambian su valor, de acuerdo con una distribución normal o gaussiana.

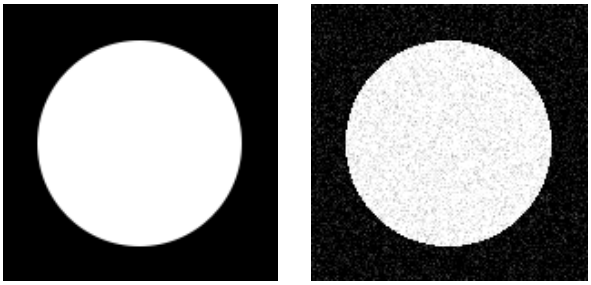

Figura 10. Ejemplo de ruido Gaussiano.

Ruido sal y pimienta

El hecho que define este tipo de ruido es que el pixel ruidoso en cuestión no tiene relación alguna con los pixeles que lo rodean. Generalmente, este tipo de ruido, afectará únicamente a una pequeña cantidad de pixeles de la imagen. La imagen resultante tendrá algunos pixeles blancos sobre otros negros y viceversa, de ahí el término sal y pimienta.

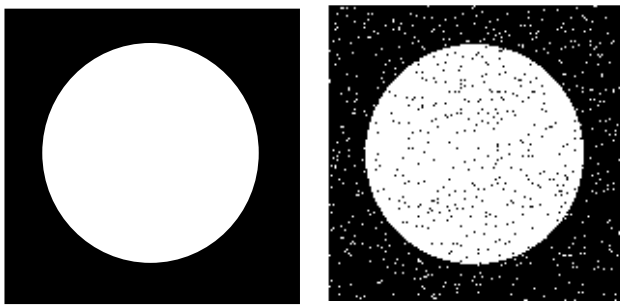

Figura 11. Ejemplo de ruido sal y pimienta. Basada en González y Woods (2006)

2.2.5 Suavizado de imágenes

Los operadores de suavizado se utilizan para eliminar o difuminar el ruido. Por ejemplo, si el ruido de una imagen es un fenómeno de alta frecuencia se puede reducir con filtros pasa bajos.

Filtros Gaussianos

Son una familia de filtros lineales con los pesos determinados de acuerdo a la forma de una función gaussiana. Son muy utilizados para la detección de bordes además de la eliminación de ruido gaussiano.

2.2.6 Métodos para transformaciones geométricas

La mayoría de las transformaciones geométricas dan lugar a pixeles que no coinciden con la posición de los de la imagen original (ver sección 2.2.2). El problema es asignar el nivel adecuado de gris existente en la imagen original o interpolando a partir de los pixeles vecinos. Algunos ejemplos de la interpolación de imágenes son el escalamiento, rotación, cambio de perspectiva y deformaciones o transformaciones no lineales. Existen varias

alternativas para interpolar una imagen, las más comunes son: interpolación por vecino más cercano e interpolación bilineal.

Interpolación por vecino más cercano

Este método consiste en asignar el nivel digital de un pixel de la imagen transformada a un pixel de la imagen corregida cuyo centro geométrico esté más cercano a su homólogo en la imagen transformada. Para ello se calculan las distancias entre el centro del pixel de la imagen corregida de coordenadas *(m, n),* y los centros de los cuatro pixeles más cercanos en la imagen transformada, tomando el nivel digital correspondiente a aquel pixel que proporcione la menor distancia.

Interpolación bilineal

En este método se consideran los niveles de gris de los cuatro pixeles más cercanos en la imagen transformada al pixel de la imagen corregida. Primero, se tiene en cuenta los valores en los pixeles conocidos que rodean a uno dado en una vecindad de los 2x2 pixeles más cercanos. Se toma el promedio ponderado de estos 4 pixeles y se calcula el valor interpolado. El resultado está más suavizado que las imágenes obtenidas utilizando el método del pixel más cercano.

2.3 Clasificación de imágenes

La clasificación de glifos mayas puede verse como un caso específico de clasificación de imágenes, ya que a partir de la imagen de un glifo determinado el objetivo es catalogarlo asignándole una etiqueta la cual es el identificador de Thompson.

El problema de clasificación de imágenes consiste básicamente en asignar una etiqueta de un conjunto determinado de categorías a una imagen de entrada dada. Aunque el ser humano es capaz de resolver esta tarea con relativa facilidad, es uno de los principales problemas en el área de visión por computadora y tiene una gran variedad de aplicaciones en diversas áreas.

Todos los algoritmos de clasificación se basan en el supuesto de que la imagen en cuestión representa una o más características y que cada una de ellas pertenece a una de varias clases distintas. Las clases pueden ser especificadas a priori, clasificación supervisada, o agrupadas automáticamente, clasificación no supervisada.

Los algoritmos de clasificación suelen emplear dos fases de procesamiento: entrenamiento y prueba. En la fase de entrenamiento, se aíslan las propiedades características de las imágenes y, a partir de ellas, se crea una descripción única de cada categoría, de manera que sea fácil distinguir una categoría de otra. Posteriormente, en la etapa de prueba, ese espacio de propiedades obtenidas en el entrenamiento, se utiliza para clasificar los nuevos elementos.

A pesar de que la tarea de reconocer objetos es relativamente fácil para los seres humanos aún es un reto desde la perspectiva de visión por computadora debido a la manera en que una computadora interpreta una imagen. Retomando el concepto de representación digital de una imagen, una imagen para la computadora es un arreglo multidimensional de números reales y por este motivo, existen diferentes factores que pueden afectar el funcionamiento del modelo de clasificación como la orientación de los objetos representados en las imágenes, variaciones de tamaño, deformaciones, oclusión de los objetos, mala iluminación en las imágenes, entre otros.

2.3.1 Clasificación por vecinos más cercanos

El método de clasificación por vecinos más cercanos o mejor conocido como KNN, es la aproximación más sencilla de un sistema de clasificación de imágenes. Es uno de los métodos más sencillos para y brinda una idea general acerca del problema de clasificación de imágenes.

La fase de entrenamiento del algoritmo consiste en almacenar los vectores de características y las etiquetas de clase de cada uno de los ejemplos de entrenamiento. El objetivo es etiquetar de manera automática, imágenes que no estuvieron presentes en el entrenamiento.

En la etapa de prueba, cada vector característico de las imágenes del conjunto de prueba es comparado uno a uno con los presentes en el entrenamiento, calculando la distancia entre ellos. Comúnmente se utiliza la distancia euclidiana, aunque existen muchas otras que pueden emplearse, dependiendo del tipo de problema. Posteriormente, se seleccionan los k arreglos más cercanos al ejemplo en cuestión y finalmente es clasificado con la etiqueta de clase que más se repite en los seleccionados.

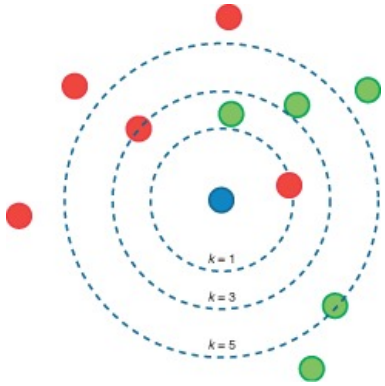

Figura 12. Ilustración de un modelo de clasificación KNN.*

**Machine learning methods in chemoinformatics - Scientific Figure on ResearchGate. Recuperado de: https://www.researchgate.net/260436143_fig2_Figure-3-Illustration-of-a-kNN- classification-model-For-k1-the-model-will-classify*

Este método supone que los vecinos más cercanos proporcionan la mejor clasificación utilizando todos los atributos; el problema de dicha suposición es la probabilidad que existan muchos atributos irrelevantes que dominen sobre los que sí son relevantes provocando que la exactitud de este algoritmo sea afectada por la presencia de ruido o características irrelevantes.

Una posible solución puede ser asignar pesos a las distancias de cada atributo, dándole mayor importancia a los atributos relevantes. Otra posibilidad consiste en identificar y eliminar los atributos que se consideren irrelevantes.

La elección de la cantidad de los k vecinos depende fundamentalmente de los datos; generalmente, valores grandes de k reducen el efecto de ruido en la clasificación, pero crean límites entre clases parecidas. Un buen k puede ser seleccionado mediante una optimización de uso.

2.3.2 Validación cruzada

En un problema de aprendizaje supervisado, al modelo se le proporciona un conjunto de datos con el que es ejecutado el entrenamiento y otro conjunto de datos desconocidos con el cual el modelo es probado. La validación cruzada es una técnica utilizada para evaluar los resultados estadísticos de la precisión del modelo y asegurar que dichos resultados son independientes de la partición de datos de entrenamiento.

Consiste en dividir los conjuntos de datos en K subconjuntos. Uno de los subconjuntos se utiliza como datos de prueba y el resto (K-1) como datos de entrenamiento. El proceso de validación cruzada es repetido durante k iteraciones, con cada uno de los posibles subconjuntos de datos de prueba.

Finalmente, se realiza la media aritmética de los resultados de cada iteración para obtener un único resultado.

Entrenamiento	Entrenamiento	Entrenamiento	Entrenamiento	Prueba
Entrenamiento	Entrenamiento	Entrenamiento	Prueba	Entrenamiento
Entrenamiento	Entrenamiento	Prueba	Entrenamiento	Entrenamiento
Entrenamiento	Prueba	Entrenamiento	Entrenamiento	Entrenamiento
Prueba	Entrenamiento	Entrenamiento	Entrenamiento	Entrenamiento

Figura 13. Esquema de validación cruzada de 5 iteraciones.

Este método es más preciso que una sola evaluación, puesto que se evalúa a partir de K combinaciones de datos de entrenamiento y de prueba. En la práctica, la elección del número de iteraciones depende de la medida del conjunto de datos. Lo más común es utilizar la validación cruzada de 10 iteraciones.

