Available Online at www.ijcsmc.com

International Journal of Computer Science and Mobile Computing

A Monthly Journal of Computer Science and Information Technology

ISSN 2320–088X

IJCSMC, Vol. 3, Issue. 4, April 2014, pg.1162 – 1170

RESEARCH ARTICLE

Forbidden Twin Substantiate Consuming Hash Generation and SVM

Mrs. Deepa R
PG Scholar, SRM University, Chennai
Saideepa84@gmail.com

Mrs. Noor Alleema N
Department of Information Technology, SRM University, Chennai
noor25nrs@gmail.com

Abstract:- Forbidden images are highly available in every field. To carryover many unauthorized job. It's highly necessary to control over the forbidden image including removal, insertion, and replacement of objects, and abnormal color modification, and for fix the forbidden area. Thus the forbidden images can be identified through process center image processing. This includes pre-processing, feature extraction and classification. Noise removal of segmentation is obtained in preprocessing global and local features is extracted through PCA and SIFT. Secret keys are introduced for hash structure and feature extraction. While being robust against content-preserving image processing, the hash is superficial to wicked interfering and, therefore, applicable to image endorsement. The hash of an experiment image is matched with that of an allusion image. When the hash distance is greater than a threshold τ_1 and less than τ_2, the acknowledged image is umpired as a false. By decomposing the hashes, the type of image forbidden and location of forbidden areas can be determined.

Index Terms - Forbidden image detection, SIFT, PCA, image hash, SVM.S

I. Introduction

Today's technology permits digital media to be reformed and manipulated by anyways which was terrible for earlier days. Techniques are required to robust against inventive image processing and transmission errors. In case of insecure environment any invariant person can alter the image during the transmission. To provide security, image authentication techniques is introduced to prevent intrusion and forbidden. The security of image hash functions is introduced by integrating a secret key in producing the hash without the information of the key the hash values should not be easily forbidden or predictable. Image content is a compact exemplification of extra file and later used for endorsement. The endorsement methods have wide applicability in law, commerce, journalism and national defense. In existing system has robust hashing system is developed for identifying image forbidden including removal, insertion, and replacement of objects, and abnormal color modification, and for locating the forged area. Both global and local features are used in forming the hash sequence. The global features are based on Zernike moments rep-resenting luminance and chrominance characteristics of the image as a whole. The local features include position and texture information of salient regions in the image.

II. Related Works On Digital Image Forbidden

This section introduces the techniques and methods currently available in the area of digital image forbidden detection. Currently, most acquisition and manipulation tools use the JPEG standard for image compression. As a result one of the standard methodologies is to use the blocking fingerprints introduced by JPEG compression, as reliable indicators of possible image forgery. Not only do this inconsistencies helps to determine possible tampering, but they can also be used to light into what method of forbidden was used. Many passives schemes have been developed based on these fingerprints to Resampling [4], Copy-Paste[5,6], Luminance-Level [7], Double Compression JPEG[8,16], ANN[9], and Wavelet Transformation Coefficient [10]. Above mentioned methodologies are derived from one another and they all contain constraints in implementations and limitations in performance.

The first step extracts a feature vector from the image, whereas the second stage compresses this feature vector to a final hash value. In the feature extraction step, the two-dimensional image is mapped to a one-dimensional feature vector. This feature vector must capture the perceptual qualities of the image. That is, two images that appear identical to the human visual system should have feature vectors that are close in some distance metric. Likewise, two images that are clearly distinct in appearance must have feature vectors that differ by a large distance. For such feature vector extraction, many algorithms could be used, e.g. [16], [17], [18], [19],[20]. For the rest of the paper, we will refer to this visually robust feature vector as the "intermediate hash". The second step then compresses this intermediate hash vector to a final hash value. In this paper, we assume the availability of such an intermediate hash vector that has been extracted from the image and a model on its distribution. The methods previously proposed for this step include using error correction decoding for compression of binary intermediate hash vectors [16] and dither-based compression via distributed source coding [21]. While compression is their primary goal [16], [21], no *explicit* attempt was made to ensure that perceptually identical images are compressed to the same hash value. The second step will involve clustering of the intermediate hash vector of an input source (image) and the intermediate hash vectors of its perceptually identical versions. In this paper, we present a solution to the second step by developing such a clustering algorithm based on the distribution of intermediate hash vectors [22]. Another important issue is the length (or granularity) of the final hash required to cluster images within a specified distance. Underestimating this length can adversely aspect the perceptual qualities of the hash. A significant contribution of our work is that this length is determined as a natural outcome of our proposed clustering algorithm.

