일반논문 (Regular Paper)
방송공학회논문지 제17권 제6호, 2012년 11월 (JBE Vol. 17, No. 6, November 2012)
http://dx.doi.org/10.5909/JBE.2012.17.6.1048
ISSN 1226-7953(Print)

MSE 추정에 기반한 적응적인 시간적 공간적 비디오 디노이징 필터

김 창 수[a)], 김 종 호[a)], 최 윤 식[a)‡]

Video De-noising Using Adaptive Temporal and Spatial Filter Based on Mean Square Error Estimation

Changshou Jin[a)], Jongho Kim[a)], and Yoonsik Choe[a)‡]

요 약

본 논문에서는 영상에 포함되어 있는 잡음을 효율적으로 제거하기 위해 원본 영상과 잡음이 포함된 영상 사이의 mean square error (MSE) 추정에 기반한 적응적인 시공간 디노이징 필터(Adaptive Temporal and Spatial De-noising Filter : ATSF)를 제안하였다. 제안하는 디노이징 필터는 잡음이 포함되어 있는 영상에 블록 단위로 적용되며, 시간적 필터인 Multi-Hypothesis Motion Compensated Filter (MHMCF)를 사용하고, 공간적 필터로는 bilateral filter를 사용하였다. 각각의 블록에 대해 시간적 필터와 공간적 필터 중에서 최적의 필터를 선택하기 위해서 잡음이 포함되지 않은 원본 영상과 잡음이 포함된 입력 영상 사이의 MSE를 추정하는 기법을 제안하였다. 디노이징 단계에서 원본 영상이 주어지지 않기 때문에 MSE를 추정하기 위해서, 본 논문에서는 MHMCF가 적용된 블록의 MSE를 수학적으로 예측하고, bilateral filter가 적용된 블록의 MSE를 통계적 선형 모델을 통해 예측하였다. 이렇게 예측된 MSE를 비교하여 더 작은 MSE를 갖는 필터를 선택적으로 매 단위 블록마다 적용하게 된다. 제안된 방법은 시간적 필터와 공간적 필터를 적응적으로 적용함으로써 기존의 디노이징 방법에 비해 객관적 화질 뿐만 아니라 주관적인 화질에서 우수한 성능을 보여준다.

Abstract

In this paper, an adaptive temporal and spatial filter (ATSF) based on mean square error (MSE) estimation is proposed. ATSF is a block based de-noising algorithm. Each noisy block is selectively filtered by a temporal filter or a spatial filter. Multi-hypothesis motion compensated filter (MHMCF) and bilateral filter are chosen as the temporal filter and the spatial filter, respectively. Although there is no original video, we mathematically derivate a formular to estimate the real MSE between a block de-noised by MHMCF and its original block and a linear model is proposed to estimate the real MSE between a block de-noised by bilateral filter and its original block. Finally, each noisy block is processed by the filter with a smaller estimated MSE. Simulation results show that our proposed algorithm achieves substantial improvements in terms of both visual quality and PSNR as compared with the conventional de-noising algorithms.

Keyword : mean square error, video de-noising, blocking effects.

---

a) 연세대학교 전기전자공학부(School of Electrical and Electronic Engineering, Yonsei University)
‡ Corresponding Author : 최윤식 (Yoonsik Choe)
E-mail: yschoe@yonsei.ac.kr
Tel: +82-2-2123-2774 Fax: +82-2-313-0960

※ "This research was supported by the MKE(The Ministry of Knowledge Economy), Korea, under the "ITRC(Information Technology Research Center)" support program supervised by the NIPA(National IT Industry Promotion Agency)" (NIPA-2012-H0301-12-1008)
· Manuscript received August 6, 2012 Revised November 9, 2012 Accepted November 9, 2012

Ⅰ. Introduction

Noise is always introduced into video sequences at different video processing steps. Noise in digital video not only degrades video visual quality but also reduces compression efficiency and influences subsequent processing tasks such as feature extraction, object detection, prediction and object tracking. So de-noising techniques are necessary and important in practical video processing systems.

Generally, de-noising algorithms are classified to three types: 2-D spatial filters [1]-[6], 1-D temporal filters [7],[8] and 3-D spatio-temporal filters [9]-[11]. De-noising algorithms can perform in pixel domain and transform domain. De-noising in transform domain needs a heavy computation, so de-noising algorithms which perform in pixel domain are most widely used.

2-D spatial de-noising algorithms, including 2-D Kalman filter [1], 2-D Wiener filter [2], wavelet shrinkage [3], non-local means [4], Bilateral filtering [5],[6], etc, utilize spatial correlation to suppress noise. 1-D temporal filters utilize temporal correlation between video frames to remove noise. Existing 1-D temporal filters include adaptive temporal filter based on motion compensation [7], Multi-hypothesis motion compensated filter (MHMCF) [8] , etc. 3-D spatio-temporal filters, such as 3-D Kalman filter [9], spatio-temporal shrinkage [10], spatio-temporal joint filtering scheme (JNT) [11], etc, exploit both spatial and temporal correlation to suppress noise by operating data on 3 dimensions.

Bilateral filter [5] is a 2-D nonlinear filter which performs spatial averaging with relatively slight edge smoothing. However, bilateral filtering brings detail elimination and blurring effects. Moreover it can not remove high-frequency noise well. So the de-noising performance of bilateral filter is poor when an image has a lot of edges, textures and details. In MHMCF [8], temporal correlation is exploited fully. But on the conditions of scene change, occlusion, objects have big motions, objects come into frame suddenly and so on, serious blocking effects are involved in the de-noised results and noise remains because of low temporal correlation between video frames. In JNT [11], the noisy video was first filtered by 1-D kalman filter and 2-D wiener filter respectively and then average the two de-noising results as the final result. JNT inherits errors from poor de-noising results of kalman filter and wiener filter.

In this paper, we propose a novel adaptive temporal and spatial filter (ATSF) which effectively exploits both temporal and spatial correlation based on mean square error (MSE) estimation. In the proposed algorithm, MHMCF is used as a temporal filter and bilateral filter is used as a spatial filter. Although there is no original video, we propose a linear model to estimate the spatial filtering MSE. And an MSE estimation approach to estimate the temporal filtering MSE is given in this paper. De-noising of ATSF proceeds block by block and each noisy block is processed by the filter with a smaller estimated MSE between a de-noised block and its original block. By using ATSF, noisy blocks which contain a lot of details, textures or edges are adaptively filtered by MHMCF and those which have low temporal correlation between video frames are adaptively filtered by bilateral filter. As a result, there are no blocking effects and remained noise involved in the de-noising results of ATSF. Moreover, edges and details are preserved well by ATSF.

