CHAPTER 2
THEORETICAL BACKGROUND AND LITERATURE REVIEW

There are two different classes of image authentication- one is active protection schemes, and another is passive detection techniques. In an active protection scheme, digital signatures or watermarks are attached to the image during the formation process of the image. The manipulated region can be identified in the image with the help of this signature or watermark. However, for image authentication using an active protection scheme, prior knowledge of the image is important. Another image authentication technique is passive detection. This technique uses clues left by the different device components of the digital camera during the image acquisition process. A detailed literature review of different passive detection techniques has been done in this chapter. This review is done in two different sections- the first is dedicated to passive detection techniques for CMF and the second is for image splicing. This chapter also summarizes the research gaps reported in the literature. Except for the literature review, the theoretical background of the research has been discussed in this chapter. This includes publicly available datasets used for the validation of the proposed models and performance measures used for the evaluation purpose.

An image is processed through various camera components before the final image is produced in the form of pipelining after capturing it from any digital camera [12]. The captured image is modified by a processing algorithm each time when it passes through any component (Figure 2.1). These processing algorithms may leave some intrinsic fingerprint clues to detect the tempered area. Based on these clues, methods are reported in various literature. Considering these reported works of literature forgery detection

methods are divided into two categories i.e. copy-move forgeries detection and spliced image forgery detection.

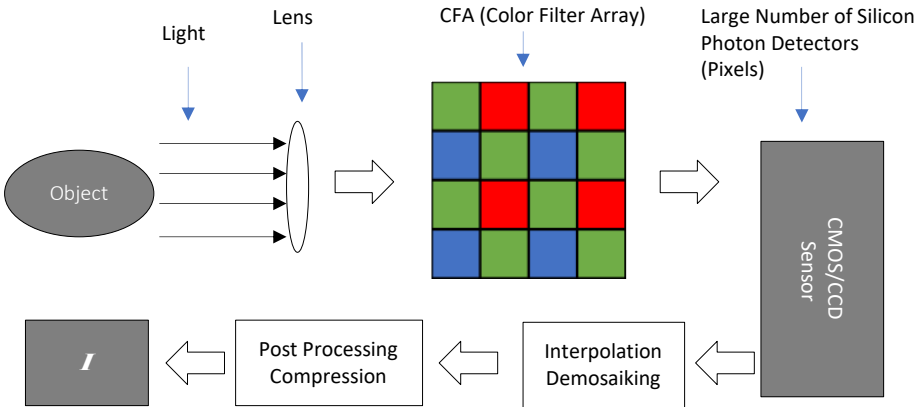

Figure 2.1: An Image acquisition pipeline

2.1 Literature review on Copy-Move Forgery Detection

This section reviews some state-of-the-art techniques for the detection of copy-move forgery (CMFD). These methods can be divided into three types- one is block-based methods, the second is key-point feature extraction-based methods and the third is data-driven approaches (i.e. machine learning and deep learning techniques). Literature reported in journals and reputed conferences is classified into three classes in the below subsections based on types of methods.

2.1.1 Block-based Approaches

The first one is block-based techniques. The major steps that are performed in the block-based matching technique are- block division, feature extraction and then feature matching as shown in Figure 2.2.

