

## 基于运动矢量多级分析的视频全局运动估计

赵亚湘<sup>\*①</sup> 樊晓平<sup>①②</sup> 刘少强<sup>①</sup>

<sup>①</sup>(中南大学信息科学与工程学院 长沙 410075)

<sup>②</sup>(湖南财政经济学院信息管理系 长沙 410205)

**摘要:** 基于运动矢量场的视频全局运动估计相较于基于像素的估计方法具有较低的计算复杂度,因而广泛应用于视频分割及视频压缩等领域中。然而噪声和前景目标等外点区域的存在,降低了全局运动估计的准确性。为了提高全局运动估计的准确度,该文提出一种基于运动矢量多级分析的全局运动估计算法,该算法根据局部运动与全局运动的运动特性差异自适应地滤除前景目标区域,由邻域矢量间相似性度量检测出纹理平滑周期区域,最后滤除孤立的噪声区域,由滤波得到的内点区域求解全局运动参数。实验结果表明,该方法能有效地滤除外点区域,提高全局运动估计的准确性。

**关键词:** 图像处理; 全局运动估计; 运动矢量场; 多级外点滤除; 运动参数模型

中图分类号: TP391.41

文献标识码: A

文章编号: 1009-5896(2012)07-1538-05

DOI: 10.3724/SP.J.1146.2011.01269

## Global Motion Estimation Based on the Multi-stage Analysis of Motion Vectors

Zhao Ya-xiang<sup>①</sup> Fan Xiao-ping<sup>①②</sup> Liu Shao-qiang<sup>①</sup>

<sup>①</sup>(School of Information Science and Engineering, Central South University, Changsha 410075, China)

<sup>②</sup>(Department of Information Management, Hunan University of Finance and Economics, Changsha 410205, China)

**Abstract:** Global motion estimation based on motion vector field has lower complexity than pixel-based method, so it is widely used in video segmentation and compression. However, outlier motion vectors, caused by image noise or foreground objects, reduce the accuracy of motion vector-based global motion estimation. In this paper, a global motion estimation algorithm based on the motion vector multi-stage processing is proposed to improve the estimation accuracy. The proposed method adaptively removes foreground objects by comparing the motion characteristics differences between the local motion and global motion area. For each block considered, the motion similarity between the neighboring blocks is exploited to detect the cycle smooth area. The isolated noise area is also filtered out. Finally, the inlier motion vectors are used to estimate the global motion parameters. Experimental results show that the proposed scheme filters effectively outlier motion vectors and improves the accuracy of global motion estimation.

**Key words:** Image processing; Global motion estimation; Motion vector field; Multi-stage outlier filter; Motion parameter model

### 1 引言

视频序列中的全局运动是由于摄像机移动或镜头缩放而引起的图像的整体变化,局部运动是前景目标相对于摄像机的运动而引起的局部像素的变化。全局运动估计的目的是要从视频序列中找出摄像机运动的规律,目前已广泛应用于视频分割<sup>[1,2]</sup>、视频压缩<sup>[3]</sup>及视频稳像<sup>[4]</sup>等领域中。

全局运动估计可以在像素域内实现<sup>[5,6]</sup>,也可在

压缩域内实现<sup>[7,8]</sup>。基于压缩域方法利用视频压缩码流中的运动矢量,相较于基于像素的方法大大减少了计算量,因而得到越来越广泛的应用<sup>[9]</sup>。但压缩码流中的运动矢量,由于局部运动的影响,或由于运动估计时部分运动矢量不能真实地反映实际运动方向<sup>[10]</sup>,会使全局运动估计产生偏差,从而降低全局运动估计的准确性。因此,有必要在全局运动估计前对运动矢量场进行处理,剔除局部运动和背景中不可靠运动,提高全局运动估计的准确性。

文献[11]通过比较中心块与其8邻域块运动矢量间的幅度差值来剔除估计误差较大的运动矢量,从

2011-12-02 收到, 2012-02-29 改回

国家自然科学基金(60870010, 60776834)资助课题

\*通信作者: 赵亚湘 yxzhaocs@sina.com

而消除局部运动的影响，该方法采用固定阈值，不适合运动对象较小的场景。Chen等人<sup>[12]</sup>在此基础上提出了一种级联的运动矢量外点滤除方法，该方法将运动矢量的8邻域分成3种不同的滤波区域，对每一区域分别计算中心块与邻域矢量间的幅度及相位差值来实现滤除外点。该方法由于只考虑了8邻域范围内差异，不能有效地去除前景对象的整体运动，对于大运动目标具有一定的局限性。Chung等人<sup>[13]</sup>利用块匹配估计时生成的绝对误差和(SAD)来分析运动矢量的可靠性，因为需要在块匹配运算时找出多个极小值点，因此对块匹配算法和其应用都有一定的限制。