This paper is organized as following. In section Ⅱ, related works are introduced. In section Ⅲ, the proposed adaptive temporal and spatial filter based on mean square error estimation is described in detail. The simulation results and analysis are given in section Ⅳ. Finally, conclusions are given in section Ⅴ.

Ⅱ. Related works

In this paper, our degradation model at location (i,j) in

kth frame is as following:

$$G_k(i,j) = F_k(i,j) + N_k(i,j) \tag{1}$$

where G_k is the noisy kth frame, F_k is the noise-free original kth frame and N_k is the additive white Gaussian noise term with zero mean and variance $\sigma^2_{n_k}$.

1. Multihypothesis motion compensated filter

In MHMCF[8], a noisy frame is processed block-by-block by weighted averaging the hypothesis of a noisy block and the noisy block. Let a noisy N×N block in kth noisy frame G_k be denoted as g_k, its original N×N block in kth original frame F_k be denoted as f_k and the zero-mean white gaussian noise which contaminates f_k be denoted as n_k, then there is:

$$g_k = f_k + n_k \tag{2}$$

The reference frames include the previous de-noised M frames $\widehat{F_{k-1}}, \widehat{F_{k-2}} \ldots, \widehat{F_{k-M}}$, where M is the number of how many previous de-noised frames is used to make the temporal hypothesis of a noisy block g_k. The predictions of a noisy block are denoted as $p_{k-m}, m = 1, \ldots, M$. The residue between p_{k-m} and f_k is denoted as r_m and the value of r_m at position (i,j) is defined as

$$r_m(i,j) = f_k(i,j) - p_{k-m}(i+v_i, j+v_j) \tag{3}$$

where $m = 1, \ldots, M$. $\vec{\nu} = (v_i, v_j)$ is the motion vector between p_{k-m} and g_k, which is gained by performing block-based motion estimation. The mean value of r_m is denoted as $\overline{r_{k-m}}$ and the variance of r_m is denoted as $\sigma^2_{r_{k-m}}$.

The de-noised result is given by

$$\hat{f}_k = \sum_{m=1}^{M} w_m \cdot p_{k-m} + w_0 \cdot g_k + d \tag{4}$$

In MHMCF, the author defines an objective function of MHMCF, ϵ ,to be

$$\begin{aligned} \epsilon &= (\acute{f} - \hat{f}_k \cdot 1)^T cov^{-1}(r)(\acute{f} - \hat{f}_k \cdot 1) \\ &= (\acute{f} - w^T \acute{f} \cdot 1 - d \cdot 1)^T cov^{-1}(r)(\acute{f} - w^T \acute{f} \cdot 1 - d \cdot 1) \end{aligned} \tag{5}$$

where $r = [-n_{k,\ldots,} f_k - p_{k-m}]^T, \acute{f} = [f_k + n_{k,\ldots,} p_{k-m}]^T$ and $w = [w_{0,\ldots,} w_m]^T$. Minimizing ϵ by putting $\delta\epsilon/\delta w = 0$ and $\delta\epsilon/\delta d = 0$, the optimal w and d are

$$w_0 = \frac{\sigma^{-2}_{n_k}}{\sum_{m=1}^{M} \sigma^{-2}_{r_{k-m}} + \sigma^{-2}_{n_k}} \tag{6}$$

$$w_m = \frac{\sigma^{-2}_{r_{k-m}}}{\sum_{m=1}^{M} \sigma^{-2}_{r_{k-m}} + \sigma^{-2}_{n_k}} \tag{7}$$

$$d = \sum_{m=1}^{M} w_m \times \overline{r_{k-m}} \tag{8}$$

where $\sigma^2_{n_k}$ is the variance of noise n_k. w_0 and w_m are the weights of a noisy block g_k and its prediction blocks p_{k-m} ($m = 1, \ldots, M$), respectively. Then d is a constant which makes MHMCF be an unbiased estimator.

As MHMCF is a purely temporal filter, it has a very good visual quality with an edge and detail preservation property. Temporal correlation can be fully used by exploiting several reference frames to make hypothesis for every noisy block.

However, as MHMCF processes a noisy frame block by block, blocking effects are involved in the de-noised results of MHMCF more or less. The more reference frames are used, the more serious blocking effects will be involved in de-noised results. Under the conditions of scene change, occlusion, objects have big motions, objects come into frame suddenly and so on, serious blocking effects are involved in de-noised results and noise remains. As a recursive de-noising filter, blocking effects, remained noise and other errors are transmitted on temporal dimension recursively. MSE quantifies the difference between a de-noised block and its origianl block. So de-noised blocks which contain blocking effects, remained noise and other errors accompany large MSEs.

2. Bilateral filter

Bilateral filter [5] is a non-linear 2-D spatial filter which removes noise by weighted averaging noisy pixels in the local filtering window. Bilateral filter smooths out noise while edges are preserved well because the weights depends on two values. One is spatial distance between a noisy pixel and the center noisy pixel in the local filtering window, which is reflected by the first exponential term. The other is intensity distance between a noisy pixel and the center noisy pixel in the local filtering window, which is reflected by the second exponential term.

In equation (9), a noise corrupted pixel at position (i,j) in the kth noisy frame G_k is defined as $G_k(i,j)$ and $\widehat{F}_k(i,j)$ is the de-noised pixel of $G_k(i,j)$. The filtering window size is (2r+1)×(2s+1), $G_k(p,q)$ is a neighboring noisy pixel of $G_k(i,j)$ in the local filtering window.

In equation (10), C is the normalization constant which makes coefficient summation equals to 1. The parameters σ_d and σ_r control the fall-off of the weights in spatial and intensity domains, respectively. Given specific application, design of parameter values are possible. In [4], Buades proposes a non-local means filter, the pixel weights are determined by spatial distances and similarities between neighboring blocks and the local block. When the block size reduces to one pixel, the no-local means filter becomes bilateral filter.

With very good edge preservation property, bilateral filter has low computation complexity. However, as a spatial filter, bilateral filtering brings detail elimination and blurring effects. Moreover high-frequency noise remains in de-noised results after bilateral filtering. So an image which contains a lot of details, textures or edges gets a poor de-noised result by using bilateral filtering. As mentioned earlier, poor de-noising results accompany large MSEs.

To combine the advantages of MHMCF and bilateral filter, we propose a novel adaptive temporal and spatial filter (ATSF) based on mean square error (MSE) estimation, which removes noise effectively and has a novel good visual quality.

