

应用于多视点视频编码的基于深度特征的 图像区域分割与区域视差估计

杨海涛¹ 常义林¹ 霍俊彦¹ 熊联欢² 林四新²

(¹ 西安电子科技大学 ISN 国家重点实验室, 陕西 西安 710071; ² 华为技术有限公司, 广东 深圳 518129)

摘要 基于多视点视图深度特征, 提出一种通过简单块匹配运算划分多视点视图区域并估计区域视差的算法。首先基于深度对象的概念确定图像中具有不同深度的区域数量以及这些区域对应的区域视差, 再根据误差最小化准则初步确定每个图像块所属区域。当区域中图像块数量小于某个阈值时, 采用区域合并算法将该区域中的每个图像块合并到与它的视差最为接近的其它图像区域, 通过迭代形成最终的有效图像区域划分。实验表明, 该算法能够以图像块为基本单元有效地划分各深度层区域, 并准确估计对应的区域视差。

关键词 图像信号处理; 区域分割; 区域视差估计; 深度对象; 多视点视频; 图像二值化

中图分类号 TN911.73 文献标识码 A

Depth Characteristic-Based Image Region Partition and Regional Disparity Estimation for Multi-View Video Coding

Yang Haitao¹ Chang Yilin¹ Huo Junyan¹ Xiong Lianhuan² Lin Sixin²

(¹ State Key Laboratory of Integrated Services Networks, Xidian University, Xian, Shaanxi 710071, China)
(² Huawei Technologies Co., Ltd., Shenzhen, Guangdong 518129, China)

Abstract Based on depth characteristic of multi-view images, an image region partition and regional disparity estimation algorithm via block matching is proposed. Firstly, based on the concept of the depth object, the number of regions is determined and corresponding regional disparities are estimated. Then, the initial image region partition is obtained by minimizing the block matching error. When the number of the blocks in a region is less than a threshold, the region is eliminated, and all the blocks in this region are reassigned to other regions according to the region merging algorithm. The region merging process is iteratively executed to obtain valid region partition results. Experimental results show that the proposed algorithm can effectively partition image regions with different depths and accurately estimate corresponding regional disparities.

Key words image signal processing; region partition; regional disparity estimation, depth object, multi-view video; image binarization

1 引 言

多视点视频是由处于空间不同位置的相机阵列拍摄同一场景时得到的一组视频序列信号, 是一种新型的具有立体感和交互操作功能的视频, 可应用于任意视点电视 (Free-viewpoint television), 三维

立体电视 (3DTV), 沉浸式电视会议 (Immersive teleconference) 以及视频监控等多种正在兴起的媒体业务^[1]。

对象深度, 即拍摄场景中真实对象与摄像机间距离, 是多视点视频的重要技术参量。当已知两个

收稿日期: 2007-08-10; 收到修改稿日期: 2007-12-13

基金项目: 国家自然科学基金(60772134)、高等学校学科创新引智计划(B08038)和西电-华为多媒体通信联合实验室合作专项基金资助课题。

作者简介: 杨海涛(1983—), 男, 博士研究生, 主要从事图像处理和多视点视频编码等方面的研究。

E-mail: htyang@mail.xidian.edu.cn

导师简介: 常义林(1944—), 男, 博士生导师, 主要从事多媒体通信和网络管理等方面的研究。

E-mail: ylchang@xidian.edu.cn

摄像机的空间几何参量、成像参量和其中一个视点摄像机所成图像时,可通过对象深度值合成另一个视点视图中该对象所成的图像。根据投影几何理论,由于深度不同,摄像机所成图像中各个对象的视差也不相同,且两者存在确定的对应关系。多视点视频编码中的视差估计与单视点视频编码中的位移估值同等重要,通过它可以有效地去除视点间的冗余度,提高多视点视频编码效率。因此有效划分不同深度层区域并准确估计不同深度层区域的视差成为多视点视频编码中的关键技术。

