Image Similarity Comparison Based on Hong Kong Historic Carvings

Dailin Gan
Supervisor: Prof. Jeff J.F. Yao

1. Abstract

We present a classification method for analyzing the similarities between Hong Kong historic carvings based on wavelet decomposition and hierarchical clustering. It may not be easy to recognize the time when historic carvings were made if some background information was lost. However, if newly found carvings could be compared with fully analyzed carvings, the similarities between those carvings may provide hints of the dates when new carvings were made. After extracting image patterns with the help of MATLAB Image Segmenter APP, we used 2-dimensional wavelet decomposition techniques to obtain image information. Then, we compared image information by t-test and grouped those provided images by hierarchical clustering. We demonstrate the effectiveness of our approach by the p-value calculated from t-test, and the clustering trees. The results showed that our methods might be helpful for image comparison and clustering.

2. Introduction

How to decide the dates when newly found carvings were made is always of great concerns. However, this problem could be quite hard if related information of carvings is impaired or missing, which could stop us from speculating the time period. In this report, we proposed techniques which may address this challenge by virtue of wavelet decomposition and hierarchical clustering. We first extracted line patterns in our images and standardized them into the same size. Then, 10 equal size sub-images were randomly draw for each image, and the wavelet coefficients of each sub-image were obtained with the same resolution. Furthermore, image distance could be calculated for each pair of sub-images based on their wavelet coefficients (Daubechies, 1992). Finally, hierarchical clustering could be conducted based on image distance matrix. This report will continue as follows. Firstly, detailed information about materials and specific methods will be discussed. Besides, the results will be shown with analysis. Finally, discussion and potential reflection will be provided in the end.

3. Materials and Methods

Thirteen images of Hong Kong historic carvings were provided, and their line patterns were previously recognized and enhanced manually. All 13 images were used in this project

and their re-scaled binary images can be found in the Appendix.

3.1. Image Pre-processing

Line patterns were extracted in the Image Segmenter application within MATLAB. We only need to specify the foreground and background of each image, and line patterns will be recognized. The chosen image will be turned into a binary image, and MATLAB code will be generated automatically. Apart from that, we standardized each image into the same size and randomly draw 10 equal size sub-images for each of them. In our settings, the standardized binary input images are all 400×400, and the size of randomly chosen sub-images is 350×350. It should be noted that since carving images may be obtained with different sizes, the comparison between images with different may be hard. Therefore, our method may be effective when the size of standardized images is 400×400. However, if the size is switched into 200×200, for instance, the final results may be biased. It should also be noted that overlapping between sub-images is allowed. Totally, we had $13 \times 10 = 130$ sub-images in our experiment.

3.2. Wavelet Decomposition

For 1-dimension signal processing, we can treat the signal as a function, which could be reconstructed as the summation of wavelet functions multiplying their coefficients, in which the coefficients could be used to represent the signal information. The formula for 1-dimensional wavelet decomposition is that

$$f(t) = \sum_{j,k} \gamma(j,k)\psi_{j,k}(t). \tag{1}$$

In 2-dimension cases, we can similarly treat images as a function with two variables and each image could be reconstructed as a summation of wavelet functions multiplying their coefficients for a given resolution (Mallat, 1989). The formula is that

$$\begin{aligned} f(x,y) = & \frac{1}{\sqrt{MN}} \sum_{m} \sum_{n} W_{\varphi}(j_0,m,n)\varphi_{j_0,m,n}(x,y) \\ & + \frac{1}{\sqrt{MN}} \sum_{i=H,V,D} \sum_{j=j_0}^{\infty} \sum_{m} \sum_{n} W_{\Psi}^{i}(j,m,n)\Psi_{j,m,n}^{i}(x,y). \end{aligned} \tag{2}$$

It should be noted that the summation will go to infinity and we only consider finitely many cases.

