IMPROVED WATERMARK DETECTION ROBUST TO CAMCORDER CAPTURE BASED ON QUADRANGLE ESTIMATION

Min-Jeong Lee\textsuperscript{a}, Kyung-Su Kim\textsuperscript{a}, Young-Ho Suh\textsuperscript{b}, and Heung-Kyu Lee\textsuperscript{a}

\textsuperscript{a}Korea Advanced Institute of Science and Technology  
Department of Electrical Engineering and Computer Science  
\textsuperscript{b}Convergent Contents Research Department, Contents Research Division  
Electronics and Telecommunications Research Institute  
Republic of Korea

ABSTRACT

In this paper, we propose a blind watermark detection scheme which is robust against composite geometric distortions. The composite geometric distortions mean that an image or a video is rotated, perspective-projected, and resized at the same time. It commonly happens during capturing a movie in a theater by a camcorder illegally. Local auto-correlation function (LACF) with a novel estimation model is utilized for calculating composite geometric distortions. We demonstrate robustness of the presented scheme against composite geometric distortions for digital cinema.

Index Terms— Robust watermarking, local auto-correlation function, composite geometric distortions, camcorder capture.

1. INTRODUCTION

Many pirate copies of new digital movies have been found on Internet before their official release. These copies filmed a movie with a camcorder at various angles according to its location of the pirate. They are translated, rotated, scaled, and projected during camcording, therefore, even though watermarking technology is employed to protect copyright of digital cinema contents, it is vulnerable to those distortions. Thus, the watermark should survive after all kinds of geometric distortions including complex distortions in digital cinema application.

Several papers described watermarking for digital cinema. Leest et al. [1] proposed a watermarking scheme which exploited the temporal axis to embed watermark by changing the luminance value of each frame and hence was inherently robust to geometrical distortions. However, since a flickering effect from the luminance change between frames is noticeable by the human eyes, the viewing quality is not satisfactory. Delannay et al. [2] investigated the restoration of geometrically distorted images occurred by the camera acquisition angles. The use of the unmodified contents for detection made their method to be impractical. Moreover, the complex task of time-synchronization between the original movie and the captured one was required. Lubin et al. [3] embedded watermark into low spatial-temporal frequency domain for invisible, robust, and secure watermarking. To determine spatial-temporal regions of video sequences in the embedding procedure, a vision model-based masking computation was employed, but not satisfied the real-time requirement. Our previous work [4] presented a blind watermarking scheme for digital cinema using local auto-correlation function (LACF) to resist to single projective transform. Its estimation model was designed for the projections with only one vanishing point such as vertical projection or horizontal projection. (See Fig. 1(b) and 1(c)). However, projective transform with more than one vanishing point may occur during camcorder capture in practice. The rectangular video frames are transformed into quadrangles through camcorder capture as shown in Fig. 1(d).

In this paper, we propose a robust watermarking scheme which is robust against composite geometric distortions during camcorder capture. The scheme is performed without an original video during the detection. Local auto-correlation function (LACF) with a novel mathematical model is utilized for estimating composite geometric distortions. The rest of paper is organized as follows. In Sec. 2, we introduce watermark design for our watermark detection scheme. In Sec. 3, we describe watermark detection scheme based on LACF. In Sec. 4, we present experiments to demonstrate the effectiveness of the proposed scheme on practical camcorder capture attack and Sec. 5 concludes.

2. WATERMARK DESIGN

The watermark pattern in the presented scheme is used in two ways. One is to estimate geometric distortions and recover from distortions, and the other is to extract the embedded message. In order to accomplish both roles, the watermark pattern should have period-

![Examples of projective distortions.](image-url)
icity for LACF to calculate geometric distortions. First of all, the basic pattern is generated using a secret key and a message and consists of 2-D rectangular pattern. Then the basic pattern is repeated to vertical axis and horizontal axis to get the periodicity. After the periodic watermark pattern \( W \) is obtained, the pattern is embedded in an additive spread-spectrum way with perceptual scaling.