Detection of digital forbidden having enormous number applications related Forensic science document questioning section although which is very helpful for media, publication, law, military, medical image science application, satellite image, research and World Wide Web publications.

III. Proposed Scheme

Hashes produced with the proposed method are robust against common image processing operations including bright-ness adjustment, scaling, small angle rotation, and JPEG coding and noise contamination. Collision probability between hashes of different images is very low. The proposed scheme has a reasonably short hash length and good ROC performance.

A robust hashing method is developed for detecting image forgery including removal, insertion, and replacement of objects, and abnormal color modification, and for locating the forged area. Both global and local features are used in forming the hash sequence. The global features are based on Zernike moments rep-resenting luminance and chrominance characteristics of the image as a whole. The local features include position and texture information of salient regions in the image.

Merits of proposed system different images have significantly different hash values, and secure so that any unauthorized party cannot break the key and coin the hash. Demerits of proposed system with the widespread use of image editing software, ensuring credibility of the image contents have become an important issue.

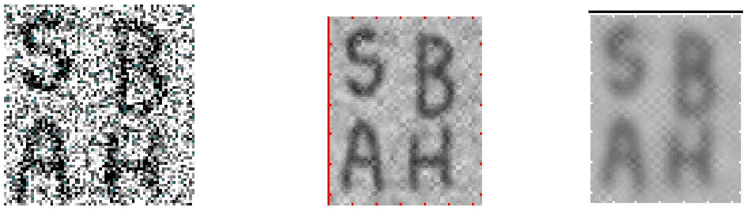
IV. Brief Explanation of Useful Tools and Techniques

1. Pre-Processing

Preprocessing methods use a small neighborhood of a pixel in an input image to get a new brightness value in the output image. Such pre-processing operations are also called filtration.

1.1 Noise Removal from Image (Bilateral Filter)
1.2 Segmentation and Level Algorithms

1.1 Noise Removal from Image (Bilateral Filter)



A "noisy" image   Smoothed   Continued smoothing

Fig1. Noise Removal of an Image

Imagine an image with noise. For example, the image on the left below is a corrupted binary (black and white) image of some letters 60% of the pixels are thrown away and replaced by random gray values.

- Noise Means Any Unwanted Signal.
- One Person's Signal Is another One's Noise.
- Noise Is Not Always Random And Randomness Is An Artificial Term.
- Noise Is Not Always Bad.

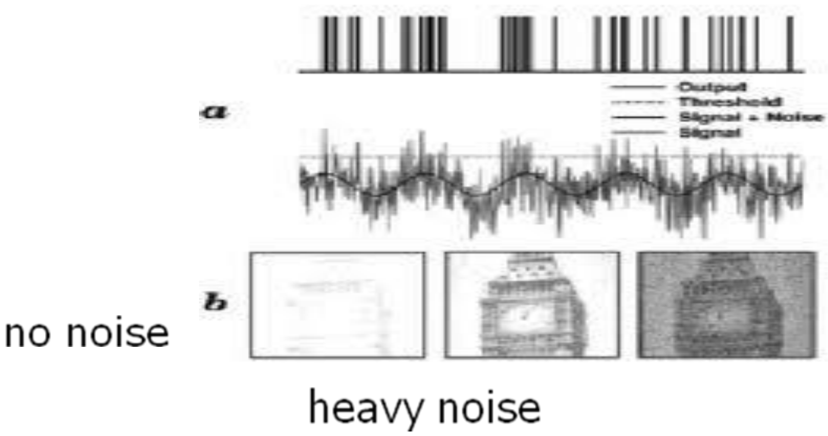


Fig2. Differentiate of Noise

A **bilateral filter** is non-linear, edge-preserving and **noise-reducing smoothing filter for images.** The intensity value at each pixel in an image is replaced by a weighted average of intensity values from nearby pixels. This weight can be based on a Gaussian distribution. Crucially, the weights depend not only on Euclidean distance of pixels, but also on the radiometric differences. (For example,): the range difference such as color intensity, depth distance, etc. This preserves sharp edges by systematically looping through each pixel and adjusting weights to the adjacent pixels accordingly.