视频区域分割是指将视频序列中图像按一定的准则分割成区域,目的是分离出视频对象,可用于基于对象的视频编码、模式识别与智能检索等多个领域。传统的单视点视频区域分割方法主要包括基于灰度、颜色、边缘以及运动等信息的分割^[2~6]。在多视点视频中,能够获得视差或深度信息,由于深度信息接近语义对象,可以利用它找到图像中不同的对象区域,从而更容易实现区域分割,因此出现了多种基于深度特征的图像区域分割算法。文献[7]提出联合颜色与深度信息使用改进的递归最短生成树(M-RSST)算法分割图像区域。文献[8]基于深度信息选取初始轮廓,再通过最小化轮廓能量精确定位区域边界。文献[9]在视差粗匹配的基础上,分离出前景与背景,再对前景物体进行精细匹配。文献[10]在得到所有图像块视差估计结果后,应用高斯混合模型(Gaussian mixed model)分类器得到区域分割结果。这些区域分割算法的一个共同特点是需要对每帧图像估计得到基于像素或图像块的视差场,再进行不同深度层区域的分割。

如果能够避免计算和分割图像视差场,直接提取图像中各对象的深度特征进而找到其对应的区域实现区域分割,可以大大降低区域分割算法的计算复杂度。基于此思想,本文提出一种新颖的基于深度特征的多视点视频区域分割与区域视差估计算法。算法以图像块为基本单元,可用于多视点图像、视频压缩与传输,多视点视频对象提取,以及各种分级视差估计。

2 算法原理

为了清楚地描述所提出的区域分割与区域视差估计算法,引入深度对象与区域视差的概念。将多视点视频图像中具有近似相同视差的区域称为深度对象,而能够代表深度对象内所有像素深度特征的视差就称为区域视差。一个深度对象并非仅对应于一个

客观世界中的真实对象,它可能对应场景中深度特性较为一致的几个对象。另外,深度对象与运动对象的概念也不相同,对应于同一个深度对象的几个真实场景中的对象可以具有各自不同的运动参量。

基于深度对象与区域视差的概念,图1给出深度对象分割与区域视差估计算法原理。算法首先使用原图像低通滤波结果作为阈值对输入的不同视点图像进行二值化处理,对得到的二值化图像对计算视点偏移-匹配误差函数(VO-MF),并根据该函数中所有局部极小值点获得一组初始区域视差估计值。接着,将原图像划分为 $M \times M$ 大小的图像块,对每一个图像块遍历每个初始区域视差估计值,计算块匹配误差,选择具有最小匹配误差的区域视差作为该图像块的视差,同时标记该图像块所属区域,具有相同视差的图像块集合构成了一个初始深度对象。通过区域合并将包含较少图像块的深度对象合并到其它深度对象中。最后,为了提高深度对象及对应区域视差表示的有效性,对分割结果中值滤波得到平滑的深度对象。

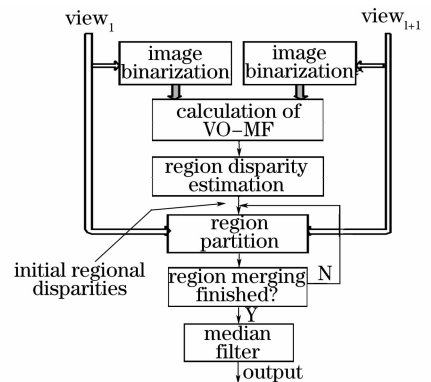


图1 深度对象分割与区域视差估计算法原理图

Fig. 1 Schematic diagram of the proposed depth object segmentation and regional disparity estimation algorithm

3 深度对象分割与区域视差估计

以多视点视频序列集 Ballroom^[13]中同一时刻两个不同视点中图像 $I_r(x, y)$ 与 $I_c(x, y)$ 作为参考图像与目标图像,讨论提出的深度对象分割与区域视差估计算法。

3.1 图像二值化

为了降低区域视差估计的运算复杂度和提高区域视差估计的精度,需要对多视点视图做二值化处理。首先对原图像 $I(x, y)$ 低通滤波得到模糊化的图像 $\tilde{I}(x, y)$:

$$\tilde{I}(x, y) = F[I(x, y)], \quad (1)$$

其中 F 采用文献[11]中的低通滤波器：

$$F = \frac{1}{4} \begin{bmatrix} 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \end{bmatrix}, \quad (2)$$

再使用这个模糊化的图像作为阈值对原图像做二值

化处理,得到二值化图像 $I^B(x,y)$ ：

$$I^B(x,y) = \begin{cases} 1, & \text{if } [I(x,y) \geq \bar{I}(x,y)] \\ 0, & \text{otherwise} \end{cases} \quad (3)$$

