FUDMA Journal of Sciences (FJS)
ISSN online: 2616-1370
ISSN print: 2645 - 2944
Vol. 3 No. 4, December, 2019, pp 449 –458

DETECTION AND LOCALIZATION OF SPLICING FORGERY IN DIGITAL VIDEOS USING CONVOLUTIONAL AUTO-ENCODER AND GOTURN ALGORITHM

***Sulaiman, N., Bagiwa, M. A., Aliyu, S., Shafii, K., Usman, A. M., Mohammed, S. and Abdulsalam, A. J.**

Department of Computer Science, Ahmadu Bello University, Zaria

*Corresponding Author's Email: nuwair68@gmail.com 08132979195

ABSTRACT

In present days, individuals in the society are becoming increasingly relying on multimedia content, especially digital images and videos, to provide a reliable proof of the occurrence of events. However, the availability of powerful and user-friendly video editing tools makes it easy even for a novice to manipulate the content of a digital video which may be used as evidence during digital investigation. This has led to great concern regarding the authenticity of digital videos. Several techniques have been developed to detect video forgeries, but only few focused on video splicing forgery detection. However, those few techniques that focused on splicing forgery detection tend to depreciate in terms of accuracy of detection when the video is compressed. Therefore, it is important to devise new technique that can detect and localize splicing forgery for both compressed and uncompressed video. In this paper, a hybrid technique for the detection and localization of splicing forgery in both compressed and uncompressed digital videos using convolutional auto-encoder and Generic Object Tracking Using Regression Network (GOTURN) algorithm is proposed. The parameters of the auto-encoder are learned during the training phase on original video frames. During the testing phase, the auto-encoder reconstructs original frames with small reconstruction error and forged frames with large reconstruction error. The forged material is then tracked in subsequent frames using GOTURN algorithm. The result of the experiments demonstrates that the proposed detection technique can adequately detect video splicing with an Area Under the Receiver Operating Characteristics (AUROC) value of 0.9307.

Keywords: Passive, Chroma key, Authentication, Frames

INTRODUCTION

The wide-spread of advanced and low cost video cameras and cell phones has caused rapid increase in the amount of digital data being generated every single day. The information provided by the contents of these digital images and videos form the basis of several important and consequential decisions in the fields of criminal or forensic investigations, intelligence services, politics, journalism and legal proceedings in the court of law (Raahat D. Singh & N. Aggarwal, 2017). Meanwhile, with the wide availability of handy and highly powerful media editing tools, it has become much easier to tamper the content of digital media without leaving any visual traces. This has made it challenging to accurately authenticate multimedia content. Digital video forgery can be performed in many ways such as frame insertion or deletion (Li, Zhang, Guo & Wang, 2016), frame duplication (Wu, Jiang, Sun, & Wang, 2014), copy move (Cozzolino, Poggi, & Verdoliva, 2014; Kaur & Kaur, 2016; Ramesh Chand Pandey, Singh, & Shukla, 2014), inpainting (Saxena, Subramanyam, & Ravi, 2016), chroma key forgery (Bagiwa, Abdul Wahab, Idna Idris, Khan, & Choo, 2016; Junyu, Yanru, Yuting, Bo, & Xingang, 2012) and splicing (D'Avino, Cozzolino, Poggi, & Verdoliva, 2017; Singh & Aggarwal, 2017).

Splicing, sometimes called partial manipulation, is a kind of forgery in which frames of two different videos either from the same or different sources are interpolated together to generate a new video (Singh & Aggarwal, 2017). The steps for video splicing are shown in Figure 1.

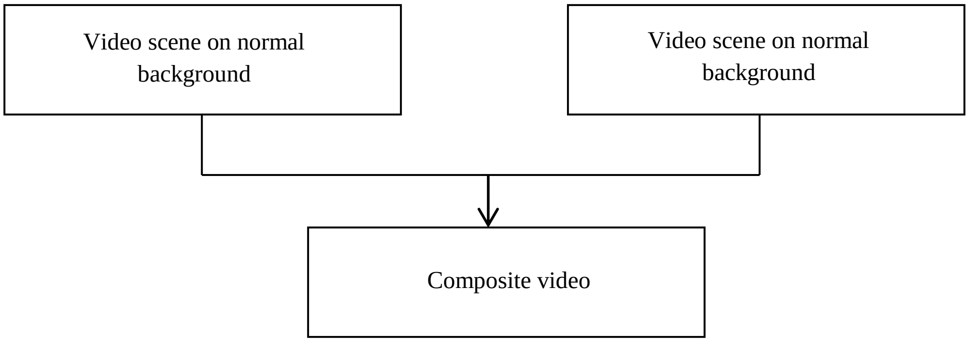
Video scene on normal background
Video scene on normal background
Composite video

Figure 1: Video splicing process (Bagiwa et al., 2016)

In video splicing, regions of different video frames are combined to create new fraudulent video frames. For instance, an object from one frame could be copied and pasted onto another frame, with malicious intention. There are two types of splicing; inter-frame and intra-frame. In inter-frame splicing the source and target frames are different while in intra-frame splicing the source and the target frames are the same. Figure 2 presents examples of video content forged using splicing attack.

