Available online www.ejaet.com

European Journal of Advances in Engineering and Technology, 2015, 2(5): 27-31

Review Article ISSN: 2394 - 658X

Comparison of Techniques for Plagiarism Detection in Document Images: A Review

Neha Bansal[1], Manish Mahajan[2] and Shashi Bhushan[1]

[1]Department of Information Technology, CEC Landran, Mohali, Punjab, India
[2] Department of Computer Science Engineering, CGC Landran, Mohali, Punjab, India
bansal.neha03@gmail.com

ABSTRACT

Plagiarism detection created providing similar options are detected inside identical distance of options associated to connect blocks. Copy-move Plagiarism may be a specific style of image meddling wherever a vicinity of the image is traced and glued on another part usually to hide unwanted parts of the image. Hence, the goal of conversion of copy-move plagiarism is to detect image that are same or extraordinarily similar detection of malicious manipulation with digital pictures (digital forgeries) is the topic of this paper. Particularly, we have a tendency to specialise in detection of a special style of digital plagiarism – the copy-move attack within which a locality of the image is traced and glued elsewhere within the image with the intent to hide a vital image feature. In this given paper, we have a tendency to review many strategies projected to realize this goal.

Key words: Plagiarism detection, forgeries, copy-move plagiarism

---

INTRODUCTION

The latest imaging technologies have given forgers need tools for ever-changing and victimization the contents of digital pictures to the aim of adding deceptive object to the photographs with no noticeable options [1]. From now, it's recommended by several researchers to ascertain pictures credibleness to notice these activities which might be found in several applications like criminal investigation, medical imaging, journalism, intelligence services and police investigation systems. As a result, there's a speedy increase of the digitally manipulated forgeries in thought media and on the net. This trend indicates serious vulnerabilities and reduces the believability of the digital pictures. Therefore, developing techniques to verify the integrity and also the credibleness of the digital pictures became vital, particularly considering the photographs bestowed as proof during a court of law, as news things, as a section of a medical history, or as a money document. During this sense, image tamper detection is one amongst the first goals in image forensics. Recently, several authors studied the matter of detective work image forgeries; forward that notwithstanding the tampered pictures. Therefore, digital plagiarism detection techniques are developed to justify the Plagiarism issue as a necessary method in image process [2]. Many analysis studies were conducted in several worrying fields to reinforce the present techniques for copy-moving Plagiarism [3], which includes activity or adding a section within the image or displaying propaganda [4]. The common plagiarism technique in digital pictures is divided into 3 main groups: Copy-Paste (i.e., Splicing), Image Retouching, and Copy-Move (i.e., cloning). For example, retouching technique that works on manipulating the digital image by ever-changing its options while not creating noticeable modifications of the content of the image.

Meanwhile, image junction on the opposite hand, build use of the initial image with further pictures to come up with a tampered copy [5-6], such methodology work on adding some a part of alternative pictures to the initial image in order that forgers hide or modify the content of the image. Additionally, image biological research, that works by repetition an explicit {part of |a {part of |a a part of}} a picture and shifting it to a different part of constant image in order that forgers will hide or duplicate some part of the image [7]. Hence, current effort in developing reliable strategies for image plagiarism detection has gained attention of the many researchers.

Detection methodology found within the literatures is classified into active methodology and passive methodology [8-9]. a vigorous detection methodology like watermarking, that consists of adding image details so as to explain digital change of state like name, date, signature, etc. whereas the passive methodology consists of detective work forgeries or duplicated objects in pictures while not considering the data of the initial pictures [10]. The most goal

of this methodology is too specific however detective work forgeries square measure potential with none wants of original image watermark. In this paper, the main focus is on sleuthing copy-move (i.e., cloning) image plagiarism at the side of describing the problems related to the plagiarism detection.

PLAGIARISM DETECTION SYSTEM

Usually it's potential to spot the duplicated object by computing and comparison these premises with the entire image. However new plagiarism detection techniques square measure still lacking of up so far malicious activities. Such assumption came from the power of forgers to alter the pure mathematics of the duplicated object simply by modifying the image's options. Therefore, a replacement copy-move plagiarism detection technique is required so as to balance the new malicious activities on digital pictures [11]. Furthermore, the event of analysis in digital forensics has finally determined the appropriate solutions for determination a lot of comprehensive problems associated with copy-move plagiarism. Consequently it's rising that generalized solutions and techniques, building standardized knowledge sets, benchmarks, analysis criteria etc. square measure still required to be projected to appreciate the new frameworks minimizing the possibilities for digital forgeries. Thus, several sensible and precise techniques, solutions are projected that analysis can introduce within the next section. The elemental issues that analysis found within the literature is categorised into the natural, plagiarism detection, flow mapping, and supply identification.