2.3.3 Redes neuronales artificiales

El uso de redes neuronales artificiales o RNAs en el campo del procesamiento de imágenes es muy favorable; la imagen digitalizada original o pre procesada se utiliza como entrada de una red, obteniéndose como salida una clasificación o identificación de la imagen o bien, de algunos segmentos de ella.

Las RNAs están inspiradas originalmente en el objetivo de modelar sistemas neuronales biológicos, pero desde un inicio se han diferenciado entre sí y se

han convertido en una cuestión de ingeniería y de lograr buenos resultados en tareas de aprendizaje automático.

Aunque las redes neurales artificiales no se acercan a la complejidad del cerebro, existen dos similitudes esenciales entre las redes neuronales biológicas y las redes neuronales artificiales. En primer lugar, los bloques de construcción de ambas redes son dispositivos simples que se encuentran altamente interconectados, aunque en la realidad, las neuronas artificiales son mucho más simples que las neuronas biológicas. En segundo lugar, las conexiones existentes entre las neuronas determinan la función de la red.

Neurona artificial

La neurona artificial es una unidad básica de procesamiento en una red neuronal (Figura 14).

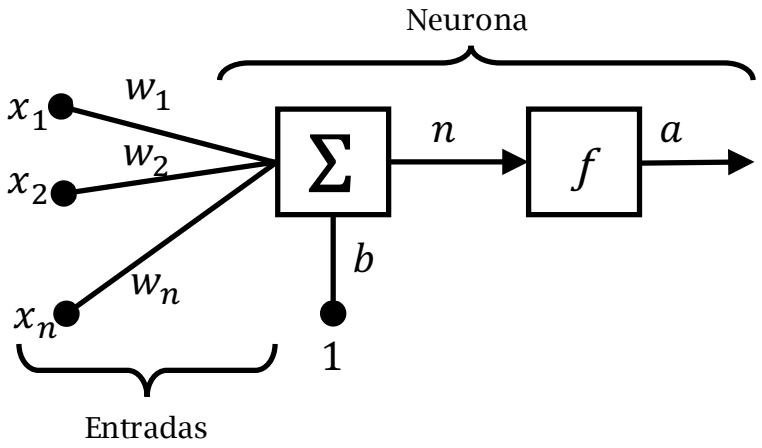

Figura 14. Esquema de una neurona artificial. Basado en Demuth, Beale, Jess, y Hagan (2014)

El funcionamiento de una neurona artificial comienza con un receptor, donde llegan una o varias señales de entrada x, generalmente provienen de otras neuronas y cada una de ellas son atenuadas o amplificadas con un factor de

peso w. Los valores son enviados a elemento sumador, donde se efectúa la suma algebraica de las señales de entrada ponderándolas de acuerdo a su peso. La otra entrada es multiplicada por un umbral y también es enviado al elemento sumador aplicando la siguiente expresión:

$$n = \sum x_i w_i + b$$

Que también puede ser expresada en forma matricial y puede interpretarse como el producto escalar de los vectores de entrada $\mathbf{X}$ y los pesos $\mathbf{w}$.

$$n = \mathbf{X}\mathbf{w} + b$$

Finalmente, la salida n del sumador es enviada al elemento de función de activación o de transferencia f, que aplica una función no lineal a la salida del sumador que produce la salida a de la neurona. La salida de una neurona es calculada entonces de la forma:

$$a = f(\mathbf{X}\mathbf{w} + b)$$

Los valores w, b son parámetros ajustables de la neurona. Por lo general, la función de activación es elegida por el diseñador y luego los parámetros w, b son ajustados por alguna regla de aprendizaje para que la relación entre la entrada y salida de la neurona cumpla con algún objetivo específico.

El modelo neuronal presentado es el utilizado en casi todas las redes neuronales artificiales, variando únicamente el tipo de función activadora. La elección de la función de transferencia es importante ya que de ella depende la salida obtenida.

Funciones de activación o transferencia

Las neuronas artificiales tienen diferentes estados de activación; algunas de ellas solamente dos: encendido y apagado (similar a las neuronas biológicas), pero otras pueden tomar cualquier valor dentro de un conjunto determinado.

La función de activación calcula el estado de actividad de una neurona; transformando la entrada global en un valor o estado de activación, cuyo rango normalmente va de 0 a 1 o bien, de –1 a 1.

La función de transferencia puede ser una función lineal o no lineal (

Tabla 1). Una función de transferencia particular se elige para satisfacer alguna especificación del problema que la neurona está intentando resolver.

Nombre	Función
Umbral	$a = 0 \quad n < 0$ $a = 1 \quad n \geq 0$
Escalón	$a = -1 \quad n < 0$ $a = 0 \quad n = 0$ $a = 1 \quad n > 0$
Lineal	$a = n$
Lineal acotada	$a = 0 \quad n < 0$ $a = \mathrm{n} \quad 0 \leq \mathrm{n} \leq 1$ $a = 1 \quad n > 1$
Lineal acotada simétrica	$a = -1 \quad n < -1$ $a = \mathrm{n} \quad -1 \leq n \leq 1$ $a = 1 \quad n > 1$
Sigmoide	$a = \frac{1}{1 + e^{-n}}$
Unidad lineal rectificada (ReLU)	$a = 0 \quad n < 0$ $a = n \quad n \geq 0$

Tabla 1. Funciones de activación más comunes

Niveles o capas de una red neuronal

La distribución de neuronas dentro de la red se realiza formando niveles o capas, con un número determinado de neuronas en cada una de ellas. Se pueden distinguir tres tipos de capas (Figura 15):

- De entrada: Capa que recibe directamente la información proveniente de las fuentes externas de la red.
- Ocultas: Son internas a la red y no tienen contacto directo con el entorno exterior. Las neuronas de las capas ocultas pueden estar interconectadas de distintas maneras, lo que determina las distintas topologías de redes neuronales.

- De salida: Transfieren información de la red hacia el exterior.

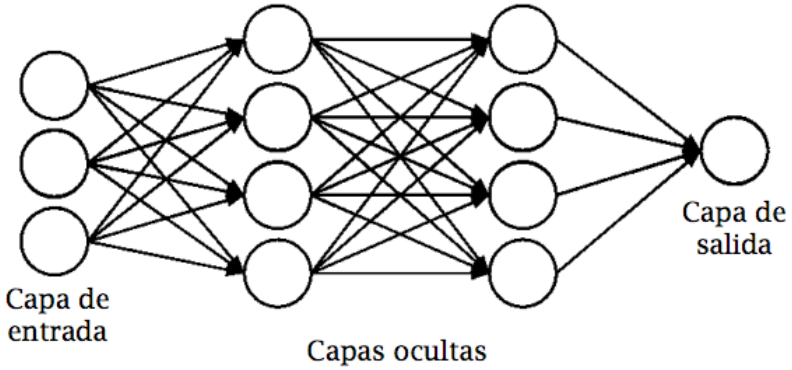

Figura 15. Ejemplo de red neuronal con 2 capas ocultas

Aquí se presenta un ejemplo de topología de red neuronal que utiliza capas totalmente conectadas, en la que cada nodo o neurona únicamente está conectada con neuronas de un nivel superior. Hay que notar que hay más conexiones que neuronas en sí; en este sentido, se dice que una red es totalmente conectada si todas las salidas desde un nivel llegan a todos y cada uno de los nodos del nivel siguiente.

Aprendizaje con redes neuronales

El proceso de aprendizaje se puede dividir en tres grupos de acuerdo a sus características:

- Supervisado: Se presenta a la red un conjunto de patrones de entrada y su salida esperada. Los pesos se van modificando de forma proporcional al error que se produce entre la salida de la red y la salida esperada.
- No supervisado: Se presenta a la red un conjunto de patrones de entrada sin información sobre la salida esperada. El proceso de entrenamiento

en este caso deberá ajustar sus pesos en base a la relación existente entre los datos de entrada.

- Por refuerzo: Este tipo de aprendizaje se ubica entre medio de los dos anteriores. Se le presenta a la red un conjunto de patrones de entrada y se le indica a la red si la salida obtenida es o no correcta. Sin embargo, no se le proporciona el valor de la salida esperada. Este tipo de aprendizaje es muy útil en aquellos casos en que se desconoce cuál es la salida exacta que debe proporcionar la red.

2.3.1 Redes neuronales convolucionales

Una red neuronal convolucional o CNN está formada por colecciones de neuronas organizadas estructuradamente mediante capas. La entrada a la red neuronal convolucional es una imagen y a su salida se obtiene la puntuación de cada una de las clases sobre cada uno de los pixeles de la imagen de entrada.

Las redes neuronales convolucionales consisten en múltiples capas con distintos propósitos. Al principio se encuentra la fase de extracción de características, compuesta de neuronas convolucionales y de reducción de muestreo. Al final de la red se encuentran neuronas que realizan la clasificación final sobre las características extraídas previamente.

En las redes neuronales simples descritas en la sección anterior, cada neurona estaba completamente conectada a cada una de las neuronas de la siguiente capa, es decir, cada neurona de la capa oculta calcula una función que depende de los valores de cada nodo en la capa de entrada.

La etapa de extracción de características tiene el efecto de filtrar la imagen de entrada. Esto transforma los datos de tal manera que ciertas características se vuelven más dominantes en la imagen de salida al tener un valor numérico más alto asignados a los pixeles que las representan. Los diferentes filtros tienen habilidades de procesamiento de imágenes específicas, como por ejemplo la detección de bordes que se puede realizar con filtros que resaltan el gradiente en una dirección en particular.

