

基于 Surfacelet 变换的 3D Context 模型视频去噪新方法

肖弘智¹, 闫敬文^{1*}, 屈小波²

(1.数字信号与图像处理广东省重点实验室, 汕头大学电子工程系, 广东 汕头 515063
2.厦门大学通信工程系, 福建 厦门 361005)

Xiao Hong-zhi¹, Yan Jing-wen^{1*}, Qu Xiao-bo²

(1.Key laboratory of Digital Signal and Image Processing of Guangdong Province,
Dept. of Electronic Engineering, Shantou University, Shantou 515063, China
2. Dept. of Communication Engineering, Xiamen University; Xiamen 361005; China)

摘要: 本文提出一种基于 3D Context 模型的 Surfacelet 域视频去噪新方法。Surfacelet 变换(ST)是一种新的 3D 变换,具有多方向分解、各向异性和低冗余度等性质。本文首次采用 3DCMST,实现对视频进行 3D 分解。根据视频信号 ST 域内系数和噪声分布的特征,将 2D Context 模型拓展到 3D,构造了基于 Context 的 3D 噪声模型,实现了真正意义 3D 去噪声处理。在 ST 域内利用系数的 3D 相关性和方向性信息,按照能量分布将 ST 系数分成多个立方块,对各立方块进行能量估计和阈值估计。实验结果表明,本文算法噪声抑制效果明显优于分层 2D 去噪声方法和其它现有的 3D 方法,显著提高了去噪视频的 PSNR 值。从视觉效果来看,本文算法在去除噪声的同时,能很好的保留视频图像细节,运动物体非常平滑,有效解决传统算法中存在的拖影、闪烁等问题,尤其适合于包含剧烈运动和丰富纹理图像的视频。

关键词: 视频去噪; Surfacelet 变换; 三维 Context 模型; 方向滤波器组

中图分类号: TN911.73

文献标识码: A

A Novel Video De-noising Method with 3D Context Model Based on Surfacelet Transform

Abstract: In this paper, we propose a novel video de-noising method with 3D Context Model based on Surfacelet Transform (3DCMST). ST is a new 3D transform. It has the useful properties of directional decomposition, perfect reconstruction and low redundancy. 3DCMST is introduced to the 3D decomposition of video signal for the first time in this paper. In order to take advantage of the characteristic of the coefficients and noise distribution, the Context model is extended from 2D to 3D, which can accomplish true 3D de-noising processing and utilize the 3D space-time information of video signal adequately. The ST coefficients are divided into several parts according to their energy distribution by 3D Context model and each part has independent energy estimate and threshold. Experimental results show that the proposed method achieves better de-noising performance than other 3D or hierarchical 2D de-noising method, and improve the peak-signal-to-noise (PSNR) value remarkably. In terms of visual quality, the proposed method can effectively preserve the video detail, and the trajectory of motion object is very smooth, which is especially adequate to process the video frames with acute movement and plenty of texture.

Key words: video De-noising; Surfacelet Transform; 3D Context model; Directional Filter Banks

1 引言

视频图像在采集和传输过程中常常会受到各种噪声的影响,导致图像质量变差。视频去噪声处理是研究能有效滤除视频噪声又不模糊视频的方法。传统的视频去噪方法基本上可以按空间域、时域、频域和小波域来进行划分[1]。

由于简单的空域滤波不考虑时域信息,而帧之间同一位置的噪声存在随机性,因此容易导致滤波后相邻帧之间明显的闪烁现象。简单的时域滤波会由于运动的存在导致匹配失败或误差,出现噪声残留或“鬼影”现象。频域滤波利用低通滤波器来去除集中在高频段的噪声,不可避免会丢失部分细节信息,使视频变得比较模糊。传统的小波域去噪通常是各帧去噪,没有考虑各帧之间的运动相关性,会发生运动物体拖尾的现象。

新的视频去噪算法可分为两类。第一类算法是在传统算法的基础上增加运动检测和运动估计,利用各帧之间的运动相关性,能够较好地解决运动图像拖尾和运动边缘模糊的问题[1][2]。如果运动补偿能够完全匹配,去噪的效果就类似于静止图像的多帧去噪。但是当视频图像中存在运动轨迹模糊(比如孔径问题)或者光照剧烈变化等现象时,这类算法的效果比较差,鲁棒性不足[3]。第二类算法是把视频信号看成一个特殊的三维信号,即两个空间维和一个时间维,采用三维变换把它当成一个整体来处理。文献[4][5]对这类算法进行了研究,如文献[5]中的视频去噪算法采用三维双树复数小波变换。这类算法对视频中的运动物体进行整体处理,有效解决了前述算法中运动物体拖尾、闪烁以及算法鲁棒性不足等问题。

