Research Journal of Applied Sciences, Engineering and Technology 4(11): 1478-1483, 2012
ISSN: 2040-7467
© Maxwell Scientific Organization, 2012
Submitted: November 25, 2011 Accepted: January 04, 2012 Published: June 01, 2012

Multimedia Data Content Detection System

[1,2]Wenyin Ni, [1]Jianxin Wang and [2]Shunping Chen
[1]School of Information Science and Engineering, Central South University,
[2]Department of Computer Science and Technology, Changsha University,
Changsha 410003, China

Abstract: This study proposes an image content fast detection algorithm based on the extraction of the most robust features of the standard samples according to the characteristics of the image edge neighborhood grayscale variation. Moreover, a multimedia data content detection system mode based on Perceptual Hash is designed. Against the traditional method, the system has two improvements of the matching accuracy and fast query, and can be applied to multimedia video stream. Experimental results demonstrated the feasibility of this system.

Key words: Content detection, multimedia, perceptual hash

INTRODUCTION

The content query based on the multimedia database (MMDBMS) by the Perceptual Hash techniques, i.e. which changed the passive traditional inquiry in the database (DBMS) and active presentation in multimedia mode, is currently one of the main treatment methods of multimedia database management (Niu and Jiao, 2008). The main idea is: construct the Perceptual Hash by the DBMS characteristic value, which meansthecharacteristic valuesof data points in the smalldifferences mappingto the samehashvalue in a greaterprobability, oppositely thecharacteristic values of data points in the greatdifferences mappingto the differenthashvalue in a greaterprobability. So searching the neighbor point just can be exhaustive analysis the data points with the same hash value, its time complexity is $o(n^p \log_{1/p} n)$, $p<< 1$and $0< p< 1/2$(Gionis *et al*., 1999).

The invariant feature of the image statistical elements, such as the brightness of the histogram (Schneider, 1996), the variance value of the pixel, the wavelet coefficient (Kailasanathan and Naini, 2001), is a kind of the main method to structure the Perceptual Hash. Though these methods has better robustness on the slight disturbance image, the Perceptual Hash sensitivity is not up to the requirements when the image content is attacked and its safety can not be assured by reflect the image content changes well.

Another way is using the parameters which has an important impact on the image content and determine the perception characteristics of image in a HMM (Hidden Markov Models), for example the binary DC wavelet (Venkatesan *et al*., 2000), the low-frequency DCT coefficients (Mihcak and Venkatesan, 2001), the rotation invariance factor of Fourier transform (Swaminathan *et al*., 2006) and the strongest singular vector of singular value decomposition (SVD) (Kozat *et al*., 2004) etc.. These methods showed a strong robustness to the processing operations of unobvious perception image, but not good robust performance for local tampered or changed content.

The literature (Lin and Chang, 2001) proposed the invariance of the relationship between the DCT coefficients of different blocks in the same location even after quantization process to structure Perceptual Hash. The use of this relationship can effectively distinguish the JPEG compression and other malicious operations, but is not robustness well to the other conventional image processing. The literature (Lu and Liao, 2003) proposed a structure Perceptual Hash scheme to identify image, but this scheme can not resist the geometric attacks when the global changes didn affect the image perception (such as small angle rotation and twist). All of their main idea is using the coefficients keeps constant after wavelet transform or discrete cosine transform to construct the Perceptual Hash value.

The literature (Monga and Evans, 2006) used the typical two-stage Hash framework to extract the feature point, did NMF (non-negative matrix factorization) to sub images which split with the random square and did NMF to the images composed of the coefficient matrix and the matrix once again, get the non-negative matrix factorization vector and the corresponding random weight vector by choosing the coefficient matrix elements and base moment of array elements, at last construct the Perception Hash value by inner product through factorization the weight vector and non negative matrix vector. This method is more suitable for image tamper detection against other methods.