Una característica que distingue a las CNNs de las redes neuronales simples es el hecho de que los pesos de la red se comparten entre diferentes neuronas dentro de las capas ocultas. Recordando que cada neurona en la red calcula primero una combinación ponderada de sus entradas, se puede ver este proceso como la evaluación de un filtro lineal sobre los valores de entrada. En este contexto, el intercambio de los pesos a través de múltiples neuronas en una capa oculta se traduce en la evaluación del mismo filtro sobre varias ventanas sub de la imagen de entrada.

En este sentido, se puede ver la CNN como el aprendizaje más efectivo a un conjunto de filtros. Utilizando el mismo conjunto de filtros sobre la imagen entera obliga a la red para aprender una codificación general o representación de los rasgos característicos. Finalmente, el elemento distintivo en una CNN es la presencia de una capa de submuestreo o pooling. El objetivo es reducir la dimensión de las respuestas convolucionales y además añadir un pequeño grado de invariancia a traslaciones en el modelo.

En una CNN típica, se dispone de múltiples capas, alternando capas de convolución, pooling y normalización pudiendo llegar a construir una arquitectura profunda multicapa (

Figura 16).

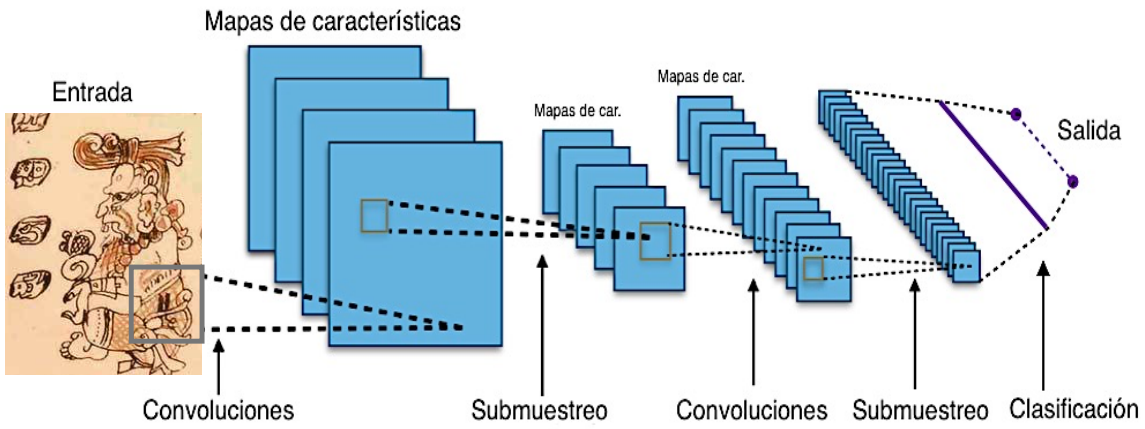

Figura 16. Arquitectura de una red neuronal convolucional

3

Trabajo relacionado

Este capítulo está dividido en tres secciones generales: Primero se resumen los trabajos que se han dedicado a resolver tareas de clasificación de diferentes tipos de caracteres. En la segunda parte, se habla sobre los trabajos que, aunque no pertenecen al área de computación, proveen información relevante acerca de la manera en que los glifos han sido catalogados a lo largo del tiempo en que estos han sido investigados. Finalmente se exponen los trabajos en el área de computación que se han dedicado a trabajar con glifos mayas y su clasificación automática, además se explican las técnicas que utilizaron y los resultados que obtuvieron.

3.1 Clasificación de caracteres

La clasificación de caracteres en general puede considerarse una tarea similar a la del reconocimiento de glifos mayas. En esta sección se presentan algunos trabajos que se considera representan el estado del arte y que sirvieron como base para el desarrollo de este trabajo de tesis.

3.1.1 Reconocimiento de caracteres manuscritos

En los últimos años, la investigación en el área de dígitos escritos a mano ha avanzado gracias a la utilización de redes neuronales convolucionales. Chellapilla, Larson, Simard, and Czerwinski (2005) publicaron un artículo donde aseguran vencer al ser humano en tener una tasa de error menor en el reconocimiento de los 10,000 ejemplos de prueba que conforman la base de datos MNIST, LeCun (1998). En este trabajo utilizan una red neuronal convolucional CNN, para producir una solución estadística de fuerza bruta.

Sin embargo, al igual que con cualquier algoritmo de aprendizaje automático, los resultados sólo pueden ser tan buenos como los datos de entrenamiento con los que se enseñan. La base de datos MNIST, tiene 60,000 ejemplos de entrenamiento, y otros 10,000 ejemplos de prueba. Esta cantidad se incrementó a través del desarrollo de un método para expandir los datos de entrenamiento aplicando una serie de transformaciones y distorsiones.

Con esas dos etapas de pre-procesamiento consiguieron aumentar el tamaño y la variación del conjunto de datos, lo que mejoró considerablemente la precisión del sistema. En lugar de intentar producir una representación de un símbolo, se pueden generar tantos ejemplos posibles de variantes de un símbolo, y luego utilizar esas variantes como datos de entrenamiento.

3.1.2 Clasificación de símbolos antiguos

Existe una cantidad muy pequeña de trabajos acerca del reconocimiento de símbolos de lenguas antiguas y los enfoques utilizados para resolver estas tareas han sido en realidad más simples que los empleados para los caracteres de lenguas modernas.

Ntzios, Gatos, Pratikakis, Konidaris, y Perantonis (2005), describen un sistema de reconocimiento óptico de caracteres griegos manuscritos, basándose en la detección de características de lo que ellos llaman “cavidades abiertas” y “cavidades cerradas”. El porcentaje de reconocimiento de caracteres que alcanzan con su propuesta es del 68%.

Arrivault, Richard, Fernandez-Maloigne, y Bouyer (2005),presentan una versión más refinada del trabajo mencionado previamente, en el cual utilizan un enfoque estadístico basado en descriptores de los símbolos combinado con un método de clasificación estructural con grafos que representan la descripción de los atributos de los caracteres. A pesar de los esfuerzos en mejorar la tasa de clasificación, no se muestra mejora alguna al combinar los dos enfoques que proponen. En estos trabajos, el porcentaje de reconocimiento de caracteres griegos es baja y resulta poco útil para ser utilizada en el área de glifos mayas debido principalmente a que la escritura griega es más bien alfabética.

Otro caso de reconocimiento de sistemas de escritura antigua es el caso de los jeroglíficos egipcios. Franken y van Gemert (2013) se encargan de recolectar y segmentar manualmente una colección de casi 4,000 jeroglíficos egipcios. Posteriormente, segmentan manual y automáticamente cada símbolo y asignan un orden de lectura a cada uno. Para la tarea específica de clasificación utilizan 5 descriptores diferentes obteniendo una tasa de reconocimiento del 85%.

3.2 Catálogos de glifos mayas

Para formalizar el estudio de un texto desconocido es necesario hacer uso de signos estandarizados. Para que la transcripción sea estandarizada se necesita crear un catálogo de signos con todas sus variantes y alógrafos. (Knórozov et al. (1999).

Aunque dentro del área de estudio de la escritura maya existen diferentes catálogos de glifos, para propósitos de este trabajo utilizaremos los dos más importantes y en los que se ha basado la mayoría de investigaciones de escritura maya.

3.2.1 Catálogo de Thompson

Thompson estableció lo que es considerado por diversos investigadores como el catálogo de glifos mayas más importante debido a que es el más completo que ha sido compilado. Tras su publicación en 1962, adquirió gran relevancia entre los especialistas, especialmente durante las dos décadas después de su divulgación; fue una herramienta crítica durante aquel periodo debido a que es una herramienta bastante buena para la identificación de signos; adicionalmente, existía una mínima cantidad de signos se podían identificar y leer con certeza y permitía a los mayistas referirse fácilmente a algún signo por medio de su identificador de Thompson.

Este catálogo recopila glifos pertenecientes a los tres códices, monumentos y vasijas con alrededor de 10,000 ejemplos. Contiene 862 signos divididos en 3 categorías: 370 afijos, 356 elementos principales y 87 formas antropomorfas denominadas retratos o cabezas.

Aunque existen otros catálogos de que han sido compilados y publicados, el de Thompson ha sido el más empleado y comúnmente el más referido.

3.2.2 Diccionario de Montgomery

John Montgomery presentó el primer diccionario visual de glifos mayas publicado desde el desciframiento completo de la escritura. Ofrece un catálogo con más de 1,200 glifos simples y compuestos que se encuentra organizado alfabéticamente por palabras, frases o sílabas.

Cada una de las entradas en el diccionario contiene elementos visuales de cada glifo, su transcripción fonética, índices de referencia para el inglés y español, su equivalente en lengua maya, la función gramatical y su significado correspondiente.

Adicionalmente, el diccionario incorpora los identificadores del catálogo de Thompson para la mayoría de glifos individuales e incluso, para el caso de los glifos compuestos, señala a partir de cuales símbolos individuales está formado cada uno de ellos.

3.3 La computación en el área de los glifos mayas

Desde hace ya varios años, han habido algunos intentos para introducir el uso de herramientas computacionales en el área de la investigación de la escritura maya.

El primer registro de dicha tentativa, fue el trabajo de Ringle y Smith-Stark (1983) quienes hicieron una propuesta de enfoques computacionales para el

desciframiento de la escritura maya. Aunque su trabajo no logró el objetivo principal, encontraron que la cantidad de glifos que hasta ese momento habían sido reconocidos era de 950 y no 862, como propuso Thompson.

Sidorov, Pichardo-Lagunas, y Chanona-Hernandez (2009) propusieron la creación de una base de datos donde se encuentran almacenados todos los signos pertenecientes al diccionario de Montgomery. Con el propósito de brindar un sistema que permitiera utilizar y administrar un sistema de base datos de la escritura maya útil para los especialistas en el área e incluso para aquellas personas sin conocimientos previos del tema. Basado en el diccionario de Montgomery, ofrece imágenes e información detallada de cada uno de los glifos, facilita el proceso de búsqueda y clasificación de los glifos de la escritura maya.

