

# 多媒体传感器网络中基于颜色空间的图像融合方案

王玉斐<sup>1</sup>,王汝传<sup>1,2</sup>,曾 鸣<sup>1</sup>,黄海平<sup>1</sup>,孙力娟<sup>1</sup>,肖 甫<sup>1</sup>

(1.南京邮电大学计算机学院,江苏南京 210003;2.南京大学计算机软件新技术国家重点实验室,江苏南京 210093)

**摘 要:** 基于彩色图像的 YUV 颜色空间和相邻节点间的视角相关性,将同一场景的监控任务分配给 3 个相关度较大的传感器节点,每个节点仅需处理亮度分量或色度分量.使用深度信息模型,以及基于自适应四叉树分割和分块空间变换的方法,对解码后的亮度和色度分量进行融合,实现监控场景的彩色图像重构.仿真实验结果表明,该方法有效可行,在视频传感器节点存储量、传输量和场景监控质量之间能取得良好的折中.

**关键词:** 多媒体传感器网络;颜色空间;图像融合;四叉树分割

**中图分类号:** TP37 **文献标识码:** A **文章编号:** 0372-2112 (2009) 08-1659-05

## An Image Fusion Method Based on Color Space for Multimedia Sensor Networks

WANG Yu-fei<sup>1</sup>, WANG Ru-chuan<sup>1,2</sup>, ZENG Ming<sup>1</sup>, HUANG Hai-ping<sup>1</sup>, SUN Li-juan<sup>1</sup>, XIAO Fu<sup>1</sup>

(1. College of Computer, Nanjing University of Posts and Telecommunications, Nanjing, Jiangsu 210003, China;

2. State Key Laboratory for Novel Software Technology, Nanjing University, Nanjing, Jiangsu 210093, China;)

**Abstract:** The monitor task of the same scene was assigned to three highly-correlated video sensor nodes based on YUV color space and the view correlation of adjacent nodes. Each sensor node only needed to take charge of compressing and transmitting luminance or chrominance part. With the proposed depth information model, luminance and chrominance parts were fused using fusion method based on adaptive quadtree partitioning and space transform. And then the color image of the scene was reconstructed. The experimental and simulation results show that our method is effective and feasible. A well tradeoff is achieved between store, transmission cost and scene monitoring quality.

**Key words:** wireless multimedia sensor networks; color space; image fusion; quadtree partitioning

### 1 引言

无线多媒体传感器网络<sup>[1]</sup> (Wireless Multimedia Sensor Networks, WMSNs)是由无线互连的设备组成的网络,能够无所不在地从监测环境中获取多媒体信息,如视频和音频、静态图像等,同时也应能够实时存储、处理、关联和融合不同来源的多媒体数据<sup>[1]</sup>.与无线传感器网络<sup>[2]</sup>一样,节点能力受限的挑战.

单个或多个节点视频序列的冗余数据获得高质量视频信息的角度出发,研究者提出了许多图像融合算法<sup>[4~7]</sup>,但是没有考虑到节点能量和网络能力问题,不适用于无线多媒体传感器网络.文献[8]提出一种基于相关性的图像融合算法,将同一场景的监控任务分配到

相关度较大的 2 个视频传感器节点上,考虑和利用了节点的有限资源及相邻节点间的冗余信息,并利用极线约束性质进行场景图像的重构.但该方法是在系统层面上的,没有考虑到低层像素级的冗余及视频压缩技术,并且在场景重构中的假设条件限制了其在具有层次深度的场景中的应用.

本文基于 YUV/YC<sub>b</sub>C<sub>r</sub> 颜色空间,提出将监控同一物理场景的任务分配给三个相关性<sup>[9]</sup>较大的视频传感器节点上,中间视角节点负责处理场景图像的亮度分量,两侧节点负责处理场景图像的色度分量.具有更多资源和更高处理能力的上层节点分别对亮度和色度分量图像进行解码,以中间视角为基准,融合亮度和色度分量,重构出场景彩色监控图像序列.

