IMAGE SPLICING DETECTION USING CAMERA RESPONSE FUNCTION CONSISTENCY AND AUTOMATIC SEGMENTATION

Yu-Feng Hsu and Shih-Fu Chang

Department of Electrical Engineering
Columbia University
{yfhsu,sfchang}@ee.columbia.edu

ABSTRACT

We propose a fully automatic spliced image detection method based on consistency checking of camera characteristics among different areas in an image. A test image is first segmented into distinct areas. One camera response function (CRF) is estimated from each area using geometric invariants from locally planar irradiance points (LPIPs). To classify a boundary segment between two areas as authentic or spliced, CRF cross fitting scores and area intensity features are computed and fed to SVM-based classifiers. Such segment-level scores are further fused to form the image-level decision. Tests on both the benchmark data set and an unseen high-quality spliced data set reach promising performance levels with 70% precision and 70% recall.

1. INTRODUCTION

With the ease of digital image manipulation, like copy-and-paste (splicing), image forgery has become a critical concern in many applications. Verification of content integrity has become increasingly important. An intuitive and promising approach to image forgery detection is to examine the consistency of inherent physics-based attributes among different parts of a single image. These attributes can be natural scene related, for example, lighting, shadow, and geometry. Or they can be imaging device properties such as camera response function (CRF), demosaicking filter, and sensor noise statistics. Any image that fails to show natural scene consistency and natural imaging consistency may be considered as forgery. Such approaches are passive and thus more general - no active mechanisms are needed to generate and embed additional signatures into images at the sources.

2. PREVIOUS WORK

Studies of physical characteristics of imaging devices, specifically cameras, have received more attention in recent years. [1] presents a typical camera imaging pipeline. Components such as CCD sensor, demosaicking filter, and camera response function (CRF), may possess unique characteristics to camera models or even to camera units. Recovery of such inherent characteristics is useful for image source identification. Some prior works proposed methods for demosaicking filter estimation [2, 3], sensor fixed pattern noise estimation [4], CRF estimation [5, 6], and natural image quality from multiple color channels [7]. Such physical attributes, when estimated from an image, can be used as the "fingerprint" of the capturing source.

Given the fingerprints, splicing detection can readily take place by checking the inconsistency among different image areas. The aforementioned works lead to detection schemes based on demosaicking inconsistency [2, 3], sensor noise [8], CRF abnormality [9], and CRF inconsistency [10].

Spliced image detection by CRF consistency checking has been shown promising in our prior work [10]. However, the application was limited because manual selection of suspicious areas was required. In this paper, we remove this dependency on manual input by incorporating automatic image segmentation and extending the statistical classification framework for splicing detection. In the following, we present the problem formulation, CRF consistency checking scheme, and fusion strategies to obtain image-level decisions from segment-level scores. Performance evaluations over a benchmark data set and a high-quality spliced data set will both be presented, followed by discussions of remaining challenges.

3. PROPOSED APPROACH

Consistency checking is motivated by the fact that spliced images typically contain suspicious areas with distinct device signatures from other areas in the spliced image. To detect such spliced areas automatically, two solutions are needed - automatic image segmentation and robust recovery of camera signatures from each area.

In most scenarios, manual input is not available and thus an automated process is demanded. This brings image segmentation into picture, followed by crucial tasks of device signature extraction and consistency checking. In this paper, we choose a popular segmentation tool, Normalized Cuts [11], though other methods such as Mean Shift [12] may also be considered. For device signature and consistency measure, we use the single channel CRF estimation proposed in [6] and a cross fitting scheme extended from our prior work in [10].

Instead of consistency checking between every possible

pair of segmented areas, we focus on the authenticity test of boundary segments, since splicing not only causes image area inconsistency, but also creates anomalous points near splicing boundaries. Incorporating such points is key to our approach. Given the segmentation results of an image, we propose to first detect suspicious boundary segments, and from there infer if the whole image contains any spliced content. Such an inferencing process may also consider the structural relations among segment boundaries in the image.