En cuanto al área de clasificación para la escritura maya, el primer intento por resolver la tarea automáticamente fue el trabajo de E. Roman-Rangel, Pallan, Odobez, y Gatica-Perez (2009). Su trabajo consiste en la implementación de un descriptor de forma para el reconocimiento de objetos en una imagen.

Utilizando un conjunto de datos de imágenes y dibujos de alta resolución de todos los monumentos mayas dentro del territorio mexicano, su metodología comienza con un procedimiento para la extracción de bordes, posteriormente se selecciona un conjunto de puntos mínimo que represente a dicho contorno. El conjunto de puntos seleccionados es descrito en forma de histogramas y para cada punto se calcula una distribución dependiendo de la posición en la que se encuentre con respecto de los demás en el conjunto, haciendo que el descriptor sea sensible a los puntos más cercanos entre sí. Mientras se tenga un mayor número de puntos que representan a un borde, mayor precisión tendrá el descriptor obtenido.

Sus experimentos fueron realizados con glifos de 22 clases diferentes, en su mayoría signos silábicos. La manera en que realizan la parte de la clasificación es muy sencilla. los descriptores de cada glifo son organizados dentro de una base de datos junto con su identificador de Thompson y es realizada una consulta con base en una medida de similitud establecida, obteniendo los *n* glifos más cercanos al que se está comparando.

Los resultados que obtuvieron son con una precisión del 32% para la clasificación. Tras la evaluación del descriptor, concluyen que el descriptor es eficaz en varios casos, pero aún enfrenta desafíos significativos debido a la complejidad y variedad de los glifos mayas.

Edgar Roman-Rangel, Pallan Gayol, Odobez, y Gatica-Perez (2011), continúan su investigación y proponen un descriptor mejorado. Basándose en el descriptor de contexto que fue descrito en el trabajo anterior, incorporan algunas mejoras: se añaden variaciones en el grosor de los contornos, debido a que existen variaciones en los trazos de los glifos que al ser discretizados a un conjunto de puntos generan pérdida de información relevante; se añade más información del glifo además de los contornos al conjunto de puntos, lo que enriquece la descripción y finalmente se computan descriptores únicamente para un subconjunto uniformemente muestreado del conjunto original. Cada punto obtenido es llamado pivote y está descrito utilizando información de todos los puntos presentes en el conjunto inicial.

Con esta nueva propuesta, los autores, lo gran aumentar la precisión de la clasificación de 24 clases de glifos mayas en un 20%, comparado con el descriptor que habían propuesto inicialmente, llegando a un 53.8% en el mejor de los casos. Los resultados que presentan aún presentan algunas fallas en

algunos casos específicos, especialmente considerando que el descriptor propuesto es muy sensible a imágenes que presentan reflexión.

Finalmente, Hu, Gayol, Krempel, Odobez, y Gatica-Perez (2014) proponen un método de recuperación automática de glifos mayas que integra la información de la forma del símbolo y además de información de contexto. La información espacial, sigue la pauta de los descriptores anteriores, mientras que la información de contexto es añadida convirtiendo cada bloque de glifos en una cadena de Markov de primer orden, la cual es una serie de eventos, en la cual la probabilidad de que ocurra un evento depende del evento inmediato anterior. Las cadenas de este tipo recuerdan el último estado o evento ocurrido lo que condiciona las posibilidades de los eventos futuros; además es aplicado un modelo de co-ocurrencia para cada bloque de glifos.

A diferencia de los experimentos de clasificación descritos en los trabajos previos, en este caso la clasificación de un glifo *x,* considera todos los glifos observados en la cadena, con el efecto de que un glifo similar que normalmente co-ocurre con glifos que son visualmente probables en las posiciones vecinas recibirá una probabilidad mayor. Esto fue implementado utilizando el algoritmo de Viterbi, con 25 clases de glifos diferentes el porcentaje de precisión obtenido supera ligeramente el 65% en el mejor de los casos.

Los trabajos expuestos en esta sección demuestran que aún hace falta explorar más técnicas para el reconocimiento automático de glifos mayas, ya que aún existen deficiencias para poder resolver la tarea. En primer lugar, los experimentos fueron realizados principalmente con afijos sin tomar en consideración los retratos y símbolos principales. Otro punto a destacar es

que la cantidad de clases que han sido consideradas para la clasificación es muy pequeña (25 clases) en comparación con la cantidad de glifos que se conocen y que han sido catalogados hasta ahora (más de 1,000). Finalmente, los trabajos descritos únicamente abordan el problema desde la perspectiva de la descripción de glifos, dejando en segundo plano los algoritmos de clasificación.

Referencia	% Exactitud
Roman-Rangel, E., Pallan, C., Odobez, J. M., & Gatica-Perez, D. (2009). *Retrieving ancient Maya glyphs with Shape Context.*	32%
Roman-Rangel, E., Pallan Gayol, C., Odobez, J. M., & Gatica-Perez, D. (2011). *Searching the past: an improved shape descriptor to retrieve maya hieroglyphs.*	53.8%
Hu, R., Gayol, C., Krempel, G., Odobez, J., & Gatica-Perez, D. (2014). *Automatic Maya hieroglyph retrieval using shape and context information.*	65%

Tabla 2. Resumen de resultados de trabajos relacionados

4

Desarrollo

En los capítulos anteriores, se ha revisado la teoría que concierne al modelo de degradación de imágenes para conseguir aproximación de imágenes ruidosas a imágenes reales; el problema de clasificación de imágenes, así como dos algoritmos utilizados para solucionar esta tarea, KNN y redes neuronales. También se han explicado algunas aplicaciones existentes en el área del reconocimiento de símbolos y los trabajos especializados en el reconocimiento de glifos mayas.

Los trabajos que han incursionado en el área del reconocimiento de caracteres cuentan con una vasta cantidad de ejemplos para lograr resultados óptimos al entrenar modelos de aprendizaje supervisado y modelos de aprendizaje profundo, lo que sitúa al trabajo del reconocimiento de glifos mayas como un problema difícil de resolver. La metodología propuesta a continuación pretende brindar una solución a la escasez de datos disponibles para el caso de los glifos mayas.

4.1 Corpus

La tarea de seleccionar fuentes que fueran de utilidad para componer el corpus no fue sencilla por diversas razones. En primer lugar, la cantidad de

datos digitalizados acerca de glifos mayas es limitada, debido a que muchos de los escritos que aún sobreviven han perdido su calidad visual con el paso del tiempo, los que han sido segmentados de las inscripciones se encuentran parcialmente ausentes debido a la oclusión. En segundo lugar, a pesar de que se han contabilizado más de 1000 signos diferentes, muchos de ellos solo cuentan con uno o muy pocos ejemplos.

Es esencial contar con un corpus representativo si no en tamaño, sí al menos en variedad de símbolos, y que además se adecúe a la finalidad del trabajo, que se cuente con la transcripción de los símbolos en él y que además los datos contenidos en él sean mayormente aprobados por los expertos en el dominio.

4.1.1 Descripción del corpus

El corpus utilizado para el desarrollo de este trabajo está conformado por datos de tres fuentes diferentes que se ilustran en la tabla 1.

Fuente principal	Catálogo de Thompson	Diccionario de Montgomery	Códices
Conjunto de datos	Mayan epigraphic database	Diccionario de los glifos mayas con descripción visual	Maya codex dataset

Tabla 3. Conjuntos de datos

The Mayan epigraphic database (1998)

Este conjunto de datos está basado en el diccionario de Thompson (1962), cuenta con imágenes de cada uno de los glifos y su valor numérico asociado;

el valor numérico asociado es único para cada glifo y a su vez está relacionado con el identificador del catálogo de Thompson.

Cuenta con un total de 1,104 imágenes pertenecientes al catálogo de Thompson, 450 de ellas son símbolos principales.

Diccionario de los glifos mayas con descripción visual estructural (2009)

Base de datos relacional que contiene glifos de la escritura maya, organizada por la escritura fonética de las palabras, frases o silabas. Incluye información general de cada glifo como su imagen, traducción, transcripción, asociación por medio del número de Thompson, referencia fonética y algunas notas descriptivas.

Debido a que este sistema está basado principalmente en el diccionario de Montgomery (2002), cuenta con glifos simples y compuestos, tiene un total 1,241 registros, de los cuales 192 son símbolos principales.

4.1.2Etiquetado del corpus

Aunque todos los corpus se encuentran etiquetados, fue necesario realizar una re-etiquetación manual para encontrar los símbolos que fueran equivalentes dentro de los tres conjuntos con el motivo de reunir más de una instancia por símbolo principal.

Figura 17. Equivalencias de símbolos en los conjuntos de datos

A pesar de las diferencias de trazo que se encuentran en los símbolos de la Figura 17, los rasgos que caracterizan a cada par de símbolos son claros mientras que en la Figura 18 se ejemplifican tres pares de símbolos que, a pesar de que están etiquetados con el mismo identificador de Thompson, sus características difieren a un nivel mayor y por tales motivos tuvieron que ser descartados.

Figura 18. Ejemplos de pares de símbolos sin equivalencia

4.2 Aumento del conjunto de datos

Bengio et al. (2010) mencionan que, dentro del área de reconocimiento de caracteres manuscritos, un algoritmo de aprendizaje profundo se beneficia más de lo que llaman “ejemplos fuera de la distribución”.

Partiendo de esta idea, implementamos algunas de las transformaciones de caracteres que utilizan para conseguir una distribución de datos que cubren un dominio bastante mayor que el conjunto de glifos que se tenía en un principio. Considerando que la mayoría de los glifos digitalizados son de carácter privado y que además muchos de los signos mayas tienen sólo unos pocos ejemplos conocidos, la generación de distorsiones automáticas resulta de crucial importancia para lograr un rendimiento aceptable en la clasificación.

