



INAOE

UNA NUEVA REPRESENTACIÓN DE JEROGLÍFICOS BASADA EN SU DESCOMPOSICIÓN EN PARTES PRIMITIVAS

Por:

Laura Alejandra Pinilla Buitrago

Tesis sometida como requisito parcial
para obtener el grado de:

**DOCTORA EN CIENCIAS EN EL ÁREA
DE CIENCIAS COMPUTACIONALES**

en el

Instituto Nacional de Astrofísica, Óptica y Electrónica

Junio, 2019
Tonantzintla, Puebla

Supervisada por:

Dr. Jesús Ariel Carrasco Ochoa, INAOE

©INAOE 2019

Derechos reservados

El autor otorga al INAOE el permiso de reproducir
y distribuir copias en su totalidad o en
partes de esta tesis.



RESUMEN

La recuperación automática o semi-automática de jeroglíficos ha sido un tópico de interés multidisciplinario donde intervienen áreas como visión por computadora, reconocimiento de patrones y humanidades (e.g. arqueología, antropología, sociología, etc.), teniendo por objetivo facilitar y dar soporte a la búsqueda manual de jeroglíficos en catálogos impresos. Para esta tarea, los jeroglíficos deben ser representados computacionalmente de acuerdo a su contenido visual. En la literatura, las representaciones de jeroglíficos más exitosas para tareas de recuperación han sido construidas siguiendo el modelo BoVW (del inglés *Bag of Visual Words*). En esta tesis se propone una nueva representación de jeroglíficos, la cual resulta de introducir modificaciones en las etapas del modelo BoVW. Dicha representación permite, de acuerdo a nuestros experimentos, obtener mejores resultados en recuperación de jeroglíficos que las representaciones más exitosas del estado del arte.

ABSTRACT

Automatic or semi-automatic hieroglyph retrieval has been a topic of multi-disciplinary interest, where areas like computer vision, pattern recognition and humanities (archeology, anthropology, sociology, etc.) collaborate to facilitate and support the manual search of hieroglyphs from printed catalogs. For this task, hieroglyphs must be represented according their visual content. In the literature, the most successful hieroglyph representations for retrieval have been built following the BoVW (Bag of Visual Words) model. In this PhD. research, a new hieroglyph representation is proposed. This hieroglyph representation results from introducing some modifications into the stages of the BoVW model. According to our experiments, the proposed hieroglyph representation allows obtaining better retrieval results than those results obtained by the most successful hieroglyph representations of the state of the art.

CONTENIDO

RESUMEN	I
ABSTRACT	III
1. INTRODUCCIÓN	1
1.1. Motivación y justificación	3
1.2. Hipótesis	4
1.3. Objetivos	4
1.4. Contribuciones	5
1.5. Descripción del documento	5
2. CONCEPTOS BÁSICOS	7
2.1. Forma adelgazada y sus elementos	7
2.2. Representación bajo el modelo BoVW	11
2.3. Recuperación de Información	12
3. TRABAJO RELACIONADO	17
3.1. Trabajos que representan jeroglíficos bajo el modelo BoVW	18
3.2. Discusión	22

4. REPRESENTACIÓN DE JEROGLÍFICOS PROPUESTA	25
4.1. Podado de jeroglíficos adelgazados	26
4.2. Extracción de descriptores locales a partir fragmentos del jeroglífico	32
4.3. Reemplazando cada descriptor local por sus k -palabras visuales más cercanas en el vocabulario	36
5. RESULTADOS EXPERIMENTALES	41
5.1. Bases de datos y configuración experimental	42
5.2. Evaluación de la poda propuesta	46
5.3. Evaluación del reemplazo de cada descriptor local por sus k -palabras visuales más cercanas.	50
5.4. Evaluación del cálculo de descriptores locales desde fragmentos del jeroglífico	54
5.5. Evaluación de la representación de jeroglíficos propuesta	59
6. CONCLUSIONES Y TRABAJO FUTURO	67
6.1. Conclusiones	67
6.2. Contribuciones	68
6.3. Trabajo futuro	69
6.4. Publicaciones	70
Anexo A. Descriptores locales	81
A.1. <i>Shape-Context:SC</i>	81
A.2. <i>Histogram of Orientation Shape-Context: HOOSC</i>	82

ÍNDICE DE FIGURAS

2.1. Tipos de pixeles en una forma adelgazada de un jeroglífico.	9
2.2. Tipos de ramas en una forma adelgazada de un jeroglífico.	10
2.3. Representación de jeroglíficos bajo el modelo BoVW	12
3.1. Representación para la recuperación de jeroglíficos en la literatura.	17
4.1. Nueva representación de jeroglíficos propuesta siguiendo el modelo BoVW	27
4.2. Forma adelgazada de un jeroglífico	28
4.3. Región reconstruida a partir de una rama en la forma adelgazada .	29
4.4. Podado de ramas finales en la forma adelgazada	30
4.5. Forma adelgazada y forma adelgazada y podada de un jeroglífico .	32
4.6. Descriptor de un punto de interés vs descriptor de un fragmento del jeroglífico.	33
4.7. Pixeles vecinos de un pixel p_i en P' para la extracción de fragmentos del jeroglífico	34
4.8. Fragmentos del jeroglífico extraídos desde el pixel p_i	35
4.9. Descriptores locales en términos del vocabulario visual.	38

5.1. Jeroglífico por clase para Maya-I y Maya-II.	43
5.2. Petroglifo por clase para Petroglyph-I y Petroglyph-II.	44
5.3. Jeroglífico por clase para Egyptian.	45
5.4. Jeroglíficos de consulta y sus primeros cinco jeroglíficos más parecidos en la base de datos, para Maya-I y Maya-II, utilizando la representación de jeroglíficos propuesta.	62
5.5. Jeroglíficos de consulta y sus primeros cinco jeroglíficos más parecidos en la base de datos, para Petroglyph-I y Petroglyph-II, utilizando la representación de jeroglíficos propuesta.	63
5.6. Jeroglíficos de consulta y sus primeros cinco jeroglíficos más parecidos en la base de datos para Egyptian, utilizando la representación de jeroglíficos propuesta.	64
5.7. mAP vs recuerdo para (a) Maya-I, (b) Maya-II, (c) Petroglyph-I, (d) Petroglyph-II y (e) Egyptian utilizando el 5%, 10% y 20% de puntos para BoVW, BoP y Bok-NF, y HOOSC128 _{xy} como descriptor local.	65
A.1. Descriptor <i>Shape-Context</i> (SC)	82
A.2. Descriptor <i>Histogram of Orientation Shape-Context</i> (HOOSC)	84

ÍNDICE DE TABLAS

5.1. Información general de las bases de datos de jeroglíficos	43
5.2. mAP promedio obtenido al representar jeroglíficos bajo el modelo BoVW desde jeroglíficos adelgazados sin podar y desde jeroglíficos adelgazados y podados	48
5.3. mAP promedio obtenido utilizando la representación de jeroglíficos BoP construida desde jeroglíficos adelgazados sin podar y desde jeroglíficos adelgazados y podados	49
5.4. mAp promedio obtenido por $\text{BoP}_{(k,k)}^*$, $\text{BoP}_{(1,k)}^*$ y $\text{BoP}_{(k,1)}^*$	53
5.5. mAP promedio obtenido por BoP^* and $\text{BoP}_{(k,1)}^*$	55
5.6. mAP promedio obtenido por $\text{Bo}k\text{-VWF}$	57
5.7. mAP promedio obtenido por $\text{BoP}_{(k,1)}^*$ y $\text{Bo}k\text{-VWF}$	58
5.8. mAP promedio obtenido por BoVW, BoP y $\text{Bo}k\text{-VWF}$	60
5.9. $p@n$ ($n=1,3,5$) obtenido por BoVW, BoP y $\text{Bo}k\text{-VWF}$	61

INTRODUCCIÓN

Parte del trabajo diario que llevan a cabo historiadores, arqueólogos y antropólogos consiste en etiquetar y catalogar jeroglíficos¹. A menudo, etiquetar y catalogar jeroglíficos conlleva un proceso de recuperación manual, tratando de encontrar para cada jeroglífico aún no identificado, el o los jeroglíficos visualmente más parecidos entre los que han sido previamente etiquetados y organizados en catálogos impresos. Este proceso es una tarea no trivial que implica gran consumo de recursos y que comúnmente sólo puede ser llevada a cabo por expertos. Es por ello que, en los últimos años, se ha trabajado de forma interdisciplinaria entre las áreas de visión por computadora, artes y humanidades, teniendo por objetivo el desarrollo de herramientas automáticas o semi-automáticas que soporten y faciliten el proceso de etiquetado y catalogación de material que contiene escritura antigua. Estas herramientas realizan entre otras, tareas de detección [Deufemia and Paolino, 2014; Domingo et al., 2017; Franken and van Gemert, 2013; Frauel et al., 2006; Roman-Rangel and Marchand-Maillet, 2015; Roman-Rangel et al., 2013], clasificación [Can et al., 2016, 2017, 2018; Deufemia et al., 2012; Guo et al., 2016; Seidl and Breiteneder, 2011; Seidl et al., 2015a,b; Zhu et al., 2009] y recuperación [Gatica-Perez et al., 2012; Hu et al., 2015; Roman-Rangel and

¹En este trabajo, un jeroglífico es un símbolo que forma parte de un sistema de escritura antiguo.

Marchand-Maillet, 2013a, 2016; Roman-Rangel et al., 2011b; Zhu et al., 2011] de jeroglíficos. Esta tesis se enfoca en el problema de recuperación de jeroglíficos.

Para que un sistema sea capaz de recuperar jeroglíficos, éstos deben ser representados de acuerdo a su contenido visual. La representación de jeroglíficos enfrenta múltiples desafíos: los jeroglíficos son a menudo imágenes binarias (o en escala de grises) donde la información disponible está limitada a la forma. La edad del material antiguo y la superposición entre algunos símbolos, usualmente generan ruido y pérdida parcial de información. Además, podemos encontrar jeroglíficos similares que pertenecen a diferentes categorías y otros muy diferentes dentro de la misma categoría, lo anterior producto del periodo histórico, la región geográfica, el estilo y habilidad de la persona que los dibujó.

En la literatura, se han presentado varios trabajos en recuperación de jeroglíficos [Gatica-Perez et al., 2012; Hu et al., 2014; Roman-Rangel and Marchand-Maillet, 2013a, 2016; Roman-Rangel et al., 2009, 2011a, 2016; Zhu et al., 2011], donde los jeroglíficos son representados siguiendo principalmente dos enfoques:

Representación de jeroglíficos basada en descriptores [Roman-Rangel et al., 2009; Zhu et al., 2011]. La representación de jeroglíficos basada en descriptores se realiza mediante el cálculo de descriptores *globales* ó *locales* [Heider et al., 2011; Kazmi et al., 2013; Lisin et al., 2005; Tian, 2013]. Los descriptores *globales* codifican las características visuales del jeroglífico como un todo, a través de un único descriptor (i.e. un único vector). Los descriptores *locales* extraen las características visuales del jeroglífico por partes (regiones) o desde diferentes perspectivas (e.g. puntos de interés, ángulos), obteniendo un descriptor para cada región, punto de interés o ángulo; es decir, cada jeroglífico es representado por múltiples descriptores locales; en este caso, la similitud entre dos jeroglíficos se calcula encontrando la mejor correspondencia descriptor-a-descriptor [Jonker and Volgenant, 1987; Papadimitriou and Steiglitz, 1982].

Representación de jeroglíficos bajo el modelo de bolsa de palabras visuales BoVW (del inglés *Bag of Visual Words*) [Gatica-Perez et al., 2012; Hu et al., 2014; Roman-Rangel and Marchand-Maillet, 2013a, 2016; Roman-Rangel et al., 2011a]. La representación de jeroglíficos bajo el modelo BoVW [Sivic and Zisserman, 2003] se obtiene haciendo uso de descriptores locales y construyendo, a partir de los mismos, un vocabulario visual. Ese vocabulario visual resulta de agrupar en c grupos (donde c es definido por el usuario) a los descriptores locales calculados desde jeroglíficos en el conjunto de entrenamiento, siendo el centroide de cada grupo una palabra visual en el vocabulario. Así, los descriptores locales pueden ser expresados en función de un conjunto finito de palabras visuales, reemplazando a cada descriptor local por su palabra visual más cercana en el vocabulario. Finalmente, cada jeroglífico se representa mediante un vector que contiene el número de ocurrencias de las palabras visuales en el jeroglífico.

De estos enfoques, los trabajos más exitosos representan jeroglíficos bajo el modelo BoVW. Por lo tanto, en este trabajo de investigación se estudian diferentes estrategias para enfrentar algunas de sus limitaciones, de modo que lo anterior resulte en una nueva representación de jeroglíficos, a través de la cual se logren alcanzar mejores resultados en recuperación de jeroglíficos respecto a los obtenidos por trabajos del estado del arte.

1.1. Motivación y justificación

La representación de jeroglíficos para tareas de recuperación es un problema desafiante, ya que podemos encontrar jeroglíficos parecidos que pertenecen a diferentes categorías y otros muy diferentes que pertenecen a la misma categoría. Además, los jeroglíficos comúnmente se encuentran sobre materiales con muchos años de antigüedad y abandono, lo que ha provocado pérdida parcial de información

en algunos jeroglíficos [Roman-Rangel et al., 2009].

Como ya se mencionó, los mejores resultados en recuperación de jeroglíficos se han obtenido cuando éstos son representados bajo el modelo BoVW [Gatica-Perez et al., 2012; Hu et al., 2014; Roman-Rangel and Marchand-Maillet, 2013a; Roman-Rangel et al., 2011a]. Sin embargo, estos resultados son susceptibles a mejora. Por lo tanto, es necesario proponer una nueva representación de jeroglíficos, a través de la cual se solventen algunas de las limitaciones de las representaciones existentes construidas bajo el modelo BoVW, de modo que se logren obtener mejores resultados en recuperación de jeroglíficos respecto a los obtenidos en trabajos previos reportados en la literatura.

1.2. Hipótesis

Es posible obtener una representación para jeroglíficos a partir de características obtenidas desde y entre partes primitivas (puntos, fragmentos o regiones), de modo que dicha representación permita obtener mejores resultados en recuperación de jeroglíficos respecto a los métodos reportados en la literatura.

1.3. Objetivos

El *objetivo general* de esta tesis es:

Proponer una nueva representación de jeroglíficos basada en su descomposición en partes primitivas, tal que, la nueva representación permita obtener resultados en recuperación de jeroglíficos estadísticamente mejores, que aquellos obtenidos previamente en el estado del arte.

Para dar cumplimiento al *objetivo general*, definimos los siguientes *objetivos específicos*:

1. Proponer un método para descomponer jeroglíficos en partes primitivas, tal que, las partes primitivas extraídas proporcionen información útil para la representación de jeroglíficos.
2. Proponer un método para obtener información desde una o múltiples partes primitivas.
3. Proponer un método para organizar la información desde y entre partes primitivas, dando lugar a una nueva representación de jeroglíficos. Tal que, utilizando dicha representación podamos obtener resultados en recuperación de jeroglíficos estadísticamente mejores a los obtenidos en trabajos previos reportados en la literatura.

1.4. Contribuciones

- Una nueva representación de jeroglíficos siguiendo el modelo BoVW, la cual se construye mediante la extracción de fragmentos del jeroglífico. La representación codifica información sobre la distribución y las relaciones de vecindad entre las palabras visuales en los jeroglíficos.

1.5. Descripción del documento

En el capítulo 2 se presentan los conceptos básicos requeridos para entender el resto del documento.

En el capítulo 3 se describen los trabajos más relevantes relacionados con la representación de jeroglíficos en tareas de recuperación.

En el capítulo 4 se describe la representación de jeroglíficos propuesta.

En el capítulo 5 se presentan los resultados experimentales.

Por último, en el capítulo 6 se presentan las conclusiones de este trabajo de investigación y posibles direcciones como trabajo futuro.

CONCEPTOS BÁSICOS

En este capítulo se introducen los conceptos básicos necesarios para entender el resto de este documento.

En la Sección 2.1 se presentan los conceptos que serán utilizados para el podado de jeroglíficos.

Puesto que la representación que se propone como resultado de la presente investigación doctoral está basada en el modelo BoVW, en la Sección 2.2 describimos el modelo BoVW.

En la Sección 2.3 se describen los conceptos generales relacionados con recuperación de información. Además, se definen algunas de las métricas comúnmente utilizadas en la literatura para evaluar esta tarea.

2.1. Forma adelgazada y sus elementos

Para representar un jeroglífico, usualmente se utiliza su forma adelgazada, la cual se define como sigue.

Definición 2.1. (*Forma adelgazada*) Sea G un jeroglífico, su forma adelgazada G' se obtiene adelgazando hasta un pixel de grosor cada línea o trazo de G .

Antes de definir los elementos que pueden ser identificados en una *forma*

adelgazada, definiremos lo que se entenderá en esta tesis por vecinos de un pixel.

Definición 2.2. (*Vecinos de un pixel*) En un jeroglífico, un pixel con coordenadas (k, l) tiene dos vecinos verticales, dos vecinos horizontales y cuatro vecinos diagonales. Estos vecinos son los siguientes

$$\begin{array}{l} \text{Vecinos verticales} \left\{ \begin{array}{l} (k-1, l) \\ (k+1, l) \end{array} \right. \\ \text{Vecinos horizontales} \left\{ \begin{array}{l} (k, l-1) \\ (k, l+1) \end{array} \right. \\ \text{Vecinos diagonales} \left\{ \begin{array}{l} (k-1, l-1) \\ (k-1, l+1) \\ (k+1, l-1) \\ (k+1, l+1) \end{array} \right. \end{array}$$

A continuación definimos algunos de los elementos que pueden ser identificados en una forma adelgazada.

Sea G' la forma adelgazada de un jeroglífico, entonces:

Definición 2.3. (*Pixel final*) Un pixel es un *pixel final* si tiene un único vecino en G' .

Definición 2.4. (*Pixel de unión*) Un pixel es un *pixel de unión* si tiene tres o más vecinos en G' .

Definición 2.5. (*Pixel de curva*) Un pixel es un *pixel de curva* si tiene exactamente dos pixeles vecinos en G' .

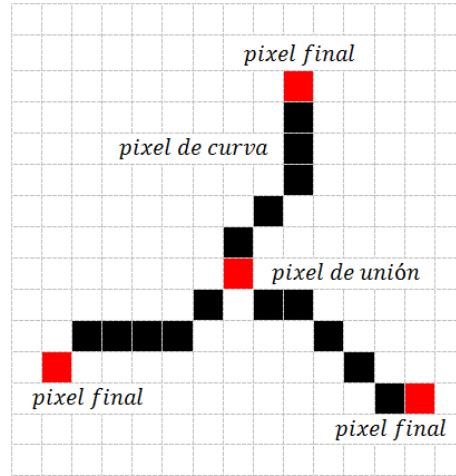


Figura 2.1: Tipos de píxeles en una forma adelgazada de un jeroglífico: *píxeles finales* y *pixel de unión* marcados en rojo, *píxeles de curva* marcados en negro.

En la Figura 2.1 se muestra una forma adelgazada, donde los píxeles en rojo corresponden a *píxeles finales* y *píxeles de unión*, y los píxeles en negro son *píxeles de curva*.

Definición 2.6. (*rama*) Una rama en G' , denotada como b , corresponde a una secuencia de N píxeles $b = \{p_1, \dots, p_N\}$, tal que p_i es vecino de p_{i+1} , para $i = 1, \dots, N - 1$. Además, los píxeles desde p_2 hasta p_{N-1} en b , son estrictamente *píxeles de curva*.

En una forma adelgazada G' , las ramas pueden ser de diferente tipo.

Definición 2.7. (*Rama simple*) Una rama b es una *rama simple* si ambos p_1 y p_N son *píxeles finales* (ver Figura 2.2a).