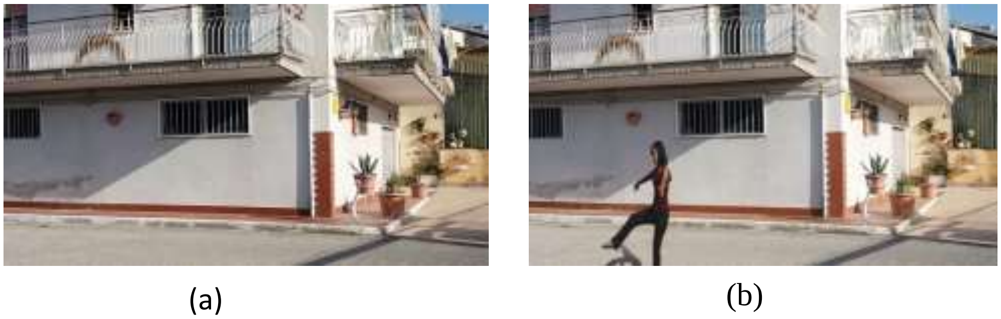

Figure 2: Examples of region-level splicing. (a) Actual scene (b) A girl spliced into the actual scene. (http://www.grip.unina.it/web-download.html)

The authenticity of digital videos has a significant role as it is popularly used as supporting evidence and historical records in various applications that may be related to law enforcements, military, surveillance, insurance claims and commercial purpose (Ramesh C. Pandey, Singh, & Shukla, 2016). Splicing forgery detection techniques such as (D'Avino et al., 2017) and (R. D. Singh & N. Aggarwal, 2017) have been proposed, but the detection accuracy of some such as (D'Avino et al., 2017) decreases when the video is compressed. However, digital video needs to be compressed to take up less space and save money on additional bandwidth charges. Thus, developing an easy-to-use forensic technique that can reliably detect splicing forgery in digital videos which is independent on video compression is important. In this paper, we present a system that can help detect traces of splicing forgery in both compressed and uncompressed digital video.

The rest of this paper is organized as follows. Heading “Related work” reviews the literature on splicing forgery detection in digital videos. Heading “Proposed detection technique” outlines the proposed detection technique using convolutional auto-encoder and GOTURN. The experimental results are reported and discussed in Heading “Experimental results and analysis”. A comparative summary is presented in the Heading “Comparison of the proposed technique with the existing work” and the paper is concluded in Heading “Conclusion”.

Related work

In this section we review recent literature on video forgery detection and localization. We focus especially on techniques that detect splicing and those that detect chroma key forgery because the main difference between splicing and chroma key is that, in splicing forgery, videos with normal background are used for the composition while in chroma key videos with green or blue background are used for the composition (Bagiwa et al., 2016).

Junyu et al. (2012), proposed a technique for detecting and localizing videos composited with blue screen technique using inconsistencies in statistical distributions of quantized discrete cosine transform (DCT) coefficients between foreground and background in the composited videos. The performance of the technique was reported to achieve an average detection accuracy of 88% when fifty composited videos were used. Furthermore, the detection accuracy of the method decreases when the quality of the composited video is worse than the background.

Bagiwa et al. (2016) proposed a technique that uses statistical correlation of blurring artefact extracted from the suspected video to detect and localize chroma key background. The technique has the ability to detect video composition carried out using other composition methods, such as splicing, when the videos used have different blurring qualities. The performance of the technique has been reported to achieve a true positive detection rate of 91.12% and a false positive detection rate of 1.95%. However, the detection accuracy of the technique decreases when the background of the video used for the composition is either blue or green.

D'Avino et al. (2017) proposed a system for detecting and localizing video splicing forgery using residual based features extracted from video frames. The auto-encoder is trained to learn how to reproduce the original input with minimum error by retaining all relevant information in the hidden layer, so that in the existence of spliced areas the reconstruction error increases triggering detection. The auto-encoder was trained using 50 original frames extracted from the video dataset available online at http://www.grip.unina.it/web-download.html and evaluated using 100 frames. The detection accuracy of the technique was presented using receiver operating curves (ROC) of average True positive rate (TPR) and false positive rate (FPR). The result of the technique indicated that it does performed quite well but the performance decreases when the video used is compressed. The number of unrolling steps of the long short-term memory (LSTM) recurrent neural network has no effects on the performance of the system.

