Open Access **Baghdad Science Journal** **P-ISSN: 2078-8665**
2020, 17(2 Special Issue)NICST:694-700 **E-ISSN: 2411-7986**

DOI: http://dx.doi.org/10.21123/bsj.2020.17.2(SI).0694

Content Based Image Retrieval (CBIR) by Statistical Methods

Fathala Ali [*] ***Alyaa Hashem Mohammed***

Received 8/10/2019, Accepted 23/6/2020, Published 23/6/2020

Abstract:

An image retrieval system is a computer system for browsing, looking and recovering pictures from a huge database of advanced pictures. The objective of Content-Based Image Retrieval (CBIR) methods is essentially to extract, from large (image) databases, a specified number of images similar in visual and semantic content to a so-called query image. The researchers were developing a new mechanism to retrieval systems which is mainly based on two procedures. The first procedure relies on extract the statistical feature of both original, traditional image by using the histogram and statistical characteristics (mean, standard deviation). The second procedure relies on the T- test to measure the independence between more than images, (coefficient of correlate, T- test, Level of significance, find the decision), and, through experimental test, it was found that this proposed method of retrieval technique is powerful than the classical retrieval System.

Key words*:* Content Based Image Retrieval, Histogram statistical characteristics, Test of- T, Trademark Image Retrieval .

Introduction:

With the rapid increase in the amount of registered trademark images around the world, trademark picture recovery (TIR) has developed to guarantee that new trademarks don't rehash any of the immense number of trademark picture put away in the trademark enlistment framework, hence trademark recovery framework ought to be intended to guarantee this uniqueness of trademark pictures (1).Content-based picture recovery (CBIR), is the use of PC vision procedures to the picture recovery issue, that is the issue of scanning for computerized pictures in huge databases .CBIR is also known as : Query by Image Content (QBIC) and Content-Based Visual Information Retrieval (CBVIR), CBIR is a functioning zone of research with a considerable lot of consideration (2).

The research deals with methods of retrieving images from trademarks and methods of measuring the similarity of trademarks in addition to the most important definitions. The research also addressed the use of two important criteria depending on the content, namely the graph scale and the statistical standard (test-t) of the image, in addition to the most important conclusions and recommendations.

Administration and Economics, Al-Mustansiriya University, Baghdad, Iraq.
[*]Corresponding author: aiiiii_f@yahoo.com
[*]ORCID ID: https://orcid.org/0000-0003-1751-2486

The study aims to shed light on the process of retrieving the image from a set of images based on the graphic content, using the histogram and testing (T), through the application of the Matlab program.

Methods for Trademark Image Retrieval:

Trademark Image Retrieval (TIR) framework utilize the visual substance of a picture. In average Trademark Retrieval System the visual substance of the pictures in database are removed and portrayed by multi-dimensional component vectors. These component vectors of the picture in the database structure the element database.

Trademark Image Retrieval strategies can be comprehensively ordered as(3):

Shading based trademark picture recovery.
Surface based trademark picture recovery.
Shape based trademark picture recovery

Shading is one of the most significant highlights of the picture that can make the acknowledgment of pictures conceivable by people prevalently. Typically, hues are characterized in three-dimensional shading spaces. These could either be Red, Green, and Blue (RGB), Hue, Saturation, and Value (HSV), or Hue, Saturation, Brightness trademark picture retrieval.(HSB). In a Trademark Retrieval framework, shading data of pictures is spoken to by shading histograms. A

shading histogram is a kind of reference diagram, where each bar represents a specific shading in the shading space utilized.

Surface is the inborn property of all surfaces that portrays visual examples and each having the properties of homogeneity. It contains significant data about the basic game plan of the surface like mists, leaves, blocks, and so on for portraying the surface; three standards are utilized statistical, structure and phantom strategies.

Measurable systems describe surfaces utilizing the factual properties of the dim degrees of the focuses/pixels containing a surface picture. Commonly, these properties are registered utilizing the dim level event lattice of the surface, or the wavelet change of the surface.

Contributions:

Researchers focus on Trademark similarity defined in terms of visual similarity and skip the conceptual/ semantic similarities. Visual similarities of trademarks include color, shape and texture aspects.

The accompanying Contributions:

*Large-scale dataset a benchmark: has been stretched out with more trademarks and better question tests with which trademark recovery frameworks can be tried and looked at.

