

基于分层视差估计的立体图象编码

安平 张兆扬

(上海大学通信与信息工程学院电子信息系, 上海 200072)

摘要 基于立体视频数据压缩的目的,提出了一种基于分层视差估计/补偿的立体图象编码方案.该方案是采用 JPEG 标准独立编码参考图象,并利用视差估计/补偿技术编码目标图象.应用分层马尔可夫随机场(MRF)模型、率失真(RD)模型以及交叠块匹配的混合视差估计/补偿算法,可得到光滑准确的视差场.与通常的变尺寸块匹配(VSBM)相比,该算法得到的视差补偿图象的峰值信噪比(PSNR)可提高 2.5dB 左右;双向半像素精度的视差估计/补偿的性能要比单向整像素搜索提高 3dB.实验结果表明,该立体图象编码方案能有效地压缩立体图象数据,并可推广到立体序列图象的编码压缩中.

关键词 立体图象编码 视差估计 马尔可夫随机场(MRF) 率失真(RD)代价函数

中图分类号: TN911.1 **文章标识码**: A **文章编号**: 1006-8961(2002)10-1077-06

Stereo Image Coding Based on Hierarchical Disparity Estimation

AN Ping, ZHANG Zhao-yang

(School of Communication and Information Engineering, Shanghai University, Shanghai 200072)

Abstract To compress the stereo video data, we propose an algorithm for stereo image coding based on hierarchical disparity estimation/compensation. A stereo image pair consists of a reference image and a target image. The reference image is coded by method using JPEG standard, while the target image is coded using the disparity estimation/compensation technique. We apply the wavelet transformation to code the created residual image, and use DPCM (difference pulse coding modulation) to code the disparity fields. Hierarchical Markov Random Field (MRF) model, Rate-Distortion (RD) model and the overlapped block matching/compensation are applied in the hybrid disparity estimation method to obtain a smooth and consistent disparity field. In terms of peak signal to noise ratio (PSNR) of the disparity compensated image, the proposed disparity estimation algorithm achieves about 2.5 dB higher PSNR, as compared to conventional variable size block matching. The experimental results also show that the disparity estimation/compensation in half pixel accuracy and dual-direction can obtain a 3dB higher performance than that of integer pixel accuracy only in horizontal direction. In the same quality of reconstructed image, our coding system can obtain 0.387 bits reduction in terms of coding bit rate of the target image compared to VSBM based scheme, and 1.003 bits reduction compared to the one using JPEG standard to code the target image. The proposed hierarchical model for disparity estimation effectively improves the efficiency of stereo image coding.

Keywords Stereo image coding, Disparity estimation, Markov random field (MRF), Rate-distortion (RD) cost function

0 引言

随着三维(3D)技术的日趋成熟,由立体或多视点摄像系统采集的 3D 视频信号的应用正在迅速发展,预计 3D 视频信号将是未来多媒体通信的重要内容.和 2D 系统相比,3D 系统传输和存贮的至少是翻番的数据量,所以必须研究切实有效的立体图象压缩技术.

3D 视频信号的一种表征方法是采用立体图象

基金项目:国家自然科学基金资助项目(69972027)

收稿日期:2001-08-24;改回日期:2002-03-04

对,包含左右两路信号,一路为参考图象,另一路为目标图象,图象对间有很强的双目视相关性.在立体图象编码中,除了去除图象的空间冗余度外,进一步利用双目相关性,比分别独立编码两幅图象(如用JPEG方法)能获得更高的编码效率^[1].压缩立体图象的方法主要是采用视差估计/补偿技术,以消除图象对间的冗余度^[2],其中一项关键技术就是立体视差估计.所谓视差是指立体摄像图象对之间的几何差异,沿视差矢量,图象对之间具有高度的相似性,从而可用参考图象对目标图象进行视差补偿预测,目标图象只需用视差矢量(DV)及视差补偿差(DCD)表征即可.

