Distinctive Image Features from Scale-Invariant Keypoints-SIFT算法译文

从尺度不变的关键点选择可区分的图像特征

David G.Lowe

温哥华不列颠哥伦比亚省加拿大英属哥伦比亚大学计算机科学系

Lowe@cs.ubc.ca

2003年1月10日接受，2004年1月7日修改，2004年1月22日采用

摘要：本文提出了一种从图像中提取独特不变特征的方法，可用于完成不同视角之间目标或场景的可靠匹配的方法。这种特点对图像的尺度和旋转具有不变性。并跨越很大范围的对仿射变换，三维视点的变化，添加的噪音和光照变化的图像匹配具有鲁棒性。特征是非常鲜明的，场景中的一个单一特征和一个许多图像的大型特征数据库也有很高的概率进行正确匹配。本文还介绍了一个使用该功能来识别目标的方法。通过将个别特征与由已知目标特征组成的数据库进行快速最近邻算法的匹配，然后使用Hough变换来识别属于单一目标的聚类（clusters），最后通过最小二乘解执行一致的姿态参数的核查确认。这种识别方法可以在有力确定对象之间的聚类和遮挡的同时实现近实时性能。

关键词：不变特征，目标识别，尺度不变性，图像匹配

1. 引言

图像匹配是计算机视觉领域中很多问题的关键，包括目标和场景识别、多幅影像进行三维构建、立体对应（correspondence）、运动追踪等。本文描述的图像特征很实用，因为它具备很多可以将一个目标或场景的不同影像进行匹配的特性。这些特征对于图像尺度和旋转具有不变性，并在光照变化和三维相机视点变化的情况下具有部分的不变性。它在空间域和频率域都可以很好地定位，减少了遮挡（occlusion）、聚类和噪音的影响。有了有效的算法，海量的特征就可以从典型的图像中提取出来。另外，这些特征是非常鲜明的，使一个单一特征可以无误地与大型数据库中的特征进行匹配，为目标和场景识别提供了基础。

通过一个层叠的过滤算法将提取这些特征的代价最小化，这样，最昂贵的运算仅在最初测试通过处。下面是生成图像特征集计算的一些主要步骤：

1） 尺度空间极值探测：第一阶段对整个尺度和图像位置进行搜索。通过使用高斯差分函数来有效地识别对于尺度和方向具有不变性的可能的兴趣点。

2） 关键点定位：在每一个候选区，都可以确定一个详细模型的位置和尺度。基于关键点的稳定性进行选择。

3） 定向任务：基于局部图像的梯度方向，给每个关键点指定一个或多个方向。所有随后的图像数据操作都是将每个特征的方向、尺度和位置进行相关变换得到的，因此这些变换具有不变性。

4） 关键点描述子：局部梯度是在每个关键点附近的区域所选尺度上测量得到的。这些可以转化成为一个允许显著的局部形状变化和光照变化的表示法。

这种方法被命名为尺度不变的特征转换法（SIFT），因为它可以基于局部特征把图像数据转换到尺度不变的坐标上。

该方法的一个重要方面是它生成了大量特征，它们密集的覆盖了整个图像尺度和位置。一幅500*500像素的典型图片可以产生约2000个稳定的特征（这个数字依赖于图像内容和几个参数的选择）。特征的数量对目标识别尤为重要，要具备探测杂乱背景下的小目标的能力，要求每个目标至少有三个特征被正确匹配才是可靠的识别。

对于图像匹配和识别，SIFT特征被第一个从一组参考图像中提取并存储在数据库中。一个新的图像通过将这幅新图像中的各个特征与原有数据库进行一一对比并基于欧氏距离找到候选的匹配特征。本文将讨论可以在大型数据库中快速执行的快速近邻算法。

关键点描述子是非常鲜明的，可以使单个特征在大型特征数据库中以很大概率进行正确匹配。然而，在杂乱的图像中，很多背景中的特征不能与数据库进行正确匹配，产生了很多错误的配对。通过确定与新图像在目标、目标的位置、尺度和定向一致的关键点的子集，可以将正确的匹配从匹配的全集中过滤出来。多种功能恰好与这些参数一致的可能性比任一个特征匹配错误的可能性要小很多。确定这些一致的聚类，可以通过一个高效的广义Hough变换的散列表快速执行。

每个拥有三个及三个以上特征与目标一致的聚类，它们的姿态都要进行下一步更精细的确认。首先，最小二乘估计是用于目标姿态的仿射近似。其他已识别的与此姿态相一致的图像特征以及异常值都忽略不计。最后，通过一个精细的计算可以得出一组可以表明目标存在的详细特征，并给出符合的准确度和可能的错误匹配数。经过所有的这些实验，可以得出这个结论：目标匹配的成功率很高。

2. 相关研究

使用一组局部兴趣点来进行图像匹配的发展可以追溯到1981年Moravec在立体匹配中使用的角探测器。Moravec的探测器在1988年被Harris和Stephens改进，在小的图像变动和近边缘区域具有了更高的重复性。Harris还展示了它在高效运动追踪和由运动恢复进行三维建模中的价值（Harris，1992），Harris的角探测器自此在很多其他的图像匹配工作中被广泛的使用。尽管这个特征探测器被称为角探测器，但它并不是只能选择角，而是可以在一个确定尺度的各个方向上选择所有具有大的梯度的图像位置。

该方法的最初应用是立体或短距离运动追踪，而后来被扩展到解决一些更困难的问题。Zhang等人在1995年在每个角的周围使用相关窗口来选择可能的匹配，使得Harris的角进行大幅图像范围的匹配成为可能。计算精确场景中两个视角间的几何约束的基础矩阵，移除异常值，同时移除那些与多数方法不一致的配对。同年（1995），Torr研发了一种类似的方法来进行大间距的运动匹配，使用几何约束来移除图像中移动刚体的异常值。

1997年，Schmid和Mohr的开创性工作展示了不变的局部特征匹配可以被扩展到解决一般的图像识别问题，即使用一个特征与大型图像数据库进行匹配。他们还使用Harris角探测器来选择兴趣点，但他们使用的是一个图像局部区域的旋转不变的描述子来代替相关窗口。这是特征可以在两幅图像之间进行任意方向变化时进行匹配。此外，他们还证明多特征匹配可以通过识别一致的匹配特征聚类，在遮挡和混杂的情况下完成一般的识别工作。