Singh andAggarwal (2017) proposed a technique for detecting and localizing of upscale-crop and splicing forgery by examining pixel correlation and noise inconsistency artefacts. Two detectors (Modified-Gallagher and Fractional MG) that expose traces of resampling in digital content by examining pixel-level anomalies are proposed in the technique. The technique can detect the presence of resampling artefacts with an accuracy of over 98%, regardless of the size of the tempered region or scaling factors used to perform resampling of the given content.

In the literature reviewed so far, it can be seen that the detection accuracy of some techniques decreases when compressed videos or videos with low quality are used. Therefore, in this paper, we propose a technique that can detect and localize splicing forgery in digital video with improved detection accuracy when compressed videos are used.

Proposed detection technique

The detection framework of the proposed technique is explained in three main stages of pre-processing, feature extraction, and post-processing as shown in Figure 3

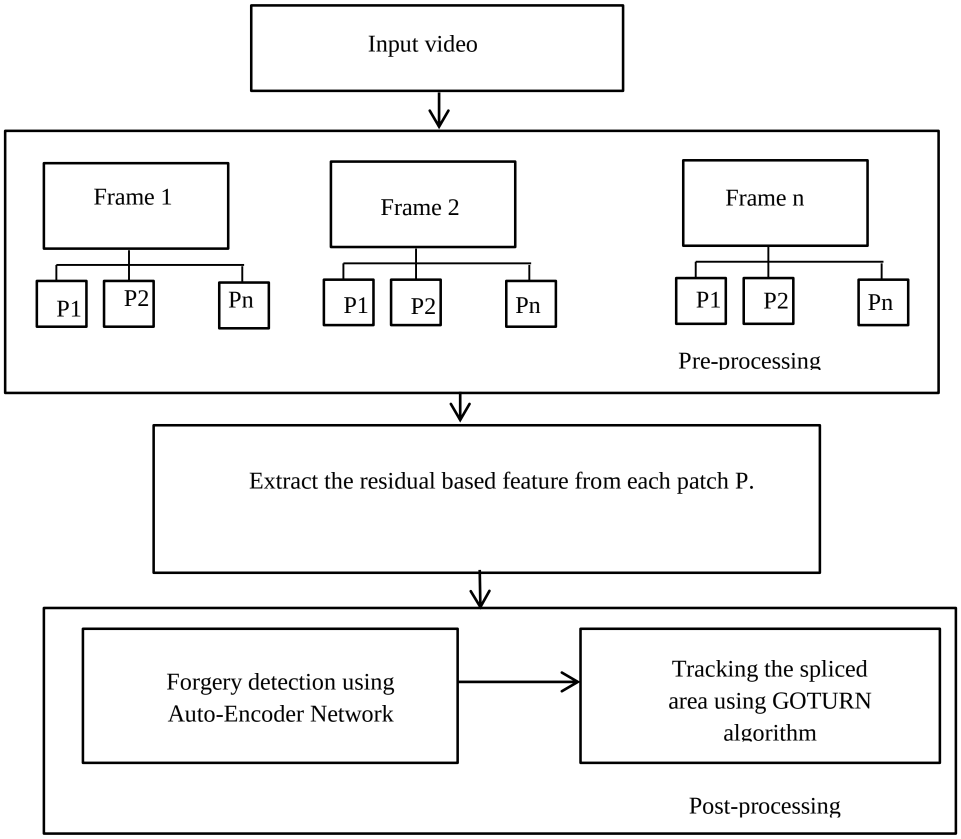
Input video
Frame 1
Frame 2
Frame n
P1
P2
Pn
Pre-processing
Extract the residual based feature from each patch P.
Forgery detection using Auto-Encoder Network
Tracking the spliced area using GOTURN algorithm
Post-processing

Figure 3: proposed splicing forgery detection framework

Pre-processing

In the pre-processing stage, the input video is divided into frames. The frames are then further converted into a grayscale frames. Each frame is then analyzed in sliding window, using image patches of 128 X 128 pixels taken with stride 8. From each patch, a residual based feature is extracted to capture the statistical differences of spliced material with respect to the original video.

Feature extraction

Feature extraction is the process of generating descriptive features to be used in selection and classification tasks. The end result of the extraction task is a set of features, commonly called a feature vector, which constitutes a representation of the image (Choras, 2007).

The residual based feature is extracted based on the following three main steps as used in (D'Avino et al., 2017).

1. Computation of residuals through high-pass filtering

In order to compute the residual image, a linear high-pass filter of the third order is used as shown in equation (1).

$$r_{i,j} = x_{i,j-1} - 3x_{i,j} + 3x_{i,j+1} - x_{i,j+2} \dots (\qquad 1$$

Where x and r are origin and residual images, and i; j indicate spatial coordinates.

