RESEARCH ARTICLE OPEN ACCESS

Mitigation of Rotational Constraints in Image Based Plagiarism Detection Using Perceptual Hash

Sushila Aghav[1], Ankit Kumar[2], Ganesh Gadakar[3],
Aditya Mehta[4], Amey Mhaisane[5]
Department of Computer Engineering,
MIT College of Engineering,
Pune-38, India

ABSTRACT
Image plagiarism can be detected using similarity between the two images. The perceptual hash function calculates similar hash value for similar images. This paper proposes a new modification to perceptual hashing (pHash) based approach for Image Retrieval. By the simplified use of a distance function two perceptual hash values can be compared and, it can be understood whether the images being compared are perceptually different or not. Existing Image Plagiarism detection techniques are not sensitive to image rotation. We propose an algorithm that compares the various degrees of rotation of the source image to the perceived similar image in an efficient way. Perceptual image hash functions along with the rotation check can be used for the identification of plagiarised images, authentication or integrity verification of images.
Keywords:- Perceptual Hash algorithm (pHash), Distance function, Image Plagiarism, Image Retrieval, Distance function..

I. INTRODUCTION

Image Based plagiarism is copying an image or portions of an image from the Internet or from classroom resources without permission or proper acknowledgment and passing it off as your own original work.
A perceptual hash algorithm is a fingerprint of a multimedia file derived from various features from its contents. Perceptual hashes of similar files are similar to each other and this can be used to detect the degrees of similarity between images.
Other types of hashing algorithms that can be used for image similarity search include Average Hashing and Difference Hashing.

1. Average Hash- This approach crushes the image into a grayscale 8x8 image and sets the 64 bits in the hash based on whether the pixel's value is greater than the average colour for the image. It is also known as Mean Hash.

2. Difference Hash- This approach uses the difference between each pixel of the reduced image. It tracks the gradients in each image for the purpose of image matching.

3. Perceptual Hash- This approach describes a class of comparable hash functions. A distinct fingerprint is generated from the features of the given image, for comparison.

Out of these results the best result has been shown by Perceptual Hashing.

II. EXISTING METHODOLOGY

The existing methodology to use perceptual hashing is as follows:

To convert an image into its corresponding hash value we follow these steps,

1. **Reduce size.** Shrink the image to 8x8 so that there are 64 total pixels. The aspect ratio need not be maintained. By not keeping the aspect ratio the hash value will match any variation of the image irrespective of the aspect ratio or the scale.

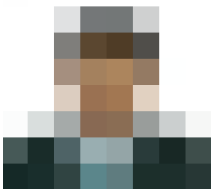
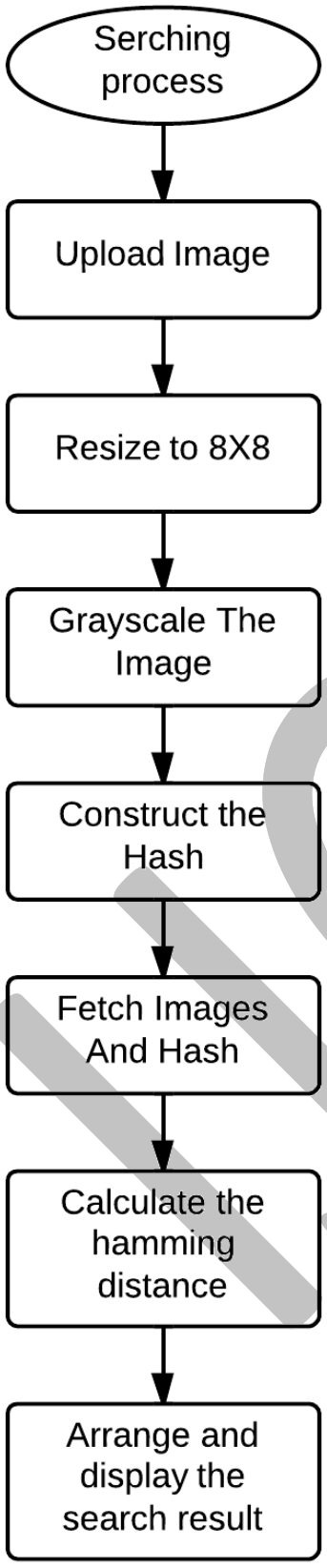


Image 1. Reduced size



Flowchart 1. Existing methodology

2. **Reduce color.** The 8x8 picture is converted to a grayscale. This changes the hash from 64 pixels (64 red, 64 green, and 64 blue) to 64 total colors.