An overview of our proposed framework is given in Fig. 1. The test image is first segmented into distinct areas using Normalized Cuts [11], after which CRF estimation is conducted in each of the areas. To check if a boundary segment between two neighboring areas is authentic or spliced, we apply cross fitting between CRF from one area and data samples from another and use these fitting scores to represent a boundary segment along with other features extracted from the two areas. Support Vector Machine (SVM) is learned over these feature vectors to classify the boundary segment as authentic or spliced. The segment-level classification results are then fused to infer whether an image is spliced or authentic, and locate the suspicious areas if it is classified as spliced.

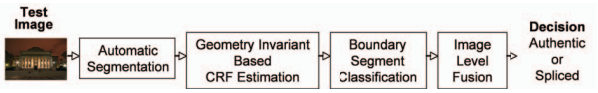


Fig. 1. A system for automatic local spliced area detection

3.1. Image Segmentation

Normalized Cuts (NCuts) requires the number of desired areas to be predetermined, typically set from 2 to 20. In general, over-segmentation is to be avoided so that the resulting areas are reasonably large and the boundaries sufficiently long. One potential drawback, however, is the inability to detect small spliced areas. Fig. 2 shows the results of one image with boundaries plotted in yellow, when setting the number of regions to 8. As mentioned above, we only target at neighboring areas and the boundary segment in between. In Fig. 2, one such segment is displayed in blue and its two sides in red and green. The two sides shall be denoted as area A and area B in the rest of the paper. We also dilate the boundary to form a boundary segment (denoted as area E). The union of areas A, B, E is then denoted as area O.

Each boundary segment is categorized into one of the following types depending on the properties of its two sides. (1) Authentic: Both sides (area A and B defined above) are from the same camera; thus the segment under consideration is authentic. (2) Spliced: Both areas are untampered but are from different cameras. In this case, the segment is a hit on the splicing boundary. (3) Ambiguous: The actual splicing boundary is close to or partially overlapped with the automatic segment, and thus cuts through one or both neighboring areas. In other words, the automatic boundary segment is a partial hit and one or both of the neighboring areas contain content from two cameras.

From the spliced image detection point of view, there is no need to distinguish Spliced from Ambiguous cases since they both indicate the presence of the splicing operation. However, at the segment level, due to the ill-defined authenticity of ambiguous segments, we will only use the Authentic and Spliced classes to train a detector in our learning-based method.

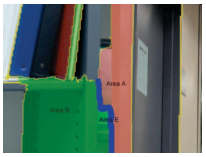


Fig. 2. Sample segmentation results by NCuts. A thickened boundary segment is shown to indicate the potential splicing boundary.

3.2. Camera Response Function Estimation

The next step is to extract camera-specific features from the segmented regions and then verify their consistency. The camera response function (CRF) nonlinearly transforms CCD sensor signals, or irradiance, to the output recorded by digital memory, or brightness. As different cameras transform the irradiance signals differently, CRF is a natural camera signature. It is often denoted as $R = f(r)$, where r is the irradiance and R the brightness. The simplest form of CRF is a gamma function, $R = f(r) = r^{\alpha_0}$. More complex models usually achieve better approximation. In this paper, we use the Generalized Gamma Curve Model (GGCM) proposed in [6], extending the exponent to a polynomial function of r: $R = f(r) = r^{\sum_{i=0}^{n} \alpha_i r^i}$. Considering the tradeoff between modeling power and computational complexity, we choose a first order GGCM, $n = 1$.

In [6], locally planar irradiance points (LPIPs) are used to extract information solely related to the CRF. Namely, if a point has a locally planar irradiance geometry, $r = ax + by + c$, then the second order partial derivatives in the irradiance domain would all be zero, and the following equation holds:

$$\frac{R_{xx}}{R_x^2} = \frac{R_{xy}}{R_x r_y} = \frac{R_{yy}}{R_y^2} = \frac{f''(r)}{(f'(r))^2} = \frac{f''(f^{-1}(R))}{(f'(f^{-1}(R)))^2} \quad (1)$$

This quantity, denoted as $A(R)$, does not carry any information about the geometry of r: $\{a, b, c\}$, rather, it contains information only for estimating CRF (f). With further manipulation we get another quantity $Q(R)$,

$$Q(R) = \frac{1}{1 - A(R)R} \quad (2)$$

which is also independent of irradiance geometry. It is hence termed Geometry Invariant (GI). With a first order GGCM, $Q(R)$ is related to the parameters of CRF by

$$Q(R) = \frac{(\alpha_1 r \ln(r) + \alpha_1 r + \alpha_0)^2}{\alpha_0 - \alpha_1 r} \quad (3)$$

The CRF is estimated by first extracting LPIPs, computing the GIs, and iteratively looking for the optimal GGCM parameters to fit the computed GI values. In [6], we have shown that with improved derivative computation, Bayesian LPIP selection, flattened error metrics, and cross channel similarity, the CRF estimation method achieves an excellent accuracy - the average RMSE as low as 0.0224. Note this method may be reliably applied to a single color channel as well as images of multiple color channels.