The bilateral filter is defined as

$$I^{\text{filtered}}(x) = \frac{1}{W_p} \sum_{x_i \in \Omega} I(x_i) f_r(\|I(x_i) - I(x)\|) g_s(\|x_i - x\|),$$

Where the normalization term

$$W_p = \sum_{x_i \in \Omega} f_r(\|I(x_i) - I(x)\|) g_s(\|x_i - x\|)$$

Ensures that the filter preserves image energy and

- I^{filtered} is the filtered image;
- I is the original input image to be filtered;
- x are the coordinates of the current pixel to be filtered;
- Ω is the window centered in x;
- f_r is the range kernel for smoothing differences in intensities. This function can be a Gaussian function;
- g_s is the spatial kernel for smoothing differences in coordinates. This function can be a Gaussian function;
- The bilateral filter is a powerful alternative to the iteration-based (WLS,RE,AD) filters for noise removal.
- We have shown that this filter emerges as a single Jacobi iteration of a novel penalty term that uses ‘long-distance’ derivative.
- We can further speed the bilateral filter using either the GS or the sub-gradient approaches.
- We have generalized the bilateral filter for treating piece-wise linear signals.

1.1.1 Noise Removal (Bilateral filter)

It extends the concept of Gaussian smoothing by weighting the filter coefficients with their corresponding relative pixel intensities. Pixels that are very different in intensity from the central pixel are weighted less even though they may be in close proximity to the central pixel. This is effectively a convolution with a non-linear Gaussian filter, with weights based on pixel intensities. This is applied as two Gaussian filters at a localized pixel neighborhood, one in the spatial domain, named the *domain filter*, and one in the intensity domain, named the *range filter*. A very intuitive mathematical approach is given in as follows.

Let *f: Re →Re* be the original brightness function of an image which maps the coordinates of a pixel (x, y) to a value in light intensity. Then for any given pixel *a* at (x , y) within a neighborhood of size n, which has a_0 as its centre, its coefficient assigned by the range filter *r(a)* is determined by the following function:

$$r(a_i) = e^{-\frac{[f(a_i)-f(a_0)]^2}{2\sigma_r^2}}$$

Similarly, its coefficient assigned by the domain filter *g(a)* is determined by the closeness function below:

$$g(x, y; t) = e^{-\frac{x^2+y^2}{2t}}$$

Where t is the scale parameter for the central pixel of the neighborhood a_0, its new value, denoted by $h(a_0)$.

$$h(a_0) = k^{-1} \sum_{i=0}^{n-1} f(a_i) \times g(a_i) \times r(a_i)$$

K is the normalization constant to maintain zero-gain and is defined as follows

$$k = \sum_{i=0}^{n-1} g(a_i) \times r(a_i)$$

Pixels close to the central pixel a_0 in both space and intensity contribute more than those further away in space and intensity.

1.2 Segmentation and Level Algorithms

We address the difficulty of image segmentation methods based on the popular level set framework to handle an arbitrary number of regions. While in the literature some level set techniques are available that can at least deal with a fixed amount of regions greater than two, there is very few work on how to optimize the segmentation also with regard to the number of regions. Based on a variational model, we propose a minimization strategy that robustly optimizes the energy in a level set framework, including the number of regions. Our evaluation shows that very good segmentations are found even in difficult situations.

1.2.1 Image Segmentation

- Segmentation divides an image into its constituent regions or objects.
- Segmentation of images is a difficult task in image processing. Still under research.
- Segmentation allows extracting objects in images.
- Segmentation is unsupervised learning.
- Model based object extraction, e.g., template matching, is supervised learning.

