A New Video Watermarking Scheme Resistant to Collusion and Synchronization Attacks

Ki-Jung Kim
Department of Information and Communication Engineering ,
Shingyeong University, Hwaseong City, Kyungki Do, Korea

ABSTRACT

A new video watermarking scheme with robustness against collusion and synchronization attacks is presented. We propose to embed only a few copies of the watermark along the temporal axis into frames, which are located at the borders of each two different plotlines of the video. As a result, each change of the video plotline is transformed into pulse, which is used for watermark embedding and extraction. In addition, since we embed a watermark only into a small number of frames, the distortions of the video are reduced to minimum. Experimental results show the robustness of the proposed scheme.

Keywords*: Video watermarking, geometric attacks.*

1. INTRODUCTION

The rapidly growing digital multimedia market has revealed an urgent requirement for development of effective copyright protection schemes. In this context, digital watermarking has been proposed to protect the copyright of multimedia data by inserting secret signal into them. The fundamental description of common watermarking techniques is given in the Cox's paper [1]. Though digital watermarking has been mostly devoted to audio and image signals, watermarking of other types of data is now being investigated and digital video is one of those new objects of interest. Currently, watermarking digital video is mostly considered as watermarking a sequence of still images [2]. In terms of this approach, two major embedding schemes are employed: either the same watermark is embedded in each video frame, or different watermarks are inserted into all video frames.

The class of malicious attacks that apply in this case is known as frame collusion and two types of linear collusion attacks are usually defined: Type I collusion attack and Type II collusion attack [3]. Type I collusion attack arises when dissimilar frames over the entire video are repeatedly embedded, via linear combination, with the same watermark. Assuming independent frames and the same watermark in all of them, the idea is to estimate this watermark from the non-similar frames over the whole video [3-5]. The estimation can then be normalized in some way and subtracted from every frame in order to cancel the watermark. The main advantage of this attack is its simplicity. However, as it is noted in [4], the energy of the frame is much higher than the energy of the watermark, and a simple averaging of frames will not provide a good estimate of the watermark. Type II collusion attack takes place when big numbers of visually similar frames are embedded, via linear combination, with different watermarks [3-6]. This attack is based on the fact that most consecutive video frames are very similar, especially for static scenes. Conversely, it is assumed that the watermark data are not correlated between frames. This attack can be seen as a temporal low-pass filtering of the watermarked video using a sliding averaging window. This is also very simple attack, but the problem with applying this attack is that it may exert a considerable blurring effect on the video [4].

Malicious attacks also include temporal synchronization attacks, such as frame dropping and insertion of arbitrary frames in order to prevent watermark correct detection [4,6]. They are especially dangerous for schemes when different watermarks are inserted in all frames because even a simple frame drop or insertion succeeds in confusing synchronization.

In this paper, we concentrate our attention mainly on resiliency to collusion attacks that are especially applicable to watermarks embedded into video production. Furthermore, we also consider such temporal synchronization attacks as frame dropping and insertion of arbitrary frames.

2. WATERMARK EMBEDDING

The first thing in embedding is to find the locations to put the watermarks in. For this purpose we propose to use the observation that each plotline of the movie is characterized by its own mean gray level. Suppose that the mean gray level of the n^{th} video frame is defined by

* *Corresponding author. E-mail : kjkim36@empal.com*
Manuscript received Apr. 21, 2009 ; accepted Jun. 21, 2009

$$f_n = \frac{1}{D_1 D_2} \sum_{i=1}^{D_1} \sum_{j=1}^{D_2} s_{ij}^n \qquad (n = 1,..., T)$$

where s_{ij}^n is the gray level of the i^{th}, j^{th} pixel of the n^{th} frame and D_1, D_2 are its dimensions, T is the total number of the video frames.

Then the mean gray level P_m (see Fig. 1) of the m^{th} plotline can be written as

$$P_m = \frac{1}{R_m} \sum_{n=q_m+1}^{Q_m} f_n, \; q_m = \sum_{k=0}^{m-1} R_k,$$

$$Q_m = q_m + R_m, \; (m = 1,..., M)$$

where R_m is the total number of frames within the m^{th} plotline, M is the total number of plotlines.

