Refining PRNU-Based Detection of Image Forgeries

Xufeng Lin and Chang-Tsun Li
Department of Computer Science
University of Warwick
Coventry, CV4 7AL, UK

***Abstract*—Photo Response Non-Uniformity (PRNU) noise can be considered as a spread-spectrum watermark embedded in every image taken by the source imaging device. It has been effectively used for localizing the forgeries in digital images. The noise residual extracted from the image in question is compared with the reference PRNU in a sliding-window based manner. If their normalized cross correlation, which servers as a decision statistic, is below a pre-determined threshold (e.g., by Neyman-Pearson criterion), the center pixel in the window is declared as forged. However, the decision statistic is calculated over the forged and the non-forged regions when the sliding window falls near the boundary of the two different regions. As a result, the corresponding pixels of the forged region are probably wrongly identified as genuine ones. To alleviate this problem, we analyze the correlation distribution in the problematic region and refine the detection by weighting the decision threshold based on the altered correlation distribution. The effectiveness of the proposed refining algorithm is confirmed through the results of detecting three different kinds of realistic image forgeries.**

I. Introduction

Detecting image forgeries is an interesting while very challenging task due to the variety of image manipulations a user can perform with increasingly powerful image editing softwares. Active techniques, such as digital watermarking, are effective in verifying the authenticity of an image, but the requirement of originally embedding into the protected image limits their widespread use in practice. Therefore, there has been growing interest in passive techniques. As one of the most promising passive techniques, Photo Response Non-Uniformity (PRNU) noise arises primarily from the manufacturing imperfections and the inhomogeneity of silicon wafers. It can be considered as an intrinsic watermark embedded in every image captured by the source device. Its uniqueness to individual camera and stability against environmental conditions make PRNU noise a powerful and robust tool for exposing image forgery [1]–[6]. Its capability of detecting image forgeries irrespective of the specific type of forgery arouses wide attention of the researchers in the field of digital forensics. In [1], a PRNU-based technique was proposed either for detecting the image forgeries in user-selected region or automatically identifying forged image regions. It was then refined in [2], where both the false acceptance rate (FAR) (i.e., misidentifying a tampered block as non-tampered) and the false rejection rate (FRR) (i.e., wrongly labeling non-tampered blocks as tampered) are considered. Chierchia et al. proposed several improvements concerning either the better estimation of PRNU noise [7] or the higher resolution of the detection results [6], [8].

Since PRNU pattern, by its very nature, is a very weak noise-like signal, its reliable detection requires jointly processing a large number of pixel samples, e.g., in a sliding-window manner. As pointed out in [3], [6], when the sliding window falls near the boundary of the tampered and the non-tampered regions, the decision statistic becomes a weighted average of two different contributions and probably leads to a high FAR. This problem can be alleviated by means of hard [3] or soft [6] image segmentation to obtain the boundary information before detection. These two algorithms share the same essence of making use of the extra structure information of the image content, but their drawbacks are twofold: Firstly, the detection result heavily depends on the quality of image segmentation or the pilot image [9], but an accurate image segmentation or high-quality pilot image is not easy to obtain. The second and most critical drawback is that they become helpless in detecting some *occlusive* forgeries, where objects in the original scene are hidden by placing a homogeneous background on them, e.g., a person is masked by a grass patch or an airplane is covered by a sky patch.

This work also aims at improving the resolution of PRNU-based image forgery detection, but it approaches the issue from a different perspective. Starting from an initial detection, we model how the decision statistic changes as the sliding window moves across the boundary of two different regions (i.e., tampered and non-tampered) and adjust the decision threshold accordingly to achieve a more satisfactory detection. In what follows, Section II revisit the background of detecting image forgeries based on PRNU noise, Section III presents the proposed algorithm and Section IV validates the proposed algorithm by detecting realistic image forgeries. Finally, Section V concludes this work.