Harris角探测器对图像尺度的变化非常敏感。因此，对于不同尺度的图像匹配，Harris的角反射器并不能提供很好的基础。本文作者（Lowe）在1999年的早期工作中扩展了这种局部特征方法来实现尺度不变性。这个工作还阐述了一种新的局部描述子，可以降低对局部图像变形的敏感度（如三维视点的变换），同时找到更加鲜明的特征。本文提出了对这一方法更加深入的研发，并分析了这些早期的工作，在稳定性和特征不变性上进行了大量改进。

在之前的研究中，关于在尺度变换下表征（representation）的稳定识别占了很大的篇幅。最早在这个领域进行研究的有Crowley和Parker，1984年，他们在尺度空间发现了一种表征可以识别峰和脊，并把它们与树结构联系起来。然后，就可以在任意尺度变换的图像间进行树结构的匹配。在近期基于图像匹配的工作中，Shokoufandeh等人在1999年使用小波系数提出了一种更加鲜明的特征描述子。Lindeberg在1993-1994年对为特征探测识别一个合适并且一致的尺度这一问题进行了深入研究。他称之为尺度选择问题，我们在下面使用了这一结论。

最近，有了大量令人印象深刻的将局部特征扩展为全局仿射变换不变量的工作（Baumberg，2000；Tuytelaars和Van Gool，2000；Mikolajczyk和Schmid，2002；Schaffalitzky和Zisserman，2002；Brown和Lowe,2002）。这使得在变化的正射三维投影平面上的特征匹配具备了不变性，多数情况下采用对图像局部仿射框架进行重采样的方法。然而，还没有一个方法实现了完全的仿射不变性，由于充分勘探仿射空间的成本过高，因此他们用一个非仿射不变的方式对最初特征、尺度和位置进行选择。仿射框架与尺度不变的特征相比，对噪音更加敏感，因此，实践中除非在仿射变形与平面倾斜程度大于40度时（Mikolajczyk，2002），仿射特征比尺度不变的特征重复率要低。对于很多应用，更宽的仿射不变性可能并不重要，因为为了获得三维目标的非平面变化和遮挡的影响，瞄准视角至少每30度旋转一下视点（也就是说对于最靠近的瞄准视角，识别也是在15度以内进行的）。

尽管本文中的方法不具备完全的仿射不变性，但它使用了一种独特的方法来使局部描述子可以随着描述子很小的变化来显著地改变相关特征的位置。这种方法不仅使描述子可以在相当大范围的仿射变形时进行可靠地匹配，还可以使特征在非平面的三维视点变化时具有更好的鲁棒性。另一个优点是它可以提取出更多的有效特征，并可以识别大量特征。另一方面，在非常大尺度的视角变化下，仿射不变性是匹配平面非常有价值的属性，以后的研究应该在一个有效稳定的方式下，将这一点与非平面的三维视点不变性很好地结合的条件下开展。

还有许多其他的被推荐进行识别的特征类型，有的可以用于协助本文所述方法在不同环境中进行进一步的匹配工作。其中一种是利用图像轮廓或区域边缘的特征，可被用来减少目标边界附近的聚类背景所带来的干扰。Matas等人在2002年称他们的最大稳定极端区域可以产生大量具有良好稳定性的匹配特征。Mikolajczyk等人在2003年使用局部边缘（edge）而忽略附近的无关边缘，发现了一种新的描述子，即使在重叠背景聚类上狭窄形状的边界附近也可以在寻找稳定的特征。Nelson和Selinger在1998年使用基于图像轮廓分组的局部特征得到了很好的结果。类似的，Pope和Lowe在2000年使用的是基于图像轮廓的等级分类的特征，尤其是对于缺少详尽纹理的目标非常有用。

对于视觉识别的研究历史包括致力于不同的可被用作特征测量的其他图像属性数集的工作。Carneiro和Jepson在2002年描述了一种基于相位的局部特征，它们用相位来表示而不是局部空间频率的量级，这种方法更有利于光照不变量的提高。Schiele和Crowley在2000年建议使用多维直方图来概括图像区域内的测量值的分布。这种特征对于纹理明显的形状畸变的目标尤为有效。Basri和Jacobs在1997年证明了提取局部区域边界对于识别的价值。其他可以吸纳的有用属性有诸如颜色、运动、图形背景识别、区域形状描述子和立体景深提示等。当有对鲁棒性有提高的可以增强匹配成功率的新特征类型时，只要它们的计算成本对其他特征的影响较小，都可以简单地被局部特征方法采纳作为额外的特征。因此，以后的系统可能会由很多特征类型组合而成。

3. 尺度空间极值的发现

引言中已经提到了，我们使用一种高效的先识别候选位置然后进一步确认的层叠过滤方法来探测关键点。关键点探测的第一步是识别同一目标在不同视角下可被重复分配的位置和尺度。使用被称为尺度空间的尺度连续函数，通过搜索对所有尺度的稳定特征进行搜索，可以完成对图像尺度变换具有不变性的位置探测。（Witkin，1983）。

Koenderink和Lindeberg分别在1984年和1994年提出经过一系列合理的假设，尺度空间唯一可行的核就是高斯函数。因此，被定义为一幅图像尺度空间函数的L(x,y,σ)是由尺度可变的高斯函数G(x,y,σ)和输入图像I(x,y)的卷积产生：

$$L(x,y,\sigma)=G(x,y,\sigma)*I(x,y).$$

其中*为x和y之间的卷积运算。而

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

为了在尺度空间中高效地探测出稳定的关键点位置，我们假设（Lowe，1999）使用尺度空间在高斯差分中的极值与图像卷积。可以计算得到两个相邻的由常数乘系数k分离的尺度的差值：

$$D(x,y,\sigma)=(G(x,y,k\sigma)-G(x,y,\sigma))*I(x,y)=L(x,y,k\sigma)-L(x,y,\sigma). \quad (1)$$

有很多选择这个函数的理由。首先，这是一个高效计算的函数，因为平滑图像L需要计算尺度空间特征描述的任何情况，而D只需计算简单的图像减法。

另外，Lindeberg于1994年研究表明，高斯差分函数可以提供高斯拉普拉斯的尺度规范化解得近似值$\sigma^2\nabla^2G$。Lindeberg展示了拉普拉斯在真实尺度不变性所要求的σ^2因素下的标准化。在更加精细的实验对比中，Mikolajczyk于2002年发现，与其他可能的图像函数如梯度法、Hessian法和Harris角函数相比，$\sigma^2\nabla^2G$的最大值和最小值产生了最稳定的图像特征。