*An examination of visual highlights and a gauge.

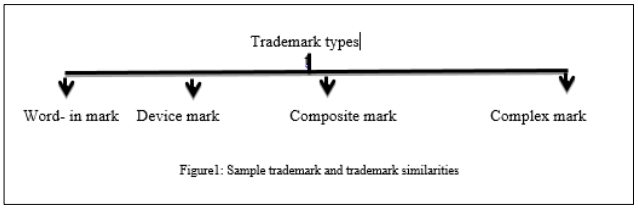
Trademark Similarity measuring Methods

Trademarks are particular visual images with high reputational esteem, because of the view of value and development related with them.

Plate 1: Photo of some trademark examples

.

So, there is a need for trademark protection by providing a solution to prevent infringement. This issue can be tended to by creating trademark recovery frameworks equipped for looking at the visual similitude of trademarks (Plate 1).



Scheme of the visualsimiltude of tradmarks

Digital image

The digital image is symbolized by a two-dimensional matrix, often consisting of small square pieces called image elements (points). The locations of these elements in the matrix correspond to the locations of the original image points represented by space coordinates, while the values of those elements are proportional to the light intensity value at those points, and The image can be represented mathematically, as shown in the following equation below

$f(x,y)=r(x,y).\ i(x,y)$

whereas

f (x, y) represents the digital image element in the location (x, y).

r (x, y) represents the reflection from the location (x, y), $1 \leq r(x, y) \leq 0$.

i (x, y) represents the intensity of the incident light.

Content-based image retrieval (CBIR) (3)

Content-based image retrieval (CBIR) is one of the vision application techniques. Image retrieval can be described as a task that searches for images in the image database, depending on a set of properties. CIBR is an important field of image processing and it has various applications in multimedia, crime prevention, internet, digital libraries, and entertainment. A number of previous works have been done for CBIR based on various features. Some works based on a single feature and some are based on the two or more combining features to get the result.

Image statistical Analysis (3)

One of the most important tasks of image analysis is to determine the necessary information extracted for the purpose of statistical data analysis, and image analysis includes the classification of the image into two types, namely wave classification depends on statistical characteristics and non-wave classification which largely depends on the degree of similarity between patterns

Average (4,5)

It represents the rate of clarity of the image or the specific region thereof, which can be found depending on a set of values over its number, and

the higher the image rate, the more clear it is, otherwise it is opaque.

Standard deviation*(4,5)*

The rate of deviations from the sharpness of image points from their average is called standard deviation, where the more the standard deviation or the variance of the image is, it is heterogeneous and dispersed, and vice versa is more homogeneous and not dispersed

Image histogram

It is a type of iterative expression that expresses the distribution of the color value in a digital imag. Looking at the graphic representation of a specific image, the viewer can quickly glimpse the total distribution of color values.

T- test (4,5)

This test is used when researchers want to compare the averages of two groups or two samples independent of each other

Formulate a null hypothesis: be as follows

There were no statistically significant differences at the level of significance (0.05, 0.01 or .....) between the first group (X) and the second group (Y) in the variable.