II. Background

In this section, we will revisit the algorithm proposed in [2]. Let noise residual $w \in \mathbb{R}^N$ be the difference of the observed image $g \in \mathbb{R}^N$ and its denoised version $\hat{f} = \mathcal{F}(g)$:

$$\begin{aligned} w &= g - \hat{f} \\ &= (1+k)f + \theta - \hat{f} \\ &= gk + (f-g)k + (f-\hat{f}) + \theta \\ &= gk + v, \end{aligned} \tag{1}$$

where k is the PRNU noise, f is the noise-free image, θ is an additive noise accounting for all other interferences and v is the sum of θ and the two additive terms introduced by

the denoising filter [10]. When an image region is tampered with, the PRNU signal in the noise residual w of that region is lost. Therefore, image forgeries can be exposed by identifying the image regions where PRNU signal is absent. Like in [2], we formulate the problem of detecting PRNU signal in noise residual w as a binary hypothesis test

$$\begin{cases} H_0 : w = v \\ H_1 : w = z + v, \end{cases} \tag{2}$$

where $z = gk$ is the signal of interest (also called the reference PRNU) and v is PRNU-irrelevant noise. For a target pixel q_i, a decision statistic ρ_i is calculated based on the normalized cross correlation (NCC) between w_{N_i} and z_{N_i}:

$$\rho_i = \frac{\sum_{j\in N_i}(w_j - \bar{w})(z_j - \bar{z})}{\sqrt{\sum_{j\in N_i}(w_j - \bar{w})^2}\sqrt{\sum_{j\in N_i}(z_j - \bar{z})^2}}, \tag{3}$$

where N_i is the pixel indices within the $n \times n$ sliding window centered at q_i. To reveal the forgery, ρ_i is then compared with a threshold γ_1:

$$\hat{u}_i = \begin{cases} 1, \rho_i < \gamma_1 \\ 0, \text{otherwise} \end{cases} \tag{4}$$

where $\hat{u}_i \in \{0, 1\}$ is a binary value indicating the forgery (1 for forgery and 0 for genuine pixel). γ_1 is usually selected according to the Neyman-Person criterion to ensure a small false acceptance rate (FAR), i.e., $Pr(\hat{u}_i = 0|u_i = 1)$, with $u_i \in \{0, 1\}$ the ground truth. However, even for the non-forged pixels, the NCC coefficients might happen to be very low in the image areas of dark, saturated or highly textured. Based on the relationship between the correlations and the local image features, this problem is addressed in [2] by estimating correlation distribution $p(x|H_1)$ under hypothesis H_1 and correcting the tampered pixels for which the false rejection rate (misidentifying non-tampered as tampered) higher than a threshold γ_2, i.e.,

$$\int_{-\infty}^{\gamma_1} p(x|H_1)dx > \gamma_2, \tag{5}$$

to non-tampered.

We would like to spend more words on the estimation of the correlation distribution under hypothesis H_0 and H_1 (i.e., $p(x|H_0)$ and $p(x|H_1)$), respectively. It was observed in our experiments that if the reference PRNU (i.e., z in Equation (2)), is preprocessed by the Wiener Filtering in DFT domain [2] or our recently proposed spectrum equalization algorithm [11], $p(x|H_0)$ fits quite well with the Gaussian distribution $\mathcal{N}(0, 1/d)$, where $d = n \times n$ is the number of pixels within the square sliding window. For the estimation of $p(x|H_1)$, we use the Gaussian model like in [5] rather than the generalized Gaussian model in [2] due to its simplicity and effectiveness.

III. Proposed Method

We assume that the d-dimensional signal within the sliding window, either for the estimated z_{N_i} or w_{N_i}, is standardized to have zero mean and unit variance, which means each element, z_j or w_j ($j \in N_i$), is independently drawn from the identical normal distribution $\mathcal{N}(0, 1)$. Presumably, each element in the standardized signal can be modeled as the sum of the true PRNU signal and other irrelevant interferences:

$$\begin{cases} w_j = x_j + \alpha_j \\ z_j = y_j + \beta_j, \end{cases} \tag{6}$$

where x_j follows a Gaussian distribution $\mathcal{N}(0, \sigma^2)$ and α_j conforms to $\mathcal{N}(0, 1-\sigma^2)$. Likewise, $y_j \sim \mathcal{N}(0, \lambda\sigma^2)$ and $\beta_j \sim \mathcal{N}(0, 1-\lambda\sigma^2)$. Here, σ^2 and $\lambda\sigma^2$ can be viewed as the quality of the true PRNU signal in w_{N_i} and z_{N_i}, respectively. Note that λ accounts for the different qualities of the PRNU signal in w_{N_i} and z_{N_i}. With the standardized signal, the decision statistic ρ_i in Equation (3) is simplified as

$$\rho_i = \frac{1}{d}\sum_{j\in N_i}(x_j y_j + \alpha_j y_j + \beta_j x_j + \alpha_j \beta_j). \tag{7}$$