D和$\sigma^2\nabla^2G$的关系可以从热扩散公式来理解（参数以σ而不是常见的$t=\sigma^2$形式）：

$$\frac{\partial G}{\partial\sigma}=\sigma\nabla^2G$$

由此，我们可以看出∇^2G可以用在两个相邻的尺度kσ,σ求解最终的差分近似为$\frac{\partial G}{\partial\sigma}$：

$$\sigma\nabla^2G=\frac{\partial G}{\partial\sigma}\approx\frac{G(x,y,k\sigma)-G(x,y,\sigma)}{k\sigma-\sigma}$$

因此，

$$G(x,y,k\sigma)-G(x,y,\sigma)\approx(k-1)\sigma^2\nabla^2G.$$

这表明，当高斯差分函数的尺度被常系数区分开后，它就包含了拉普拉斯尺度不变性要求的σ^2尺度标准化。等式中的系数（k-1）是所有尺度的常数，因此不影响极值的位置。K越接近1，估计误差就越趋向于0，但是实践中我们发现估值对极值探测的稳定性和即使是最显著的尺度差距的定位，如$k=\sqrt{2}$几乎没有影响。

构建D(x,y,σ)的有效方法如图1所示。初始图像与高斯函数递增地卷积形成图像，通过尺度空间的常系数k被分开，如左图的堆放的层。我们将尺度空间中的每个组（如σ的两倍）分为整数，间距为s，所以$k=2^{1/s}$。我们必须在每个组的堆中建立s+3幅模糊的图像才能完成覆盖全部组的极值探测。临近的图像尺度相减便产生了高斯差分图像，如右图所示。一旦完成了所有组的处理，我们就用σ代替初始值2σ（顶层的堆中会产生2幅图像）以每行每列的第二个像素对高斯图像进行重采样。相对于σ，采样的精度与第一个组没有差别，但计算量被很大程度上地降低了。

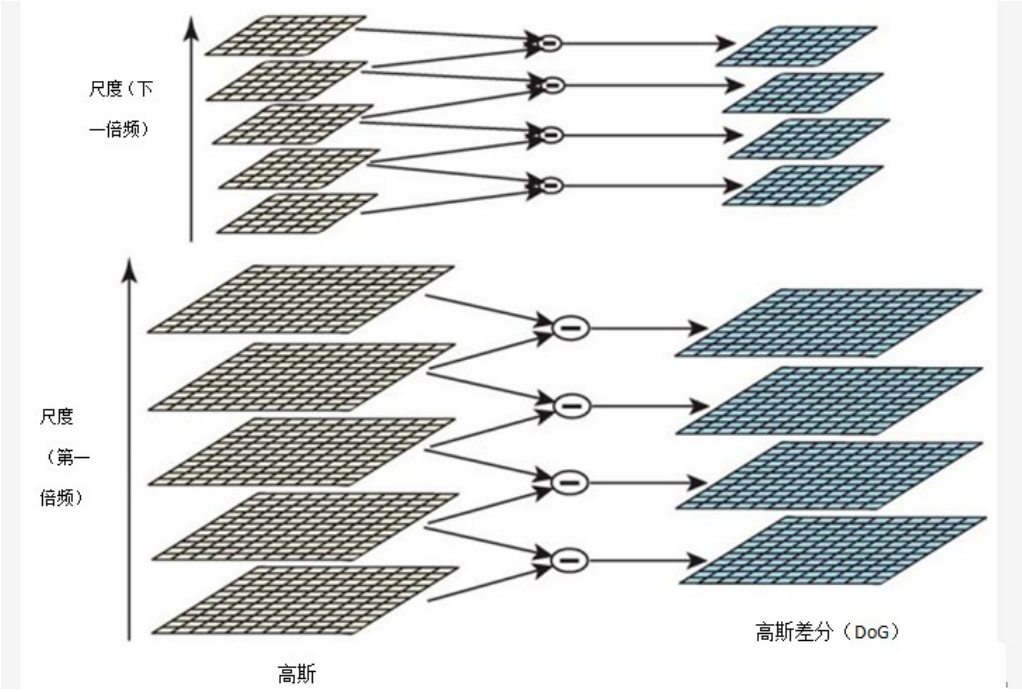

图1. 对于每个尺度空间的组，初始图像与高斯函数多次卷积所得尺度空间如图像左边所示。相邻的高斯图像相减产生了右边的差分高斯图像。每个组后，高斯图像被降采样2倍，重复该过程。

3.1 局部机制探测

为了探测到D(x,y,σ)的局部最大值和最小值，每个样本点都要和它当前图像的八个近邻已经上下尺度上的各九个近邻相比较（如图2）。只有在它比所有近邻大或者小时才会被选择。因为在前几次检查中大多数的样本点会被排除，因此，这个检查的代价相对较小。

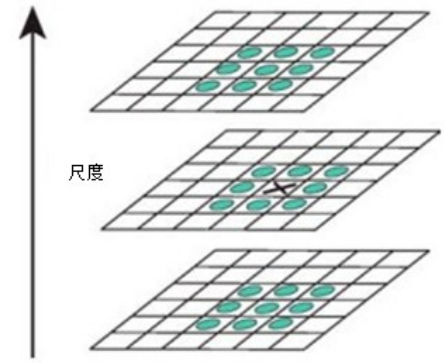

图2. 在现下的尺度和邻近度（记为圆），通过将一个像素（记为叉）与其临近的3*3区域内的26个像素进行对比，得到高斯差分图像的最大值和最小值。

确定图像和尺度空间中样本的频率非常重要，需要对极值进行可靠地探测。不幸的是，无法找到可以探测到所有极值的最小采样间隔，因为极值之间可以任意程度的接近，无规律可循。可以想象黑色背景上有一个白色的圆圈，在尺度空间的最大值处为圆形高斯差分函数区域的正中心，它与圆的大小和位置匹配。对于一个被拉长的椭圆形，它的每个端点（end）都有一个最大值。最大值的位置是一个图像的连续函数，对于那些中间被拉长的椭圆形将有从一个极值到两个极值的过渡，在过渡中，极值会任意的接近彼此。

因此，我们必须使用一个权衡效率和完整性的方案。实际上，正如我们所想，也被我们的实验所证实，相邻近的极值对图像很小的摄动是很不稳定的。我们可以通过对很大范围内采样频率的研究和使用那些在匹配任务的逼真模拟中提供了最可靠结果（的数据）来决定最好的选择。