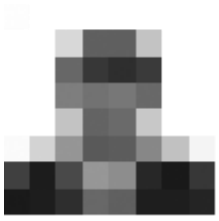


Image 2. Reduced color

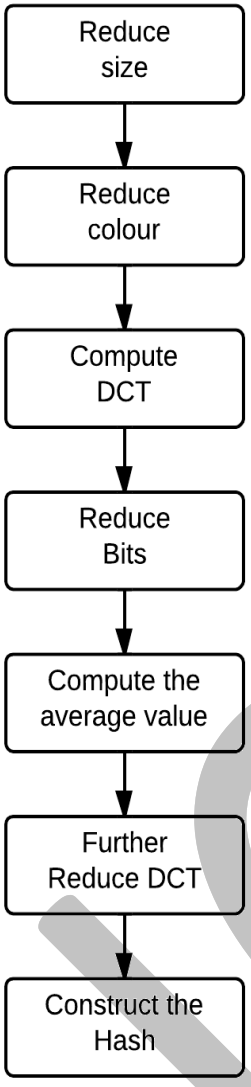
3. **Average the colors**. The average of the 64 values is computed.

4. **Compute the bits.** Set each bit depending on if the value is above or below the average value.

5. **Construct the hash.** The 64 bit integer value can be generated into a hash value with any consistent method for conversion.
   Generated hash value from the image = 8f373714acfcf4d0

To check for plagiarism, we take the suspected plagiarized image and perform the above conversion on the suspected image. The steps to be followed after that are,

6. **Calculate hamming distance.** Calculate the hamming distance between the newly generated hash value and those that have been saved in the database.



Flowchart 2. Hash construction

7. **Find least value.** Find the image and check if the hamming distance between the generated hash and the hash from the database is the least value from the set.

8. Check if this distance is below the threshold value. If it is then the image is plagiarized, otherwise it is not plagiarized.

III. PROPOSED METHODOLOGY

The current implementation of the algorithm has a weakness to rotational changes that may be performed on the image. So, in addition to the existing methodology we propose certain modifications. The addition to the existing algorithm is as follows,