收稿日期:2008-08-27;修回日期:2009-02-10

基金项目:国家自然科学基金(No. 60573141, No. 60773041);国家 863 高技术研究发展计划(No. 2006AA01Z201, No. 2006AA01Z219, No. 2006AA01Z439);江苏省高技术研究计划(No. BG2006001);2006 江苏省软件专项;南京市高科技项目(2007 软资 106);江苏省博士后基金(No. 0801019C)

## 2 基于 YUV/YC<sub>b</sub>C<sub>r</sub> 颜色空间的传感任务分配

基于彩色图像的 YUV/YC<sub>b</sub>C<sub>r</sub> 颜色空间,充分考虑到视频传感器节点的有限资源和人眼视觉系统的特点,将同一场景的视觉监控任务分配到相关度<sup>[9]</sup>较大的3个视频传感器节点上,每个 camera 节点只传输 YUV/YC<sub>b</sub>C<sub>r</sub> 其中一个分量,中心 camera 负责传输监控场景图像的亮度分量,称为 Y-camera;两侧 camera 负责传输场景图像的色度分量,分别称为 U/C<sub>b</sub>-camera 和 V/C<sub>r</sub>-camera.三个 camera 各自独立编码,不需要进行通信,从而降低通信开销,考虑到人眼视觉系统(HVS)对色度信号的敏感程度要弱于亮度信号,为了进一步降低传输的数据量,对色度信号采用更低的采样率.

假设监控某一场景,按 4:4:4 采样,每秒数字视频所占用的比特数为 B,采用我们提出的方法,将任务分配给三个相关度较大的监控该场景的传感器节点上,每个节点所承担的数据量减为 B/3.若对色度信号采用更低的采样率,采用 4:1:1 的采样格式,每 4 个 Y 采样点对应于 1 个 U 采样点和 1 个 V 采样点.则负责亮度分量的节点所承担的数据量仍为 B/3,而 U-camera 和 V-camera 所承担的数据量则均减为 B/12.

## 3 基于 YUV/YC<sub>b</sub>C<sub>r</sub> 颜色空间的图像融合

Sink 节点在收到同一物理场景图像的亮度和色度分量后,将不同视角的亮度和色度分量融合成一幅亮度分量视角的彩色场景图像.由于色度分量图像几乎没有灰度值以及特征轮廓不鲜明,不能采用依赖于图像灰度的准确性的搜索算法匹配方法和特征提取;因此,基于空间变换进行亮度和色度分量图像之间的配准.基于四叉树分割的自适应分块,根据场景图像的细节多少和复杂度对图像进行分块匹配.

### 3.1 场景深度信息模型

在大部分的场景中是很难获得场景详细准确的深度图像,同时,若对场景取分辨率很高的深度信息,也会增加网络的存储、计算和传输负担.我们提出一种模糊深度信息模型,将场景深度分为三个层级:远景深度  $depth\_back$ ,主目标深度  $depth\_object$  和近景深度  $depth\_forward$ .远景深度和主目标深度可以通过超声波传感器测距等方法获得,假定场景主目标近似位于背景和近景的中间位置,可以得到近景深度  $depth\_forward$ :

$$\begin{aligned} &= depth\_object - (depth\_back - depth\_object) \\ &= 2 \times depth\_object - depth\_forward \end{aligned}$$

对于同一场景,在视频传感器节点固定不变的情况下,远景深度是不变的,只需在第一次传输时,将远景深度和主目标深度数据传递到 sink 节点;当主目标

朝向视频传感器或背向视频传感器移动,  $depth\_object$  相应变化时,仅需发送当前  $depth\_object$ .

### 3.2 基于空间变换和四叉树分割的分量图像融合方法

#### 3.2.1 基于四叉树的亮度分量图像分割

基于矩形的封闭性和易表达性,采用分块的思想,首先对亮度分量图像进行基于四叉树的特征级区域分割,将其分割为大小不等的块.令 Y 表示亮度图像区域,根据区域分割的基本公式,易知最终被划分为 n 个子区域  $Y_1, Y_2, \dots, Y_n$ ,满足以下条件:(1)  $\bigcup_{i=1}^n Y_i = Y$ ; (2)  $Y_i$  是一个连通的区域,  $i = 1, 2, \dots, n$ ; (3)  $Y_i \cap Y_j = \emptyset, i \neq j$ ; (4)  $P(Y_i) = \text{TRUE}, i = 1, 2, \dots, n$ ; (5)  $P(Y_i \cap Y_j) = \text{FALSE}, i \neq j$ .