对图 2 中的参考图像 $I_r(x,y)$ 与目标图像 $I_c(x,y)$ 进行二值化处理,得到二值化图像 $I_r^B(x,y)$ 与 $I_c^B(x,y)$, 如图 3 所示。



图 2 (a) 参考图像 $I_r(x,y)$, (b) 目标图像 $I_c(x,y)$

Fig. 2 (a) Reference image $I_r(x,y)$, (b) target image $I_c(x,y)$

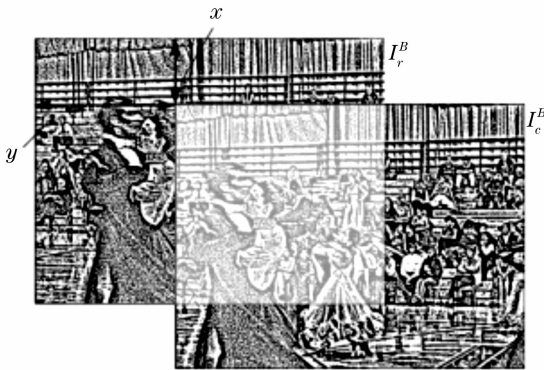


图 3 视点偏移-匹配误差函数计算

Fig. 3 Calculation of view offset-mismatch function

3.2 区域视差估计

不失一般性,考虑相邻视点视图间同时存在水平方向与垂直方向的视差,对于得到的二值化

图像 $I_r^B(x,y)$ 与 $I_c^B(x,y)$ 按照图 3 所示逐像素改变相对偏移量,并以(4)式所示的重叠区域的平均绝对差值(MAD)作为视点偏移-匹配误差函数：

$$F_{VO-MF}(x,y) = \frac{1}{(h-y)(w-x)} \times \sum_{i=0}^{w-x-1} \sum_{j=0}^{h-y-1} |I_r^B(i+x,j+y) - I_c^B(i,j)|, \quad (4)$$

$0 \leq x \leq X, 0 \leq y \leq Y$

其中 w 与 h 分别是图像的宽度与高度; i,j 分别表示图像中像素点的水平与垂直坐标; x,y 取整像素值,是 $I_r^B(x,y)$ 与 $I_c^B(x,y)$ 间的相对偏移量; X,Y 规定了偏移量的取值范围。

为了便于观察,对 $F_{VO-MF}(x,y)$ 取倒数得到 $1/[F_{VO-MF}(x,y) + c]$,并将 $1/[F_{VO-MF}(x,y) + c]$ 曲面及对应等高线在图 4 中给出。 c 为一小常量,避免产生

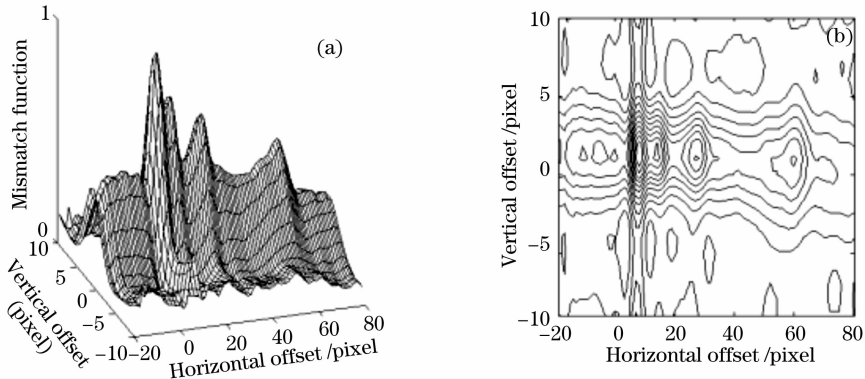


图 4 (a) $1/[F_{VO-MF}(x,y) + c]$ 曲面, (b) 相应的等高线图

Fig. 4 (a) $1/[F_{VO-MF}(x,y) + c]$ curved surface, (b) corresponding contour map