Definición 2.8. (*Rama final*) Una rama b es una *rama final* si p_1 es un *pixel de unión* y p_N es un *pixel final*, o viceversa (ver Figura 2.2b).

Definición 2.9. (*Rama interna*) Una rama b es una *rama interna* si ambos p_1 y p_N son *píxeles de unión* (ver Figura 2.2c)

Definición 2.10. (*Rama cerrada*) Una rama b es una *rama cerrada* si $p_1 = p_N$ (ver Figura 2.2d).

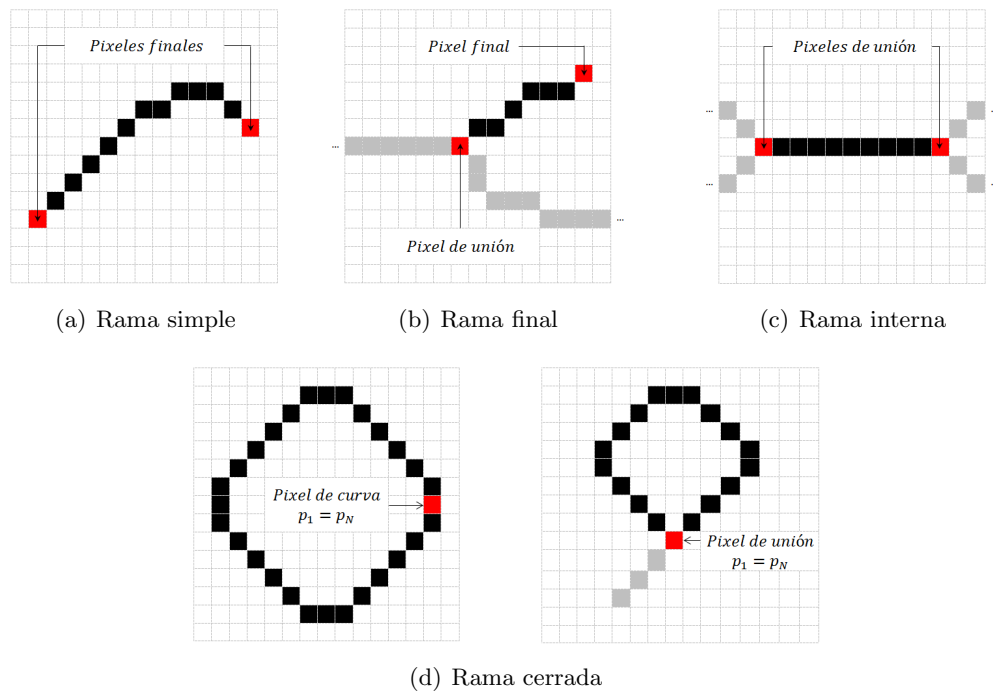


Figura 2.2: Tipos de ramas en una forma adelgazada de un jeroglífico: Los píxeles en rojo corresponden a los píxeles p_1 y p_N , mientras los píxeles en negro corresponden a los píxeles p_2, \dots, p_{N-1} .

Para algunos de los métodos propuestos en este trabajo es necesario el cálculo de la transformada de distancia para los píxeles de un jeroglífico G , la cual se define como sigue:

Definición 2.11. (*Transformada de distancia*) la transformada de distancia para un jeroglífico G , denotada como DT^G , asocia a cada píxel del jeroglífico (i.e. *foreground*), con coordenadas (k, l) , la distancia mínima desde dicho píxel hasta el píxel más cercano del *background* [Haralick and Shapiro, 1992]. Cualquier distancia puede ser utilizada, la más común es la distancia Euclidiana.

2.2. Representación bajo el modelo BoVW

En la literatura, la representación de jeroglíficos bajo el modelo BoVW ha sido ampliamente utilizada en tareas de recuperación debido a su simplicidad y efectividad [Gatica-Perez et al., 2012; Hu et al., 2014; Roman-Rangel and Marchand-Maillet, 2013a, 2016; Roman-Rangel et al., 2011a]. El modelo BoVW permite representar el contenido del jeroglífico (codificado a través de descriptores locales) utilizando un conjunto finito de patrones denominados palabras visuales [Sivic and Zisserman, 2003]. La metodología comúnmente utilizada para representar jeroglíficos bajo el modelo BoVW trabaja como se describe a continuación: Dado un jeroglífico, a partir de este, normalmente se obtiene su forma adelgazada. Después, se selecciona un conjunto de puntos de interés a menudo mediante selección aleatoria uniforme sobre la forma adelgazada del jeroglífico. Para cada punto de interés se calcula un descriptor local (e.g. SC del inglés *Shape-Context* [Belongie et al., 2000, 2002], GSC del inglés *Generalized Shape-Context* [Mori et al., 2005], HoG del inglés *Histogram of Oriented Gradients* [Dalal and Triggs, 2005], HOOSC del inglés *Histogram of Orientation Shape-Context* [Roman-Rangel et al., 2011a]). Los descriptores locales extraídos a partir de jeroglíficos en el conjunto de entrenamiento son agrupados en c -grupos, usualmente utilizando c -means [Hartigan and Wong, 1979]. El centroide de cada grupo, representa a todos los descriptores locales que caen en el grupo y es considerado como una palabra visual en el vocabulario. Haciendo uso del vocabulario visual, cada descriptor local es reemplazado por la palabra visual más cercana en el vocabulario. Finalmente, la representación de un jeroglífico bajo el modelo BoVW es obtenida construyendo un vector que contiene en cada entrada, la frecuencia de una palabra visual en el jeroglífico. En la Figura 2.3 se ilustra el esquema general para representar un jeroglífico bajo el modelo BoVW.

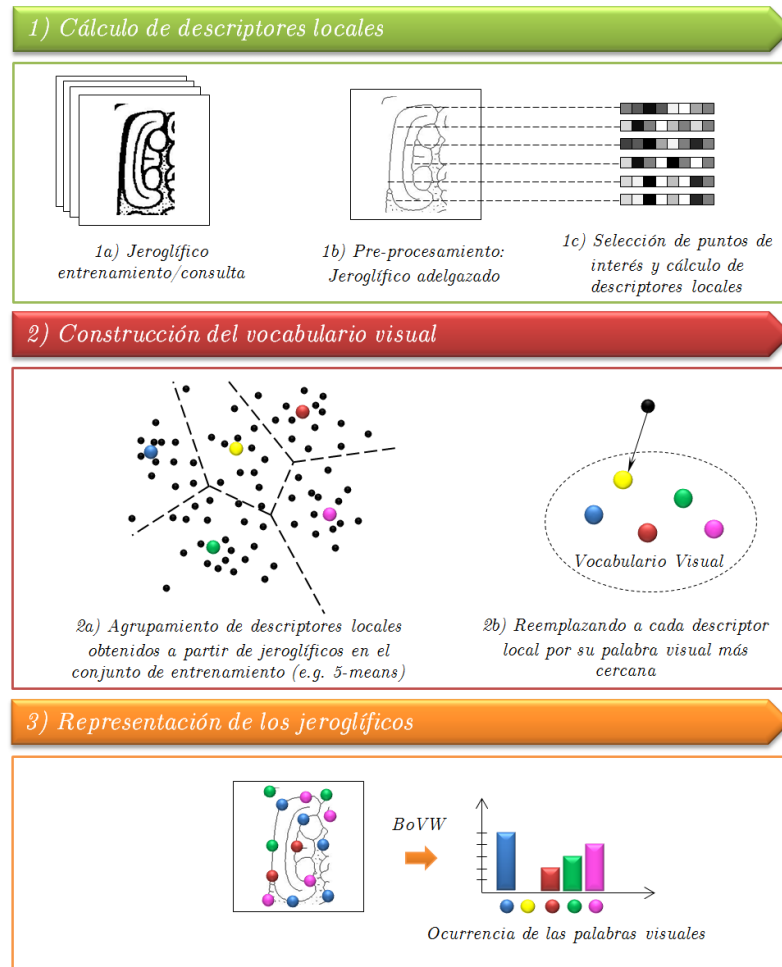


Figura 2.3: Representación de jeroglíficos bajo el modelo BoVW: **Etapa 1.** Extracción de descriptores locales. **Etapa 2.** Construcción del vocabulario visual. **Etapa 3.** Representación del jeroglífico.

2.3. Recuperación de Información

El término “recuperación de información” se refiere a encontrar en colecciones de datos, información relevante que satisface algún criterio de búsqueda.

En este trabajo, nos limitamos a la recuperación de imágenes basada en

contenido (CBIR del inglés *Content-Based Image Retrieval*) [Datta et al., 2008; Smeulders et al., 2000], específicamente en recuperación de jeroglíficos . El mecanismo de búsqueda utilizado en esta tesis para recuperar jeroglíficos consiste de cuatro etapas:

1. Construir una representación vectorial para cada uno de los jeroglíficos en la colección de datos.
2. Representar al jeroglífico de consulta, de manera similar en que fueron representados los jeroglíficos en la colección de datos.
3. Calcular la similitud entre el jeroglífico de consulta y cada jeroglífico de la colección, considerando la representación obtenida en (1) y (2).
4. Ordenar los jeroglíficos de la colección y presentarlos al usuario en orden descendente de acuerdo al parecido con el jeroglífico de consulta.

Para calcular la similitud entre dos jeroglíficos se pueden utilizar diferentes métricas de distancia. En este trabajo, se utilizará la distancia de Manhattan.

Dado un jeroglífico *query* q y un jeroglífico p en la colección, representados por sus vectores h^q y h^p , de dimension k , la distancia de Manhattan para medir la similitud entre q y p :

$$d(q, p) = \sum_{1 \leq i \leq k} |h^q(i) - h^p(i)| \quad (2.1)$$

La calidad de la recuperación se evalúa mediante diferentes métricas. A continuación, definiremos las métricas comúnmente utilizadas en la literatura para evaluar la recuperación de jeroglíficos [Baeza-Yates and Ribeiro-Neto, 1999]:

Definición 2.12. (*Precisión*) La *precisión* es la fracción de jeroglíficos recuperados

que son relevantes al jeroglífico de búsqueda.

$$precision = \frac{\text{Número de jeroglíficos relevantes recuperados}}{\text{Número total de jeroglíficos recuperados}} \quad (2.2)$$

Definición 2.13. (*Recuerdo*) El *recuerdo* es la fracción de jeroglíficos relevantes para el jeroglífico de consulta *query* que fueron recuperados.

$$recall = \frac{\text{Número de jeroglíficos relevantes recuperados}}{\text{Número total de jeroglíficos relevantes disponibles}} \quad (2.3)$$

Definición 2.14. (*precision at n*) La *precisión at n* ($P@n$) es la fracción de jeroglíficos recuperados en el rango n que son relevantes al jeroglífico de búsqueda.

$$p@n = \frac{\text{Número de jeroglíficos relevantes recuperados}}{n} \quad (2.4)$$

Definición 2.15. (*Precisión promedio*) La *precisión promedio* (AP) es el promedio de la precisión alcanzada por los jeroglíficos recuperados que son relevantes al jeroglífico de consulta, donde los jeroglíficos recuperados están ordenados en función del valor de similitud.

$$AP = \frac{\sum_{n=1}^K (p@n * rel(n))}{K} \quad (2.5)$$

donde K es el número total de jeroglíficos recuperados y $rel(n)$ se define como:

$$rel(n) = \begin{cases} 1 & \text{Si el } n\text{-ésimo elemento recuperado es relevante,} \\ 0 & \text{en otro caso} \end{cases}$$

Definición 2.16. (*Media de la precisión promedio*) La *media de la precisión promedio* (mAP) para un conjunto de jeroglíficos de consulta es el promedio de

las puntuaciones de *precisión promedio* para cada consulta.

$$mAP = \frac{\sum_{q=1}^Q AP(q)}{Q} \quad (2.6)$$

Donde Q es el número de consultas.

TRABAJO RELACIONADO

En este capítulo presentamos una revisión del estado del arte, describiendo los trabajos directamente relacionados con la representación de jeroglíficos para tareas de recuperación. En la literatura, los trabajos que llevan a cabo tareas de recuperación de jeroglíficos, representan a los jeroglíficos siguiendo principalmente dos enfoques: (1) Representación basada en descriptores y (2) representación bajo el modelo BoVW. En la Figura 3.1, mostramos a través de un esquema los trabajos en la literatura que siguen cada uno de esos enfoques.

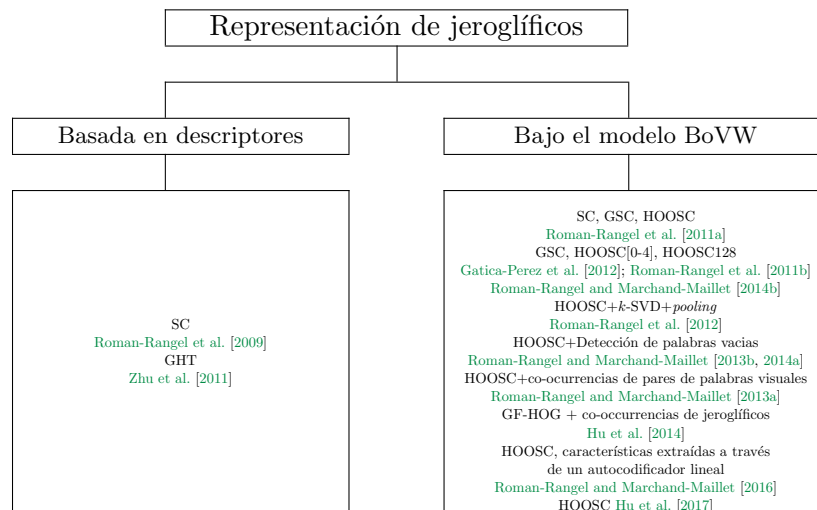


Figura 3.1: Representación para la recuperación de jeroglíficos en la literatura.

La representación de jeroglíficos basada en descriptores se realiza a través del cálculo de descriptores *locales* o *globales* [Roman-Rangel et al., 2009; Zhu et al., 2011]. Zhu et al. [2011] utilizan GHT (*Generalized Hough Transform*) [Ballard, 1981; Hough, 1962] para representar petroglifos, e introducen una novedosa métrica de distancia que permite de manera efectiva y eficiente calcular la similitud entre dos petroglifos. Roman-Rangel et al. [2009] investigan la efectividad del descriptor SC (*Shape Context*) [Belongie et al., 2000, 2002] para la representación de jeroglíficos Mayas, y proponen una mejora en la función de costo utilizada para calcular la similitud entre los jeroglíficos.

La representación de jeroglíficos bajo el modelo BoVW se construye haciendo uso de descriptores locales junto con el modelo BoVW [Sivic and Zisserman, 2003]. En la literatura, la representación de jeroglíficos bajo el modelo BoVW, ha mostrado ser la más efectiva y simple en tareas de recuperación [Gatica-Perez et al., 2012; Hu et al., 2014; Roman-Rangel and Marchand-Maillet, 2013a; Roman-Rangel et al., 2011a]. Por lo tanto, en la Sección 3.1 describimos detalladamente aquellos trabajos que representan jeroglíficos con propósitos de recuperación siguiendo el modelo BoVW. Finalmente, en la Sección 3.2 presentamos una discusión sobre el trabajo relacionado.

3.1. Trabajos que representan jeroglíficos bajo el modelo BoVW

En [Roman-Rangel et al., 2011a] la representación de jeroglíficos Mayas se construye bajo el modelo BoVW usando los descriptores SC, GSC y HOOSC, este último descriptor propuesto por los autores, combinando los descriptores SC y HoG. A partir del contorno del jeroglífico se extrae un conjunto de puntos de interés mediante selección aleatoria uniforme. El conjunto de puntos de interés

corresponde a un porcentaje del total de puntos en el contorno. Para cada punto de interés se calcula un descriptor local (i.e. SC, GSC, ó HOOSC) en función de los puntos en el conjunto de puntos de interés. Una vez se han calculado los descriptores locales, la representación de los jeroglíficos se construye siguiendo el esquema general del modelo BoVW. En sus experimentos, los autores muestran que el uso de descriptores HOOSC bajo el modelo BoVW en tareas de recuperación de jeroglíficos supera a los resultados obtenidos utilizando descriptores SC o GSC.

En la literatura se han propuesto algunas mejoras a la representación introducida en [Roman-Rangel et al., 2011a], las cuales consisten en: 1) Obtener la forma adelgazada de los jeroglíficos antes de seleccionar los puntos de interés [Gatica-Perez et al., 2012; Roman-Rangel et al., 2011b], lo anterior, con el objetivo de eliminar contornos dobles que introducen ruido en la representación. 2) Roman-Rangel et al. [2011b] muestran que se obtiene una descripción más precisa, cuando para cada punto de interés el descriptor local HOOSC es calculado en función de todos los puntos en el contorno del jeroglífico, en lugar de considerar únicamente a los puntos en el conjunto de puntos de interés. 3) Utilizar una configuración diferente en el diagrama logarítmico-polar utilizado en HOOSC, prestando atención a las regiones en el diagrama que contienen la mayor porción de información del jeroglífico (las regiones más próximas alrededor de cada punto de interés). Lo anterior da lugar a descriptores HOOSC de menor dimensión y con mayor poder discriminativo [Roman-Rangel and Marchand-Maillet, 2014b; Roman-Rangel et al., 2011b]. En [Roman-Rangel et al., 2011b] y [Roman-Rangel and Marchand-Maillet, 2014b] los mejores resultados en recuperación se obtienen con los descriptores denotados como HOOSC4 y HOOSC128, respectivamente. 4) Roman-Rangel et al. [2011b] proponen incorporar explícitamente al final de cada descriptor local HOOSC la posición relativa del punto de interés (en el intervalo $[0, 1]$) a partir del cual, el descriptor local está siendo calculado, con el objetivo

de reforzar la habilidad que tiene el descriptor para capturar información sobre la distribución de los puntos de interés en el jeroglífico

En [Roman-Rangel et al., 2012] se utiliza un método de codificación *sparse* llamado k -SVD [Aharon et al., 2006]. En lugar de reemplazar a cada descriptor local por su palabra visual más cercana en el vocabulario, el método k -SVD codifica cada descriptor local HOOSC mediante una combinación lineal de las palabras visuales en el vocabulario que mejor representan a dicho descriptor local (llamadas funciones base). Donde un parámetro define el número máximo de funciones base a ser consideradas. k -SVD regresa para cada descriptor local un vector, donde cada coeficiente en dicho vector refleja el grado de similitud del descriptor local con cada una de sus funciones base, utilizando Manhattan como métrica de distancia. Para obtener la representación final de cada jeroglífico se hace uso de diferentes técnicas de *pooling*. Las técnicas de *pooling* evaluadas fueron: AVP (*Average Pooling*) [Boureau et al., 2010a], Max-NWP (*Max-N Weight Pooling*) [Rigamonti et al., 2011], Max-NBP (*Max-N Binary Pooling*) [Boureau et al., 2010b] y Max-NIP (*Max-N Integer Pooling*) [Boureau et al., 2010a]. En términos generales, los resultados obtenidos de la combinación de k -SVD y *pooling* son solamente competitivos a los obtenidos cuando los jeroglíficos son representados mediante el simple conteo de ocurrencias de las palabras visuales en el jeroglífico.

En [Roman-Rangel and Marchand-Maillet, 2013b, 2014a] se introduce un método para detección de palabras vacías, mediante una métrica que calcula la entropía de las palabras visuales en el vocabulario. Las palabras visuales con mayor entropía son consideradas palabras irrelevantes y eliminadas del vocabulario. La detección de palabras vacías permite reducir la dimensión del vector que representa a cada jeroglífico, manteniendo la calidad de la representación. Roman-Rangel and Marchand-Maillet [2014a] presenta una metodología, a través de la cual se estima automáticamente la cantidad de palabras visuales que pueden ser eliminadas del

vocabulario visual sin que la calidad de la representación, ni el rendimiento en recuperación de jeroglíficos se vea afectado.