其中,  $P(Y_i)$  是定义在集合  $Y_i$  的点上的逻辑谓词,表示该集合内的所有像素点是否具有相同的亮度级别,定义为  $P(Y_i): \text{Max}(Y_i) - \text{Min}(Y_i) \leq \alpha \times 255, 0 < \alpha < 1$ ,  $\alpha$  选取的依据是图像细节信息的多少及需要的达到的分割分辨率.从整幅亮度图像 Y 开始,如果  $P(Y) = \text{FALSE}$ ,就将其分割为 4 个区域,对分割得到的区域反复执行四叉树分割算法,每一步所做的工作是:(1)对于任何区域  $Y_i$ ,如果  $P(Y_i) = \text{FALSE}$ ,就将其拆分为 4 个相连的象限区域;(2)将  $P(Y_m \cup Y_n) = \text{TRUE}$  的任意两个相邻区域  $Y_m$  和  $Y_n$  进行聚合;(3)当再无法进行聚合或拆分时操作停止.

#### 3.2.2 基于深度信息模型的分割区域空间变换

在获得 Y 分量图像的区域分割图之后,对每个矩形区域取矩形的四个顶点,从左上角以逆时针顺序依次记为  $c_1, c_2, c_3, c_4$ ,若使用  $m \times n$  矩阵表示 Y 分量图像,那么设这四个点在 Y 分量图像矩阵中的位置分别为:  $c_1: (m_1, n_1), c_2: (m_2, n_2), c_3: (m_3, n_3), c_4: (m_4, n_4)$ .设点  $c_i$  对应的世界坐标系中的坐标为  $(X_i, Y_i, Z_i)$ ,根据下面的步骤得到 Y 分量图像(正视角)与色度分量图像(偏视角)像素点之间的空间转换关系:

**Step1:** 根据将一幅视图通过透视变换将三维空间中的点投影到图像平面的投影公式,对于正视角 Y 分量图像有  $c_i = \frac{1}{s} A [R | t] C_i$ ,其中, A 是内参矩阵,不依赖于场景,可以根据视频传感器的内部参数计算,假设所有的视频传感器节点具有相同的内参矩阵,旋转平移矩阵  $[R | t]$  是外参矩阵,对不同角度监控同一场景的视频传感器节点,具有不同的外参矩阵.记  $T = \frac{1}{s} A [R | t] = (\alpha_1 \ \alpha_2 \ \alpha_3 \ \alpha_4)$ ,称为正视角 Y 分量图像

的转换矩阵,  $\alpha_i (i = 1, 2, 3, 4)$  是  $3 \times 1$  的列向量,有:

$$[m_i \ n_i \ 1]^T = (\alpha_1 \ \alpha_2 \ \alpha_3) [X_i \ Y_i \ Z_i \ 1]^T \quad (1)$$

将式(1)展开可得:

$$\begin{bmatrix} X_i & Y_i & 1 \end{bmatrix}^T = (\alpha_1 \quad \alpha_2 \quad \alpha_4)^{-1} \cdot (\begin{bmatrix} m_i & n_i & 1 \end{bmatrix}^T - Z_i \alpha_3) \quad (2)$$

**Step2:** 假设偏视角色度分量图像中与点  $(m_i, n_i)$  对应的点为  $(p_i, q_i)$ , 这两点在世界坐标系中具有相同的世界坐标  $(X_i, Y_i, Z_i)$ , 转换矩阵为  $T = (\beta_1 \quad \beta_2 \quad \beta_3 \quad \beta_4)$ ,  $\beta_i (i=1,2,3,4)$  也是  $3 \times 1$  的列向量, 同样可以得到:

$$\begin{bmatrix} X_i & Y_i & 1 \end{bmatrix}^T = (\beta_1 \quad \beta_2 \quad \beta_4)^{-1} \cdot (\begin{bmatrix} p_i & q_i & 1 \end{bmatrix}^T - Z_i \beta_3) \quad (3)$$

**Step3:** 由式(2)、(3)两可得

$$\begin{bmatrix} p_i & q_i & 1 \end{bmatrix}^T = (\beta_1, \beta_2, \beta_4) \cdot (\alpha_1, \alpha_2, \alpha_4)^{-1} \cdot (\begin{bmatrix} m_i & n_i & 1 \end{bmatrix}^T - Z_i \cdot \alpha_3) + Z_i \cdot \beta_3 \quad (4)$$

**Step4:** 根据 3.1 节提出的场景深度信息模型, 对于亮度分割图像的每个矩形区域的顶点  $(m_i, n_i)$ , 通过式(3), 我们首先令  $Z_i$  取主目标深度  $depth\_object$ , 计算得到一个初步的映射结果; 再令  $Z_i$  取远景深度  $depth\_back$  和近景深度  $depth\_forward$  得到另外两个映射结果. 对于每个  $(m_i, n_i)$ , 在偏视角色度图像中得到的三个映射点, 从而, 对正视角亮度分量图像中一个矩形分割区域的四个顶点, 根据不同的深度值, 在偏视角色度分量图像中得到三个不同的映射区域, 这限制了映射结果可能出现的范围.