3.2 尺度采样的频率

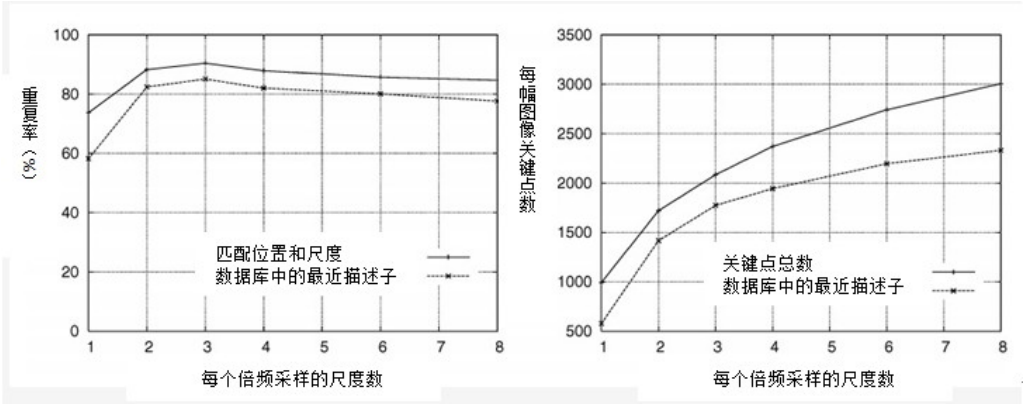

图3. 第一幅图上一条线是关键点在一幅被转换的图像的同一位置和尺度被重复探测的百分率，作为一个组采样的尺度数的值。下面的那条线是关键点的描述子与大型数据库正确匹配的百分率。第二幅图表示的是在一幅典型图像中被探测到的关键点总数，以此作为尺度采样的数值。

如图3和图4所示为采样频率所决定的取极大值的稳定性实验。这些图（以及本文中的大多数模拟）是基于一组32幅不同范围的真实图像的匹配工作，图像包括外景、人脸、航空影像和工业图像（经研究发现图像域对结果无任何影响）。每幅图像都经过了一系列的变换，包括旋转、缩放、仿射拉伸、明亮度对比度变化和增加图像噪声。改变是综合的，这样才有可能精确地推断初始图像的每个特征在转换后的图像中如何呈现，从而可以对每个特征测量正确的重复率和位置的准确性。

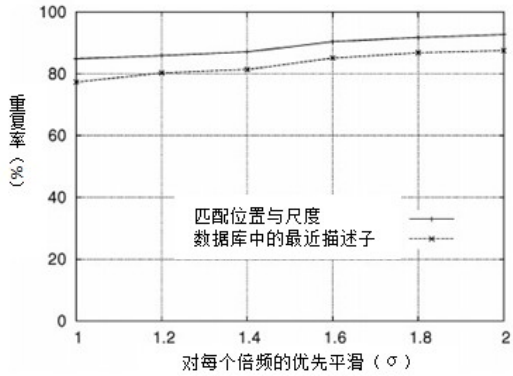

图4. 图中顶部的那条线表现的是关键点位置在转换的图像中被重复探测到的百分率，被作为对每个组第一级优先图像平滑的函数。

图3所示为用于检查在极值探测前采样的图像函数每个组尺度数变化的效果。在这个情况下，每幅图像都经过随机角旋转和初始图像0.2-0.9倍的随机缩放，然后被重采样。降分辨率图像的关键点与初始图像进行匹配，因此所有关键点的尺度将被呈现在匹配图像中。另外，添加了1%的图像噪声，也就是说在[0,1]范围内变化的像素值将被随机添加一个在[-0.01,0.01]内等间距变化的随机数字（相当于降低图像像素6比特的准确度）。

图3中的首行为关键点在一幅转换后图像进行匹配，位置和尺度被探测到的百分率。本文中所有的例子，我们都将匹配尺度定义为正确尺度的$\sqrt{2}$倍，匹配位置为σ个像素以内，σ为关键点的尺度（定义为高斯差分函数中使用的最小高斯函数的标准差）。图中下面的一行为使用最近邻匹配法，与有40,000个关键点的大型数据库正确匹配的关键点数，详细过程在第六章讲述（这表明关键点被重复定位对识别和匹配工作非常有利）。这幅图表示当每个组采样3个尺度时，重复率最高，这也是本文中其它实验所使用的采样尺度数。

采样的尺度是不是越高重复率就越好，这一点可能让人觉得有些奇怪。这是因为在很多被探测到的局部极值结果中，这些（尺度高的）结果稳定性较差，因此在转换图像中被探测到的几率也就降低了。可以由图3中的第二幅图看出，关键点被探测出来的平均数以及每幅图像中正确的匹配数。关键点数随采样尺度增加而提高。由于目标识别成功与否更多的是依赖于关键点正确匹配的数量，而不是它们匹配的正确率，因此对于很多应用而言，选择较大的尺度采样才是最佳选择。然而，计算成本也会随之增大，因此本文中的实验我们选择使用每个组3个采样尺度。

总而言之，这些实验表明高斯差分函数的尺度空间有很多的极值，但是完全的探测到它们成本很高。幸运的是，我们只使用一些较大的采样尺度就可以探测到很多有用而稳定的子集。

3.3 空间域采样的频率

我们刚决定尺度空间每组的采样频率，接下来要确定与平滑尺度相关的图像域中的采样频率。极值可能任意程度上的接近彼此，这里有一个类似的堆采样频率和探测率的权衡。图4所示为优先平滑函数的决策实验，应用于建立每个组的尺度空间代表前。同样，图中顶部的那条线表示关键点探测的重复率，结果显示重复率随σ的增大而增大。然而，使用大的σ对效率有所影响，所以我们选用σ=1.6来实现近似最佳的结果。这个值在本文中（包括图3中的结果）被普遍应用。

当然，如果我们在极值探测前对图像进行预平滑处理，我们就有效地剔除了最高的空间频率。这样，要充分利用输入，相比初始图像，图像可以被扩展来获取更多的采样点。在建立金字塔第一层之前，我们使用线性插值使输入图像的大小加倍。对原始图像使用亚像素补偿滤波可以有效的等价运算，但图像加倍的实现更加有效。我们假设原始图像有至少σ=0.5的模糊（防止显著混淆现象的最小值），因此相对新的像素空间，加倍的图像有σ=1.0。这意味着在创建第一组的尺度空间前，增加小量的平滑是必要的。图像加倍使稳定的关键点数增加了近4倍，但使用更大的扩展系数没有更明显的提高。