Lo que significa que, aunque en el conjunto de datos exista una cantidad mínima de instancias por glifo, la generación de distorsiones permite obtener una cantidad mayor, garantizando que nunca se presente la misma imagen dos veces. Es de suma importancia mencionar que estas distorsiones no toman en cuenta las diferencias estilísticas significativas y las cuales sólo pueden ser aprendidas de un conjunto de datos inicial representativo; sin embargo, con la aplicación de las técnicas de distorsión, es posible generar pequeñas variaciones en la apariencia de las imágenes y ayudan a evitar que la red se sobreajuste al conjunto de datos.

Como se vio anteriormente, el modelo de degradación de la imagen sirve para aproximar una imagen que, tras algunos procesos de degradación y restauración, pueda asemejarse a su imagen original. Siguiendo esta aproximación, se propone generar más instancias de la imagen de un glifo a partir de diferentes métodos de transformación e inyección de ruido.

Los modelos de degradación y perturbación que se proponen fueron aplicados a imágenes binarias de 64 x 64 pixeles y se clasificaron en cinco categorías:

- Operaciones morfológicas
- Transformaciones geométricas
- Modelos de ruido
- Filtros
- Degradaciones combinadas

4.2.1 Operaciones morfológicas

Para modificar el grosor de los símbolos, se aplicaron dos operaciones morfológicas para cada imagen con diferentes elementos de estructura.

Dilatación

Esta operación sirve para aumentar el grosor de cada símbolo, se emplearon 12 diferentes elementos estructurales de tamaños 3x3 y 2x2 (ver Anexo).

Se realizaron un total de 12 dilataciones por cada elemento original con diferentes elementos de estructura. A continuación se muestran algunos ejemplos del glifo T001 dilatado con diferentes elementos de estructura (Figura 19).

Figura 19. Ejemplo de dilatación del glifo T001

Erosión

Esta operación permite disminuir el grosor de las imágenes. Se emplearon únicamente dos elementos de estructura, ambos de tamaño 2x2 (ver Anexo), ya que los utilizados en el proceso de dilatación provocaban la desaparición del glifo completo y por tal motivo no fueron considerados en esta operación.

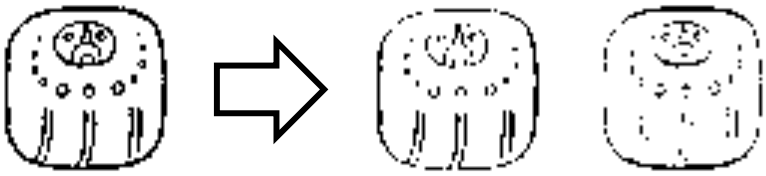

Figura 20.Ejemplo de erosión

4.2.2 Transformaciones geométricas

Las transformaciones modifican la imagen original, dando como resultado imágenes donde los pixeles no coinciden con la posición de la imagen original.

Rotación

Cada imagen original se giró con un ángulo máximo de ±15º. El ángulo y dirección del giro fueron establecidos aleatoriamente. Se utilizó el método de interpolación cúbica y aquellas imágenes que resultaban fuera del foco de la imagen, fueron escaladas para que de manera que cupieran en una imagen de 64x64. Se generaron 11 instancias por imagen.

Figura 21. Ejemplo de rotación

Permutación de pixeles

La imagen original fue recorrida con una máscara de tamaño 3x3. En cada una de las convoluciones, el pixel central de la máscara se intercambió con un píxel aleatorio entre sus vecinos más cercanos, a su izquierda, derecha, arriba o abajo. Se utilizó en parámetro de control para limitar la cantidad de vecinos cercanos a permutar, permitiendo conseguir 5 diferentes variaciones de la imagen.

Figura 22. Ejemplo de permutación de pixeles

Arremolinado

Esta transformación, distorsionó las imágenes originales simulando un pequeño remolino que comienza desde el centro hacia los extremos de la imagen (ver Anexo).

Se produjeron 4 imágenes arremolinadas por glifo, cada una con ángulos de ± 70º y ± 60º.

Figura 23. Ejemplo de arremolinado

4.2.3 Modelos de ruido

Ruido Gaussiano

Con este modelo, a la imagen original se le añadió aleatoriamente ruido blanco gaussiano con valores de media 0 y varianza 0.01. Para este modelo se crearon 3 imágenes ruidosas.

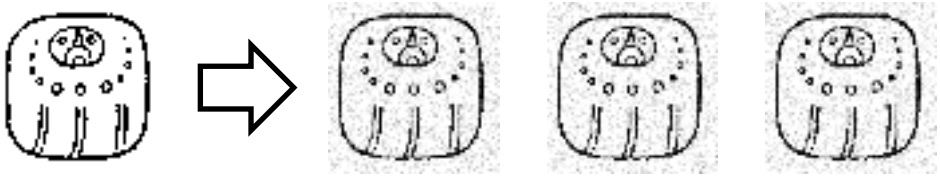

Figura 24. Ejemplo de ruido gaussiano

Ruido sal y pimienta

A la imagen original, se le agregaron pixeles blancos y negros aleatoriamente. La cantidad de ruido se moduló por un parámetro de control que indica en qué medida la imagen original es afectada. Se obtuvieron 5 imágenes transformadas por glifo.

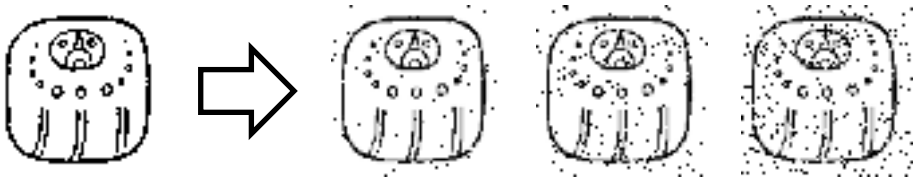

Figura 25. Ejemplo de ruido sal y pimienta

4.2.4 Filtros

Desenfoque

En este módulo, la imagen se recorrió con una máscara donde el valor final de cada pixel final fue obtenido de los pixeles encontrados en una cierta longitud establecida y desplazándose en dirección de un ángulo aleatoriamente establecido. Se crearon 3 variaciones por glifo.

Figura 26. Ejemplo de desenfoque

Suavizado Gaussiano

Las imágenes originales, se filtraron con un kernel isotópico gaussiano. El tamaño del filtro se estableció aleatoriamente y la cantidad de ruido

introducido fue modulado con un parámetro de control. Tres variaciones por glifo fueron generadas a partir de este módulo

Figura 27. Ejemplo de suavizado gaussiano

Tras la aplicación de todos los modelos anteriores, 48 instancias nuevas fueron generadas a partir de cada imagen original del conjunto de datos con las degradaciones simples. Sin embargo, esta cantidad aún es pequeña para lograr una clasificación satisfactoria. De modo que se propusieron nuevas degradaciones combinando las detalladas previamente para la obtención de un conjunto de datos aún mayor.

4.2.5 Degradaciones combinadas

Este tipo de degradación es el resultado de la combinación de dos transformaciones simples. Por cada variación obtenida en ellas, fue originada una nueva instancia a partir de la selección aleatoria de un método de degradación simple, excluyendo el ya existente en la imagen.

Un caso especial de transformación combinada fue la erosión. En el proceso de combinación, la dilatación fue excluida de manera que las dos variantes resultantes de la erosión simple además de ser mezcladas aleatoriamente con los métodos restantes fueran forzosamente mezcladas con la dilatación. Dando como resultado 2 variantes nuevas, resultado de un proceso de la aplicación de erosión y posteriormente, dilatación.

Figura 28. Combinación de erosión y dilatación

La tabla 2 sintetiza la cantidad de transformaciones, tanto simples como combinadas que se aplicaron para cada imagen original. Dado que en la aleatoriedad en algunos métodos genera siempre una imagen diferente, fue añadida una variante combinada más con degradaciones aleatorias para que finalmente, sumando las 99 variantes más la imagen original resultara un total de 100 instancias por glifo.

Transformaciones	Simples	Combinadas
Dilatación	12	12
Erosión	2	4
Rotación	11	11
Permutación de pixeles	5	5
Arremolinado	4	4
Ruido gaussiano	3	3
Ruido sal y pimienta	5	5
Desenfoque	3	3
Suavizado gaussiano	3	3
Total	**48**	**50**

Tabla 4. Resumen de transformaciones realizadas

4.3 Clasificación de glifos mayas

Como se vio en la sección 3, la clasificación de glifos mayas es una tarea que aún se encuentra sin ser resuelta satisfactoriamente. Los métodos que se han

propuesto para resolverla están enfocados principalmente en la manera en la que los glifos son descritos y los resultados que se han obtenido hasta ahora no son muy alentadores.

Con la idea de proponer un enfoque diferente, se abordó la tarea como un problema de clasificación de caracteres cualquiera, se cubrió la deficiencia en términos de cantidad de datos y se consideró a los algoritmos de clasificación como lo más importante en lugar de la representación de la imagen.

En primer lugar, fue necesario considerar un algoritmo sencillo como line base. El algoritmo KNN, descrito en el capítulo 2, funciona como punto de referencia, considerándolo como el algoritmo más sencillo para clasificación de imágenes. La implementación de dicho algoritmo se realizó utilizando además el método de validación cruzada de 10 pliegues. En cada pliegue, el conjunto de datos fue dividido en 10 secciones, donde 9 de ellas se seleccionaron para realizar el entrenamiento y la restante para la evaluación.