### 3.2.3 $Y, C_b, C_r$ 分量图像融合

通过分割区域空间变换的映射结果, 以正视角的亮度分量图像为基准, 使用偏视角的  $C_b$  和  $C_r$  分量进行“涂色”融合得到监控场景的彩色视图. 首先给出两个定义:

**定义 1:** 对于一个由四个点组成的点集, 由水平和垂直边组成的, 包含该点集中所有点的最小矩形, 称为该点集的外闭包. 如图 7 所示, 点集  $\{P_1, P_2, P_3, P_4\}$  的外闭包为矩形  $ABCD$ .

**定义 2:** 对于一个由四个点组成的点集, 由水平和垂直边组成的, 使得该点集中的所有点均在其各边延长线所组成的十字区域之外的最大矩形, 称为该点集的内闭包. 如图 1 所示, 点集  $\{P_1, P_2, P_3, P_4\}$  的内闭包

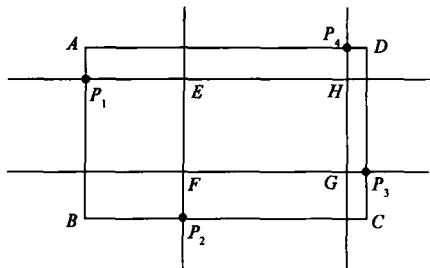


图1 点集  $\{P_1, P_2, P_3, P_4\}$  的外闭包与内闭包

为矩形  $EFGH$ .

利用三个深度信息, 对亮度图像矩形分割区域的四个顶点, 可以分别计算出对应于不同深度的, 三个由映射点组成的四点点集. 求这三个点集的外闭包, 以及这三个外闭包的交集, 最后选取交集的色度分量与对应的亮度分量叠加融合即可.

针对三个外闭包的交集为空的情况, 使用以下的策略进行“取色”: (1) 若三个外闭包交集不为空, 那么选取公共区域的色度分量; (2) 若只有其中两个外闭包有交集, 那么取该公共区域的色度分量; (3) 若存在两个互不相交的公共区域, 则选取相交闭包中心距离最近的那两个闭包的公共区域的色度分量; (4) 若三个闭包两两不相交, 则选择对应于主目标深度的红色闭包内的色度分量.

视频传感器节点之间监控角度相差不太大时, 可以忽略矩形分割映射到偏视角后的边长变化. 只计算每个分割矩形的左上角顶点的映射, 根据边长估计其他 3 个顶点的映射, 简单计算三个映射矩形的中心位置的均值, 取该均值位置的色度分量.

### 4 算法复杂度分析

首先分析基于四叉树的亮度分量图像分割的时间复杂度, 为方便讨论, 设图像的分辨率为  $M \times M (M = 2^N)$ , 对每个分块进行进一步分割的概率  $p$  是常数, 对每个分块进行操作的计算量 (如判断是否可分) 为常数  $c$ , 记对  $2^k \times 2^k$  大小的图像进行空间变换映射的时间复杂度为  $g(k)$ ,  $2^k \times 2^k$  大小的图像进行四叉树分割的平均时间复杂度为  $f(k)$ . 对每个分块进行四叉树分割的平均时间可近似为进一步分割的时间和不进一步分割时所进行的操作时间两部分, 即:  $f(k+1) = p \cdot [c + 4f(k)] + (1-p) \cdot c = 4p \cdot f(k) + c$

令  $f(0) = 1$ , 解得  $f(k) = 4^k p^k + c \sum_{i=0}^{k-1} 4^i p^i$ , 取  $k = N$ , 可得基于四叉树的亮度分量图像分割的时间复杂度为

$$O(f(N)) = \begin{cases} O(M^2), & p > \frac{1}{4} \\ O(\log(M)), & p = \frac{1}{4} \\ O(1), & p < \frac{1}{4} \end{cases}$$

同样的方法分析可得:

$$O(g(N)) = \begin{cases} O(M^2), & p > \frac{1}{4} \\ O(\log(M)), & p = \frac{1}{4} \\ O(1), & p < \frac{1}{4} \end{cases}$$

本算法在场景相对简单即  $p$  取值较小时, 有较好

的时间复杂度. 对于复杂场景, 对应于像素个数 ( $M \times M$ ), 本算法也能达到线性复杂度.