4. 准确的关键点定位

完成了像素与其近邻的比较就可以得到关键点的候选值，下一步就是完成附近数据位置、尺度和主曲率的精细配置（fit）。这个信息使低对比度的点（对噪音敏感）或定位在边角的差点被淘汰。

这个方法的初步（Lowe，1999）简单地实现了关键点定位于中心样本点的位置和尺度。然而，Brown最近改进了此方法（Brown和Lowe，2002）。通过局部样本点的三维二次方程配置来决定最大值的插值位置。他的实验表明这一改进很大程度地提高了匹配和稳定性。他的方法对尺度空间方程D(x,y,σ)使用了泰勒级数展开（到二阶）变换，把样本点作为原点。

$$D(x) = D + \frac{\partial D^T}{\partial x} x + \frac{1}{2} x^T \frac{\partial^2 D}{\partial x^2} x$$

（2）

其中D和它的导数是样本点的估值，而 $\hat{x}$ 为这一点的补偿。通过对函数求关于x的偏导并设为零得到极值的位置：

$$\hat{x} = -\frac{\partial^2 D}{\partial x^2}^{-1} \frac{\partial D}{\partial x}.$$

如Brown建议的，Hessian法和对D求导都是对相邻样本点使用差分来求估计值的。可以在最小的成本下得到3*3的线性系统的结果。如果 $\hat{x}$ 在任何维度的补偿大于0.5，就意味着极值与另一个样本点更为接近。这时，样本点改变，并进行插值取代该点。最终的补偿值 $\hat{x}$ 加到样本点的位置上来获取极值位置的插值估计值。

极值处的函数值 $D(\hat{x})$ 对排除低对比度的不稳定极值非常有用。这个可以通过用公式（3）代替（2）得到。

$$D(\hat{x}) = D + \frac{1}{2} \frac{\partial D^T}{\partial x} \hat{x}$$

对于本文中的实验来说，所有极值的 $D(\hat{x})$ 小于0.03的都要被排除（如前假设，我们设图像像素值在[0,1]的范围内）。

图5所示为在自然图像上进行关键点选择的结果。为了防止太多的聚类，我们使用了一个233*189像素的低分辨率图像，关键点被变现为矢量形式，给出了每个关键点的位置、尺度和方向（方向的指定见下文）。图5（a）所示为原始图像，后面的图像对其进行了降对比度。图5（b）所示为高斯差分函数探测到的所有最大值和最小值。而（c）所示为除去 $D(\hat{x})$ 值小于0.03所剩的729个关键点，（d）部分将在后面的章节中介绍。

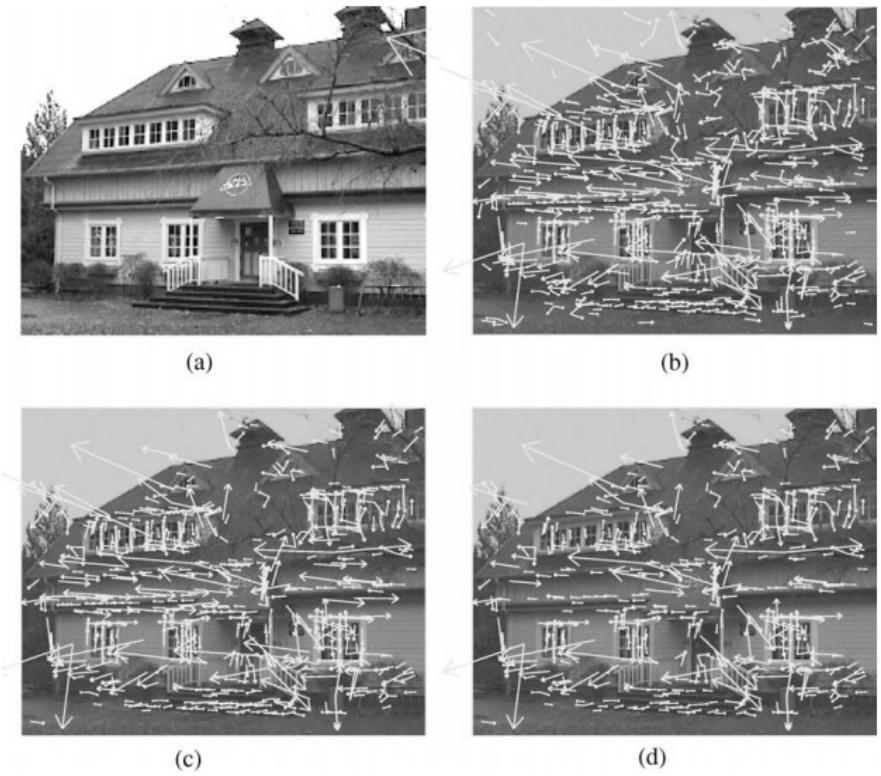

图5. 本图表示的是关键点选择的阶段。（a）233*189个像素的初始图像。（b）高斯差分公式最大值和最小值确定出的832个关键点位置，关键点被显示为矢量形式，表示尺度，方向和位置。（c）对最小值对比设置临界值后，还剩下729个关键点。（d）附加一个主曲率极限后，最终剩下的536个关键点。

4.1 排除角反射

对稳定性而言，只去除低对比度的关键点是不够的。即便在边缘处的点具有很差的决策性并且对很小的噪声很不稳定，高斯差分函数也会有很强的反应。高斯差分函数中一个定义不好的峰值将会对边缘处产生很大的主曲率，而在垂直方向上产生很小的主曲率。主曲率可以通过一个2*2的Hessian矩阵来计算。H在关键点的位置和尺度上。

$$H = \begin{bmatrix} Dxx & Dxy \\ Dxy & Dyy \end{bmatrix}$$

（4）

通过对近邻样本点的差分来估计导数值。

H的特征值与D的主曲率成比例。借用Harris和Stephens（1988）的方法，我们可以明确地避免特征值的计算，而只关心它们的比值。设α为最大量级的特征值，而β为最小量级的。然后我们可以通过求H的迹来获得特征值的和，从行列式获得它们的积：

$$Tr(H) = D_{xx} + D_{yy} = \alpha + \beta,$$

$$Det(H) = D_{xx}D_{yy} - (D_{xy})^2 = \alpha\beta$$

行列式不可能为负，曲率符号不同时，点不为极值，舍去。设r为最大量级特征值和最小特征值之比。所以α=rβ。接下来，

$$\frac{Tr(H)^2}{Det(H)} = \frac{(\alpha+\beta)^2}{\alpha\beta} = \frac{(\gamma\beta+\beta)^2}{\gamma\beta^2} = \frac{(\gamma+1)^2}{\gamma},$$

