

SIFT 特征匹配算法研究*

王国美¹, 陈孝威¹

(贵州大学 计算机科学与技术学院, 贵州 贵阳 550025)

摘要: SIFT 特征匹配算法是目前特征匹配研究领域的热点, 其匹配能力较强, 可以处理图像间发生平移、旋转、仿射变换的匹配, 对任意角度拍摄的图像也具备较稳定的特征匹配能力。实验结果表明, 该算法具有较强的匹配能力和鲁棒性, 是一种较好的特征匹配算法。

关键词: 虚拟现实; 特征匹配; 图像拼接; 全景图

中图分类号: TP37 **文献标识码:** A **文章编号:** 1671-5322(2007)02-0001-05

图像匹配是虚拟现实、计算机视觉等领域的一个研究热点。目前, 有关研究人员对图像匹配技术进行了大量的研究, 提出了很多匹配算法: 有基于面积的方法^[1], 有基于比值的方法^[2], 有相位相关算法^[3]等等, 但是这些匹配算法都有一个共同点: 图像间的焦距要一致, 不能有尺度缩放, 旋转不能太大, 变形不能太明显, 还有光照变化、仿射变换等方面受到限制。随着图像技术和计算机技术的发展, 出现了基于特征的图像匹配技术。这种技术的优点是能处理具有不同特性的图像和图像间变形复杂的情况。缺点是特征的检测困难, 算法稳定性较差。针对特征匹配算法存在的不足, 经过计算机视觉多年的发展, 特征提取技术越来越稳定, 特别是尺度空间的特征检测器^[4-6]甚至可以稳定地对两幅位移很大的图像进行准确的特征检测和匹配。

在基于特征的匹配技术中, 其首要任务是提取稳定的特征, 并进行描述。常用的方法有基于空间关系的匹配算法、基于不变量描述子的匹配算法、金字塔和小波算法等等。其中文献[5]提出的 SIFT (Scale Invariant Feature Transform 即尺度不变特征变换) 特征匹配算法是目前国内外特征匹配研究领域取得比较成功的一种算法, 该算法匹配能力较强, 能提取稳定的特征, 可以处理两

幅图像之间发生平移、旋转、仿射变换、视角变换、光照变换情况下的匹配问题, 甚至在某种程度上对任意角度拍摄的图像也具备较为稳定的特征匹配能力, 从而可以实现差异较大的两幅图像之间的特征的匹配。

本文在研究以上各种匹配算法的基础上, 对匹配能力较强、适用范围较大的 SIFT 特征变换匹配算法进行比较详细和深入的介绍, 并用 Matlab 7.0 实现了该算法。

1 SIFT 特征匹配算法

SIFT 特征匹配算法是 David G. Lowe 在 2004 年总结了现有的基于不变量技术的特征检测方法的基础上, 提出的一种基于尺度空间的、对图像缩放、旋转甚至仿射变换保持不变性的特征匹配算法。SIFT 特征是图像的局部特征, 该特征对旋转、尺度缩放、亮度变化保持不变性, 对视角变化、仿射变换、噪声也保持一定程度的稳定性。

SIFT 特征匹配算法分两个阶段来实现: 第 1 阶段是 SIFT 特征的生成, 即从多幅待匹配图像中提取出对尺度缩放、旋转、亮度变化无关的特征向量; 第 2 阶段是 SIFT 特征向量的匹配。

1.1 SIFT 特征向量的生成

在生成 SIFT 特征向量之前, 先对图像进行规

* 收稿日期: 2006-12-21

基金项目: 国家教育部春晖计划基金(200203)

作者简介: 王国美(1975-), 女, 贵州遵义市人, 博士研究生, 主要研究方向为虚拟现实技术、多媒体技术。

一化处理,然后扩大图像为原来的两倍,预滤波剔除噪声,得到高斯金字塔的最底层即第 1 阶的第 1 层。一幅图像 SIFT 特征向量的生成算法总共包括 4 步。

1.1.1 尺度空间极值检测

尺度空间理论最早出现于计算机视觉领域,当时其目的是模拟图像数据的多尺度特征^[7]。随后 Koendetink 利用扩散方程来描述尺度空间滤波过程,并由此证明高斯核是实现尺度变换的唯一变换核^[8]。Lindeberg^[9]、Babaud^[10]等人通过不同的推导进一步证明高斯核是唯一的线性核。因此,尺度空间理论的主要思想是利用高斯核对原始图像进行尺度变换,获得图像多尺度下的尺度空间表示序列,对这些序列进行尺度空间特征提取。