计算溢出。发现图 4(a) 曲面中一些局部极大值点与某些真实场景中对象所对应图像区域的视差有关。换言之, 场景中一个或几个对象所在图像区域的视差可由某个局部极大值点所对应的视点偏移量表示。因此可以简单地通过搜索图 4(a) 曲面中的所有局部极大值点, 也就是 $VO-MF(x, y)$ 中的所有局部极小值点, 确定图像中深度对象的数量以及这些深度对象所对应的区域视差向量, 将这组区域视差向量记为 $RD_O = \{RD_{O,1}, RD_{O,2}, \dots, RD_{O,N}\}$ 。

图像二值化处理能够有效地提取图像高频信息, 有利于找出 $F_{VO-MF}(x, y)$ 中较多的局部极小值点, 从而提高区域分割与区域视差估计算法性能。

显然, $F_{VO-MF}(x, y)$ 中的全局极小值即为多视点视图对之间的全局视差估计值^[12]。

3.3 深度对象分割

分析视点偏移-匹配误差函数可知, 当二值化多视点视图 $I_r^b(x, y)$ 与 $I_c^b(x, y)$ 之间的偏移量与目标视图 $\bar{I}_c(x, y)$ 中某一深度对象的视差值相等时, 该函数中就会出现一个局部极小值。根据这一特点, 我们提出一种基于深度对象概念的区域划分算法。首先将 $I_c(x, y)$ 划分为 $M \times M$ 大小的图像块, 然后按照偏移量 $RD_{O,i} (i = 1, \dots, N)$ 计算图像块 B_k 与其在参考视图中对应块的匹配误差, 选择使匹配误差最小化的区域视差 $RD_{O,i}$ 作为该图像块的视差 $RD_{Opt}^{B_k}$, 即

$$RD_{Opt}^{B_k} = \operatorname{argmin}_{RD_{O,i} \in RD_O} \{MAD_{B_k}(RD_{O,i})\}, \quad (5)$$

$$i = 1, 2, \dots, N$$

式中匹配误差 $MAD_{B_k}(RD_{O,i})$ 为

$$MAD_{B_k}(RD_{O,i}) = \frac{1}{M \times M} \sum_{(x,y) \in B_k} |I_c^b(x, y) - I_r^b[x + \Delta x(RD_{O,i}), y + \Delta y(RD_{O,i})]|, \quad (6)$$

式中 (x, y) 为图像块 B_k 内像素点坐标, $\Delta x(\cdot)$ 与 $\Delta y(\cdot)$ 分别得到所输入向量的水平与垂直分量。

$I_c(x, y)$ 中与 B_k 具有相同最佳匹配视差 $RD_{Opt}^{B_k}$ 的所有图像块的集合构成了一个图像区域, 从而将

$I_c(x, y)$ 分割为一组深度对象, 记为 $DO_O = \{DO_{O,1}, DO_{O,2}, \dots, DO_{O,N}\}$ 。

通过以上区域划分, 可能存在某些区域中图像块数量较少, 不能够有效表示图像中深度对象的情况。因此当某区域中图像块数量小于阈值 T 时, 将该区域拆分、合并到其它区域。区域合并方法具体描述如下:

若具有最佳匹配视差 $RD_{O,l}$ 的区域 $DO_{O,l}$ 中的图像块数量

$$N_{B_j} \leq T = k(BN_{I_c}/N), \quad (7)$$

$$B_j \in DO_{O,l}$$

则拆分 $DO_{O,l}$, 将其中所有图像块合并到其他区域

$$RD_{Opt}^{B_j} = \operatorname{argmin}_{RD_{O,i} \in RD_O} \{MAD_{B_j}(RD_{O,i})\}, \quad (8)$$

$$i = 1, 2, \dots, N, i \neq l$$

(7) 式中阈值 T 可随当前区域总数 N 自适应调整; BN_{I_c} 为 $I_c(x, y)$ 中图像块总数; 参量 k 的取值范围为 $(0, 1)$, 其值将决定算法最终划分得到深度对象的大小。

迭代执行上面的区域合并过程, 直到所有区域中图像块数量均大于阈值 T , 得到最终深度对象分割结果。

4 实验结果与分析

从 MPEG 组织提供的多视点视频序列集^[13] 中选择聚焦良好, 且各视点间没有亮度、色度偏差的多视点图像对 (同一时刻, 两个不同视点的图像) BR 与 AK, 对深度对象分割与区域视差估计算法进行测试, 并选择 16×16 像素大小的图像块作为深度对象分割的基本单元。表 1 给出 BR 与 AK 图像对参量, 表 2 给出区域视差估计结果, 包括区域分割数量与每一个区域的区域视差向量。图 5、图 6 分别是 BR 与 AK 图像对中的目标图像及其区域分割结果。整体看来, 提出的深度对象分割算法能够有效地划分图像中各深度层区域。