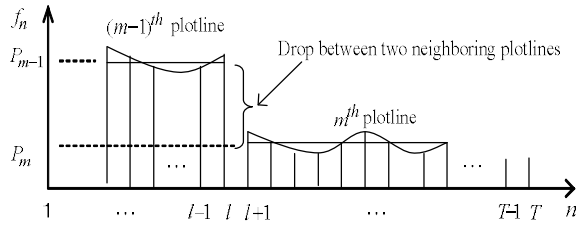

Fig. 1 The dependence of the mean gray level of the plotline with respect to its number

In order to contrast one plotline from another we propose to take an absolute difference

$$d_l' = \left| f_{l+1} - f_l \right| \qquad (l = 1,..., T-1) \qquad (1)$$

and compare its values with some threshold h. If d_l' exceed threshold h, then the value may be considered as pulses, corresponding to the very outsets of plotlines.

Figure 2 illustrates the m^{th} plotline, which begins with such pulse. So for watermark embedding within one plotline we propose to use frames located right after this pulse.

Let F_l^m be the l^{th} frame of the m^{th} plotline. Then the second step is to replace the sequence of frames F_{l+2}^m, … , $F_{l+N_f}^m$ with the same watermark-free frame F_{l+1}^m as shown in Fig. 2.

Finally the watermark embedding procedure (see Fig 3) is reduced as

$$\tilde{F}_n^m = F_n^m + \alpha W \qquad (n = l+1, ..., l+N_W)$$

where $\tilde{F}_n^m$ is the n^{th} watermarked frame, α is the embedding gain, W is the watermark.

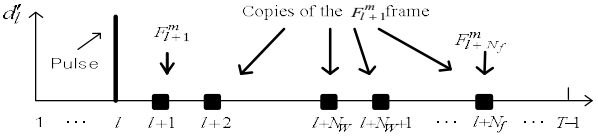

Fig. 2 Replacement of the frames $F_{l+2}^m, \ldots, F_{l+N_f}^m$ with the same watermark-free frame F_{l+1}^m

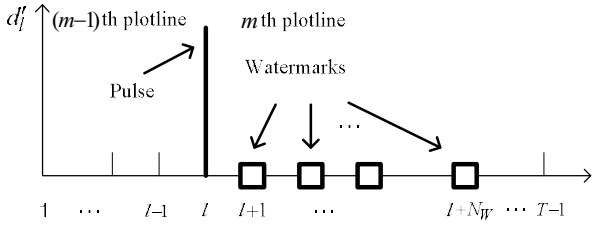

Fig. 3 The absolute difference d_l'

By embedding several copies of the watermark within each plotline we create some watermark redundancy, which increases the robustness of our scheme against such attacks as frame dropping and frame insertion. Note that we propose to embed these copies into each plotline keeping the watermark redundancy on a low level in compare with the schemes [2]-[6].

3. WATERMARK EXTRACTION

Let us make an assumption that the watermarked video is undergone by collusion attacks. Taking into account that Type I collusion exploits the high watermark redundancy, this kind of attack can not be effectively applied against inserted watermarks by definition since in our case the redundancy can be kept on a low level. However Type II collusion may smooth away in some degree the pulses, representing the outsets of each plotline, like low frequency filter

$$\hat{F}_l^m = \frac{1}{L} \sum_{n=l}^{l+L-1} \tilde{F}_n^m \qquad (2)$$

where $\hat{F}_l^m$ is the l^{th} estimated watermark-free frame from this sliding averaging, $\tilde{F}_n^m$ is the n^{th} watermarked frame, L is the total number of averaging frames (window size).

Really, in accordance with (2) L frames are averaged into one frame and as a result the values of the absolute difference will be spread among L frames. Let, for example, L be set to 3. Then, after this attack, the pulse will be converted into L pulses represented by the sequence $\hat{d}_l'$ as shown in Fig. 4

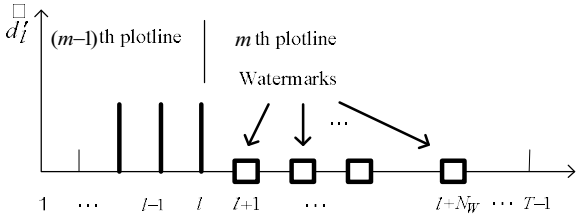

Fig. 4 The smoothed pulse after Type II collusion attack when $L = 3$

To restore the pulse (see Fig. 5) we propose to take the difference d_l'' of the difference: $\hat{d}_l'$

$$d_l'' = \hat{d}_l' - \hat{d}_{l+1}'. \qquad (3)$$

Note that a number of the pulse obtained after calculating d_l' and d_l'' is the same.