求取视差的方法很多,如图象特征法、像素递归法、光流法及块匹配法等,但这些方法都难以获得光滑而准确的视差场^[3].固定块匹配法(FSBM)很难得到好的匹配结果,重建目标图象中块效应严重;可变尺寸块匹配(VSBM)虽然能改善匹配效果,但在块尺寸较小时,也会导致不准确的匹配,尤其是在物体边界处,误匹配严重.为了保证对应的精确性,在匹配过程中,应该利用一些适当的约束条件,如平滑约束和偏振光学几何约束^[4].文献[5]应用平滑约束,提出了一种基于马尔可夫随机场(MRF)模型的视差估计算法,可以得到较光滑而准确的视差场,然而,由于遮挡效应,常规MRF模型法在物体边界处往往不能求得准确的视差估计.

鉴于上述所存在的问题,提出一种基于分层视差估计的立体图象编码方法,其采用JPEG标准独立编码参考图象,并利用视差估计/补偿技术来编码目标图象.其中的视差估计基于分层MRF模型、交叠块匹配(OBM)和率失真(RD)模型,可以得到光滑而准确的稠密视差场,按此视差场对目标图象进行视差补偿能得到高质量的重建目标图象.

1 立体视觉模型及视差

立体摄像系统一般分为立体平行摄像系统和立体汇聚摄像系统.如果两部摄像机的光轴被设置为平行排列的,则称为立体平行摄像系统;若两个光轴相交,汇聚于感兴趣物体上,则称为立体汇聚摄像系统^[6].

图1是立体平行摄像系统的简示图,其说明了立体图象的概念以及视差的直观定义.2个摄像机的光轴相互平行,并与基线(两摄像机光心间的连线)垂直.空间对象点P在投影图象对平面中的投

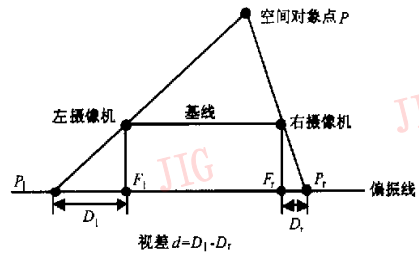


图1 立体图象及视差

影点为 P_1 和 P_2 ,称为对应点,左右投影点与两个焦点 F_1 和 F_2 之间的距离分别记为 D_1 和 D_2 .视差定义为同一空间点在两个图象平面中的投影位置差矢量,若以右图象为参考,则视差 $d = D_1 - D_2$;若以左图象为参考,则视差 $d = D_2 - D_1$.

利用视差的一些特性有助于提高视差匹配的效率和精度.视差对满足偏振几何学约束,即对应点处在同一条偏振线上,对于立体平行系统而言,图象对的扫描线与偏振线重合,则在视差估计时主要考虑水平方向的对应,从而降低了视差估计的复杂度;一般情况下,视差对应是唯一的(透明物体除外),这是唯一性约束;视差除了物体边界处,视差场应是局部光滑的,即邻域内的视差是相近的,这就是视差对应的平滑约束,MRF模型是建立视差对应平滑模型的理想工具.

2 视差估计

2.1 分层MRF模型及交叠块匹配

设 I_1 (参考图象)和 I_2 (目标图象)是立体图象对, I_2 的视差场为 D ,遮挡(在 I_2 中出现,在 I_1 中未出现的内容)场为 O .根据 Bayes 理论和 MAP(maximum a posteriori)准则,可将视差估计问题规范为^[8]

$$\begin{aligned} (\hat{D}, \hat{O}) &= \arg \max_{(D,O)} P(D,O|I_1,I_2) \\ &= \arg \min_{(D,O)} [U(I_2|I_1,D,O) + U(D|O) + U(O)] \quad (1) \end{aligned}$$

其中, $P(\cdot)$ 为后验概率函数, $U(\cdot)$ 为能量函数, $\arg \max P(\cdot)$ 表示使 P 最大的 D, O 值, $\arg \min U(\cdot)$ 表示使 U 最小的 D, O 值.