二维高斯核的定义如公式(1)所示,其中 σ 代表了高斯正态分布的方差:

$$G(x,y,\sigma) = \frac{1}{2\pi\sigma^2} e^{-(x^2+y^2)/2\sigma^2} \quad (1)$$

对于二维图像 $I(x,y)$,在不同尺度下的尺度空间表示 $L(x,y,\sigma)$ 可由图像 $I(x,y)$ 与高斯核 $G(x,y,\sigma)$ 的卷积得到,如公式(2)所示:

$$L(x,y,\sigma) = G(x,y,\sigma) \times I(x,y) \quad (2)$$

在公式(2)中, L 表示尺度空间。 (x,y) 代表图像 I 上的点。 σ 是尺度因子,其值越小则表征该图像被平滑得越小;其值越大则表征该图像被平滑得越大。大尺度对应于图像的概貌特征,小尺度对应于图像的细节特征。因此,选择合适的尺度因子平滑是建立尺度空间的关键。

在这一步里面,主要是建立高斯金字塔和 DOG(Difference of Gaussian)金字塔,然后在 DOG 金字塔里面进行极值检测,以初步确定特征点的位置和所在尺度。

(1)建立高斯金字塔

为了得到在不同尺度空间下的稳定特征点,将图像 $I(x,y)$ 与不同尺度因子下的高斯核 $G(x,y,\sigma)$ 进行卷积操作,构成高斯金字塔。

高斯金字塔有 o 阶,一般选择 4 阶,每一阶有 s 层尺度图像, s 一般选择 5 层。高斯金字塔的构成如图 1 所示。

在高斯金字塔的构成中要注意,第 1 阶的第 1 层是放大 2 倍的原始图像,其目的是为了得到更多的特征点;在同一阶中相邻两层的尺度因子比例系数是 k ,则第 1 阶第 2 层的尺度因子是 $k\sigma$,

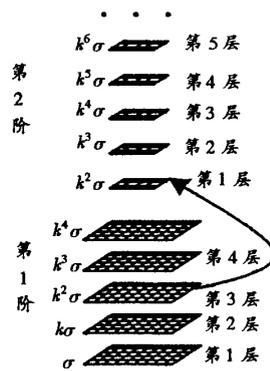


图 1 高斯金字塔

Fig. 1 Gaussian pyramid

然后其它层以此类推则可;第 2 阶的第 1 层由第一阶的中间层尺度图像进行子抽样获得,其尺度因子是 $k^2\sigma$,然后第 2 阶的第 2 层的尺度因子是第 1 层的 k 倍即 $k^3\sigma$ 。第 3 阶的第 1 层由第 2 阶的中间层尺度图像进行子抽样获得。其它阶的构成以此类推。

(2)建立 DOG 金字塔

DOG 即相邻两尺度空间函数之差,用 $D(x,y,\sigma)$ 来表示,如公式(3)所示:

$$D(x,y,\sigma) = (G(x,y,k\sigma) - G(x,y,\sigma)) \times I(x,y) = L(x,y,k\sigma) - L(x,y,\sigma) \quad (3)$$

DOG 金字塔通过高斯金字塔中相邻尺度空间函数相减即可,如图 2 所示。在图中,DOG 金字塔的第 1 层的尺度因子与高斯金字塔的第 1 层是一致的,其它阶也一样。

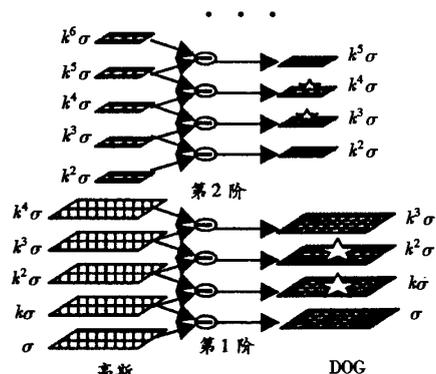


图 2 DOG 金字塔

Fig. 2 DOG pyramid

(3)DOG 空间的极值检测

在上面建立的 DOG 尺度空间金字塔中,为了检测到 DOG 空间的最大值和最小值,DOG 尺度

空间中中间层(最底层和最顶层除外)的每个像素点需要跟同一层的相邻 8 个像素点以及它上一层和下一层的 9 个相邻像素点总共 26 个相邻像素点进行比较,以确保在尺度空间和二维图像空间都检测到局部极值,如图 3 所示。