Durante el entrenamiento del algoritmo, cada imagen fue representada como un vector de características, donde cada posición en él correspondía a cada uno de los pixeles. Ante un nuevo patrón o imagen a clasificar del conjunto de prueba fue calculada la distancia correspondiente para cada una de las imágenes de entrenamiento y se seleccionaron diferentes K vecinos más cercanos.

La medida de distancia empleada fue la comparación pixel a pixel, es decir comparando la i-ésima posición del vector x contra el vector y. Finalmente, la imagen fue clasificada dependiendo de la clase mayoritaria presente en el conjunto de K vecinos cercanos.

4.3.1 Implementación de red neuronal convolucional

Las redes neuronales son consideradas como herramientas bastante poderosas para la clasificación de imágenes en general. Por otro lado, las redes neuronales convolucionales son más adecuadas para tareas de reconocimiento de caracteres que las redes tradicionales. En los enfoques tradicionales, una red neuronal se entrena en la salida de una etapa de extracción de características. En cambio, con las redes neuronales convolucionales, el entrenamiento se realiza en la matriz de la imagen sin procesar, es decir, no se realiza ninguna fase de representación de datos, aparte de la segmentación, de ser necesaria.

Una red ordinaria de perceptrón multicapa toma un vector de entrada, y como salida proporciona un vector de clasificación. Por otra parte, las redes neuronales convolucionales más sencillas toman una matriz tridimensional de entrada, la cual es la representación digital de la imagen. Posteriormente, es recorrida con una máscara para finalmente pasar por lo que realmente es un perceptrón multicapa, de manera que las capas convolucionales iniciales actúan como un extractor de características, mientras que las capas posteriores conectadas actúan como un clasificador Simard, Steinkraus, and Platt (2003)

Arquitecturas de red

Como se ha mencionado, las redes neuronales convolucionales poseen una capacidad mayor para la detección de características visuales, lo que permite que las tareas de clasificación de imágenes sean resueltas más eficazmente.

La estrategia general de una red convolucional es extraer características simples a una resolución alta y luego convertirlas en funciones más complejas

con una resolución menor. Siguiendo las arquitecturas de red descritas por Simonyan and Zisserman (2014), se exploraron 4 diferentes arquitecturas de CNN.

La arquitectura general consiste en una capa convolucional relativamente pequeña seguida por una capa de activación y una capa de max-pooling. El proceso es repetido en cada capa hasta que finalmente se calcula la probabilidad de pertenencia por clase utilizando un clasificador softmax. El tamaño de kernel utilizado fue de 3x3 con un salto de 1, lo que permite conservar el tamaño de la imagen a través de la red neuronal. En las capas de max-pooling se utilizaron ventanas de 2x2 con un salto de 2. En todas las capas tanto convolucionales como completamente conectadas se añadió a la salida una activación de tipo ReLU y un dropout de 0.5 para evitar el sobreajuste.

Las diferentes configuraciones de red son resumidas en la tabla siguiente. Cada modelo está nombrado de acuerdo a su cantidad de capas y se utiliza un índice adicional para diferenciar entre modelos con el mismo número de capas.

<table>
<tr><th>CNN6-1</th><th>CNN6-2</th><th>CNN7-1</th><th>CNN7-2</th></tr>
<tr><td colspan="4">Entrada: imagen 64x64</td></tr>
<tr><td>Capa convolucional 32
Capa convolucional 32</td><td rowspan="2">Capa convolucional 64</td><td rowspan="2">Capa convolucional 64</td><td rowspan="2">Capa convolucional 64</td></tr>
<tr><td colspan="4">Maxpool</td></tr>
<tr><td>Capa convolucional 32
Capa convolucional 32</td><td rowspan="2">Capa convolucional 128</td><td rowspan="2">Capa convolucional 128</td><td rowspan="2">Capa convolucional 128</td></tr>
<tr><td colspan="4">Maxpool</td></tr>
<tr><td rowspan="4"></td><td rowspan="2">Capa convolucional 512</td><td>Capa convolucional 512</td><td rowspan="2">Capa convolucional 192</td></tr>
<tr><td>Capa convolucional 512</td></tr>
<tr><td colspan="3">Maxpool</td></tr>
<tr><td></td><td></td><td>Capa convolucional 32</td></tr>
<tr><td></td><td></td><td>Maxpool</td></tr>
<tr><td rowspan="2">Capa totalmente conectada 256</td><td>Capa totalmente conectada 4096</td><td>Capa totalmente conectada 4096</td><td>Capa totalmente conectada 1024</td></tr>
<tr><td>Capa totalmente conectada 4096</td><td>Capa totalmente conectada 4096</td><td>Capa totalmente conectada 1024</td></tr>
<tr><td colspan="4">Capa totalmente conectada n-clases</td></tr>
<tr><td colspan="4">Softmax</td></tr>
</table>

Tabla 5. Arquitecturas CNN propuesta

5

Experimentos y resultados

5.1 Diseño de experimentos

La clasificación fue dividida en 2 tipos: clasificación con datos sintéticos y clasificación con datos reales.

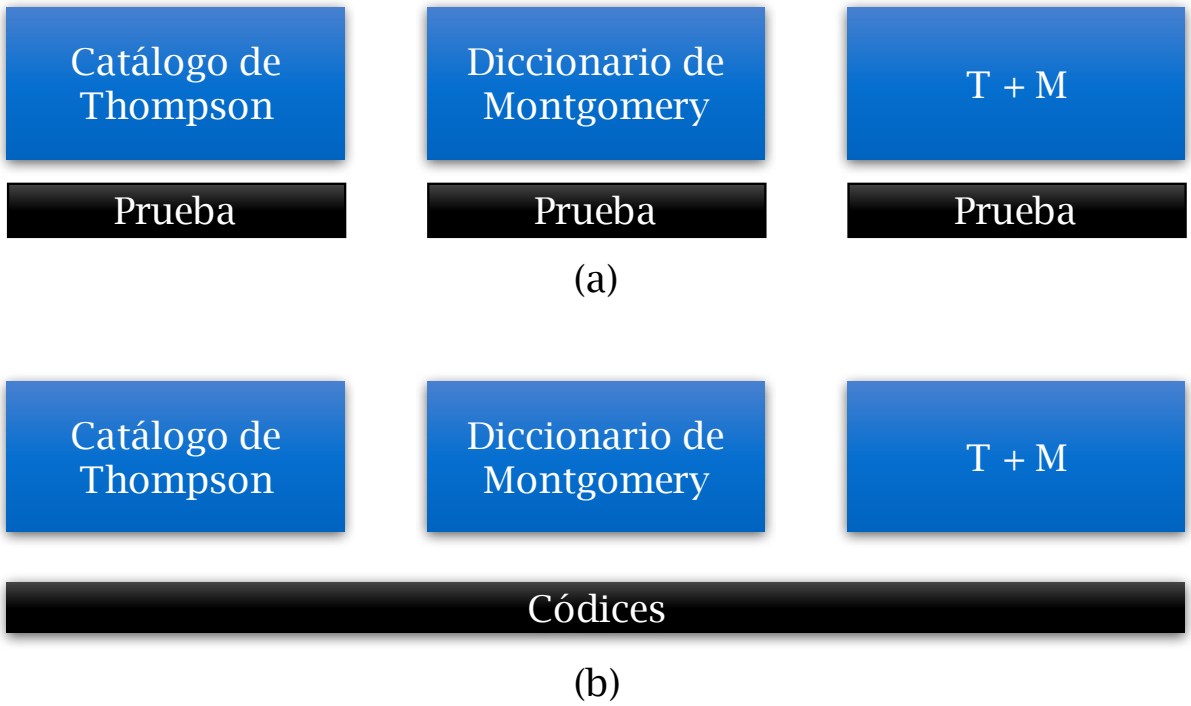

Figura 29. Diseño de experimentos. (a) Datos sintéticos. (b) Datos reales

Como se ilustra en la figura anterior, los tipos de clasificación, se subdividen en 3 experimentos donde cada tipo de experimento utiliza un conjunto de datos diferente.

5.1.1 Clasificación de datos sintéticos

Los experimentos de clasificación de datos sintéticos fueron diseñados tras aplicar las transformaciones descritas en el capítulo 4 a los conjuntos de datos basados en el catálogo de Thompson y el diccionario de Montgomery. Se añadió además un tercer conjunto, resultado de la intersección de los glifos presentes en ambos catálogos.

Conjunto de datos / Cantidad	Thompson	Montgomery	Thompson + Montgomery
Glifos originales	450	194	154
Glifos con transformaciones	45,000	19,400	15,400
Clases	450	194	77

Tabla 6. Distribución de glifos (1)

Durante la etapa de entrenamiento se utilizó el 80% de cada conjunto de datos, dejando el 20% restante para pruebas.

5.2 Clasificación de datos reales

La necesidad de evaluar los resultados obtenidos con la metodología que se propuso con un conjunto de datos reales y poder evaluar el desempeño del

clasificador condujo a la obtención del conjunto de datos de códices Mayas. Hu et al. (2014)

Este conjunto de datos, contiene una representación de alta calidad de los datos de jeroglíficos mayas, extraídos de los tres códices mayas sobrevivientes: Dresde, Madrid y Paris. Lo conforman 174 glifos de alta calidad reconstruidos y segmentados de varios bloques, junto con el identificador de Thompson correspondiente para cada uno de ellos; cuenta con un total de 32 símbolos principales.

Los 32 símbolos principales presentes en este corpus fueron identificados en los conjuntos de Thompson y Montgomery y fueron mezclados para la obtención de un tercer conjunto de entrenamiento.

Conjunto de datos / Cantidad	Thompson	Montgomery	Thompson + Montgomery
Glifos originales	32	32	64
Glifos con transformaciones	3,200	3,200	6,400
Clases	32	32	32

Tabla 7. Distribución de glifos (2)

En los 3 casos, cada conjunto de datos fue empleado para la etapa de entrenamiento mientras que para la etapa de prueba se utilizó exclusivamente el conjunto perteneciente a los códices, esto con la finalidad de verificar que las transformaciones aplicadas realmente se asemejaran a las imágenes de un conjunto de datos más cercano a la realidad.