If x and y are from two different cameras (i.e., under hypothesis H_0), using the Central Limit Theorem (CLT), ρ_i follows a Gaussian distribution $\mathcal{N}(\mu_0, \Sigma_0)$, where $\mu_0 = 0$ and $\Sigma_0 = 1/d$. While under hypothesis H_1, we have $y_i = \sqrt{\lambda}x_i$. Therefore, Equation (7) can be rewritten as

$$\rho_i = \frac{1}{d}\sum_{j\in N_i}(\sqrt{\lambda}x_j^2 + \sqrt{\lambda}\alpha_j x_j + \beta_j x_j + \alpha_j \beta_j). \tag{8}$$

It is known that x_j^2/σ^2 follows the Chi-square distribution with 1 degree of freedom, $\chi^2(1)$. So based on the assumption that x_j, α_j and β_j are mutually independent, we can easily arrive at

$$\rho_i \sim \mathcal{N}(\mu_1, \Sigma_1), \tag{9}$$

where

$$\begin{cases} \mu_1 = \sqrt{\lambda}\sigma^2 \\ \Sigma_1 = (1 + \lambda\sigma^4)/d. \end{cases} \tag{10}$$

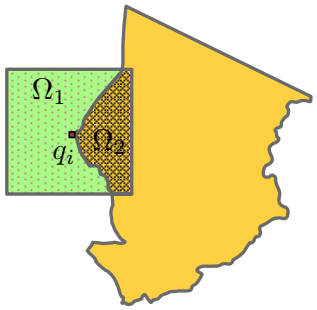


Fig. 1: The square sliding window across the non-tampered region Ω_1 and the tampered region Ω_2.

Equation (9) is the decision statistic distribution if the sliding window falls completely on the non-tampered region. To see how the problematic region violates the distribution, let us look at Fig. 1, which shows a sliding window across the non-tampered region Ω_1 (in green background) and the tampered region Ω_2 (in yellow background). The decision

statistic therefore becomes a weighted average of two different contributions:

$$\rho_i = \frac{1}{n_1}\sum_{j\in\Omega_1}(\sqrt{\lambda}x_j^2+\sqrt{\lambda}\alpha_j x_j+\beta_j x_j+\alpha_j\beta_j) + \frac{1}{n_2}\sum_{j\in\Omega_2}(x_j y_j+\alpha_j y_j+\beta_j x_j+\alpha_j\beta_j), \quad (11)$$

where n_1 and n_2 are the number of pixels in Ω_1 and Ω_2, respectively. x_j and y_j are independent in the second summation term for region Ω_2. Therefore,

$$\rho_i \sim \mathcal{N}(\mu,\Sigma) \quad (12)$$

where

$$\begin{cases}\mu = \sqrt{\lambda}\sigma^2 n_1/d\\ \Sigma = 1/d+\lambda n_1\sigma^4/d^2,\end{cases} \quad (13)$$

To drop the unknown λ and σ^2, we finally rewrite Equation (13) as the form of weighting μ_1 and Σ_1

$$\begin{cases}\mu = n_1\mu_1/d\\ \Sigma = (n_1 d\Sigma_1+d-n_1)/d^2,\end{cases} \quad (14)$$

where $d = n_1+n_2$. This is the final expression of the distribution of decision statistic in the boundary region. As shown in Fig. 2, if $n_1 = 0$, which means the sliding window entirely falls in the tampered region, the decision statistic ρ_i conforms to $\mathcal{N}(\mu_0,\Sigma_0)$. As the sliding window moves away and completely falls in the non-tampered region, ρ_i follows the distribution $\mathcal{N}(\mu_1,\Sigma_1)$. After analyzing the correlation distribution in the problematic area near the boundary of the tampered and the non-tampered regions, we therefore propose an algorithm to alleviate the miss detection problem as follows:

1) Calculate the correlation ρ_i between the noise residual w_{N_i} and the estimated reference PRNU signal z_{N_i} within the sliding window N_i centered at pixel q_i.
2) Estimate the expected correlation $\bar{\rho}_i$ and the variance $\sigma^2_{H_1}$ of the NCC coefficients under hypothesis H_1 using the correlation predictor proposed in [2];
3) Select two thresholds γ_1 and γ_2 to obtain an initial detection result $\hat{u}_i$ using Equation (4) and (5).
4) Obtain the number n_i of pixels in the sliding window centered at pixel q_i that belong to the non-tampered region by convoluting a $n\times n$ matrix of ones with the initial detection result.
5) Calculate a new threshold γ_1^i for each pixel by solving the following equation

$$\frac{1}{\sigma_i}\int_{-\infty}^{\gamma_1^i} e^{-\frac{(t-\mu_i)^2}{2\sigma_i^2}}dt = \int_{-\infty}^{\gamma_1} e^{-\frac{t^2}{2}}dt, \quad (15)$$

where

$$\begin{cases}\mu_i = \bar{\rho}_i n_i/d\\ \sigma_i^2 = (\sigma^2_{H_1} d n_i + d - n_i)/d^2.\end{cases} \quad (16)$$

The purpose of Equation (15) is to guarantee the same desired FAR in the boundary region as in other regions according to the altered distribution as formulated in Equation (14).

6) Label a pixel q_i as tampered ($\hat{u}_i = 1$) if $\rho_i < \gamma_1^i$.
7) Label a tampered pixel ($\hat{u}_i = 1$) as non-tampered if

$$\frac{1}{\sqrt{2\pi}\sigma_{H_1}}\int_{-\infty}^{\gamma_1} e^{-\frac{(t-\bar{\rho}_i)^2}{2\sigma^2_{H_1}}}dt > \gamma_2. \quad (17)$$

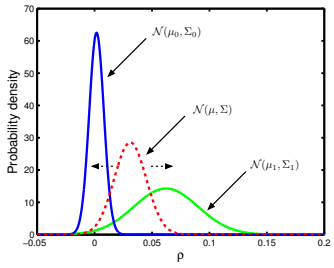


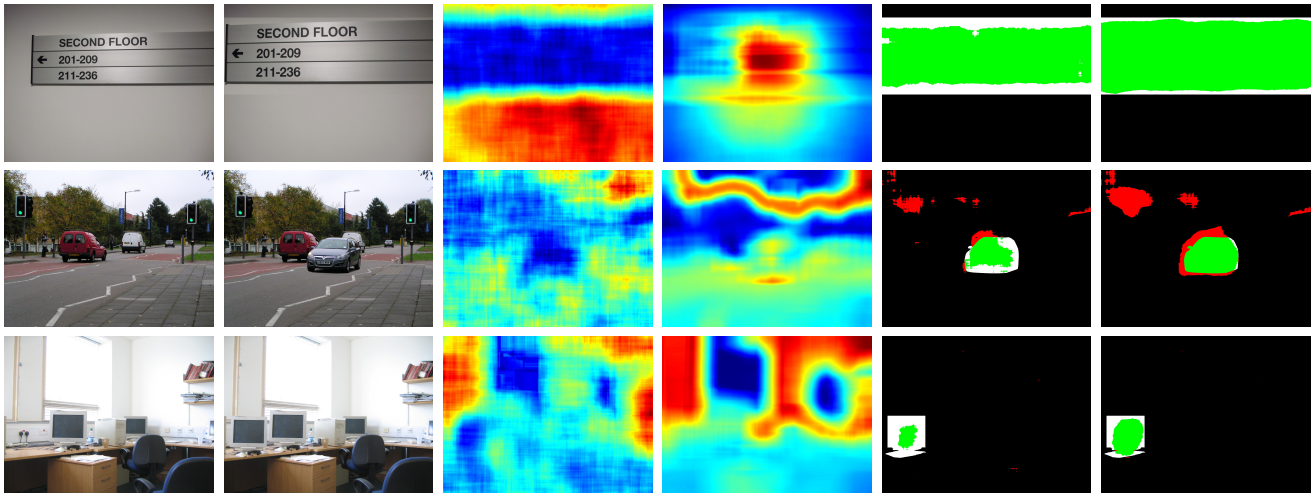
Fig. 2: How the correlation distribution changes as the sliding window moves across the boundary region.