由于FSBM简单,易于实现,因此被广泛用于视差估计,其基本思想是将目标图象分割为固定尺寸的块,并找出每一个块在参考图象中的最佳匹配.然而,由于噪声及遮挡等因素,FSBM会导致不精确的视差估计,并且在重建的目标图象中有较严重的

块状干扰.若增大块尺寸,则会增加相应的估计误差,对于立体图像编码而言,这就使得视差补偿差图象的码率增大;若减小块尺寸,则表征视差场的码率太大,并且得到的视差往往并非真正的视差,难以作为分割对象的依据.改进 FSBM 的方法之一就是采用 VSBM,将具有较高估计误差的块分割为更小的块.VSBM 可以在有效的视差估计/补偿、视差场的表示以及视差补偿差之间取得折衷.可是,在子块较小时,VSBM 也会产生不一致的视差估计,导致补偿差图象中沿子块边界处有较多的高频能量.

为进一步改善视差估计的效率,将 VSBM 和 MRF 模型结合起来,建立分层的 MRF 模型,以更有效地表示视差场,减少物体边界处的块效应.此外,还把交叠块匹配(OBM)引入到视差估计中,在匹配时加权了相邻块内容,减少了因局部最小化而造成的视差误匹配;在视差补偿时,用相邻视差和当前视差对遮挡子块进行加权补偿,从而提高了补偿效率,降低了重建图象中的块效应.

应用式(1)定义的 MRF 视差能量函数的前两项,并结合 OBM,给出基于分层 MRF 和交叠块匹配(HMOM)视差估计的能量函数,式(1)中的第 3 项由于不是视差的直接函数,故可忽略.

将目标图象 I_2 等间隔分割为方块,即 $I_2 = \{b_{2,mn}, (m, n) \in \Omega\}$. 其中, $b_{2,mn}$ 表示第 (m, n) 个块, Ω 表示离散网格,即 $\Omega = \{(m, n) | 0 \leq m < N_x, 0 \leq n < N_y\}$, N_x 和 N_y 分别表示垂直和水平方向的块数.第 (m, n) 个扩展块为 $s_{2,mn}$, 块的分割标志(或遮挡标志)为 o_{mn} , 以二值表示.若第 (m, n) 个块的估计误差较大,则 $o_{mn} = 1$. 须将该块分割为小块,以提高视差估计的精度.这里采用二叉树分割结构,设 $b_{2,mn}^{(l)}$, $s_{2,mn}^{(l)}$ 为第 l 层的块及扩展块,则它们的子块分别为 $b_{2,mn}^{(l+1,k)}$, $s_{2,mn}^{(l+1,k)}$, $k = \{0, 1, 2, 3\}$. 总的能量函数如下

$$U(D, O | I_1, I_2) = \sum_{(m,n) \in \Omega} [(1-o_{mn}) \cdot u_{mn} + o_{mn} \cdot su_{mn}]$$

$$u_{mn} = (1-\alpha \|w \cdot (s_{2,mn} - s_{1,mn} - d_{mn})\|^2 + \alpha \sum_{\gamma} (1-o_{mn}^{(\gamma)}) (d_{mn} - d_{mn}^{(\gamma)})^2) \quad (2)$$

$$su_{mn} = \sum_{l+1,k} \{ (1-\alpha) \|w \cdot (s_{2,mn}^{(l+1,k)} - s_{1,mn}^{(l+1,k)} - d_{mn}^{(l+1,k)})\|^2 + \alpha \left[\sum_{\gamma^{l+1,k}} (1-o_{mn}^{(\gamma^{l+1,k})}) (d_{mn}^{(l+1,k)} - d_{mn}^{(\gamma^{l+1,k})})^2 + (d_{mn}^{(l+1,k)} - d_{mn}^{(l)})^2 \right] \}$$

其中, w 为双线性扩展窗, $d_{mn}^{(l,k)}$ 和 $o_{mn}^{(l,k)}$ 分别表示 $b_{2,mn}^{(l,k)}$ 的视差量及分割状态, γ 为预设的视差场邻域, α 是常量权,用以控制视差的光滑度;下标“ \rightarrow ”表示沿视

差的位移.

u_{mn} 和 su_{mn} 的第 1 项均表征立体图像对间的相似性约束,第 2 项为平滑约束,在视差估计时, u_{mn} 和 su_{mn} 两者选一.首先在最大块尺寸的初层,用 u_{mn} 进行基于 MRF-OBM 的视差估计,如果匹配误差大于预设的阈值,则对此块进行二叉树分割,其分割的条件是 4 个子块的编码代价小于母块的编码代价,但如果 4 个子块的编码代价大于母块,就维持原块,不进行分割,重复此分割过程,直到当前块的编码代价小于子块编码代价的总和或块尺寸足够小为止.用 su_{mn} 对子块进行视差匹配.