III. Proposed algorithm

The proposed algorithm, ATSF processes a noisy video frame by frame and block by block. Each noisy frame is divided into N×N blocks for MSE estimation and filtering. ATSF has four main sections: Temporal filtering MSE estimation (MSE between a block de-noised by MHMCF and its original block), spatial filtering MSE estimation (MSE between a block de-noised by bilateral filter and its original block), spatial filtering and temporal filtering. Noise level is detected for every noisy frame in the homogeneous region and it is assumed that blocks in a frame have a same noise variance. For every noisy block, there are two kinds of filtering options: temporal filtering (MHMCF) and spatial filtering (bilateral filter). Temporal filtering MSE estimation is performed before temporal filtering, estimated

$$\hat{F}(i,j)=\frac{1}{C}\sum_{p=i-r}^{p=i+r}\sum_{q=j-s}^{q=j+s}\exp(-\frac{|(p-i)^2+(q-j)^2|}{2\sigma_d^2}).\exp(-\frac{|G(p,q)-G(i,j)|^2}{2\sigma_r^2}).G(p,q) \tag{9}$$

$$C=\sum_{p=i-r}^{p=i+r}\sum_{q=j-s}^{q=j+s}\exp(-\frac{|(p-i)^2+(q-j)^2|}{2\sigma_d^2}).\exp(-\frac{|G(p,q)-G(i,j)|^2}{2\sigma_r^2}) \tag{10}$$

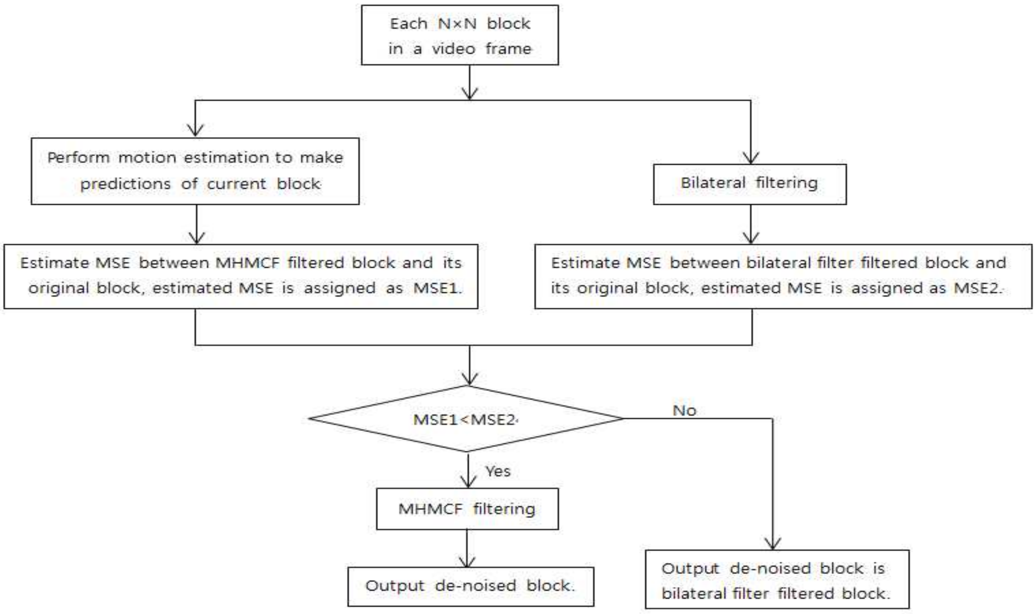


그림 1. ATSF 알고리즘 흐름도
Fig. 1. Block diagram of the proposed ATSF algorithm

MSE is assigned as $MSE1$. Spatial filtering MSE estimation is performed after spatial filtering, estimated MSE is assigned as $MSE2$. If $MSE1$ is smaller than $MSE2$, MHMCF filtering is performed to the current noisy block. If not, the final de-noising result is the block de-noised by bilateral filter. A block diagram of proposed algorithm is illustrated in Fig.1. Let ATSF using 1HMCF (MHMCF uses one reference frame) and bilateral filter be presented by ATSF1R and ATSF using 2HMCF (MHMCF uses two reference frames) and bilateral filter be presented by ATSF2R. The four main sections of the proposed algorithm ATSF are as following:

1. Temporal filtering MSE estimation

As discussed in section Ⅱ-1,

$$g_k = f_k + n_k$$

$$\hat{f}_k = \sum_{m=1}^{M} w_m \cdot p_{k-m} + w_0 \cdot g_k + d$$

Mean square error between a de-noised block $\hat{f}_k$ and its original block f_k contains two parts as (11). The first term is the variance part and the second term is the square of bias.

$$\begin{aligned} MSE &= E[\,\| \hat{f}_k - f_k \|^2] \\ &= E[(\hat{f}_k - f_k - E(\hat{f}_k - f_k) + E(\hat{f}_k - f_k))^2] \\ &= E[(\hat{f}_k - f_k - E(\hat{f}_k - f_k))^2 + (E(\hat{f}_k - f_k))^2 + 2 \bullet (\hat{f}_k - f_k - E(\hat{f}_k - f_k)) \bullet E(\hat{f}_k - f_k)] \\ &= E[(\hat{f}_k - f_k - E(\hat{f}_k - f_k))^2] + E[(E(\hat{f}_k - f_k))^2] \\ &= Var(\hat{f}_k - f_k) + [E(\hat{f}_k - f_k)]^2 \end{aligned} \quad (11)$$

For MHMCF, the bias part of the MSE is as (12).

$$\begin{aligned} & E(\hat{f}_k - f_k) \\ & = E(\sum_{m=1}^{M} w_m \cdot p_{k-m} + w_0 \cdot g_k + d - f_k) \\ & = E(\sum_{m=1}^{M} w_m \cdot p_{k-m} + w_0 \cdot g_k + \sum_{m=1}^{M} w_m \cdot E(f_k - p_{k-m}) - f_k) \\ & = w_0 \cdot E(f_k) + \sum_{m=1}^{M} w_m \cdot E(f_k) - E(f_k) \\ & = E(f_k) - E(f_k) \\ & = 0 \end{aligned} \tag{12}$$

From the above derivation we know that the bias part of the MSE between a block de-noised by MHMCF $\hat{f}_k$ and its original block f_k equals to zero. So the MSE between a block de-noised by MHMCF $\hat{f}_k$ and its original block f_k equals to $Var(\hat{f}_k - f_k)$,i.e,

$$MSE = E[\, \| \hat{f}_k - f_k \|^2] = Var(\hat{f}_k - f_k) \tag{13}$$

And following derivation shows the approximation of it.