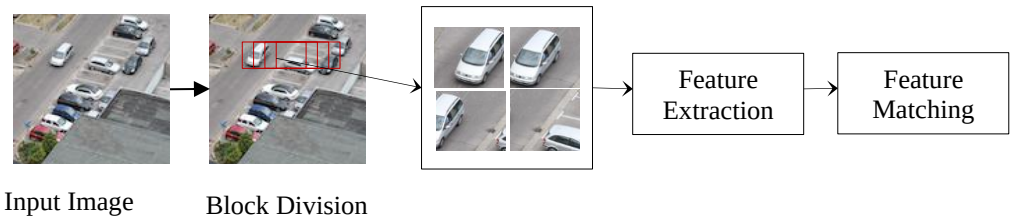

Figure 2.2: Steps involved in Block-Based CMFD Techniques

Jessica Fridrich et al. [27] was the first who proposed a block-based method for CMFD by two different matching techniques. In this method, the image is extracted into $b \times b$ sized overlapping patches and each patch are sorted into lexicographical order. Two or more identical patches are saved as duplicated regions in exact matching while the shift vector is used in robust matching. No publicly available dataset has been used for validation purpose. Also, this method fails when any type of geometric transform has been performed over the duplicated region. Weiqi Luo et al. [28] proposed another method based on block matching. In this method, three colour features and four statistical features are extracted and then sorted in lexicographical order. Identical feature blocks are saved as duplicated regions. Here also, no publicly available dataset has been used for validation purposes. This method also fails with the geometric transformation of duplicated regions. Babak Mahadian [29] proposed another method using blur moment features. In this method, the image is first divided into overlapping blocks and then 24 blur moment features have been extracted from each overlapping block. Then principal component analysis (PCA) is applied to features and features are sorted using a K-dimensional tree. This method works well in the case of additive noise, but the computation cost of the method is very high. This method is not robust for geometric transformation too. Toqueer Mahmood et al. [30] proposed a method using stationary wavelet transform (SWT) and local binary patterns (LBP). In this method image is first transformed into the SWT then the transformed image is divided into circular overlapped blocks. The LBP feature descriptors are extracted from overlapped blocks. Then these features are matched using Euclidian distance. The problem with this method is that this doesn't work with affine transformation.

Except for these, [29]–[35], [35] methods are also based on block matching based. In these methods, the image is divided into blocks and the features are extracted from

blocks then these extracted features are matched together using different matching techniques. Problems with these techniques are- computation cost is very high and not robust with geometrical transformation.

2.1.2 Key-point-based Approaches

The second type of CMFD is keypoint feature-based techniques. In this technique major steps performed are- preprocessing, keypoint feature extraction from the image and feature matching as shown in Figure 2.3.

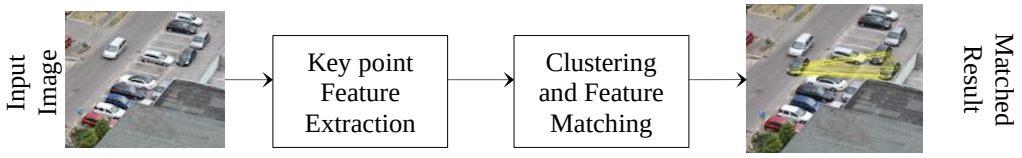

Figure 2.3: Steps involved in Keypoint Matching Based CMFD Techniques

Irene Amerini et al. [36] proposed a keypoint-based CMFD technique in 2011. In this method, scale-invariant feature transform (SIFT) keypoints are extracted from the image of 180 features then instead of using direct Euclidian distance a ration of distance pair has been calculated and made a cluster of nearest neighbours using generalized 2NN for matching the key points in the image and detected the duplicated regions in the image. The problem with this technique is that it gives lower accuracy result in the case of small, duplicated regions. Fan Yang et al. [37] CMFD Based on Hybrid Features proposed a method using SIFT and KAZE keypoint features. In this method, hybrid features are extracted from the image and then are matched with the gNN matching technique. Although it works with some geometrical transform of the duplicated region but fails with smooth images. A lot of hybrid approaches are reported in the literature by combining block-based and keypoint features based [38]. All these either have the problem of computation cost or performance.

2.1.3 Data-driven Approaches

The third type of CMFD approach is data-driven based. In this method some training data are used to train a classification model and using this model forged images can be predicted as forged or original. Also, some of the data-driven approaches use ground truth mask for training purpose and this type of model localize the duplicate region in the forged image (Figure 2.4).

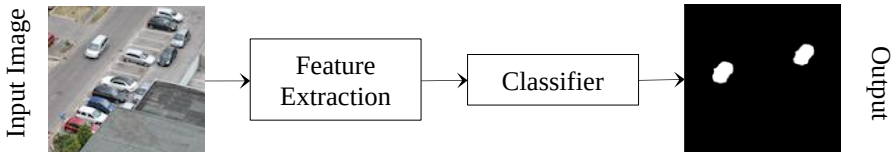

Figure 2.4: Steps involved in Keypoint Matching Based CMFD Techniques

Yaqi Liu et al. [39] proposed a data-driven-based deep learning approach for feature extraction from key points. Then k-NN is used on extracted features for feature mapping which is used for matching. Another approach of deep learning is by Yue Wu et. al [40]. In this technique convolution layers are used to extract features then the correlation is performed on point-wise extracted features then these features are deconvoluted by upsampling to localize the result. Mohamed A. Elaskily et. al. [41] proposed another deep learning-based CMFD. This method is a classification of an image into forged and original not the localization of the forged region (i.e. image-wise classification is performed instead of pixel-level classification). Other problems with other approaches are performance and robustness with geometrical transformation.