本文结合运动的全局性和局部性的特点，提出了一种基于运动矢量多级分析的全局运动估计算法，该算法根据运动矢量的分布特性分级处理，滤除局部运动区域和运动矢量不可靠区域，从而有效地求解全局运动参数。

## 2 运动矢量的多级分析

视频序列中的运动通常由占较大比例的背景像素的全局运动和前景目标的局部运动组成，它们具有不同的运动特征。同时，序列中的纹理平滑周期区域由于没有足够的梯度信息或者由于纹理相似性，在块匹配运动估计时常常会产生多个匹配块，从而使估计出的运动矢量与实际的运动方向存在偏差。全局运动估计时，要尽量利用可信背景区域，而剔除不可信任区域，包括前景目标区域和平滑周期区域。将背景可信区域作为内点区域参与运动参数估计，而将后两类区域作为外点区域予以滤除。

### 2.1 前景目标及噪声区域滤除

根据背景运动是图像中大部分像素的运动，前景目标运动是图像中局部像素的运动的特点，同时根据背景区域具有基本一致运动的特性，本文应用运动矢量场提出一种有效的局部运动区域滤除方法。

由运动矢量场 $\{\mathbf{MV}_i\}$ ，统计每一运动矢量 $\mathbf{MV}_i$ 出现的次数，建立2维运动矢量直方图 $H(\mathbf{MV}_i)$ ，将出现次数最多的运动矢量 $\mathbf{MV}_m$ 作为初始的全局运动方向。因为图像像素的运动可以看成是全局运动和像素的局部运动的叠加，因此将每一子块运动矢量 $\mathbf{MV}_i$ 相对于初始全局方向 $\mathbf{MV}_m$ 进行矢量校正，得到 $\mathbf{CV}_i = \mathbf{MV}_i - \mathbf{MV}_m$ 。因为 $\mathbf{MV}_m$ 反映了全局运动大致方向，因此与全局运动基本一致的背景区域经过校正后其对应的矢量幅值相对较小，而场景中运动对象或者受噪声干扰块对应的矢量幅值相对较大。因此，计算 $\mathbf{CV}_i$ 的模值 $m\mathbf{CV}_i = |\mathbf{CV}_i|$ ，根据

$m\mathbf{CV}_i$ 利用均值聚类算法对运动矢量场进行分类。定义准则函数为

$$J = \sum_{i=1}^c \sum_{m\mathbf{CV}_i \in G_i} |m\mathbf{CV}_i - \mu_i|^2 \quad (1)$$

其中 $G_i$ 为类 $i$ 的集合， $\mu_i$ 为类 $i$ 的均值。本文令 $c=2$ ，即将运动矢量分为两类。设分割阈值为 $T$ ， $M = \max\{m\mathbf{CV}_i\}$ ， $H(i)$ 为具有 $m\mathbf{CV}_i$ 值的个数，聚类中心定义为 $\mu_1 = T/2$ ， $\mu_2 = (T + M)/2$

则式(1)变为

$$J = \sum_{m\mathbf{CV}_i=0}^{T-1} H(i)(m\mathbf{CV}_i - \mu_1)^2 + \sum_{m\mathbf{CV}_i=T}^M H(i)(m\mathbf{CV}_i - \mu_2)^2 \quad (2)$$

令 $\partial J / (\partial T) = 0$ ，可近似得到阈值 $T$ 为

$$T = \frac{2}{\sum_{m\mathbf{CV}_i=0}^M H(i)} \left[ \sum_{m\mathbf{CV}_i=0}^M H(i) \cdot m\mathbf{CV}_i - \frac{M}{2} \sum_{m\mathbf{CV}_i=M/2}^M H(i) \right] \quad (3)$$