$$\begin{aligned} Var(\hat{f}_k - f_k) &= Var(\sum_{m=1}^{M} w_m \cdot p_{k-m} + w_0 \cdot g_k + d - f_k) \\ &= Var(\sum_{m=1}^{M} w_m \cdot (p_{k-m} - f_k) + w_0 \cdot (g_k - f_k)) \end{aligned} \tag{14}$$

We assume that $p_{k-m} - f_k$ $(m = 1,...,M)$ and n_k are independent from each other. Combine (6), (7) and (14), then

$$\begin{aligned} Var(\hat{f}_k - f_k) &= \sum_{m=1}^{M} w_m^2 \cdot Var(p_{k-m} - f_k) + w_0^2 \cdot Var(n_k) \\ &= \left(\sum_{m=1}^{M} Var(p_{k-m} - f_k)^{-1} + Var(n_k)^{-1} \right)^{-1} \end{aligned} \tag{15}$$

As $p_{k-m} - f_k$ $(m = 1,...,M)$ and n_k are independent from each other,

$$\begin{aligned} Var(p_{k-m} - g_k) &= Var(p_{k-m} - f_k - n_k) \\ &= Var(p_{k-m} - f_k) + Var(n_k) \end{aligned} \tag{16}$$

Thus,

$$\begin{aligned} MSE &= E[\, \| \hat{f}_k - f_k \|^2] \\ &= \left(\sum_{m=1}^{M} (\, Var(p_{k-m} - g_k) - Var(n_k))^{-1} + Var(n_k)^{-1} \right)^{-1} \end{aligned} \tag{17}$$

Let the estimated MSE between a block de-noised by MHMCF $\hat{f}_k$ and its original block f_k be presented by $MSE1$, then $MSE1$ can be calculated as following:

$$MSE1 = (\sum_{m=1}^{M} (\, Var(p_{k-m} - g_k) - Var(n_k))^{-1} + Var(n_k)^{-1})^{-1} \tag{18}$$

where $m = 1,2...,M$.

2. Spatial filtering MSE estimation

For spatial filtering MSE estimation, a linear model is proposed to estimate the MSE between a block de-noised by bilateral filter and its original block. In the linear model, there are three variables, $x1$, $x2$ and $x3$. $x1$ is the noise variance $Var(n_k)$, $x2$ and $x3$ are the variance of a noisy block $Var(g_k)$ and the variance of its de-noised block which is processed by bilateral filter, respectively.

It is reasonable that when noise variance is fixed, the larger variance difference between a noisy block and its de-noised block is, the larger MSE between a de-noised block and its original block there will be.

Table 1. shows correlation coefficients between $x2 - x3$ and the MSE between a block de-noised by bilateral filter and its original block at different noise variances. From table 1, it can be seen that there is a strong linear relation between $x2 - x3$ and the real MSE at different noise variances. So we propose a linear model to estimate the MSE as following: (In this paper, noise variance of zero-mean white Gaussian noise is assumed to be between

표 1. 다른 노이즈 바이런스 에서 X2-X3 와 리얼 MSE 사이의 상관 계수 (4,680 블록)
Table 1. correlation coefficient between x2-x3 and real MSE at different noise variances (4680 blocks)

Noise variance	Correlation coefficient	Noise variance	Correlation coefficient
10	0.825	160	0.890
20	0.877	170	0.894
30	0.896	180	0.907
40	0.901	190	0.918
50	0.908	200	0.914
60	0.909	210	0.906
70	0.908	220	0.896
80	0.905	230	0.890
90	0.902	240	0.886
100	0.918	250	0.884
110	0.907	260	0.887
120	0.892	270	0.886
130	0.883	280	0.881
140	0.880	290	0.879
150	0.881	300	0.877

0 and 300.)

$$MSE = x^T \cdot \beta \tag{19}$$

where $\beta^T = [\beta_0\ \beta_1\ \beta_2\ \beta_3]$, $x^T = [1\ x1\ x2\ x3]$.

To train a reliable parameter β, we got a large amount of data sets to make a matrix X, whose ith row is the ith observed $x^T = [1\ x1\ x2\ x3]$ and its corresponding real MSE vector which is denoted by y. Then we used the popular estimation method "residual sum of squares" to fix the linear regression model. The residual sum-of-square can be written as:

$$RSS(\beta) = (y - X \cdot \beta)^T \cdot (y - X \cdot \beta) \tag{20}$$

which is a quadratic function of the parameter β.

To find stationary points, we put the first derivative or gradient of $RSS(\beta)$ to zero:

$$X^T \cdot (y - X \cdot \beta) = 0 \tag{21}$$

Assuming that X has full rank then $X^T \cdot X$ is positive definite. The following solution can the be obtained :

$$\begin{aligned} \hat{\beta}^T &= [(X^T \cdot X)^{-1} \cdot X^T \cdot y]^T \\ &= [2.819\ \ -0.255\ \ 0.379\ \ -0.390] \end{aligned} \tag{22}$$

Let the estimated MSE between a block de-noised by bilateral filter and its original block be presented by $MSE2$. Then by using the proposed linear model, $MSE2$ can be calculated as following:

$$MSE2 = x^T \cdot \hat{\beta} = 2.819 - 0.255 * x1 + 0.379 * x2 - 0.39 * x3 \tag{23}$$

In our experimental study, the values of parameter vector $\hat{\beta}$ are determined by bilateral filter and change very little when block size changes.

3. Temporal filtering

MHMCF is chosen as the temporal filter of our proposed algorithm. When MHMCF uses more reference frames, a better de-noising result produces. The de-noising performance increment between 1HMCF and 2HMCF is large but the de-noising performance increment between 2HMCF and 3HMCF is very limited. Sometimes the de-noising performance of 3HMCF is even worse than that of 2HMCF. And adding one reference frame will bring a large computation complexity increment due to motion estimation. So in this paper, ATSF1R which uses 1HMCF and bilateral filter and ATSF2R which uses 2HMCF and bilateral filter are introduced.

4. Spatial filtering

Bilateral filter [5] is chosen as the spatial filter of the proposed algorithm. Thanks to the contribution of Multi-resolution Bilateral Filter for Image De-noising [6], we know that the optimal value of the σ_d is relatively independent of the noise variance and the optimal value of σ_r of the bilateral filter is linearly proportional to the standard deviation of the noise. In our experiments, σ_d=1.8 and $\sigma_r=3\times\sigma_{n_k}$ produced very good results for [5], so we fix the parameters σ_d=1.8 and $\sigma_r=3\times\sigma_{n_k}$ in this paper.