2.1.4 Research Gaps and Findings

Based on the above literature, we have identified some points of the reported literature. These points are summarized in below Table 2.1. In this table A, S and R represent the Availability of the dataset, the number of images in the dataset and the resolution per image, respectively.

Table 2.1: Related works and their comparison on different parameters

Type	Method	Features	Matching Technique	Dataset			Evaluation Metrics	Robustness against transformations	Remark
				A	S	R			
Block-Based	Fridrich et al. [27]	DCT	Lexicographical Sort	No	-	-	-	Doesn't work with transformation	Computation Cost is very high
	L. Weiqi et al. [28]	Colour and Statistical	Lexicographical Sort	No	100	300x400	Accuracy, FNR	Doesn't work with transformation	Lower computation than previous
	Mahmood et al. [30]	SWT and LBP	Euclidian Distance	Yes	10400	768x512	TDR, FDR	Scale-Invariant	Lower Computation than previous
Keypoint Based	Amerini et al. [36]	SIFT	Nearest Neighbour Clustering	Yes	2000	800x600	TPR and FPR	Scale-Invariant	Poor Performance in case of the small-duplicated region
	Yang et al. [37]	SIFT and KAZE	Multiple copied feature matching	Yes	-	3000x2300	P, R, F1	Rotation and Scale Invariant	Poor performance with smooth images
Data-Driven	Y. Wu et al. [40]	Deep Features	Classification Pixel wise	Yes	>10000	256x256	P, R, F1	Robust against affine transformation, blurring and JPEG compression	Performance is not satisfactory
	Elaskily et al. [41]	Deep Features	Classification Image wise	Yes	>3000	< 3888 × 2592	A, TPR, FPR	-	Only Image-Level Classification

From the table, it can be summarized that the CMFD method should be robust against geometric transformation and computation cost should be low for image-level detection as well as pixel-level analysis. Key-point feature extraction techniques-based methods are robust against geometric transformation but then the problem with these methods is poor performance. A deep learning data-driven approach can be used here to overcome these challenges.

2.2 Literature review on Spliced Image Detection

Image splicing is another type of digital image forgery. This section elaborates on some latest research works associated with image splicing detection. These works are presented and analyzed based on their performance on the publicly available dataset and their overall cost. These works are either based on statistical or data-driven approaches. Statistical works are based on clues/traces left by the component of digital cameras during the image acquisition process and data-driven approaches are detection mechanisms based on machine learning.

2.2.1 Data-driven Techniques

Most of the works discussed here use an approach where features from blocks of the pre-processed image are extracted and these features are trained using the machine learning classification approach.

Zhao et al. [42] suggested that if image splicing detection is tough in one color space then probably it may be easier in another color space, so they derived a passive technique of image splicing detection in which a chroma channel is used to extract 4 Gray level run length number with different directions as feature vectors. This approach uses a support vector machine (SVM) classifier to classify forged images. The method also shows that these extracted features have better performance than those features extracted from individual Red, Green and Blue luminance channels. The experiment was performed on CASIA v1.0 and COLUMBIA datasets where accuracy measured on Cb and Cr were 94.3%, 94.7% and 82.1%, 85% respectively.

Another splicing detection scheme is based on discrete cosine transform (DCT) and Local Binary Pattern (LBP) in which Amani *et. al.* [43] recommended a novel approach of a passive technique for image splicing forgery detection. In this technique

RGB input image is first converted into YCbCr color space then its chrominance channel is divided into overlapping blocks, from each overlapping block LBP image is derived. These LBP images are transformed from the spatial domain to its 2D DCT frequency domain, from which DCT coefficients are used as a feature vector. These feature vectors are given to the SVM classifier to classify forged and authentic images. In this approach three datasets CASIA v1.0, v2.0 and COLUMBIA were taken, and the performance measured was 97%, 97.5% and 96.6% respectively.