This study proposed a scheme which based on the neighborhood gray difference is great on the characteristics image edge and the Perceptual Hash value

Corresponding Author: Wenyin Ni, School of Information Science and Engineering, Central South University, Changsha 410083, China

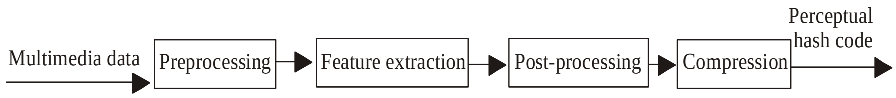

Fig. 1: The data flow diagram of image perception

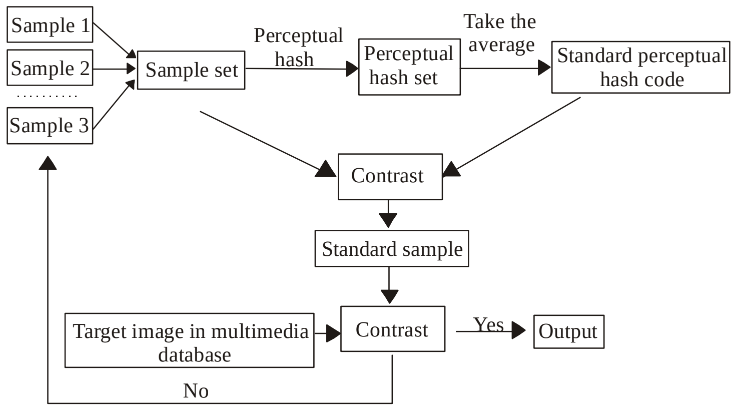

Fig. 2: The detection system of image content

is easy to determine, that choosing the most robust features of perception, i.e., the standard sample, from our sample set and making the standard sample and the content of the multimedia database image been rapidly detected. The improving of the algorithm has two main aspects both of the matching accuracy and fast query opposite to the traditional method, and the experimental results prove the feasibility of the method.

The multimedia database image content detecting system: Based on the information processing theory of cognitive psychology, we simulate the one-way mapping process from multimedia data sets to multimedia awareness summary sets by using Perceptual Hash for multimedia image digital abstract.

Figure 1 is a data flow diagram of image perception, the main purpose of preprocessing of multimedia data is primarily to remove the characteristic parameters low to human activity, and choose the most significant features of perception to facilitate computer processing, at last output the Perceptual Hash code after compression under the premise of keep good robustness.

In actual usage, we use the detection system of multimedia database content as shown in Fig. 2. Firstly, we structure Perceptual Hash set by sample set consisting of the sample 1, sample 2 ... sample k with the methods described as Fig. 1.

Each sample provides an approximation to the input image. The input image some parts, such as the background, may be well approximated by many similar samples in Perceptual Hash set. Parts which are better approximated by many other similar samples are less salient as they are commonplace among the relevant images in the Perceptual Hash set. Parts are more salient which is poor approximated. We measure the saliency S(x,y) as the absolute error $|I_{w,i}(x, y)$, I = 1, 2, …, k$|$ averaged across all samples $I_{w,i}(x, y)$, i = 1, 2, …, k and normalized to the range [0,1]:

$$E(x,y) = \frac{1}{k}\sum_{i=1}^{k}\left|I_{w,i}(x,y) - Q(x,y)\right| \tag{1}$$

$$S(x,y) = E(x,y) / \max_{x,y} E(x,y) \tag{2}$$

Through the comparison of S(x, y), we could choose the most robust feature perceptual content from the Perceptual Hash set, i.e., standard sample, standard sample is the most relevant digital abstract with our perceptual content. Because of its application purpose is to detection of the multimedia data content, there is no need to consider the safety. Instead of using the mean absolute error, we could use the other measure of goodness of fit, for example mean squared error. The experiment proves that they produce roughly similar results.

Next, we can extend the system to the application of streaming video content detection by contrast the standard sample and the video frames of streaming, which showed in Fig. 3.

Part of the detection algorithm: Due to limited space, here only part of the algorithm:

- Preprocess the target image
- Normalize the original image to the size of 512*512