From the 3rd order one-dimensional mask M = [1 -3 3 -1] of equation (1), a two dimensional mask (filter) of size 4x4 is developed by (Mary & Begum, 2015) as shown in figure 4. This filter is used for computing image residuals in the proposed technique because it gives better edge map than Laplacian based edge detection algorithms (Mary & Begum, 2015).

$$\begin{matrix} 0 & 0 & -1 & 0 \\ 0 & 0 & 3 & 0 \\ -1 & 3 & -6 & 1 \\ 0 & 0 & 1 & 0 \end{matrix}$$

Figure 4: 4 x 4 Third derivative filter

2. Quantization and truncation of the residuals

In this step, quantization and truncation is performed on the extracted residuals using equation (2):

$$\hat{r}_{i,j} = trunc_T\left(round\left({}^{r_{i,j}}/_{q}\right)\right) \qquad 2$$

where q > 0 is a quantization step, T is the truncation value and $\boldsymbol{\hat{r}_{i,j}}$ is the truncated residuals.

The truncation function with threshold T > 0 is defined by (Jessica & Jan, 2012) as;

$$trunc_T(x) = \begin{cases} x, \; x \in [-T, T] \\ Tsign(x), \quad otherwise \end{cases} \quad \forall x \ni R \quad \dots \qquad 3$$

In the proposed technique, we use T = 2 and q =1 as used in (Cozzolino et al., 2014) to limit the matrix size.

3. Computation of a histogram of co-occurrences

Finally, the row wise and column wise co-occurrence are computed using equation (4) and equation (5) respectively.

$$C = (k_0 + k_1 + k_2 + k_3) = \sum_{i,j} I\left(\hat{r}_{i,j} = k_0, \widehat{r}_{i+1,j} = k_1, \hat{r}_{i+2,j} = k_2, \hat{r}_{i+3,j} = k_3\right) \dots (\qquad 4)$$

$$C = (k_0 + k_1 + k_2 + k_3) = \textstyle\sum_{i,j} I\left(\hat{r}_{i,j} = k_0, \widehat{r}_{i,j+1} = k_1, \hat{r}_{i,j+2} = k_2, \hat{r}_{i,j+3} = k_3\right) \qquad 5$$

The column-wise co-occurrences are then pooled with the row-wise co-occurrences, based on symmetry considerations. The co-occurrences is then converted to a vector form and normalized to zero mean and unit norm to obtain the final feature vector x.

Post-processing

The post-processing stage is divided into two namely; anomaly detection and tracking of the spliced area. Anomaly detection is a technique used to identify unusual patterns that do not conform to expected behaviour, called outliers. It has many applications such as fraud detection, forgery detection, etc. In this paper, the auto-encoder is used to detect anomaly or forgery in the video.

Convolutional auto-encoder (CAE) is trained using feature vectors that are extracted from ten (10) video frames, one from each original video dataset. At the testing phase, the CAE succeeds in reconstructing original frames with small reconstruction error while in the presence of spliced material; the CAE reconstructs it with a large reconstruction error. The detection accuracy of the proposed method is then measured using ROC curve.

If a patch of a frame is forged, a bounding box (bbox) around it is defined and GOTURN algorithm (Held, Thrun, & Savarese, 2016) is used to keep track of it in the remaining video frames. The network architecture of GOTURN is presented in figure 5.

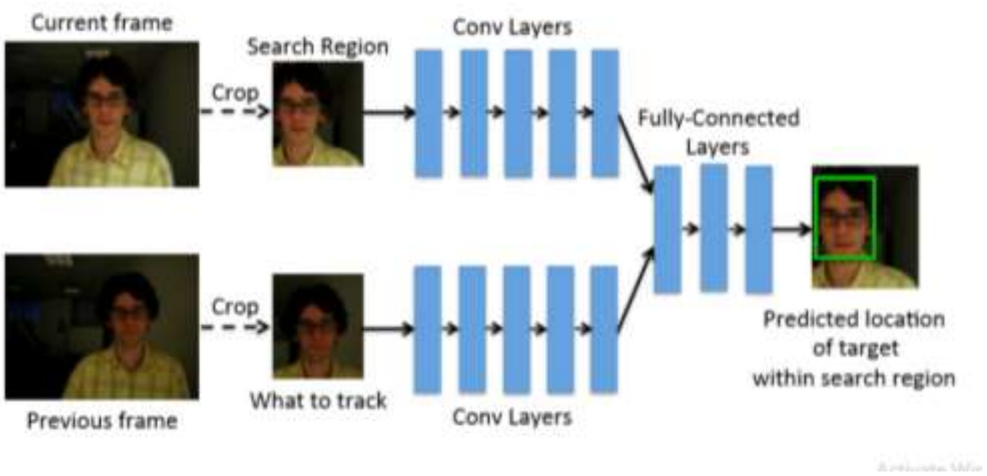

Figure 5: GOTURN network architecture **(**Held *et al*., 2016)

The target object and the search region are inputted into a sequence of convolutional layers. The output of these layers is a set of features that capture a high-level representation of the image. The outputs are then fed through a number of fully connected layers. The fully connected layers compare the features from the target object to the features in the current frame to find where the target object has moved.