3. WATERMARK DETECTION

Since a blind detector is used, an approximation of the embedded watermark is obtained by employing Wiener filtering as denoising filter. Subtracting the denoised frame from the captured frame, we obtain an approximate version of the embedded watermark pattern. Both estimating geometric distortion and extracting watermark are proceeded using this extracted pattern.

In watermark detector, the detection process performs as follows: 1) find geometric distortions using LACF and the estimation model and 2) recover the watermark from the distortions and extract the embedded message.

3.1. Estimating Geometric Distortion

As a result of camcorder capture, the shapes of captured movie scenes in the rectangular frames are generally quadrangles. A rectangle is transformed into a quadrangle by perspective projection. Let \( x = (x_1, x_2, x_3)^T \) be the homogeneous point coordinates of the original frame and \( x' = (x_1', x_2', x_3')^T \) be the homogeneous point coordinates of the geometrically distorted frame. The projective transformation is a linear transformation on homogeneous 3-vectors represented by a non-singular \( 3 \times 3 \) matrix [5]:

\[
    x' = Hx, \text{ where } H = \begin{pmatrix} h_{00} & h_{01} & h_{02} \\ h_{10} & h_{11} & h_{12} \\ h_{20} & h_{21} & h_{22} \end{pmatrix} \tag{1}
\]

Note that \( H \) is a homogeneous matrix, since as in the homogeneous representation of a point, only the ratio of the matrix elements is significant. There are eight independent ratios among the nine elements of \( H \), and it follows that a projective transformation has eight degrees of freedom (DOF). Thus, four pairs of point-to-point correspondence in the original and distorted frames are required to determine eight DOF. In this paper, four corner points of the original video frame are selected as the original points and four corner points of the distorted video frame are chosen as the corresponding distorted points. Since both the embedder and the detector know the coordinates of four original points, it only needs to know those of four distorted points.

3.1.1. Local Auto-Correlation Function(LACF)

In the previous work [4], local auto-correlation function (LACF) was employed for estimating projective transform. It computed the auto-correlation function of two local areas of the image that are parallel to each other instead of computing auto-correlation of the whole image. Since the estimation model used only two parallel area for LACF, the performance of estimation was limited to the cases of projections with only one vanishing point. However, the camcorder capture raises projections with more than one vanishing point which \( R_{b1} \) and \( R_{b2} \) are needed for LACF. The width and the height of the video frame denote as \( M \) and \( N \) respectively. The LACF on the estimated watermark pattern \( W' \) is modeled as

\[
    lacf_R(x, y) = \sum_{i=-w_R+1}^{w_R-1} \sum_{j=-h_R+1}^{h_R-1} W' (x + d_x(R) + \frac{i}{2}, y + d_y(R) + \frac{j}{2})^2 \tag{2}
\]

where \( w_R \) and \( h_R \) are the width and the height of the region \( R \), \( d_x(R) \) and \( d_y(R) \) are the distance of \( x \)-axis and \( y \)-axis from the upper-left corner point for selecting region \( R \) for LACF. The distance and the size of regions are adaptively selected by the size of the used basic pattern and lower bounds of projective distortion in a practical point of view. The calculation of LACF is accelerated by FFT-based fast equation as follows:

\[
    lacf_R = \frac{\text{IFFT}(\text{FFT}(R) \cdot \text{FFT}(R^*))}{|R|^2} \tag{3}
\]

where “*”-operator denotes complex conjugation. The LACF yields multiple periodic peaks since the periodic watermark pattern was embedded. These local auto-correlation peak (LACP) is detected by applying an adaptive threshold as

\[
    lacp > \mu_{lacf} + \alpha_{lacf} \sigma_{lacf} \tag{4}
\]

where \( \mu_{lacf} \) and \( \sigma_{lacf} \) denote the average and standard deviation of the LACF, respectively. \( \alpha_{lacf} \) is a value which is related to the false positive error rate. Presetting the maximum false positive error rate, we calculate \( \alpha_{lacf} \) and obtain the threshold. Figure 3 shows the...
LACF results of the watermarked image undergone composite projective distortion. Four LACF results show the different angles and intervals, however, the angles and intervals between LACPs in the same LACF result are equal to each other. Based on these four pairs of angles and intervals, we construct the coordinates of the distorted frame.