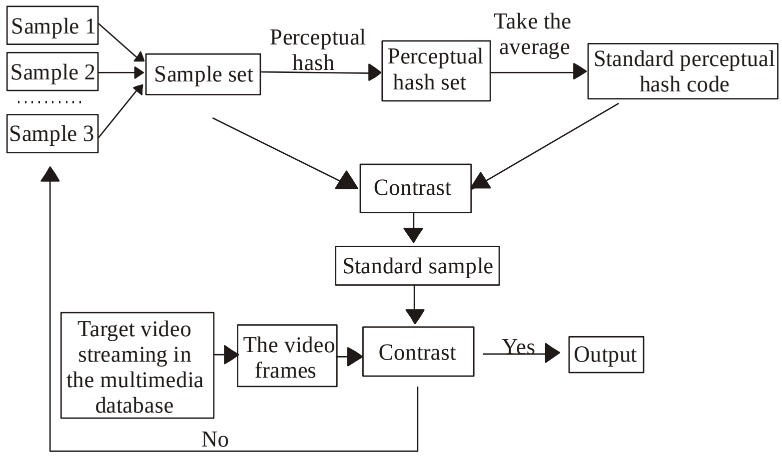

Fig. 3: The detection system of the video streaming content

- Segment the original mage into multiple blocks, each block of size n*n pixels. Remember I_1, I_2, I_l, which I_l is the perceptual hash code of the L block
- Filling I_1, I_2, I_l into a hash table after module
- Normalize the standard sample to the size of n*n, then calculate the sample's perceptual hash code, remember I
- Determine the position of I is empty or not, empty means fail, then return; not empty means succeed, output corresponding positionI_l

Analysis: In the stage of the image pre-processing, we could use the 8 direction of optimal edge method by characteristics of the image's neighborhood grayscale varies large on the edge to get the edge enhancement image G (Luo *et al*., 2004):

$$g(x,y) = d_k = \max \{d_i \mid i = 0,1, \ldots, 7\} \tag{3}$$

In which, $d_k > \varepsilon$, and d_i is the 3*3 difference value as the original image F in the direction of I at point . In the direction k, we could extract the maximum value to construct the thinning edge image G' from the edge enhancement image G. The threshold value t(i, j) corresponding to the edge point is calculated by the following formula (4):

$$t(I, j) = (f_{max} + f_{min}) /2$$

$$(f_{max} = \max \{f (p, q) \mid (p,q) \in S\}$$

$$f_{min} = \min \{f(p, q) \mid (p, q) \in S\} \tag{4}$$

Mathematically, S is t the neighborhood of edge point (I, j).

In the matching stage of the hash table, each image block calculate the hash value through the hash function g_1, g_2 ,…, g_1 , which is randomly selected K mapping among H and then put these image block into the hash table corresponding to their hash value location. We just search for the corresponding position of $g_1(q)$, $g_2(q)$,…, $g_1(q)$ in the hash table. The values of K and L will affect the speed and accuracy of the experiment. K increases, then the search speed is increased, but the accuracy will drop. Conversely, the increase of L would make the search accuracy increased, but the corresponding image block will also increase, thereby reducing the speed of search.

Suppose H is a group of sensitive mapping r_1, r_2, p_1, p_2 on metric space (S,d_x), q is to be retrieved and there presence k and l to make the following two events established by presumably rate:

- If there is $p^* \in B(q,r_1)$, then we could get some j = 1, 2, ..., l to ensure $g_j(p^*) = g_l(q)$;
- The number of image block whose distance to q is greater than r_2 and have the same hash value with q is less than 3L, i.e.:

$$\sum_{j=1}^{l} \left| (S - B(q, r_2)) \cap g_j^{-1}(g_j(q)) \right| < 3l \tag{5}$$

The average number of hash map above is $O(n^p \log_{1/p2} n)$, n is the number of image block, d is the dimension of data block , $\rho = \log 1/p_1 / \log 1/p_2$, and the space complexity is $O(dn + n^{1+p})$. In practical use, the values of K and L could be estimated by derived the number of data points from the data set (Andoni and Indyk, 2008).

Fig. 4: The original images

Fig. 5: The images after pro-processing

Fig. 6: The matching thumbnail (with RGB and without RGB)

EXPERIMENTAL RESULTS AND ANALYSIS

Experimental environment: Intel dual-core CPU 1.6GHz, 3GB DDR3 1066, algorithm using c #.

Experiment 1: In order to analyze the performance of this program, in Fig. 4, we randomly selected three 3264*2448 original image with subtle photography angle change, using 160*168 matching figure as content search. Figure 5 shows the images after pro-processing. Figure 6 shows the matching thumbnail (with RGB and without RGB). In experiments we added cut, Gaussian noise, blurring, scaling, etc. with a collection composition. The output results verify the feasibility.