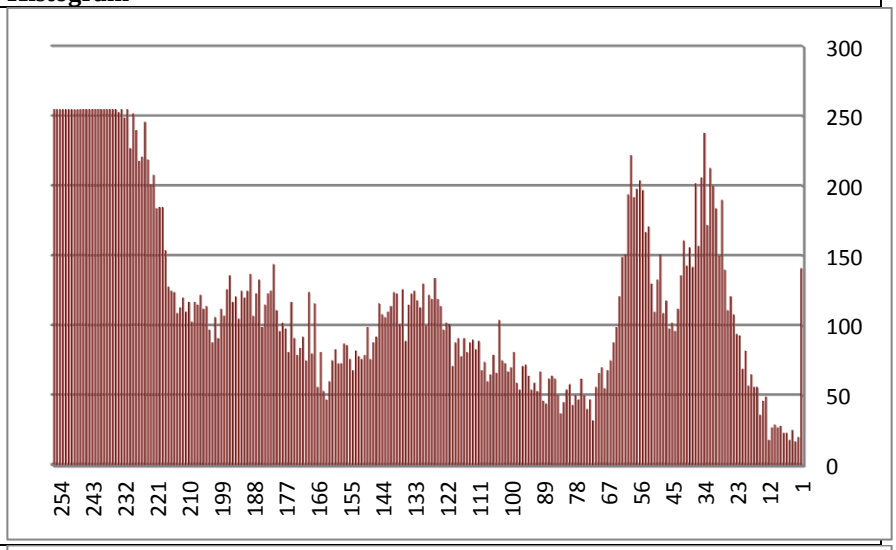
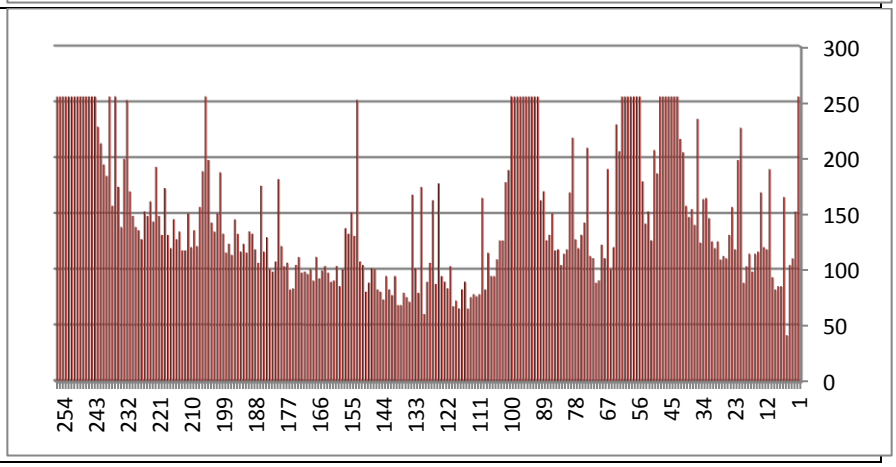
Formulation of alternative hypothesis: be as follows

There are statistically significant differences at α = 0.05, 0.01 or ...) between the first group and the second group in the variable.