Wei Wang *et. al.* [44] proposed an approach for splicing detection based on the Gray Level Co-occurrence Matrix(GLCM) features of the threshold image. First, the image is converted into YCbCr color space. The authors introduced that chrominance channels are more sensitive than luminance. They showed the edges of spliced images are sharper than the original ones. In this approach, the authors took GLCM of the chrominance channel. Since gray values of edges in these channels are not big, they threshold it to a reasonable value to reduce the size of GLCM features. To reduce the dimension of the feature vector and increase the accuracy of the classifier, a Boost Feature Selection (BFS) method was used. Then these feature vectors are trained using LIBSVM classifier to detect the forged image. In this approach, GLCM features are used only, while orientation information of an image is not used. The highest accuracy rate achieved by this approach was 90.5% with 50 dimensions.

Zhongwei *et. al.* [45] proposed an approach for image splicing detection based on Markov features in DCT and DWT domain. First, from the input image Markov features are extracted from its DCT and DWT transform coefficients, then a feature selection method is used to reduce the dimension of the feature vector using SVM-RFE which reduces its computational costs. Finally, the SVM classifier is used to classify the spliced

and authentic image. The maximum accuracy rate in this approach achieved was 93.55% on COLUMBIA dataset while 89.76% on the CASIA v2.0 dataset.

Ghulam Muhammad et al. [46] have given a technique using a steerable pyramid transform and LBP. In this technique, a color image is first converted into YCbCr color space and its chrominance channels are transformed into steerable pyramid transform. From its sub-bands, the LBP transform is extracted, and a histogram of each LBP transformed sub-bands fused to classify the images. SVM classifier is in the proposed technique to classify images into spliced and authentic. Though the accuracy in this approach was higher in this method as the approach achieved 97.33% on the CASIA v2.0 dataset, the feature of LBP histogram is used in this approach whereas scale and orientation information of the image is left.

Agarwal et al. [47] proposed a splicing image detection scheme using a multi-scale entropy filter and local phase quantization. In this proposed method, a color image is converted first into a YCbCr image then from its chrominance channel, an entropy filter is used to highlight the boundary of the forged image. Then LPQ operator is used to providing internal statistics of the image using its phase information. Histogram of each feature is then fused and is given to the SVM classifier for distinguishing non-forged and forged images. In this work, the authors have discussed that method works well for both copy-move forgery and splicing detection. Since multiple entropy filter sizes have been used in the method on both chrominance channel dimension of the feature vector has been increased. Two class problems can be solved by the SVM classifier well when the size of the dataset is balanced and small. In this approach accuracy rates measured on CASIA v1.0, CASIA 2.0 and COLUMBIA were 95.41%, 98.33% and 91.14% respectively. It shows that the method doesn’t perform well in the absence of texture patterns in the image.

Abrahim *et. al.* [48] presented a framework to identify the spliced image by exploiting image texture features. The proposed framework uses different texture features and color features of the image such as LBP, Histogram oriented Gradient (HoG) and Higher-order statistical features. Then these features are combined for feature level fusion to make a feature vector. These feature vectors are trained using Artificial Neural Network (ANN) to classify the forged image. Another model also introduced in this framework is majority voting in which different features are directed feed into an ANN classifier. Though the accuracy rate has been increased in the approach effective cost and time have been also increased.

Recently, deep learning techniques are applied in almost every application of computer vision and image processing. Deep learning is defined as a system that is artificially intelligent and mimics the functionality of neurons of the human brain to perform the task of decision-making by analyzing the data and concluding a pattern out of it. It works on unsupervised and supervised learning algorithms and follows the hierarchical leveling of ANN of machine learning. Convolutional Neural Network (CNN) [49] is a very prominent architecture of deep learning used for the prediction of images belonging to a class. This architecture is also used with some external features for the detection and localization of the forged region in an image [50]. A deep learning CNN architecture [51] is used for the detection of forged and authentic images. This model can classify images into fake or authentic images but is unable to define the region of forgery. Another technique [52] uses hybrid LSTM (Long-Short Term Memory) and Encoder-Decoder architecture. In this model, the authors utilize features of two different architecture for the localization of the forged region. Also, they have provided a huge volume of public datasets for the training and testing of deep learning models.

2.2.2 Statistical Techniques

This section presents a brief review of Image forgery detection using clues left by the component of cameras. During the image acquisition process, each operation in the image acquisition pipeline leaves an intrinsic peculiar fingerprint. Maximum image forensic tools reported in the literature have relied on these exploited assets.