表 1 测试图像对参量

Table 1 Parameters of testing image pairs

Image pair	Spatial resolution	Block size	Video sequence of image pair	View index of reference image	View index of target image	Temporal index	Camera arrangement
BR	640×480	16×16	Ballroom	4	5	64	parallel
AK	640×480	16×16	Akko & Kayo	6	8	160	parallel

表 2 区域视差估计结果
Table 2 Regional disparity estimation results

Image pair	Number of Regional disparities	Regional disparity vectors (d_x, d_y)
BR	4	(7,0), (13,0), (26,0), (59,0)
AK	4	(23,0), (27,-1), (37,0), (47,0)



图 5 BR 中的目标图像(a)与分割结果(b)

Fig. 5 Target image of BR (a) and the region partition result (b)



图 6 AK 中的目标图像(a)与分割结果(b)

Fig. 6 Target image of AK (a) and the region partition result (b)

为了验证估计得到的区域视差的准确性,以各图像对中的参考图像作为唯一的编码参考图像,采用基于 H. 264 标准的 JMVM 参考软件^[14]对各测试图像对中的目标图像编码。编码器量化参量(QP)设置为 29,采用全搜索模式进行运动估计,并使用率失真优化(RDO)与基于内容的自适应二进制算术编码(CABAC)进行模式选择与符号编码。分别使用 JMVM 中位移向量预测器的输出与前面实验中估计得到的区域视差向量作为图像块位移估值的初始搜索点,设置不同的位移向量搜索范围进行测试。

表 3、表 4 中的实验结果表明,使用区域视差作为初始搜索点可以显著降低编码比特开销,且能够保持重建图像质量基本不变甚至有所提高。由于图像深度特性,或者说图像区域视差取值为图像自身属性,因此编码效率改善与图像内容密切相关。表 2 中的区域视差估计结果表明,图像对 AK 的区域视差较大,因此在搜索范围较小的情况下, JMVM

无法在参考图像中找到编码块的匹配块,多数编码块选择帧内预测编码模式,导致编码效率严重下降。而使用本文算法估计得到的区域视差能够在位移估值过程中为每一个编码块提供较准确的初始搜索点,增加帧间预测编码块的数量,进而显著改善编码效率。由于图像对 BR 的区域视差较小, JMVM 中原有的位移向量预测器仍然能够为目标图像中部分编码块提供较准确的位移向量预测值,因此基于区域视差编码相对于 JMVM 的改善并不十分显著。从上面分析可以看出,在参考图像与目标图像存在较大视差的情况下,基于区域视差进行帧间位移估值可以得到更高的编码效率。

从整体来看,使用区域视差作为位移估值的初始搜索点进行编码,所需的比特开销与重建图像质量基本不受搜索范围改变的影响,也就是说仅需要使用小的搜索范围即可得到与使用大搜索范围的 JMVM 参考软件相同甚至更好的编码效率。这说明该算法估计得到的区域视差具有较高精度。

表 3 JMVM 与本文算法编码 BR 结果对比

Table 3 Comparison of coding results of BR using JMVM and the proposed algorithm

Search range /pixel	Bit expenditure /bit			PSNR /dB		
	JMVM	Proposed	Reduction /%	JMVM	Proposed	Gain
4	119096	97856	17.83	38.53	38.43	-0.10
8	115096	97488	15.30	38.53	38.43	-0.10
16	106944	96912	9.38	38.51	38.44	-0.07
32	98664	97040	1.65	38.46	38.43	-0.03
64	97184	96808	0.39	38.43	38.44	0.01
96	97336	96960	0.39	38.42	38.43	0.01

表 4 JMVM 与本文算法编码 AK 结果对比

Table 4 Comparison of coding results of AK using JMVM and proposed algorithm

Search range /pixel	Bit expenditure /bit			PSNR /dB		
	JMVM	Proposed	Reduction /%	JMVM	Proposed	Gain
4	168320	89784	46.66	39.75	40.48	0.73
8	148544	89784	39.56	39.93	40.50	0.57
16	124000	89360	27.93	40.15	40.50	0.35
32	90440	89504	1.03	40.48	40.51	0.03
64	89920	89448	0.52	40.50	40.52	0.02
96	89992	89504	0.54	40.50	40.51	0.01