利用 $T$ 对运动矢量场进行分类，若 $m\mathbf{CV}_i < T$ ，则认为对应块是背景块，划入集合 $G_1$ 。若 $m\mathbf{CV}_i \geq T$ ，则认为对应块是运动对象块，划入集合 $G_2$ 。得到内点掩膜图：

$$\text{InMap}_i = \begin{cases} 1, & \mathbf{MV}_i \in G_1 \\ 0, & \mathbf{MV}_i \in G_2 \end{cases}$$

### 2.2 纹理平滑周期区域滤除

经过上一级处理后，运动对象区域得到了较好的滤除，但对于纹理平滑周期区域，由于它们的运动矢量与背景区域运动矢量的幅值差异并不明显，它们经常被划入内点集合 $G_1$ 中，影响全局运动估计的精度，因此有必要进一步处理。相对于矢量幅值差异的变化，矢量相位间的差值更能反映出平滑周期区域与可靠背景区域间的差异。

对于集合 $G_1$ 中的每一子块 $i$ ，定义它与周围8邻域子块间的矢量相似值 $S_{ij}$ ：

$$S_{ij} = \cos(\mathbf{MV}_i, \mathbf{MV}_j) = \frac{\mathbf{MV}_i \cdot \mathbf{MV}_j}{|\mathbf{MV}_i| |\mathbf{MV}_j|} \quad (4)$$

式中 $j \in N_i$ ， $N_i$ 为子块 $i$ 的8邻域集合。

当中心块运动矢量 $\mathbf{MV}_i$ 为零矢量时，式(4)改写为 $S_{ij} = \frac{\mathbf{MV}_j \cdot \mathbf{x}_j}{|\mathbf{MV}_j|}$ 。

当 $S_{ij}$ 大于某一阈值 $T_s$ 时，认为 $\mathbf{MV}_i$ 与 $\mathbf{MV}_j$ 间差异较小，为相似的运动矢量，令对应的 $d_{ij} = 1$ ，否则 $d_{ij} = 0$ 。计算 $d_i = \sum_{j \in N_i} d_{ij}$ ，定义子块权值 $w_i = 1 / \sum_{j \in U_i} e^{-\alpha d_i}$ ， $U_i$ 为以子块 $i$ 为中心的 $3 \times 3$ 区

域,  $\alpha$  为可调收敛因子。计算  $w_i d_i$ , 并对其进行排序, 取  $T\%$  范围内的较大值作为内点值, 对应块设置为内点块, 保留在集合  $G_1$  中, 其余块作为外点划入集合  $G_2$  中。调整相应的掩膜  $\text{InMap}_i$  值。

### 2.3 孤立内点区域滤除

全局运动估计时, 我们强调的是参与估计块的运动矢量的可靠性, 因此应尽量滤除不可信块。经过前两级处理后, 运动对象和平滑周期区域得到了有效的滤除, 但是内点区域中还存在着一些孤立块, 它们或是运动对象边缘, 或是噪声块, 它们被外点区域包围, 对于这样的块, 我们也把它划入外点集合  $G_2$  中。即对于  $G_1$  中每一子块  $i$ , 计算其 8 邻域内的掩膜值  $\text{InMap}$  和, 即  $\text{sum} = \sum_{j \in N_i} \text{InMap}_j$ 。当  $\text{sum} < 2$  时, 将子块  $i$  划入集合  $G_2$  中, 并调整相应的  $\text{InMap}_i$  值。

## 3 全局运动参数模型及估计

### 3.1 全局运动参数模型

全局运动可以用运动参数模型来描述, MPEG-7 中规定了几类摄像机的运动模型<sup>[14]</sup>, 它们分别是 2 参数平移, 4 参数几何, 6 参数仿射和 8 参数透射模型。其中 8 参数模型是最基本的一种, 其它参数模型可以由 8 参数模型简化得到。它表示为

$$\left. \begin{aligned} x'_i &= \frac{m_0 + m_1 x_i + m_2 y_i}{m_6 x_i + m_7 y_i + 1} \\ y'_i &= \frac{m_3 + m_4 x_i + m_5 y_i}{m_6 x_i + m_7 y_i + 1} \end{aligned} \right\} \quad (5)$$

式中  $(x_i, y_i), (x'_i, y'_i)$  分别是像素点在当前帧和参考帧中的坐标,  $(m_0, \dots, m_7)$  是要估计的模型参数。