In our case, we chose the Haar wavelet function and level 2 resolution. At level 1 resolution, images will be decomposed into four parts, which are approximated image A_1, horizontal detail H_1, vertical detail V_1 and diagonal detail D_1. In the next resolution, A_1 will be decomposed into another four parts, A_2, H_2, V_2 and D_2. Here, A, H, V and D are all vectors. We only use A_1 and A_2 as the wavelet decomposition coefficients to represent an image's information. Because in the original formula of 2-dimensional wavelet decomposition formula, H, V and D are treated as residual terms and only A contains most of the image information. We applied this decomposition method for all 130 sub-images and we chose to use their image coefficient vectors A_1 and A_2 in the following calculations.

3.3. Similarity Comparison

Considering different sub-images' wavelet coefficients may have different scale which may affect the performance of our method, we firstly normalized coefficients in A_1 and A_2 for each image. The normalising formula is

$$A_i^{new} = \frac{A_i}{\sqrt{\sum_j (A_j)^2}}, \tag{3}$$

where A_i represents the i-th term in the original vector A. A^{new} is the vector that we used for further calculation.

Then we defined the image distance based on Euclidean distance, which is

$$d(I_1, I_2) = ||A_1 - B_1||_2^2 + ||A_2 - B_2||_2^2 \tag{4}$$

where I_1 and I_2 represent two different images. A and B stand for approximated images for I_1 and I_2 respectively. A_i is the approximated image at level i-th resolution, where $i = 1, 2$ in our case.

For each pair of 130 sub-images, we calculated their image distance and obtained a 130×130 distance matrix D. Then, we divided all the distance into 2 groups, which are called Image-Within (IW) and Image-Between (IB). In IW group, the distance was obtained from a pair of sub-images which were from the same original image. While in IB group, the distance was calculated from a pair of sub-images which were from different original images. In total, there are 585 samples in IW group and 7800 samples in IB group. Furthermore, we assumed that both groups follow normal distribution where mean and variance were estimated based on their sample mean and sample variance. Then, we want to use t-test to verify whether there is significant difference between IW and IB group.

3.4. Hierarchical Clustering

With the distance matrix D, hierarchical clustering was performed with the defined metric, which is also called the farthest neighbor

$$d(r, s) = max(dist(x_{ri}, x_{sj})), i \in (1, 2, ..., n_r), j \in (1, 2, ..., n_s) \tag{5}$$

where r and s are two groups. x_{ri} and x_{sj} are elements in r and s respectively. This metric chooses the largest distance between all the pairs from the two groups r and s. The clustering process will stop until all the elements are clustered into one group.

We conduct image clustering based on two samples. The first clustering tree is based on 130 sub-images, where the distance matrix D is used to draw the tree. The second clustering is based on 13 original images, where another distance matrix K is used to draw the tree. The distance matrix K is defined in the following.

Suppose A and B are two different original images and $A_1, A_2, \ldots, A_{10}, B_1, B_2, \ldots, B_{10}$ are their sub-images respectively. We can obtain a 10×10 distance matrix based on these

sub-images, where each entry is $d(A_i, B_j)$, $i = 1, 2, \ldots, 10; j = 1, 2, \ldots, 10$. Then, the average value of all the entries in the 10×10 matrix is an entry in the matrix K, indicating the distance between the original images A and B. Results of clustering trees and related heat maps with relevant explanation will be provided in the next section.

4. Results

4.1. P-value results

For one single experiment, we randomly draw 10 sub-images for each original image, and calculated the image distance as well as the p-value. We repeated the experiment ten times, and the averaged p-value is $3{,}8116 \times 10^{-21} < 0{,}05$ for the t-test between IW group and IB group. We may claim at 0.05 significance level that two groups are significantly different and can be separated easily. For illustration purpose, one experiment was chosen for the following analysis. The distributions of image-between & image-within can be found in Figure 1, where blue stands for image-within sample, and red stands for image-between sample. In Figure 2, the box plot is also provided. It may also show that our method could distinguish the image-within group and image-between group effectively.