1.2.2 Use Segmentation

- After a successful segmenting the image, the contours of objects can be extracted using edge detection and/or border following techniques.
- Shape of objects can be described.
- Based on shape, texture, and color objects can be identified.
- Image segmentation techniques are extensively used in similarity searches.

2. Feature Extraction

2.1 SIFT Algorithm
2.2 PCA Algorithm

2.1 SIFT Algorithm

Scale-invariant feature transform (or SIFT) is an algorithm in computer vision to detect and describe local features in images. The detection and description of local image features can help in object recognition. The SIFT features are local and based on the appearance of the object at particular interest points, and are invariant to image scale and rotation. They are also robust to changes in illumination, noise, and minor changes in viewpoint. In addition to these properties, they are highly distinctive, relatively easy to extract and allow for correct object identification with low probability of mismatch. They are relatively easy to match against a (large) database of local features but however the high dimensionality can be an issue, and generally probabilistic algorithms such as k-d trees with best bin first search are used.

Object description by set of SIFT features is also robust to partial occlusion; as few as 3 SIFT features from an object are enough to compute its location and pose. Recognition can be performed in close-to-real time, at least for small databases and on modern computer hardware.

1. Build Gaussian Scale Space
Definition: a convolution of an image with a variable scale (sigma) Gaussian
Pyramidal construction
Octaves as levels of pyramid
Octave represents doubling of sigma

i. Algorithm samples 3 scales per octaves plus 2 extra (5 total images) each sample is a convolution.
ii. Each octave is half the size of the previous one.
iii. An image is doubled in size for the first octave (gives ore keypoints).
iv. No limit to the number of octaves (except image size)

2. Build Difference of Gaussian Space (DoG)
Subtract adjacent images in Gaussian Space
This approximates Laplacian of Gaussian
3. Find key points (SIFT features) on the DoG
Minima or maxima of 26 neighboring points (9 above, 8 at, 9 below); see figure in slides
4. Localize key points
Precise location of a key point determined by fitting a 3D (x, y, sigma) quadratic curve to the sample points around a key point
Uses Taylor expansion (3x3 linear system)
If the new offset is larger than 0.5 in any dimension, repeat this process with the closer sample point
5. Filtering

Check for contrast

v. Look at the DoG value at the key point
vi. If < 0.03, throw the key points out Check for "well-defined peak" (edge-iness)
vii. Ratio of two derivative expressions (trace and determinant of Hessian matrix)

6. Orientation assignment
Uses the standard arctan() on the points in the Gaussian image where the key point came from (determined from key point's scale) to calculate orientation
Orientation(s) assigned from the histogram of orientations in the region around the keypoint
Multiple orientations improve stability of keypoints
7. Descriptor

viii. Orientation histogram of gradient magnitudes in the region around the keypoint (16x16 sample area)
ix. Coordinates of descriptor and gradient orientations rotated relative to keypoint orientation to achieve rotation invariance of the descriptor
x. Each point weighted by a Gaussian function
xi. One orientation histogram for each 4x4 sample region (gives 4x4 histograms)
xii. Each histogram has 8 orientations (thus 4x4x8=128 total elements in the descriptor)
xiii. Trilinear interpolation used to distribute the value of each gradient sample to adjacent histogram bins (reduces boundary effects)
xiv. Descriptor vector normalized to unit length, capped, and renormalized again to reduce effects of illumination change Distribution of values in the vector is more important than magnitudes.

2.2 PCA Algorithm

Principal component analysis (PCA) is a statistical procedure that uses orthogonal transformation to convert a set of observations of possibly correlated variables into a set of values of linearly uncorrelated variables called principal components. The number of principal components is less than or equal to the number of original variables. This transformation is defined in such a way that the first principal component has the largest possible variance (that is, accounts for as much of the variability in the data as possible), and each succeeding component in turn has the highest variance possible under the constraint that it is orthogonal to (i.e., uncorrelated with) the preceding components. Principal components are guaranteed to be independent if the data set is jointly normally distributed. PCA is sensitive to the relative scaling of the original variables. Depending on the field of application, it is also named the discrete Karhunen–Loève transform (KLT) in signal processing, the Hotelling transform in multivariate quality control, proper orthogonal decomposition (POD) in mechanical engineering, singular value decomposition (SVD) of $\mathbf{X}$ (Golub and Van Loan, 1983), Eigen value decomposition (EVD) of $\mathbf{X}^T\mathbf{X}$ in linear algebra, factor analysis, Eckart–Young theorem (Harman, 1960), or Schmidt–Mirsky theorem in psychometrics, empirical orthogonal functions (EOF) in meteorological science, empirical Eigen function decomposition (empirical component analysis (Lorenz, 1956), quasi harmonic modes (Brooks et al., spectral decomposition in noise and vibration, and empirical modal analysis in structural dynamic