5 结 论

提出一种基于场景深度特征的多视点视频深度区域分割与区域视差估计算法,其优点在于不需要计算图像的视差场,直接提取图像中各区域的深度特征,通过最小化图像块匹配误差完成分割过程,因此与其它区域分割算法相比能够显著降低运算量。实验表明该算法能够有效地分割多视点视频中的深度对象,并准确地估计各深度对象的区域视差。

参 考 文 献

- 1 ISO/IEC JTC1/SC29/WG11. Requirements on multi-view video coding v. 6 [R]. ISO/IEC JTC1/SC29/WG11, Doc. N8064, Apr. 2006
- 2 Fuhui Long, Dagan Feng, Hanchuan Peng *et al.*. Extracting semantic video objects [J]. *IEEE Computer Graphics and Applications*, 2001, **21**(1): 48~55
- 3 P. Salembier, F. Marqués. Region-based representations of image and video; segmentation tools for multimedia services[J]. *IEEE Transactions on Circuits and Systems for Video Technology*, 1999, **9**(8): 1147~1169
- 4 Arun Hampapur, Kamesh Jain, Terry Weymouth. Digital video segmentation[C]. *Proceedings of the second ACM international conference on Multimedia*, 1994. 357~364
- 5 Guo Ping, Lu Hanqing. A study on Bayesian probabilistic image automatic segmentation[J]. *Acta Optica Sinica*, 2002, **22**(12): 1479~1483

郭 平,卢汉清. 贝叶斯概率图像自动分割研究[J]. 光学学报,

2002, **22**(12): 1479~1483

- 6 Liu Guixi, Shao Mingli, Liu Xianhong *et al.*. Video moving object auto-extraction in real scene [J]. *Acta Optica Sinica*, 2006, **26**(8): 1150~1155
- 刘贵喜,邵明礼,刘先红 等. 真实场景下视频运动目标自动提取方法[J]. 光学学报, 2006, **26**(8): 1150~1155
- 7 Anastasios D. Doulamis, Nikolaos D. Doulamis, Klimis S. Ntalianis *et al.*. Unsupervised semantic object segmentation of stereoscopic video sequences [C]. *Proceedings of IEEE International Conference on Information Intelligence and Systems*, 1999. 527~533
- 8 Klimis S. Ntalianis, Nikolaos D. Doulamis, Anastasios D. Doulamis *et al.*. An active contour-based video object segmentation scheme for stereoscopic video sequences [C]. *Proceedings of IEEE International Electromechanical Conference*, 2000, **2**: 554~557
- 9 Nicole Atzpadin, Serap Askar, Peter Kauff *et al.*. New concept for joint disparity estimation and segmentation for real-time video processing[C]. *Picture Coding Symposium*, 2003. 23~25
- 10 Polin Lai, Antonio Ortega, Purrin Pandit *et al.*. Adaptive reference filtering for MVC[R]. ISO/IEC JTC1/SC29/WG11 and ITU-T Q6/SG16, Doc. JVT-W065, Apr. 2007
- 11 Jeng-Hung Luo, Chung-Neng Wang, Tihao Chiang. A novel all-binary motion estimation (ABME) with optimized hardware architectures[J]. *IEEE Transactions on Circuits and Systems for Video Technology*, 2002, **12**(8): 700~712
- 12 Yo-Sung Ho, Kwan-Jung Oh, Pil-Kyu Park *et al.*. Global disparity compensation for multi-view video coding[R]. ISO/IEC JTC1/SC29/WG11, Doc. M13581, Jul. 2006
- 13 Yeping Su, Anthony Vetro, Alioscha Smolic. Common test conditions for multiview video coding[R]. ISO/IEC JTC1/SC29/WG11 and ITU-T Q6/SG16, Doc. JVT-T207, Jul. 2006
- 14 Anthony Vetro, Yeping Su, Hideaki Kimata *et al.*. Joint multiview video model (JMVM) 1.0 [R]. ISO/IEC JTC1/SC29/WG11 and ITU-T Q6/SG16, Doc. JVT-T208, Jul. 2006