2.2 基于 RD 模型的四叉树(Quadtree)分割

Quadtree 表征每个初始图像块的分割状态,若该块满足分割条件,则将其分成 4 个子块,即 Quadtree 伸展出 4 棵子树,直到子块尺寸达到最小为止,此时,Quadtree 达到最高层的叶子结点.

RD 性能常作为编码优化问题的依据之一,视差估计的目的是为了在立体图像编码的码率和重建目标图像失真之间取得折中.用 RD 代价函数来确定 Quadtree 的伸展方向.块 $b_{mn}^{(l,k)}$ 及其 4 个子块的编码代价函数 $C_{mn}^{(l)}$ 和 $C_{mn}^{(l+1)}$ 定义如下

$$C_{mn}^{(l)}(\lambda) = D_{mn}^{(l)} + \lambda R_{mn}^{(l)} \quad (3)$$

$$C_{mn}^{(l+1)}(\lambda) = \sum_{k=0}^3 D_{mn}^{(l+1,k)} + \lambda R_{mn}^{(l+1,k)}$$

其中, D 和 R 分别表示块的失真和码率, λ 是 Lagrange 乘子,取大于 0 的数.当 $C_{mn}^{(l+1)} < C_{mn}^{(l)}$ 时,进行块分割,Quadtree 上由该块对应的结点伸展出 4 棵子树,否则,该结点停止伸展.

为简化算法,对 D 和 R 作如下处理: $R = R(DV) + R(DCD)$, $R(DCD) \propto D(DCD)$ (“ \propto ”表示相当于),故 $C(\lambda) = R(DV) + \lambda D(DCD)$. 其中, DV 为视差矢量,设以等长码表示; DCD 为块的视差补偿差,可用均方误差(MSE)来表示.则代价函数简化为:

$$C_{mn}^l(\lambda) = \frac{\lceil \log_2 W \rceil}{B_l \times B_l} + \frac{B_l - 1}{B_l} \sum_{i=0}^{B_l-1} \sum_{j=0}^{B_l-1} \frac{(b_{2,mn}^{(l)}(i, j) - b_{1,mn}^{(l)} - d_{mn}^{(l)}(i, j))^2}{B_l \times B_l} \quad (4)$$

其中, W 为搜索窗长度, $B_l \times B_l$ 为第 l 层块的尺寸,“ $\lceil \cdot \rceil$ ”表示向下取整.此外,Quadtree 应计入子块的码率中,以“1”表示分割结点,“0”表示叶子结点,则在 Quadtree 中需用 4bit 表示一个分割结点.

2.3 视差估计过程

根据以上的分层 MRF 模型及 RD 四叉树分割模型,以半像素精度进行视差匹配,具体过程如下:

(1) 对参考图象 I_1 作半像素精度的双线性内插^[7],内插后的参考图象记为 I_{1i} .

(2) 将目标图象 I_2 划分为初始块,在分割层次,置初值 $l=0$,遮挡标志置初值 $o_{mn}^l=0$.

(3) 用式(2)在 I_{1i} 中搜索初始块的最佳匹配,若产生的 MSE 小于阈值 T ,则该块不分割,否则到下一步作进一步分割.

(4) 在第 $l+1$ 层,用式(2)对 4 个子块进行视差估计,并计算第 l 层母块及第 $l+1$ 层 4 个子块的 RD 代价函数.若 $C_{mn}^{l+1} < C_{mn}^l$,则进行母块分割, $l=l+1$;反之则不必分割,置 $o_{mn}^l=1$.