3. Classification

3.1 SVM (Support Vector Machine)
*3.2 Hash Generati*on

3.1 Support Vector Machine

In machine learning, **support vector machines** (**SVMs**, also **support vector networks** are supervised learning models with associated learning algorithms that analyze data and recognize patterns, used for classification and regression analysis. Given a set of training examples, each marked as belonging to one of two categories, an SVM training algorithm builds a model that assigns new examples into one category or the other, making it a non-probabilistic binary linear classifier. An SVM model is a representation of the examples as points in space, mapped so that the examples of the separate categories are divided by a clear gap that is as wide as possible.

New examples are then mapped into that same space and predicted to belong to a category based on which side of the gap they fall on.

- SVMs are helpful in text and hypertext categorization as their application can significantly reduce the need for labeled training instances in both the standard inductive and transductive settings.
- Classification of images can also be performed using SVMs. Experimental results show that SVMs achieve significantly higher search accuracy than traditional query refinement schemes after just three to four rounds of relevance feedback.
- SVMs are also useful in medical science to classify proteins with up to 90% of the compounds classified correctly.
- Hand-written characters can be recognized using SVM.

3.2 Hash Function

A hash function is any algorithm that maps data of arbitrary length to data of a fixed length. The values returned by a hash function are called hash values, hash codes, hash sums, checksums or simply hashes. Recent development of internet payment networks also uses a form of 'hashing' for checksums, and has brought additional attention to the term.

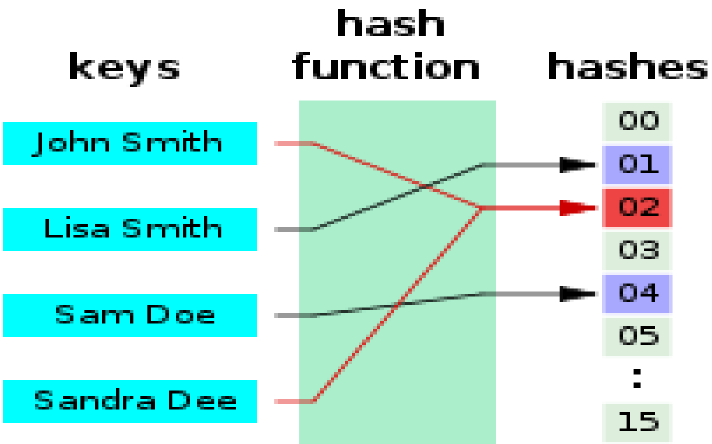


Fig3. Hash Function

The image is first rescaled to a fixed size with bilinear interpolation, and converted from RGB to the YCbCr representation. Y and are used as luminance and chrominance components of the image to generate then hash. The aim of rescaling is to ensure that the generated image hash has a fixed length and the same computation complexity

