

基于频率敏感三维自组织映射的视差估计算法

任云 程福林 黎洪松

(桂林电子科技大学信息与通信学院 桂林 541004)

摘要 提出基于频率敏感三维自组织映射的立体视频视差估计算法,视差预测采用基于亮度分类频率敏感三维自组织映射(Classified Frequency Sensitive Three-Dimensional Self-Organizing Map, CFS-3DSOM)的视差模式识别(Disparity pattern recognition, DPR)算法。其基本思想是对视差图像的低亮度区域和高亮度区域分别进行预测,在训练模式库时引入频率敏感方法。实验表明,与传统基于块的视差估计算法相比,CFS-3DSOM-DPR算法的视差预测图像的平均峰值信噪比提高了0.78~1.78dB,时间减少了70%。

关键词 视差估计,自组织映射,视差模式识别,立体视频编码

中图分类号 TN919.8 **文献标识码** A

Disparity Estimation Algorithm Based on Frequency Sensitive Three-dimensional Self-organizing Map

REN Yun CHENG Fu-lin LI Hong-song

(School of Information and Communication Engineering, Guilin University of Electronic Technology, Guilin 541004, China)

Abstract A disparity estimation algorithm was presented based on frequency sensitive three-dimensional self-organizing map in this paper. It uses disparity pattern recognition (DPR) algorithm based on classified three-dimensional self-organizing map (CFS-3DSOM). The main idea is predicting disparity separately for low light area and high light area of disparity images, and adding frequency sensitive method when training pattern library. Experiment results show that the average peak signal noise ratio (PSNR) of disparity prediction image of CFS-3DSOM-DPR algorithm increases 1.78dB compared to disparity estimation algorithm based on block.

Keywords Disparity estimation, Self-organizing map, Disparity pattern recognition, Stereo video coding

1 引言

传统视差估计算法可分为基于区域的视差估计和基于特征的视差估计。基于区域的视差估计的基本思想是把图像中的某一点或某一块作为基本匹配单元,在另一幅图像中搜索对应的点或块,从而得到各个点的视差,这类算法可以得到稠密的视差估计。根据匹配策略的不同,基于区域的视差估计可以分为局部法和全局法。代表性的局部法是基于块^[1-2]的方法,其实现复杂度低,有广泛的应用。典型的全局算法有动态规划法、图割法^[3]和置信度传播法^[4]等,这类方法能得到比局部法更好的结果,但是其计算复杂度高,不利于硬件实现。基于特征的视差估计^[5-7]是利用图像的自身特征来实现匹配的,此类算法特征的提取比较复杂,且只能得到稀疏的视差估计。

根据人脑神经元的组织原理, Kohonen 提出了自组织映射(Self-Organizing Map, SOM)算法^[8],这是一种具有自组织特性的人工神经网络算法。该算法模拟了人脑神经系统对某一图形或某一频率特定兴奋的特征,在模式识别和数据挖掘等领域得到了深入的研究和广泛的应用^[9-12],是一种高效的数据聚类方法。SOM算法的核心是寻找最佳分类,即通过对大量的样本序列进行不断学习和训练,得到最优匹配的模式库。

为了克服传统视差估计算法的不足,改善视差补偿预测的性能,本文提出了基于频率敏感三维自组织映射的视差估计算法,其视差预测编码方案采用基于亮度分类频率敏感三维自组织映射(CFS-3DSOM)的图像模式识别作为视差预测器。该算法的基本思想是对视差图像的低亮度区域和高亮度区域分别训练模式库,即低亮度模式库和高亮度模式库;编码时,根据图像块的亮度采用不同的模式库进行预测。同时模式库训练引入频率敏感方法以减少模式库中无效模式的数量。实验表明,相比于传统基于块的视差估计算法,CFS-3DSOM-DPR算法的视差预测图像的平均峰值信噪比提高了0.78~1.78dB。

2 CFS-3DSOM 算法

2.1 频率敏感三维自组织映射算法

传统 SOM 算法中,虽然模式库中每个模式都有可能被调整,但每个模式成为响应模式的概率是不一样的,模式响应的概率越大,该模式及其邻域内的模式调整的机会也就越大;反之,调整的机会就越小。最优模式库要求模式库中的每个模式都能得到充分利用,这需要每个模式都能适应信源的输入模式矢量。

针对传统 SOM 算法存在的不足,本文采用了频率敏感

本文受国家自然科学基金资助项目(61261035),桂林电子科技大学研究生教育创新项目(GDYCSZ201451)资助。

任云(1992—),女,硕士生,主要研究方向为视频通信、神经网络算法, E-mail: 2431006613@qq.com;程福林(1990—),男,硕士生,主要研究方向为视频通信、神经网络算法;黎洪松(1963—),男,博士,教授,主要研究方向为视频通信、智能信息检测、处理和控制在。