(5) 对所有的块作分层视差估计及分割处理,直到 $C_{mn}^{l+1} > C_{mn}^l$ 或子块尺寸达到最小为止.

3 立体图象编码

立体图象编码方案如图 2 所示.

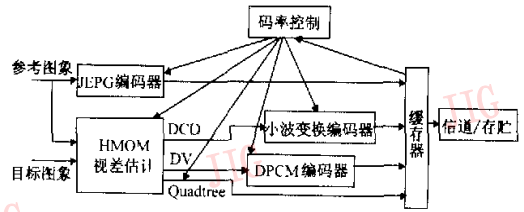


图 2 基于 HMOM 视差估计的立体图象编码框图

对目标图象,需要编码传送的数据有:DV 场, DCD 帧以及 Quadtree. 对 DV 场作 Zigzag 扫描后,进行 DPCM 编码;根据 DV 场和 Quadtree,对所有的遮挡块($o_{mn}^l=1$)作交叠块补偿处理,对非遮挡块作块补偿处理,补偿残差(DCD)帧采用小波变换编码.在解码端,根据解码的参考图象、DV 场、DCD 帧以及 Quadtree,恢复目标图象.

4 实验结果及分析

对两对立体图象对进行模拟实验(图 3,图 4),图 3 为实验所用的原始图象对,其中 Room 搜索窗



图 3 测试图象对

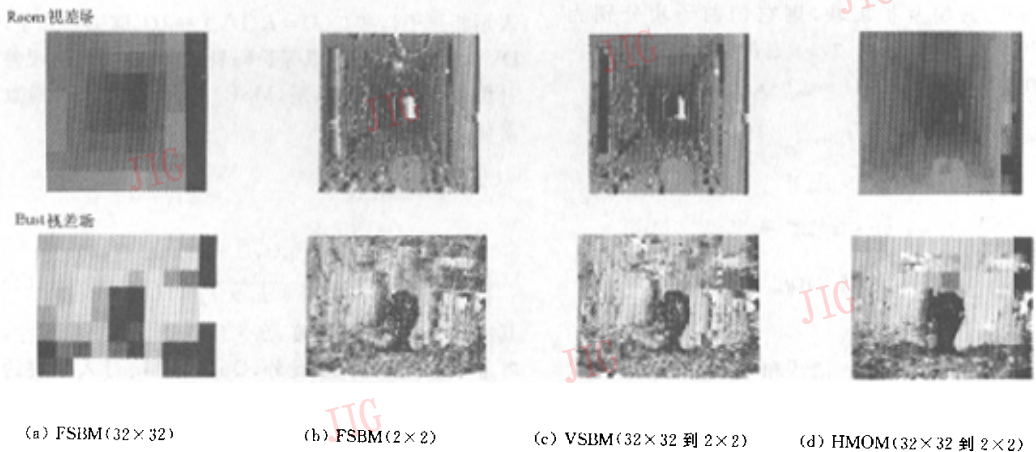


图 4 各种视差估计算法的视差场比较

取 $(0, \pm 16)$, Bust 搜索窗取 $(0, \pm 30)$. 基于 FSBM 的算法, 采用大尺寸块 (32×32) 和小尺寸块 (2×2) 所得视差场分别如图 4(a)、(b) 所示, 大尺寸块的 FSBM, 有利于避免局部最小, 作为初始视差估计是较合适的, 但其视差场缺少细节, 会导致视差补偿差的增大; 而小尺寸块的 FSBM 视差场则充满了噪声, 由于匹配仅依据亮度的相似性, 这些视差往往没有反映真正的对象深度信息, 误匹配严重. 正如前面所分析的, VSBM 得到的视差场性能优于 FSBM, 视差场受噪声干扰的程度大为减轻, 但当子块变小时, 由于噪声、遮挡等干扰, 视差场也会出现不一致现象, 如图 4(c) 所示. 图 4(d) 是基于 HMOM 算法的视差场, 该视差场的光滑性和精确性得到很大提高. 光滑而精确的视差场对图象编码是非常有利的, 从而提高了整个立体图象编码的效率. 图 5 比较了 VSBM 和 HMOM 算法的视差估计/补偿率的失真性能, 在相同比特率下, HMOM 得到的补偿目标图象的信噪比可提高 2.5dB 左右.