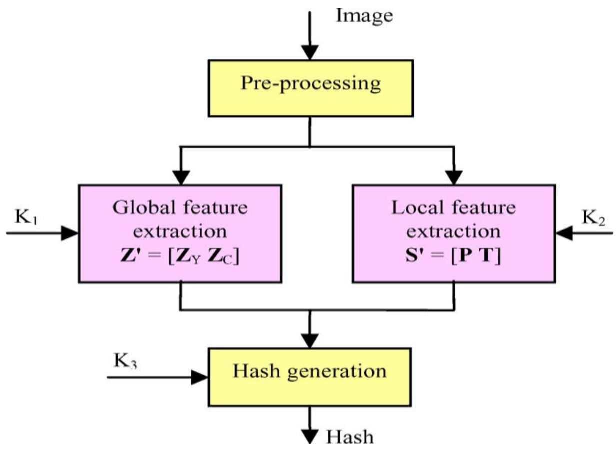


Fig.4 Block diagram of the proposed image hashing method

a. Original Image

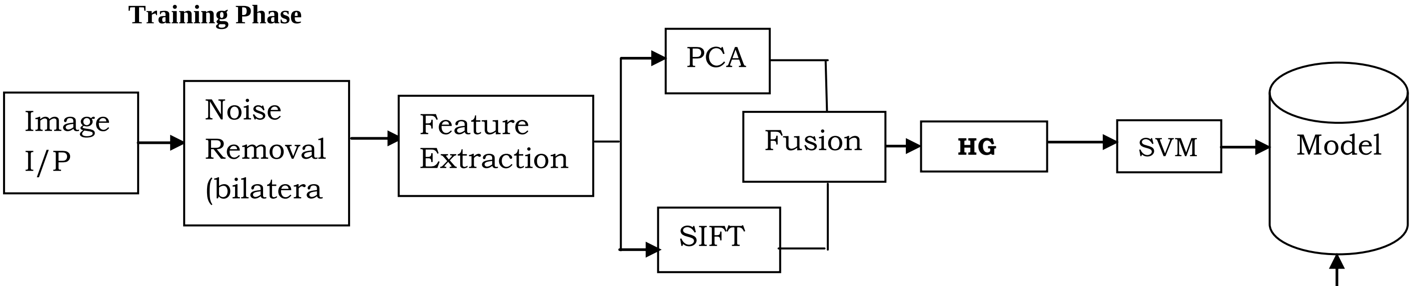
b. Input Image of Forbidden

c. Feature Map

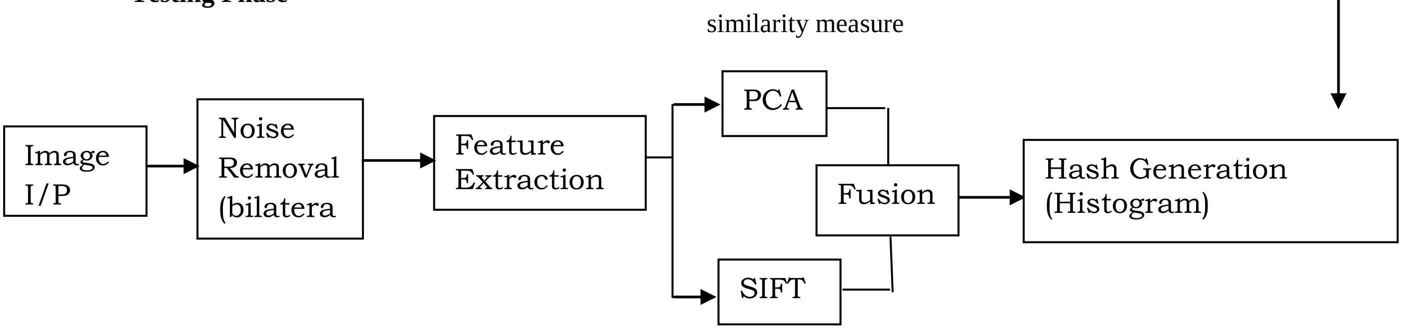
d. Marked Tampered

4. Architecture Diagram

Training Phase



Testing Phase



V. Conclusion

In this paper, an image hashing method is developed using both global and local features. The global features are based on SIFT algorithm representing the luminance and chrominance characteristics of the image as a whole. The local features include PCA algorithm regions in the image. Hashes produced with the proposed method are robust against common image processing operations including brightness adjustment, scaling, small angle rotation, JPEG coding and noise contamination. Collision probability between hashes of different images is very low. The proposed scheme has a reasonably short hash length and good ROC performance. The method described in this paper is aimed at image authentication. The hash can be used to differentiate similar, forged, and different images. At the same time, it can also identify the type of forbidden and locate fake regions containing salient contents. In the image authentication, a hash of a test image is generated and compared with a reference hash previously extracted from a trusted image. When the hash distance is greater than the threshold but less than, the received image is judged as a fake. By decomposing the hashes, the nature of image forgery and locations of forged areas can be determined. It should be stressed that the success of image authentication using the proposed scheme depends to a large extent on the Hashes produced with the proposed method are robust against common image processing operations including brightness adjustment, scaling, small angle rotation, and JPEG coding and noise contamination. Collision probability between hashes of different images is very low. The proposed scheme has a reasonably short hash length and good ROC performance.