三维自组织映射 (Frequency Sensitive Three-Dimensional Self-Organizing Map, FS-3DSOM) 算法, 可以在模式矢量响应次数增大时, 降低其再次响应的概率, 从而达到人为的“公平”。

基于 FS-3DSOM 算法的模式库训练步骤如下:

(1) 设定 SOM 网络的大小为 (N, M) , 其中 N, M 分别为模式库的大小和模式库中模式矢量的大小。

(2) 将视差图像分割为 $M=8 \times 8$ 的图像块, 每个图像块为一个训练矢量, 共得到 L 个训练矢量 $\{X(t), t=0, 1, \dots, L-1\}$ 。设置初始化模式库为 $\{W_j(0), j=0, 1, \dots, N-1\}$, 并将模式矢量排列成三维立体结构。

(3) 设置初始化邻域为 $N_j(0), j=0, 1, \dots, N-1$, 设置响应计数器 $c_j=1$, 其中 $j=0, 1, \dots, N-1$ 。

(4) 输入训练矢量 $\{X(t), t=0, 1, \dots, L-1\}$ 。

(5) 将失真测度设定为:

$$\tilde{d}_j(t) = f(c_j) d_j(t)$$

其中, $f(c_j)$ 为频率敏感函数, 通常采用式(1)的指数函数:

$$f(c_j) = c_j^s \quad (1)$$

它随着频率敏感参数 c_j 的增大而单调递增, s 为频率敏感指数。均方误差 $d_j(t)$ 为:

$$d_j(t) = \|X(t) - W_j(t)\|^2 \quad (2)$$

分别计算出该训练矢量与模式库中各模式矢量的失真 $\tilde{d}_j(t)$, 并选择获胜模式矢量 j^* 为具有最小失真的模式矢量。

(6) 按式(2)调整获胜模式矢量 j^* 及其三维邻域 $N_{j^*}(t)$ 范围内的模式矢量。

$$W_j(t+1) = \begin{cases} W_j(t) + \alpha(t)[X(t) - W_j(t)], & j \in j^*, N_{j^*}(t) \\ W_j(t), & \text{其他} \end{cases}$$

其中, $N_{j^*}(t)$ 为邻域函数, 通常选用单调递减函数:

$$N_{j^*}(t) = A_0 + A_1 e^{-t/T_1}$$

其中, A_0, A_1 分别为获胜模式矢量 j^* 的最小邻域和最大邻域, T_1 为邻域衰减常数。在训练的初始阶段, 邻域半径较大, 而随着训练矢量的增多, 网络逐渐趋于稳定, 只需对获胜节点进行较细微的权值调整, 因而邻域半径不断缩小。 $\alpha(c_j)$ 为学

习速度函数, 它反映了模式矢量调整的幅度大小, 一般选用单调递减函数:

$$\alpha(c_j) = A_2 e^{-c_j/T_2}$$

其中, A_2 为训练开始时的最大学习速度, T_2 为学习衰减常数。

(7) 设置 $c_j^* = c_j^* + 1$, 返回步骤(4), 直到训练完所有的训练矢量。

2.2 分类模式库

在模式库训练过程中, 差别大的训练矢量之间存在着相互的负面影响, 这是无效模式产生的一个重要原因。高亮度区域与低亮度区域的差别较大, 训练过程中容易相互影响。针对该问题, 在训练模式库时, 首先利用均值将训练矢量集分为高亮度区域和低亮度区域, 分别训练模式库, 以保证 SOM 算法在两个区域上的聚类性能。模式库的分类步骤如下:

步骤 1 将视差图分割成 8×8 的子块, 得到总的训练矢量集 $\{X(t), t=0, 1, \dots, L-1\}$ 。

步骤 2 计算训练矢量 $X(t)$ 的均值 $\bar{X}(t)$, 根据式(3):

$$\begin{cases} X(t) \in X_1, & \text{if } \bar{X}(t) < T_h \\ X(t) \in X_2, & \text{if } \bar{X}(t) > T_h \end{cases} \quad (3)$$

进行分类。其中 T_h 为高亮度区域和低亮度区域均值判定限值, X_1 和 X_2 分别表示高亮度区域和低亮度区域的训练矢量集。

步骤 3 对 X_1 和 X_2 两个训练矢量集分别使用 FS-3DSOM 算法训练得到高亮度模式库和低亮度模式库。

3 基于 CFS-3DSOM 的立体视频视差编码方案

基于 CFS-3DSOM 的立体视频视差估计可分为视差模式库训练和视差模式识别两个阶段。在视差模式库训练阶段, 首先将输入的训练视差序列进行分块处理, 然后采用 SOM 算法对分块后的图像序列进行学习训练, 同时调整相关参数, 以获得最优的视差模式库。在视差模式识别阶段, 先将测试视差序列进行分块处理, 然后在视差模式库中寻找与输入的子图像块最匹配的视差模式。