IV. Simulation Results

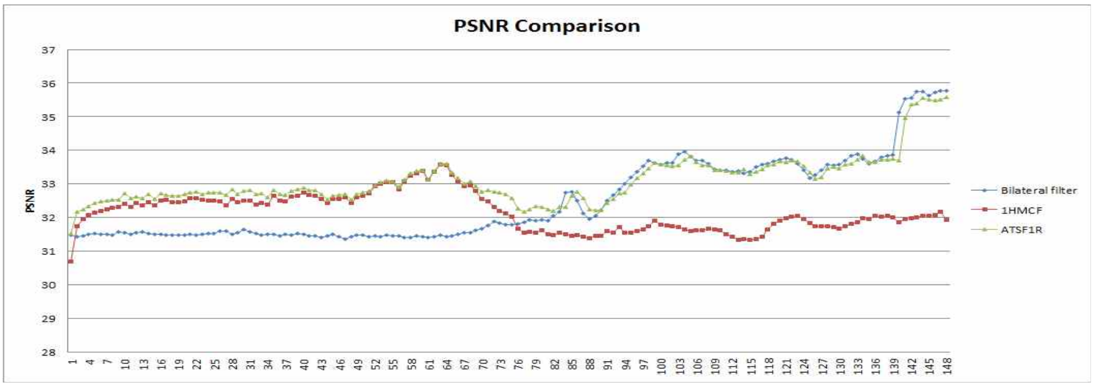
In our experiments, we used 3 different video sequences: "Football", "Foreman" and "Mobile". We added zero-mean white Gaussian noise with the variance $\sigma_n^2(0)$=65, 130 and 260 to the luminance component of each video and processed the sequences with different filters: JNT [11], MHMCF [8], Bilateral filter [5], proposed ATSF1R and ATSF2R. In JNT, 1HMCF, 2HMCF, ATSF1R and ATSF2R, block size 16×16 is used for motion estimation and motion compensation, search window size is 48×48. For spatial filtering, 5×5 window size is used for bilateral filter and the wiener filter of JNT.

1. Average PSNR comparison

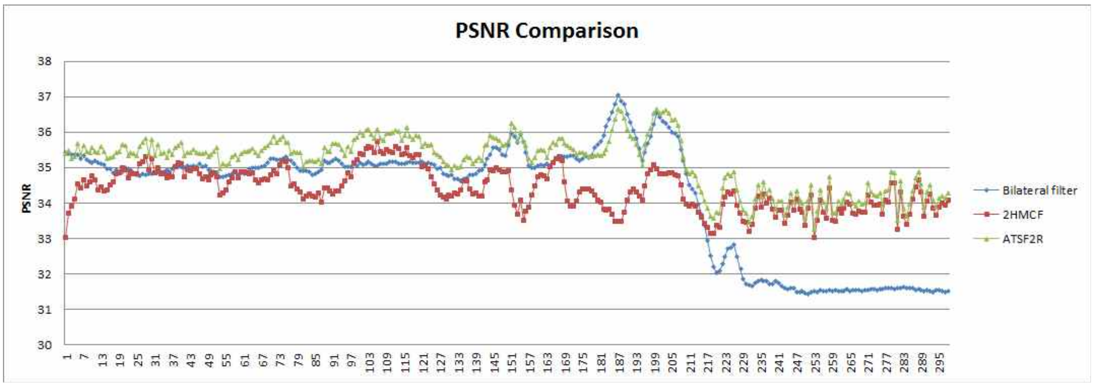
From Table 2, it can be seen that ATSF1R and ATSF2R could effectively suppress noise. PSNR of ATSF1R is up to 1.5dB better than that of 1HMCF. PSNR of ATSF2R is up to 0.86dB better than that of 2HMCF. And ATSF2R obtains highest PSNRs. Compared with ATSF1R and ATSF2R, JNT has considerably low average PSNRs at different noise variances.

표 2. JNT, 1HMCF, 2HMCF, bilateral filter 과 제안한 ATSF1R, ATSF2R 의 PSNR 비교
Table 2. PSNR comparison between JNT, 1HMCF, 2HMCF, bilateral filter, proposed ATSF1R and ATSF2R

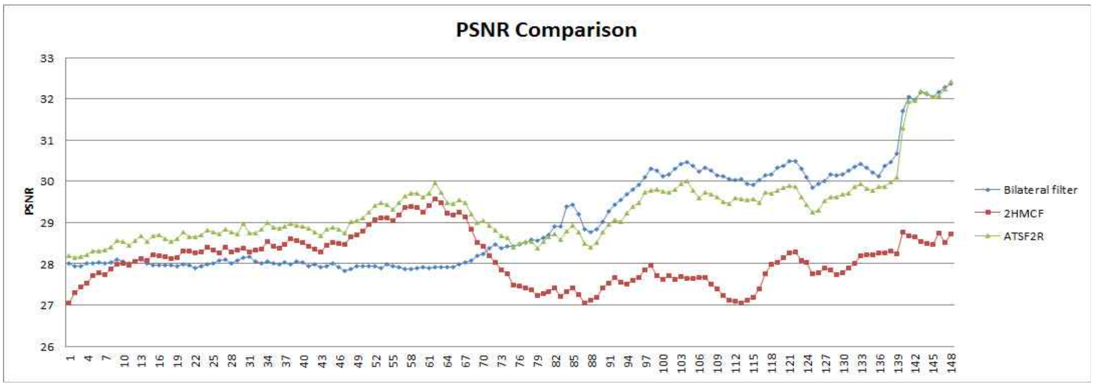
Noise Variance $\sigma_n^2(0)$	Noised video	JNT	1HMCF	2HMCF	Bilateral filer	Proposed ATSF1R	Proposed ATSF2R
Football(150 frames), resolution 352*288							
65.0	29.99	31.43	32.15	32.30	32.48	33.10	33.16
130.0	27.01	29.90	29.87	30.14	30.82	31.24	31.24
260.0	24.03	28.38	27.38	28.11	29.04	29.32	29.31
Foreman(300 frames) resolution 352*288							
65.0	29.99	33.60	33.54	34.40	34.24	34.74	35.17
130.0	27.00	32.06	31.42	32.58	32.54	32.88	33.41
260.0	24.04	30.42	29.32	30.85	30.73	30.88	31.70
Mobile and Calendar(300 frames), resolution 352*288							
65.0	30.01	28.75	31.31	31.97	30.02	31.47	32.05
130.0	27.03	27.64	28.95	29.75	27.79	29.16	29.87
260.0	24.07	26.36	26.80	27.64	25.70	27.07	27.79