绝大多数的三维变换都是可分一维变换的组合,不是真正意义上的三维变换,这类变换得到的子带会混淆三维信号的方向信息,不适合处理视频。文献[5]中采用的三维双树复数小波把信号分解为6个方向,有较好的方向选择性,但是当视频运动非常剧烈、图像纹理非常丰富时,其方向数显然不足。Lexing Ying 和 Candes 等人提出的 Curvelet 变换的三维形式[6],即 3D-Curvelet,具有良好的方向选择性,但是其过大的冗余度(约40倍)使得算法效率非常低。

ST 是 Yue Lu 和 M.N.Do 在 2007 年提出的一种新的三维变换[7],具有多方向分解、各向异性、高效率的树结构滤波器组、完全重建和低冗余度等性质,非常适合于视频处理。本文在国内首先提出基于 3D Context 模型的 ST 域视频去噪算法(3DCMST),把视频信号作为特殊的三维信号,对信号的空间信息和时间信息整体处理。针对 ST 系数的特点,在阈值选取中采用 3D Context 模型。把传统的 Context 模型从二维扩展到三维后,能够充分利用信号的三维时空信息,按照能量相近与否把 ST 系数矩阵分成多个部分,每一个部分都有各自的能量估计和相应的阈值。实验表明,该方法能达到更好的视觉效果和更高的 PSNR 值,尤其适用于包含剧烈运动物体和丰富纹理图像的视频。

2 Surfacelet 变换

2.1 三维方向滤波器组

1992 年, Bamberger 和 Smith 首先提出方向滤波器组 (DFB, Directional Filter Banks) 的概念[8]。2002 年, M.N. Do 和 M. Vetterli 提出一种“真”二维的图像表示方法, 即 Contourlet 变换[9]。Contourlet 变换首先利用拉普拉斯金字塔对图像进行多尺度分解, 再通过二维方向滤波器组来实现完整的变换。

DFB 和 Contourlet 变换都只能处理二维信号, 为了处理三维甚至多维信号, Yue Lu 和 M.N.Do 在 2007 年提出一种新的多维方向滤波器组设计方法——NDFB (N-dimensional Directional Filter Banks)。本文只讨论三维情况下的 NDFB, 即 3D-DFB。图 1 为 3D-DFB 的频率分割示意图。

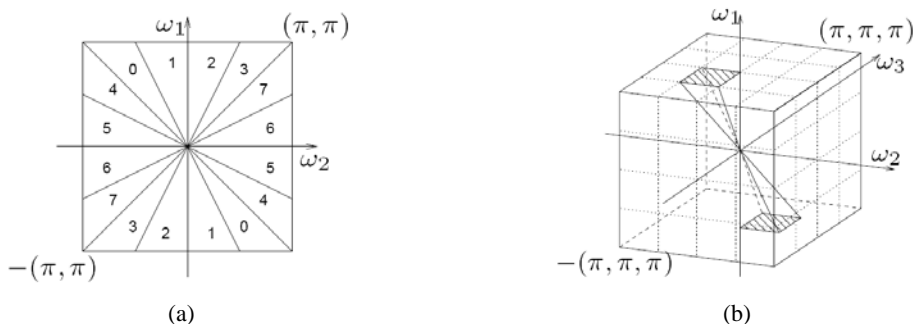


图 1. 二维和三维方向滤波器组的频率分解示意图

Fig.1 Frequency decomposition of the 2D-DFB and 3D-DFB

(a) 方向数为 8 的二维方向滤波器组的频率分解图

(b) 3D-DFB 的频率分解图

(a) Frequency decomposition of the 2D-DFB with 8 directions

(b) Frequency decomposition of the 3D-DFB

3D-DFB 的结构图如图 2 所示, 输入 $x[n]$ 为一个三维信号, 其三个轴分别为 n_1, n_2, n_3 。

$x[n]$ 首先通过一个沙漏滤波器 $p_{0,0}^{0,0}(\omega)$, 该滤波器由三通道非十进滤波器组组成。输出 $y[n]$

继续通过一个二维方向滤波器组 $IRC_{12}^{(l_2)}$, 对 $x[n]$ 的 (n_1, n_2) 平面进行处理, 得到 2^{l_2} 个子块

$z_i[n]$, 其中 $0 \leq i \leq 2^{l_2}$ 。 $z_i[n]$ 再分别通过另一个二维方向滤波器组 $IRC_{13}^{(l_3)}$, 对 $x[n]$ 的