DATA PLAGIARISM DETECTION

Plagiarism detection strategies become far more sophisticated to touch upon the newest Plagiarism techniques. This back to the provision of digital writing tools, alteration, and manipulation become terribly straightforward and as a result Plagiarism detection becomes a posh and threatening downside [12]. Image Plagiarism detection is manipulated in varied ways in which with several straightforward operations like affine transforms like translation, scaling, etc., compensation operations like brightness, colours, distinction changes, etc., suppression operation like noise extraction, filtering, compression, etc [9]. What is more, a lot of complicated operations also are potential like compositing, blending, matting, cropping, and montage resulting in visually untraceable artifacts in a picture. The automated and methodology of sleuthing the cast pictures has become an enormous difficult downside for researchers and therefore the same downside is true for each multimedia system contents.

PLAGIARISM DETECTION TECHNIQUES

Mahdian and Saic [13] used blur moment invariants to represent image regions as a result of they can't be tormented by blur degradation and additive noise. Their methodology begins with tilting of pictures by blocks of a selected size. They described every block with blur invariants. The feature vector for every block is of length seventy two. This square measure normalized additional to boost the duplication detection skills of the algorithmic program. They applied principal part transformation to scale back the dimension of feature vector. For blocks similarity analysis, they used k-d tree illustration. Employing a bound threshold price, they found similar blocks. Once the similar blocks square measure found, they have to be verified. They verified this by finding the neighbourhood of comparable blocks that also are identical. Two similar blocks with non-identical neighbourhood square measure thought of as false positive. By victimisation this methodology, they have detected copy-move plagiarism for pictures that have blurred duplicated region. They might additionally discover duplicated regions with modified distinction values. However, there square measure some false alarms that square measure common in several of the projected strategies. Also, the computation time of the algorithmic program is relatively high.

Wang et al [14] conducted a study on copy-move plagiarism detection by victimisation Hu moments. They developed the algorithmic program to be a lot of economical and additionally strong to numerous post-processing techniques like blurring, lossy JPEG compression. They reduced the size of the image by victimisation Gaussian pyramid. They divided the image into many mounted sized blocks that square measure overlapping. They applied Hu moments to the blocks and calculated the Eigen values. They sorted these vectors lexicographically and a neighbourhood threshold is chosen to scale back false detections. They performed finding matching blocks by victimisation mathematical morphological techniques. Their methodology is self-made in sleuthing copy-move plagiarism even once post-processing is completed. Mohamadian and Pouyan [15] delineate new methodology of sleuthing copy-move plagiarism by victimisation SIFT algorithmic program at the side of Zernike moments. They used SIFT algorithmic program to perform traditional copy-move Plagiarism detections. However SIFT can't be accustomed discover flat derived regions. To account for this, they used Zernike moments. The method begins with SIFT feature points extraction. When extraction, they used these feature to seek out potential matches. To avoid false alarms of Plagiarism, they used stratified clump. This involves clump of feature points into a tree structure supported bound threshold price. By this methodology, they were able to scale back false alarms as a result of they thought of that image is solid only if 2 clusters square measure matched with a minimum of 3 similar feature points. However, this feature reduces the chance of sleuthing flat forgeries. Their methodology was able to determine the potential geometric transformations performed.

Popescu and Farid [16] were able efficiency discover copy-move plagiarism to with by applying PCA (Principal part Analysis). Their methodology is comparable to DCT approach and higher in capturing discriminating options. The given image is regenerated from coloured to gray scale. They divided the image into many tiny sized blocks, that square measure described into vectors. Then they organized it lexicographically before matching. This can be far better than the brute-force methodology of finding matches. They used PCA methodology to represent the various blocks in an alternate method. PCA is capable of sleuthing even minor variations attributable to noise or lossy compression. Their methodology is just for gray scale pictures. However, the strategy is created to figure for coloured pictures also by process the image for every colour channel, which yields 3 duplication maps. Then PCA is applied to every map severally to discover the forgeries. Their methodology contains a smart potency in sleuthing copy-move plagiarism and additionally offers less variety of false positives. However, the potency falls because the block size decreases and additionally if the standard of the image is low.

Ting and Rang-ding [17] projected a copy-move plagiarism detection methodology victimisation Singular price Decomposition (SVD). Their developed algorithmic program is computationally less complicated and is powerful to post-processing techniques. They used the correlation between the derived and glued regions and probe for identical regions. Within the opening move, they divided the image into many tiny overlapping blocks. Then, they applied SVD to each block and extracted distinctive singular values feature vector for every block. Victimisation these vectors, they found the matching blocks by remodelling every block options into k-d tree. They used a threshold price to extend the lustiness and eliminate pseudo-matching. A natural image won't have identical regions with coherent orientation. So, the obtained matched blocks square measure associate degree proof for copy-move plagiarism. They used lines to attach 2 identical blocks in an exceedingly figure that clearly shows the tampered regions.