5.3 Análisis de resultados

	Conjunto de datos		
	Thompson	Montgomery	T+M
# clases	450	194	77
Algoritmo			
KNN- 60	75%	84%	68%
CNN6-1	**99.78%**	99.85%	99.76%
CNN6-2	99.68%	99.79%	99.68%
CNN7-1	99.61%	99.75%	99.89%
CNN7-2	99.57%	**99.86%**	**99.92%**

Tabla 8. Resultados clasificación de datos sintéticos

Comenzando por el algoritmo KNN, es claro que la manera en que las imágenes son clasificadas es mala, aunque podría ser mejorada si la medida de similitud que se emplea se modificara. La comparación pixel a pixel afecta de manera significativa al desempeño del clasificador ya que, si una de las imágenes fuera desplazada un pixel o girada ligeramente, la medida de similitud determinaría que las imágenes son diferentes al no lograr un empate entre pixeles.

Una observación importante es la diferencia de resultados entre las arquitecturas propuestas. En primer lugar, los resultados de los experimentos con los datos sintéticos para el conjunto de Thompson la arquitectura que mejor exactitud proporciona es la CNN6-1 la cual resulta ser la más sencilla, en cambio, en el conjunto de datos de Montgomery la arquitectura más compleja CNN7-2, es la que tiene el mejor desempeño. Esto podría indicar que como la cantidad de clases en el conjunto de datos de Thompson es mucho

mayor en comparación con el de Montgomery, las instancias que se tienen por clase no son suficientes para discriminar entre las 450 clases que se tienen, mientras que en el conjunto de datos de Montgomery la relación instancias por clase contra número de clases es más proporcional y por tal motivo una arquitectura más profunda funciona de mejor manera.

Para el caso del tercer conjunto de datos, con la mezcla de ambos conjuntos de datos es claro que la arquitectura más compleja supera en exactitud a las demás y la tasa de reconocimiento es la mayor entre los tres tipos de experimentos. Nuevamente, la razón por la cual esta configuración proporciona mejores resultados que las demás puede atribuirse a la cantidad de ejemplos por clase con los que se cuenta. En los dos experimentos anteriores la cantidad de instancias por clase era de tan sólo 100 glifos, mientras que en este caso al unir los dos conjuntos se obtienen 200 instancias lo que permite que la red más profunda pueda generar una mejor representación de los glifos y además una mejor distinción entre clases.

	Conjunto de datos		
	Thompson	Montgomery	T+M
# clases	450	194	77
Algoritmo			
KNN- 60	17%	23%	35%
CNN6-1	58%	**72%**	77%
CNN6-2	57%	69%	75%
CNN7-1	61%	61%	**78%**
CNN7-2	**63%**	63%	77%

Tabla 9. Resultados de clasificación de datos reales

Comparando los resultados obtenidos entre la clasificación de datos sintéticos y datos reales utilizando redes neuronales convolucionales, se puede observar en primer lugar, que las arquitecturas al ser entrenadas con los datos sintéticos tienden a sobre ajustarse al conjunto de datos derivando en un porcentaje de exactitud bastante elevado.

A pesar de que la cantidad de clases en los tres experimentos era menor que en la clasificación de datos sintéticos las arquitecturas propuestas logran una tasa de reconocimiento un poco baja.

Contrario a lo sucedido en la clasificación de datos sintéticos, el entrenamiento con el conjunto de datos de Thompson la arquitectura con la mejor tasa de clasificación fue la CNN7-2; mientras que en el conjunto de datos de Montgomery fue la más sencilla, CNN6-1. Una explicación a estos resultados puede ser que el conjunto de datos de Montgomery contiene glifos que se asemejan un poco más a los presentes en el conjunto de datos de los códices, mientras que los glifos dentro del catálogo de Thompson son símbolos estandarizados (sección 3.2) y su semejanza con los reales es menor.

En los resultados obtenidos con la combinación de ambos conjuntos de datos se observa una mezcla de los dos resultados anteriores. La arquitectura CNN7-1 presenta un resultado de exactitud ligeramente mayor que CNN6-1 y CNN7-2. Al aumentar la cantidad de ejemplos por clase, las redes de 7 capas superan rápidamente a las de 6 capas, lo que quiere decir que, aunque los glifos del catálogo de Thompson tienen una semejanza menor con los de los códices, la cantidad de variaciones sí aportan una mejora en el desempeño de la red.

Para obtener una mejor interpretación de los resultados para ambos tipos de clasificación fue necesario obtener la significancia estadística de cada uno de los experimentos realizados. Este proceso se realiza utilizando diferentes

pruebas que permiten cuantificar hasta qué punto la variabilidad de la muestra en los experimentos es responsable de los resultados obtenidos.

La significancia estadística depende principalmente de dos conceptos: la magnitud de la diferencia entre resultados y el tamaño de la muestra de estudio, en nuestro caso el conjunto de datos. Con respecto a la magnitud de la diferencia entre resultados, es importante considerar que a mayor diferencia entre las diferentes arquitecturas será más evidente que la diferencia de resultados entre los algoritmos es significativa, como es el caso de los resultados probados en el conjunto de datos reales. En los resultados obtenidos en las pruebas con datos sintéticos la diferencia entre cada uno es mínima, lo que indica que las posibilidades de detectar diferencias se minimizan. Por otra parte, respecto al tamaño de la muestra, se puede deducir que mientras mayor sea el conjunto de datos, más fácil será detectar diferencias entre las imágenes; entonces, cuando las diferencias son muy pequeñas entre las instancias es necesario contar con una cantidad mayor de datos.

Los cálculos realizados para la obtención de la significancia estadística son obtenidos a partir de un experimento de referencia, se seleccionó como referencia el algoritmo con el mejor porcentaje de exactitud para cada uno de las diferentes variantes de experimentos. La primera columna, en todos los casos, corresponde al algoritmo con el mejor desempeño. (Tablas 10,11)

Arquitectura	Tamaño dataset	Símbolos reconocidos correctamente	Exactitud	Resultado
Thompson				
CNN6-1	**45,000**	**44,901**	**99.78%**	**N/A**
CNN6-2	45,000	44,856	99.68%	Significativo en un 99.62%
CNN7-1	45,000	44,825	99.61%	Significativo en un 100.00%
CNN7-2	45,000	44,807	99.57%	Significativo en un 100.00%
Montgomery				
CNN7-2	**19,400**	**19,373**	**99.86%**	**N/A**
CNN6-1	19,400	19,371	99.85%	No significativo
CNN6-2	19,400	19,359	99.79%	No significativo
CNN7-1	19,400	19,352	99.75%	Significativo en un 98.48%
Thompson + Montgomery				
CNN7-2	**15,400**	**15,388**	**99.92%**	**N/A**
CNN6-1	15,400	15,363	99.76%	Significativo en un 99.96%
CNN6-2	15,400	15,351	99.68%	Significativo en un 100.00%
CNN7-1	15,400	15,383	99.89%	No significativo

Tabla 10. Significancia estadística experimentos datos sintéticos

Arquitectura	Tamaño dataset	Símbolos reconocidos correctamente	Exactitud	Resultado
Thompson				
CNN7-2	**3,200**	**2,016**	**63.00%**	**N/A**
CNN6-1	3,200	1,856	58.00%	Significativo en un 100.00%
CNN6-2	3,200	1,824	57.00%	Significativo en un 100.00%
CNN7-1	3,200	1,952	61.00%	No significativo
Montgomery				
CNN6-1	**3200**	**2304**	**72.00%**	**N/A**
CNN6-2	3200	2208	69.00%	Significativo en un 99.15%
CNN7-1	3200	1952	61.00%	Significativo en un 100.00%
CNN7-2	3200	2016	63.00%	Significativo en un 100.00%
Thompson + Montgomery				
CNN7-1	**6400**	**4992**	**78.00%**	**N/A**
CNN6-1	6400	4928	77.00%	No significativo
CNN6-2	6400	4,800	75.00%	Significativo en un 99.99%
CNN7-2	6400	4,928	77.00%	No significativo

Tabla 11. Significancia estadística experimentos datos reales

Finalmente, los resultados obtenidos en los experimentos realizados con el conjunto de datos de entrenamiento sin las transformaciones propuestas demostraron que las arquitecturas de redes neuronales convolucionales se benefician de las variaciones realizadas y por tanto la tasa de reconocimiento es mayor tanto para la clasificación de datos sintéticos como de datos reales. (Tabla 12)

	Conjunto de datos	
	T+M	T+M/Codex
# clases	77	32
Algoritmo		
CNN6-1	**98.70%**	**38%**
CNN6-2	96.63%	36%
CNN7-1	94.80%	11%
CNN7-2	97.40%	16%

Tabla 12. Clasificación sin transformaciones

6

Conclusiones

La metodología propuesta para el reconocimiento de glifos mayas ha cumplido el objetivo principal de este trabajo de tesis. Aunque todavía quedan deficiencias que cubrir, es importante destacar que se mejoró el nivel de precisión, además, la cantidad de clases identificadas es ligeramente mayor a las aproximaciones descritas en la sección 3.3. (Tabla 13)

Referencia	# Clases	Aleatorio	Exactitud
Roman-Rangel et al. (2009)	22	4.5%	32%
Roman-Rangel et al. (2011)	24	4.2%	54%
Hu et al. (2014)	25	4.0%	65%
Metodología propuesta	**32**	**3.1%**	**78%**

Tabla 13. Comparación entre métodos.