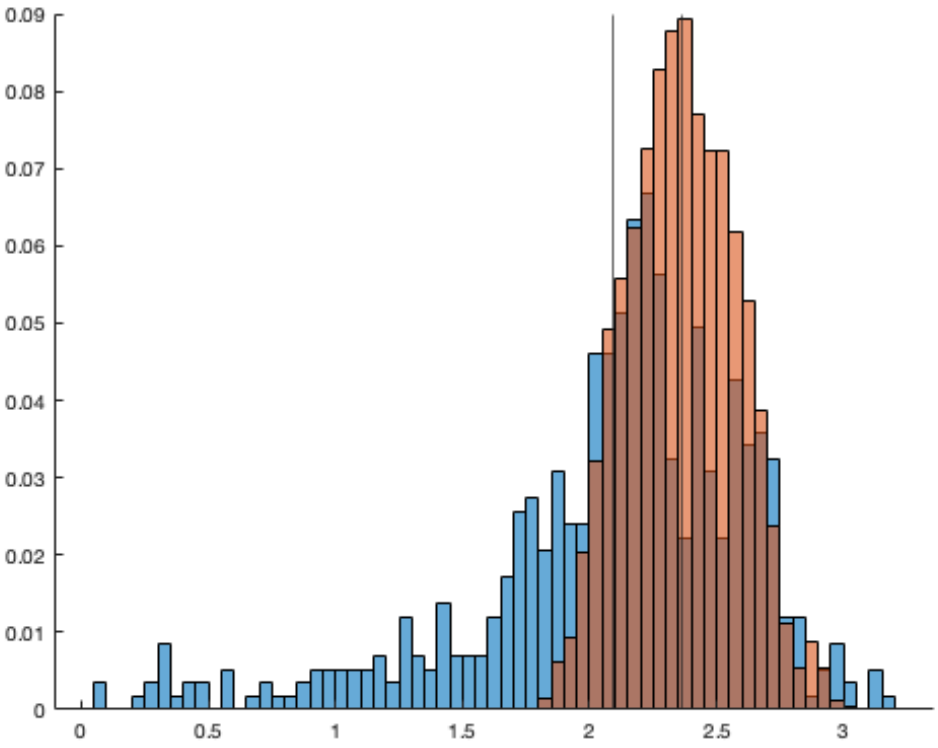


Figura 1: Distributions of image-within & image-between samples

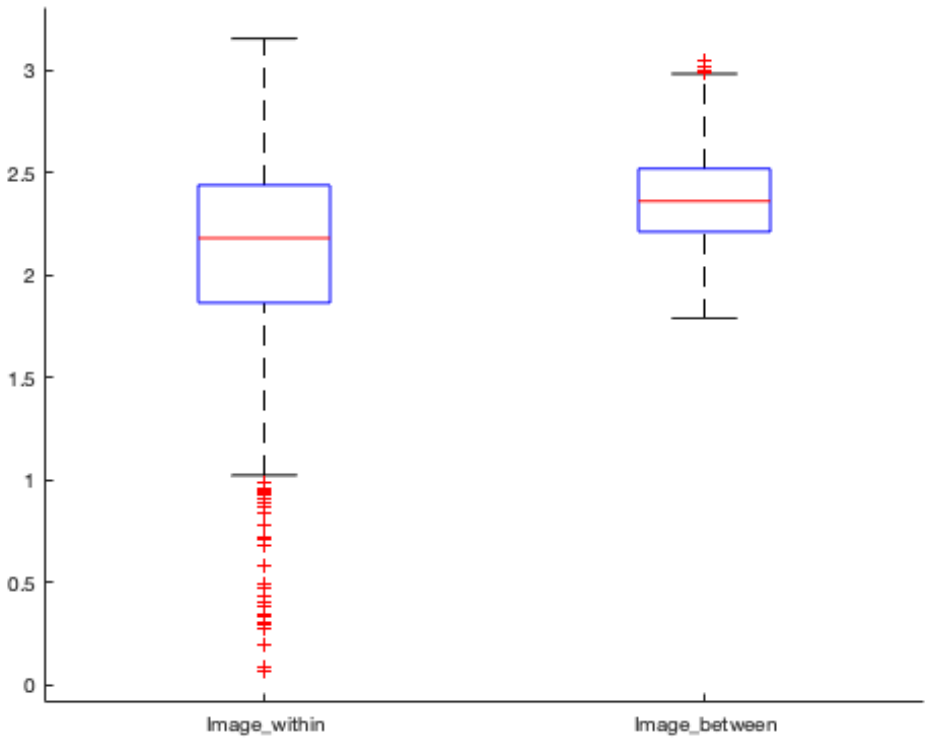


Figura 2: Box plot of image-within & image-between samples

4.2. Clustering results

4.2.1. Heat map

Figure 3 is the symmetric distance heat map of 130 sub-images. The x-axis and y-axis both represent 130 sub-images. Dark blue means that the distance between two sub-images are small while dark red means the distance is large. It should be noted that $\{1, 2, \ldots, 10\}$ represents one image, $\{11, 12, \ldots, 20\}$ represents another one and so on. It could be noticed that the diagonal 10×10 squares represent the original 13 images. Besides, it can be found that light red and blue colors are mainly concentrated along the diagonal squares, which means the differences within each original image may be small.

For illustration purpose, another averaged heat map of 13 images is also provided in Figure 4, where each square is calculated based on a 10×10 square in Figure 3. It could be consistently found that the diagonal entries are relatively smaller than the off-diagonal entries.