Experimental results and analysis

To assess the performance of the proposed video splicing detection technique, experiment was performed on videos obtained from http://www.grip.unina.it/web-download.html created by (D'Avino et al., 2017). The dataset comprises 10 short videos with splicing created using chroma key compositing by means of green screen acquisitions. Table 1 shows some synthetic information on the videos and in particular the total number of frames and the number of forged frames.

Table 1: Characteristics of video dataset

S/N	Name	Number of frames	Number of forged frames	Camera
1	Tank	335	191	Nokia Lumia520
2	Man	399	207	Apple iphone 7
3	Cat	281	136	Huawei p7 mini
4	Helicopter	488	292	Apple iphone 5
5	Hen	373	169	Huawei p9 plus
6	Lion	294	228	Samsung GT 18150
7	Ufo	306	96	Motorolo Moto G
8	Tree	302	240	Huawei p8 lite
9	Girl	371	162	Samsung J5
10	Dog	310	186	Nokia Lumia520

All videos were cropped at the same size of 720×1280 pixels. Figure 6 and Figure 7 show individual frames extracted from five original and five forged videos respectively.

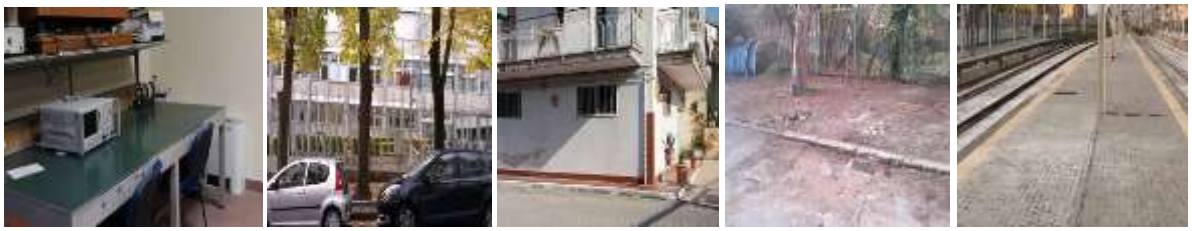

Figure 6: Individual frames extracted from five original videos

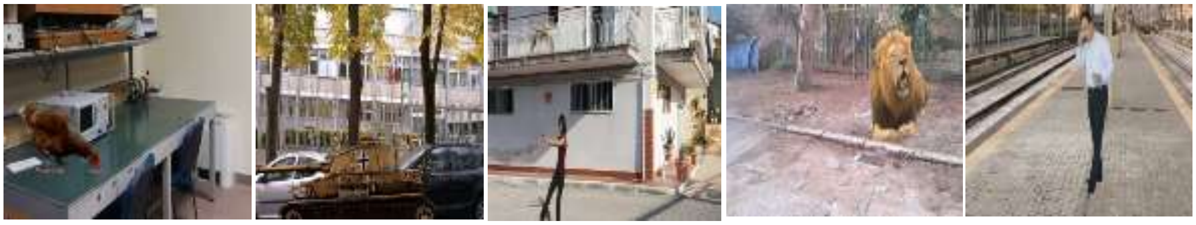

Figure 7: Individual frames extracted from five forged videos

The proposed detection technique is implemented in Tensorflow. If a patch of a frame is found to be forged, a bounding box (bbox) around it is defined as shown in Figure 8.

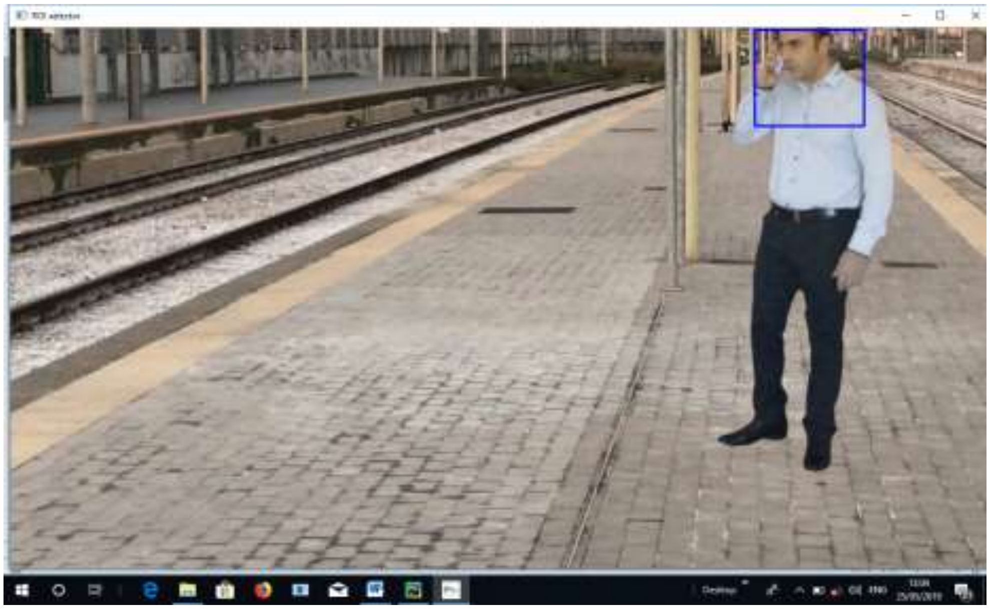

Figure 8: Bounding box around a forged patch

After a bbox is defined around the forged patch, GOTURN is then used to keep track of that patch in the remaining video frames as demonstrated in figure 9.