Physical traces are based on inconsistent lighting conditions [17], [18], [53]–[56], perspective or geometry projection [19], [57], [58]. In these techniques vanishing points of perspective projection are computed which is built on the assumption of a pinhole camera model that centrally projects 3D space points on a 2D plane. Forgery in the image can be found in the case of inconsistent vanishing points. Although these physical traces are effective but are unable to localize accurate forged regions in an image. To capture color or natural images most of the image acquisition devices use Bayer Color Filter Array (CFA). This filter gives color components (Red, Green, Blue) to each pixel for the incoming light. In this way, a color image has interpolation of the periodic structure of the CFA, which is also known as demosaicing. The absence of periodic interpolation structures identifies a forged location in fake images [20]–[22], [59]. A very fundamental limitation of this type of detection is the periodic structure can be destroyed by compressing the image. Though JPEG is an old compression technique, it is still used by default in many image acquisition devices and web services of social media. Computer-generated graphics are also converted to JPEG-type images for the uploading on web or publishing in news content. This compression technique transforms distinct 8x8 blocks of the image from spatial domain to frequency domain (DCT) and obtained coefficients are quantized using a quantization table. Since compression is performed on 8x8 distinct blocks, content copied to the original image to forged may lead to blocking artifacts that can reveal the forged region in the fake image [23]–[25], [60].

Another assumption is that image pixels have minor variations due to the imperfection in camera sensors which leads to homogeneous and unique patterns of noise. This pattern is also robust with mild post-processing such as compression and scaling of the image. This pattern is used to localize the spliced or fake content in the image. Most of the techniques [13], [14], [16], [61], [62] only improve the quality of the estimated noise pattern.

Mahadian and Saic [16] have proposed a method to detect forged regions based on noise level estimation of different regions in the image. This work divides the image into various partitions based on the different noise levels. In this work, the image is first decomposed into a wavelet transform. Then Median absolute deviation (MAD) of the detailed coefficient (diagonal component) is computed on fixed-size non-overlapping blocks. Based on this MAD value, noise levels are distinguished. Then post-processing operation Block-Merging is performed to segment different regions. The problem with this technique is that the technique has block merging segmentation to divide partition, but the threshold value is fixed instead of automatic thresholding. Also, the technique doesn't care about the probability when edges may consider as noise.

Lyu and Pan [14] suggest an effective method to expose the spliced region using local noise level inconsistency. In this work noise level is computed using projection kurtosis concentration of natural image in the bandpass domain. These noise level variances are calculated in different bandpass domains such as DCT, PCAS, HAAR, and RAND. To detect forged regions first noise variance is calculated locally for all pixels. Then morphological operations are performed to connect similar neighboring pixels. In this work, there is a lack of automatic segmentation technique used to distinguish the forged and non-forged regions.

A noise level function (NLF) based image splicing detection technique is proposed by Yao et. al. [63]. The authors claim that the noise distribution used in this work is intensity-dependent. A model described here is NLF that fits actual noise characteristics. In this method, the image is first segmented into edge and non-edge regions. Then NLF of both regions is computed. Based on the predefined threshold of NLF value, regions are connected. Thereafter forged and non-forged regions are divided.

Another method has been proposed using NLF by Zhu and Li [13]. In this work special type of forgery (image splicing) is detected using noise variance and sharpness value of non-overlapping blocks. In this method, the image is first divided into non-overlapping blocks. Then noise variance and sharpness value of each block are calculated. Based on the relationship between these noise variance and sharpness values the NLF finds. Using this NLF distance map can be generated to find the minimum distance of the block from the fitted curve. Based on this distance forged region can be detected. To segment the forged region here convolution operation of a filter with the mapped image is performed.

2.2.3 Research Gaps and Findings

The above-mentioned data-driven approaches in section 2.2.1 conclude that some approaches use color and texture features like GLCM, LBP and HoG [48] [46] and some of them use frequency-based features like DCT and DWT [45]. The following points summarize the limitations of the above-mentioned methods.

- The global features used in the approaches have the advantage of ease to compute, faster and compact. Except [45] and [48], none used orientation and scale features.
- Information regarding translation and rotation is not extracted in the methods. Features for smooth edges from images are not identified leading to loss of information.