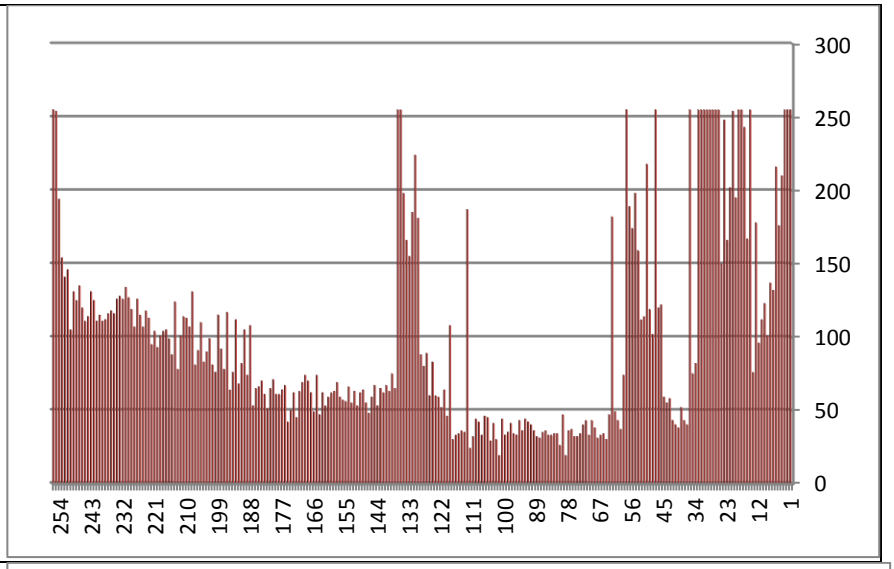
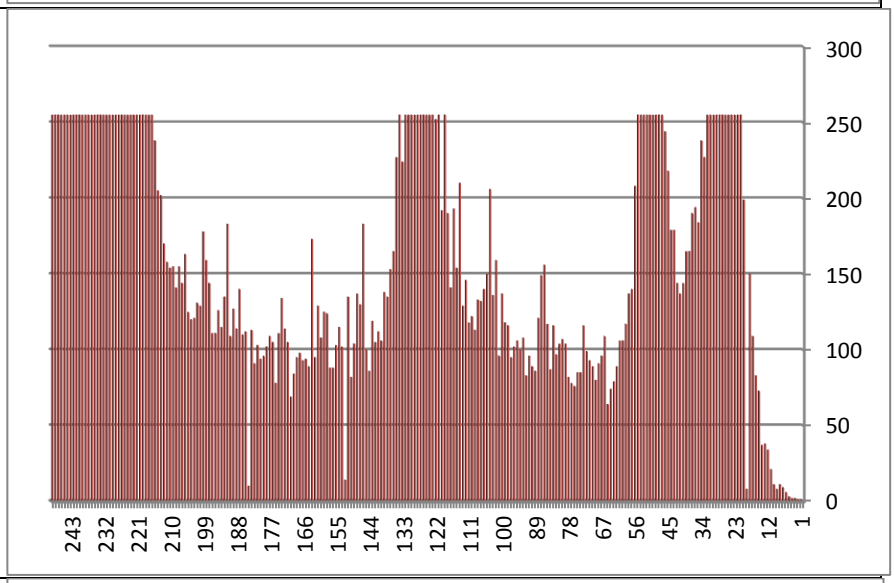
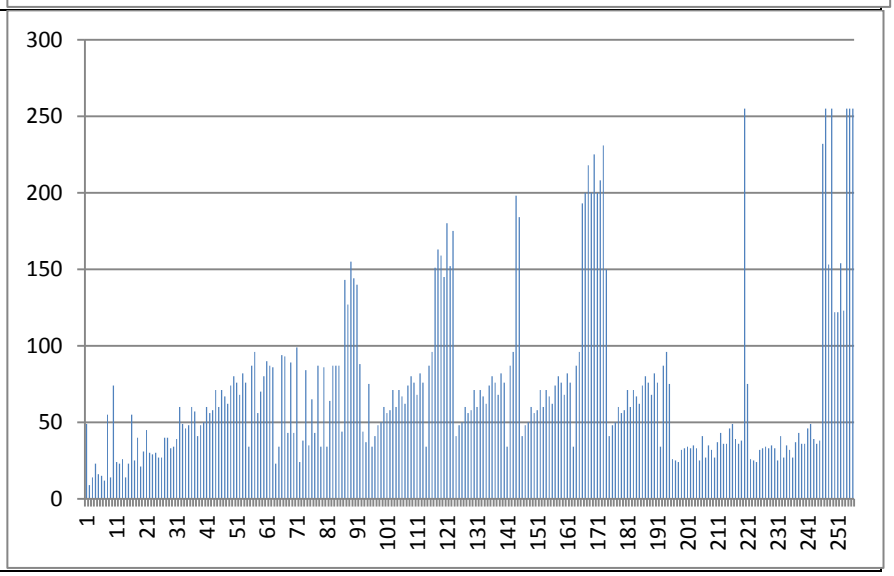
Practical

In this research, 7 pictures were studied, an image (A) that represents the basic image, while the rest of the images (B, C, D, E, F, G) represent traditional images(Table 1).