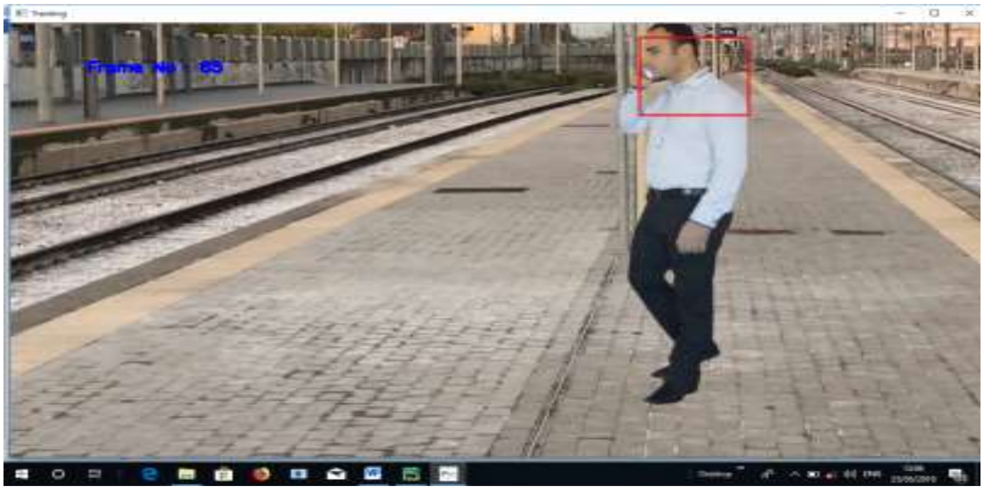

Figure 9: Tracking of the forged patch in frame 67

If the patch is not found in the frame the bbox will disappear as indicated in figure 10.

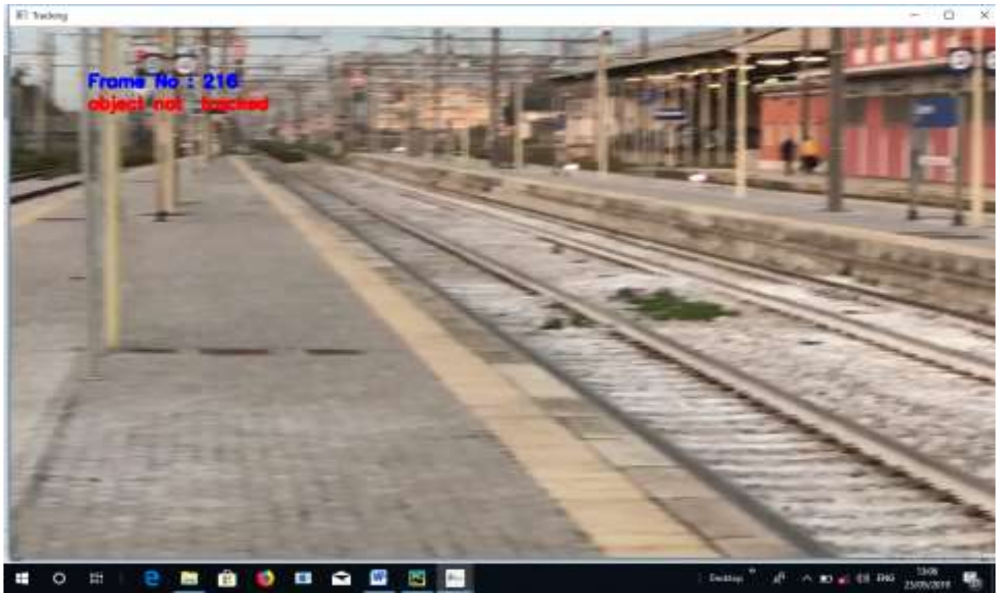

Figure 10: Tracking of forged patch in frame 216 (object not found)

The performance of the proposed technique was measured on 10 frames from each video in the dataset used. For each video, the true positive rates (TPR) and false positive rates (FPR) are computed and the results are summarized in Table 2 and the ROC curve is plotted and presented in Figure 11.