To classify the image into forged or non-forged classes these approaches use SVM or ANN-based machine learning techniques. Though SVM can handle large feature space it is not efficient with a large number of observations

From the above-explained statistical methods for image splicing detection in section 2.2.2, it can be said that the major limitation with these techniques is dependent on the quality level difference of compression. If this difference is sufficient then only localization will be better otherwise performance will be degraded. From the above-mentioned reasons, it can be concluded that mathematical models constructed using traces from individual steps of the acquisition pipeline can work on a single application with a single assumption. They are not able to detect and localize all types of forgeries with a single trace. And techniques based on noise inconsistency as clues perform better than others. Though noise inconsistency-based techniques have advantages over other clues, they have limitations in their post-processing operation. Segmentation of forged region in these methods are based on fixed threshold values. Also, most of these techniques don't care when edges are considered as noise during the estimation of noise inconsistency which results in false detection of the forged region sometimes. Table 2.2 explains briefly the properties of the above-mentioned works of literature. Here meanings of symbols are as given:

P1: The algorithms work with Natural Images,
P2: Morphological operations are mentioned in the work,
P3: The method can distinguish spliced regions,
P4: Block-wise Noise estimation,
P5: The algorithm has pixel-level analysis of spliced and non-spliced pixels,
P6: Block-level Analysis,
P7: Don't need prior knowledge about the image,
P8: Works for small-size images.

Table 2.2: Advantages and Limitations of various state-of-the-art techniques for Image Splicing Detection

Methods	P1	P2	P3	P4	P5	P6	P7	P8
Lyu et al. [14]	✓	×	×	✓	×	✓	✓	✓
N. Zhu and Z. Li [13]	✓	✓	✓	✓	×	✓	×	×
H. Yao et al. [63]	×	×	✓	×	✓	×	✓	✓
Riess et al. [16]	×	×	×	✓	×	✓	✓	×
Pospescu and H. Farid [64]	×	×	×	×	✓	×	×	✓

The most important challenge with these detection techniques is the estimation of the qualified noise pattern. Other limitations of these techniques are the sequence of morphological processes used for segmentation of forged regions (i.e. manual operations are used) and defining block size for estimation of noise patter (if the block size is taken large, the small fake regions will not be able to detect).

2.3 Dataset Used for Experimental Study

Digital image forgery detection is still an open problem. A lot of techniques have been already reported to detect the forged image and to localize the manipulated region in the forged image. These techniques are evaluated on some benchmarking datasets. These datasets are publicly available. These datasets are used in this thesis to evaluate the proposed models.

2.3.1 CoMoFoD

Figure 2.5: The instances of the CoMoFoD dataset

CoMoFoD is a benchmark publicly available dataset that contains copy-move forged images. This dataset can be used for the evaluation of the CMFD techniques. The dataset consists of various geometrically transformed copy-move forgery attacks i.e. scaling and rotation. Except these, the dataset has a collection of various mild processing copy-move forgery attacks i.e. brightness enhancement and contrast adjustment.

Table 2.3: Details of CoMoFoD dataset

S. No.	Transformation	Resolution	Forged Images	Ground Truth
1	Natural (Without Transformation)	512 × 512	40	40
2	Scaling	512 × 512	40	40
3	Rotation	512 × 512	40	40
4	Distortion (Skew)	512 × 512	40	40
5	Combination	512 × 512	40	40
6	JPEG Compression	512 × 512	1800	1800
7	Image Blurring	512 × 512	600	600
8	Noise Addition	512 × 512	600	600
9	Brightness Change	512 × 512	600	600
10	Colour Reduction	512 × 512	600	600
11	Contrast Adjustment	512 × 512	600	600
		Total	5000	5000

Table 2.3 contains the type of transformation used for forgery, resolution of the image and number of images and respected ground truth. The instances of the datasets are visualized in Figure 2.5. The visual contains three rows. The first row shows the original image, the second row shows the forged image of the corresponding original image and the last row shows the ground truth mask of the respected forged image.

2.3.2 CMFD

CMFD is also a publicly available standard dataset for the evaluation of CMFD techniques. Similar to CoMoFoD, this dataset also contains some geometrical transformations of the forged region in the manipulated image. These transformations are only scaling and rotation of the forged region. The rotation angle ranges from -25 to 360

and the scaling factor ranges from 0.25 to 1.25. More details of the dataset are given in Table 2.4.

Table 2.4: Details of CMFD dataset

S. No.	Transformation	Resolution	Forged Images	Ground Truth
1	Natural (Without Transformation)	700 × 1000	100	100
2	Scaling	700 × 1000	320	320
3	Rotation	700 × 1000	600	600
		Total	1020	1020

The instances of the datasets are visualized in Figure 2.6. The visual contains three rows. The first row shows the original image, the second row shows the forged image of the corresponding original image and the last row shows the ground truth mask of the respected forged image.