En [Roman-Rangel and Marchand-Maillet, 2013a], la representación resultante no solo toma en cuenta la frecuencia de las palabras visuales en el jeroglífico, sino que además considera las relaciones entre palabras visuales, contando el número de veces en que palabras visuales co-ocurren en una vecindad. Esto da lugar a una representación de jeroglíficos que permite alcanzar mejores resultados en recuperación respecto a los obtenidos cuando los jeroglíficos son representados con SPM (*Spatial Pyramid Matching*) [Lazebnik et al., 2006], n -gramas [Zhang and Mayo, 2010] y el modelo BoVW general como en [Roman-Rangel et al., 2011a; Sivic and Zisserman, 2003].

Hu et al. [2014, 2015] proponen un método de recuperación automática de jeroglíficos que integra información de la forma del jeroglífico e información del contexto del jeroglífico según su distribución en un bloque. Un bloque es una porción de un texto antiguo que contiene más de un jeroglífico. Para codificar información del contexto, cada bloque se convierte a una cadena de Markov de primer orden y se obtiene información de las co-ocurrencias entre pares de jeroglíficos vecinos en el bloque [Thompson, 1962; Vail and Hernández]. La recuperación de jeroglíficos se lleva a cabo haciendo uso de los descriptores HOOSC y GF-HoG [Hu and Collomosse, 2013] (*Gradient Field Histogram of Orientation Gradient*), cada uno bajo el modelo BoVW junto con la información del contexto. Los resultados en términos de recuperación muestran que introducir información del contexto, permite que se obtengan mejores resultados en recuperación de jeroglíficos, respecto a los obtenidos cuando se hace uso únicamente del modelo BoVW.

En [Roman-Rangel and Marchand-Maillet, 2016], cada jeroglífico es adelgazado y un conjunto de puntos de interés se obtiene mediante selección aleatoria uniforme.

A partir de cada punto de interés se extrae una región de $T \times T$ píxeles con centro en cada punto de interés, siendo T la distancia promedio entre todos los puntos de interés. A partir de esas regiones, los descriptores locales son calculados a través de un auto-codificador (auto-encoder). Una vez los descriptores locales son obtenidos, éstos se utilizan bajo el modelo BoVW, representando a cada jeroglífico según la frecuencia de cada palabra visual. Los resultados alcanzados en términos de recuperación usando los descriptores locales extraídos con el auto-codificador no logran superar a los obtenidos utilizando HOOSC, ambos bajo el modelo BoVW.

En [Hu et al., 2017], Los jeroglíficos son represados mediante el calculo de descriptores locales HOOSC en jeroglíficos adelgazados, junto con el modelo BoVW. Para cada jeroglífico de consulta se regresa una lista ordenada de los jeroglíficos en la base de datos, para los cuales se obtuvieron los mayores valores de similitud. Este trabajo es similar a los presentados en [Hu et al., 2014; Roman-Rangel et al., 2011a].

3.2. Discusión

A partir de la revisión de los trabajos reportados en la literatura sobre representación de jeroglíficos para tareas de recuperación, podemos observar que en la mayoría de estos trabajos [Gatica-Perez et al., 2012; Hu et al., 2014; Roman-Rangel and Marchand-Maillet, 2013b, 2014b, 2016; Roman-Rangel et al., 2011b, 2012] fue obtenida la forma adelgazada para cada jeroglífico, ya que haciendo uso de esta se logran obtener mejores resultados en recuperación de jeroglíficos respecto a los obtenidos a partir del contorno sin adelgazar o cuando se hace uso de detectores de bordes.

Por otra parte, la mayoría de las representaciones de jeroglíficos se construyen a partir de características obtenidas desde puntos de interés extraídos mediante selec-

ción aleatoria uniforme [Gatica-Perez et al., 2012; Hu et al., 2014; Roman-Rangel and Marchand-Maillet, 2014b; Roman-Rangel et al., 2011a,b]. Esta selección ha mostrado ser mejor, ya que la representación resultante permite obtener mejores resultados en recuperación de jeroglíficos, respecto a cuando se utilizan puntos de interés extraídos mediante métodos de detección de puntos de interés [Roman-Rangel and Marchand-Maillet, 2015; Roman-Rangel et al., 2013] tales como DoG (Difference of Gaussian)[Lowe, 2004] ó Harris-Laplace [Mikolajczyk and Schmid, 2004],

Debido al éxito que ha tenido la representación de jeroglíficos bajo el modelo BoVW, la mayoría de trabajos en la literatura (ver Figura 3.1) siguen este enfoque. No obstante, la mayoría de estos trabajos únicamente han estudiado el uso de diferentes tipos de descriptores locales. Sin embargo, como ya se mencionó, la representación de jeroglíficos bajo el modelo BoVW tiene varias limitaciones, siendo muy pocos los trabajos en la literatura que han intentado resolverlas [Roman-Rangel and Marchand-Maillet, 2013a; Roman-Rangel et al., 2012]. Entre las limitaciones del modelo BoVW podemos mencionar:

- El reemplazo de cada descriptor local por su palabra visual más cercana en el vocabulario, puede dar lugar a que descriptores locales muy parecidos sean reemplazados por palabras visuales diferentes, lo que puede producir que jeroglíficos parecidos tengan representaciones diferentes.
- La representación para cada jeroglífico resulta del conteo de ocurrencias de las palabras visuales, es decir, no se considera información sobre las relaciones de vecindad ni sobre la distribución de las palabras visuales dentro de el jeroglífico.

Por lo tanto, en esta investigación doctoral se propondrán mejoras al modelo BoVW que contribuyan a mejorar la calidad de la representación de jeroglíficos,

de modo que se puedan obtener mejores resultados en recuperación de jeroglíficos respecto a los obtenidos en los trabajos del estado del arte.

REPRESENTACIÓN DE JEROGLÍFICOS PROPUESTA

En este capítulo se introduce la representación de jeroglíficos propuesta en esta investigación, la cual es construida siguiendo el modelo BoVW (mostrado en la Figura 2.3), pero introduciendo mejoras en cada una de sus tres etapas.

La primera etapa del modelo BoVW corresponde a la extracción de descriptores locales. Un paso que generalmente se considera parte de esta etapa consiste en obtener la forma adelgazada del jeroglífico. No obstante, esta forma adelgazada usualmente contiene ramas no deseadas producto de perturbaciones, variaciones o ruido presente en el jeroglífico. Incluir estas ramas durante el cálculo de los descriptores locales podría afectar negativamente la representación de los jeroglíficos. Con el objetivo de eliminar las ramas no deseadas, en la Sección 4.1 se propone que éstas sean podadas de la forma adelgazada de los jeroglíficos.

Una vez que la forma adelgazada ha sido obtenida, los descriptores locales son usualmente calculados a partir de un conjunto de puntos de interés. No obstante, la forma del jeroglífico resulta básicamente de diferentes trazos que se intersectan entre sí. Por lo tanto, en la sección 4.2 se propone que los descriptores locales sean calculados a partir de fragmentos de la forma adelgazada.

En la segunda etapa del modelo BoVW cada descriptor local es reemplazado por

la palabra visual más cercana de un vocabulario visual, es decir, su palabra visual más parecida en el vocabulario. No obstante, reemplazar a cada descriptor local por su palabra más cercana en el vocabulario puede resultar en que descriptores locales parecidos sean reemplazados por palabras visuales diferentes, lo que podría afectar la representación de los jeroglíficos. Para enfrentar este problema, en la Sección 4.3 se propone reemplazar a cada descriptor local por sus k -palabras visuales más cercanas en el vocabulario.

En la tercera etapa, la representación del jeroglífico se obtiene mediante el conteo de ocurrencias de las palabras visuales. No obstante, el número de ocurrencias de las palabras visuales para dos jeroglíficos diferentes podría ser muy parecido o el mismo, ya que la disposición de las palabras visuales en el jeroglífico no es considerada en el cálculo de la representación. Una manera de considerar la disposición de las palabras visuales en el jeroglífico es mediante el cálculo de co-ocurrencias de palabras visuales [Roman-Rangel and Marchand-Maillet, 2013a]. No obstante, en la literatura, las co-ocurrencias han sido calculadas considerando a la palabra visual más cercana que reemplazó a cada descriptor local. En la Sección 4.3 se propone introducir el reemplazo de cada descriptor local por sus k -palabras visuales más cercanas en el cálculo de las co-ocurrencias y la frecuencia de las palabras visuales, para construir la representación de los jeroglíficos.

En la Figura 4.1 se ilustra el esquema general para la representación de jeroglíficos propuesta, donde las modificaciones introducidas en cada una de las tres etapas del modelo BoVW están marcadas con rectángulo rojo (Figura 2.3).

4.1. Podado de jeroglíficos adelgazados

En la literatura, los jeroglíficos son generalmente adelgazados[Frauel et al., 2006; Guo et al., 2016; Roman-Rangel et al., 2011b]. No obstante, como ya se

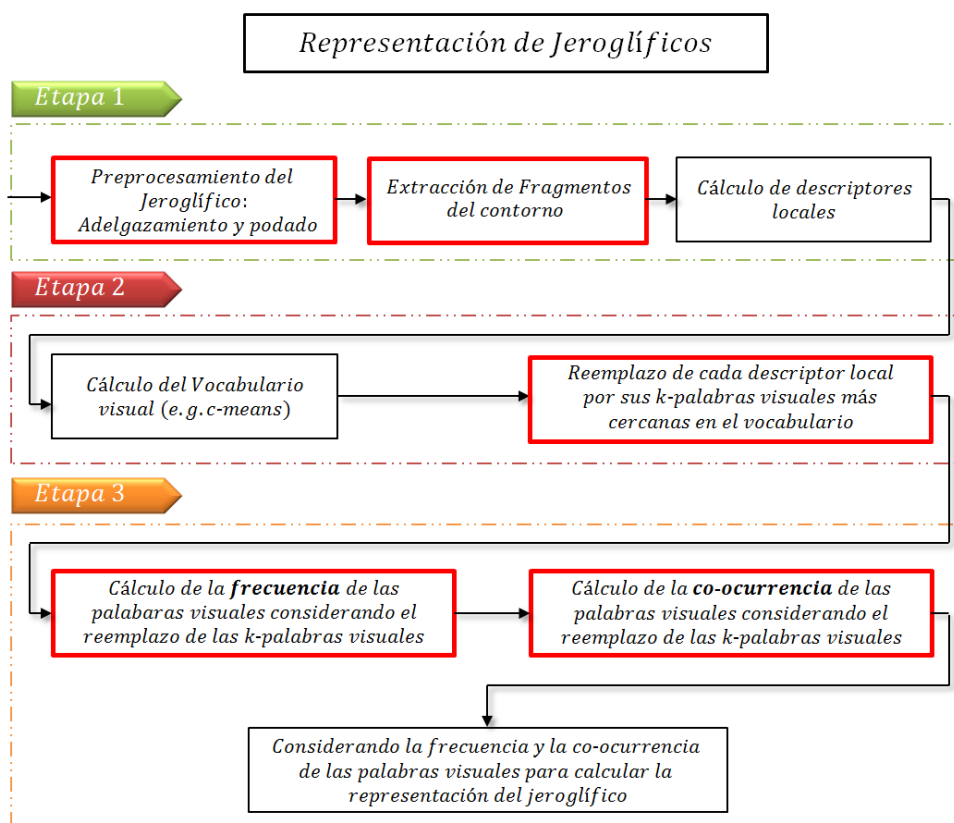


Figura 4.1: Nueva representación de jeroglíficos propuesta siguiendo el modelo BoVW.

mencionó, la forma adelgazada de un jeroglífico puede contener ramas no deseadas producto de perturbaciones en el jeroglífico. En la Figura 4.2, se muestra un jeroglífico y su forma adelgazada. Como se puede apreciar, la forma adelgazada contiene múltiples ramas no deseadas (ramas marcadas en rojo), las cuales no representan información útil del jeroglífico. Por lo tanto, es necesario eliminar estas ramas no deseadas, de modo que se pueda evitar la mayor cantidad de información ruidosa que pueda afectar la representación de los jeroglíficos.

En este trabajo proponemos podar las ramas no deseadas, para ello se ana-

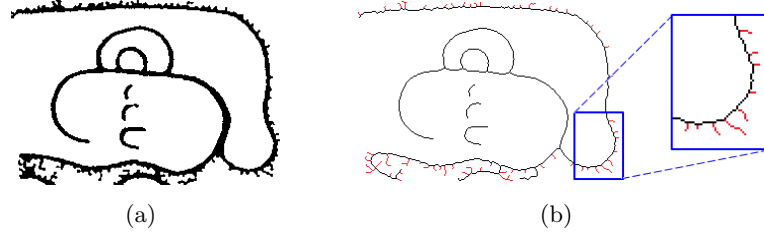


Figura 4.2: (a) Jeroglífico Maya. (b) Forma adelgazada obtenida a partir del jeroglífico Maya. Las ramas no deseadas están marcadas en rojo.

lizarán las *ramas simples* (ver definición 2.7) y las *ramas finales* (ver definición 2.8) de la forma adelgazada del jeroglífico. La poda de una rama dependerá de la proporción del área reconstruida a partir de la rama, respecto al área total del jeroglífico antes de obtener su forma adelgazada (*foreground* del jeroglífico). Es decir, se eliminan las ramas simples y ramas finales cuya área reconstruida represente una pequeña proporción del área total del jeroglífico original. Para esto se utiliza un umbral, el cual indica el área mínima que debe poder reconstruirse a partir de una rama para que sea considerada importante y que por lo tanto, no sea podada.

La región reconstruida a partir de una rama se obtiene de las regiones reconstruidas a partir de cada uno de los pixeles de la rama. La región reconstruida a partir de un pixel p_i (con coordenadas (k, l)), denotada como $r(p_i)$, corresponde a todos los pixeles dentro del disco con origen en (k, l) y radio $DT^G(k, l)$.

Sea G un jeroglífico, G' su forma adelgazada y $b = \{p_1, \dots, p_N\}$ una rama de G' . Si b es una *rama simple*, la región reconstruida a partir de b , denotada como $R(b)$, resulta de la unión de las regiones reconstruidas a partir de cada pixel $p_i \in b$. La región reconstruida a partir de una *rama final* b , resulta de la unión de las regiones reconstruidas a partir de cada pixel $p_i \in b$, menos la región reconstruida a partir de p_1 , siendo p_1 el pixel de unión. La región reconstruida a partir de p_1

no es considerada, ya que p_1 forma parte de otras ramas en la forma adelgazada del jeroglífico, y por lo tanto, ese pixel no será eliminado.

La región reconstruida a partir de una *rama simple* se define en la Ecuación 4.1

$$R(b) = \bigcup_{i=1}^N r(p_i) \quad (4.1)$$

La región reconstruida a partir de una *rama final* se define en la Ecuación 4.2

$$R(b) = \left[\bigcup_{i=2}^N r(p_i) \right] - r(p_1) \quad (4.2)$$

En la Figura 4.3 se ilustra la región reconstruida (región sombreada) a partir de una *rama simple* (Figura 4.3a) y una *rama final* (Figura 4.3b). En esta figura, los pixeles p_1 y p_N en las ramas se muestran en rojo y los pixeles p_2, \dots, p_{N-1} en verde. Para facilitar la ilustración de la región reconstruida a partir de una rama, se muestra únicamente la región reconstruida para algunos pixeles en cada una de las ramas. En la Figura 4.3b la región reconstruida a partir de p_1 no está sombreada, ya que no forma parte de la región que no podría ser reconstruida en caso de que dicha rama fuese eliminada, ya que al ser p_1 un pixel de unión, este no sería eliminado en el podado de la rama.

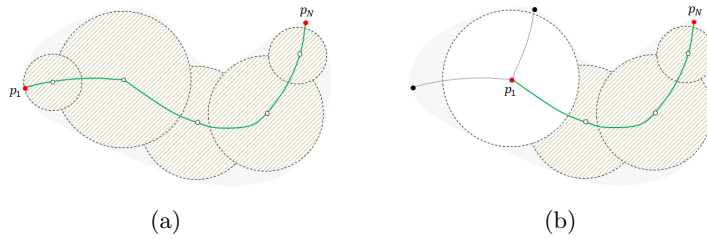


Figura 4.3: Región reconstruida a partir de una (a) rama simple y (b) rama final.

Una rama b en la forma adelgazada G' de un jeroglífico G es no deseada y

debe ser eliminada, si el número de píxeles en $R(b)$ es menor a un porcentaje α del número de píxeles en el jeroglífico G , donde α es definido por el usuario.

Eliminar simultáneamente todas las ramas finales no deseadas podría conducir a que la forma adelgazada que representa a una región importante sea eliminada en su totalidad o casi por completo. Lo anterior sucede en aquellas regiones representadas a través de una forma adelgazada que contiene múltiples ramas finales no deseadas compartiendo un mismo pixel de unión. En la Figura 4.4 se ilustra este caso, donde la forma adelgazada obtenida a partir de una región está formada por tres ramas finales etiquetadas como k , l y m , las cuales comparten un mismo pixel de unión, marcado en rojo. Si k , l y m son ramas finales no deseadas, su eliminación simultánea daría lugar a la pérdida casi total de la forma adelgazada que representa a dicha región (ver Figura 4.4b). Por el contrario, si únicamente la rama final no deseada con la menor región reconstruida es eliminada, es decir, la rama final etiquetada como l , las ramas resultantes etiquetadas como k y m dan lugar a una rama simple, a partir de la cual se puede obtener una mayor región reconstruida, que las regiones reconstruidas desde cada una de las ramas finales en la Figura 4.4a, de modo que es más probable que la forma adelgazada formada por la rama simple (ver Figura 4.4c) sea preservada para representar a dicha región.

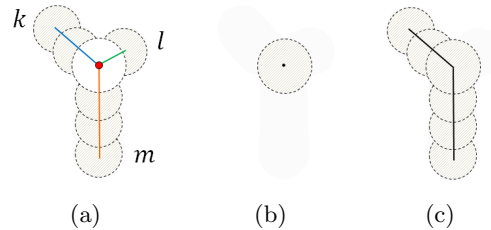


Figura 4.4: Podado de ramas finales en la forma adelgazada: (a) Forma adelgazada, (b) eliminación de todas las ramas finales, (c) rama simple formada por las ramas preservadas al eliminar solamente la rama con menor contribución.

Considerando lo anterior, para la eliminación de las ramas no deseadas en la forma adelgazada de un jeroglífico se propone llevar a cabo la poda de manera iterativa. En cada iteración se eliminan las ramas no deseadas que son ramas simples y ramas finales. En el caso de ramas finales no deseadas que comparten el mismo pixel de unión, únicamente la rama final no deseada con la menor región reconstruida es eliminada. Lo anterior se repite hasta que en una iteración ninguna rama sea eliminada.

En el algoritmo 1 se presenta el pseudocódigo para llevar a cabo la poda de jeroglíficos propuesta.

Algorithm 1: prunThinHieroglyph(G')

Input: G' : Jeroglífico adelgazado
Output: G' : Jeroglífico adelgazado y podado

```

1 repeat
2   foreach rama final o rama simple  $b$  en  $G'$  do
3     if  $b$  es una rama no deseada then
4       if  $b$  es una rama simple then
5          $G' \leftarrow$  eliminar de  $G'$  la rama simple  $b$ ;
6       else
7         //  $b$  es una rama final;
8         if  $b$  comparte su pixel de unión con otras ramas no deseadas en  $G'$  then
9            $G' \leftarrow$  eliminar de  $G'$  la rama final no deseada con la menor región
              reconstruida, entre las que comparten el pixel de unión con  $b$ ;
10        else
11           $G' \leftarrow$  eliminar de  $G'$  la rama final  $b$ ;
12 until ninguna rama sea eliminada;
```

En la Figura 12 se ilustra la forma adelgazada de un jeroglífico y la forma adelgazada después de aplicar el podado. Puede notarse que las ramas no deseadas producto de perturbaciones y ruido en el jeroglífico fueron eliminadas, preservando los trazos que representan la información más importante del jeroglífico a través de su forma adelgazada.