3.3. CRF Consistency Measure via Cross Fitting

To check if a boundary segment is authentic or spliced, we compute cross fitting errors using the estimated CRFs and GIs (i.e. (Q, R) values) of the selected LPIPs:

$$\mathbf{s}_{ij}=\{(Q_i(R)_n-\frac{(\alpha_{1,j}r_n\ln(r_n)+\alpha_{1,j}r_n+\alpha_{O,j})^2}{\alpha_{O,j}-\alpha_{1,j}r_n})^2|n\leq N_i\};i,j\in\{A,B\}$$

$$\mathbf{s}_{k}=\{(Q_k(R)_n-\frac{(\alpha_{1,k}r_n\ln(r_n)+\alpha_{1,k}r_n+\alpha_{O,k})^2}{\alpha_{O,k}-\alpha_{1,k}r_n})^2|n\leq N_k\};k\in\{O,E\}$$

where N_i is the total number of LPIPs from area i.

If areas A and B are indeed from different cameras, we expect to see large cross-fittting errors $\mathbf{s}_{ij}$'s. Anomalous distributions of (Q, R) samples from areas E and O are also expected since they are not from a single camera. We compute the first- and second-order moments of these cross-fitting errors, resulting in the first set of twelve features of a segment:

$$\mathcal{F}_{fitting,1}=[\mu(\mathbf{s}_{AA}),\mu(\mathbf{s}_{BB}),\mu(\mathbf{s}_{AB}),\mu(\mathbf{s}_{BA}),\mu(\mathbf{s}_{E}),\mu(\mathbf{s}_{O})];$$

$$\mathcal{F}_{fitting,2}=[\sigma(\mathbf{s}_{AA}),\sigma(\mathbf{s}_{BB}),\sigma(\mathbf{s}_{AB}),\sigma(\mathbf{s}_{BA}),\sigma(\mathbf{s}_{E}),\sigma(\mathbf{s}_{O})]$$

Our experiments in [6] showed that images with a large coverage of R values usually result in much more accurate CRFs. Therefore we add the averages and the ranges of R's from each area as a second feature set:

$$\mathcal{F}_{R}=[\mu(\mathbf{R}_{A}),\mu(\mathbf{R}_{B}),\mu(\mathbf{R}_{E}),\mu(\mathbf{R}_{O});\Delta\mathbf{R}_{A},\Delta\mathbf{R}_{B},\Delta\mathbf{R}_{E},\Delta\mathbf{R}_{O}]$$

Hence each segment is represented by the combined 20-dimensional feature vector described above.

3.4. SVM Classification

Before reaching image level decisions, we first classify the segments as authentic or spliced using SVM bagging. Since authentic segments outnumber spliced ones, as shown in Table 1, we divide the whole authentic pool into P subsets, each with a similar number of samples as spliced ones. We then train P classifiers out of these evenly populated sample bags. At the test stage, every test segment gets P classification labels and P distance values to the decision boundaries, $[l_p, d_p]$, $p = 0 \dots P-1$. These distances are fused with a sigmoid warping function to obtain the final decision:

$$l_{fuse} = sign(\frac{1}{P}\sum_{p=0}^{P-1}\frac{1}{1+\exp^{(-d_p/\omega)}} - 0.5) \quad (4)$$

In our experiment, P is set to 5, and ω is determined empirically through cross-validation. The decision threshold can be changed to obtain different operation points in the precision-recall curve as shown in Fig. 3(a).

3.5. Image Level Fusion

To get a global decision for the image, naively averaging over all d_p's would not be appropriate, since an image with only one spliced segment is certainly spliced, but its single positive d_p will vanish if it is to be averaged with other segment scores.