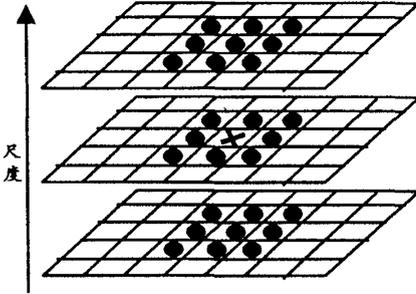


图 3 尺度空间极值检测
Fig. 3 Detection of scale - space extrema

在图 3 中,标记为叉号的像素若比相邻 26 个像素的 DOG 值都大或小,则该点将作为一个局部极值点,记下它的位置和对应尺度。

1.1.2 精确定位特征点位置

由于 DOG 值对噪声和边缘较敏感,因此,在上面 DOG 尺度空间中检测到局部极值点还要经过进一步的检验才能精确定位为特征点。下面对局部极值点进行三维二次函数拟和以精确定特征点的位置和尺度,尺度空间函数 $D(x, y, \sigma)$ 在局部极值点 (x_0, y_0, σ) 处的泰勒展开式如公式(4)所示。

$$D(x, y, \sigma) = D(x_0, y_0, \sigma_0) + \frac{\partial D^T}{\partial X} X + \frac{1}{2} X^T \frac{\partial^2 D}{\partial X^2} X \quad (4)$$

其中 $X = (x, y, \sigma)^T$, $\frac{\partial D}{\partial X} = \begin{bmatrix} \frac{\partial D}{\partial x} \\ \frac{\partial D}{\partial y} \\ \frac{\partial D}{\partial \sigma} \end{bmatrix}$,

$$\frac{\partial^2 D}{\partial X^2} = \begin{bmatrix} \frac{\partial^2 D}{\partial x^2} & \frac{\partial^2 D}{\partial x \partial y} & \frac{\partial^2 D}{\partial x \partial \sigma} \\ \frac{\partial^2 D}{\partial y \partial x} & \frac{\partial^2 D}{\partial y^2} & \frac{\partial^2 D}{\partial y \partial \sigma} \\ \frac{\partial^2 D}{\partial \sigma \partial x} & \frac{\partial^2 D}{\partial \sigma \partial y} & \frac{\partial^2 D}{\partial \sigma^2} \end{bmatrix}$$

$$\frac{\partial^2 D}{\partial \sigma^2} = \frac{(D_{k+1}^{x,y} - D_k^{x,y}) - (D_{k-1}^{x,y} - D_k^{x,y})}{4}$$

($k \rightarrow$ 当前层, $k - 1 \rightarrow$ 下层, $k + 1 \rightarrow$ 上一层)

$$\frac{\partial^2 D}{\partial x \partial \sigma} = \frac{(D_{k+1}^{x+1,y} - D_{k+1}^{x-1,y}) - (D_{k-1}^{x+1,y} - D_{k-1}^{x-1,y})}{4}$$

⋮

公式(4)中的一阶和二阶导数是通过附近区域的差分来近似求出的,列出其中的几个,其它的二阶导数以此类推。通过对公式(4)求导,并令其为 0,得出精确的极值位置 X_{max} ,如公式(5)所示:

$$X_{max} = - \left(\frac{\partial^2 D}{\partial X^2} \right)^{-1} \frac{\partial D}{\partial X} \quad (5)$$

在上面精确确定的特征点中,同时要去掉低对比度的特征点和不稳定的边缘响应点,以增强匹配稳定性、提高抗噪声能力。

去除低对比度的特征点:把公式(5)代到公式(4)中,只要前两项,得到公式(6):

$$D(X_{max}) = D + \frac{1}{2} \frac{\partial D^T}{\partial X} \quad (6)$$

通过式(6)计算出 $D(X_{max})$,若 $|D(X_{max})| \geq 0.03$,则该特征点就保留下来,否则就丢弃。

去除不稳定的边缘响应点:海森矩阵如公式(7)所示,其中的偏导数是上面确定的特征点处的偏导数,它也是通过附近区域的差分来近似估计的。

$$H = \begin{bmatrix} D_{xx} & D_{xy} \\ D_{xy} & D_{yy} \end{bmatrix} \quad (7)$$