4.2.2. Hierarchical Tree

Figure 5 is the clustering tree of all 130 sub-images. The horizontal axis stands for the indices of the total 130 sub-images. The upside-down U-shaped lines represent the links

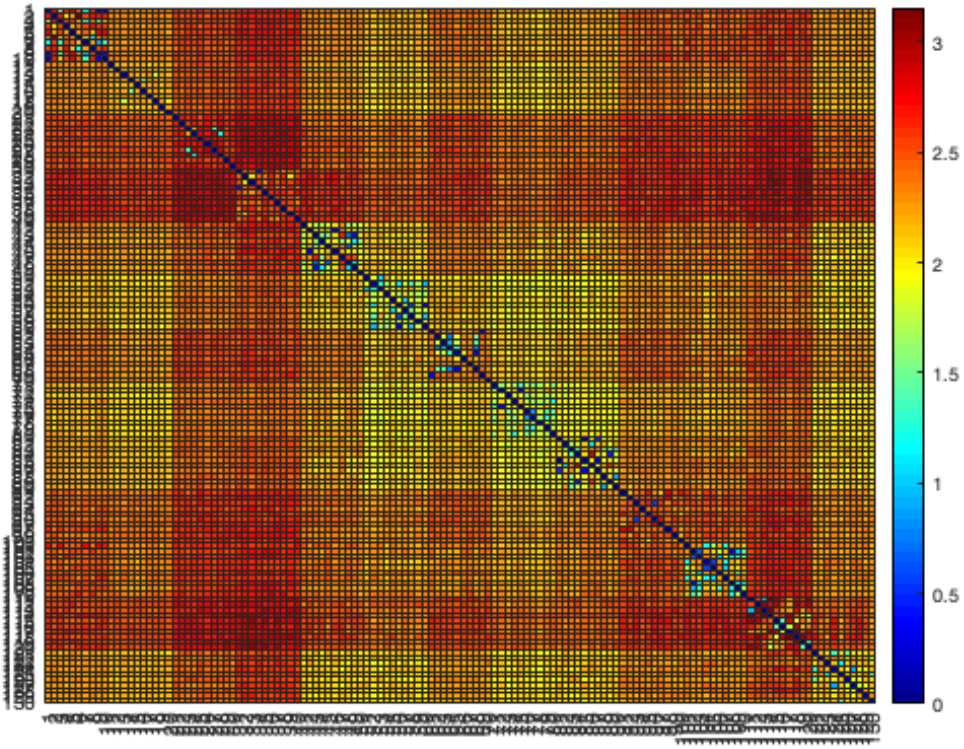


Figura 3: 130 sub-images heat map

between two groups. The vertical axis means the distance between groups. We may treat the clustering problem as a semi-supervised problem since we know that at least 10 sub-images from the same original image should be grouped together if our method is appropriate. Equivalently speaking, we should find the pattern that 10 sub-images from the same original image are grouped together.

In order to show whether 10 sub-images from the same original image are grouped together, a summary table is provided with 13 clusters. It can be noticed in Figure 6 that most images have the pattern that one of those clusters contain at least half of their sub-images, which is the expected semi-supervised pattern. Therefore, at least our method could recognise those sub-images from the same original images.