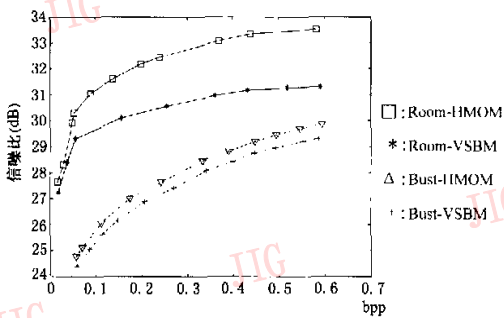


图 5 VSBM 与 HMOM 视差补偿率失真比较

从理论上说, 立体平行摄像系统只产生水平方向的视差, 但由于仪器精度问题, 实际在垂直方向也存在微小的视差, 故对 Room 和 Bust 分别选择 $[\pm 1, \pm 16]$ 和 $[\pm 1, \pm 30]$ 搜索窗, 可显著提高视差

匹配的精确度, 减少 DCD 帧误差. 此外, 图象对应点不一定正好处在整像素位置上, 在半像素精度上进行视差估计能进一步提高视差估计和补偿的效率. 采用 HMOM 法, 双向半像素精度视差估计/补偿的编码效率与单向整像素点编码效率的比较见图 6, 双向半像素精度的视差估计/补偿性能比单向整像素搜索提高 3dB 以上.

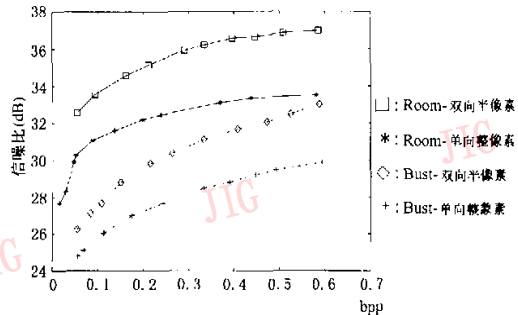


图 6 双向半像素与单向整像素视差估计补偿率失真性能比较
(Room 搜索窗: 双向 $[\pm 1, \pm 16]$, 单向 $[\pm 0, \pm 16]$;
Bust 搜索窗: 双向 $[\pm 1, \pm 30]$, 单向 $[\pm 0, \pm 30]$)

与采用 JPEG 标准压缩目标图象相比, 基于视差补偿法可显著提高编码效率. 本系统在信源端采用基于 HMOM 视差估计/补偿, 进一步提高了编码效率. 表 1 比较了 Room 及 Bust 在解码图象的峰值信噪比 (PSNR) 分别为 38.66dB 和 35.42dB 时几种方法的码率分布. 其中, 基于视差补偿法分别列出了 VSBM 及 HMOM 的数据. HMOM 法与 JPEG 法相比, Room 和 Bust 的码率分别降低 1.003bit 和 0.798bit; 与 VSBM 法相比, 则码率分别降低 0.387bit 及 0.310bit.

立体图象编码方案能有效地压缩立体图象数据. 此方法可进一步推广到立体序列图象的编码压缩中.

表 1 用视差补偿法及 JPEG 法编码目标图象的码率比较

单位: bpp

图象	基于视差补偿的系统编码方案								JPEG 编码方案 比特率	解码图象 PSNR (dB)
	基于 HMOM 法				基于 VSBM 法					
	总比特率	DV	Quadtree	DCD	总比特率	DV	Quadtree	DCD		
Room	0.157	0.043	0.004	0.110	0.544	0.328	0.108	0.108	1.160	38.66
Bust	0.492	0.229	0.033	0.230	0.802	0.457	0.094	0.251	1.290	35.42