(n_1, n_3) 平面进行处理。最后总共得到 $2^{l_2+l_3}$ 个方向子块 $z_{i,j}[n]$, 其中 $0 \leq i \leq 2^{l_2}$ 且

$0 \leq j \leq 2^{l_3}$ 。

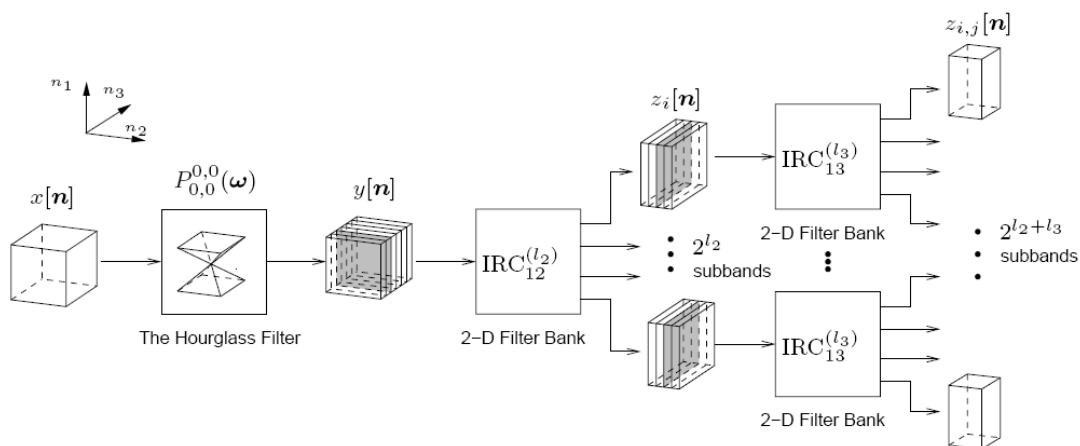


图 2. 3D-DFB 的结构图

Fig.2 The structure of the 3D-DFB

2.2 多尺度分解

因为 3D-DFB 只能处理信号的高频部分，在信号通过 3D-DFB 之前，需要先进行多尺度分解。在 Contourlet 变换中，图像的多尺度分解由拉普拉斯金字塔 (LP, Laplacian Pyramid) 实现。而在 ST 中，采用一种新的塔式结构来实现，如图 3 所示。图中 $D_i(\omega)(i = 0,1)$ 表示高通滤波器， $L_i(\omega)(i = 0,1)$ 表示低通滤波器， $S(\omega)$ 为反混频滤波器，可以避免上抽样操作带来的混频现象。多尺度分解和 3D-DFB 结合就构成了一个完整的 ST，其完整的结构图如图 3 所示。

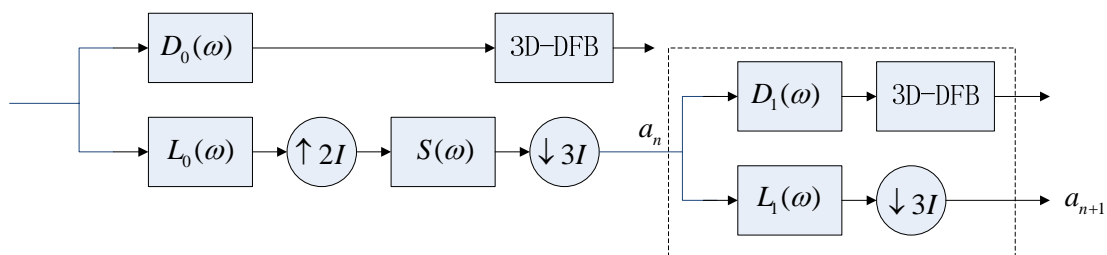


图 3. ST 结构图

Fig.3 The structure of the ST

2.3 ST 的性质

ST 具有多方向分解、高效率的树结构滤波器组、完全重建和低冗余度等性质，它能够用不同尺度、不同频率的方向子块准确地捕获三维信号中的面状奇异，并且变换后的 ST 系数能量非常集中，或者说 ST 对于曲面有更“稀疏”的表示。图 4 为对一个三维图形进行 ST 分解后，选取不同的方向子块进行重建，重建后的图形如图 4 (b) - (d) 所示。从图中可以看出，ST 可以选取不同的方向，并进行完全重建。