只取决于特征值的比而不是它们各自的值。当两个特征值相等时，（r+1）^{2}/r最小，并随r的增加而增加。因此，要看主曲率是否低于某个极限，我们只需要检查：

$$\frac{Tr(H)^2}{Det(H)} < \frac{(\gamma+1)^2}{\gamma}.$$

这个计算非常高效，当对每个关键点只需进行小于20次的浮点运算检测。本文中的实验使用的r值为10，这意味着认为关键点在主曲率间的比值大于10。图5中（c）到（d）的转换即为这个运算的结果。

5. 定向任务

通过局部图像属性给每个关键点指定一个的方向，关键点描述子可以与这个方向相关，从而实现图像旋转的不变性。这个方法和Schmid和Mohr（1997）的方法相比，他们的每个图像属性都是基于一个旋转不变的测量。他们方法的缺点就是它限制了可用的描述子，并因为没有要求所有测量都基于一个一致的旋转而丢失了图像信息。

下面的实验使用了很多方法来指定局部方向，下面的方法为找到最多稳定结果的。关键点的尺度是用来选择尺度最接近的高斯平滑图像L的，这样所有的计算都是在一个尺度不变条件下进行的。对于每个图像样本L(x，y)，在这个尺度下，梯度量级m（x，y）和方向θ（x，y）是用像素差预计算出来的：

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

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

一个方向直方图是用关键点周围区域的样本点的梯度方向组成的。方向直方图有36个柱子，覆盖360度方位角。每个加到直方图的样本都被它的梯度量级定权，再被该处关键点尺度1.5倍的σ的高斯圆形窗口定权。

方向直方图的峰值与局部梯度的主方向相对应。直方图中的最高峰值首先被探测到，然后是占最高峰80%以上的局部峰值也会在该方向建立一个关键点。因此，对于有多个相似量级的位置，可以在该位置和尺度创建多个不同向的关键点。只有15%的点会被指定多个方向，但它们对匹配的稳定性意义很大。最后，得到一个与3个直方图值配准（fit）的过每个峰值最接近的更准确峰位的插值抛物线。

图6所示为在不同数量的图像噪声下，位置、尺度和方向指定的实验稳定性。如前，图像被随机地旋转缩放过。顶端的线为关键点位置和尺度指定的稳定性。第二条线为当方向指定在15度以内的匹配稳定性。上面两条线之间的差距可以看出，即使加了10%的像素噪声。方向指定保留了时间95%的准确性（相当于相机有小于3比特的准确度）。正确匹配的方向量测变化为25度左右，当有10%的噪声时，升为3.9度。图6最下面一条线为一个关键点描述子与一个有40,000个关键点的数据库匹配正确的最终准确率（下文讨论）。如图所示，SIFT特征对大量的像素噪声具有抵抗性，而错误的主要原因在初始位置和尺度的探测。

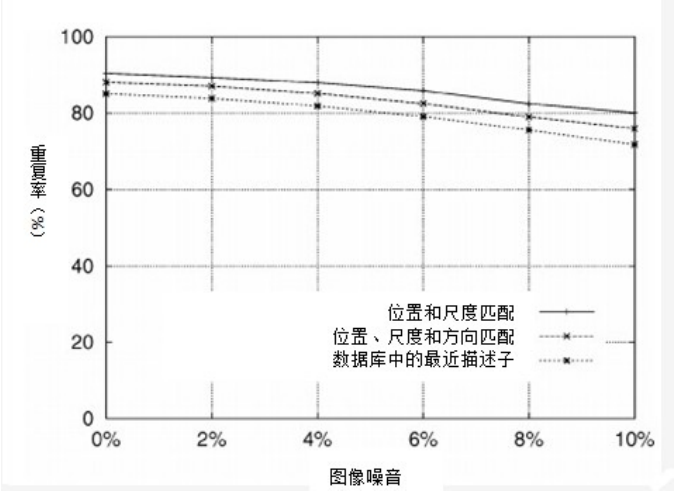

图6. 图中顶行显示的百分率为像素噪声的函数，是可重复检测的关键点的位置和尺度。第二行显示的是之后的重复率，同样要求方向一致。最下一行为最后与大型数据库正确匹配的描述子的百分率。

6. 局部图像描述子

之前的操作已经指定了图像的位置，尺度和每个关键点的方向。这些参数定义（impose）了一个可重复的局部二维坐标系统，可以在里面定义局部图像区域，从而为这些参数提供不变式。下一步就是为局部图像区域计算描述子，它要非常鲜明，同时对于剩下的变化尽可能的保持不变性，如光照或三维视点的改变。

一个明显的方法就是在合适的尺度的关键点周围的局部图像亮度进行采样，使用归一化的相关方法进行匹配。然而，简单的图像块的相关性对变化非常敏感，，从而导致样本的误匹配，如仿射变化或三维视点变化或非刚性变形。Edelman等人在1997年提出了一个更好的方法。他们提出的方法是基于生物视觉的，尤其是主视觉皮层中复杂的神经细胞。这些复杂的神经细胞对某个方向和空间频率的梯度变化反应，但梯度在视网膜上的位置却是在一个可以接受的范围（field）内变化而不是精确地固定。Edelman等人假设这些复杂神经细胞的函数使得我们进行匹配和一定视点范围内三维目标的识别。他们展示了详细的实验，通过三维计算机目标和动物形状的模型表明在允许位置变化下的匹配梯度比在三维旋转下的分类结果要好得多。（They have performed de-tailed experiments using 3D computer models of

object and animal shapes which show that matching gradients while allowing for shifts in their position results in much better classification under 3D rotation.）比如说，在使用复杂的细胞模型后，三维目标在20度景深下旋转的识别准确率从35%的梯度相关性升为94%。我们的下面的实践正是受这个思想的启发，但使用的是另一种计算机制来允许位置变化。