Suppose that some watermarks are embedded into N_W frames as shown in Fig. 5. For their extraction, picking out the pulses is the first requirement. This procedure can be done by using (1) and (3). Then, the obtained values of the difference d_l'' must be compared with threshold h, as shown in Fig. 5. If $d_l'' > h$ a decision is made that the pulse has been found and, consequently, the procedure of the watermark extraction is instructed to be begun.

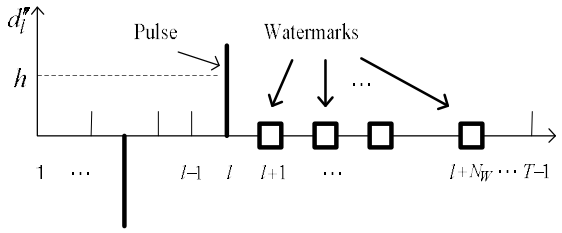

Fig. 5 Restored pulse after calculating the difference d_l''

Second, within the m^{th} plotline, we proposed to average several identical frames $F^m_{l+N_W+1}$, … , $F^m_{l+N_f}$ into one frame and subtract it from the watermarked frames F^m_{l+1}, … , $F^m_{l+N_W}$ with further summation of all the extracted copies of the watermark:

$$\hat{W}_l^m = \frac{\alpha}{N_W} \sum_{n=l+1}^{l+N_W} \left(\hat{F}_n^m - \frac{1}{N_f - N_W} \sum_{k=l+N_W+1}^{l+N_f} F_k^m \right).$$

The final version of the extracted watermark $\hat{W}$ is formed as a result of summation of the watermarks extracted from all plotlines of the video. To make a decision about the watermark presence in a possibly video the normalized correlation ρ is calculated between W and $\hat{W}$. If ρ is found to exceed threshold $\hat{h}$, the watermark is determined to have been successfully extracted; otherwise, the watermark recognized as not existing, or the extraction has failed. In order that the proposed scheme effectively withstand Type II collusion, it is necessary to fulfill the following condition

$$N_f - N_W > L. \quad (4)$$

It may be seen under this condition the sliding averaging (2) will not disturb the embedded watermark.

4. EXPERIMENTAL RESULTS

To evaluate the effectiveness of the proposed scheme a number of experiments have been carried out. Watermark embedding and extraction models have been implemented for the video sequences. The experiments were made based on 20 different TV clips of $T = 1500$ frames and of 640×480 pixels and 256 gray levels, representing about one minute of video at 25 frames per second. During experimental research the watermarks were embedded into different number of plotlines. We disturbed the video with collusion attacks, frame dropping attack and insertion of arbitrary frames attack.

First, we tested our observation that each plotline is really characterized by its own mean gray level. Figure 6a illustrates four plotlines represented by the sequence f_n where $n = 1, ..., 500$. It can be seen from the figure that the gray level values of one plotline considerably differ from the other. The typical frames of the clip “Nature”, representing the first and the second plotlines, are shown in figure 7a and 7b, respectively. The calculated absolute difference d_l', corresponding to figure 6a, is shown in figure 8a. It is seen the pulses in d_l' correspond to the outset of each plotline.

It was experimentally proved that these pulses can be used to localize the watermarks. However these pulses may be smoothed away after Type II collusion. For organizing this attack we used the expression (2) and obtained a sequence $\hat{f}_n$ by averaging matrices $\hat{F}_n^m$ when L set to 18 (see Fig. 6b).

The absolute difference $\hat{d}_l'$, corresponding to Fig. 6b, is shown in Fig. 8b. It is seen that only trailing edges of these smoothed pulses correspond to outsets of the plotlines. To restore the pulses we calculated $\hat{d}_l''$ in accordance with (4) and obtained positive pulses which can be used for the watermarks extraction (see Fig. 9a and 9b where d_l'' and $\hat{d}_l''$ correspond to Fig. 8a and 8b, respectively).

Figure 10 illustrates two types of the watermarks (a) and (b), generated with mean zero and variance one which were embedded in video frames with the gain strength $\alpha = 3$.