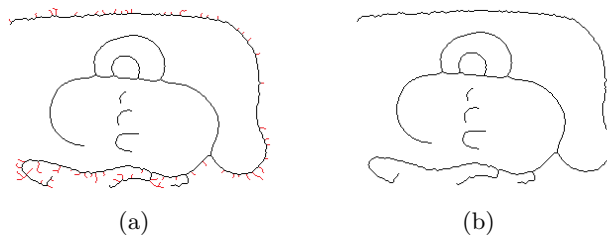


Figura 4.5: Jeroglífico: (a) Forma adelgazada, (b) Forma adelgazada y podada.

4.2. Extracción de descriptores locales a partir fragmentos del jeroglífico

En la literatura, los trabajos que representan jeroglíficos siguiendo el modelo BoVW utilizan descriptores locales calculados a partir de puntos de interés (o regiones, e.g. *grids*, cuadrículas). Estos descriptores resultan al codificar la información del jeroglífico completo respecto a cada punto de interés [Belongie et al., 2002; Roman-Rangel et al., 2011a] (ver Figura 4.6a). No obstante, un jeroglífico está formado básicamente por trazos, muchos de estos intersectados entre sí. Por este motivo, proponemos una representación de jeroglíficos a partir de fragmentos de la forma adelgazada del jeroglífico, donde cada descriptor local resulta al codificar únicamente la información del fragmento (ver Figura 4.6b).

Según nuestro conocimiento, ningún trabajo en la literatura representa jeroglíficos utilizando descriptores locales calculados desde fragmentos del jeroglífico adelgazado. No obstante, proponer una estrategia que permita extraer fragmentos del jeroglífico es una tarea no trivial, ya que no es sencillo determinar automáticamente dónde comienza y termina cada fragmento. Por lo tanto, proponemos una forma de extraer fragmentos del jeroglífico.

La extracción de fragmentos propuesta inicia con la selección de un subconjunto

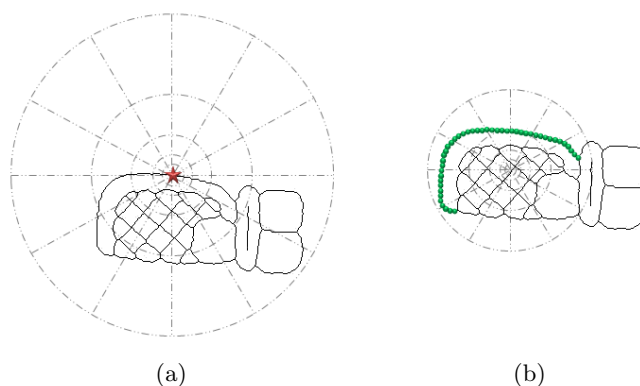


Figura 4.6: a) Descriptor de un punto de interés: el descriptor local resulta de codificar la información del jeroglífico completo respecto al punto de interés (marcado con la estrella roja). b) Descriptor de un fragmento del jeroglífico: el descriptor local resulta de codificar únicamente los píxeles del fragmento del jeroglífico (marcado en verde).

de píxeles de la forma adelgazada y podada del jeroglífico G' . Este subconjunto de píxeles, denotado como P' , se usará para extraer fragmentos del jeroglífico. P' contendrá al menos un porcentaje γ de los píxeles en G' , siendo γ un parámetro definido por el usuario. Inicialmente P' contiene a los *píxeles finales* y *píxeles de unión* en G' . Si el número de píxeles en P' es menor al porcentaje especificado γ del número de píxeles en G' , otros píxeles son seleccionados de G' y agregados a P' hasta que el porcentaje γ haya sido alcanzado. Estos píxeles son seleccionados de las ramas en G' de manera proporcional a la longitud de cada rama¹ respecto al número de píxeles en G' . Una vez calculado el número de píxeles a seleccionar desde cada rama, éstos son seleccionados equidistantemente a lo largo de cada rama.

Con el objetivo de capturar detalles del jeroglífico a diferentes niveles, una vez que P' ha sido obtenido, a partir de cada píxel en P' se extraen varios fragmentos

¹La longitud de una rama corresponde al número de píxeles en la rama

del jeroglífico como se describe a continuación.

Para simplificar la explicación de la manera en que los fragmentos del jeroglífico son extraídos, primero definiremos cuando un par de píxeles en P' son píxeles vecinos. Dado un par de píxeles (p_i, p_j) en P' , decimos que son píxeles vecinos si y solo si, en el jeroglífico G' hay al menos un camino de píxeles consecutivos que conecta al píxel p_i con el píxel p_j , de modo que en dicho camino, no hay más píxeles de P' , excepto p_i y p_j . En la Figura 4.7 los píxeles marcados con un círculo representa a los píxeles en P' , los píxeles p_j, p_m y p_n (marcados en verde) son vecinos de p_i (marcado en rojo), mientras que los pares de píxeles (p_j, p_n) , (p_j, p_m) y (p_m, p_n) no son píxeles vecinos en P' .

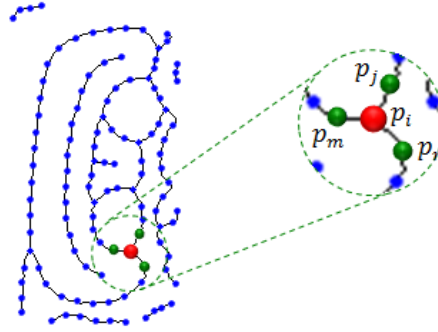


Figura 4.7: Píxeles en P' marcados en azul. Píxel p_i en P' marcado en rojo, y píxeles vecinos de p_i ($N_{p_i} = \{p_j, p_m, p_n\}$) marcados en verde.

Para cada píxel p_i en P' obtenemos varios fragmentos del jeroglífico adelgazado G' como se describe a continuación:

1. Sea N_{p_i} el conjunto de los píxeles vecinos del píxel p_i en P' . El primer fragmento del jeroglífico obtenido desde $p_i \in P'$, denotado como $F_{p_i}^1$, está formado por p_i y todos los píxeles consecutivos en G' que conectan a p_i con cada uno de sus píxeles vecinos en N_{p_i} , incluyendo los píxeles en N_{p_i} .
2. El segundo fragmento del jeroglífico G' obtenido desde el píxel $p_i \in P'$,

denotado como $F_{p_i}^2$, está formado por la unión del conjunto de pixeles del fragmento $F_{p_i}^1$ con el conjunto de todos los pixeles consecutivos en G' que conectan a cada pixel p_j en N_{p_i} con sus pixeles vecinos en N_{p_j} , incluyendo a los pixeles en N_{p_j} .

- De la misma manera se pueden seguir extrayendo fragmentos del jeroglífico a partir del pixel $p_i \in P'$.

En la Figura 4.8 se pueden observar cinco fragmentos del jeroglífico (marcados en verde) extraídos a partir de un pixel p_i (marcado en color rojo).

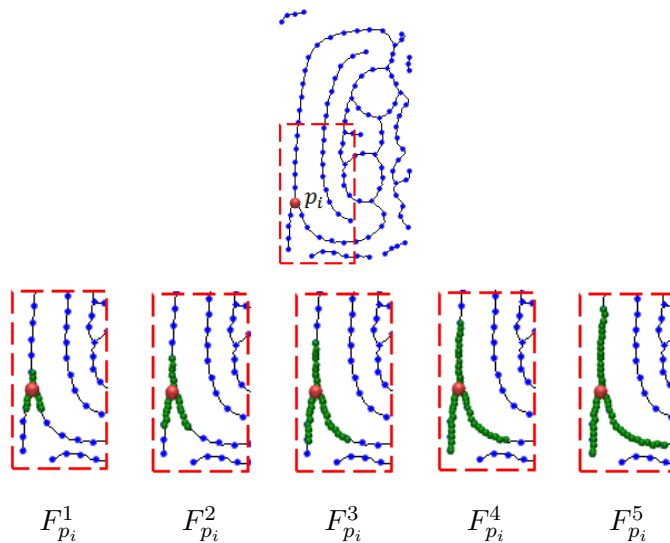


Figura 4.8: Fragmentos del jeroglífico extraídos desde el pixel p_i . Los fragmentos $F_{p_i}^1$, $F_{p_i}^2$, $F_{p_i}^3$, $F_{p_i}^4$, $F_{p_i}^5$ están marcados en verde.

Una vez extraídos los fragmentos del jeroglífico, la información de cada uno de los fragmentos es codificada a través de un descriptor local de forma (e.g. HOOSC). Intuitivamente, el descriptor para cada fragmento debería ser calculado desde el pixel p_i , es decir, el pixel a partir del cual el fragmento fue extraído. No obstante, a partir de dos pixeles diferentes en P' se pueden extraer fragmentos

iguales. Entonces, si el descriptor local para esos fragmentos es calculado desde p_i , lo anterior resultaría en descriptores locales diferentes para fragmentos iguales. Por lo tanto, para cada fragmento, el descriptor local es calculado a partir del punto medio del rectángulo mínimo que contiene al fragmento, considerando únicamente a los pixeles que forman parte del fragmento del jeroglífico. De esta manera, para dos fragmentos iguales, el descriptor local resultante sería el mismo.

4.3. Reemplazando cada descriptor local por sus k -palabras visuales más cercanas en el vocabulario

En el modelo BoVW (ver Figura 2.3), el vocabulario visual resulta de agrupar en c -grupos los descriptores locales obtenidos a partir de jeroglíficos en el conjunto de entrenamiento. El centroide w_i , $i = 1, \dots, c$, de cada grupo es considerado como la palabra visual que representa a todos los descriptores locales en el i -ésimo grupo. Así, una vez que se obtiene el vocabulario visual, cada descriptor local es reemplazado por su palabra visual más cercana en el vocabulario. Idealmente, descriptores locales similares deberían ser reemplazados por la misma palabra visual. Sin embargo, algunos descriptores locales parecidos pueden pertenecer a diferentes grupos y, por lo tanto, estos descriptores locales a pesar de ser similares son reemplazados por palabras visuales diferentes, lo cual puede afectar la representación de los jeroglíficos.

En la Figura 4.3 se ilustra el problema de reemplazar a cada descriptor local por su palabra visual más cercana en el vocabulario. En esta figura los puntos azules representan a los descriptores locales, las líneas punteadas ilustran los límites de cada grupo obtenidos durante la construcción del vocabulario visual, cada grupo es producto del proceso de agrupamiento (usualmente usando c -means) de los descriptores locales calculados desde jeroglíficos en el conjunto de entrenamiento

y los puntos negros representan el centroide de cada grupo, el cual corresponde a una palabra visual en el vocabulario visual. Los dos puntos azules resaltados a través de estrellas rojas denotan dos descriptores locales, que a pesar de su similitud, son reemplazados por palabras visuales diferentes en el vocabulario, debido a que estos descriptores locales caen en diferentes grupos (ver Figura 4.3a). Para abordar este problema, proponemos reemplazar a cada descriptor local por sus k -palabras visuales más cercanas en el vocabulario ($k > 1$), en lugar de que cada descriptor local sea reemplazado por su palabra visual más cercana. Este reemplazo múltiple permite que descriptores locales parecidos, preserven su similitud a través de las palabras visuales que comparten. Por ejemplo, en la Figura 4.3b, para el mismo par de descriptores locales parecidos marcados con estrellas rojas, se puede apreciar que si estos descriptores locales son reemplazados por sus k -palabras visuales más cercanas en el vocabulario ($k = 3$ en el ejemplo), a pesar de que los descriptores locales caen en diferentes grupos, la similitud entre ellos se preserva a través de las palabras visuales que comparten. En este ejemplo, para ambos descriptores locales, sus tres palabras visuales más cercanas, son las mismas.

Una vez que los descriptores locales han sido reemplazados por palabras visuales en el vocabulario, la representación para un jeroglífico siguiendo el modelo BoVW se obtiene calculando la frecuencia de las palabras visuales en el jeroglífico (ver Figura 2.3). No obstante, la frecuencia de las palabras visuales puede ser muy similar para jeroglíficos diferentes. Por lo tanto, considerar información sobre la disposición de las palabras visuales en el jeroglífico podría ayudar a mejorar la calidad de su representación. Siguiendo esta idea, en [Roman-Rangel and Marchand-Maillet, 2013a] se incorpora información sobre la disposición de las palabras visuales mediante el cálculo de co-ocurrencias de palabras visuales. Específicamente, se obtiene información de los pares de palabras visuales que

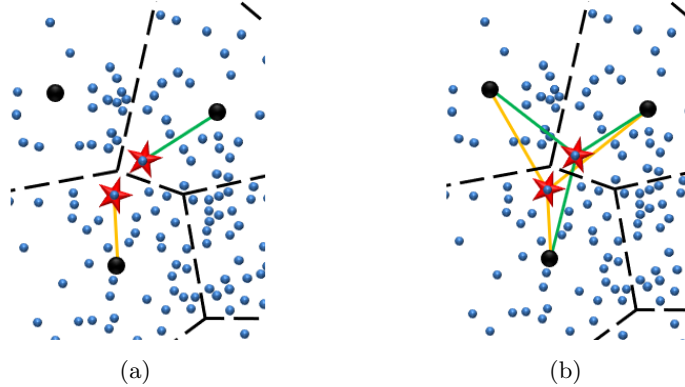


Figura 4.9: (a) Descriptores locales parecidos (puntos azules marcados con estrellas rojas) reemplazados por diferentes palabras visuales, su palabra visual más cercana en el vocabulario. (b) Descriptores locales parecidos reemplazados por sus k -palabras visuales más cercanas en el vocabulario ($k = 3$ en el ejemplo).

co-ocurren en una vecindad. No obstante, la co-ocurrencia de palabras visuales es calculada considerando que cada descriptor local fue reemplazado por su palabra visual más cercana en el vocabulario, lo cual como ya se mencionó, presenta algunos inconvenientes. Por lo tanto, en esta sección introducimos, además, una forma de calcular la frecuencia de las palabras visuales y la co-ocurrencia de las palabras visuales, pero considerando para ambos casos, el reemplazo de cada descriptor por sus k -palabras visuales más cercanas en el vocabulario.

La *frecuencia de una palabra visual* w_i en un jeroglífico G , denotada como $f(w_i)$, se calcula como el número de veces en que la palabra visual w_i aparece como una de las k -palabras visuales que reemplazaron a cada descriptor local x_t calculado a partir de un punto de interés p_t en G .

La *co-ocurrencia de un par de palabras visuales* (w_p, w_q) en un jeroglífico G se calcula como el número de veces en que la palabra visual w_p co-ocurre junto con la palabra visual w_q en una cierta vecindad en G . Sea w_p una de las k -palabras visuales que reemplazaron al descriptor local x_m calculado a partir

del pixel p_m y w_q una de las k -palabras visuales que reemplazaron al descriptor local x_n calculado a partir del pixel p_n ², consideramos que un par de palabras visuales (w_p, w_q) co-ocurren en el jeroglífico G si la distancia entre p_m y p_n no es mayor que un umbral δ_G , donde δ_G se calcula como en [Roman-Rangel and Marchand-Maillet, 2013a], usando la distancia Euclidiana promedio entre todos los puntos en P' en el jeroglífico G (ver Ecuación 4.3).

$$\delta_G = \frac{\sum_i^{|P'|} \sum_j^{|P'|} d(p_i, p_j)}{|P'|} \quad (4.3)$$

donde $d(p_i, p_j)$ es la distancia Euclidiana entre dos puntos p_i y p_j .

Ya que hemos definido cuándo un par de palabras (w_p, w_q) co-ocurren, para cada jeroglífico G , se construye una matriz de co-ocurrencias de pares de palabras visuales, denotada M^G , la cual es de tamaño $c \times c$. M^G contiene en cada posición p, q , el número de veces en que el par de palabras (w_p, w_q) co-ocurren en G .

Una vez ha sido calculada la matriz de co-ocurrencias de pares de palabras visuales para cada jeroglífico en el conjunto de entrenamiento, se calcula una matriz de co-ocurrencias de pares de palabras visuales para cada clase cl . Esta matriz, denotada M^{cl} , la cual es de tamaño $c \times c$, resulta de sumar todas las matrices de co-ocurrencias de pares de palabras visuales de los jeroglíficos que pertenecen a la clase cl . Así, M^{cl} contiene en cada posición p, q el número de veces en que el par de palabras visuales (w_p, w_q) co-ocurren en la clase cl . M^{cl} es normalizada para que el número de co-ocurrencias sea independiente del número de jeroglíficos en cada clase cl . La normalización se realiza dividiendo cada entrada $M_{p,q}^{cl}$ por el número de veces en que la palabra visual w_p co-ocurre en la clase cl con cualquier otra palabra visual w_j en el vocabulario (ver Ecuación 4.4).

² p_m y p_n son pixeles en P' , es decir los pixeles desde donde fueron extraídos los fragmentos del jeroglífico

$$M_{p,q}^{cl} = \frac{M_{p,q}^{cl}}{\sum_{j=1}^c M_{p,j}^{cl}} \quad (4.4)$$

Cada jeroglífico G en el conjunto de entrenamiento es representado como un vector h^G . Este vector contiene en su j -ésima posición la suma de las co-ocurrencias de los pares de palabras visuales (w_q, w_j) en la clase cl , siendo w_q cualquiera de las c palabras visuales en el vocabulario. Cada co-ocurrencia es ponderada por la frecuencia de la palabra visual w_q en G . De este modo, la j -ésima entrada en h^G tomará valores más grandes, mientras mayor sean tanto la frecuencia de w_q en G , como la co-ocurrencia de los pares de palabras visuales (w_q, w_j) en la clase cl (ver Ecuación 4.5).

$$h_j^G = \sum_{q=1}^c (M_{q,j}^{cl} \cdot f(w_q)) \quad (4.5)$$

En tareas de recuperación, un jeroglífico de consulta no puede ser representado de la misma manera que un jeroglífico de entrenamiento, ya que la clase cl para un jeroglífico de consulta es desconocida. Por lo tanto, para representar a un jeroglífico de consulta siguiendo la forma que se acaba de describir para la representación de jeroglíficos (ver Ecuación 4.5), éste es considerado como si fuera el único elemento de su clase.

RESULTADOS EXPERIMENTALES

La representación de jeroglíficos propuesta resulta de introducir modificaciones en las etapas del modelo BoVW [Roman-Rangel et al., 2011b]. Por lo tanto, en este capítulo se evalúa el efecto que tiene, en la calidad de la representación de jeroglíficos, el introducir en dicha representación, cada una de las modificaciones, para posteriormente comparar los resultados de recuperación obtenidos por la representación de jeroglíficos propuesta, la cual considera todas las modificaciones propuestas, contra la representación de jeroglíficos más exitosa reportada en la literatura para tareas de recuperación, la cual es denotada por sus autores como BoP [Roman-Rangel and Marchand-Maillet, 2013a]. Esos resultados son además comparados contra los resultados en recuperación obtenidos por la representación de jeroglíficos construida bajo el modelo BoVW [Roman-Rangel et al., 2011b], la cual pese a no ser la representación más exitosa en el estado del arte, es comúnmente utilizada para efectos de comparación. Antes de presentar los resultados experimentales, en la Sección 5.1 se describen las bases de datos y la configuración experimental utilizada. En el primer experimento, presentado en la Sección 5.2, se evalúa la calidad de la representación de jeroglíficos cuando ésta es construida desde jeroglíficos adelgazados y podados como se propone en la Sección 4.1, respecto a la obtenida desde jeroglíficos adelgazados sin podar.

Posteriormente, en la Sección 5.3, se evalúa el efecto que tiene en la calidad de la representación de jeroglíficos, el reemplazo de cada descriptor local por sus k -palabras visuales más cercanas en el vocabulario, lo cual fue propuesto en la Sección 4.3, en lugar de reemplazar cada descriptor local por su palabra visual más cercana en el vocabulario. En la Sección 5.4, se evalúa el efecto que tiene en la calidad de la representación de jeroglíficos, el cálculo de los descriptores locales a partir de fragmentos, lo cual fue propuesto en la Sección 4.2, en lugar de que los descriptores locales sean calculados desde puntos de interés. Finalmente, en la Sección 5.5, se comparan los resultados en recuperación obtenidos por la representación de jeroglíficos propuesta, la cual es construida al incluir todas las modificaciones propuestas, contra las representaciones más exitosas reportadas previamente en la literatura.