5. Discussion

Results showed that our method can successfully distinguish sub-images from the same image group, and it may have the potential to accomplish hierarchical clustering in order to give a detailed classification result. There are some problems which may affect performance in the clustering process. Firstly, we resized each image into the same size, which may not conform to the reality. Because each original image may be photographed with

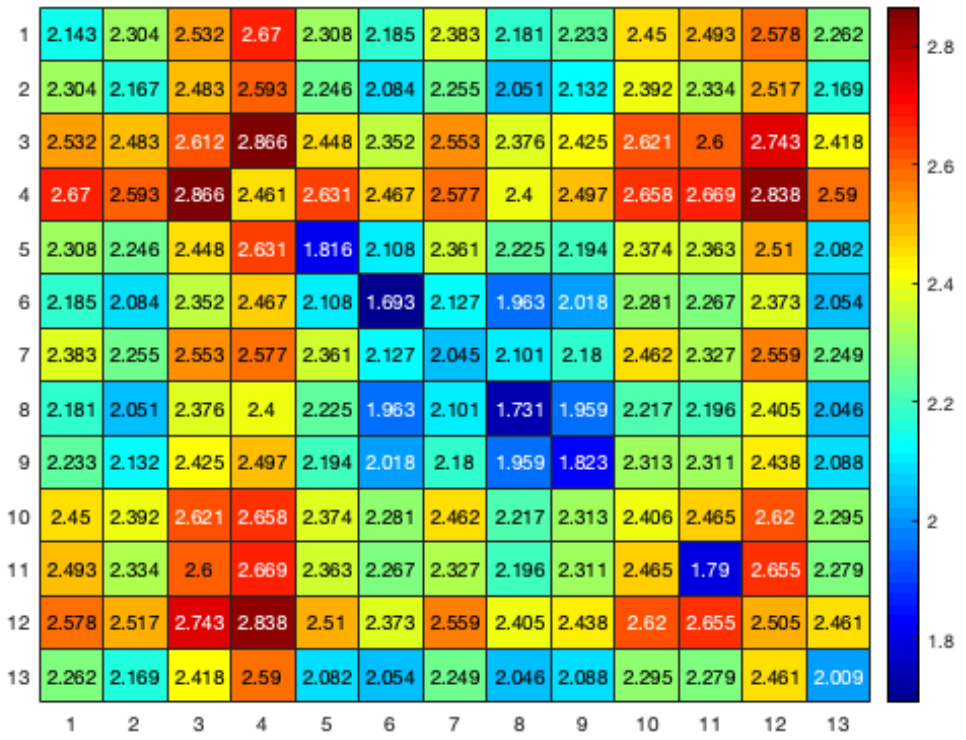


Figura 4: 13 original images heat map

different distance from the camera and the scale of their line patterns may be different, which means those patterns may not be compared at the same level. Secondly, sub-image difference may exist in carving picture. For a given original image, four sub-images may also be quite different, which may cause self-variation. Hence, more modifications may be required in order to solve the image self-variation problem.