Performance of our scheme was measured by the percentage of the extracted watermarks in the attacked video depending on such parameters as window size L, a number of the plotlines M, a number of the watermark-free frame N_f, a number of the watermarked frames N_W within one plotline. An extraction rate of 100% indicates that the attack is ineffective. An extraction rate of 0% indicates that our scheme is unable to extract the watermark in the attacked video.

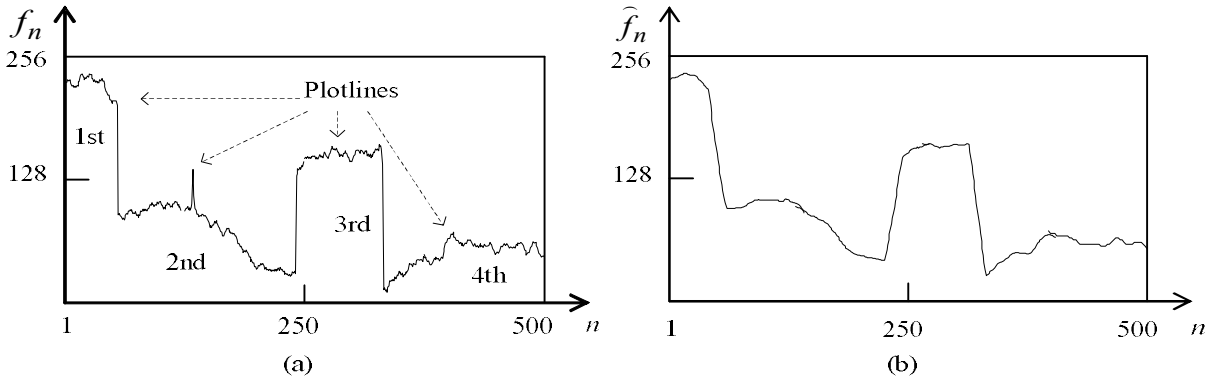

Fig. 6 The sequence of 500 frames before (a) and after (b) averaging attack when $L = 18$

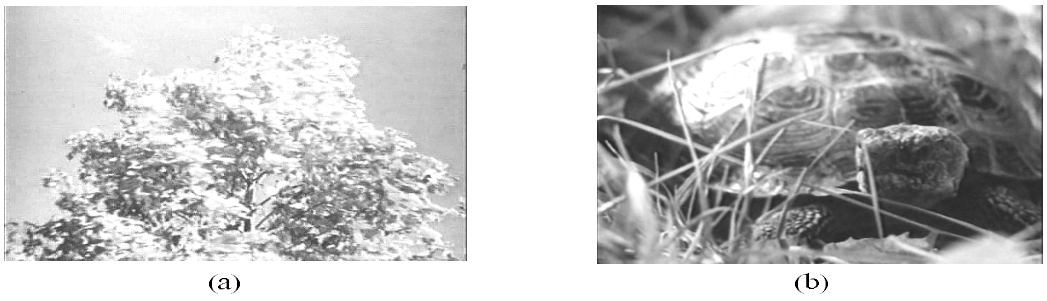

Fig. 7 The real frames corresponding to the first and the second plotlines shown in Fig. 6a

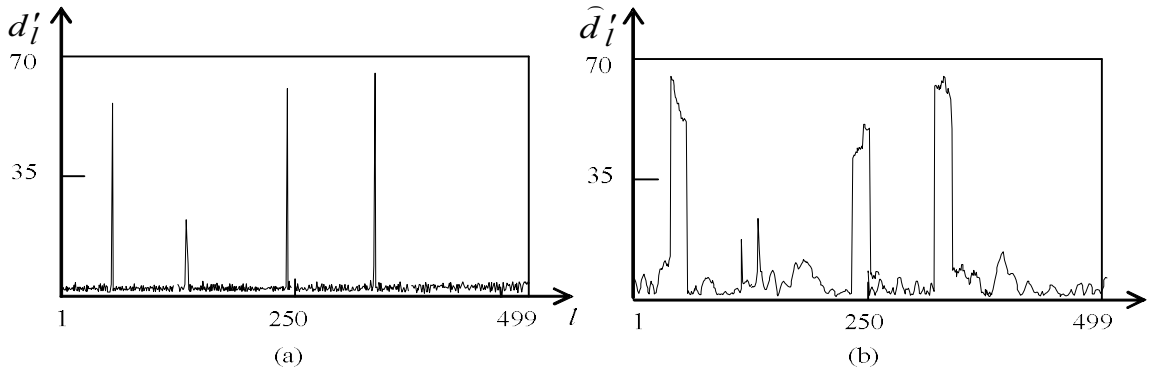

Fig. 8 The absolute difference d'_n before (a) and after (b) Type II collusion attack