5.1. Bases de datos y configuración experimental

Para los experimentos se utilizaron cinco bases de datos que han sido utilizadas previamente en la literatura para evaluar la representación de jeroglíficos. La primera base de datos, denotada *Maya-I*, contiene 240 instancias de jeroglíficos Mayas distribuidos en 24 clases (10 jeroglíficos por clase) [Roman-Rangel et al., 2013]. La segunda base de datos, denotada *Maya-II*, contiene 1,043 instancias distribuidas en 25 clases [Roman-Rangel et al., 2011b]. La tercera y cuarta base de datos, denotadas como *Petroglyph-I* y *Petroglyph-II*, contienen 1,223 y 1,066 petroglifos distribuidos en 18 y 26 clases [Seidl et al., 2015a], respectivamente. La última base de datos, denotada *Egyptian*, está compuesta por 4,210 jeroglíficos egipcios distribuidos en 171 clases [Franken and van Gemert, 2013]. Dado que la base de datos *Egyptian* contiene jeroglíficos etiquetados como “DESCONOCIDO” y algunas clases tienen muy pocos jeroglíficos, los jeroglíficos etiquetados como

“DESCONOCIDO” y las clases con menos de 10 jeroglíficos fueron eliminados, dando lugar a una base de datos compuesta por 3,674 jeroglíficos distribuidos en 59 clases. En la Tabla 5.1 se presenta la información general de las bases de datos utilizadas en los experimentos. En las Figuras 5.1-5.3 se presenta un jeroglífico por clase para cada una de las bases de datos utilizadas en nuestros experimentos.

Tabla 5.1: Bases de datos utilizadas en nuestros experimentos

Nombre	N° de jeroglíficos	N° de clases
Maya-I	240	24
Maya-II	1,043	25
Petroglyph-I	1,223	18
Petroglyph-II	1,066	26
Egyptian	3,674	59

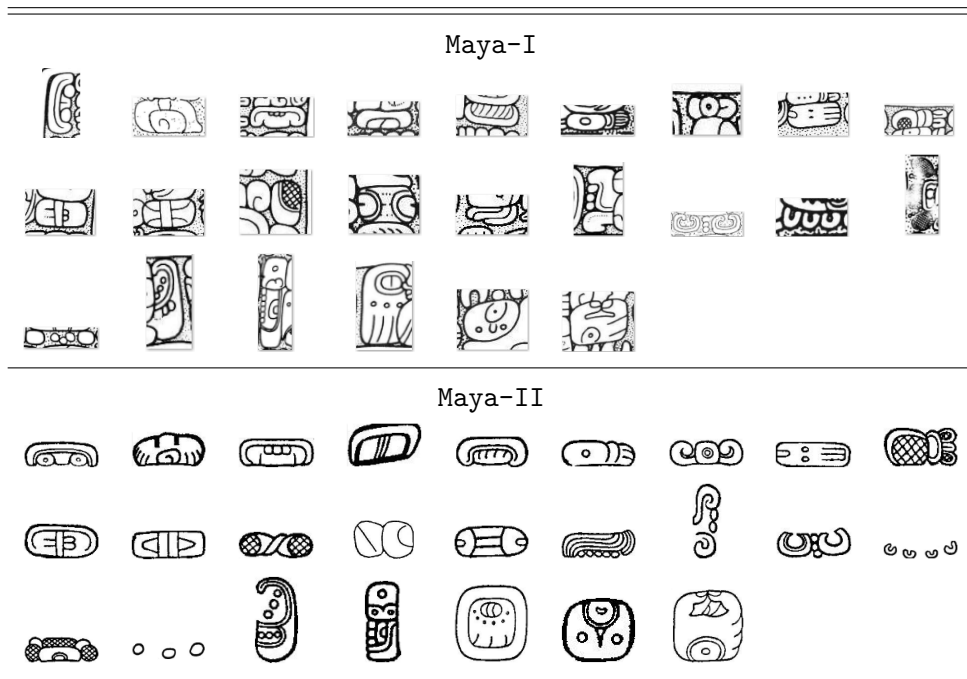


Figura 5.1: Jeroglífico por clase para Maya-I y Maya-II.

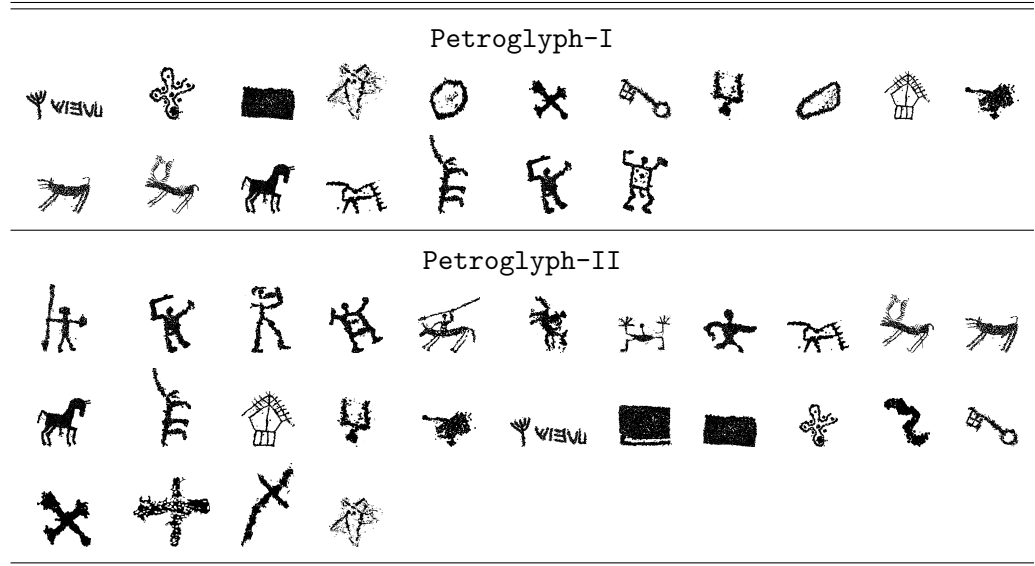


Figura 5.2: Petroglifo por clase para Petroglyph-I y Petroglyph-II.

Debido a que las bases de datos contienen jeroglíficos de diferentes tamaños, cada jeroglífico fue redimensionado utilizando la función *imresize* de matlab, de modo que el lado más largo de cada jeroglífico sea de 256 píxeles.

Los jeroglíficos en cada una de las bases de datos fueron pre-procesados de la misma manera que se ha hecho en trabajos previamente reportados en la literatura [Franken and van Gemert, 2013; Guo et al., 2016; Roman-Rangel et al., 2011b; Seidl, 2016; Wieser et al., 2017], como se describe a continuación. Las bases de datos Maya-I, Maya-II, Petroglyph-I y Petroglyph-II contienen imágenes binarias. Para los jeroglíficos en estas bases de datos, primero se aplicó el filtro de ruido propuesto en [Vincent, 1994], para eliminar pequeñas perturbaciones en los jeroglíficos producidas durante el proceso de digitalización. Posteriormente, para obtener la forma adelgazada de cada jeroglífico, de modo que cada rama en el jeroglífico adelgazado tenga un píxel de grosor, se aplicó el algoritmo de adelgazamiento propuesto en [Lam et al., 1992]. Como se puede apreciar en la

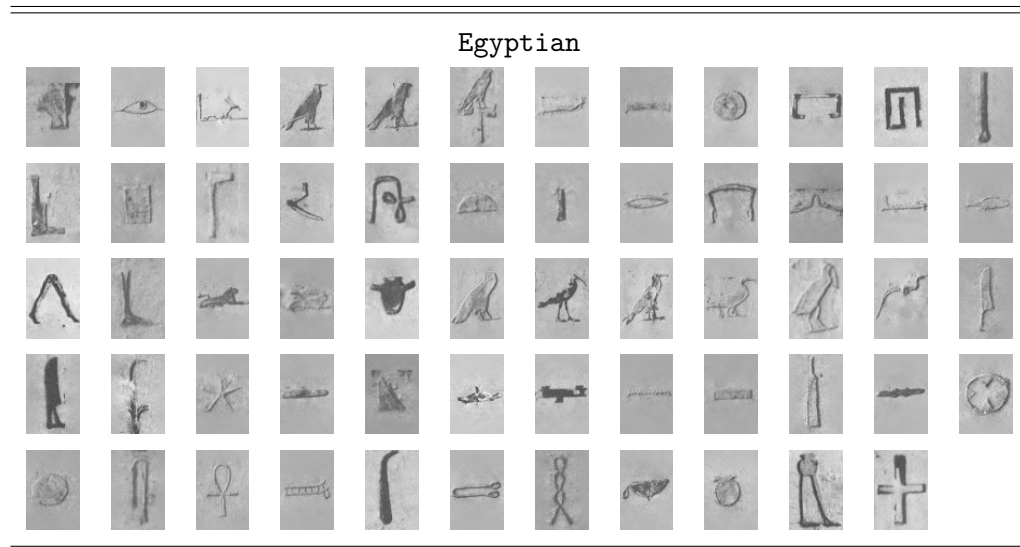


Figura 5.3: Jeroglífico por clase para *Egyptian*.

Figura 5.3, *Egyptian* contiene jeroglíficos en escala de grises y por esta razón no se puede aplicar un algoritmo de adelgazamiento. Para obtener una forma adelgazada de los jeroglíficos Egipcios, se utilizó el detector de bordes de Canny [Canny, 1986]. Finalmente, cada jeroglífico fue colocado en el centro de una imagen negra de 256×256 píxeles.

Para todos los experimentos se aplicó validación cruzada de cinco pliegues (*five-fold cross-validation*). En cada experimento, el 20% (un pliegue) y 80% (4 pliegues) de los jeroglíficos en cada categoría se utilizaron como jeroglíficos de consulta (i.e. *queries*) y catálogo de jeroglíficos, respectivamente. La calidad de la representación de jeroglíficos en tareas de recuperación se evaluó en términos del **mAP**[Baeza-Yates and Ribeiro-Neto, 1999], que se define como el promedio de las medias de precisión para un conjunto de consultas o *queries*. A su vez, la precisión promedio de cada consulta o *query* es calculada como el promedio de la precisión alcanzada por las instancias recuperadas (i.e. jeroglíficos en el

catálogo) ordenadas en términos de su similitud con el jeroglífico de consulta. Los resultados reportados se obtienen repitiendo cada experimento tres veces por partición, reportando el mAP promedio. Cada experimento se ejecuta varias veces por partición porque el conjunto inicial de centros iniciales para el algoritmo de agrupamiento *c*-means, utilizado para construir el vocabulario visual, es diferente en cada ejecución, además cuando se obtienen descriptores locales a partir de puntos de interés, éstos puntos son obtenidos mediante selección aleatoria uniforme. Esa aleatoriedad provoca que el vocabulario visual sea diferente en cada ejecución. Es importante mencionar que para construir la representación, el vocabulario visual se obtiene considerando únicamente a los jeroglíficos del catálogo (*training*), sin considerar a los de consulta (*queries*).

Uno de los parámetros del modelo BoVW es el tamaño del vocabulario visual. Para la construcción del vocabulario se evaluaron diferentes tamaños del mismo: 500, 1000, 1500, 2000 y 2500. En la mayoría de los casos, los mejores resultados se obtuvieron con un vocabulario de 1500 palabras visuales. Por este motivo, en todos los experimentos se se presentan a continuación se utilizó un vocabulario visual formado por 1500 palabras visuales.

5.2. Evaluación de la poda propuesta

En esta sección se evalúa el efecto de la poda propuesta en la Sección 4.1 sobre la calidad de la representación de jeroglíficos. Para esto, se comparan los resultados obtenidos en términos de recuperación, cuando los jeroglíficos son representados desde su forma adelgazada sin podar, usando las representaciones de jeroglíficos BoVW[Roman-Rangel et al., 2011b] y BoP[Roman-Rangel and Marchand-Maillet, 2013a] (esta última como ya se mencionó es la representación de jeroglíficos más exitosa reportada en la literatura), respecto a las representaciones construidas

desde jeroglíficos adelgazados y podados utilizando la poda propuesta en la Sección 4.1.

El podado de ramas no deseadas en un jeroglífico adelgazado introducido en la Sección 4.1 requiere del parámetro α , para decidir si las *ramas simples* y las *ramas finales* en un jeroglífico adelgazado son ramas no deseadas y por lo tanto deben ser eliminadas. Para este experimento, el valor del parámetro α fue determinado experimentalmente, mediante prueba y error. Se probaron valores desde 0,005 hasta 0,03, con incrementos de 0,005. Los mejores resultados se obtuvieron con $\alpha = 0,005$ y $\alpha = 0,01$ para las bases de datos de jeroglíficos Mayas (i.e., **Maya-I**, **Maya-II**) y petroglifos (i.e., **Petroglyph-I**, **Petroglyph-II**), respectivamente. La base de datos **Egyptian** no es incluida en este experimento, ya que a partir de las ramas que resultan del detector de bordes de Canny no es posible reconstruir región alguna, de modo que, la poda propuesta en la Sección 4.1, no puede ser aplicada en esta base de datos.

El modelo BoVW requiere como parámetro el número de puntos de interés que se utilizarán para calcular los descriptores locales. En la literatura, para este parámetro, denotado β , se han utilizado $\beta = 5\%$, 10% y 20% del total de píxeles en el jeroglífico adelgazado. Por lo tanto, en este trabajo, los experimentos son reportados utilizando estos valores para β y calculando diferentes variantes del descriptor local HOOSC: HOOSC[Roman-Rangel et al., 2011a], HOOSC4[Roman-Rangel et al., 2011b], HOOSC128[Roman-Rangel and Marchand-Maillet, 2014b] y HOOSC128_{xy}, ya que en la literatura [Franken and van Gemert, 2013; Roman-Rangel et al., 2011a] se ha demostrado que estos descriptores locales permiten alcanzar una mejor calidad en la representación de jeroglíficos, y por lo tanto son ampliamente utilizados. HOOSC128_{xy} es similar al HOOSC128 pero agregando al final de cada descriptor local, la posición relativa (x, y) del punto de interés.

En la Tabla 5.2 y la Tabla 5.3 se presentan los resultados obtenidos en recupe-

Tabla 5.2: mAP promedio obtenido al representar jeroglíficos bajo el modelo BoVW desde jeroglíficos adelgazados sin podar (BoVW) y desde jeroglíficos adelgazados y podados (BoVW*), usando diferentes porcentajes de puntos de interés (5 %, 10 % y 20 %) y diferentes variantes del descriptor local HOOSC: HOOSC, HOOSC4, HOOSC128 y HOOSC128_{xy}.

Descriptor		Porcentaje de puntos de interés					
		5 %		10 %		20 %	
		BoVW	BoVW*	BoVW	BoVW*	BoVW	BoVW*
Maya-I	HOOSC	0.178	0.216	0.198	0.236	0.239	0.257
	HOOSC4	0.234	0.241	0.287	0.303	0.334	0.339
	HOOSC128	0.186	0.216	0.226	0.263	0.283	0.301
	HOOSC128 _{xy}	0.292	0.320	0.341	0.381	0.396	0.415
Maya-II	HOOSC	0.176	0.232	0.259	0.298	0.328	0.338
	HOOSC4	0.219	0.253	0.313	0.340	0.378	0.387
	HOOSC128	0.246	0.287	0.346	0.367	0.405	0.406
	HOOSC128 _{xy}	0.334	0.382	0.442	0.467	0.494	0.495
Petroglyph-I	HOOSC	0.170	0.172	0.200	0.209	0.229	0.248
	HOOSC4	0.164	0.164	0.184	0.193	0.210	0.220
	HOOSC128	0.169	0.177	0.195	0.215	0.224	0.246
	HOOSC128 _{xy}	0.173	0.184	0.202	0.226	0.231	0.257
Petroglyph-II	HOOSC	0.139	0.142	0.160	0.163	0.187	0.195
	HOOSC4	0.135	0.136	0.149	0.155	0.171	0.180
	HOOSC128	0.136	0.138	0.153	0.167	0.178	0.197
	HOOSC128 _{xy}	0.149	0.153	0.173	0.187	0.198	0.213

ración de jeroglíficos en términos del mAP promedio para las representaciones de jeroglíficos BoVW y BoP, respectivamente; ambas representaciones construidas desde jeroglíficos adelgazados sin utilizar y utilizando la poda propuesta, bajo la configuración de parámetros descrita. Cada tabla está dividida en cuatro bloques, un bloque por cada base de datos utilizada (Maya-I, Maya-II, Petroglyph-I, Petroglyph-II). Para cada bloque horizontal, las filas presentan los resultados obtenidos con los descriptores locales HOOSC, HOOSC4, HOOSC128 y HOOSC128_{xy}, utilizando diferentes porcentajes de puntos de interés $\beta = 5\%$, 10% y 20% .

Tabla 5.3: mAP promedio obtenido utilizando la representación de jeroglíficos BoP [Roman-Rangel and Marchand-Maillet, 2013a] construida desde jeroglíficos adelgazados sin podar (BoP) y desde jeroglíficos adelgazados y podados (BoP*), usando diferentes porcentajes de puntos de interés (5 %, 10 % y 20 %) y diferentes variantes del descriptor local HOOSC: HOOSC, HOOSC4, HOOSC128 y HOOSC128_{xy}.

Descriptor	Porcentaje de puntos de interés						
	5 %		10 %		20 %		
	BoP	BoP*	BoP	BoP*	BoP	BoP*	
Maya-I	HOOSC	0.247	0.294	0.328	0.395	0.454	0.449
	HOOSC4	0.353	0.375	0.469	0.498	0.572	0.558
	HOOSC128	0.293	0.330	0.387	0.408	0.495	0.467
	HOOSC128 _{xy}	0.422	0.469	0.544	0.557	0.600	0.602
Maya-II	HOOSC	0.444	0.530	0.575	0.617	0.648	0.652
	HOOSC4	0.495	0.542	0.622	0.640	0.685	0.681
	HOOSC128	0.530	0.563	0.629	0.631	0.665	0.661
	HOOSC128 _{xy}	0.612	0.649	0.685	0.699	0.718	0.718
Petroglyph-I	HOOSC	0.343	0.380	0.410	0.468	0.485	0.537
	HOOSC4	0.323	0.359	0.367	0.427	0.425	0.487
	HOOSC128	0.331	0.378	0.389	0.445	0.440	0.491
	HOOSC128 _{xy}	0.348	0.388	0.403	0.454	0.440	0.489
Petroglyph-II	HOOSC	0.247	0.258	0.337	0.368	0.406	0.453
	HOOSC4	0.227	0.242	0.292	0.336	0.359	0.412
	HOOSC128	0.236	0.267	0.299	0.350	0.363	0.410
	HOOSC128 _{xy}	0.254	0.282	0.320	0.358	0.359	0.398

Cada celda de la Tabla 5.2 muestra el mAP alcanzado al utilizar la representación de jeroglíficos bajo el modelo BoVW construida desde jeroglíficos adelgazados sin podar (BoVW) y desde jeroglíficos adelgazados y podados (BoVW*). Por su parte, cada celda de la Tabla 5.3 muestra el mAP alcanzado al utilizar la representación de jeroglíficos BoP construida desde jeroglíficos adelgazados sin podar (BoP) y desde jeroglíficos adelgazados y podados (BoP*). En cada base de datos, para cada descriptor local y cada porcentaje de puntos de interés, se resalta en negritas el mejor mAP promedio obtenido. Como se puede observar en la Tabla 5.2, el podado de ramas no deseadas en jeroglíficos adelgazados permite, en todos los casos, mejorar la calidad de la representación de jeroglíficos BoVW. En la Tabla 5.3 los mejores resultados de recuperación se obtienen, en la mayoría de casos con la representación de jeroglíficos BoP*, la cual es construida desde jeroglíficos adelgazados y podados.