(a)



(b)



(c)

그림 2. 각기 다른 노이즈 분산을 갖는 영상에서의 프레임별 PSNR 비교. y 축은 각 프레임의 PSNR을 나타냄. x축은 영상의 프레임 인덱스를 나타냄. (a) bilateral filter, 1HMCF, 그리고 ATSF1R을 사용하여 노이즈를 제거한 "football" 영상 (150 프레임). 영상의 노이즈 분산은 65임. (b) bilateral filter, 2HMCF, 그리고 ATSF2R을 사용하여 노이즈를 제거한 "foreman" 영상 (300 프레임). 영상의 노이즈 분산은 65임. (c) bilateral filter, 2HMCF, 그리고 ATSF2R을 사용하여 노이즈를 제거한 "football" 영상 (150 프레임). 영상의 노이즈 분산은 260임

Fig.2. PSNR comparison for each frame at different noise variances. x axis donates PSNR of each frame, y axis donates frame index of video. (a)"football" video (150 frames) de-noised by bilateral filter, 1HMCF and ATSF1R. Video was corrupted with $\sigma_n^2(0)$ = 65. (b)"foreman" video (300 frames) de-noised by bilateral filter, 2HMCF and ATSF2R. Video was corrupted with $\sigma_n^2(0)$ = 65. (c)"football" video (150 frames) de-noised by bilateral filter, 2HMCF and ATSF2R. Video was corrupted with $\sigma_n^2(0)$ = 260

2. PSNR comparison for each frame

At the latter half of "football" video, footballers start to run. There is occlusion, scene change and some blocks with big motions. So PSNRs of latter half de-noised video frames of 1HMCF and 2HMCF are low. But for bilateral filter, there are no a lot of details, textures or edges, so PSNRs of latter half de-noised video frames of bilateral filter are high.

Walls and a lot of details appear at the last part of "foreman" video, there are a lot of edges and details in the last part frames. So PSNRs of the last de-noised video frames of bilateral filter are very low. But for 1HMCF and 2HMCF, there is no occlusion, scene change or blocks with big motions. So PSNRs of the last de-noised video frames of 1HMCF and 2HMCF are high.

In Fig.2, it can be seen that PSNRs of ATSF1R are bigger than PSNRs of both 1HMCF and bilateral filter when PSNRs of 1HMCF and bilateral filter are similar. And PSNRs of ATSF2R are bigger than those of both 2HMCF and bilateral filter when PSNRs of 2HMCF and bilateral filter are similar. When PSNR difference between 1HMCF (or 2HMCF) and bilateral filter is large, the PSNR of ATSF1R (or ATSF2R) follows the bigger PSNR. From Fig.2(a), (b) and (c), it can be seen that ATSF based on MSE estimation perfectly works on different noise levels.

3. Visual quality comparison

Fig.3. and Fig.4. show de-noising performances of different filters on condition that there is low temporal correlation between video frames for a part of an image. There exists a hand suddenly in 153th frame and there is no a hand in previous frames. So for the hand there is low temporal correlation can be used for de-noising. As a result, the de-noised hands of 1HMCF and 2HMCF are poor.

In Fig.3, for 2HMCF and JNT, except the de-noised

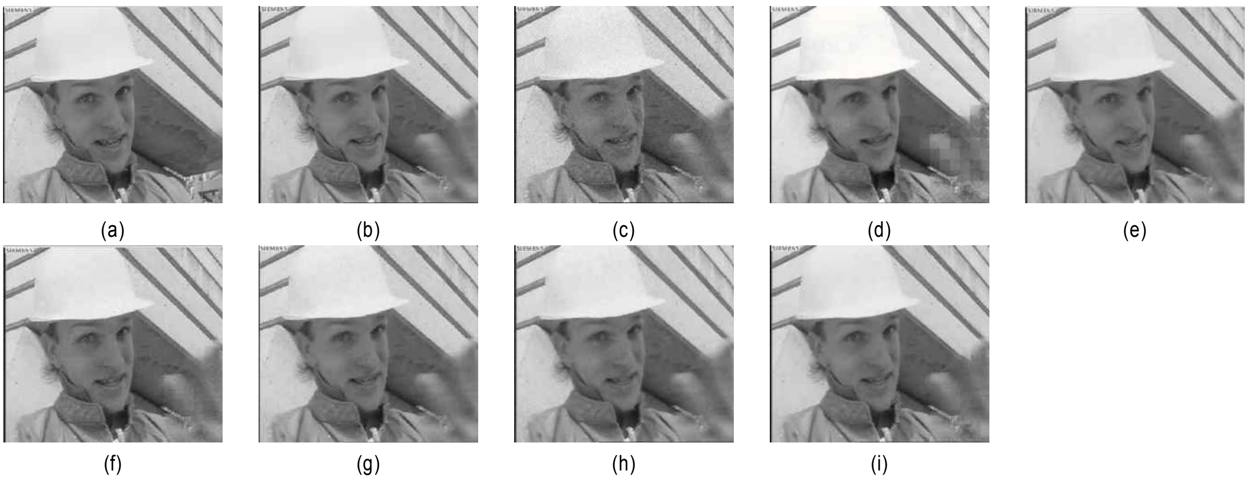


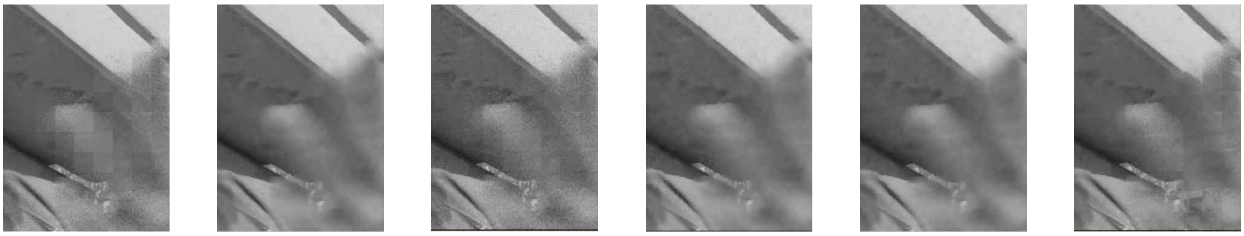
그림 3. 원본“foreman”영상과 디노이징 필터가 적용된“foreman”영상의 153 번째 프레임 (a)원본 영상의 152 번째 프레임 (b)원본 영상의 153 번째 프레임 (c)노이즈가 포함된 153 번째 프레임 (d) 2HMCF가 적용된 153 번째 프레임 (e) ATSF2R이 적용된 153 번째 프레임 (f)1HMCF가 적용된 153 번째 프레임 (g) ATSF1R가 적용된 153 번째 프레임 (h) Bilateral filter가 적용된 153 번째 프레임 (i)JNT가 적용된 153 번째 프레임