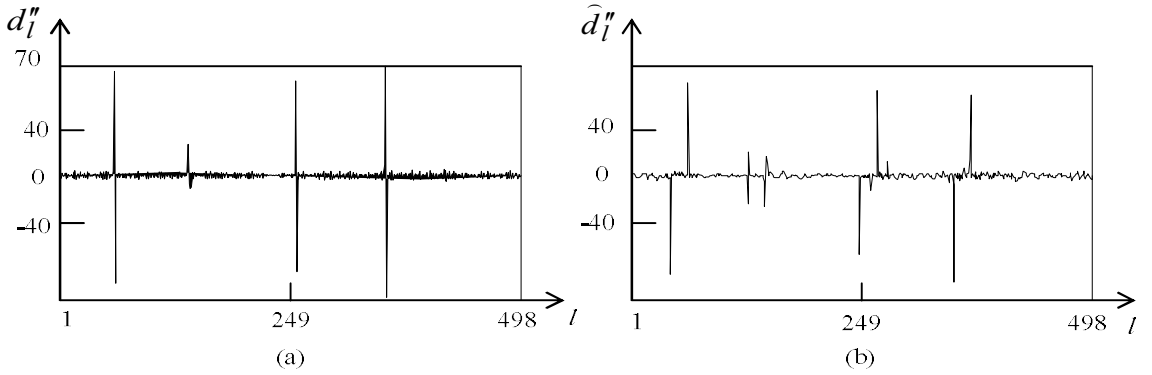

Fig. 9 The difference d''_l of the absolute difference d'_l before (a) and after (b) Type II collusion attack

Table 1. Typical values of the difference d_l'' obtained after processing 15 different plotlines

№ of plotline	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
d_n''	80	79	81	76	44	72	86	47	76	51	74	46	89	94	67

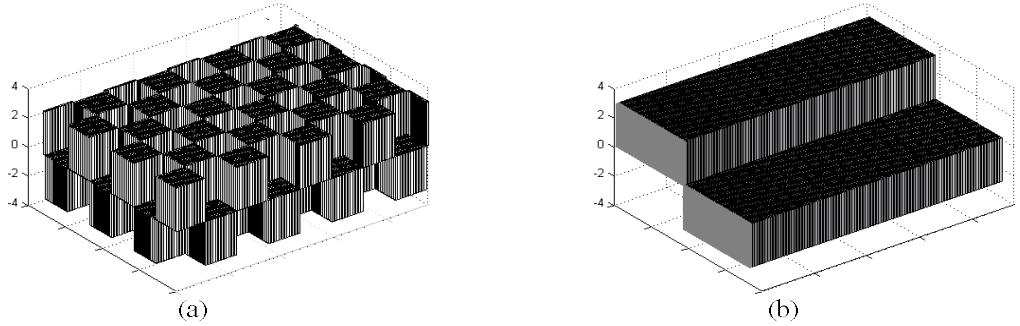

Fig. 10 Two types of watermarks which were embedded in video frames

The threshold h (see Fig. 4) was set to 40 on the base of real values the difference d_l''. Some typical values of d_l'' are presented in Table 1 which were obtained after processing 15 different plotlines.

The normalized correlation ρ was compared with the threshold $\hat{h}$ the value of which was also found experimentally and set to 0.6.

Table 2. Dependence of the percentage of extracted watermarks on L, M, N_f and N_W

L	Percentage of extracted watermarks			
	$M=1$		$M=4$	
	$N_f=8$ $N_W=4$	$N_f=18$ $N_W=9$	$N_f=8$ $N_W=4$	$N_f=18$ $N_W=9$
2	100	100	100	100
4	100	100	100	100
6	0	100	100	100
8	0	100	87	100
10	0	100	43	100
12	0	0	0	100
14	0	0	0	54
16	0	0	0	35
18	0	0	0	0

First of all we tested the robustness of our scheme against Type I collusion attack. The obtained results demonstrated that our scheme does react on this kind of attack at all: the extraction rate was 100%. As it was already mentioned, the most dangerous attack for our scheme is Type II collusion. In Table 2 the dependence of extracted watermarks percentage on L, M, N_f and N_W, which was calculated after processing 20 video clips, is presented.