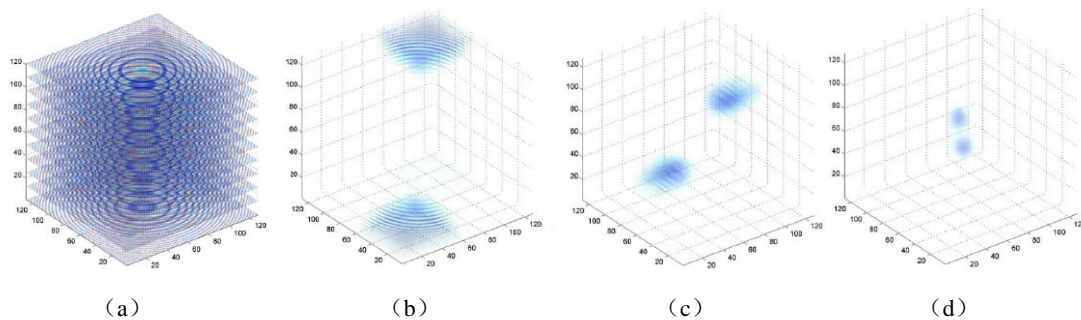


图 4. ST 的不同方向子块

Fig.4 The different directional subblocks of ST

在实际应用中，还要考虑到算法的运算效率。表 1 为针对不同信号，ST 的程序运行时间。运行环境为 CPU P4 2.8GHz，内存 1.5GB，编译环境为 Visual C++6.0。

表 1. ST 运行时间 (秒)

Tab.1 The running time of ST (second)

数据大小	层数×方向数			
	3×1	3×4	3×16	3×64
$N \times N \times N$	3×1	3×4	3×16	3×64
128×128×128	0.73	0.89	1.00	1.11
192×192×192	2.88	3.78	4.03	4.33
256×256×256	8.58	8.82	8.87	9.71

3 3DCMST 视频去噪算法

3.1 ST 系数分析

在阈值去噪算法中，阈值的选取最为关键。比如小波域图像去噪，常根据小波分解后各层系数的分布情况而选择不同的阈值。这种自适应阈值去噪算法在去除噪声的同时还能较好地保留图像细节，因而得到了广泛的应用。在讨论如何确定 3DCMST 去噪算法的阈值之前，先对 ST 系数进行分析。

图 4 是对 Mobile 视频进行 ST 分解后，最精细层和最粗糙层的某方向子块的系数分布，图中 Y 表示系数值， N 表示系数个数。从图中可以看出，视频经 ST 分解后的系数变得较为稀疏，能量集中在有限的变换域系数上，其余大部分变换域系数的幅值接近于零。