Fig.3. Original, noisy and de-noised 153th frames of "foreman" video sequence, $\sigma_n^2(0)$ = 65. (a) Original 152th frame. (b)Original 153th frame. (c)Noisy 153th frame. (d)De-noised 153th frame by 2HMCF. (e)De-noised 153th frame by ATSF2R. (f)De-noised 153th frame by 1HMCF. (g)De-noised 153th frame by ATSF1R. (h)De-noised 153th frame by Bilateral filter. (i)De-noised 153th frame by JNT

footballers with big motions, the de-noised images look clean. The de-noised image of ATSF2R looks cleanest. Edges and details are preserved very well by ATSF2R. For ATSF1R, the de-noised image contains some blurring effects. For 1HMCF, besides the de-noised hand, the other parts of de-noised image contain some noise.

In Fig.4, the de-noised hands of 1HMCF, 2HMCF and JNT contain blocking effects and remained noise. The blocking effects of 2HMCF are more serious than those of 1HMCF and JNT. For JNT, the blocking effects and noise in the de-noised hand inherit from 1-D kalman filter de-noised hand. The de-noised hands of ATSF1R and ATSF2R contain no blocking effects and noise in them is removed very well by ATSF1R and ATSF2R.

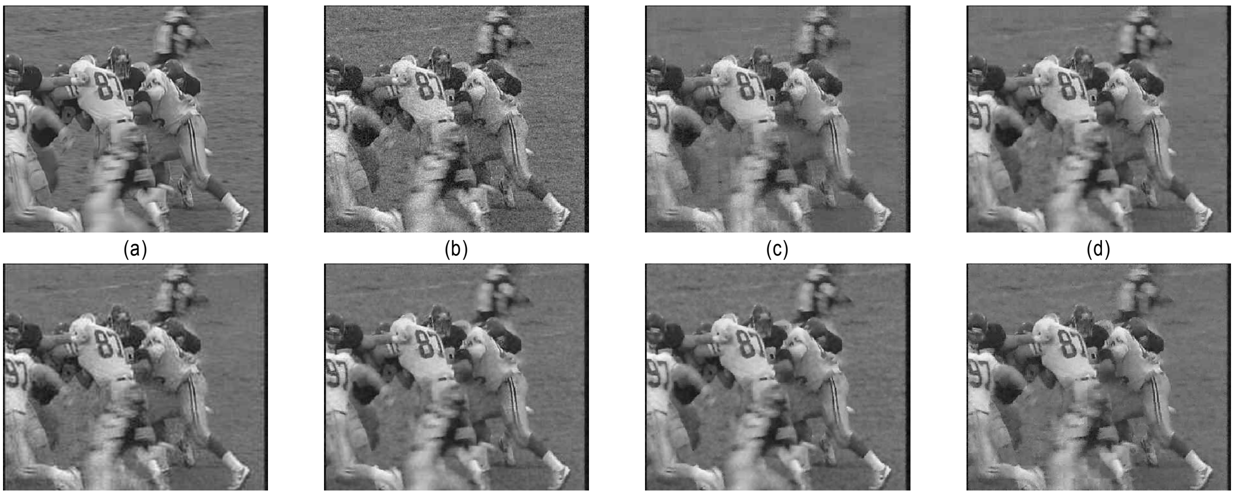
Fig.5. and Fig.6. show de-noising performances of different filters on condition that objects have big motions



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

그림 4. 디노이징 필터가 적용된 “foreman”영상 153 번째 프레임의 "손" (a) 2HMCF가 적용된 손 (b) ATSF2R가 적용된 손 (c) 1HMCF가 적용된 손 (d) ATSF1R가 적용된 손 (e) Bilateral filter가 적용된 손 (f) JNT가 적용된 손

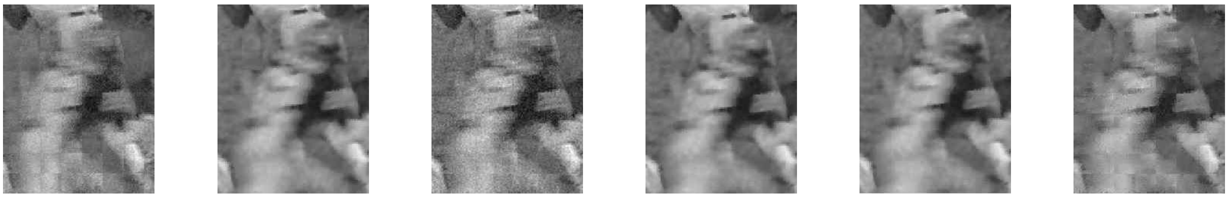
Fig.4. De-noised hand in 153th frame of "foreman" video sequence by different filters, $\sigma_n^2(0)$ = 65. (a) De-noised hand by 2HMCF. (b)De-noised hand by ATSF2R. (c)De-noised hand by 1HMCF. (d)De-noised hand by ATSF1R. (e)De-noised hand by Bilateral filter. (f) De-noised hand by JNT



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

그림 5. 원본“football”영상과 디노이징 필터가 적용된“football”영상의 113 번째 프레임 (a) 원본 영상의 113 번째 프레임 (b) 노이즈가 포함된 113 번째 프레임 (c) 2HMCF가 적용된 113 번째 프레임 (d) ATSF2R가 적용된 113 번째 프레임 (e) 1HMCF가 적용된 113 번째 프레임 (f) ATSF1R가 적용된 113 번째 프레임 (g) Bilateral filter가 적용된 113 번째 프레임 (h) JNT가 적용된 113 번째 프레임

Fig.5. Original, noisy and de-noised 113th frames of "football" video sequence, $\sigma_n^2(0)$ = 260. (a) Original 113th frame. (b) Noisy 113th frame. (c) De-noised 113th frame by 2HMCF. (d) De-noised 113th frame by proposed ATSF2R. (e) De-noised 113th frame by 1HMCF. (f) De-noised 113th frame by ATSF1R. (g) De-noised 113th frame by Bilateral filter. (h) De-noised 113th frame by JNT



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

그림 6. 디노이징 필터가 적용된"football"영상 113 번째 프레임의 "footballer" (a) 2HMCF가 적용된 "footballer" (b) ATSF2R이 적용된 "footballer" (c) 1HMCF가 적용된 "footballer"(d) ATSF1R가 적용된 "footballer" (e) Bilateral filter가 적용된 "footballer" (f) JNT가 적용된 "footballer"

Fig.6. De-noised "footballer" in 113th frame of "football" video sequence by different filters, $\sigma_n^2(0)$ = 260. (a) De-noised "footballer" by 2HMCF. (b) De-noised "footballer" by ATSF2R. (c) De-noised "footballer" by 1HMCF. (d) De-noised "footballer" by ATSF1R. (e) De-noised "footballer" by Bilateral filter. (f) De-noised "footballer" by JNT

and noise level is high. In the 113th frame of "football" video, footballers start to run. There are some blocks with big motions in the 113th frame.