Currently we adopt a simple method - the splicing score of the image equals the maximal segment-level splicing score contained in the image. This score is then thresholded in order to classify an image to be authentic or spliced. Varying the threshold will result in different operation points in the precision-recall curve shown in Fig 3(b). More sophisticated fusion strategies may consider the structural relationships among boundary segments to detect a spliced object, instead of isolated suspicious segments.

Table 1. Numbers of test segments and images.

Segments			Images	
Auth	Splc	Amb	Auth	Splc
828	219	249	84	89

4. EXPERIMENTS AND DISCUSSIONS

4.1. Data Sets

Our benchmark data set consists of 363 uncompressed images [10]: 183 authentic and 180 spliced. Authentic images are taken with four cameras: Canon G3, Nikon D70, Canon EOS 350D, and Kodak DCS330. Each spliced image has content from exactly two cameras, with visually salient objects from one image (eg. a yellow rubber duck) copied and pasted onto another image using Adobe Photoshop without any post processing. We also made best efforts to ensure content diversity. This data set will be referred to as the Basic data set.

Base on initial empirical tests, we set the number of segmentation areas to be 8. Output boundary segments from NCuts are categorized into three sets: Authentic, Spliced, and Ambiguous, using the definitions in Sec. 3.1. The breakdown is reported in Table 1.

Ambiguous segments are excluded from our segment-level experiment, trimming down the number of boundary segments within each image to 7$\sim$10. In image-level tests, however, they are included to examine the effect of partial alignment.

Both segment level and image level performances are evaluated over the Basic data set. Standard validation procedures are used to randomly partition the data set into training and test sets. The partitioning is done at the image level so that segments from the same image will not be included in both training and test sets. In order to see how well our classifier generalizes, we tested our detector on 21 authentic images and 38 high-quality spliced images using some advanced image manipulation tools recently developed in Microsoft Research Asia. These images are typically JPEG compressed, with advanced matting or color adjustment in addition to copy-and-paste, therefore it is a much more realistic and challenging set for splicing detection. We denote this test set as the Advanced data set. Note we tested our method over the new advanced spliced images without re-training the SVMs in order to evaluate the generalization capability of the method.

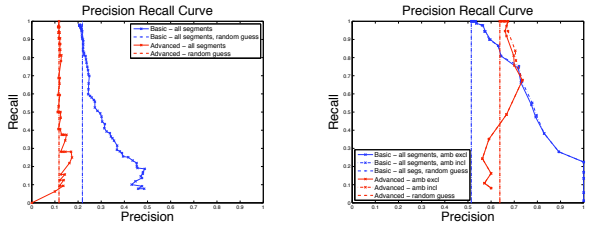
4.2. Results

As shown in Fig. 3, although the segment-level classification accuracy is only slightly better than random guess (Fig. 3(a)), our simple fusion scheme for image-level classification proves to be powerful - 70% precision, 70% recall (Fig. 3(b)). The curves when excluding and including ambiguous segments are almost identical, meaning that such ill-defined instances do not play a major role in image level decisions.

When the detector is tested over the unseen Advanced data set, we observe performance decrease at the segment level as anticipated: the PR curve is almost only as good as random guess (Fig. 3(a)). Despite this degradation, at the image level, a precision-recall of 70% and 70% can still be obtained, comparable to that of the Basic data set (Fig. 3(b)).

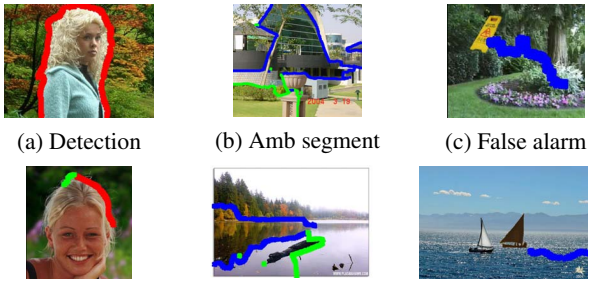
This is encouraging in that it promises satisfactory detection even when the trained detector is applied to new images that are subject to significantly different splicing processes and additional post-processing operations (e.g., compression).