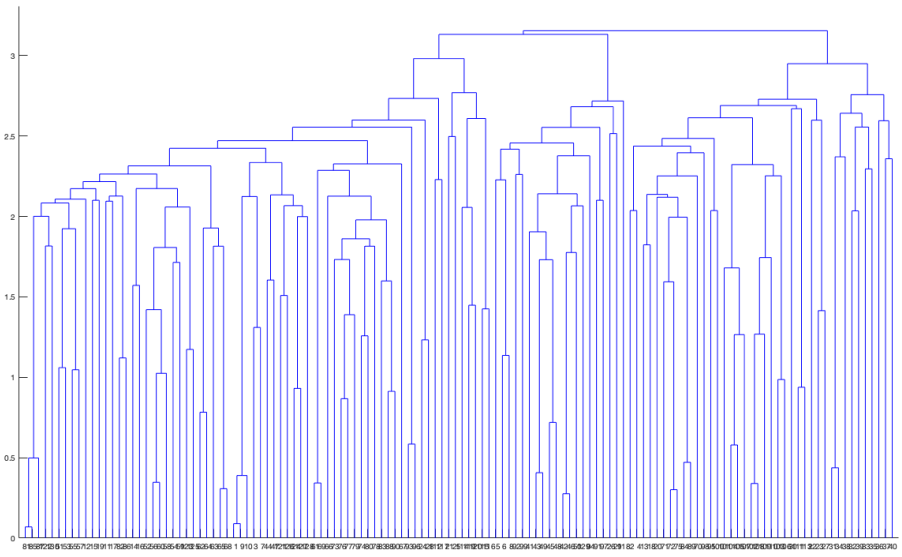


Figura 5: 130 sub-images clustering tree

Counts	Image_1	Image_2	Image_3	Image_4	Image_5	Image_6	Image_7	Image_8	Image_9	Image_10	Image_11	Image_12	Image_13
Cluster_1	0	0	1	0	0	0	0	0	0	0	0	0	0
Cluster_2	0	0	0	0	0	0	0	0	0	0	0	2	0
Cluster_3	0	0	2	0	0	0	0	0	0	0	0	0	0
Cluster_4	3	0	0	0	**8**	0	0	0	0	**5**	0	0	1
Cluster_5	2	3	0	0	0	0	1	3	2	3	**10**	0	0
Cluster_6	0	0	0	0	0	0	0	0	0	0	0	1	0
Cluster_7	0	0	3	0	0	0	0	0	0	0	0	0	0
Cluster_8	0	0	0	0	0	0	0	0	0	0	0	2	0
Cluster_9	**5**	**7**	2	0	2	**10**	**9**	**7**	**8**	2	0	0	**9**
Cluster_10	0	0	0	3	0	0	0	0	0	0	0	0	0
Cluster_11	0	0	0	**7**	0	0	0	0	0	0	0	0	0
Cluster_12	0	0	2	0	0	0	0	0	0	0	0	0	0
Cluster_13	0	0	0	0	0	0	0	0	0	0	0	**5**	0

Figura 6: 13 clusters summary table

6. Acknowledgement

This report is written as a summary of the work done in the summer of 2019 under the supervision of Prof. Jeff J. F. Yao. I sincerely thank Prof. Yao for his valuable suggestions and kind support and thank Prof. Chan Lung Sang for kindly providing carving pictures and giving some insightful advice in this project.

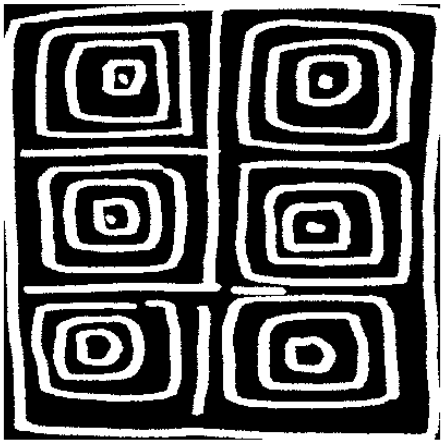
7. References

[1] Daubechies, I. *Ten Lectures on Wavelets*, CBMS-NSF Regional Conference Series in Applied Mathematics. Philadelphia, PA: SIAM Ed, 1992.

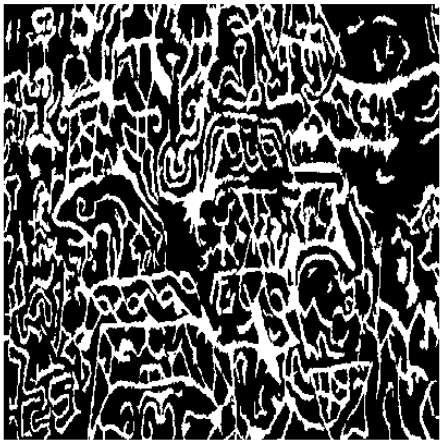
[2] Mallat, S. G. "A Theory for Multiresolution Signal Decomposition: The Wavelet Representation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*. Vol. 11, Issue 7, July 1989, pp. 674–693.

8. Appendix

This section includes the binary extracted and re-scaled carving images.