5.3. Evaluación del reemplazo de cada descriptor local por sus k -palabras visuales más cercanas.

En la Sección 4.3, se propone que cada descriptor local sea reemplazado por sus k -palabras visuales más cercanas en el vocabulario, en lugar de su palabra visual más cercana. Adicionalmente, se introdujo una manera para que las k -palabras visuales que reemplazaron a cada descriptor local sean consideradas en el cálculo de la frecuencia y de las co-ocurrencias de palabras visuales. En esta sección se evaluará, si el reemplazo de cada descriptor local por sus k -palabras visuales más cercanas ayuda a obtener mejores resultados en recuperación que los obtenidos por la representación de jeroglíficos construida desde jeroglíficos adelgazados y podados BoP*, en donde cada descriptor es reemplazado únicamente por su palabra visual más cercana en el vocabulario. Considerando lo anterior, en todos

los experimentos de esta sección, la representación de jeroglíficos se obtendrá a partir de jeroglíficos adelgazados y podados ¹. Para esto, primero se evaluará la representación de jeroglíficos que utiliza el reemplazo de cada descriptor por sus k -palabras visuales más cercanas en el vocabulario utilizando diferentes valores para k , así como tres opciones para el cálculo de las co-ocurrencias de palabras visuales. Posteriormente, en un segundo experimento, se compara la representación de jeroglíficos que usa el reemplazo de cada descriptor local por sus k -palabras visuales más cercanas en el vocabulario contra la representación BoP*, utilizando las diferentes variantes del descriptor local HOOSC.

Al igual que en la Sección 5.2, las pruebas en recuperación de jeroglíficos se llevaron a cabo utilizando diferentes porcentajes de puntos de interés ($\beta = 5\%$, 10% y 20%), y un vocabulario visual de 1500 palabras visuales. En el primer experimento de esta sección se evalúan diferentes valores para k , $k = 3, 5$ y 7 , y para el cálculo de las co-ocurrencias entre un par de palabras visuales (w_p, w_q) , además de considerar que un par de palabras visuales (w_p, w_q) co-ocurren si w_p y w_q son cualquiera de las k -palabras visuales que reemplazaron al descriptor local respectivo². También se prueban dos opciones más:

1. Considerar como w_p únicamente a la palabra visual más cercana que reemplazó al descriptor local respectivo y como w_q considerar a las k -palabras visuales que reemplazaron al descriptor local respectivo; y
2. Considerar como w_q únicamente a la palabra visual más cercana que reemplazó al descriptor local respectivo y como w_p considerar a las k -palabras visuales que reemplazaron al descriptor local respectivo.

¹Como ya se mencionó, para la base de datos **Egyptian** la poda propuesta no puede ser aplicada. Por lo tanto los resultados en recuperación obtenidos desde jeroglíficos adelgazados con o sin poda, son los mismos.

²Las palabras visuales w_p y w_q vienen desde dos puntos diferentes, cuya distancia no es mayor a un umbral δ_G . Donde δ_G se calcula como en la Ecuación 4.3

Estas opciones para el cálculo de las co-ocurrencias son denotadas como $\text{BoP}^*_{(k,k)}$, $\text{BoP}^*_{(1,k)}$ y $\text{BoP}^*_{(k,1)}$, respectivamente.

En la Tabla 5.4 se presentan los resultados obtenidos en recuperación de jeroglíficos, en términos del mAP promedio, para las tres opciones para el cálculo de las co-ocurrencias: $\text{BoP}^*_{(k,k)}$, $\text{BoP}^*_{(1,k)}$ y $\text{BoP}^*_{(k,1)}$, utilizando diferentes porcentajes de puntos de interés, diferentes valores para k y descriptores locales HOOSC[Roman-Rangel et al., 2011a]. La Tabla 5.4 está horizontalmente dividida en cinco bloques, un bloque para cada base de datos. En cada bloque horizontal se presentan en las fila los resultados de recuperación en jeroglíficos obtenidos siguiendo cada una de las opciones para calcular las co-ocurrencias: $\text{BoP}^*_{(k,k)}$, $\text{BoP}^*_{(1,k)}$ y $\text{BoP}^*_{(k,1)}$, utilizando diferentes porcentajes de puntos de interés ($\beta = 5\%$, 10% y 20%) y para cada uno de esos porcentajes diferentes valores para k ($k = 3, 5$ y 7). En la Tabla 5.4, para cada base de datos, cada porcentaje de puntos de interés y cada valor de k , el mejor mAP promedio alcanzado está resaltado en negrita. De los resultados mostrados se puede observar que, en la mayoría de los casos, el reemplazo de cada descriptor local por sus k -palabras visuales más cercanas en el vocabulario y el cálculo de las co-ocurrencias entre palabras $\text{BoP}^*_{(k,1)}$ obtiene los mejores resultados. Además, se puede notar que el mAP promedio incrementa a medida que el k incrementa, sin embargo, el incremento del mAP promedio es mayor entre $k = 3$ y $k = 5$ que entre $k = 5$ y $k = 7$. Por esta razón, para los siguientes experimentos utilizaremos el reemplazo de cada descriptor local por sus k -palabras visuales más cercanas en el vocabulario y el cálculo de las co-ocurrencias entre palabras $\text{BoP}^*_{(k,1)}$ con $k = 5$.

Con el objetivo de evaluar si el reemplazo de cada descriptor local por sus k -palabras visuales más cercanas en el vocabulario y el cálculo de las co-ocurrencias entre palabras visuales $\text{BoP}^*(k, 1)$ permite mejorar la calidad de la representación

Tabla 5.4: mAP promedio obtenido por $\text{BoP}^*_{(k,k)}$, $\text{BoP}^*_{(1,k)}$ y $\text{BoP}^*_{(k,1)}$ para Maya-I, Maya-II, Petroglyph-I, Petroglyph-II y Egyptian, utilizando diferentes porcentajes de puntos de interés (5 %, 10 % y 20 %), diferentes valores para k ($k = 3, 5$ y 7) y HOOSC como descriptor local

Método		Porcentaje de puntos de interés								
		5 %			10 %			20 %		
		$k = 3$	$k = 5$	$k = 7$	$k = 3$	$k = 5$	$k = 7$	$k = 3$	$k = 5$	$k = 7$
Maya-I	$\text{BoP}^*_{(k,k)}$	0.463	0.503	0.518	0.523	0.539	0.544	0.563	0.576	0.575
	$\text{BoP}^*_{(1,k)}$	0.499	0.541	0.550	0.542	0.552	0.554	0.570	0.580	0.578
	$\text{BoP}^*_{(k,1)}$	0.417	0.471	0.507	0.518	0.560	0.581	0.562	0.594	0.609
Maya-II	$\text{BoP}^*_{(k,k)}$	0.675	0.701	0.707	0.701	0.706	0.703	0.704	0.707	0.702
	$\text{BoP}^*_{(1,k)}$	0.694	0.711	0.713	0.704	0.705	0.700	0.702	0.703	0.697
	$\text{BoP}^*_{(k,1)}$	0.643	0.681	0.700	0.701	0.725	0.737	0.716	0.732	0.739
Petroglyph-I	$\text{BoP}^*_{(k,k)}$	0.480	0.521	0.537	0.555	0.571	0.574	0.588	0.593	0.592
	$\text{BoP}^*_{(1,k)}$	0.494	0.526	0.535	0.555	0.565	0.564	0.580	0.581	0.576
	$\text{BoP}^*_{(k,1)}$	0.443	0.481	0.506	0.540	0.568	0.583	0.595	0.615	0.625
Petroglyph-II	$\text{BoP}^*_{(k,k)}$	0.381	0.432	0.451	0.475	0.496	0.498	0.477	0.519	0.519
	$\text{BoP}^*_{(1,k)}$	0.402	0.444	0.456	0.477	0.493	0.493	0.506	0.510	0.508
	$\text{BoP}^*_{(k,1)}$	0.329	0.375	0.403	0.452	0.484	0.500	0.510	0.531	0.542
Egyptian	$\text{BoP}^*_{(k,k)}$	0.734	0.731	0.722	0.839	0.840	0.836	0.876	0.875	0.868
	$\text{BoP}^*_{(1,k)}$	0.739	0.731	0.703	0.842	0.839	0.833	0.875	0.872	0.868
	$\text{BoP}^*_{(k,1)}$	0.722	0.743	0.751	0.832	0.842	0.847	0.877	0.883	0.886

de jeroglíficos, comparamos los resultados en recuperación obtenidos por $\text{BoP}_{(k,1)}^*$ con $k = 5$ contra los obtenidos por BoP^* , ya que de acuerdo a nuestros experimentos BoP^* es la representación que permite obtener los mejores resultados en términos de recuperación, superando a la representación de jeroglíficos más exitosa reportada en la literatura (i.e. BoP [Roman-Rangel and Marchand-Maillet, 2013a]). Los resultados en recuperación obtenidos por BoP^* y $\text{BoP}_{(k,1)}^*$ se presentan en la Tabla 5.5.

La Tabla 5.5 está horizontalmente dividida en cinco bloques, un bloque por cada base de datos. En cada bloque horizontal, las filas presentan los resultados en recuperación de jeroglíficos obtenidos haciendo uso de los descriptores locales HOOSC , HOOSC4 , HOOSC128 y HOOSC128_{xy} , utilizando diferentes porcentajes de puntos de interés (5 %, 10 % y 20 %) y para cada uno de esos porcentajes las representaciones de jeroglíficos BoP^* y $\text{BoP}_{(k,1)}^*$ (con $k = 5$). De los resultados mostrados en la Tabla 5.5, se puede observar que, independientemente del descriptor local y del porcentaje de puntos de interés utilizado, los resultados obtenidos por $\text{BoP}_{(k,1)}^*$ siempre superan a los obtenidos por BoP^* .

5.4. Evaluación del cálculo de descriptores locales desde fragmentos del jeroglífico

En esta sección evaluaremos si utilizar descriptores locales calculados a partir de fragmentos del jeroglífico en la representación $\text{BoP}_{(k,1)}^*$, siguiendo la propuesta presentada en la Sección 4.2, en lugar de utilizar descriptores locales desde puntos de interés, permite alcanzar aún mejores resultados en recuperación de jeroglíficos.

Debido a que la extracción de fragmentos propuesta en la Sección 4.2 permite obtener diferente número de fragmentos, en esta sección, primero se presentan los resultados experimentales considerando diferente número de fragmentos. Una

Tabla 5.5: mAP promedio obtenido por BoP* y BoP*_(k,1) (nuestra propuesta con $k = 5$) para Maya-I, Maya-II, Petroglyph-I, Petroglyph-II y Egyptian, usando diferentes porcentajes de puntos de interés ($\beta = 5\%$, 10% y 20%) y diferentes variantes del descriptor local HOOSC: HOOSC, HOOSC4, HOOSC128 y HOOSC128_{xy}.

Descriptor	Local	Porcentaje de puntos de interés					
		5 %		10 %		20 %	
		BoP*	BoP* _(k,1)	BoP*	BoP* _(k,1)	BoP*	BoP* _(k,1)
Maya-I	HOOSC	0.294	0.471	0.395	0.560	0.449	0.594
	HOOSC4	0.375	0.527	0.498	0.649	0.558	0.683
	HOOSC128	0.330	0.462	0.408	0.559	0.467	0.603
	HOOSC128 _{xy}	0.469	0.604	0.557	0.679	0.602	0.728
Maya-II	HOOSC	0.530	0.681	0.617	0.725	0.652	0.732
	HOOSC4	0.542	0.671	0.640	0.739	0.681	0.761
	HOOSC128	0.563	0.687	0.631	0.722	0.661	0.742
	HOOSC128 _{xy}	0.649	0.746	0.699	0.773	0.718	0.788
Petroglyph-I	HOOSC	0.380	0.481	0.468	0.568	0.537	0.615
	HOOSC4	0.359	0.439	0.427	0.529	0.487	0.574
	HOOSC128	0.378	0.479	0.445	0.547	0.491	0.579
	HOOSC128 _{xy}	0.388	0.490	0.454	0.555	0.489	0.586
Petroglyph-II	HOOSC	0.258	0.375	0.368	0.484	0.453	0.531
	HOOSC4	0.242	0.335	0.336	0.445	0.412	0.502
	HOOSC128	0.267	0.375	0.350	0.467	0.410	0.504
	HOOSC128 _{xy}	0.282	0.391	0.358	0.472	0.398	0.510
Egyptian	HOOSC	0.688	0.743	0.812	0.842	0.863	0.883
	HOOSC4	0.636	0.704	0.782	0.818	0.841	0.867
	HOOSC128	0.664	0.736	0.786	0.823	0.833	0.862
	HOOSC128 _{xy}	0.729	0.782	0.831	0.857	0.872	0.894

vez se determina la cantidad de fragmentos que permite alcanzar los mejores resultados en recuperación de jeroglíficos, estos resultados son comparados contra los resultados obtenidos por $\text{BoP}_{(k,1)}^*$, construida a partir de puntos de interés.

Al igual que en las secciones previas, las pruebas en recuperación de jeroglíficos se llevaron a cabo utilizando un vocabulario visual de 1500 palabras visuales. En el primer experimento de esta sección se evalúan diferentes cantidad de fragmentos T ($T = 9, 11, 13$ y 15) usando diferentes valores de γ (para γ se usan los mismos valores que para la selección de los puntos de interés, es decir, $\gamma = 5\%$, 10% y 20%). En la Tabla 5.6, se presentan los resultados en recuperación obtenidos para cada una de las bases de datos (*Maya-I*, *Maya-II*, *Petroglyph-I*, *Petroglyph-II* y *Egyptian*) utilizando la representación de jeroglíficos $\text{BoP}_{(k,1)}^*$ pero con descriptores locales calculados desde fragmentos, en lugar de descriptores locales calculados desde puntos de interés. Dicha representación será denotada como *Bo k -VWF*. Para los resultados presentados en la Tabla 5.6, se utilizaron descriptores locales HOOSC128_{xy} , ya que en términos generales, en los experimentos anteriores, con este descriptor local se han alcanzado los mejores resultados en recuperación de jeroglíficos para las diferentes bases de datos.

La Tabla 5.6 está dividida en cinco bloques horizontales, un bloque para cada base de datos. Para cada bloque, cada fila presenta los resultados cuando se utilizan los diferentes porcentajes de pixeles del jeroglífico adelgazado (5% , 10% y 20%) y para cada porcentaje los diferentes valores de T . En esta tabla, para cada base de datos y cada porcentaje de puntos del jeroglífico adelgazado, el valor resaltado en negrita corresponde al mejor mAP promedio obtenido.

De los resultados presentados en la Tabla 5.6 se puede observar que para cada base de datos, los mejores resultados en términos del mAP promedio se obtienen utilizando el 20% de pixeles del jeroglífico adelgazado. De estos resultados, en la mayoría de los casos, los mejores resultados se alcanzan para $T = 13$. Considerando

Tabla 5.6: mAP promedio obtenido por *Bok*-VWF para *Maya-I*, *Maya-II*, *Petroglyph-I*, *Petroglyph-II* y *Egyptian*, utilizando diferentes valores para γ (5 %, 10 % y 20 %), diferente cantidad de fragmentos ($T = 9, 11, 13$ y 15) y HOOSC128_{xy} como descriptor local.

	γ	Cantidad de niveles de fragmentos			
		$T = 9$	$T = 11$	$T = 13$	$T = 15$
<i>Maya-I</i>	5 %	0.690	0.684	0.672	0.653
	10 %	0.757	0.748	0.738	0.737
	20 %	0.703	0.727	0.758	0.754
<i>Maya-II</i>	5 %	0.836	0.830	0.825	0.816
	10 %	0.839	0.845	0.841	0.831
	20 %	0.837	0.844	0.850	0.843
<i>Petroglyph-I</i>	5 %	0.520	0.516	0.513	0.508
	10 %	0.593	0.582	0.577	0.569
	20 %	0.562	0.581	0.603	0.609
<i>Petroglyph-II</i>	5 %	0.418	0.412	0.404	0.397
	10 %	0.504	0.499	0.495	0.489
	20 %	0.503	0.514	0.537	0.526
<i>Egyptian</i>	5 %	0.814	0.807	0.805	0.799
	10 %	0.883	0.875	0.869	0.863
	20 %	0.879	0.891	0.902	0.894

lo anterior, independientemente del porcentaje de píxeles del jeroglífico adelgazado, para *Bok*-NF utilizaremos $T = 13$.

Con el objetivo de evaluar si utilizar descriptores locales calculados desde fragmentos permite obtener mejores resultados de recuperación que al utilizar descriptores locales calculados desde puntos de interés, comparamos los resultados en recuperación obtenidos por $\text{BoP}^*_{(k,1)}$ (con $k = 5$) contra los obtenidos por *Bok*-VWF (con $k = 5$ y $T = 13$). Ambos, $\text{BoP}^*_{(k,1)}$ y *Bok*-VWF se construyen siguiendo

UNA NUEVA REPRESENTACIÓN DE JEROGLÍFICOS BASADA EN SU
DESCOMPOSICIÓN EN PARTES PRIMITIVAS

Tabla 5.7: mAP promedio obtenido para Maya-I, Maya-II, Petroglyph-I, Petroglyph-II y Egyptian por $\text{BoP}^*_{(k,1)}$ (con $k = 5$ y diferentes porcentajes de puntos de interés $\beta = 5\%$, 10% y 20%) y Bok-VWF (con $k = 5$, $T = 13$ y diferentes porcentajes de pixeles $\gamma = 5\%$, 10% y 20%), utilizando HOOSC128_{xy} como descriptor local.

	Porcentaje de puntos					
	5 %		10 %		20 %	
	$\text{BoP}^*_{(k,1)}$	Bok-NF	$\text{BoP}^*_{(k,1)}$	Bok-NF	$\text{BoP}^*_{(k,1)}$	Bok-NF
Maya-I	0.604	0.672	0.679	0.738	0.728	0.758
Maya-II	0.746	0.825	0.773	0.841	0.788	0.850
Petroglyph-I	0.490	0.513	0.555	0.577	0.586	0.603
Petroglyph-II	0.391	0.404	0.472	0.495	0.510	0.537
Egyptian	0.782	0.805	0.857	0.869	0.894	0.902

el modelo BoVW desde jeroglíficos adelgazados y podados y considerando para el cálculo de las co-ocurrencias de palabras visuales, las k -palabras visuales más cercanas que reemplazaron a cada descriptor local. Los resultados en recuperación obtenidos por $\text{BoP}^*_{(k,1)}$ y Bok-VWF se presentan en la Tabla 5.7.

En la Tabla 5.7 se presentan por fila los resultados obtenidos para cada una de las bases de datos (Maya-I, Maya-II, Petroglyph-I, Petroglyph-II y Egyptian) por $\text{BoP}^*_{(k,1)}$ y Bok-VWF con cada uno de los porcentajes (5 %, 10 % y 20 %) y utilizando descriptores locales HOOSC128_{xy} . Para cada base de datos y cada porcentaje, en negrita se resalta el mejor mAP promedio alcanzado. De los resultados mostrados en la Tabla 5.7 se puede observar que independientemente del porcentaje de puntos utilizado para extraer los descriptores locales desde puntos de interés o para extraer los descriptores locales desde fragmentos, el mejor mAP promedio para cada una de las bases de datos siempre es obtenido por Bok-VWF .

5.5. Evaluación de la representación de jeroglíficos propuesta

El objetivo de esta sección es comparar los resultados en recuperación alcanzados por la representación de jeroglíficos propuesta (i.e. *Bok-VWF*) contra los obtenidos por la representación de jeroglíficos más exitosa reportada en la literatura BoP [Roman-Rangel and Marchand-Maillet, 2013a], así como contra BoVW [Roman-Rangel et al., 2011b].