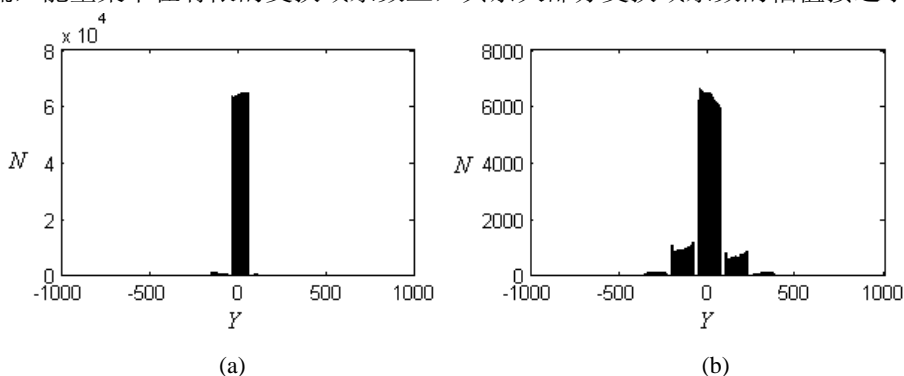


图 4. ST 系数分布

Fig.4 The distribution of ST coefficients

- (a) 最精细层 (b) 最粗糙层
- (a) The finest scale (b) The coarsest scale

图 5 是高斯白噪声进行 ST 分解后，任意选取两个方向子块的系数分布。可以看出，高斯白噪声经 ST 分解后仍是白噪声，能量均匀地分布在所有的变换域系数上。

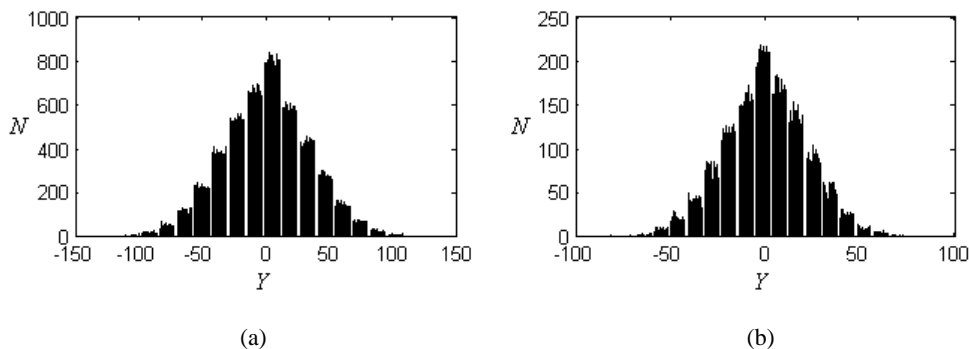


图 5. 高斯白噪声经 ST 分解后的系数分布

Fig.5 The coefficient's distribution of Gaussian Noise after ST decomposition

- (a) 最精细层 (b) 最粗糙层
(a) The finest scale (b) The coarsest scale

3.2 3DCMST 去噪算法的阈值确定

假设噪声模型为 $y(i, j, k) = f(i, j, k) + n(i, j, k)$ ，其中 $f(i, j, k)$ 为理想的不含噪声的视频经 ST 分解后的系数， $y(i, j, k)$ 是观察到的 ST 系数。 $n(i, j, k)$ 是服从独立同分布的高斯噪声，即 $n(i, j, k) \sim N(0, \sigma^2)$ ，均值为零， σ 为噪声等级。若 $f(i, j, k)$ 和 $n(i, j, k)$ 不相关，从 $y(i, j, k)$ 中恢复 $f(i, j, k)$ 就是一个经典的统计估计问题。

通过分析视频经 ST 分解后的所有方向子块，发现其系数分布都近似服从近似高斯分布或者近似拉普拉斯分布[10]，即 GGD (General Gaussian Distribution) 分布。Grace Chang 等人在[11]中专门对这种 GGD 模型的阈值做了探讨，并给出了阈值表达式：

$$T(\alpha) = \sigma_n^2 \sqrt{\alpha} \quad (2)$$

其中 σ_n 是噪声标准差， α 是信号的标准差，即 $\sqrt{\alpha} = \frac{1}{\sigma_x}$ ，所以式 (2) 或以重写为 (3)：

$$T(\alpha) = r \times \sigma_n^2 / \sigma_x \quad (3)$$

其中 r 是可调节的参数，经多种测试发现 $r = \sqrt{2}$ 时能取得最好的去噪效果。

从式 (3) 可知，如何估计噪声标准差 σ_n 和信号标准差 σ_x 就成为阈值选取的关键。文献[12]中，给出了静止图像中噪声标准差的估计式：

$$\sigma_n = \widehat{\sigma}_n = \text{median}(|y(i, j, k)|) / 0.6745 \quad (4)$$

在 ST 中，经过测试，发现式 (4) 的噪声估计也是可信的。

3.3 3D Context 模型建立

对含噪视频进行 ST 分解后的系数 $y(i, j, k)$ 是信号和噪声的混合矩阵，假设这个系数矩阵的方差是信号和噪声的能量和，即 $\text{var}(y) = \sigma_x^2 + \sigma_n^2$ 。

由于在同一方向子块中 ST 系数变化范围很大，所以用求整个矩阵方差减去噪声能量来估计信号的标准差，显然没有考虑到每个 ST 系数的情况，它只是一种平均的方法。针对这个问题，本文提出一种 3D Context 模型。Context 模型最早由文献[10]提出，主要针对静止图像处理，本文算法把传统的 Context 模型从 2D 扩展到 3D，能够充分利用信号的三维时空信息，把能量相近的 ST 系数放在一起，求出这一部分系数的方差作为这一部分系数信号和

噪声的能量和。将 ST 系数矩阵分成多个部分，每一个部分都有各自的能量估计和相应的阈值。

定义每一个 ST 系数的 context 值为 $Z(i, j, k)$ ：

$$Z(i, j, k) = \frac{1}{N} \sum_{(i, j, h) \in S} |y(i, j, k)| \quad (5)$$

S 为 ST 分解后方向子块中系数 $y(i, j, k)$ 的邻域，一般取 $3 \times 3 \times 3$ 的方块， N 为 S 内的系数个数。图 6 (b) 为 N 取 $3 \times 3 \times 3$ 时系数 $y(i, j, k)$ 的 3D Context 模型图。