In Fig.5, for 2HMCF and JNT, except the de-noised footballers with big motions, the de-noised images basically look clean. The de-noised image of ATSF2R looks cleanest. Edges and details are preserved very well by ATSF2R. For ATSF1R, the de-noised image contains some blurring effects. For 1HMCF, besides the de-noised footballers with big motions, the other parts of de-noised image contain some noise.

In Fig.6, for 1HMCF, 2HMCF and JNT, the de-noised footballers with big motions contain blocking effects and noise in them is not removed well. The blocking effects of 2HMCF are more serious than those of 1HMCF and JNT. The de-noised footballer of 1HMCF contains the most noise. The de-noised footballers of ATSF1R and ATSF2R contain no blocking effects and noise in them is removed very well by ATSF1R and ATSF2R.

From the above comparisons, it is can be seen that ATSF2R outperforms the other de-noising filters in terms of both PSNR and visual quality. The de-noised results of ATSF1R contain no blocking effects and noise can be removed very well by ATSF1R. However, there are some blurring effects in the de-noised results of ATSF1R. Without using spatial correlation, MHMCF has a poor de-noising performance when temporal correlation between video frames is poor. Bilateral filter brings blurring effects and it has a relatively poor visual quality. JNT lightens the problem that spatial wiener filter smooths out details of an image and temporal kalman filter represses the parts which have poor temporal correlation. But de-noised results of JNT inherit errors from poor de-noising results of 1-D kalman filter and 2-D wiener filter.

Ⅴ. Conclusions

In this paper, a spatio-temporal filter ATSF is developed. By adaptively filtering each noisy block based on MSE estimation, ATSF combines the advantages of MHMCF and bilateral filter. ATSF could remove noise effectively while no blocking effects are introduced in the de-noised results. Moreover, edges and details are preserved well by ATSF. The experimental results show that ATSF outperforms conventional de-noising algorithms in terms of visual quality and PSNR. And in our experimental study, ATSF2R (ATSF using two reference frames) performs best.

References

[1] J.Woods and C. Radewan,"Kalman filtering in two dimensions,"IEEE

Trans. Inf. Theory., vol. 23, no. 7, pp. 473 - 482, Jul.1977.
[2] F. Jin, P. Fieguth, L. Winger, and E. Jernigan,"Adaptive wiener filtering of noisy images and image sequences," in Proc. IEEE Int. Conf. on Image Process., pp. 349 - 352, Sep.2003.
[3] D. L. Donoho, "De-noising by soft-thresholding,"IEEE Trans. Inf. Theory., vol. 41, no. 5, pp. 613 - 627, May .1995.
[4] A. Buades, B. Coll, and J. M. Morel, "A non-local algorithm for image denoising,"in Proc. IEEE Computer Soc. Conf. on Comput. Vision and Pattern Recog., pp. 60 - 65, Jun.2005.
[5] TOMASI C., MANDUCHI R.:"Bilateral filtering for gray and color images,"Sixth Int. Conf. on Computer Vision., pp. 839 - 846, 1998.
[6] MING Z., GUNTURK B.K.:"Multiresolution bilateral filtering for image denoising", IEEE Trans. Image Process., 17, pp. 2324 - 2333, 2008.
[7] LI Y., YANFENG Q.:"Novel adaptive temporal filter based on motion compensation for video noise reduction," Int. Symp. on Communications and Information Technologies, ISCIT.', pp. 1031 - 034, 06, 2006.
[8] GUO L., AU O.C., MA M., LIANG Z.:"Temporal video denoising based on multihypothesis motion compensation, "IEEE Trans. Circuits Syst. Video Technol.,17, p. 1423, 2007.
[9] A. J. Patti, A. M. Tekalp, and M. I. Sezan,"A new motion-compensated reduce-order model Kalman filter for space-varying restoration of progressive and interlaced video,"IEEE Trans. Image Process., vol.7, no. 4, pp. 543 - 554, Apr. 1998.
[10] E. J. Balster,Y. F. Zheng, and R. L. Ewing,"Combined spatial and temporal domain wavelet shrinkage algorithm for video denoising,"IEEE Trans. Circuits Syst. Video Technol., vol. 16, no. 2, pp. 220 - 230, Feb.2006.
[11] R. Dugad and N. Ahuja,"Video denoising by combining Kalman and wiener estimates," in Proc. IEEE Int. Conf. on Image Process., pp. 152 - 156, Oct. 1999.
[12] D. H. Shin, R. H. Park, S. Yang, and J. H. Jung,"Block-based noise estimation using adaptive Gaussian filtering," IEEE Trans. on Consumer Electronics., vol. 51, No. 1, pp. 218 - 226, 2005.
[13] A. Bosco, A. Bruna, G. Messina, and G. Spampinato,"Fast method for noise level estimation and integrated noise reduction," IEEE Trans. On Consumer Electronics., vol. 51 , No. 3, pp. 1028 - 1033, 2005.

저 자 소 개

김 창 수

- 2010년 6월 : 무한이공대학교(중국) 전자정보공학과 학사
- 2012년 8월 : 연세대학교 전기전자공학과 석사
- 주관심분야 : 비디오 디노이징

김 종 호

- 2005년 8월 : 연세대학교 전기전자공학과 학사
- 2005년 9월 ~ 현재 : 연세대학교 전기전자공학과 석박사 통합과정 재학 중
- 주관심분야 : 영상 부호화, H.264/AVC, HEVC, 영상 개선 처리

최 윤 식

- 1979년 2월 : 연세대학교 전기공학과 학사
- 1984년 5월 : Case Western Reserve 대학 시스템 공학과 석사
- 1987년 5월 : Pennsylvania 주립 대학 전기공학과 석사
- 1990년 12월 : Purdue 대학 전기공학부 박사
- 1993년 3월 ~ 현재 : 연세대학교 전기전자공학 교수
- 주관심분야 : 비디오/영상 신호 처리 및 해석