En la Tabla 5.8 se presentan los resultados de recuperación, en términos del mAP promedio, alcanzados por BoVW, BoP y *Bok-NF* (la representación de jeroglíficos propuesta) para cada una de las bases de datos (*Maya-I*, *Maya-II*, *Petroglyph-I*, *Petroglyph-II* y *Egyptian*), utilizando los diferentes porcentajes (5 %, 10 % y 20 %). Para cada base de datos y cada porcentaje el valor resaltado en negrita indica la representación de jeroglíficos que logró alcanzar los mejores resultados de recuperación en términos del mAP promedio. De estos resultados se puede notar que la representación de jeroglíficos propuesta *Bok-VWF*, en todos los casos, alcanza mejores resultados de recuperación que los obtenidos por BoVW y BoP. Con el objetivo de validar estadísticamente los resultados obtenidos, se llevó a cabo la prueba de Wilcoxon entre los mAP promedio obtenidos por la representación de jeroglíficos propuesta *Bok-NF* y BoP. La prueba de Wilcoxon detectó una diferencia estadísticamente significativa entre los resultados obtenidos por *Bok-NF* y BoP.

Con el objetivo de comparar la calidad de la recuperación alcanzada por BoVW, BoP, y *Bok-NF* para cada una de las bases de datos, considerando únicamente los primeros n elementos recuperados, en la Tabla 5.9 se muestran los resultados de $p@n$ ($n = 1,3,5$) utilizando los diferentes porcentajes de puntos (5 %, 10 %

Tabla 5.8: mAP promedio obtenido por BoVW, BoP y Bok-VWF para Maya-I, Maya-II, Petroglyph-I, Petroglyph-II y Egyptian, utilizando diferentes porcentajes (5 %, 10 % y 20 %) y HOOSC128_{xy} como descriptor local.

	Porcentaje de puntos								
	5 %			10 %			20 %		
	BoVW	BoP	Bok-NF	BoVW	BoP	Bok-NF	BoVW	BoP	Bok-NF
Maya-I	0.292	0.422	0.672	0.341	0.544	0.738	0.396	0.600	0.758
Maya-II	0.334	0.612	0.825	0.442	0.685	0.841	0.494	0.718	0.850
Petroglyph-I	0.173	0.348	0.513	0.202	0.403	0.577	0.213	0.440	0.603
Petroglyph-II	0.149	0.254	0.404	0.173	0.320	0.495	0.198	0.359	0.537
Egyptian	0.372	0.729	0.805	0.569	0.831	0.866	0.700	0.872	0.902

y 20 %) y HOOSC128_{xy} como descriptor local. La Tabla 5.9 está dividida en cinco bloques horizontales, uno para cada base de datos. En cada fila del bloque se muestran los resultados obtenidos por p@1, p@3 y p@5 bajo los diferentes porcentajes de puntos. Como se puede observar en la Tabla 5.9, el p@n (n = 1,3,5) obtenido por la representación de jeroglíficos propuesta (i.e. Bok-NF) en Maya-I, Maya-II, Petroglyph-I y Petroglyph-II es mayor que el obtenido por BoVW y BoP, independiente del porcentaje de puntos. Para Egyptian, el p@n (n=1,3,5) obtenido por Bok-NF con el 5 % y el 10 % de puntos es mayor que el obtenido por BoVW y BoP, mientras que el p@n (n = 1,3,5) obtenido por BoP con el el 20 % de puntos de interés es mayor que el obtenido por BoVW y Bok-NF. Lo anterior se puede atribuir a que a diferencia de Maya-I, Maya-II, Petroglyph-I y Petroglyph-II, los jeroglíficos en Egyptian son imágenes en escala de grises, evitando que se puedan aplicar métodos que forman parte de la representación de jeroglíficos propuesta, por ejemplo el adelgazamiento del jeroglífico y el podado de ramas no deseadas. Así mismo, los fragmentos son extraídos desde los trazos obtenidos mediante el detector de bordes de Canny, en lugar de que estos provengan del jeroglífico adelgazado y podado.

Tabla 5.9: $p@n$ ($n = 1,3,5$) obtenido por BoVW, BoP y Bok-VWF para Maya-I, Maya-II, Petroglyph-I, Petroglyph-II y Egyptian, utilizando diferentes porcentajes (5 %, 10 % y 20 %) y HOOSC128_{xy} como descriptor local.

		Porcentaje de puntos								
		5 %			10 %			20 %		
		BoVW	BoP	Bok-NF	BoVW	BoP	Bok-NF	BoVW	BoP	Bok-NF
Maya-I	p@1	0.513	0.575	0.687	0.617	0.667	0.763	0.675	0.750	0.763
	p@3	0.410	0.497	0.699	0.503	0.618	0.747	0.532	0.700	0.750
	p@5	0.349	0.451	0.668	0.412	0.583	0.725	0.461	0.642	0.733
Maya-II	p@1	0.713	0.774	0.880	0.782	0.834	0.863	0.828	0.865	0.871
	p@3	0.657	0.754	0.871	0.782	0.820	0.859	0.782	0.848	0.864
	p@5	0.620	0.736	0.866	0.751	0.806	0.854	0.751	0.832	0.860
Petroglyph-I	p@1	0.277	0.464	0.634	0.419	0.583	0.633	0.527	0.634	0.672
	p@3	0.244	0.453	0.624	0.363	0.556	0.624	0.458	0.600	0.659
	p@5	0.220	0.439	0.612	0.322	0.541	0.612	0.424	0.585	0.650
Petroglyph-II	p@1	0.277	0.378	0.506	0.371	0.480	0.582	0.464	0.549	0.610
	p@3	0.226	0.362	0.483	0.306	0.449	0.562	0.405	0.521	0.592
	p@5	0.207	0.349	0.470	0.277	0.431	0.549	0.366	0.503	0.579
Egyptian	p@1	0.714	0.806	0.820	0.828	0.879	0.890	0.891	0.918	0.9058
	p@3	0.677	0.803	0.818	0.810	0.880	0.886	0.873	0.914	0.9025
	p@5	0.658	0.797	0.816	0.797	0.875	0.881	0.862	0.909	0.8994

En las Figuras 5.4, 5.5 y 5.6 se presentan para cada una de las bases de datos, tres jeroglíficos de consulta y los primeros cinco jeroglíficos recuperados para cada consulta, utilizando la representación de jeroglíficos propuesta. En cada una de esas figuras, en la primera columna se presenta el jeroglífico de consulta y en las columnas dos a seis los primeros cinco jeroglíficos recuperados. Los jeroglíficos marcados en un recuadro rojo, son aquellos jeroglíficos que no pertenecen a la misma clase del jeroglífico de consulta. Después de analizar los resultados obtenidos, pudimos observar que aunque nuestra propuesta permite, en la mayoría de los casos, obtener mejores resultados en términos del mAP y $p@n$, el mayor problema continua siendo distinguir entre jeroglíficos similares que pertenecen a categorías diferentes.

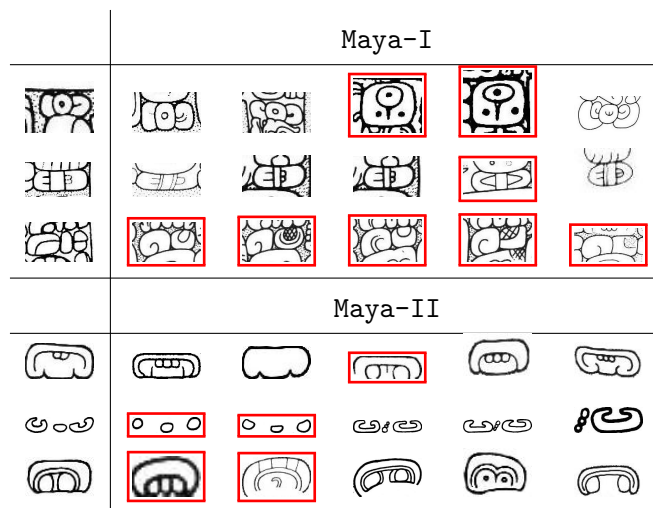


Figura 5.4: Jeroglíficos de consulta y sus primeros cinco jeroglíficos más parecidos en la base de datos, para Maya-I y Maya-II, utilizando la representación de jeroglíficos propuesta.

En la Figura 5.7 se presenta el mAP vs el recuerdo (*Standard recall*) para BoVW, BoP y Bok-NF en cada una de las bases de datos, utilizando los diferentes porcentajes de puntos (5 %, 10 % y 20 %). Como se puede apreciar en la Figura 5.7, independientemente del porcentaje de puntos, en Maya-I (ver Figura 5.7a), Maya-II (ver Figura 5.7b), Petroglyph-I (ver Figura 5.7c) y Petroglyph-II (ver Figura 5.7d), la representación propuesta Bok-NF obtiene el mejor mAP para todos los intervalos de recuerdo (*Standard recall*). El mAP obtenido por la representación propuesta Bok-NF en los diferentes intervalos del recuerdo para Egyptian (ver Figura 5.7e) es en la mayoría de veces mayor que el obtenido por BoVW y BoP. Únicamente, en el primer intervalo del recuerdo con el 20 % de puntos, el mejor mAP es alcanzado por BoP, lo cual es consistente con los resultados de p@n.











































		Petroglyph-I					
							
							
							
		Petroglyph-II					
							
							
							

Figura 5.5: Jeroglíficos de consulta y sus primeros cinco jeroglíficos más parecidos en la base de datos, para Petroglyph-I y Petroglyph-II, utilizando la representación de jeroglíficos propuesta.



Figura 5.6: Jeroglíficos de consulta y sus primeros cinco jeroglíficos más parecidos en la base de datos para **Egyptian**, utilizando la representación de jeroglíficos propuesta.

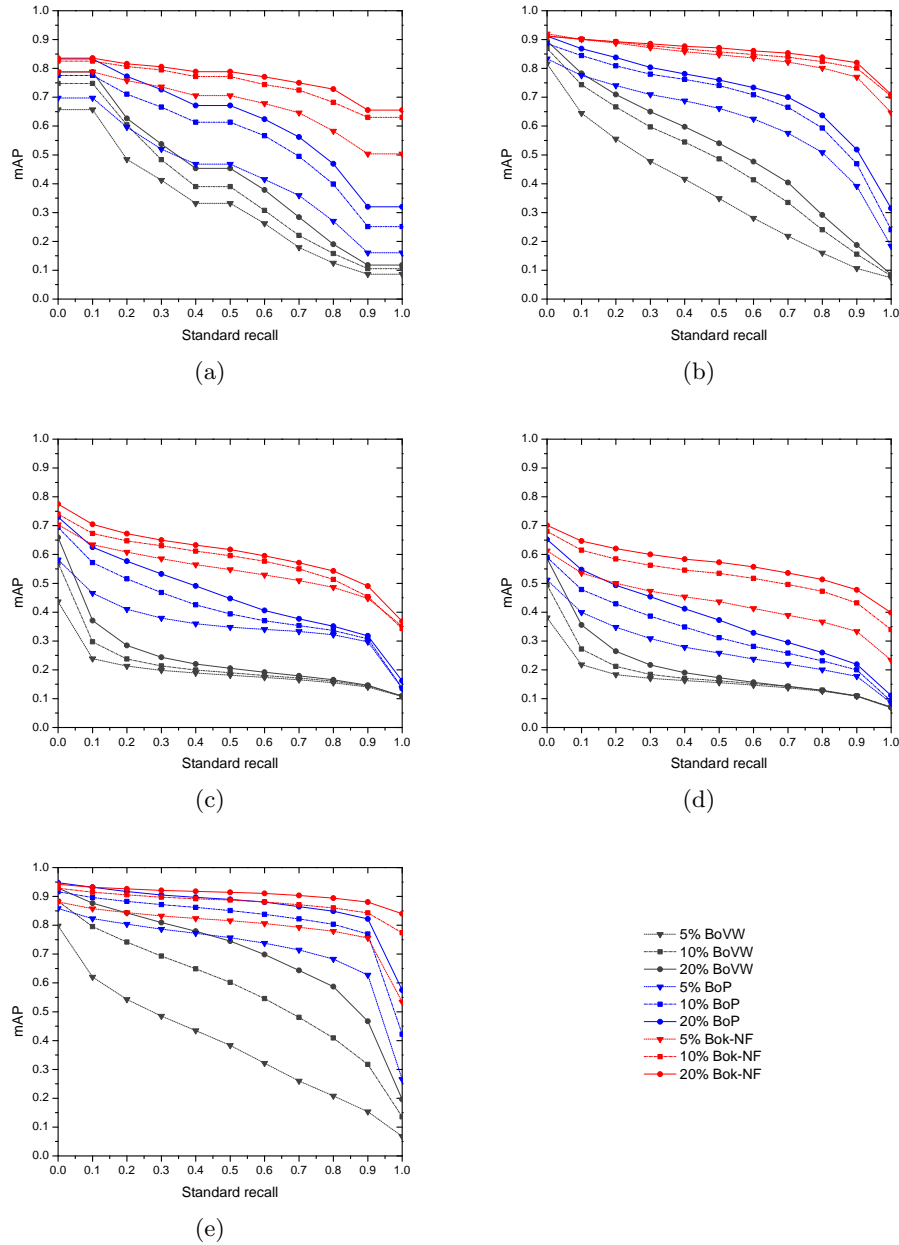


Figura 5.7: mAP vs recuerdo para (a) Maya-I, (b) Maya-II, (c) Petroglyph-I, (d) Petroglyph-II y (e) Egyptian utilizando el 5 %, 10 % y 20 % de puntos para BoVW, BoP y Bok-NF, y HOOSC128_{xy} como descriptor local.

CONCLUSIONES Y TRABAJO FUTURO

En este trabajo de investigación se propuso una nueva representación de jeroglíficos para tareas de recuperación, la cual resulta de introducir mejoras en las diferentes etapas del modelo BoVW. Primero, se propuso un método para la poda de ramas no deseadas en jeroglíficos adelgazados. Además, se propuso reemplazar cada descriptor local por sus k -palabras visuales más cercanas en el vocabulario, en lugar de reemplazar sólo por su palabra visual más cercana. Finalmente, se introdujo una manera de extraer fragmentos del jeroglífico para calcular descriptores locales considerando únicamente a los píxeles en cada fragmento.

En este capítulo, primero se presentan las conclusiones de este trabajo de investigación, posteriormente se mencionan las contribuciones y se describen brevemente algunas líneas a seguir como trabajo futuro. Finalmente, se listan las publicaciones.

6.1. Conclusiones

De los resultados experimentales podemos concluir que:

- La representación de jeroglíficos propuesta permite alcanzar resultados de recuperación estadísticamente mejores respecto a los resultados obtenidos

por las representaciones más exitosas del estado del arte.

- La poda de ramas no deseadas en jeroglíficos adelgazados permite, por si sola, mejorar la calidad de la representación de los mismos, cuando estos son representados ya sea bajo el modelo BoVW [Roman-Rangel et al., 2011b] o a través de la representación más exitosa reportada en la literatura BoP [Roman-Rangel and Marchand-Maillet, 2013a].
- Reemplazar a cada descriptor local por sus k -palabras visuales más cercanas en el vocabulario, permite mejorar aún más la calidad de la representación de jeroglíficos, respecto a cuando cada descriptor local es reemplazado por su palabra más cercana en el vocabulario.
- La extracción de fragmentos del jeroglífico y el cálculo de descriptores locales de fragmentos, permite mejorar la calidad de la representación de jeroglíficos construida desde jeroglíficos adelgazados y podados, considerando el reemplazo de cada descriptor local por sus k -palabras visuales más cercanas en el vocabulario, comparado con la representación obtenida utilizando descriptores locales obtenidos a partir de puntos de interés.

6.2. Contribuciones

Las contribuciones de este trabajo de investigación son:

- Una nueva representación de jeroglíficos siguiendo el modelo BoVW que permite obtener mejores resultados de recuperación que las representaciones más exitosas reportadas previamente en la literatura.
- Un método para podar ramas no deseadas desde jeroglíficos adelgazados, el cual permite obtener representaciones de mejor calidad, en términos de

recuperación, que las obtenidas a partir de jeroglíficos sin podar.

- Tres alternativas para calcular las co-ocurrencias entre palabras visuales considerando las k -palabras visuales más cercanas que reemplazaron a cada descriptor local.
- Un método para extraer fragmentos del jeroglífico y descriptores locales a partir de los mismos.

6.3. Trabajo futuro

Los resultados obtenidos en esta tesis abren otros campos de estudio para representar jeroglíficos siguiendo el modelo BoVW. Como trabajo futuro consideramos lo siguiente:

- Estudiar diferentes maneras de calcular las co-ocurrencias entre palabras visuales, considerando el reemplazo de cada descriptor por sus k -palabras visuales más cercanas, por ejemplo, utilizar co-ocurrencias entre grupos de palabras visuales o considerar no sólo la cercanía entre las palabras visuales sino además la disposición entre las mismas.
- Los resultados obtenidos al calcular descriptores locales HOOSC a partir de fragmentos son alentadores, por lo cual, como trabajo futuro sería interesante estudiar diferentes maneras de calcular descriptores locales a partir de fragmentos.
- Considerando la popularidad que han ganado los autocodificadores profundos, como trabajo futuro proponemos hacer uso de este tipo de extractores de características y combinarlos con nuestra propuesta para la representación de jeroglíficos.

6.4. Publicaciones

A continuación se listan los artículos publicados y/o aceptados, directamente derivados de este trabajo de investigación:

- **Artículos en revista JCR**

1. **Pinilla-Buitrago L. A.**, *et al.* Improved Hieroglyph Representation for Image Retrieval. *Journal on Computing and Cultural Heritage* (Aceptado Oct-2018).
2. **Pinilla-Buitrago L. A.**, *et al.* Bag of k -Nearest Visual Words for Hieroglyph Retrieval. *Journal of Intelligent & Fuzzy Systems* (Aceptado Oct-2018).

- **Conferencias internacionales**

3. **Pinilla-Buitrago L. A.**, *et al.* Including Foreground and Background Information in Maya Hieroglyph Representation. *10th Mexican Conference on Pattern Recognition*, LNCS 10880, pages 238-247, (2018).

BIBLIOGRAFÍA

Aharon, M., Elad, M., and Bruckstein, A. k-svd: An algorithm for designing overcomplete dictionaries for sparse representation. *IEEE Transactions on Signal Processing*, 54(11):4311–4322, Nov 2006.

Baeza-Yates, R. A. and Ribeiro-Neto, B. *Modern Information Retrieval*. Addison-Wesley Longman Publishing Co., Inc., Boston, MA, USA, 1999. ISBN 020139829X.

Ballard, D. Generalizing the hough transform to detect arbitrary shapes. *Pattern Recognition*, 13(2):111 – 122, 1981. ISSN 0031-3203.

Belongie, S., Malik, J., and Puzicha, J. Shape context: A new descriptor for shape matching and object recognition. In *In NIPS*, pages 831–837, 2000.

Belongie, S., Malik, J., and Puzicha, J. Shape matching and object recognition using shape contexts. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(4):509–522, April 2002.

Boureau, Y., Bach, F., LeCun, Y., and Ponce, J. Learning mid-level features for recognition. In *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 2559–2566, June 2010a.