Zimba and Xingming [18] projected a replacement methodology of copy-move plagiarism detection. Their methodology begins with conversion of colour image into gray scale image. Then, they applied DWT to entire image. This offers sub-bands, out of that low frequency sub-band is enough to perform detection method. They divided the image into many overlapping blocks. They performed principal part analysis - Eigen price decomposition on the blocks. They placed these feature vectors square measure placed into the matrix and sorted the entries lexicographically. This methodology of sorting makes the matching less complicated. They calculated the normalized shift vector so offset frequency. This offset frequency is subjected to morphological process to administer final results. They created this methodology a lot of economical than typical PCA methodology by reducing the image size within the starting of the method. Their algorithmic program will discover duplications involving rotation of varied degrees. They enclosed morphological operations to avoid false detections. The sole disadvantage is that the duplicated region ought to be larger than the block size, otherwise it can't be detected.

Bravo-Solorio and Nandi [19] conducted a study on copy-move detection technique to seek out forgeries involving reflection, rotation and scaling. They covered the image as block of components by slippery component by pixel with a window of specific size in an exceedingly raster-scan order. They calculated feature vectors that square measure colour-dependent. By this, they reduced the quantity of searches thereby increasing the potency. They calculated four options out of those 3 options square measure severally computed as red, inexperienced and blue elements. The fourth feature is calculated because the entropy of brightness level channel. They used this fourth feature to discard blocks with scant textural info. These options square measure listed lexicographically so matching is performed. Their methodology produces heap of matches; thence they used refinement to scale back them.

A study by Sridevi et al [20], proposes a copy-move Plagiarism detection technique in an exceedingly parallel setting. They projected this methodology principally to accomplish copy-move Plagiarism detection in period of time. Alternative strategies like PCA, DWT or SVD have high computation time; thence can't be utilized in period of time applications. Their methodology begins with dividing the gray scale image into many overlapping blocks of a specified size. Then intensity options for each block square measure extracted. The last 2 locations of the feature vectors store the block position. All this method of extracting the intensity options is taken care by one algorithmic program. They developed another algorithmic program for parallel sorting. This performs the lexicographic typing victimisation base sort methodology in an exceedingly parallel method. This type of sorting ensures straightforward detection of comparable blocks by finding the identical options. They found the duplicated regions by matching of options and these blocks square measure mapped on to the image victimisation the placement hold on within the vector. There'll be a main algorithmic program that controls of these steps. Their methodology has shown performance improvement over several alternative typical techniques. This can be accomplished by reducing the interval. They controlled the false detection rate by adjusting the block size. However, their methodology can't be applied for a colour image.

A tabulated form of different method used for plagiarism detection and their drawbacks by several researchers is given in Table -1.

Table -1 Comparison of Different Method used for Plagiarism Detection and their Drawbacks

Author	Year	Method Used	Drawbacks
Mahdian and Saic [13]	2007	Used blur moment invariants to represent image regions as a result of they can't be tormented by blur degradation and additive noise.	The computation time of the algorithmic program is relatively high.
Wang, Liu, Zhang, Dai and Wang [14]	2009	Conducted a study on copy-move plagiarism detection by victimisation Hu moments.	Their methodology is self-made in sleuthing copy-move plagiarism even once post-processing is completed.
Mohamadian and Pouyan [15]	2013	Delineate new methodology of sleuthing copy-move plagiarism by victimisation SIFT algorithmic program at the side of Zernike moments.	Their methodology was able to determine the potential geometric transformations performed.
Popescu and Farid [16]	2004	Discover copy-move plagiarism to with by applying PCA (Principal part Analysis).	The potency falls because the block size decreases and additionally if the standard of the image is low.
Ting and Rang-ding [17]	2009	Projected a copy-move plagiarism detection methodology victimisation	They used lines to attach 2 identical blocks in an exceedingly figure that clearly shows the tampered
Zimba and Xingming [18]	2011	Projected a replacement methodology of copy-move plagiarism detection	The sole disadvantage is that the duplicated region ought to be larger than the block size, otherwise it can't be detected.
Bravo-Solorio and Nandi [19]	2011	A study on copy-move detection technique to seek out forgeries involving reflection, rotation and scaling.	Their methodology produces heap of matches; thence they used refinement to scale back them.
Sridevi, bone and Sandeep [20]	2012	Proposes a copy-move Plagiarism detection technique in an exceedingly parallel setting.	They controlled the false detection rate by adjusting the block size. However, their methodology can't be applied for a colour image.