References

[1] Baxes, G. A., Digital Image Processing: Principles and Applications. New York: John Wiley & Sons,Inc, 1994.
[2] H. Farid, "Detecting digital forgeries using bispectral analysis, "AI Lab, Massachusetts Institute of Technology, Tech. Rep. AIM-1657, 1999.
[3] J. Fridrich, D. Soukal, and J. Lukás, "Detection of copy move forgery in digital images," in Proc. Digital Forensic Research Workshop, Aug. 2003.
[4] A. C. Popescu and H. Farid, "Exposing Digital Forgeries by Detecting Traces of Re-sampling," IEEE Trans. on Signal Processing, vol. 53,no.2, pp. 758–767, Feb. 2005
[5] T. Ng, S.F. Chang, and Q. Sun, "Blind Detection of Photomontage using Higher Order Statistics", IEEE International Symposium on Circuits and Systems, Canada, May 2004
[6] A. C. Popescu and H. Farid, "Exposing Digital Forgeries by Detecting Traces of Re-sampling," IEEE Trans. on Signal Processing, vol. 53,no.2, pp. 758–767, Feb. 2005.
[7] S. Murali, Anami Basavaraj S, and Chittapur Govindraj B. " Detection of Digital Imager forgery Using Luminance Level Techniques", IEEE Third National Conference of Computer Vision, Pattern Recognition, Image processing and Graphics, NCVPRIPG 2011 pp.215.
[8] Fredric, J. and J. Lukas, "Estimation of Primary Quantization Matrix in Double Compressed JPEG Images." Proceedings of DFRWS 2003. Cleveland, OH, August 2003.
[9] E.S.Gopi, N Lakshmanan, T.Gokul and S.Kumar Ganesh "Digital image forgery detection Using artificial Neural Network and Auto Regressive Coefficients" EEE/CCCGET, Ottawa, May 2006 1- 4244-0038-4 2006.
[10] Y.Sutch, B.Coskum and H.T.Sencar,N. Memon "Temper Detection Based On Regularities Of Wavelet transform Coefficients " Polytechnic University,Electrical Computer Engineering Dept., Brooklyn Ny 11201, USA
[11] Saha, S. "Image Compression – from DCT to Wavelets: A Review," May 28, 2004 http://www.acm.org/crossroads/xrds6-3/sahaimgcoding.html.
[12] Fridrich, J. and J. Lukas, "Estimation of Primary Quantization Matrix in Double Compressed JPEG Images." Proceedings of DFRWS 2003. Cleveland, OH, August 2003.
[13] Fridrich, J., R. Du, and M. Goljan, "Steganalysis Based on JPEG Compatibility," Special session on Theoretical and Practical Issues in Digital Watermarking and Data Hiding, Multimedia Systems and Applications IV. Pp. 275-280. Denver, CO, August 2001.
[14] Johnson, R. C. "JPEG2000 Wavelet Compression Spec Approved," August 18, 2004. http://www.us.design-reuse.com/news/news1917.html.
[15] A. Garg, A. Hailu, and R. Sridharan "Image Forgery Identification using JPEG Intrinsic Fingerprints" Authors copy
[16] Murali S.,Anami B. S, Chittapur G. B. "Digital Photo Image Forgery Techniques" International Journal Of Machine Intelligence ISSN: 0975-2927 & E-ISSN: 0975-9166, Volume 4, Issue 1, 2012, pp.-405.
[17] R. Venkatesan, S. M. Koon, M. H. Jakubowski, and P. Moulin,\Robust image hashing," *Proc. IEEE*