3.1.2. Calculating Coordinates

Now, it needs to calculate the coordinates using the intervals $\delta_1, \delta_2, \delta_3, \delta_4$ and the angles $\theta_1, \theta_2, \theta_3, \theta_4$ between LACPs. Note that at least the transform description of four points is needed to obtain eight DOF of coefficient matrix of projective transform in Eq. 1. Figure 4 shows the process for obtaining the coordinates of the corresponding distorted points. The original points $P_1, P_2, P_3,$ and $P_4$ are known to the detector. First, the relative coordinates are calculated from each results of LACF. It assumes that the left-upper corner point $P_1'$ is located at $(0, 0)^T$. Then $P_2', P_3'$, and $P_4'$ are calculated by

$$P_2' = \left( m \delta v_1 \cos \theta v_1, m \delta v_1 \sin \theta v_1 \right)^T$$  \hspace{1cm} (5)

$$P_3' = \left( m \delta v_2 \cos \theta v_2, m \delta v_2 \sin \theta v_2 \right)^T$$  \hspace{1cm} (6)

$$P_4' = \left( m \delta v_3 \cos \theta v_3, m \delta v_3 \sin \theta v_3 \right)^T$$  \hspace{1cm} (7)

where $m$ and $n$ are horizontal and vertical repetition times of basic watermark pattern. Next, the quadrangle $P_1'P_2'P_3'P_4'$ is translated to a center of the video frame. Under the assumption that the pirate wants to record the movie to have good visual quality as possible, the entire scene locates in the middle of the frame during camcording. So the barycentric coordinates of quadrangle $P_1'P_2'P_3'P_4'$ are fit onto the center of the video frame. The translation parameters $t = (\tau_x, \tau_y)^T$ are determined as the distance between the center coordinates of the original frame and barycentric coordinates of quadrangle $P_1'P_2'P_3'P_4'$:

$$t = (\tau_x, \tau_y)^T = \left( \frac{M}{2}, \frac{N}{2} \right)^T - \frac{1}{4} \sum P_i', \ i \in \{1, 2, 3, 4\}$$  \hspace{1cm} (8)

Then the translation form of the quadrangle $P_1'P_2'P_3'P_4'$ is given by

$$\begin{pmatrix} P_i' \\ 1 \end{pmatrix} = \begin{pmatrix} 1 & t \\ 0 & 1 \end{pmatrix} \begin{pmatrix} P_i \\ 1 \end{pmatrix}, \ i \in \{1, 2, 3, 4\}$$  \hspace{1cm} (9)

where $I$ is an $2 \times 2$ identity matrix. The corner points of the translated quadrangle $P_2', P_3', P_4'$, and $P_4'$ are then used to compute the projective matrix $H$ in Eq. 1.

3.1.3. Computing Projective Matrix

Now we compute the projective matrix of projective transform using estimated pairs of four points and Eq. 1. Let the inhomogeneous coordinates of a pair of matching point $x$ and $x'$ be $(x, y)$ and $(x', y')$ respectively. The projective transformation of Eq. 1 can be written in inhomogeneous form as

$$x' = \frac{h_{00}x + h_{01}y + h_{02}}{h_{20}x + h_{21}y + h_{22}}, \ y' = \frac{h_{10}x + h_{11}y + h_{12}}{h_{20}x + h_{21}y + h_{22}}$$  \hspace{1cm} (10)

Four correspondences of points $P_i$ and $P_i'$ $(i \in \{1, 2, 3, 4\})$ lead to eight linear equations in the entries of $H$. Since no three points are collinear, they are sufficient to solve for $H$ up to an insignificant multiplicative factor. The inverse matrix transformation $H^{-1}$ is finally computed.

3.2. Extracting Watermark

The watermark extraction is performed by normalized cross correlation. That is, the cross correlation is computed by correlating the restored watermark to the basic patterns. If the resulting correlation value exceeds a preset threshold, the hidden messages are correctly extracted. Experimentally, we determine the preset threshold depending on an error probability model which follows Gamma distribution model, since we take the maximum value from cross correlation [6].