CONCLUSION

A brief survey on the plagiarism detection strategies was conferred which will facilitate researchers explore new ideas and supply new solutions to the challenges within the field, particularly with blind strategies. An effort has been created to introduce varied promising techniques that represent cheap enhancements within the plagiarism detection strategies. There square measure techniques exhibiting improved detection accuracy, however having high machine complexness. Moreover, most of the strategies might not be that awake to the geometric transformations, like rotation and scaling. The issue of human perception is additionally not counted as an element throughout the event of those techniques. Thus there's a desire to develop techniques that square measure automatic, HVS actuated and effective against geometric transformations. In essence, this work, surveyed detection techniques for three of the foremost common plagiarism sorts, specifically copy/move, splice and retouching. Most of those are unfit by one or a lot of factors that embody restricted accuracy rate, low reliableness and high complexness additionally to their sensitivity to varied transformations and non-responsiveness to noise.

REFERENCES

[1] XZ Pan and HM Wang, The Detection Method of Image Regional Plagiarism Based DWT and DIMPCA, *Advanced Materials Research*, **2012**, 532, 692-696.
[2] B Shivakumar and SS Baboo, Automated Forensic Method for Copy-Move Plagiarism Detection based on Harris Interest Points and SIFT Descriptors, *International Journal of Computer Applications*, **2011**, 27(3), 9-17.
[3] H Yao, T Qiao, Z Tang, Y Zhao and H Mao, Detecting Copy-Move Plagiarism using Non-Negative Matrix Factorization, *Third International Conference on Multimedia Information Networking and Security (MINES)*, Shanghai, China, **2011**, 591-594.
[4] VS Pujari, and M Sohani, A Comparative Analysis on Copy Move Plagiarism Detection using Frequency Domain Techniques, *International Journal of Global Technology Initiatives*, **2012b**, 1(1), E104-E111.
[5] L Chen, W Lu, J Ni, W Sun and J Huang, Region Duplication Detection based on Harris Corner Points and Step Sector Statistics, *Journal of Visual Communication and Image Representation*, **2013**, 24(3), 244-254.
[6] MH Liu and WH Xu, Detection of Copy-Move Plagiarism Image based on Fractal and Statistics, *Journal of Computer Applications*, **2011**, 8, 61.
[7] P Yadav, Y Rathore and A Yadu, DWT Based Copy-Move Image Plagiarism Detection, *International Journal of Advanced Research in Computer Science and Electronics Engineering*, **2012**, 1(5), 56-58.
[8] VS Pujari and M Sohani, A Comparative Analysis on Copy Move Plagiarism Detection in Spatial Domain Method Using Lexicographic and Non Lexicographic Techniques, *International Journal of Electronics Communication and Computer Engineering*, **2012a**, 3(1), 136-139.
[9] B Shivakumar and Santhosh Baboo, Detecting Copy-Move Plagiarism in Digital Images: A Survey and Analysis of Current Methods, *Global Journal of Computer Science and Technology*, **2010**, 10(7), 61-65.
[10] L Chen, W Lu and J Ni, An Image Region Description Method Based on Step Sector Statistics and its Application in Image Copy-Rotate/Flip-Move Plagiarism Detection, *International Journal of Digital Crime and Forensics*, **2012**, 4(1), 49-62.

[11] A Piva, *An Overview on Image Forensics*, ISRN Signal Processing, **2013**, 1-22.
[12] HJ Lin, CW Wang and YT Kao, Fast Copy-Move Plagiarism Detection, *WSEAS Transactions on Signal Processing*, **2009**, 5(5), 188-197.
[13] B Mahdian and S Saic, Detection of Copy–Move Plagiarism using a Method based on Blur Moment Invariants, *Forensic Science International*, **2007**, 171(2), 180-189.
[14] JW Wang, GJ Liu, Z Zhang, Y Dai and Z Wang, Fast and robust forensics for image region-duplication Plagiarism, *Acta Automatica Sinica*, **2009**, 35(12), 1488-1495.
[15] Z Mohamadian and AA Pouyan, Detection of Duplication Plagiarism in Digital Images in Uniform and Non-uniform Regions, *UK Sim*, **2013.**
[16] AC Popescu and H Farid, Exposing Digital Forgeries by Detecting Duplicated Image Regions, Dept Computer Science, Dartmouth College, *Hanover*, **2004,** 515.
[17] Z Ting and W Rang-ding, Copy-move Plagiarism Detection based on SVD in Digital Image Paper, *2nd International Congress on Image and Signal Processing*, **2009**, 1-5.
[18] M Zimba, and S Xingming, DWT-PCA (EVD) Based Copy-move Image Plagiarism Detection, *International Journal of Digital Content Technology and its Applications*, **2011,** 5(1), 251-258.
[19] S Bravo Solorio and AK Nandi, Automated Detection and Localisation of Duplicated Regions Affected by Reflection, Rotation and Scaling, *Image Forensics Signal Processing*, **2011,** 91(8), 1759-1770.
[20] M Sridevi, C Mala and S Sandeep, Copy–Move Image Plagiarism Detection, *Journal of Computer Science and Information Technology*, **2012**, 52, 19-29.