## 5 仿真实验

对原始分辨率为  $1024 \times 768$  的测试序列 *breakdancers*<sup>[10]</sup> 其进行水平方向和垂直方向上的 1:2 下采样, 转换成  $256 \times 192$  的分辨率. 考虑人眼感觉颜色的方式, 以及人眼对绿色的敏感度是对红色和蓝色敏感度的两倍, 定义加权均方误差 ( $WMSE$ ) 和加权峰值信噪比 ( $WPSNR$ ):  $WMSE = 0.25 \times MSE_R + 0.5 \times MSE_G + 0.25 \times MSE_B$ ;  $WPSNR = 10 \times \lg \frac{255 \times 255}{WMSE}$ .

选择三组 *camera* 345、246 和 147, 以摄像机 4 为中间视角摄像机, 在每个摄像机对应的视频序列中选择前 99 帧, 按照 4:2:2 转换成  $YUV$  格式. 对亮度分量图像进行四叉树分割时, 取  $\alpha = 0.05$ . 假设经过编码后的视频流在无线多媒体传感器网络中是无损传输. 多媒体传感器网络中基于  $YUV/YC_bC_r$  颜色空间的图像处理与融合方法效果如图 2 所示, 可以看出, 该方法在有效减少每个传感器节点需要处理的数据量, 降低复杂度的同时, 又能在 Sink 节点获得较为理想的融合图像, 有效实现了场景彩色图像的重建.

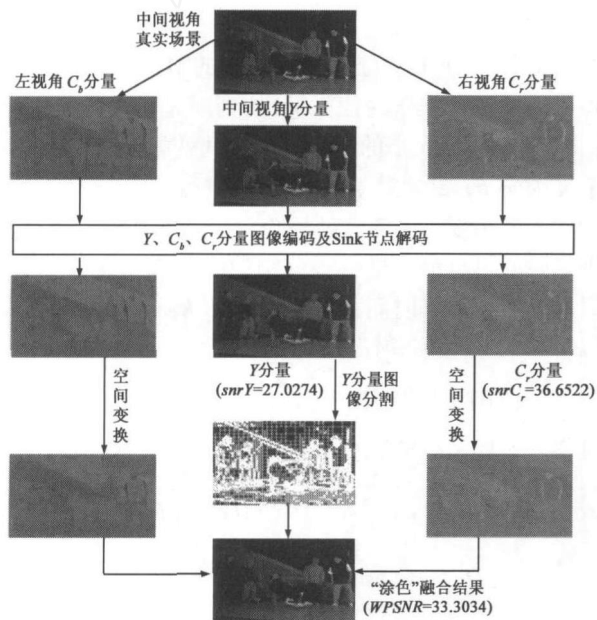


图2 基于  $YUV/YC_bC_r$  颜色空间的视频图像处理与融合仿真效果图

对转换成  $256 \times 192$  的 *camera* 3、4、5 的视频序列的亮度和色度分量, 分别使用  $4 \times 4$  的分块深度信息和深度信息模型中的深度进行融合重构. 融合重建后的中间视角彩色图像的加权均方误差如图 3 所示,  $WPSNR$  比使用  $4 \times 4$  分块深度的  $WPSNR$  低 1.7184dB, 表明在仅需传输两个深度值的情况下, 我们的深度信息模型能够达到与采用大量深度信息相近的融合质量.

测试三组不同角度的视频序列, 分别为 *camera* 345、246 和 147, 两侧 *camera* 和中心 *camera* 4 的偏离角度依次增大. 仿真结果分别如图 4 所示. 使用组合 345 的加权峰值信噪比平均比组合 246 高 0.8718dB, 比组合 147 高 1.5053dB. 这说明随着负责亮度分量和色度分量的视频传感器节点之间的偏离角度的增大, 最后融合重构的场景图像质量会逐渐降低.