图 1 给出了 CFS-3DSOM-DPR 视差编码的框图。

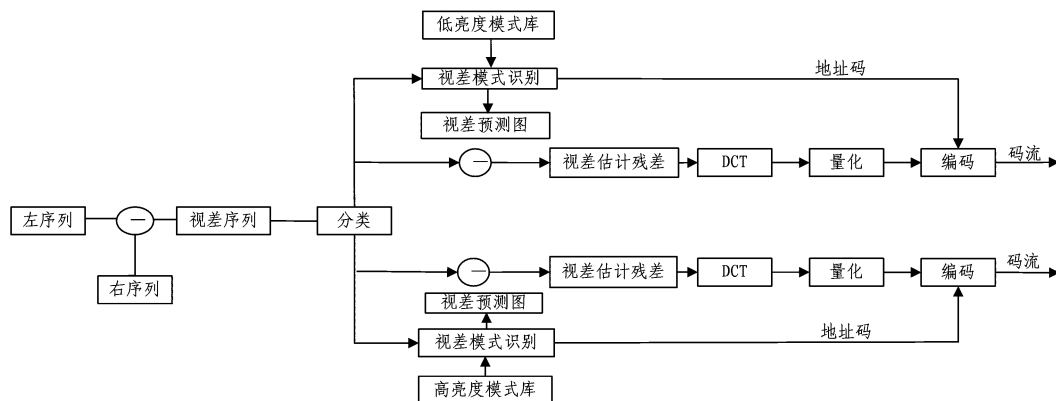


图 1 CFS-3DSOM-DPR 视差编码方案

CFS-3DSOM-DPR 视差编码步骤如下。

步骤 1 对立体视频视差序列的每帧进行分块(一般取 8×8 或 16×16), 根据亮度均值分为高亮度块和低亮度块。

步骤 2 利用 CFS-3DSOM 算法对分类后的图像块分别进行训练, 以获得最优的高亮度模式库和低亮度模式库。

步骤 3 对视差序列的每帧进行分块, 根据均值对每个图像块进行分类, 选择相应的模式库, 对每个子块图像(输入

模式矢量)进行视差模式识别。计算图像子块与视差模式库中各模式矢量之间的均方误差, 均方误差最小的模式即为匹配的视差模式, 并将匹配视差模式作为该图像子块的预测模式矢量。

步骤 4 将输入模式矢量与预测模式矢量相减, 得到预测误差矢量, 即视差估计残差。

步骤 5 对视差估计残差进行 DCT 变换、量化和编码。

步骤 6 转到步骤 3,直至对所有的视差块完成预测编码。

4 实验结果

实验采用标准立体视频测试序列 Exit, Vassar 的第 1,2 视点,图像分辨率为 480×640 ,左右视点各 8 帧。重建图像品质的客观评价采用峰值信噪比 (Peak Signal to Noise Ratio, PSNR) 来衡量, $PSNR = 10 \lg \frac{255^2}{E_{MSE}}$ dB, 其中 E_{MSE} 为原始图像与重建图像之间的均方误差。在 3DSOM-DPR 视差估计算法中,图像压缩比为 $C_R = \frac{M \times B_O}{B_C}$, 其中 M 为模式矢量的维数, B_O 为原始图像的每像素比特数, B_C 为模式矢量地址比特数。例如,模式矢量大小为 $M=8 \times 8$,模式库大小设为 2048,压缩率为 46.5。在基于块的视差估计算法中,块的大小为 $M=8 \times 8$,横向搜索范围为 $[-31, 32]$,纵向搜索范围为 $[-15, 16]$,压缩率为 46.5。

图 2 和图 3 分别给出了测试序列为 Exit 和 Vassar 时, CFS-3DSOM-DPR 算法与基于块的方法的视差预测图像的 PSNR 分布情况。表 1 列出了 CFS-3DSOM-DPR 算法与基于块的方法的视差预测图像的平均 PSNR 对比,表 2 列出了 CFS-3DSOM-DPR 算法与基于块的方法的视差预测图像的运行时间对比。