Output:
1060 (the X coordinate value of starting point)
2320 (the Y coordinate value of starting point)
884 (the X coordinate value of starting point)
2260 (the Y coordinate value of starting point)
992 (the X coordinate value of starting point)
2296 (the Y coordinate value of starting point)

Experiment 2: This program will be extended to video streaming applications, and the match image is the image of Obama (Fig. 7), in accordance with the method in Experiment 1 we deal with the sub-frame (Fig. 8). Figure 9 shows the images after pro-processing. Figure 10 shows the matching thumbnail (with RGB and without RGB). The output results also verify the feasibility.

Output:
420(the X coordinate value of starting point)
25 (the Y coordinate value of starting point)

Fig. 7: The image of Obama

Fig. 8: The sub-frame of video streaming

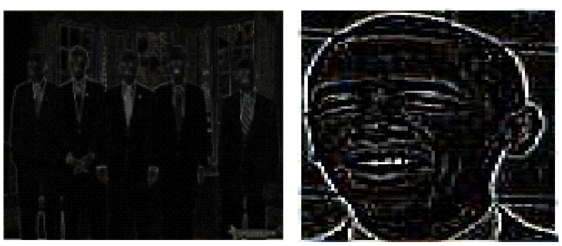

Fig. 9: The image after pro-processing

Fig. 10: The matching thumbnail(with RGB and without RGB)

CONCLUSION

In this study, we proposes an image content fast detection algorithm based on the extraction of the most robust features of the standard samples according to the characteristics of the image' edge neighborhood grayscale variation, and design a multimedia data content detection system mode based on Perceptual Hash. Principle analysis and experiments demonstrate that the system has two improvement of the matching accuracy and fast query against the traditional method, and the scheme can extend to multimedia video stream.

In order to further the matching accuracy, the next step will be to expand to the image with local tampering or content changes to enhance the method robustness.

ACKNOWLEDGMENT

This study is supported by the science and technology project of Hunan Province (No. 2010ZK 3008 and 2011GK3023) and the science and technology project of Changsha city, Hunan Province, China (No.K110720-11).

REFERENCES

Andoni, A. and P. Indyk, 2008. Near-optimal hashing algorithms for approximate nearest neighbour in high dimensions. Commun. ACM, 51(1): 117-122.

Gionis, A., P. Indyk and R. Motwani, 1999. Similarity search in high dimensions via hashing. In: Proceeding of the 25th International Conference on Very Large Data Bases, San Francisco, Morgan Kaufmann Publishers, USA, pp: 518-529.

Kailasanathan, C. and R.C. Naini, 2001. Image Authentication Surviving Acceptable Modifications Using Statistical Measures and K-Mean Segmentation. IEEE EURASIP Workshop Nonlinear Sig and Image Processing.

Kozat, S.S., R. Venkatesan and M.K. Mihcak, 2004. Robust Perceptual Image Hashing via Matrix Invariants. Proceeding of IEEE International Conference on Image Processing, Singapore, 5: 3443-3446.

Lin, C.Y. and S.E. Chang, 2001. A Robust Image Authentication Method Distinguishing JPEG Compression from Malicious Manipulation. IEEE T. Circuits Syst. Video Tech., 2(11): 153-168.

Lu, C.S. and M. Liao, 2003. Structural digital signature for image authentication. IEEE T. Multimedia, 5(2): 161-173.

Luo, S., S. Sha, D. Shen, *et al*., 2004. On-line visual counting system for production of steel bar. Mini-Micro Syst., 25(4): 671-675.

Mihcak, M.K. and R. Venkatesan, 2001. New Iterative Geometric Methods for RobuSt Perceptual Image Hashing. Proc of ACM Workshop Security and Privacy in Digital Rights Management, Philadelphia, PA.

Monga, V. and B.L. Evans, 2006. Perceptual image hashing via feature points: Performance evaluation and trade-offs. IEEE T. Image Proc., 15(11): 3453-3466.

Niu, X. and Y.H. Jiao, 2008. An overview of perceptual hashing. Acta. Electronica. Sinica, 36(7): 1405-1411.

Schneider, M.C., 1996. A Robust Content Based Digital Signature for Image Authentication. Proceeding of IEEE International Conference on Image Processing, Lausanne, Switzerland, 3(9): 227-230.

Swaminathan, A., Y. Mao and M. Wu, 2006. Robust and secure image hashing. IEEE T. Inf. Foren. Sec., 1(2): 215-230.

Venkatesan, R.K.S.M., M.H. Jakubowski, and P. Moulin., 2000. Robust Image Hashing. Proceeding of IEEE Conference on Image Processing, pp: 664-666.