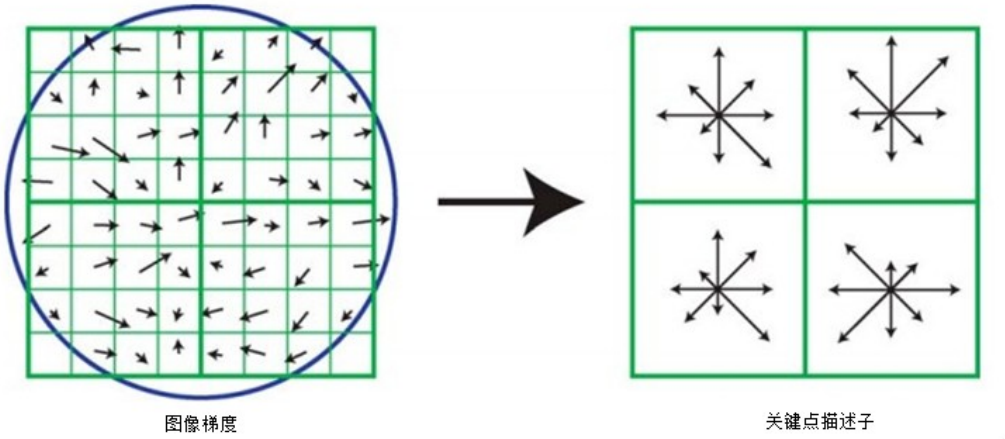

图7. 要创建一个关键点描述子首先要计算关键点位置附近区域的每个图像样本点的梯度大小和方向，如左图所示。它们由高斯窗口定权，由重叠的圆形表示。然后如右图所示，将这些样本聚集为方向直方图，每4*4个子区域概括为一个。这里的每个箭头的长度均为该区域内这个方向附近梯度模值的总和相对应。图中所示的2*2的描述子箭头就是由8*8的样本集计算出来的，本文所使用的是由16*16的样本集中计算出来的4*4的描述子。

6.1 描述子表达

图7表明关键点描述子的计算。首先，图像的梯度量级和方向是在关键点周围的采样，使用关键点的尺度来选择图像高斯模糊的程度。为了达到方向不变性，我们描述子的坐标和梯度方向都是随关键点方向进行旋转的。为了高效性，如第五章中所提到的，金字塔中所有等级的梯度都被预计算出来了。在图7的左边，它们在每个样本位置处以小箭头标出。

σ为描述子窗口宽度1.5倍的高斯定权公式用来指定每个样本点的权。这个如图7左边的圆形窗口所示，使得权重可以平滑的减弱。高斯窗口的目的是为了防止描述子在窗口位置发生微小变化下的突变，给远离描述子中心的梯度更少的关注，这些梯度对误匹配影响最大。

关键点描述子如图7右侧所示。它通过在4*4的样本区域建立方向直方图使得梯度位置可以发生较大的变化。每个方向直方图有八个方向，每个箭头的长度与该直方图输入的量级有关。一个左边的梯度样本可以变为四个样本位置，并向右边的直方图输出值，从而实现了更大的局部位置变化的目的。

当描述子在从一个直方图到另一个直方图或从一个方向平滑地变向另一个方向时发生突变，防止所有的边缘影响很重要。因此，三线性插值用来给每个梯度样本向邻近的箱（柱子）内分配值。换句话说，就是每个箱中的输入都是乘过了1-d各个方向的权值的，其中d为以直方图各柱子之间的空间为单位测量的样本到中心柱子的距离值。

描述子由保存所有方向直方图的值得矢量得到，对应于图7右边图中箭头的长度。图像显示了一个2*2阵列的方向直方图，而我们下面的实验表明每个方框里有八个方向的4*4阵列的直方图所得结果最优。因此，本文所用的为每个关键点有4*4*8=128个元素特征矢量的。

最终，为了减弱光照变化的影响，特征矢量被修改。首先，矢量被标准化为单位长度。对图像对比度的改变就是讲每个像素值乘以一个常数，这样整个梯度也会乘上同一个常数，这种对比度变化会被矢量归一化抵消掉。亮度变化中图像里的每个像素都会加一个常数，这不会影响到梯度值，因为梯度值是像素值之差。因此，描述子对于光照的仿射变化是具有不变性的。然而，非线性光照变化也可能是由于相机饱和度或光照变化影响了不同数量不同方向的三维表面。这些影响可能会造成一些梯度相关量级的巨大变化，但对梯度方向影响很小。因此，我们减少将每个单位特征矢量不大于0.2的这个限定对大的梯度量级的影响，然后对单位长度进行重归一化。这意味着匹配大梯度量级不再是一件重要的事，而更加强调方向的分布。值0.2是通过图像对相同的三维目标保留不同光照的实验得到的。

6.2 描述子测试

有两个参数可被用为变化描述子的复杂度：在直方图中的方向数r和n*n方位直方图阵列的宽n。最终描述子矢量的大小为rn^2。当描述子的复杂度增加时，在大型数据库中的区分度更好，但它对形状畸变和闭塞也更为敏感。

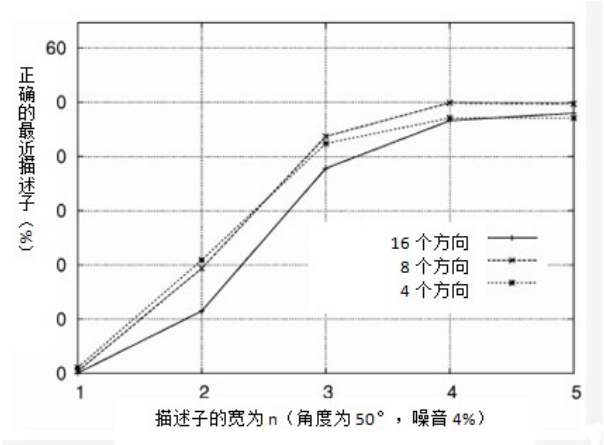

图8. 本图描述的为宽为n*n的关键点描述子以及每个直方图方向数的函数，是关键点与数据库中40,000个关键点正确匹配的百分率。

图8所示的实验结果，其中方向数和描述子大小为变化的。图是用一个视点变换得到的，其中平面相比观察者倾斜了50度，并添加了4%的噪声。这接近了可靠匹配的极限，也是在这种更加困难的情况下，描述子的表现才更为重要。结果为在40，000个关键点的数据库在找到正确匹配的最近邻单点的关键点百分数。图中显示所得，单个方向的直方图（n=1）的区分度很差，但增加直至一个4*4阵列的八方向直方图的过程中结果一直在改善。这之后，再增加方向或加大描述子只对匹配造成了影响，使得描述子对畸变更加的敏感。在其他视角角度变化和噪声情况下，结果是相似的。尽管在一些简单的情况下，区分度（从最高级）继续提高直至5*5和更高的描述子大小。但我们在本文中仍使用4*4的8方向描述子，可产生128维的特征矢量。尽管描述子的维数好像很高，但我们发现这在一系列匹配任务中比低维度表