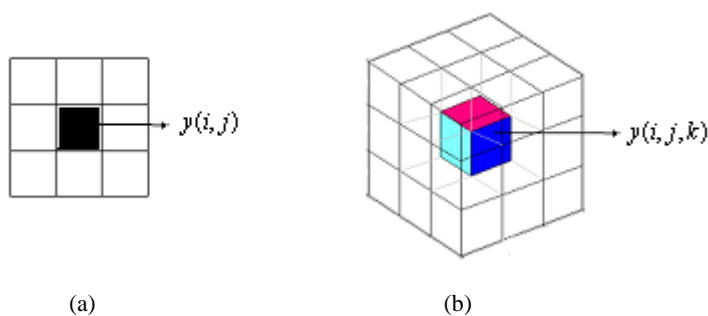


图 6. 2D 和 3D Context 模型示意图

Fig.6 The 2D and 3D Context model

(a) 系数 $y(i, j)$ 的 2D Context 模型 (b) 系数 $y(i, j, k)$ 的 3D Context 模型

(a) The 2D Context model of $y(i, j)$ (b) The 3D Context model of $y(i, j, k)$

作为一个例子，下面考察用 3D Context 模型是否能衡量变换系数能量接近情况。如图 7，对含噪视频进行 ST 分解后，任选其中一个方向子块，发现系数越小它们所对应的 context 值的分布就越密，而系数越大则 context 值的分布就越稀疏。由图 7 可见 $Z(i, j, k)$ 值越小时，分布的越密； $Z(i, j, k)$ 值越大，则分布越稀疏。因此当 context 值小于 80 时，context 值每隔 5 分一级，如图 6 窗口 a；当 context 值大于 80 时则增大分级间隔，如图 7 中窗口 b。

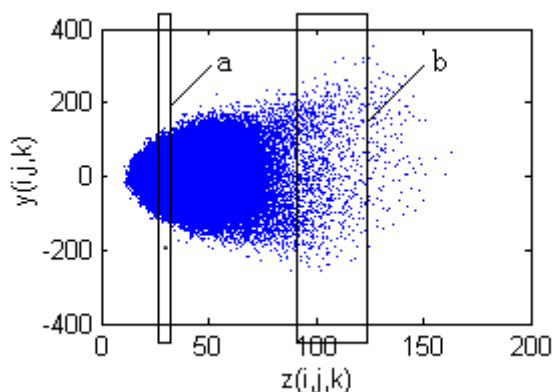


图 7. ST 方向子块系数与其对应 context 值的分布关系图

Fig.7 The distributing of ST directional block and context value

3DCMST 视频去噪算法具体步骤如下：

- 1) 对含噪视频进行 ST 分解；

2) 用 3D Context 模型将 ST 分解得到的各方向子块分成多个部分, 分别计算每一部分的方差 $\text{var}(y)$ 和噪声标准差 σ_n ;

3) 计算每一部分的信号方差 $\sigma_x^2 = \text{var}(y) - \sigma_n^2$ 及其相对应的阈值 $T(\alpha) = \sqrt{2}\sigma_n^2 / \sigma_x$;

4) 利用步骤 (3) 中得出的阈值分别对 ST 分解后各方向子块进行去噪;

5) 对去噪后的 ST 系数进行重构。

4 实验结果和分析

4.1 实验环境

本文采用 SIF 格式的标准视频序列进行实验, 包括小幅度运动的 Mobile、大幅度运动的 Football 和图像纹理丰富的 Suise。首先对视频序列添加不同大小方差的零均值高斯白噪声, 然后采用基于小波去噪算法 (WT)、基于双树复数小波去噪算法 (DTWT)、基于三维 Curvelet 去噪算法 (CT)、基于硬阈值的 ST 去噪算法 (ST) [3]以及本文的基于三维 Context 模型的 ST 去噪算法 (3DCMST) 分别对含噪视频进行去噪。实验中, 各种算法的变换均只分解三层。为了考察分解层数对本文算法的影响, ST 分别采用三层和四层的分解, 对应 3D-DFB 的方向数分别为 64、16、4 和 64、64、16、4, 只对最精细两层的方向子块进行去噪处理。

4.2 实验结果分析和讨论

表 2 给出了四种算法去噪后 PSNR 值比较, 图 8~9 为 Football 和 Suise 视频去噪效果图。从实验结果可以得到以下结论:

表 2. 四种算法去噪结果 (dB)

Tab.2 The de-noising result of the four algorithms

σ_n	Mobile			Football			Suise		
	30	40	50	30	40	50	30	40	50
Noisy	18.59	16.09	14.15	18.58	16.09	14.14	18.58	16.09	14.15
WT	23.12	22.75	21.28	21.62	20.08	18.64	26.31	24.82	23.34
CT	23.54	23.19	22.86	23.16	22.11	21.14	26.79	25.31	24.76
DTWT	24.56	23.41	22.58	23.67	22.14	20.90	27.83	26.35	24.78
ST	25.86	24.23	23.15	24.98	23.58	21.98	28.69	26.97	25.56
3DCMST(3)	26.48	25.08	23.89	26.31	24.89	23.51	29.78	28.16	26.88
3DCMST(4)	27.54	26.68	25.49	27.87	26.43	25.03	31.41	29.72	28.44