通过 2×2 的海森矩阵 H 来计算主曲率,由于 D 的主曲率与 H 矩阵的特征值成比例,根据文献[5],不具体求特征值,求其比例 ratio。设 α 是最大幅值特征, β 是次小的, $r = \frac{\alpha}{\beta}$,则 ratio 如公式(8)所示。

$$\begin{aligned} Tr(H) &= D_{xx} + D_{yy} = \alpha + \beta \\ Det(H) &= D_{xx}D_{yy} - (D_{xy})^2 = \alpha\beta \\ ratio &= \frac{Tr(H)^2}{Det(H)} = \frac{(\alpha + \beta)^2}{\alpha\beta} = \frac{(r + 1)^2}{r} \quad (8) \end{aligned}$$

通过公式(8)求出 ratio,常取 $r = 10$,若 $ratio \leq \frac{(r+1)^2}{r}$,则保留该特征点,否则就丢弃。

1.1.3 确定特征点主方向

利用特征点邻域像素的梯度方向分布特性为每个特征点指定方向参数,使算子具备旋转不变性。

$$m(x, y) = \frac{1}{\sqrt{(L(x+1, y) - L(x-1, y))^2 + (L(x, y+1) - L(x, y-1))^2}}$$

$$\theta(x,y) = \tan^{-1}((L(x,y+1) - L(x,y-1)) / (L(x+1,y) - L(x-1,y))) \quad (9)$$

公式(9)为(x,y)处的梯度值和方向。L为所用的尺度为每个特征点各自所在的尺度,(x,y)要确定是哪一阶的哪一层。在实际计算过程中,在以特征点为中心的邻域窗口内采样,并用梯度方向直方图统计邻域像素的梯度方向。梯度直方图的范围是0°~360°,其中每10°一个柱,总共36个柱。梯度方向直方图的峰值则代表了该特征点处邻域梯度的主方向,即作为该特征点的方向。

在梯度方向直方图中,当存在另一个相当于主峰值80%能量的峰值时,则将这个方向认为是该特征点的辅方向。一个特征点可能会被指定具有多个方向(一个主方向,一个以上辅方向),这可以增强匹配的鲁棒性。

通过上面的3步,图像的特征点已检测完毕,每个特征点有3个信息:位置、对应尺度、方向。

1.1.4 生成 SIFT 特征向量

首先将坐标轴旋转为特征点的方向,以确保旋转不变性。

接下来以特征点为中心取8×8的窗口(特征点所在的行和列不取)。在图4(a)中,中央黑点为当前特征点的位置,每个小格代表特征点邻域所在尺度空间的一个像素,箭头方向代表该像素的梯度方向,箭头长度代表梯度模值,图中圈内代表高斯加权的范围(越靠近特征点的像素,梯度方向信息贡献越大)。然后在每4×4的图像小块上计算8个方向的梯度方向直方图,绘制每个梯度方向的累加值,形成一个种子点,如图4(b)所示。此图中一个特征点由2×2共4个种子点组成,每个种子点有8个方向向量信息,可产生2×2×8共32个数据,形成32维的SIFT特征向量即特征点描述器,所需的图像数据块为8×8。这种邻域方向性信息联合的思想增强了算法抗噪声的能力,同时对于含有定位误差的特征匹配也提供了较好的容错性。

实际计算过程中,为了增强匹配的稳健性,文献[5]建议对每个特征点使用4×4共16个种子点来描述,每个种子点有8个方向向量信息,这样对于一个特征点就可以产生4×4×8共128个数据,最终形成128维的SIFT特征向量,所需的图像数据块为16×16。此时SIFT特征向量已经去除了尺度变化、旋转等几何变形因素的影响,再继