### 3.2 基于运动矢量的全局运动估计

定义  $x, y$  方向上的运动误差为

$$\left. \begin{aligned} \text{ex}_i &= \mathbf{M}\mathbf{V}x_i - \mathbf{M}\mathbf{V}x'_i \\ \text{ey}_i &= \mathbf{M}\mathbf{V}y_i - \mathbf{M}\mathbf{V}y'_i \end{aligned} \right\} \quad (6)$$

其中  $\mathbf{M}\mathbf{V}_i = (\mathbf{M}\mathbf{V}x_i, \mathbf{M}\mathbf{V}y_i)$  为视频帧中第  $i$  个子块的运动矢量,  $(\mathbf{M}\mathbf{V}x'_i, \mathbf{M}\mathbf{V}y'_i) = (x'_i - x_i, y'_i - y_i)$  为预测运动矢量。 $(x_i, y_i)$  为第  $i$  个子块中心像素的坐标,  $(x'_i, y'_i)$  为由式(5)求得的子块中心像素坐标。

则平方误差函数为

$$\begin{aligned} e_i^2 &= \text{ex}_i^2 + \text{ey}_i^2 = (\mathbf{M}\mathbf{V}x_i - x'_i + x_i)^2 \\ &\quad + (\mathbf{M}\mathbf{V}y_i - y'_i + y_i)^2 \end{aligned} \quad (7)$$

参数  $(m_0, \dots, m_7)$  的选择应使目标函数  $E = \sum_i e_i^2$  达到最小, 即

$$(m_0, \dots, m_7) = \arg \min_{(m_0, \dots, m_7)} \sum_i e_i^2 \quad (8)$$

实际参数计算时, 我们剔除局部运动块和背景

中不可信块, 即外点块, 将内点掩膜代入式(8), 得  $(m_0, \dots, m_7) = \arg \min_{(m_0, \dots, m_7)} \sum_i \text{InMap}_i \cdot e_i^2$ 。

利用 Newton-Raphson 算法求得参数迭代公式为<sup>[15]</sup>

$$\mathbf{m}^{(k+1)} = \mathbf{m}^{(k)} + \Delta \mathbf{m} \quad (9)$$

式中  $\mathbf{m}^{(k+1)}, \mathbf{m}^{(k)}$  分别是第  $k+1, k$  次迭代时的估计参数,  $\Delta \mathbf{m}$  为参数调节项,  $\Delta \mathbf{m} = \mathbf{H}^{-1} \mathbf{b}$ 。其中  $\mathbf{H}$  是 Hessian 矩阵,  $\mathbf{b}$  是梯度向量。它们中的各项元素分别近似为<sup>[6]</sup>

$$\left. \begin{aligned} H_{kl} &= \frac{1}{2} \sum_{i \in S_1} \frac{\partial^2 e_i^2}{\partial m_k \partial m_l} \approx \sum_{i \in S_1} \frac{\partial \text{ex}_i \partial \text{ex}_i}{\partial m_k \partial m_l} + \sum_{i \in S_1} \frac{\partial \text{ey}_i \partial \text{ey}_i}{\partial m_k \partial m_l} \\ b_k &= -\frac{1}{2} \sum_{i \in S_1} \frac{\partial e_i^2}{\partial m_k} \approx -\sum_{i \in S_1} \text{ex}_i \frac{\partial \text{ex}_i}{\partial m_k} - \sum_{i \in S_1} \text{ey}_i \frac{\partial \text{ey}_i}{\partial m_k} \end{aligned} \right\} \quad (10)$$

## 4 实验结果及分析

为了评价本文算法的有效性, 我们将本文算法(记为 MS\_GME)与文献[7](GD\_GME)、文献[11](FLT\_GME)、文献[12](CAS\_GME)和文献[16](RNS\_GME)中的基于运动矢量全局运动估计方法进行了综合对比测试, 选用峰值信噪比(PSNR)和计算时间作为性能指标进行比较。采用 6 个标准测试序列: coastguard, garden, stefen, foreman, waterfall 和 tempete。其中前 3 个序列中摄像机基本作平移运动, foreman 序列含有较大比例的前景对象, waterfall, tempete 序列包含镜头的缩放及小的运动对象。