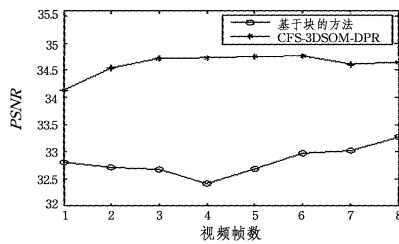


图 2 Exit 序列视差预测图像的 PSNR 分布情况

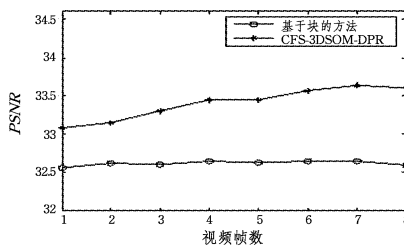


图 3 Vassar 序列视差预测图像的 PSNR 分布情况

表 1 CFS-3DSOM-DPR 算法与基于块的方法的比较结果 (平均峰值信噪比/dB)

测试序列	基于块的方法	CFS-3DSOM-DPR 算法
Exit	32.8244	34.6122
Vassar	32.6179	33.4052

表 2 CFS-3DSOM-DPR 算法与基于块的方法的比较结果 (运行时间/s)

测试序列	基于块的方法	CFS-3DSOM-DPR 算法
Exit	343.8609	85.0447
Vassar	347.7581	84.3563

由图 2 和图 3 可以看出,采用 CFS-3DSOM-DPR 算法的重建效果要明显优于基于块的方法,但如图 2 和图 3 所示,随着视频帧数的增加,PSNR 有下降的趋势,这是由所用的高亮度模式库的有限性决定的,帧数增加导致多样性增加,所

能得到的匹配块误差也会慢慢变大。但在帧数较少或者模式库足够大的情况下,CFS-3DSOM-DPR 算法具有明显的优势。由表 1 可以看出,与基于块的方法相比,由 CFS-3DSOM-DPR 算法得到的视差预测图像的平均峰值信噪提高了 0.78~1.78dB,且时间减少了约 75%。

结束语 本文提出了基于频率敏感三维自组织映射的视差估计算法,视差编码采用基于分类频率敏感三维自组织映射的视差模式识别(CFS-3DSOM-DPR)算法,以取代传统的基于块的视差估计算法。实验表明,与基于块的方法相比,该算法视差预测图像的平均峰值信噪提高了 0.78~1.78dB。下一步工作是将该算法应用于立体视频编码中,并结合具体应用做更深入的研究。

参考文献

- [1] HAUBLEIN K, REICHENBACH M, FEY D. Fast and generic hardware architecture for stereo block matching applications on embedded systems[C]// 2014 International Conference on Re-ConFigurable Computing and FPGAs (ReConFig). IEEE, 2014: 1-6.
- [2] SAHLBACH H, ERNST R, WONNEBERGER S, et al. Exploration of FPGA-based dense block matching for motion estimation and stereo vision on a single chip[C]// IEEE Intelligent Vehicles Symposium. IEEE, 2013: 823-828.
- [3] WANG Y C, TUNG C P, CHUNG P C. Efficient disparity estimation using hierarchical bilateral disparity structure based graph cut algorithm with a foreground boundary refinement mechanism [J]. IEEE Transactions on Circuits & Systems for Video Technology, 2013, 23(5): 784-801.
- [4] 赵戈, 蔺兰, 唐延东, 等. 一种基于曲率与置信度传播的视差估计方法[J]. 模式识别与人工智能, 2013, 26(12): 1154-1160.
- [5] PODDAR S, SAHU H, BANGALE M R, et al. Feature based dense disparity estimation[C]// 2015 International Conference on Industrial Instrumentation and Control (ICIC). IEEE, 2015: 950-955.
- [6] KAMENCAY P, BREZNAN M, JELSOVKA D, et al. Sparse disparity map computation from stereo-view images using segment based algorithm[C]// 2012 22nd International Conference Radioelektronika (RADIOELEKTRONIKA). IEEE, 2012: 1-4.
- [7] RAMAN S, KANOJIA G, et al. FacialStereo: Facial depth estimation from a stereo pair [C]// International Conference on Computer Vision Theory and Applications. 2014: 686-691.
- [8] KOHONEN T. Essentials of the self-organizing map[J]. Neural Networks, 2013, 37: 52-65.
- [9] LI N, CHENG X, ZHANG S, et al. Realistic human action recognition by Fast HOG3D and self-organization feature map[J]. Machine Vision and Applications, 2014, 25(7): 1793-1812.
- [10] KAMAL S, MUJEEB A, SUPRIYA M H. Novel class detection of underwater targets using self-organizing neural networks[C]// Underwater Technology (UT), 2015. IEEE, 2015: 1-5.
- [11] PARK S, RYU U S, CHOI Y, et al. A framework for baseline load estimation in demand response: Data mining approach[C]// 2014 IEEE International Conference on Smart Grid Communications (SmartGridComm). IEEE, 2014: 638-643.
- [12] MCLOUGHLIN F, DUFFY A, CONLON M. A clustering approach to domestic electricity load profile characterisation using smart metering data[J]. Applied Energy, 2015, 141: 190-199.