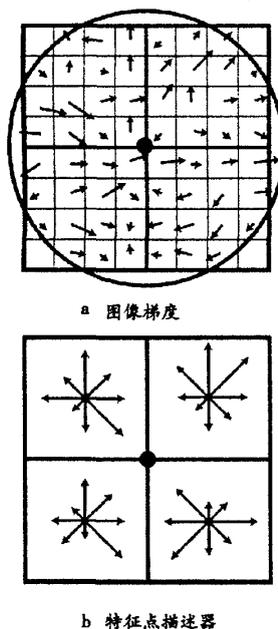


图4 图像梯度及特征点描述器

Fig.4 The image gradients and keypoint descriptor

续将特征向量的长度归一化,则可以进一步去除光照变化的影响。

当两幅图像的SIFT特征向量即特征描述器生成后,下一步就是进行特征向量的匹配。

1.2 SIFT 特征向量的匹配

首先,进行相似性度量。一般采用各种距离函数作为特征的相似性度量,如欧氏距离、马氏距离等。

通过相似性度量得到图像间的潜在匹配。本文中采用欧氏距离作为两幅图像间的相似性度量。获取SIFT特征向量后,采用优先k-d树进行优先搜索来查找每个特征点的2近似最近邻特征点。在这两个特征点中,如果最近的距离除以次近的距离少于某个比例阈值,则接受这一对匹配点。降低这个比例阈值,SIFT匹配点数目会减少,但更加稳定。

其次,消除错配。通过相似性度量得到潜在匹配对,其中不可避免会产生一些错误匹配,因此需要根据几何限制和其它附加约束消除错误匹配,提高鲁棒性。常用的去外点方法是RANSAC随机抽样一致性算法,常用的几何约束是极线约束关系。

2 实验结果及分析

本文采用 Matlab7.0 实现上述算法。首先提取 SIFT 特征,然后再进行特征匹配。在图 5 中,a 和 b 为原始图像,在拍摄时进行了相机平移、转动和变焦操作;c、d 分别为第 1、第 2 幅图像中检测出的 SIFT 特征,用箭头来表示,箭头的起点代表