Boureau, Y.-L., Ponce, J., and Lecun, Y. A theoretical analysis of feature

- pooling in visual recognition. In *27TH INTERNATIONAL CONFERENCE ON MACHINE LEARNING, HAIFA, ISRAEL*, 2010b.
- Can, G., Odobez, J.-M., and Gatica-Perez, D. Evaluating shape representations for maya glyph classification. *J. Comput. Cult. Herit.*, 9(3):14:1–14:26, Sept. 2016.
- Can, G., Odobez, J.-M., and Gatica-Perez, D. Shape representations for maya codical glyphs: Knowledge-driven or deep? In *Proceedings of the 15th International Workshop on Content-Based Multimedia Indexing, CBMI '17*, pages 32:1–32:6, New York, NY, USA, 2017. ACM.
- Can, G., Odobez, J.-M., and Gatica-Perez, D. How to tell ancient signs apart? recognizing and visualizing maya glyphs with cnns. *ACM Journal on Computing and Cultural Heritage (JOCCH)*, May 2018.
- Canny, J. A computational approach to edge detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-8(6):679–698, Nov 1986. ISSN 0162-8828.
- Dalal, N. and Triggs, B. Histograms of oriented gradients for human detection. In *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, volume 1, pages 886–893 vol. 1, June 2005.
- Datta, R., Joshi, D., Li, J., and Wang, J. Z. Image retrieval: Ideas, influences, and trends of the new age. *ACM Computing Surveys*, 40(2), April 2008.
- Deufemia, V. and Paolino, L. Segmentation and recognition of petroglyphs using generic fourier descriptors. In Elmoataz, A., Lezoray, O., Nouboud, F., and Mammass, D., editors, *Image and Signal Processing*, pages 487–494, Cham, 2014. Springer International Publishing.

- Deufemia, V., Paolino, L., and d. Lumley, H. Petroglyph recognition using self-organizing maps and fuzzy visual language parsing. In *2012 IEEE 24th International Conference on Tools with Artificial Intelligence*, volume 1, pages 852–859, Nov 2012.
- Domingo, J. D., Herrera, P. J., Valero, E., and Cerrada, C. Deciphering egyptian hieroglyphs: Towards a new strategy for navigation in museums. In *Sensors*, 2017.
- Franken, M. and van Gemert, J. C. Automatic egyptian hieroglyph recognition by retrieving images as texts. In *Proceedings of the 21st ACM International Conference on Multimedia*, MM '13, pages 765–768, New York, NY, USA, 2013. ACM.
- Frauel, Y., Quesada, O., and Bribiesca, E. Detection of a polymorphic mesoamerican symbol using a rule-based approach. *Pattern Recognition*, 39(7):1380 – 1390, 2006. ISSN 0031-3203.
- Gatica-Perez, D., Roman-Rangel, E., Odobez, J.-M., and Pallan, C. New world, new worlds: Visual analysis of pre-columbian pictorial collections. In Grana, C. and Cucchiara, R., editors, *Multimedia for Cultural Heritage*, pages 90–102, Berlin, Heidelberg, 2012. Springer Berlin Heidelberg.
- Guo, J., Wang, C., Roman-Rangel, E., Chao, H., and Rui, Y. Building hierarchical representations for oracle character and sketch recognition. *IEEE Transactions on Image Processing*, 25(1):104–118, Jan 2016.
- Haralick, R. M. and Shapiro, L. G. *Computer and Robot Vision*. Addison-Wesley Longman Publishing Co., Inc., Boston, MA, USA, 1st edition, 1992. ISBN 0201569434.

Hartigan, J. and Wong, M. Algorithm AS 136: A K-means clustering algorithm. *Applied Statistics*, pages 100–108, 1979.

Heider, P., Pierre-Pierre, A., Li, R., and Grimm, C. Local shape descriptors, a survey and evaluation. In *Proceedings of the 4th Eurographics Conference on 3D Object Retrieval*, 3DOR '11, pages 49–56, Aire-la-Ville, Switzerland, Switzerland, 2011. Eurographics Association. ISBN 978-3-905674-31-6.

Hough, P. Method and means for recognizing complex patterns. 12 1962.

Hu, R. and Collomosse, J. A performance evaluation of gradient field hog descriptor for sketch based image retrieval, 2013.

Hu, R., Gayol, C. P., Krempel, G., Odobez, J.-M., and Gatica-Perez, D. Automatic maya hieroglyph retrieval using shape and context information. In *Proceedings of the 22Nd ACM International Conference on Multimedia*, MM '14, pages 1037–1040, New York, NY, USA, 2014. ACM.

Hu, R., Can, G., Gayol, C. P., Krempel, G., Spotak, J., Vail, G., Marchand-Maillet, S., Odobez, J., and Gatica-Perez, D. Multimedia analysis and access of ancient maya epigraphy: Tools to support scholars on maya hieroglyphics. *IEEE Signal Processing Magazine*, 32(4):75–84, July 2015.

Hu, R., Pallan Gayol, C., Odobez, J.-M., and Gatica-Perez, D. Analyzing and visualizing ancient maya hieroglyphics using shape: from computer vision to digital humanities. *Digital Scholarship in the Humanities*, 32:179–194, Dec. 2017.

Jonker, R. and Volgenant, A. A shortest augmenting path algorithm for dense and sparse linear assignment problems. *Computing*, 38(4):325–340, Dec 1987.

- Kazmi, I. K., You, L., and Zhang, J. J. A survey of 2d and 3d shape descriptors. In *2013 10th International Conference Computer Graphics, Imaging and Visualization*, pages 1–10, Aug 2013. doi: 10.1109/CGIV.2013.11.
- Lam, L., Lee, S. ., and Suen, C. Y. Thinning methodologies-a comprehensive survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 14 (9):869–885, Sept 1992. ISSN 0162-8828.
- Lazebnik, S., Schmid, C., and Ponce, J. Beyond bags of features: Spatial pyramid matching for recognizing natural scene categories. In *2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'06)*, volume 2, pages 2169–2178, June 2006.
- Lisin, D. A., Mattar, M. A., Blaschko, M. B., Learned-Miller, E. G., and Benfield, M. C. Combining local and global image features for object class recognition. In *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05) - Workshops*, pages 47–47, Sept 2005. doi: 10.1109/CVPR.2005.433.
- Lowe, D. G. Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*, 60(2):91–110, Nov 2004.
- Mikolajczyk, K. and Schmid, C. Scale & affine invariant interest point detectors. *International Journal of Computer Vision*, 60(1):63–86, Oct 2004.
- Mori, G., Belongie, S., and Malik, J. Efficient shape matching using shape contexts. *IEEE Trans. Pattern Anal. Mach. Intell.*, 27(11):1832–1837, Nov. 2005. ISSN 0162-8828.
- Papadimitriou, C. H. and Steiglitz, K. *Combinatorial Optimization: Algorithms*

and Complexity. Prentice-Hall, Inc., Upper Saddle River, NJ, USA, 1982. ISBN 0-13-152462-3.

Rigamonti, R., Brown, M. A., and Lepetit, V. Are sparse representations really relevant for image classification? In *CVPR 2011*, pages 1545–1552, June 2011.

Roman-Rangel, E. and Marchand-Maillet, S. Bag-of-visual-phrases via local contexts. In *2013 2nd IAPR Asian Conference on Pattern Recognition*, pages 867–871, Nov 2013a.

Roman-Rangel, E. and Marchand-Maillet, S. Stopwords detection in bag-of-visual-words: The case of retrieving maya hieroglyphs. In Petrosino, A., Maddalena, L., and Pala, P., editors, *New Trends in Image Analysis and Processing – ICIAP 2013*, pages 159–168, Berlin, Heidelberg, 2013b. Springer Berlin Heidelberg.

Roman-Rangel, E. and Marchand-Maillet, S. Automatic removal of visual stopwords. In *Proceedings of the 22Nd ACM International Conference on Multimedia, MM '14*, pages 1145–1148, New York, NY, USA, 2014a. ACM.

Roman-Rangel, E. and Marchand-Maillet, S. *HOOSC128: A More Robust Local Shape Descriptor*, pages 172–181. Springer International Publishing, Cham, 2014b.

Roman-Rangel, E. and Marchand-Maillet, S. Shape-based detection of maya hieroglyphs using weighted bag representations. *Pattern Recognition*, 48(4): 1161 – 1173, 2015.

Roman-Rangel, E. and Marchand-Maillet, S. Indexing mayan hieroglyphs with neural codes. In *2016 23rd International Conference on Pattern Recognition (ICPR)*, pages 253–258, Dec 2016.

Roman-Rangel, E., Pallan, C., Odobez, J.-M., and Gatica-Perez, D. Retrieving ancient maya glyphs with shape context. In *2009 IEEE 12th International Conference on Computer Vision Workshops, ICCV Workshops*, pages 988–995, Sept 2009.

Roman-Rangel, E., Pallan, C., Odobez, J.-M., and Gatica-Perez, D. Analyzing ancient maya glyph collections with contextual shape descriptors. *International Journal of Computer Vision*, 94(1):101–117, Aug 2011a. ISSN 1573-1405.

Roman-Rangel, E., Pallan Gayol, C., Odobez, J.-M., and Gatica-Perez, D. Searching the past: An improved shape descriptor to retrieve maya hieroglyphs. In *Proceedings of the 19th ACM International Conference on Multimedia, MM '11*, pages 163–172, New York, NY, USA, 2011b. ACM. ISBN 978-1-4503-0616-4.

Roman-Rangel, E., Odobez, J.-M., and Gatica-Perez, D. Assessing sparse coding methods for contextual shape indexing of maya hieroglyphs. *Journal of Multimedia*, 7(2):179–192, 2012.

Roman-Rangel, E., Odobez, J.-M., and Gatica-Perez, D. Evaluating shape descriptors for detection of maya hieroglyphs. In *Pattern Recognition*, pages 145–154, Berlin, Heidelberg, 2013. Springer Berlin Heidelberg.

Roman-Rangel, E., Can, G., Marchand-Maillet, S., Hu, R., Gayol, C. P., Krempele, G., Spotak, J., Odobez, J.-M., and Gatica-Perez, D. Transferring neural representations for low-dimensional indexing of maya hieroglyphic art. In Hua, G. and Jégou, H., editors, *Computer Vision – ECCV 2016 Workshops*, pages 842–855, Cham, 2016. Springer International Publishing.

Seidl, M. *Computational Analysis of Petroglyphs*. PhD thesis, TU Wien, 8 2016.

- Seidl, M. and Breiteneder, C. Detection and Classification of Petroglyphs in Gigapixel Images – Preliminary Results. In Niccolucci, F., Dellepiane, M., Serna, S. P., Rushmeier, H., and Gool, L. V., editors, *VAST: International Symposium on Virtual Reality, Archaeology and Intelligent Cultural Heritage - Short and Project Papers*. The Eurographics Association, 2011.
- Seidl, M., Wieser, E., and Alexander, C. Automated classification of petroglyphs. *Digital Applications in Archaeology and Cultural Heritage*, 2(2):196 – 212, 2015a.
- Seidl, M., Wieser, E., Zeppelzauer, M., Pinz, A., and Breiteneder, C. Graph-based shape similarity of petroglyphs. In Agapito, L., Bronstein, M. M., and Rother, C., editors, *Computer Vision - ECCV 2014 Workshops*, pages 133–148, Cham, 2015b. Springer International Publishing.
- Sivic, J. and Zisserman, A. Video google: A text retrieval approach to object matching in videos. In *Proceedings of the Ninth IEEE International Conference on Computer Vision - Volume 2*, ICCV '03, pages 1470–, Washington, DC, USA, 2003. IEEE Computer Society. ISBN 0-7695-1950-4.
- Smeulders, A. W. M., Worring, M., Santini, S., Gupta, A., and Jain, R. Content-based image retrieval at the end of the early years. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(12):1349–1380, Dec 2000. ISSN 0162-8828. doi: 10.1109/34.895972.
- Thompson, J. E. S. *A catalog of Maya hieroglyphs / by Eric S. Thompson*. University of Oklahoma Press Norman, Okla, 1962.
- Tian, D. A review on image feature extraction and representation techniques. *International Journal of Multimedia and Ubiquitous Engineering*, 8:385–395, 01 2013.

- Vail, G. and Hernández, C. The maya codices database, version 5.0. <http://www.mayacodices.org>.
- Vincent, L. Morphological area openings and closings for grey-scale images. In O, Y.-L., Toet, A., Foster, D., Heijmans, H. J. A. M., and Meer, P., editors, *Shape in Picture*, pages 197–208, Berlin, Heidelberg, 1994. Springer Berlin Heidelberg. ISBN 978-3-662-03039-4.
- Wieser, E., Seidl, M., and Zeppelzauer, M. A study on skeletonization of complex petroglyph shapes. *Multimedia Tools and Applications*, 76(6):8285–8303, Mar 2017.
- Zhang, E. and Mayo, M. Improving bag-of-words model with spatial information. In *2010 25th International Conference of Image and Vision Computing New Zealand*, pages 1–8, Nov 2010.
- Zhu, Q., Wang, X., Keogh, E., and Lee, S.-H. Augmenting the generalized hough transform to enable the mining of petroglyphs. In *Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD '09*, pages 1057–1066, New York, NY, USA, 2009. ACM.
- Zhu, Q., Wang, X., Keogh, E., and Lee, S.-H. An efficient and effective similarity measure to enable data mining of petroglyphs. *Data Mining and Knowledge Discovery*, 23(1):91–127, Jul 2011.

Descriptores locales

Este apéndice presenta una breve explicación de los descriptores utilizados en el desarrollo de esta tesis para la representación de jeroglíficos.

A.1. *Shape-Context*

El descriptor *Shape-Context* (SC) [Belongie et al., 2000, 2002] es un descriptor local basado en contorno, el cual codifica la información de la distribución relativa de un subconjunto de puntos en el contorno respecto a cada punto en dicho subconjunto. Explícitamente, el descriptor SC se obtiene como se describe a continuación. Sea $P = \{p_1, \dots, p_M\}$ el conjunto de puntos en el contorno del jeroglífico (o en el jeroglífico adelgazado) y P' un subconjunto de n puntos $P' = \{p'_1, \dots, p'_n\}$ obtenidos a partir de P , para cada p'_i se calcula un histograma x_i , el cual contiene en la r -ésima posición el número (normalizado) de puntos en P' que caen en la r -ésima región de un diagrama logarítmico-polar con origen en p'_i . El diagrama logarítmico-polar consta de 12 regiones angulares y 5 intervalos de distancia, cubriendo en total el doble de la distancia promedio entre todos los pares de puntos en P' .

La r -ésima entrada en x_i es definida como:

$$x_i(r) = \frac{|P'_i{}^r|}{|P'| - 1} \quad r = 1, \dots, 60 \quad (\text{A.1})$$

donde $|\cdot|$ es el operador de cardinalidad, r identifica a la región en el diagrama logarítmico-polar y $P'_i{}^r$ denota al conjunto de puntos en P' que caen en la región r con respecto al punto p'_i :

$$P'_i{}^r = \{p'_k \in P' : p'_k \neq p'_i, p'_k - p'_i \in r\} \quad (\text{A.2})$$

donde $p'_k - p'_i$ denota el vector diferencia. El descriptor h_i es de dimensión 60, resultado de las 60 regiones en el diagrama logarítmico-polar. En la (ver *Figura A.1*) ilustramos al descriptor SC calculado a partir de un punto p'_i en un jeroglífico Maya.

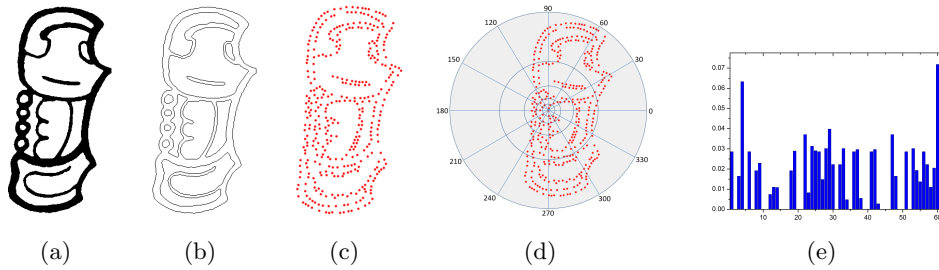


Figura A.1: Descriptor SC: (a) Jeroglífico Maya, (b) Contorno del jeroglífico, (c) Puntos obtenidos a partir del contorno, (d) Diagrama logarítmico-polar con origen en un punto p'_i , (e) Descriptor SC para un punto p'_i .

A.2. *Histogram of Orientation Shape-Context*

El descriptor *Histogram of Orientation Shape-Context* (HOOSC) [Roman-Rangel et al., 2011a] es un descriptor local de forma y apariencia, el cual resulta al combinar los descriptores SC [Belongie et al., 2002] y HoG [Dalal and Triggs,

2005].

Sea $P = \{p_1, \dots, p_M\}$ el conjunto de puntos en el contorno (o en la forma adelgazada) y $P' = \{p'_1, \dots, p'_n\}$ el conjunto de puntos de interés, obtenidos a partir de P , a menudo mediante selección aleatoria uniforme. Cada jeroglífico es representado a través de n descriptores locales HOOSC, cada descriptor local x_i $i = \{1, 2, \dots, n\}$ es un vector que describe al punto de interés p'_i en función de los puntos en P .

Cada descriptor local x_i es calculado mediante un diagrama logarítmico-polar con origen en el punto p'_i . El diagrama logarítmico-polar consta de 12 regiones angulares y 5 intervalos de distancia, cubriendo en total el doble de la distancia promedio entre todos los pares de puntos en P . Lo anterior resulta en un diagrama logarítmico-polar formado por 60 regiones $r = \{1, \dots, 60\}$ (ver Figura A.2a).

Cada región r contiene al subconjunto de puntos de P que caen en dicha región, respecto al punto p'_i , ese subconjunto de puntos es denotado como P_i^r .

$$P_i^r = \{p_j \in P : p_j \neq p'_i \in r\}$$

La región r es caracterizada a través de un histograma de orientaciones locales en P_i^r . El histograma contiene la distribución de las orientaciones locales de los puntos en P_i^r calculadas mediante un enfoque basado en kernel que estima la densidad de la orientación. Explícitamente, la densidad para el ángulo θ en la región logarítmico-polar r del punto de interés p'_i es denotada por $h_i^r(\theta)$

$$h_i^r(\theta) = \sum_{p_j \in P_i^r} \mathcal{N}(\theta, \theta_j, \sigma^2)$$

donde $\mathcal{N}(\theta, \mu, \sigma^2)$ es el valor del ángulo θ de una Gaussiana de media μ y varianza σ^2 ($\sigma = 10^\circ$). El histograma h_i^r es codificado en 8 rangos, el valor en cada

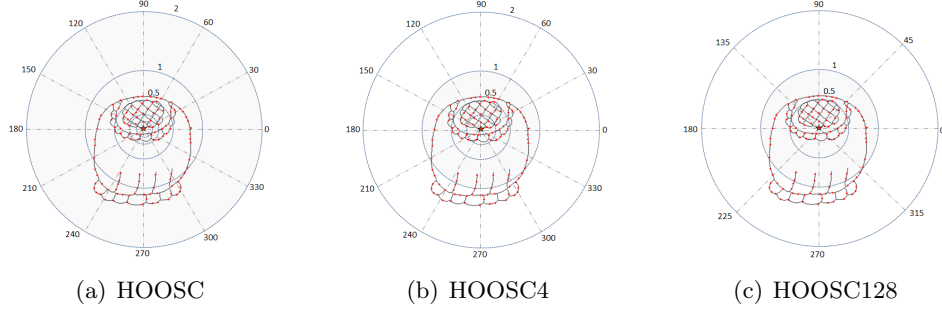


Figura A.2: Descriptor *Histogram of Orientation Shape-Context* (HOOSC)

rango es obtenido al integrar la densidad $h_i^r(\theta)$ en el intervalo $[a, b]$. El descriptor local x_i obtenido a partir del punto de interés p_i' se obtiene concatenando los 60 histogramas h_i^r , dando lugar a un descriptor de dimensión 480.

$$x_i = [h_i^1, h_i^2, \dots, h_i^{60}]$$

En la literatura, se han calculado descriptores locales HOOSC considerando diferentes configuraciones en el diagrama logarítmico-polar: HOOSC [Roman-Rangel et al., 2011a](ver *Figura A.2a*), HOOSC4 [Roman-Rangel et al., 2011b](ver *Figura A.2b*), HOOSC128 [Roman-Rangel and Marchand-Maillet, 2014b](ver *Figura A.2c*).