IV. Experiments

In this section, we will report some preliminary experiments meant to support the idea of weighting the threshold accordingly in the boundary regions to improve the detection resolution. Our experiments were carried out on three cameras, a Cannon IXY500, a Olympus C730UZ and a FujiFilm FinePix A920. BM3D [12] denoising algorithm was used to extract the noise residual from images. 50 blue sky images from each camera were used to estimate the reference PRNU noise, which is further preprocessed by the algorithm proposed in [11] to reduce the false positives and make the correlation distribution under hypothesis H_0 fit better to the theoretical distribution $\mathcal{N}(0,1/d)$. Another 20 natural images were randomly selected for each camera to train the correlation predictor proposed in [2]. All images taken by the three cameras have the same size of 1536×2048 pixels. Additionally, as shown in Fig. 3, three different kinds of forgeries were involved:

- Scaling forgery (image 1): A direction board in a image taken by FujiFilm FinePix A920 is enlarged.
- Cut-and-paste forgery (image 2): A car in a image is cut and pasted onto another image. The two images are both taken by Olympus C730UZ.
- Copy-and-move forgery (image 3): A computer in a image taken by Cannon IXY500 is copy and move to a new location in the same image.

We attempt to compare the proposed algorithm, which weights threshold controlling the FAR in the boundary region, with the method based on the constant false acceptance rate (CFAR) decision rule in [2]. From left to right, Figure 3 shows (a) original image, (b) forged image, (c) predicted correlation field, (d) actual correlation field, (e) detection result by CFAR and (f) refined detection result based on (e). The forged area is highlighted in white, the correctly detected area is



(a) (b) (c) (d) (e) (f)

Fig. 3: Forgery detection results for scaling (the first row), copy-and-move (the second row) and cut-and-paste (the third row) forgery. (a) Original image, (b) forged image, (c) predicted correlation field, (d) actual correlation field, (e) detection result by CFAR and (f) refined detection result by our proposed algorithm.

highlighted in green, while the area falsely labeled as tampered is highlighted in red. From top to down, the detection results are shown for the three kinds of image forgeries, i.e., scaling, cut-and-paste and copy-and-move forgery, respectively. The size of sliding window is set to 256×256 pixels. The two thresholds γ_1 and γ_2 are set to 0.01 and 0.05, respectively. It is worth mentioning that we did not apply any morphological operation.

The first image is a simple case, where an enlarged direction board is placed on a smooth and bright wall. It is not surprising to see both the proposed algorithm and CFAR can accurately detect the forged area. But the proposed refining algorithm clearly does a better job in the upper and the bottom boundary of the forged direction board without introducing any false positives. The second row of Fig 3 shows the detection of the cut-and-paste forgery. As can be seen, the false positives are hard to avoid due to the dark area of the traffic lights and the complex background, e.g., trees and grass. In spite of introducing slightly more false positives, most of the forged car is reliably detected. Similar result can be observed in detecting smaller tampered area, as shown in the third row of Fig. 3. The refined detection result reveals part of the forgery in the stand of the monitor, which is completely ignored by CFAR.

By revealing more possible forgeries in the boundary areas, the refined detection result apparently fits more closely to the actual shape of the tampered area, which can potentially provide the forensic investigator with more detailed and suggestive information. Like in [5], we show the upper envelope of $(\mathrm{FPR}, \mathrm{TPR})$ points in Fig. 4. What can be seen in Fig. 4 is that the superiority of the refining algorithm seems more evident in the more challenging detection tasks. The refining algorithm is slightly better than CFAR in the simplest case, while the equal error rate (EER) increases from 91.7% to 95.3% for image 2 and from 81.8% to 93.8% for image 3.

V. Conclusion

In this work, we have proposed a refining scheme for PRNU-based detection of image forgeries. We model the correlation distribution near the boundary across the tampered and the non-tampered regions and weight the threshold accordingly to achieve the desired false acceptance rate. Despite some possible false positives introduced (e.g., γ_1 is set too small), the overall better performance has been verified in the task of detecting three different kinds of realistic image forgeries. We believe that this work will facilitate forensic investigators to get a more accurate and informative detection result.

Acknowledgment

This work is the outcome of the EU project, Computer Vision Enabled Multimedia Forensics and People Identification (Project no. 690907; Acronym: IDENTITY), funded through the EU Horizon 2020 - Marie Skodowska-Curie Actions - Research and Innovation Staff Exchange action.