1) 本文算法在 PSNR 值上要明显优于其它四种算法, 比如 Mobile 序列, 本文算法比其它算法的 PSNR 提高了约 2~3dB。基于硬阈值的 ST 算法的去噪效果已经优于传统算法, 但无论是 PSNR 还是视觉效果, 其效果都不如本文算法。

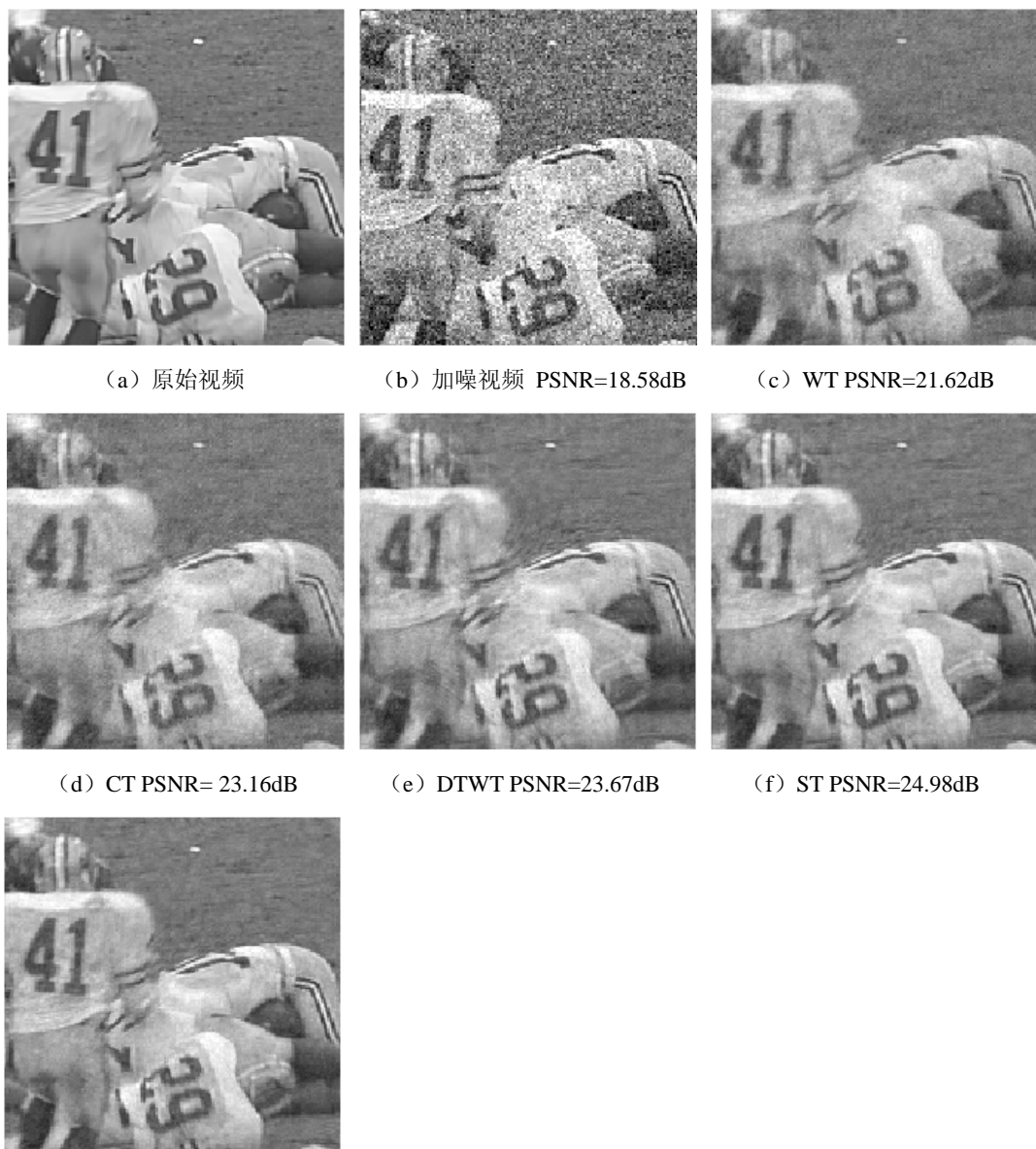
2) 本文算法在处理包含剧烈运动和丰富纹理图像的视频时, 效果更加明显。三个视频中, Mobile 属于小幅度运动序列, 视频中的所有物体作整体平移运动, 各帧之间差异较小; 而 Football 属于大幅度运动序列, 其中包含的都是杂乱无章的大面积的运动; Suise 序列的运动幅度较小, 但是包含头发、皮肤等纹理丰富的图像。从表 2 可以看出, 本文算法在处理