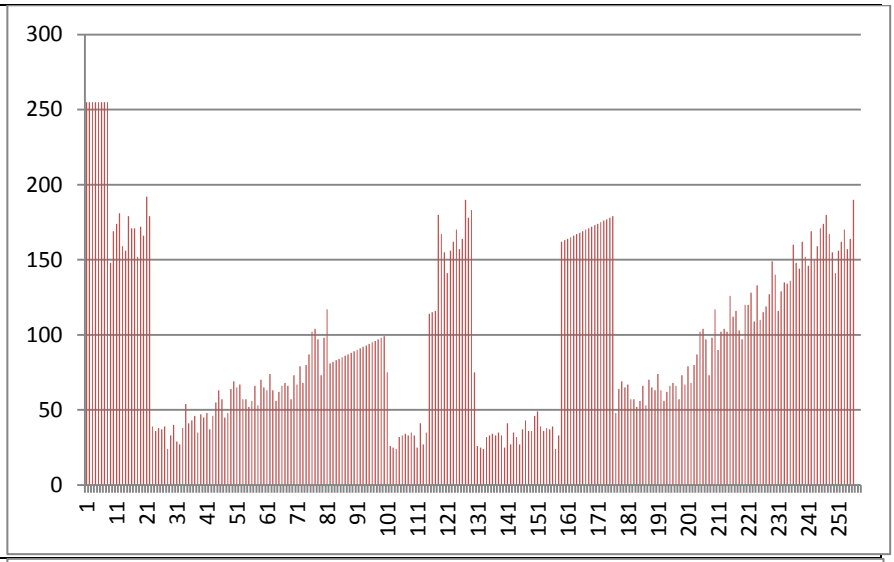
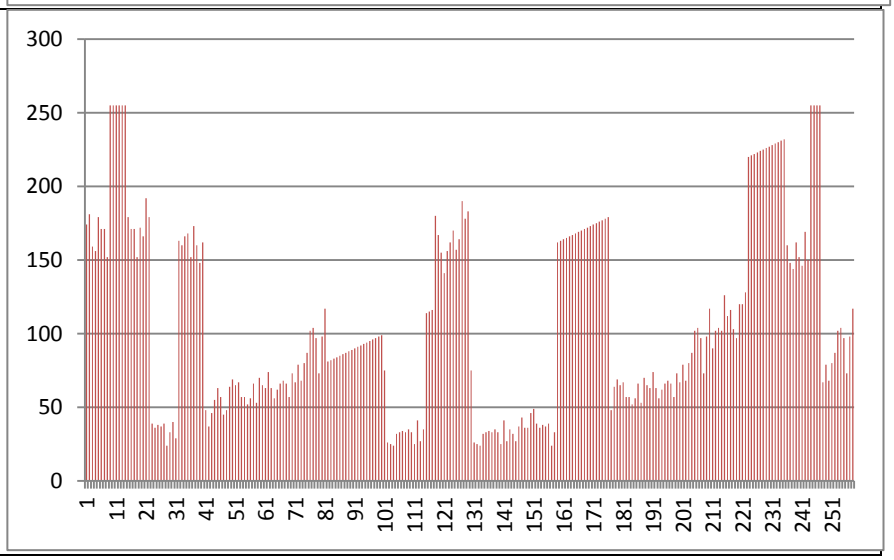
Table 1. HISTOGRAM IMAGE

	Image	Histogram
A		
B		

Baghdad Science Journal
P-ISSN: 2078-8665
E-ISSN: 2411-7986

C		
D		
E		

Baghdad Science Journal
P-ISSN: 2078-8665
E-ISSN: 2411-7986

F		
G		

The table above, it was found that the image (D) is close to the image (A) in terms of data content. Through studying the statistical characteristics of the images, it was found that the mean and the standard deviation of the main image (A) was (119.3945, 65.36921) respectively, while the mean and the standard deviation of the image (B) reached (99.6679, 66.79447). image(C) (148.4336, 60.26064), image (D) (155.5663, 65.30254), image (E) (89.608, 87.701), image (F) (48.487, 69.246), and image (G) (142.0 6, 70. 641)

In this research, a test (T) was used to find out the independence of two groups (the original image and its traditional forms) according to the following hypothesis

The hypothesis of nothingness H 0: $\mu_1 = \mu_2$

Alternative hypothesis $H_1 : \mu_1 \neq \mu_2$

To test the validity of the hypotheses, the study use a program SPSS. If the $\alpha > 0.05$, the study reject the null hypothesis (i.e. two groups (the two pictures) are independent)

Otherwise, will accept the null hypothesis at the level of significance 0.05 as shown in the Table 2 below

Table 2. T - TEST

	Coefficient of correlation	T test	Level of significance (α)
testing the independence between image 1,2	0.293	3.377	0.575
testing the independence between image 1,3	0.440	-5.226	0.891
testing the independence between image 1,4	0.6889	-5.954	0.000
testing the independence between image 1,5	0.293	2.357	0.715
testing the independence between image 1,6	0.341	-3.209	0.211
testing the independence between image 1,7	0.42	5.251	0.150

In table(2), the following decisions were reached

1- When testing the independence between image A,B , the test found that the value of calculated α is bigger than 0.05 , and therefore the study reject the Alternative hypothesis that the images are independent H 1: $\mu_A \neq \mu_B$

2- When testing the independence between image A,C, the test found that the value of calculated α is bigger than the 0.05, and therefore the study accept the Alternative hypothesis that the images are independent H 1: $\mu_A \neq \mu_C$

3- When testing the independence between image A,D the test found that the value of calculated T is smaller than the table , and therefore the study accept the null hypothesis that the images are dependent. H 0: $\mu_A = \mu_D$

4- When testing the independence between image A,E, the test found that the value of calculated α is bigger than the 0.05, and therefore can accept the Alternative hypothesis that the images are independent H 1: $\mu_A \neq \mu_E$