4. EXPERIMENTAL RESULTS

On HD-resolution clips of digital cinema, the fidelity and robustness of watermark detection are measured against camcorder capture attack. A 40 bits payload was embedded into each five-minutes clip to adhere to digital cinema initiatives [7]. The 2D basic pattern whose size is $120 \times 120$ is tiled sixteen times $(n = 16)$ to horizontal axis and nine times $(n = 9)$ to vertical axis. Thus, the watermark pattern is formed $1920 \times 1080$ dimensions and embedded in the entire frame. The parameters in Eq. 2 are set for both horizontal and vertical LACF. For horizontal LACF, $\omega_h$ is set to 120 ($= M/m$) and $h_{h}$ is set to 1080 ($= N$). The $d_4(R_{h1})$ for region $R_{h1}$ is set to 180 and the $d_4(R_{h2})$ for region $R_{h2}$ is set to 1500. Both $d_4(R_{h1})$ and $d_4(R_{h2})$ are set to zero. For vertical projection, $\omega_v$ is set to 1920 ($= M$) and $h_v$ is set to 120 ($= N/n$). The $d_4(R_{v1})$ is set to 180 and the $d_4(R_{v2})$ is set to 660. Both $d_4(R_{v1})$ and $d_4(R_{v2})$ are set to zero. $\alpha_{\text{false}}$ in Eq. 4 is set to 6.0 with presetting the maximum false positive error rate as $10^{-9}$. The average PSNR value is 46.05 dB.

4.1. Fidelity testing

Fidelity testing was performed as described in [3]. Clips were displayed using an EPSON EMP-TW1000 projector onto a wide screen.
Table 1. Results of the robustness against camcorder capture attacks in the proposed watermarking system.

<table>
<thead>
<tr>
<th>Case</th>
<th>Snapshot</th>
<th>Contour of distortion</th>
<th>Correlation</th>
<th>Previous work [4]</th>
<th>Our scheme</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td>0.06</td>
<td>0.35</td>
</tr>
<tr>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td>0.03</td>
<td>0.17</td>
</tr>
<tr>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td>0.06</td>
<td>0.21</td>
</tr>
</tbody>
</table>

Projected clips were about 2.20m and 1.24m in horizontal and vertical directions, respectively. Four expert observers participated in a two alternative forced choice experiment in which each trial consisted of two presentations of the same clip, once with and once without the watermark present. Observers viewed the screen from two picture heights and requested to indicate which clips contained the watermark. Each source clip was played in four times such trials which lasted five minutes. No observer could determine the identity of the watermarked clip surely in any case.

4.2. Robustness testing

To measure the robustness, experiments have been performed against camcorder capture in practice, which include projective transform as well as signal processing distortions. Clips are projected on a screen with same environment in Sec. 4.1 and captured with a SONY HDR-FX1 camcorder tripod-mounted subsequently. Four locations which have respective angles and distances from the screen are selected for capturing. In addition to projective distortion, the captured videos suffered affine distortion and signal processing including aspect ratio change, contrast enhancement, and gamma correction. The performance of our scheme is compared to the previous work [4]. Table 1 shows the results as normalized correlation. Using our scheme, the 40 bits of the embedded watermark are successfully extracted satisfying false positive error probability of $10^{-9}$ in all cases while our previous scheme are not. The presented scheme proves robustness to camcorder capture attack including composite projective distortions.

5. CONCLUSION

Camcorder capture attack is common in digital cinema by capturing with a camera. Since this attack includes composite geometric distortions such as perspective projection and RST. In this paper, we propose a blind watermark detection scheme which is robust against composite geometric distortions. We exploited LACF and constructed a mathematical model for estimating composite geometric distortions. Experimental results showed that the presented scheme is robust to composite geometric distortions for digital cinema.

Acknowledgement

This research is supported by Ministry of Culture, Sports and Tourism(MCST) and Korea Culture Content Agency(KOCCA) in the Culture Technology(CT) Research & Development Program 2009 and the IT R&D program of MKE/MCST/IITA. [2009-S-017-01, Development of user-centric contents protection and distribution technology]

6. REFERENCES