为了得到较准确的参考运动矢量, 我们采用文献[7,12]中同等的测试条件, 使用全搜索块匹配算法估计块运动矢量, 选用块大小为  $8 \times 8$ , 搜索范围为  $32 \times 32$ , 根据文献[12]中对不同焦距、焦距变化率、镜头旋转角度等的综合测试, 选取阈值  $T_s$  为 0.98, 内点比例阈值  $T$  为 70, 收敛因子  $\alpha = 1/2$ 。采用 8 参数透射模型, 将参数递归次数设为 1, 利用 Matlab 7.0 在 PC 机上进行了仿真实验。

图 1 采用峰值信噪比(PSNR)对两个序列 (foreman, tempete) 各帧进行了 5 种全局运动估计算法的比较, 给出了应用这 5 种方法分别进行全局运动估计、补偿差分后的 PSNR, 图像格式为 CIF 格式。从图 1 中可看出, 本文方法与 RNS\_GME 方法在两个序列中的结果相近, PSNR 较高。FLT\_GME 方法效果最差, 在两个序列中 PSNR 都较低。GD\_GME 方法在 tempete 序列中的结果较差, CAS\_GME 方法对于 foreman 序列所得的 PSNR 相对较低。

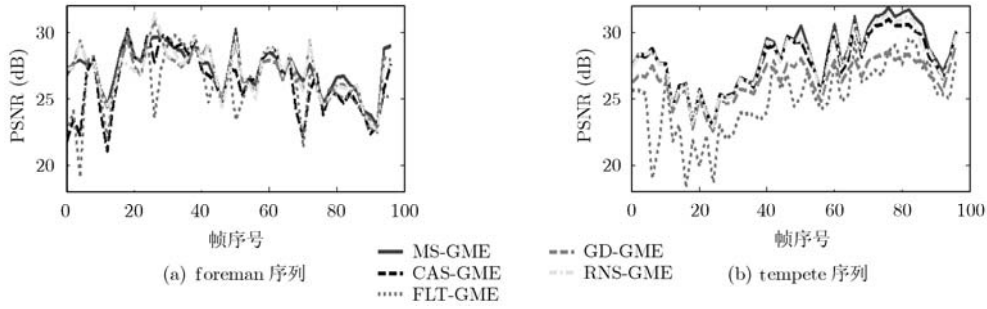


图 1 序列帧峰值信噪比较

表 1 给出了多个标准测试序列用 5 种方法求出的平均 PSNR。表 2 给出了 5 种方法对不同序列估计全局运动参数时平均每帧所花费的处理时间，包括运动矢量多级滤波、运动参数估计、补偿等的处理时间。

从表 1，表 2 中可看出，几种方法中 FLT-GME 的计算速度最快，但其信噪比性能最差，特别是在 waterfall 和 tempete 等含有较小运动对象的场景中。CAS\_GME 算法的速度次之，但对于场景中前景目标比例较大时，如 foreman 序列，估计结果稍差。本文算法平均 PSNR 最高，平均计算时间

略高于 CAS\_GME，远低于 RNS-GME，且对不同的序列均能获得较稳定的性能。RNS\_GME 算法性能与本文方法接近，但计算时间却远大于本文算法。

### 5 结束语

本文通过对运动矢量的多级分析提出了一种基于运动矢量的全局运动估计方法，该方法根据不同区域运动矢量的特点，有效地分离出前景目标、纹理平滑周期及噪声等影响全局运动估计准确性的外点区域。通过对不同视频序列的测试，本文方法均能取得较好的峰值信噪比和较短的计算时间，可对全局运动进行快速而准确的估计。

表 1 不同序列采用不同方法得到的平均 PSNR

序列 (Y 分量)	PSNR (dB)				
	MS-GME	CAS-GME	FLT-GME	GD-GME	RNS-GME
coastguard(352×288)	26.853	26.776	26.806	26.543	26.956
foreman(352×288)	27.326	26.316	26.123	27.094	27.161
waterfall(352×288)	35.095	34.859	24.247	34.709	35.484
tempete(352×288)	28.051	27.829	24.978	26.478	27.854
stefen(352×288)	24.864	24.628	22.160	24.508	24.743
garden(352×240)	22.315	22.187	21.436	22.298	21.871
平均	27.417	27.099	24.292	26.938	27.345