References

[1] J. Lukáš, J. Fridrich, and M. Goljan, "Detecting digital image forgeries using sensor pattern noise," in *Proc. SPIE*. International Society for Optics and Photonics, 2006, pp. 362–372.

[2] M. Chen, J. Fridrich, M. Goljan, and J. Lukás, "Determining image origin and integrity using sensor noise," *IEEE Trans. Inf. Forensics Security*, vol. 3, no. 1, pp. 74–90, 2008.

[3] G. Chierchia, S. Parrilli, G. Poggi, L. Verdoliva, and C. Sansone, "PRNU-based detection of small-size image forgeries," *Proc. IEEE Int. Conf. Digital Signal Process.*, pp. 1–6, 2011.

[4] C.-T. Li and Y. Li, "Color-decoupled photo response non-uniformity for digital image forensics," *IEEE Trans. Circuits Syst. for Video Technol.*, vol. 22, no. 2, pp. 260–271, 2012.

[5] G. Chierchia, G. Poggi, C. Sansone, and L. Verdoliva, "A Bayesian-MRF Approach for PRNU-Based Image Forgery Detection," *IEEE Trans. Inf. Forensics Security*, vol. 9, no. 4, pp. 554–567, April 2014.

[6] G. Chierchia, D. Cozzolino, G. Poggi, C. Sansone, and L. Verdoliva, "Guided filtering for prnu-based localization of small-size image forgeries," in *Proc. IEEE Int. Conf. Acoustics, Speech, Signal Process.* IEEE, 2014, pp. 6231–6235.

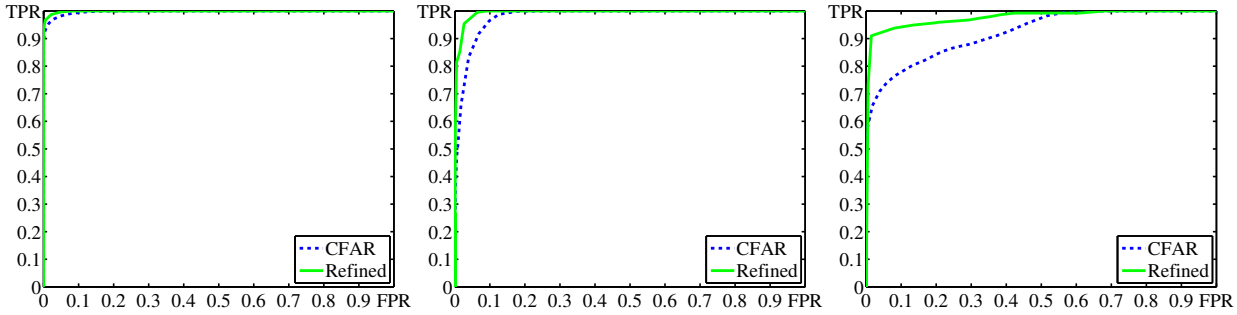


Fig. 4: ROC curves for (a) image 1, (b) image 2 and (c) image 3.

[7] G. Chierchia, S. Parrilli, G. Poggi, C. Sansone, and L. Verdoliva, "On the influence of denoising in prnu based forgery detection," in *Proc. the 2nd ACM workshop on Multimedia in forensics, security and intell.* ACM, 2010, pp. 117–122.

[8] G. Chierchia, S. Parrilli, G. Poggi, L. Verdoliva, and C. Sansone, "PRNU-based detection of small-size image forgeries," in *Proc. IEEE Int. Conf. Digital Signal Process.*, July 2011, pp. 1–6.

[9] K. He, J. Sun, and X. Tang, "Guided image filtering," in *Proc. European Conf. Computer Vision*. Springer, 2010, pp. 1–14.

[10] X. Lin and C.-T. Li, "Enhancing sensor pattern noise via filtering distortion removal," *IEEE Signal Process. Lett.*, vol. 23, no. 3, pp. 381–385, March 2016.

[11] X. Lin and C.-T. Li, "Preprocessing Reference Sensor Pattern Noise via Spectrum Equalization," *IEEE Trans. Inf. Forensics Security*, vol. 11, no. 1, pp. 126–140, 2016.

[12] K. Dabov, A. Foi, V. Katkovnik, and K. Egiazarian, "Image denoising by sparse 3-d transform-domain collaborative filtering," *IEEE Trans. Image Process.*, vol. 16, no. 8, pp. 2080–2095, Aug 2007.