Table 2: Result of Experiments on 10 Test Videos

Video	TPR	FPR
Video 1	0.79	0.05
Video 2	0.92	0.15
Video 3	0.94	0.20
Video 4	0.95	0.30
Video 5	0.97	0.45
Video 6	0.98	0.58
Video 7	0.97	0.65
Video 8	0.99	0.75
Video 9	0.99	0.80
Video 10	0.99	1.0

TPR is the ability of the proposed technique to confirm the presence of forgery and FRP is the ability of the proposed technique to confirm the absence of forgery.

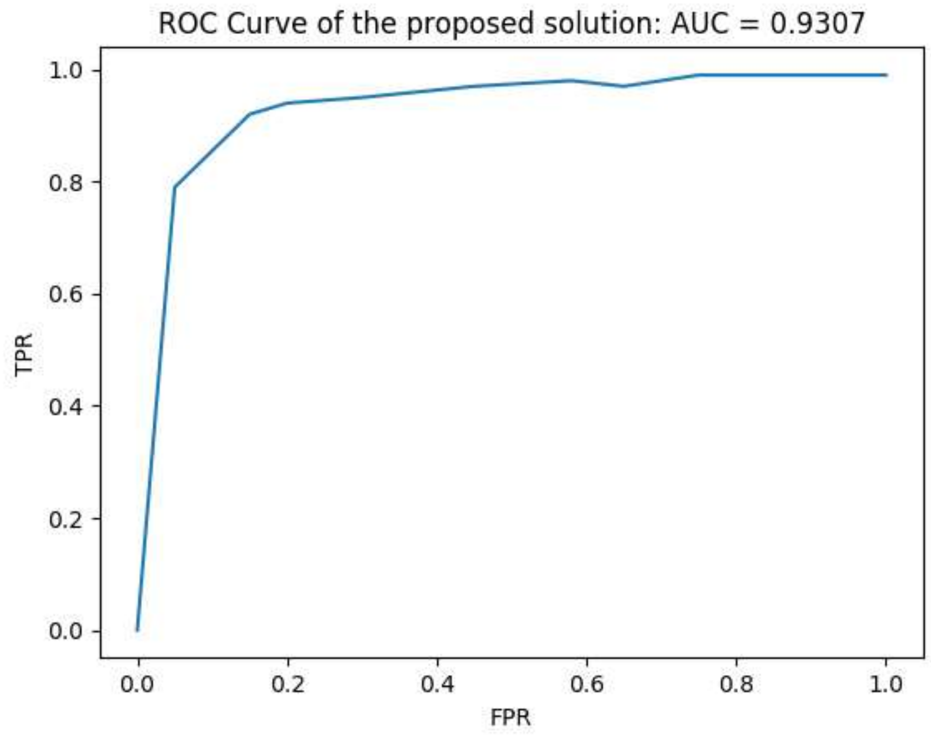

Figure 11: ROC curve of the proposed technique

From Figure 11, the AUROC is 0.9307 indicating that the proposed technique can rank a randomly chosen forged patch higher than a randomly chosen original patch with a probability of 0.9307. Therefore, the results obtained from the experiment conducted have shown that the proposed technique can be effectively used to detect digital video splicing in both compressed and non-compressed videos.

Comparison of the proposed technique with the existing work

In this section, the performance of this technique is presented as compared with existing splicing detection technique proposed in the work of (D'Avino et al., 2017) using the same video dataset. To compare this proposed technique with the work of (D'Avino et al., 2017), ROC curves were plotted and the

AUROC of each technique is calculated. The result obtained from the comparison is presented in Figure 12.