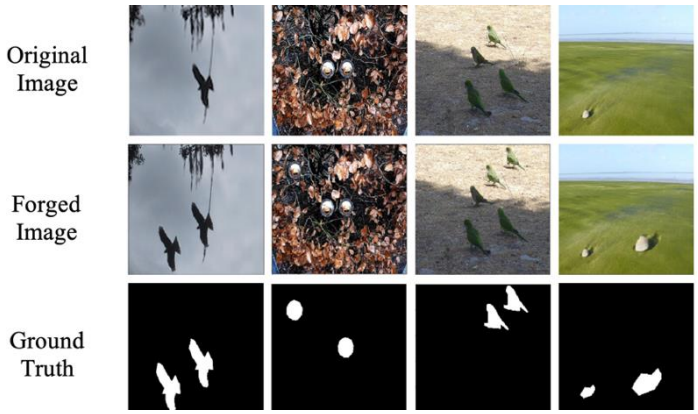

Figure 2.6: The instances of the CMFD dataset

2.3.3 CASIA v1.0 and CASIA v2.0

CASIA v1.0 and CASIA v2.0 datasets are the benchmarks for image splicing forgery. The first dataset CASIA v1.0 [65] is made by the splicing operation over the authentic dataset. The region from the original image is cropped and pasted to different original images, using Adobe Photoshop CS3 in Windows XP operating system. These images in the dataset are categorized into eight different categories (texture, nature, scene, plant, character, architecture, article and animals). The cropped region had gone through various distortion, rotation and resizing operations. The dataset has 1721 images of

384×256 images with 800 authentic and 921 spliced color images. The second dataset CASIA v2.0 is also the same as the above dataset and has a total of nine categories of images with an extra category of indoors images. The images have gone through various processing like resizing, rotation and other distortion. Also, some post-processing operations like blurring after cropping of the region were performed. The dataset has different format images (.jpg and .tif). The dataset contains 12614 images with different sizes from 240×160 to 900×600 pixels, with 7491 authentic and 5123 spliced images. Table 2.5 explains both datasets.

Table 2.5: Details of CASIA v1.0 and CASIA v2.0 datasets

S. No.	Dataset	Resolution	Number of Images	Format
1	CASIA v1.0 [65]	384×256	Total: 1721 (800 Authentic and 921 Spliced)	.jpg
2	CASIA v2.0	240×160 to 900×600	Total: 12614 (7491 Authentic and 5123 Spliced)	.jpg and .tif

2.3.4 IEEE IFS Dataset

IEEE IFS [66] is a dataset which is having both types of digital image forgeries (i.e. copy-move forgery and spliced image). It was created for IEEE IFS-TC Image Forensic Challenge. This dataset is very important to evaluate the image spliced detection techniques. This dataset has its ground truth data to verify the resultant image. The dataset is created through editing applications such as Adobe Photoshop CS5 and GNU Gimp. To manipulate images different algorithms are used such as Clone-stamp and Patch-Match for copy-move forgery whereas Alpha-Matting and Content-aware healing for image splicing. This is a dataset having 1500 images in PNG format with high-resolution images. The resolution of the image in the IFS dataset is 1024×768 pixels. The most important specialty of this dataset is the spliced region can't be seen by nude eyes. The visual demonstration of the dataset is shown in Figure 2.7.

Figure 2.7: Demonstration of IEEE IFS Dataset

2.3.5 Columbia Uncompressed Dataset (CUD)

CUD [67] is a standard and publicly available dataset dedicated to the spliced image. It has a combination of the spliced image of a different camera device. These images are combinations of Canon-Canon, Canon-Kodak, Canon-Nikon and Nikon-Kodak. This dataset has 180 images in total with high resolution and TIF format. It has edge masks of forged images. These 180 spliced images have resolution ranges from 757×568 to 1152×768. Most of the images are indoor scenes such as labs, desks, books etc. Figure 2.8 visualizes the examples of this dataset.