特征点的位置,箭头的长度代表该特征点所在的尺度,箭头的方向代表该尺度下特征点所处领域的主梯度方向;e 是特征匹配的结果。

从图中可看出,尽管两幅图像有旋转、缩放及亮度等变换,但两幅图像中找到的 SIFT 特征匹配对,其位置信息、方向性和尺度信息是准确的。

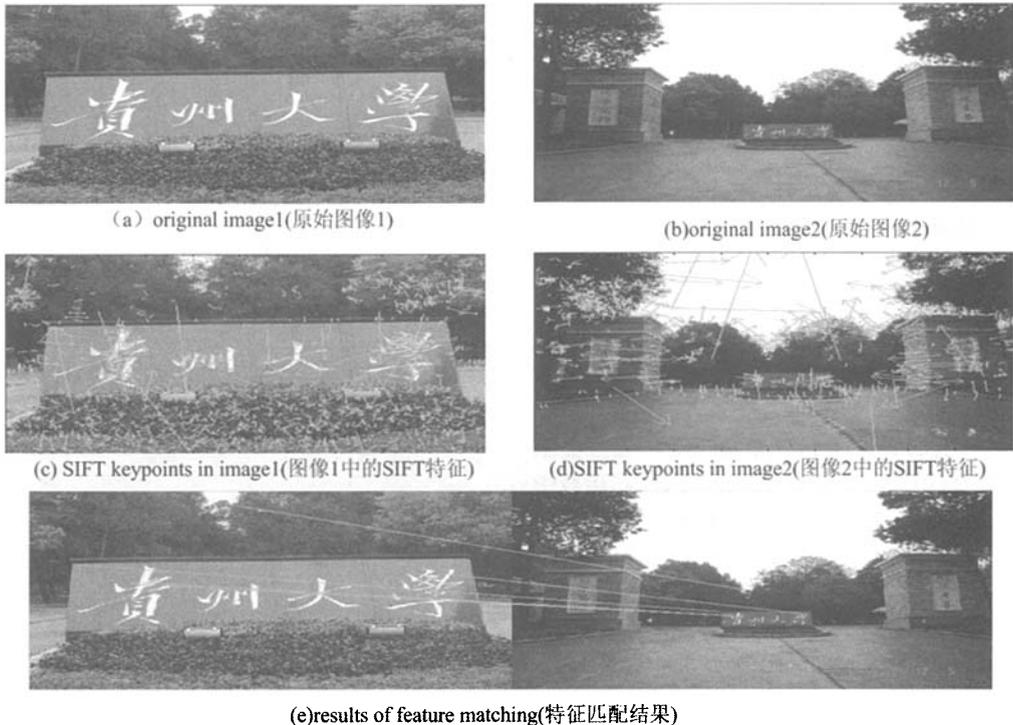


图 5 实验结果

Fig. 5 experiments

3 结论

本文对基于特征的匹配算法做了较深入的研究。在特征提取过程中,由于 SIFT 算法需要在各个尺度上进行计算,其时间复杂度相对较高,在今

后的研究中,希望在算法的实时性方面能够做出改进;在特征匹配方面,算法的实时性和特征匹配的准确度都有待于改进,在图像处理中,这仍然是一个没有完全解决的问题,这将促使新的匹配方法不断出现,也是今后需要进一步研究的方面。

参考文献:

- [1] 汪成为. 灵境(虚拟现实)技术的理论、实践及应用[M]. 南宁:广西科学技术出版社,1996.
- [2] 钟力,胡小锋. 重叠图像拼接算法[J]. 中国图像图形学报,1998(3):365-369.
- [3] Kuglin C D, Hines D C. The Phase Correlation Image alignment method[C]. IEEE Conference on Cybernetics and Society, New York,1975:163-165.
- [4] David G Lowe. Object Recognition from Local Scale - Invariant Features[C]. 7th International Conference on Computer Vision,1999:1150-1157.

(下转第 10 页)

参考文献:

- [1] 吴茂念. 广义 Fibonacci 数列几个前 n 项和式[J]. 贵州大学学报:自然科学版,2005(增刊):6-8.
 [2] 吴茂念. 广义 Fibonacci 数列一些前 n 项和式[J]. 贵州大学学报:自然科学版,2005(4):343-347.
 [3] 宋卫星,杨巧梅.“魔(n, k)方”与广义 Fibonacci 数列[J]. 数学通报,2001(4):41-43.
 [4] 康士凯. 斐波那契数列[M]. 上海:上海科技教育出版社,1992:41-49.
 [5] 高山珍. Lucas 数列的几个性质[J]. 贵州大学学报:自然科学版,2002(4):291-296.

Some Formals of Sum about General Fibonacci Sequence

WU Mao-nian

(School of Science, Guizhou University, Guizhou Guiyang 550025, China)

Abstract: The article has demonstrated an elementary proof about the sum of the first n item difference less than 6 of the generalized Fibonacci sequence. Thus, the sum of the first n item difference less than 6 of Fibonacci sequence, Lucas sequence could be obtained and its numerical value could be calculated by the general formals of these sequences.

Keywords: Fibonacci sequence; Lucas sequence; Generalized Fibonacci Sequence; the sum of the first n item

(上接第 5 页)

- [5] David G Lowe. Distinctive Image Features from Scale - Invariant Interest Points[J]. International Journal of Computer Vision, 2004, 60(2):91-110.
 [6] Harris C Stephens. A Combined Corner and Edge Detector[J]. 4th Alvey Vision Conference. 1988, 60(2):147-151.
 [7] Faugeras O, Robert L. What can two images tell us about the third one[C]. Proceedings of the Europe Conference on Computer Vision, Sweden, 1994.
 [8] Koenderink J. The structure of images[J]. Biological Cybernetics, 1984, 50:363-396.
 [9] Lindeberg T. Scale - Space for discrete Signals[J]. IEEE Transactions PAMI, 1980, 207:187-217.
 [10] Babaud J, Witkin A P, Baudin Metal. Uniqueness of the Gaussian kernel for scale - space filtering[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1996, 8(1):26-33.

Study on a Naw Algorithm of Feature Matching - SIFT

WANG Guo-me¹, CHEN Xiao-wei¹

(College of Computer Science and Technology, Guizhou University, Guizhou Guiyang 550025, China)

Abstract: A new algorithm of feature matching - SIFT(Scale Invariant Feature Transform) which has become a hot topic and difficult research area is introduced. At present the algorithm, whose matching ability is more strong and can dispose of matching problem with translation, rotation and affine distortion between images and to a certain extent is with more stable feature matching ability for images which are shot from random different angles. The experiments show that the algorithm is of stronger matching ability and robustness, which is a better feature matching algorithm.

Keywords: virtual reality; feature matching; image mosaicing; panorama