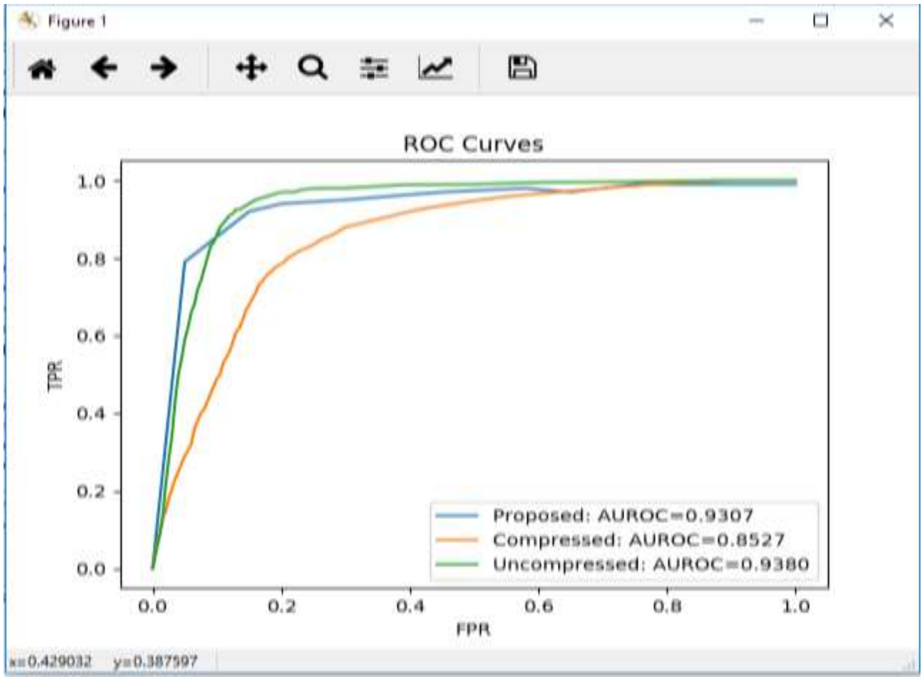

Figure 12: ROC curves of the proposed technique and (D'Avino et al., 2017) for compressed and uncompressed videos

The result of the comparison between the proposed technique and (D'Avino et al., 2017) for video splicing forgery indicated that the proposed technique recorded a higher AUROC value for Compressed videos, but marginally lower than (D'Avino et al., 2017) for Uncompressed videos. The main focus of this research is to increase the accuracy of splicing detection in compressed digital videos.

CONCLUSION

A technique for video splicing forgery detection based on CAE and GOTURN algorithm is proposed in this paper. During the training phase, the CAE learns to reproduce the pristine input, so that in the presence of spliced areas the reconstruction error increases causing detection. When a patch is found to be forged, that patch will be tracked using GOTURN in the remaining video frames. The experimental results presented have demonstrated the efficacy of the proposed method on both compressed and uncompressed videos. The proposed technique recorded 9.15% improvement when compressed video is used and 0.78% decrease in detection accuracy when the video is uncompressed. Further research may focus on increasing the detection accuracy of splicing forgery detection in both compressed and uncompressed digital videos simultaneously using different type of manipulations, feature extraction and post-processing operations.

REFERENCES

Bagiwa, M. A., Abdul Wahab, A. W., Idna Idris, M. Y., Khan, S., & Choo, K.-K. R. (2016). Chroma key background detection for digital video using statistical correlation of blurring artifact. *Digital Investigation*, 29-43.

Choras, R. S. (2007). Image feature extraction techniques and their applications for CBIR and biometrics systems. *INTERNATIONAL JOURNAL OF BIOLOGY AND BIOMEDICAL ENGINEERING, 1*, 1-11.

Cozzolino, D., Poggi, G., & Verdoliva, L. (2014). Copy-move forgery detection based on patch match. *ICIP*, 5247-5251.

D'Avino, D., Cozzolino, D., Poggi, G., & Verdoliva, L. (2017). Autoencoder with recurrent neural networks for video forgery detection. 1-8.

Davide, Cozzolino. (n.d). *GRIP Download* Retrieved from http://www.grip.unina.it/web-download.html

Held, D., Thrun, S., & Savarese, S. (2016). Learning to Track at 100 FPS with Deep Regression Networks. 26.

Jessica, F., & Jan, K. (2012). Rich models for steganalysis of digital images. *IEEE Transactions on Information Forensics and Security, 7*(3), 868–882.

Junyu, X., Yanru, Y., Yuting, S., Bo, D., & Xingang, Y. (2012). Detection of blue screen special effects in videos *International Conference on Medical Physics and Biomedical Engineering*, 1316-1322.

Kaur, R., & Kaur, E. J. (2016). Video forgery detection using hybrid techniques. *International Journal of Advanced Research in Computer and Communication Engineering, 5*, 112-117.

Mary, G. J. J., & Begum, A. R. (2015). Edge detection using third order difference equation: a new dimension *Communications on Applied Electronics (CAE), 1*, 10-14.

Pandey, R. C., Singh, S. K., & Shukla, K. K. (2014). Passive copy- move forgery detection in videos *International conference on computer and communication technology*, 301-306.

Pandey, R. C., Singh, S. K., & Shukla, K. K. (2016). Passive forensics in image and video using noise features: A review. *Digital Investigation*, 1-28.

Saxena, S., Subramanyam, A. V., & Ravi, H. (2016). Video inpainting detection and localization using inconsistencies in optical flow. . *IEEE Region 10 Conference*.

Singh, R. D., & Aggarwal, N. (2017). Detection of upscale-crop and splicing for digital video authentication. *Digital Investigation*, 31-52. doi: doi: 10.1016/j.diin.2017.01.001

Singh, Raahat D., & Aggarwal, N. (2017). Video content authentication techniques: a comprehensive survey 1-30. doi: DOI 10.1007/s00530-017-0538-9

Wu, W., Jiang, X., Sun, T., & Wang, W. (2014). Exposing video inter-frame forgery based on velocity field consistency. *International Conference on Acoustic, Speech and Signal Processing Assurance and Security (ICASSP)*, 2693-2697.