The data presented in table 2 show that the proposed scheme is capable to withstand against Type II collusion attack with high efficiency. For example, when L is set to 2 or 4 our scheme provides 100% of extracted watermarks independently of watermark redundancy and the other parameters. For comparison it should be noted that the values of extracted watermarks percentage for the scheme in [6, see Fig. 13] are within the range from 0% to 95% depending on the level of redundancy (Period [Frames]/Repeat [Frames]). It is indicated in [6] that the larger temporal redundancy, the higher extracted watermarks percentage. However, the high level of the redundancy is considered in [6] as serious disadvantage for the watermark security.

The robustness of the proposed scheme was also tested against frame-dropping attack, where each frame of the watermarked movie was dropped with probabilities $P = 0.25$ and $P = 0.5$. The watermark redundancy was set to 300/5 as in [6]. The extracted watermarks percentage for our scheme was 97% and 68%, respectively. Note that for scheme, proposed in [6], the same data are around 100% and 58%, respectively. Practically the same results were obtained for the frame insertion attack.

5. CONCLUSION

At present mainly two major embedding strategies are used in the watermarking: either a different watermark is inserted in each video frame, or the same watermark is embedded in all video frames. Both of these strategies result in a high temporal redundancy that an attacker can exploit in order to remove the watermark. In contrast to these strategies, a strategy of embedding only a few watermarks within each plotline keeping the temporal redundancy of the watermark on a very low level was proposed. As a result, the collusion attacks were not effective against our scheme since they are based on the property of watermark redundancy.

Furthermore, the robustness of our scheme against such temporal synchronization attacks as frame dropping and insertion of arbitrary frames was very high by two reasons. First, since we embed the watermark a few numbers of times, the probability of watermark distortion by this type of attacks was not large. Second, even in the case of watermark disruption within one plotline, our scheme allows the finding of copies within the other plotlines. Note that the more plotlines that are processed, the larger the value of correlation ρ . During experiments, we processed only four plotlines and obtained very high percentage of extracted watermarks. A full-length video may contain tens of thousands of plotlines. It is apparent that in this case, the percentage of extracted watermarks will tend to100%, even in the face of any attacks.

Finally, watermark were only embedded into negligible number of frames within each plotline, the distortions of the video are reduced to a minimum.

REFERENCES

[1] I.J. Cox, J. Killian, F.T. Leighton and T. Shamoon. "Secure spread spectrum watermarking for multimedia," *IEEE Trans. Image Proc., vol. 6, No. 12*, pp. 1673-1687, 1997.

[2] G. Doerr, J.L. Dugelay. "A guide tour of video watermarking," *Signal processing: Image Commun., vol. 18, no. 4*, pp 263-285, 2003.

[3] Karen Su, Deepa Kundur, Dimitrios Hatzinakos. "A Novel approach to collusion-resistant video watermarking," *Proc. SPIE Vol. 4675*, p. 491-502, 2002.

[4] Fredric Deguillaume, Gabriela Csurka, Thierry Pun. "Countermeasures for unintentional and intentional video watermarking attacks," *Proc. SPIE Vol. 3971*, p. 346-357, 2000.

[5] G. Doerr, J.L. Dugelay. "Security Pitfalls of Frame-by-Frame Approach to Video Watermarking," *IEEE Trans. on Signal Proc., vol. 52, No. 10*, pp. 2955-2964, 2004.

[6] Eugene T.Lin, Edward J. Delp. "Temporal synchronization in video Watermarking," *IEEE Trans. on Signal Proc., vol. 52, No. 10*, pp. 3007-3021, 2004.

Ki-Jung Kim

He received the B.S. degree in electronics engineering, and the M.S degree and the Ph.D. degrees in the same field from Wonkwang University, South Korea. Since 2005, he has been professor of Department of Information and Communications, Shingyeong University, Hwaseong, Korea. His main research interests include image processing, digital signal processing and multimedia communications systems.