Así mismo, se demostró que la detección específica de símbolos principales es posible. Los trabajos anteriores se habían enfocado únicamente en los glifos pertenecientes a la categoría de afijos, los cuales representan

aproximadamente 10% del total de símbolos existentes en la escritura maya, mientras que la gran mayoría pertenecen a la clase de símbolos principales.

Se demostró que el uso de redes neuronales convolucionales es una solución potencialmente útil para resolver esta tarea, permitiendo que la propia red sea quien realice el proceso de extracción de características. Aunque la tasa de clasificación no es tan alta como se esperaría, los resultados demuestran que el uso de imágenes con transformaciones puede ayudar significativamente al desempeño de la red.

Además, se verificó que es posible la clasificación de glifos mayas haciendo uso de redes neuronales convolucionales para la detección específica de los símbolos principales.

Es notable que la aplicación de las variaciones al conjunto de datos original sí ayudaron a mejorar la tasa de clasificación, pero la cantidad de instancias generadas para algunos experimentos aún es pequeña. Por lo tanto, se deduce que, al aumentar la cantidad de clases, los ejemplos por clase deben también aumentar.

Se comprobó que las redes convolucionales más profundas no necesariamente funcionan mejor para todos los casos. Esto también va ligado a la cantidad de ejemplos por clase, si el número de instancias no son suficientes para lograr un entrenamiento correcto, la clasificación resultará en un mal desempeño.

Finalmente, un aspecto importante a destacar es la escalabilidad de las redes neuronales convolucionales; cuando se entrenó con los datos sintéticos del catálogo de Thompson el tiempo de ejecución para cada red era bastante significativo, en primer lugar, por la cantidad de clases y en segundo lugar por la cantidad de capas ocultas y el número de épocas con las que fue

entrenada cada red. Es importante mencionar que, para la posible clasificación del catálogo completo de glifos, sería necesaria la recopilación de una mayor cantidad de ejemplos por clase además de la generación de sus variaciones para poder lograr una clasificación adecuada.

6.1 Trabajo futuro

Durante la realización de este trabajo se encontró que aún existen puntos por mejorar en el campo de la clasificación automática de glifos mayas. El más claro es la necesidad de más imágenes digitalizadas de glifos. Una posible solución podría ser considerar la generación automática de imágenes de dibujos a partir de fotografías de estelas o códices con el fin de aumentar la cantidad de glifos reales.

La generación de transformaciones es un área que también puede ser mejorada. Aquí se propusieron algunos ejemplos de variaciones para aplicar a los glifos, pero pudiera ser que el reconocimiento de los glifos se viera afectado debido a alguna deformación excesiva. Una posible mejora podría ser, evaluar cada una de las transformaciones con el fin de determinar cuáles son las que ayudan más en la clasificación.

Adicionalmente, los afijos pueden ser tomados en cuenta para lograr el reconocimiento de glifos compuestos. Una manera de resolver esto puede ser considerar incluir algún modelo de lenguaje o de co-ocurrencia de glifos para facilitar la detección de estructuras más complejas y no sólo de símbolos individuales.

ANEXO

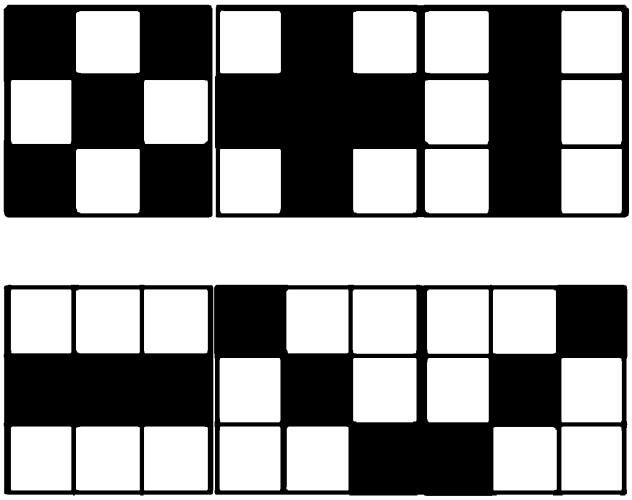

Figura 30. Elementos de estructura para dilatación. Tamaño 3x3.

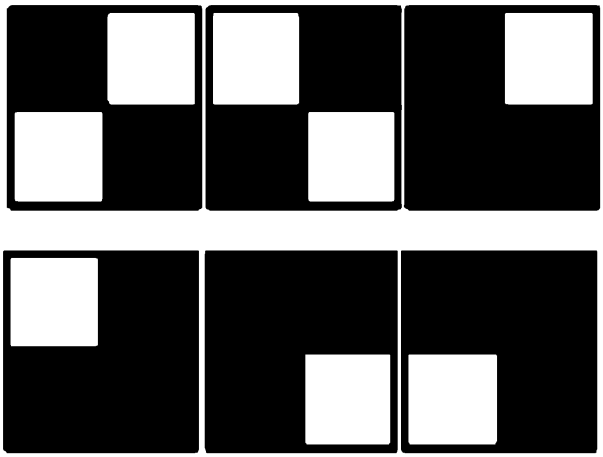

Figura 31. Elementos de estructura para dilatación. Tamaño 2x2

Figura 32. Elementos de estructura para erosión. Tamaño 2x2

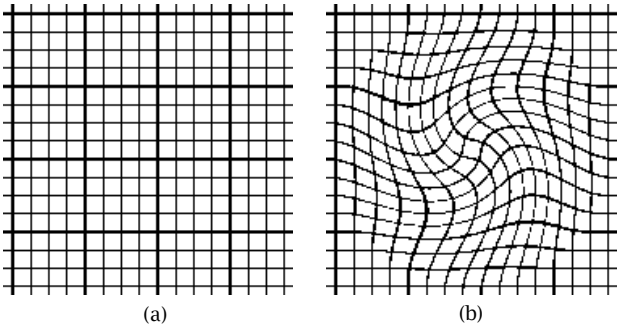

Figura 33. Arremolinado de una imagen

REFERENCIAS

Alvarado, R. (1998). The Mayan Epigraphic Database Project.

Arrivault, D., Richard, N., Fernandez-Maloigne, C., & Bouyer, P. (2005). *Collaboration between statistical and structural approaches for old handwritten characters recognition.* Paper presented at the International Workshop on Graph-Based Representations in Pattern Recognition.

Bengio, Y., Bastien, F., Bergeron, A., Boulanger-Lewandowski, N., Chherawala, Y., Cisse, M., . . . Glorot, X. (2010). *Deep self-taught learning for handwritten character recognition.* Paper presented at the NIPS* 2010 Deep Learning and Unsupervised Feature Learning Workshop.

Chellapilla, K., Larson, K., Simard, P. Y., & Czerwinski, M. (2005). *Computers beat Humans at Single Character Recognition in Reading based Human Interaction Proofs (HIPs).* Paper presented at the CEAS.

Coe, M. D. (2001). *Reading the Maya glyphs*: Thames & Hudson.

Demuth, H. B., Beale, M. H., Jess, O. D., & Hagan, M. T. (2014). *Neural Network Design*: Martin Hagan.

Franken, M., & van Gemert, J. (2013, 2013). *Automatic Egyptian hieroglyph recognition by retrieving images as texts.*

Gonzalez, R. C., & Woods, R. E. (2006). *Digital Image Processing (3rd Edition)*: Prentice-Hall, Inc.

Hu, R., Gayol, C. P., Krempel, G., Odobez, J.-M., & Gatica-Perez, D. (2014). *Automatic Maya hieroglyph retrieval using shape and context information.* Paper presented at the Proceedings of the 22nd ACM international conference on Multimedia.

Kettunen, H. J., & Helmke, C. (2008). *Introduction to Maya Hieroglyphs: Workshop Handbook.*

Knórozov, Y. V., Ochoa, P. R., Marín, E. G., & González, M. C. (1999). *Compendio Xcaret de la escritura jeroglífica maya descifrada por Yuri V. Knórosov* (Vol. 1): Universidad de Quintana Roo.

LeCun, Y. (1998). The MNIST database of handwritten digits.

Luxton, R., Balam, P., & Uribe, M. (1993). *Sueño del camino maya: el chamanismo ilustrado en Yucatán*: Fondo de Cultura Económica.

Montgomery, J. (2002). *How to read Maya hieroglyphs*: Hippocrene Books New York.

Ntzios, K., Gatos, B., Pratikakis, I., Konidaris, T., & Perantonis, S. J. (2005). *An old Greek handwritten OCR system.* Paper presented at the Document Analysis and Recognition, 2005. Proceedings. Eighth International Conference on.

Roman-Rangel, E., Pallan, C., Odobez, J. M., & Gatica-Perez, D. (2009, 2009). *Retrieving ancient Maya glyphs with Shape Context.* Paper presented at the 2009 IEEE 12th International Conference on Computer Vision Workshops, ICCV Workshops.

Roman-Rangel, E., Pallan Gayol, C., Odobez, J.-M., & Gatica-Perez, D. (2011). *Searching the past: an improved shape descriptor to retrieve maya hieroglyphs.* Paper presented at the Proceedings of the 19th ACM international conference on Multimedia.

Sidorov, G., Pichardo-Lagunas, O., & Chanona-Hernandez, L. (2009). Search Interface to a Mayan Glyph Database Based on Visual Characteristics *Natural Language Processing and Information Systems: 14th International Conference on Applications of Natural Language to Information Systems.* (pp. 222-229). Berlin, Heidelberg: Springer Berlin Heidelberg.

Simard, P. Y., Steinkraus, D., & Platt, J. C. (2003). *Best Practices for Convolutional Neural Networks Applied to Visual Document Analysis.*

Simonyan, K., & Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556.*

Thompson, J. E. S. (1962). *A catalog of Maya hieroglyphs*: University of Oklahoma Press Norman.