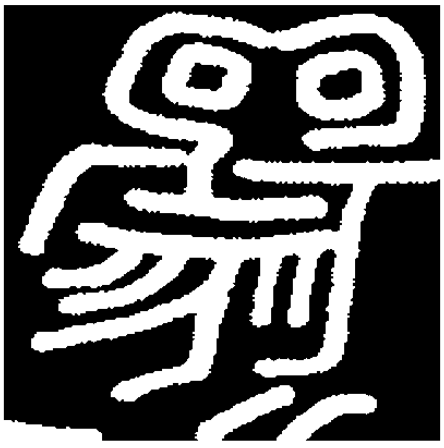
(a) Pic 1



(b) Pic 2

(c) Pic 3

(d) Pic 4

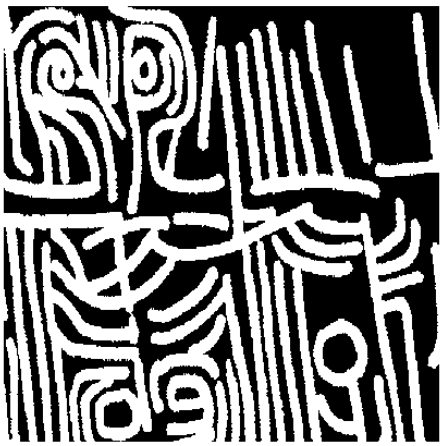


(e) Pic 5

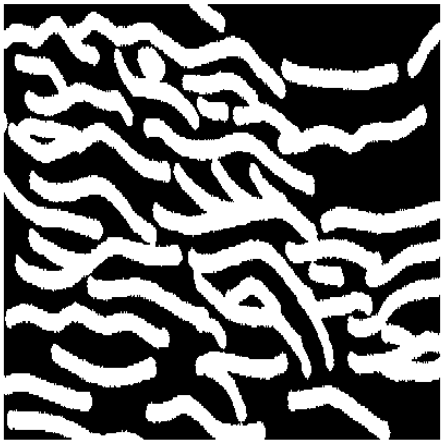
(f) Pic 6

Figura 7: Extracted and re-scaled carving images

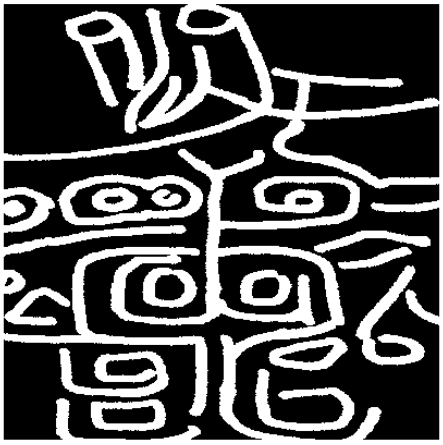
(a) Pic 7



(b) Pic 8

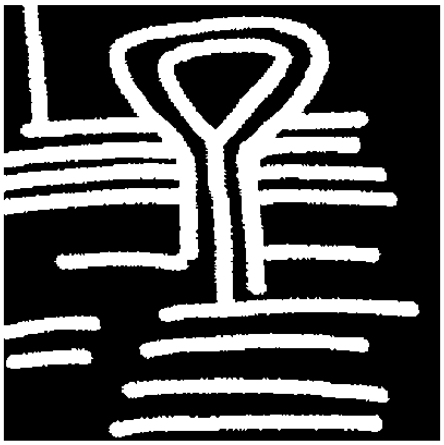


(c) Pic 9

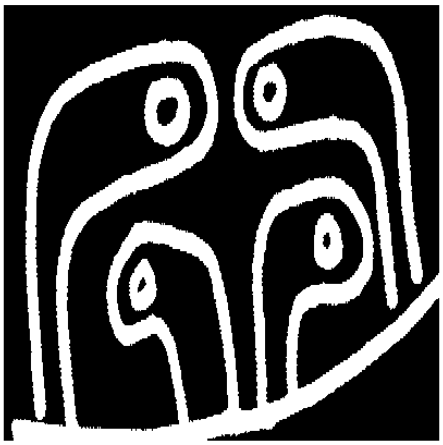


(d) Pic 10

Figura 8: Extracted and re-scaled carving images



(a) Pic 11



(b) Pic 12

(c) Pic 13

Figura 9: Extracted and re-scaled carving images