Mobile 序列时, PSNR 值比 DTWT 算法提高了约 1.5~2dB, 而在处理 Football 和 Suise 序列时, PSNR 值比 DTWT 算法提高了约 2~2.7dB。

3) 从图 8~9 的去噪效果图可以看出, 本文算法能够更好地恢复图像细节信息, 视觉效果明显优于其他算法。限于客观条件, 本文中只能给出视频中某一帧的效果图, 实际上, 当播放完整的视频序列时, 本文算法的视觉效果会更加的突出, 不存在其他算法中常见的拖影、闪烁等现象。

4) 当增加分解层数时, 本文算法的性能会得到较大提高。从表 2 可知, 当 ST 的分解层数由 3 层增加为 4 层时, PSNR 值提高了约 1~1.5dB。而其它三种算法, 尤其是 WT 和 DTWT, 在增加分解层数时, 去噪效果只是略有提升。实际上, 因为 WT 和 DTWT 的分解方向数较少, 在分解层数达到一定程度时, 再继续增加分解层数性能提高非常小。

本文在国内首先提出基于 ST 的视频去噪算法, 针对 ST 系数特点, 在阈值选取中采用 3D Context 模型。实验结果表明, 本文算法显著提高了去噪视频的 PSNR 值, 能够很好的保留视频图像细节, 运动物体非常平滑, 不存在传统算法中的拖影、闪烁等现象。因 ST 具备的多方向性、各向异性和低冗余度等性质, 使得其在视频增强、运动目标提取、三维医学影像处理和三维数据压缩等方面也有很好的应用前景。这将是我们的下一步研究工作的重点。



(g) 3DCMST PSNR=26.31dB

图 8. Football 去噪效果图

Fig.8 The de-noising result of Football



(a) 原始视频

(b) 加噪视频 PSNR=18.58dB

(c) WT PSNR=26.31dB



(d) CT PSNR= 26.79dB

(e) DTWT PSNR=27.83dB

(f) ST PSNR=28.69dB



(g) 3DCMST PSNR=29.78dB

图 9. Suisse 去噪效果图

Fig.9 The de-noising result of Suisse

参考文献:

- [1] 陈勇兵,郭晓云,沈未名. 一种鲁棒性的基于运动估计的自适应时空域视频去噪算法. 计算机应用. 2006,26(8):1882-1887.
- [2] F. A. Mujica, J.-P. Leduc, R. Murenzi, et al. A new motion parameter estimation algorithm based on the continuous wavelet transform. IEEE.Trans. Image Processing., 2000,9(4):873-888.
- [3] Yue Lu, M.N.Do. Video Processing Using The 3-Dimensional Surfacelet Ttransform. Invited Paper, Fortieth Annual Asilomar Conference on Signals, Systems and Computers, Pacific Grove, 2006:883-887.
- [4] E.Chang, A.Zakhor. Subband video coding based on velocity filters. IEEE.Trans. Circuits and Systems, 1992.
- [5] I.W.Selesnick, K.Y.Li. Video denoising using 2D and 3D dual-tree complex wavelet transforms. Proc. of SPIE conference on Wavelet Applications in Signal and Image Processing X, San Diego, USA, August2003.
- [6] Lexing Ying, Laurent Demanet, Emmanuel Candes. 3D Discrete Curvelet Transform. Applied and Computational Mathematics,California Institute of Technology,2005.
- [7] Yue Lu, M.N.Do. Multidimensional directional filterbanks and surfacelets. IEEE Trans. Image Processing., 2007,16(4):918:931.
- [8] R.H.Bamberger, M.J.T.Smith. A filter bank for the directional decomposition of images: Theory and design. IEEE Trans. Signal Process., 1992,40(4):882-893.
- [9] M.N.Do, M.Vetterli. The contourlet transform: an efficient directional multiresolution image representation. IEEE Trans. Image Processing.,2005,14(12).
- [10] S.G.Chang, B.Yu, M.Vetterli. Spatially Adaptive Wavelet Thresholding with Context Modeling for Image Denoising. IEEE Trans. Image Processing, 2000,9(9):1522:1531.
- [11] S.G.Chang, B.Yu, M.Vetterli. Image denoising via lossy compression and wavelet thresholding. IEEE Trans. Image Processing, 2000,9:1532-1546.
- [12] David L, Donoho. De-noising by thresholding. IEEE.Trans. Information Theory, 1995,41: 613-627.