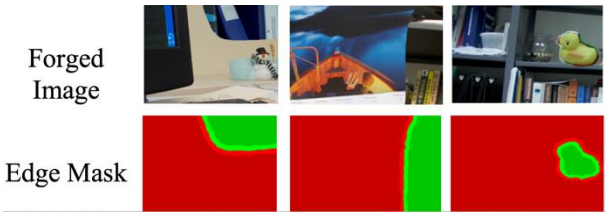

Figure 2.8: Visualization of Columbia Uncompressed Dataset

2.4 Evaluation Metrics

Evaluation metrics are used to measure the performance and quality of the given model. Contributions made in this thesis are either statistical approaches or data-driven approaches for digital image forgery detection and localization. Here, detection defines the image level analysis and localization defines pixel-level analysis of the image.

Image Level Analysis: In the image-level analysis of the forgery detection, the proposed approach classifies the given image into binary class whether the image is forged or authentic.

Pixel Level Analysis: In pixel-level analysis, the proposed approach classifies pixels of the image into two classes whether the pixel of the image is forged or authentic.

To evaluate the performance of the proposed technique, classified instances (here image and pixel both will be called as an instance for the simplicity) are first put into a confusion matrix. This confusion matrix has four cells and can be defined as a collection of positive and negative cases deduced by the proposed technique. If forged instances are taken as positive and authentic instances are taken as negative, then these cases are-

***True Positive* (*tp*):** Number of forged instances predicted as forged by the technique.
***True Negative* (*tn*):** Number of authentic instances predicted as authentic by the technique.
***False Positive* (*fp*):** Number of authentic instances predicated as forged by the technique.
***False Negative* (*fn*):** Number of forged instances predicted as authentic by the technique.

The confusion matrix can be drawn -

		Ground Truth Value	
		Positive (1)	Negative (0)
Experimental Instance Value	Positive (1)	True Positive	False Positive
	Negative (0)	False Negative	True Negative

Evaluation metrics are defined based on the above confusion matrix. Mathematically, these metrics can be formulated as:

Precision: Proportion of correctly identified pixels as positive and all identified pixels as positive from the resultant image. Similarly, the proportion of correctly identified images as positive class and all identified images as the positive class.

$$precision\ (p) = \frac{tp}{tp + fp} \tag{2.1}$$

Recall: Number of pixels corrected identified as positive from all positive in ground truth mask. Similarly, the number of images corrected was classified as a positive class from all positive class images.

$$recall\ (r) = \frac{tp}{tp + fn} \tag{2.2}$$

Accuracy: Proportion of correctly classified pixels or ratio of correctly classified images:

$$accuracy\ (a) = \frac{tp\ + tn}{tp + tn + fp + fn} \tag{2.3}$$

Specificity: Proportion of correctly identified negative pixels from resultant image and actual negative pixels of ground truth mask. Similarly, in image-level analysis- the ratio of correctly classified images as a negative class and actual negative class images.

$$specificity\ (s) = \frac{tn}{tn + fp} \tag{2.4}$$

Sensitivity: Proportion of correctly identified positive pixels from resultant image and collectively true positive and false negative pixels. Similarly, in image-level analysis- how many actual forged images are correctly classified as forged images.

$$sensitivity\ (tpr) = \frac{tp}{tp + fn} \tag{2.5}$$

Miss-rate: Miss rate is a very important performance measure which gives knowledge about misses positive classes and miss rate should be minimum. It can be defined as the number of incorrectly identified pixels as negative from all positive pixels in the ground truth mask. Similarly, the number of incorrectly identified images as negative from all positive images.

$$missrate\ (m) = \frac{fn}{tp + fn} \tag{2.6}$$

Critical Success Index: Proportion of correctly identified pixels as positive and *tp+fp+fn.* Similarly in the case of image level, the proportion of correctly identified images as positive and *tp+fp+fn.*

$$csi = \frac{tp}{tp + fp + fn} \tag{2.7}$$

F1-Score: The F1-score is a harmonic mean of precision and recall.

$$f1 - score\ (f1) = \frac{2 \times p \times r}{p + r} \tag{2.8}$$

Mathew's correlation coefficients: It can be defined as the correlation coefficient of predicted class and a true class of pixels as well as image.

$$mcc = \frac{(tp \times tn - fp \times fn)}{\sqrt{\big((tp + fp) \times (tp + fn) \times (tn + fp) \times (tn + fn)\big)}} \tag{2.9}$$

There is no single evaluation metric that describes the confusion matrix in its best way. Sometimes precision, recall, and accuracy mislead results that's why it is important to measure results using miss-rate, f1- score, and MCC values.