参考文献

- Sethuraman S *et al.* Segmentation based coding of stereoscopic image sequences[A]. In: Proceedings of SPIE on Digital Video Compression: Algorithm and Technologies [C], San Jose CA USA:1996,2668:420~429.
- Labonte F *et al.* Spatiotemporal spectral coding of stereo image sequences[J]. IEEE Trans. on Circuits and Systems for Video Technology, 1999,9(10):144~155.
- Andre' Redert, Emile Hendriks, Jan Biemond. Correspondence estimation in image pairs[J]. IEEE Signal Processing Magazine, 1999,16(3):47~54.
- Ebroul Izquierdo M. Disparity/Segmentation Analysis: Matching with an Adaptive Window and Depth-Driven Segmentation[J]. IEEE Trans. on CSVT, 1999,9(4):589~607.
- Woontack Woo, Antonio Ortega. Stereo image compression with disparity compensation using the MRF model [A]. In: Proceedings of SPIE on Visual Communications and Image Processing[C], Orlando FL USA:1996,2727:28~41.
- 侯春萍. 平面图象立体化技术的研究[D]. 天津:天津大学, 1998:27~34.
- Kenneth R. Castleman. Digital image processing [M]. New Jersey USA, Prentice Hall, 1996:117~119.



安平 1968年生,上海大学通信与信息工程学院在读博士生,研究方向为视频信号处理和三维图象编码及其应用。



张兆扬 1938年生,教授,博士生导师,上海大学通信与信息学科带头人,研究方向为数字视频处理、多媒体通信和计算机视觉。

新惠普又一成果新战略

—— 惠普掘金 EDA、MDA

9月4日,惠普举行了“EDA、MDA 引领惠普工作站”的新战略发布会,会上,惠普详细分析了EDA(电子设计自动化)、MDA(机械工程自动化)在全球乃至亚太地区良好的发展状况,同时指出,由于当今众多工作站产品的速度和容量有限,使这两大领域在最短时间内很难完成高质量的产品设计,而惠普推出的完整工作站系列以一体化解决方案的形式,使得电子、制造业的工程师可以更快、更准确地运行计算密集型应用,从而提高产品质量,增强其特性,降低制造成本。据Gartner数据显示,惠普全球工作站产品市场的占有率已经达到了27%,稳居市场头名,领先第二位市场占有率6个百分点。惠普工作站产品新战略的发布,无疑将加剧工作站产品市场的竞争。

在新战略发布会上,惠普信息产品集团中国区工作站产品部总经理陈允贤先生介绍,合并使新惠普拥有更完整的工作站产品系列,将在全球工作站产品市场占有绝对领先的地位。整合后的惠普工作站的产品线分为三个系列,分别是zx、x与b/c/j。其中zx系列是基于Itanium2架构的,可运行Windows NT, Linux, 和HP Unix三种操作系统,x系列是基于IA-32架构的个人工作站,基于PA-RISC的产品系列又分为j、c、b高、中、低三条产品线。惠普作为综合的工作站产品提供商向用户提供了三种不同的架构,从低端的桌面设计到高端科学计算领域,使用户不再局限于一种选择,有了更多考虑。

据惠普亚太区个人系统集团工作站事业部EDA市场行销总监陈威铭先生介绍,工作站应用于EDA领域已成为增长最快的应用之一,2002年亚洲市场将成为其中增长最快的区域,占全球市场份额的15.0%,预计到2003年中国将成为全球最大的EDA市场。面对着日益增长的市场,陈威铭先生信心百倍地表示:“惠普工作站已经凭借其品质在业界取得了良好的口碑,整合使新惠普获得了双方的优势,为用户提供了更大存储容量、更快的处理速度,并且允许用户选择不同的产品架构与操作系统,完善的技术支持与售后服务也将成为惠普扩大EDA市场份额的重要砝码。”

首先,惠普工作站提供了快速处理能力,性能较高;其次,惠普工作站产品提供了大容量内存,以备运行之需,拥有大容量内存的好处就是可以大大缩减启动复杂程序的时间,使整机运行速度上升一个档次;再次,惠普工作站以网络为中心的计算概念,将更便于管理,并且节约成本。另外,最值得一提的是,惠普工作站具有跨平台,多兼容性的特性,它提供了多种操作系统,包括Unix、Windows NT、Linux,其相互兼容性使得惠普工作站更具可用性,给用户更多选择。