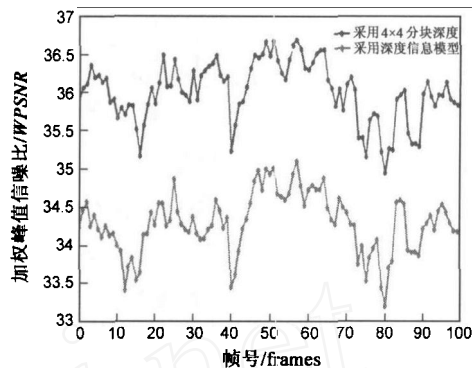


图3 采用两种深度信息的融合图像  $WPSNR$

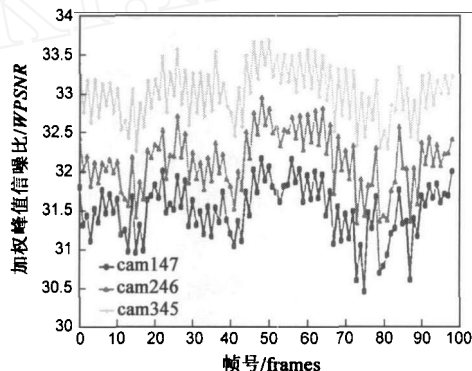


图4 基于  $YUV/YC_bC_r$  的融合图像  $WPSNR$

## 6 小结

本文基于彩色图像的  $YUV/YC_bC_r$  颜色空间, 首次提出将无线多媒体传感器网络中同一场景的监控任务分配给相关度较大的 3 个视频传感器节点上, 减少了单个节点传输和处理的数据量, 从而降低单个节点进行视频压缩的复杂度, 节约节点资源和能量. 在资源基本不受限的 Sink 节点, 提出基于自适应四叉树分割和分块空间变换的多视角亮度分量和色度分量图像融合方法, 使用深度信息模型中的两个深度信息, 融合重现场景的中间视角彩色图像. 算法复杂度分析和仿真实验结果表明, 本文所提出的方法有效可行, 在无线多媒体传感器网络的能耗及复杂度与场景监控质量之间取得了较好的折中.

## 参考文献:

- [1] Akyildiz I F, Melodia T, Chowdhury K R. A survey on wireless multimedia sensor networks [J]. *Computer Networks* (Elsevier).

- er), 2007, 51(4): 921 – 960.
- [2] Akyildiz I F, Su W, Sankarasubramaniam Y, Cayirci E. Wireless sensor networks: A survey[J]. Computer Networks(Elsevier), 2002, 38(4): 393 – 422.
- [3] 马华东, 陶丹. 多媒体传感器网络及其研究进展[J]. 软件学报, 2006, 17(9): 2013 – 2028.
- MA Hua-dong, TAO Dan. Multimedia sensor network and its research progresses[J]. Journal of Software, 2006, 17(9): 2013 – 2028. (in Chinese)
- [4] Foresti G, Snidaro L. A distributed sensor network for video surveillance of outdoor environments[A]. Proceedings of IEEE International Conference on Image Processing[C]. Rochester, New York, USA: IEEE Press, 2002. 525-528.
- [5] Marchesotti Luca, Piva Stefano, Regazzoni Carlo. An agent-based approach for tracking people in indoor complex environments [A]. Proceedings of the 12<sup>th</sup> International Conference on Image Analysis and Processing [C]. Mantova, Italy: IEEE Press, 2003. 99-102.
- [6] Ju Han, Bir Bhanu. Hierarchical multi-sensor image registration using evolutionary computation [A]. Proceedings of the 2005 Conference on Genetic and Evolutionary Computation [C]. Washington DC, USA: ACM Press, 2005. 2045 – 2052.
- [7] Dixon T D, Li J, Noyes J M etc. Scanpath analysis of fused multi-sensor images with luminance change: A pilot study [A]. The 9<sup>th</sup> International Conference on Information Fusion [C]. Florence, Italy: IEEE Press, 2006. 1 – 8.
- [8] 陶丹, 马华东. 视频传感器网络中基于相关性图像融合算法[J]. 计算机辅助设计与图形学学报, 2007, 19(5): 656 – 660.
- TAO Dan, MA Hua-dong. An image fusion algorithm based on correlation for video sensor networks[J]. Journal of Computer-aided Design & Computer Graphics, 2007, 19(5): 656 – 660.
- [9] Ma H, Liu Y. Correlation based video processing in video sensor networks [A]. Proceedings of International Conference on Wireless Networks, Communications and Mobile Computing [C]. Hawaii, USA: IEEE Press, 2005. 987 – 992.
- [10] Microsoft. Breakdancers Test Sequence[EB/OL]. <http://research.microsoft.com/vision/InteractiveVisualMediaGroup/3DVideoDownload/>. 2008

#### 作者简介:

王玉斐 男, 1982 年生于河南洛阳, 南京邮电大学计算机学院博士研究生, 主要研究方向是无线多媒体传感器网络。

王汝传 男, 1943 年生于安徽合肥, 南京邮电大学计算机学院教授、博士生导师, 主要研究方向是计算机软件、计算机通信、信息安全、无线传感器网络、Agent 技术等. E-mail: wangrc@njupt.edu.cn