Analysis of the detection results gives rise to some interesting observations. Among the correctly detected spliced images from the Advanced data set, about one third achieve both accurate spliced area segmentation and classification. One third are detected by classifying some ambiguous (i.e., partially aligned boundaries) as spliced. The remaining one third are detected as spliced due to falsely classifying some authentic segments as spliced. Two sets of example image are shown in Fig. 4. Among these images, the following observations can also be made: spliced images with a large object, eg. human face or body, are more likely to get both precise segmentation and correct detection, even when post-processing is present (Fig. 4(a)(d)). Images with similar objects and backgrounds tend to suffer from inaccurate segmentation. However in some of these cases the resultant ambiguous segments can still be useful, as shown in Fig. 4(b)(e). Lastly, images with complex textures (eg. grass and tree in Fig. 4(c) and lake reflections in Fig. 4(f)) are prone to false alarms.



(a) Segment level (b) Image level

Fig. 3. PR curves for segment- and image-level classification. The vertical lines show the results corresponding to random guess.



(a) Detection (b) Amb segment (c) False alarm

(d) Detection (e) Amb segment (f) False alarm

Fig. 4. Three types of detected image in the Advanced spliced image data set. Red denotes successfully detected spliced segments, green denotes ambiguous segments detected as spliced, and blue denotes authentic segments detected as spliced.

5. CONCLUSION

We proposed a novel spliced image detection approach based on camera response function estimation, consistency checking and image segmentation. The method is fully passive and automatic - neither active signature embedding nor manual input is needed. To the best of our knowledge, this is the first work combining automatic image segmentation with camera signature consistency checking. The proposed method is tested over two data sets, created with basic or advanced splicing techniques. Results show promising performance in detecting spliced images in both data sets, with about 70% in precision and 70% in recall.

6. ACKNOWLEDGEMENT

This work has been sponsored by NSF Cyber Trust program under award IIS-04-30258. The authors thank Tian-tsong Ng for sharing codes and helpful discussions and Junfeng He for access to the Advanced spliced image data set from MSRA.

7. REFERENCES

[1] Y. Tsin, V. Ramesh, and T. Kanade, "Statistical calibration of the CCD imaging process," in *IEEE International Conference on Computer Vision*, 2001, pp. 480–487.

[2] A.C. Popescu and H. Farid, "Exposing digital forgeries in color filter array interpolated images," *IEEE Transactions on Signal Processing*, vol. 53, no. 10, pp. 3948–3959, 2005.

[3] A. Swaminathan, M. Wu, and K.J.R. Liu, "Component forensics of digital cameras: A non-intrusive approach," in *Conference on Information Sciences and Systems*, 2006.

[4] J. Lukáš, J. Fridrich, and M. Goljan, "Determining digital image origin using sensor imperfections," *Proceedings of the SPIE*, vol. 5685, 2005.

[5] S. Lin, J. Gu, S. Yamazaki, and H.-Y. Shum, "Radiometric calibration from a single image," in *IEEE International Conference on Computer Vision and Pattern Recognition*, 2004.

[6] T.-T. Ng, S-.F. Chang, and M.-P. Tsui, "Using geometry invariants for camera response function estimation," in *IEEE International Conference on Computer Vision and Pattern Recognition*, 2007.

[7] M. Kharrazi, H. T. Sencar, and N. D. Memon, "Blind source camera identification.," in *International Conference on Image Processing*, 2004, pp. 709–712.

[8] J. Lukáš, J. Fridrich, and M. Goljan, "Detecting Digital Image Forgeries Using Sensor Pattern Noise," *Proceedings of the SPIE*, vol. 6072, 2006.

[9] Z. Lin, R. Wang, X. Tang, and H.-Y. Shum, "Detecting doctored images using camera response normality and consistency," in *IEEE International Conference on Computer Vision and Pattern Recognition*, 2005, pp. 1087–1092.

[10] Y.-F. Hsu and S.-F. Chang, "Detecting image splicing using geometry invariants and camera characteristics consistency," in *International Conference on Multimedia and Expo*, 2006.

[11] Jianbo Shi and Jitendra Malik, "Normalized cuts and image segmentation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, no. 8, pp. 888–905, 2000.

[12] C. Yang, R. Duraiswami, NA Gumerov, and L. Davis, "Improved fast gauss transform and efficient kernel density estimation," *Computer Vision, 2003. Proceedings. Ninth IEEE International Conference on*, pp. 664–671, 2003.