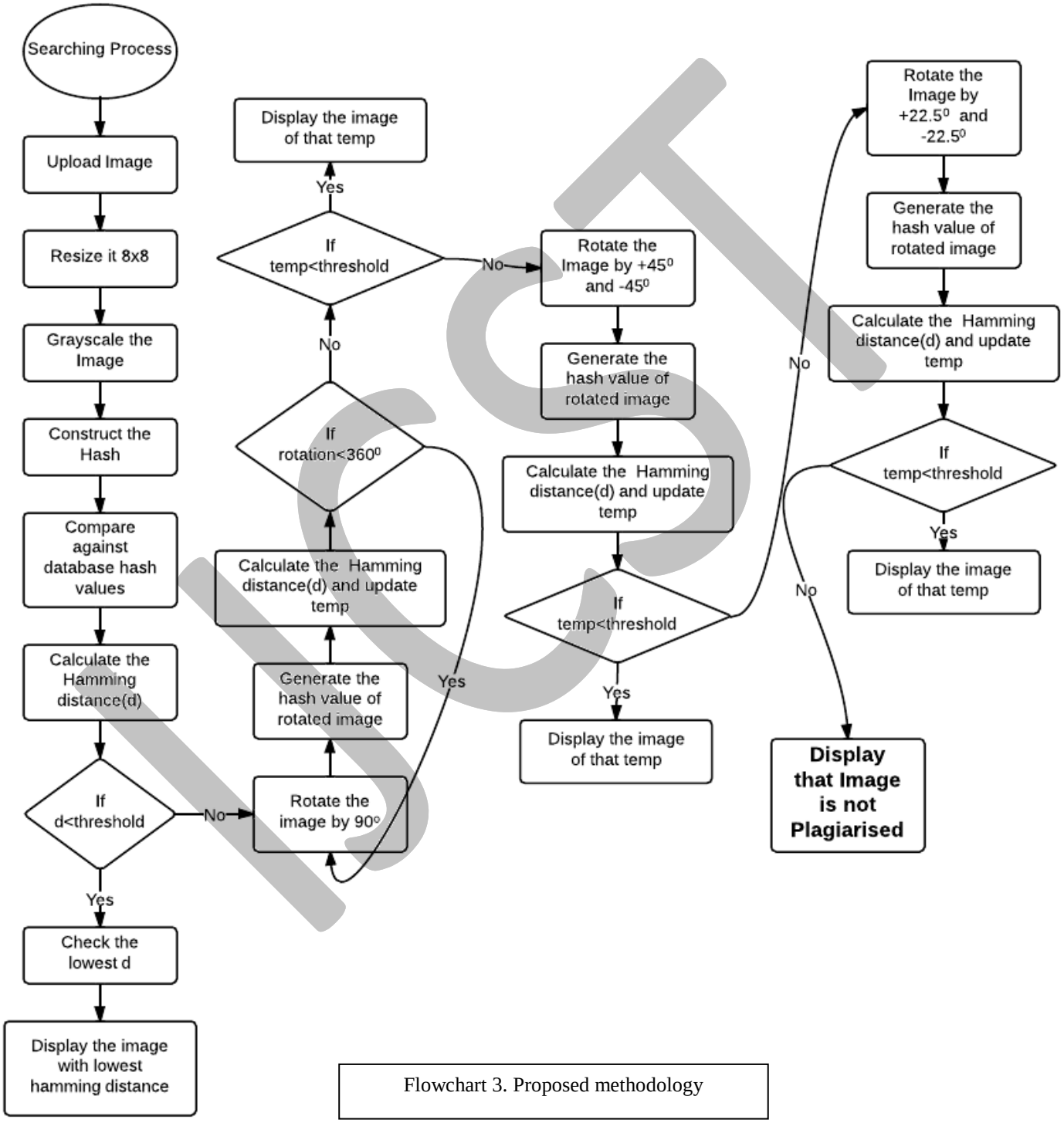
1. Rotate the image by a factor of 90º, 180º and 270º as the first step to refinement and checking for rotational modifications and repeat the process of calculating the hash values and then the hamming distances for each of these rotated images.

2. Once, this has been done find the value for the lowest hamming distance between the suspected image and all of the rotated images and perform a rotation of +45º and -45º on the image which has the least hamming distance and repeat the process stated in the existing methodology.

3. Repeat step 2 but this time by rotating the image +22.5º and -22.5º and perform the entire procedure as stated in step 2.

4. Display the image corresponding to the hash value whose hamming distance is the least.

5. Else display to the user that the image has not been plagiarized.



Flowchart 3. Proposed methodology

IV. CONCLUSIONS

The existing methodology maybe sufficient for detecting plagiarism of images when the source and suspected image have not been rotated by a large margin, but in case of rotational changes the existing methodology will fail. The proposed methodology will ensure that even if the image is rotated plagiarism is detected if it has occurred or if an attack of rotational change has been made. The algorithm makes sure that the matching time of the images is less by reducing the search field by a significant factor each time the refinement is done.

ACKNOWLEDGMENT

We would also like to thank MIT College of Engineering, Pune for the opportunity to present this paper.

REFERENCES

1. Harsh Kumar Sarohi, Farhat Ullah Khan. "Image Retrieval using Perceptual Hashing".IOSR Journal of Computer Engineering (IOSR-JCE) e-ISSN: 2278-0661, p- ISSN: 2278-8727Volume 9, Issue 1 (Jan. - Feb. 2013), PP 38-40.

2. NIU Xia-mu ,JIAO Yu-hua, "An Overview of Perceptual Hashing," ACTA ELECTRONICA SINICA.China, Beijing, Vol.36,No. 7, pp. 1405-1411,July, 2008.

3. M.B. Rao, B.P Rao, A. Govardhan, "CTDCIRS: Content based Image Retrieval System based on Dominant Color and Texture Features", International Journal of Computer Applications, Vol. 18, No.6, pp-0975-8887, 2011.

4. Neal Krawetz Ph. D Computer Science, Texas A&M University,"Perceptual and Average Hash" posted on hackerfactor.com under forensics and fotoforensics.

5. *Manimala Singha, K. Hemchandran, "Performance analysis of Color Spaces Image Retrieval",* Assam University Journal of Science & Technology: Physical Sciences and Technology, Vol. 4, 2011.