5- When testing the independence between image A,F the test found that the value of calculated T is smaller than the table , and therefore the study accepted the null hypothesis that the images are dependent. H 0: $\mu_A = \mu_F$

6- When testing the independence between image A,F the test found that the value of calculated T is smaller than the table , and therefore the study accepted the null hypothesis that the images are dependent. H 0: $\mu_A = \mu_G$

Conclusions:

1- When comparing the images using the difference coefficient , it was discovered that the coefficient contrast of the primary picture was 0.547 , and the coefficient distinction of the principal picture was 0.67. The coefficient contrast of the main picture was 0. 405, the four pictures, arriving at a distinction of 0.48.

2-There are no statistically significant differences between the target image (A) and the traditional image (D), and this indicates that (D) is the best and closest to the original image (A).

3-when comparing images (B,C,D,E,F,G) based on the average and the standard deviation, it was found that

The image (D*)* has the lowest standard deviation 65.30254, and the highest average deviation155.5663

Recommendations:

1-Use statistical methods to assess individual or family risks for a group of inherited conditions, such as genetic
disorders and birth defects. Provide information to health care providers at risk of inherited conditions.

2- Can use T-test to measure independence and comparison of medical images

3-can be utilized to the attributes of factual different pictures

Authors' declaration:

- Conflicts of Interest: None.
- We hereby confirm that all the Figures and Tables in the manuscript are mine ours. Besides, the Figures and images, which are not mine ours, have been given the permission for re-publication attached with the manuscript.
- Ethical Clearance: The project was approved by the local ethical committee in Al-Mustansiriya University.

References:

1. Jain AK, Duin RP, Mao J. Statistical pattern recognition: A review. IEEE Trans. Pattern Anal. Mach. Intell. 2000 Jan;22(1):4-37.
2. Fukunaga K. Introduction to statistical pattern recognition. Elsevier; 2013 Oct 22.
3. Shahla J H, Ahmed T , Mazen M S. Trademark Image Retrieval using Transfer Learning. J. Eng. Appl. Sci. 2019;14(18) :6897-6905.
4. Al-Achi A. The Student's t-Test: A Brief Description. Research & Reviews: Journal of Hospital and Clinical Pharmacy. 2019;5(1):1.
5. Mishra P, Singh U, Pandey CM, Mishra P, Pandey G. Application of student's t-test, analysis of variance, and covariance. Ann. Card. Anaesth. 2019 Oct;22(4):407.

Open Access **Baghdad Science Journal** **P-ISSN: 2078-8665**
2020, 17(2 Special Issue)NICST:694-700 **E-ISSN: 2411-7986**

استعادة محتوى الصور(CBIR) بواسطة اساليب احصائية

فاضلة علي جيجان **علياء هاشم محمد**

كلية الادارة والاقتصاد، الجامعة المستنصرية، بغداد، العراق.

الخلاصة:

نظام استرجاع الصور هو نظام كمبيوتر لتصفح الصور والبحث فيها واستعادتها من قاعده بيانات ضخمه من الصور المتقدمه. الهدف من أساليب استرجاع الصور المستندة إلى المحتوى (CBIR) هو أساسًا استخراج عدد محدد من الصور المتشابهة في المحتوى المرئي والدلالي ، من قاعدة بيانات كبيرة (للصور) إلى صورة الاستعلام المزعومة. كان الباحثون يطورون آلية جديدة لاسترجاع الأنظمة التي تعتمد بشكل أساسي على إجراءين. يعتمد الإجراء الأول على استخراج الميزة الإحصائية لكل من الصورة الأصلية والتقليدية باستخدام المدرج الإحصائي والخصائص الإحصائية (متوسط ,انحراف معياري). يعتمد الإجراء الثاني على قياس الاستقلال بين أكثر من صوره، (معامل الارتباط ، اختبار T ، مستوى الأهمية ، العثور على القرار) ، ومن خلال الاختبارات التجريبية وجد ان الطريقة المقترحة لتقنية الاسترجاع (T- اختبار) هو افضل من نظام استرجاع الكلاسيكية.

الكلمات المفتاحية: اختبار T، استرجاع الصور المستندة إلى المحتوى، استرجاع صورة العلامات التجارية، الخصائص الإحصائية المدرج الإحصائي.