表 2 不同序列采用不同方法计算每帧所需的平均时间及比值

序列 (Y 分量)	平均计算时间 (ms)				
	MS-GME	CAS-GME	FLT-GME	GD-GME	RNS-GME
coastguard(352×288)	64.223	61.551	58.674	69.330	176.251
foreman(352×288)	62.162	60.563	56.015	68.785	237.535
waterfall(352×288)	62.545	59.155	55.145	65.352	149.064
tempete(352×288)	62.757	59.570	55.755	67.710	162.810
stefen(352×288)	63.595	60.118	56.580	67.040	207.821
garden (352×240)	63.536	60.557	55.110	69.315	187.258
平均	63.136	60.250	56.213	67.922	186.790

## 参考文献

- [1] Chen Y M and Bajic I V. A joint approach to global motion estimation and motion segmentation from a coarsely sampled motion vector field[J]. *IEEE Transactions on Circuits and Systems for Video Technology*, 2011, 21(9): 1316-1328.
- [2] Li H J, Tang J H, Wu S, *et al.* Automatic detection and analysis of player action in moving background sports video sequences[J]. *IEEE Transactions on Circuits and Systems for Video Technology*, 2010, 20(3): 351-364.
- [3] Keller Y and Averbuch A. Fast gradient methods based on global motion estimation for video compression[J]. *IEEE Transactions on Circuits and Systems for Video Technology*, 2003, 13(4): 300-309.
- [4] Kumar S, Azartash H, Biswas M, *et al.* Real-time affine global motion estimation using phase correlation and its application for digital image stabilization[J]. *IEEE Transactions on Image Processing*, 2011, 19(5): 3406-3418.
- [5] Bin Q, Ghazal M, and Amer A. Robust global motion estimation oriented to video object segmentation [J]. *IEEE Transactions on Image Processing*, 2008, 17(6): 958-967.
- [6] Haque M N, Biswas M, Pickering M R, *et al.* An adaptive low-complexity global motion estimation algorithm[C]. Proceeding of 28th Picture Coding Symposium, Nagoya, Japan, 2010: 598-601.
- [7] Su Y, Sun M T, and Hsu V. Global motion estimation from coarsely sampled motion vector field and the applications[J]. *IEEE Transactions on Circuits and Systems for Video Technology*, 2005, 15(2): 232-242.
- [8] Dinh T N and Lee G. Efficient motion vector outlier removal for global motion estimation[C]. Proceeding of IEEE International Conference on Multimedia and Expo, Barcelona, Spain, 2011: 1-6.
- [9] Haller M, Krutz A, and Sikora T. Evaluation of pixel-and motion vector-based global motion estimation for camera motion characterization[C]. Proceeding of 10th Workshop on Image Analysis for Multimedia Interactive Services, London, United Kingdom, 2009: 49-52.
- [10] Yin H B, Fang X Z, Yang H, *et al.* Motion vector smoothing for true motion estimation[C]. Proceeding of IEEE International Conference on Acoustics, Speech and Signal Processing, Toulouse, France, 2006: 241-244.
- [11] Dante A and Brookes M. Precise real-time outlier removal from motion vector fields for 3D reconstruction[C]. Proceeding of IEEE International Conference on Image Processing, London, United Kingdom, 2003: 393-396.
- [12] Chen Y M and Bajic I V. Motion vector outlier rejection cascade for global motion estimation[J]. *IEEE Signal Processing Letters*, 2010, 17(2): 197-200.
- [13] Chung Y C and He Z H. Reliability analysis for global motion estimation[J]. *IEEE Signal Processing Letters*, 2009, 16(11): 977-980.
- [14] Information technology-multimedia content description interface. Part 3: Visual, ISO/IEC 15938-3[S]. 2001.
- [15] Press W H, Teukolsky S A, Vetterling W T, *et al.* Numerical Recipes in C: The Art of Scientific Computing [M]. Second Edition, Cambridge, United Kingdom: Cambridge University Press, 1992: 362-383.
- [16] Hartley R and Zisserman A. Multiple View Geometry in Computer Vision[M]. Second Edition, Cambridge, United Kingdom: Cambridge University Press, 2004: 117-121.
- 赵亚湘: 女, 1966年生, 讲师, 博士生, 研究方向为数字图像及数字视频处理等。
- 樊晓平: 男, 1961年生, 教授, 博士生导师, 研究方向为智能控制、虚拟现实与网络控制等。
- 刘少强: 男, 1964年生, 副教授, 研究方向为传感器技术、测控技术等。