现更好，而且匹配的计算成本在使用如下介绍的近似的最近邻方法中也很低。

6.3 仿射变化敏感度

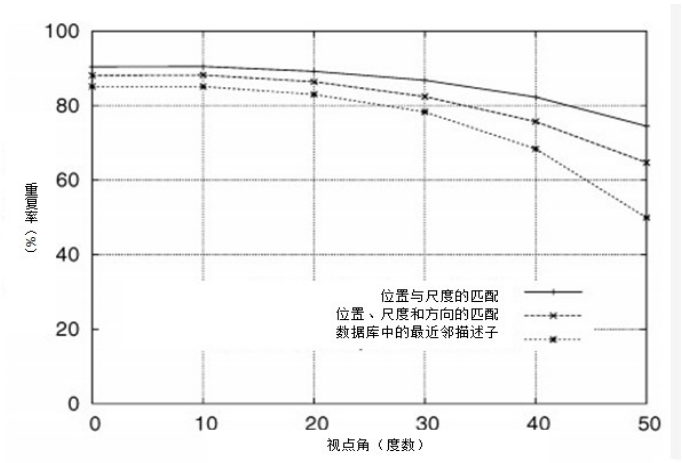

图9. 本图所示为关键点位置、方向和最终与数据库匹配的探测的稳定度，作为仿射变换的一个函数。仿射变换的程度由平面内一组等视点景深旋转来表示。

描述子对仿射变化的敏感度如图9所示。图中所示为关键点位置和尺度选择、方向分配和与一个与远离观察者的平面中进行深度旋转函数的数据库进行最近邻匹配的可靠性。可以看出，每个计算阶段中，随着仿射畸变的增加，重复率的下降，但在最后，对于一个视点变化50度时，匹配的准确度仍是保持在50%之上。

为了实现大视点角情况下可靠的匹配，如第二章所讲，可以使用一种仿射不变的探测器来对图像区域进行选择和重采样。那里提到，由于这些方法都是起源于一个非仿射不变的图像的初始特征位置，所以它们并不具有真正的仿射不变性。在看起来最具有仿射不变性的方法中，Mikolajczyk（2002）对Harris仿射探测器假设并执行了详细的实验。他发现，它的关键点重复率比这里给出的50度的视点角要低，但在角度为70度时，保持在接近40%的重复率上，在极值仿射变换中表现更好。缺点是计算成本高，关键点数量少和在噪声下设定一致仿射变换框架误差对小的仿射变换稳定性差。实际上，三维目标允许的范围是远少于对平面的，所以仿射不变性在匹配视点变化时并不是限制因素。如果要求大范围的仿射不变性，如要求表面为平面，那么一个简单的解决方案就是去采用Pritchard和Heidrich（2003）的方法，生成由训练图像的4仿射变换的版本到60度视点的变化的附加SIFT特征。这使得标准SIFT特征的使用在图像识别处理中没有增加新的运算成本，但在因素为3的特征数据库的大小增加了。

6.4 与大型数据库匹配

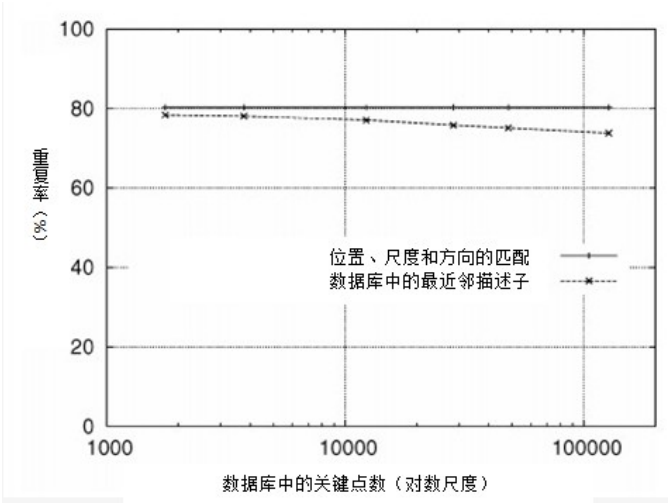

图10. 虚线表明关键点与数据库正确匹配的百分率，为数据库大小的一个函数（使用对数尺度）。实线为关键点分配正确的位置、尺度和方向。图像有随机尺度和旋转变换，30°的仿射变换和2%的图像噪声被预先加入匹配中。

一个测量特征鲜明性的重要遗留问题是匹配重复率如何作为一个匹配数据库中的特征数的函数变化。本文中大多数的例子是使用一个32幅图像，40,000个关键点的数据库而生成的。图10所示匹配重复率如何作为一个数据库大小的函数变化的。这幅图像是使用一个更大的有112幅图像的数据库生成的，视点旋转深度为30度，添加了2%的噪声，图像进行了随机地尺度变化和旋转。

虚线表明数据库中以最近邻为正确匹配的图像特征的部分，它作为数据库大小的函数以对数的形式显示出来。最左端的点是与由一张图像中得到的特征进行匹配而最右端的点是从含有112幅图像的数据库的所有特征中选择的匹配。可以看出匹配的可靠性在以干扰项数目为函数时降低了，而所有的显示表明在更大的数据库大小下更多的正确匹配将继续被找到。

实线为关键点在转换图像的正确匹配的位置和方向被识别的百分率，所以只有这些点在数据库中有机会有匹配的描述子。这条线平缓（flat）的原因是测试在整个数据库中运行了每个值，但只改变了数据库中一部分用来干扰的部分。有趣的是，两条线之间的间隔很小，表明匹配失败更多的是因为初始特征定位和方向分配的问题，而不是特征鲜明性的问题，而不是大型数据库大小的问题。

7. 目标识别的应用

如上所示，本文主要讨论的是鲜明不变性关键点的派生。为了展示它们的应用，我们给出它们在目标遮挡和聚类情况下进行识别的应用。更多关于这些特征的识别应用参见其它文献（Lowe，1999；Se等人，2002）。

目标识别首先要将每个关键点独立的与从训练图像中提取的关键点进行匹配。由于模糊的特征和从背景聚类中得到的特征，很多这些最初的匹配是不正确的。因此，首先识别那些与一个目标或其姿态一致的至少有三个特征的聚类，因为他们比那些独立特征有更高的可能被正确匹配。接下来，通过履行一个与模型合适的精细几何来检查每个聚类，并判断结果，决定采纳还是放弃解译。

7.1关键点匹配

通过在由训练图像得到的关键点数据库中识别最近邻，我们找到了每个关键点的最佳候选匹配。如第六章所述，最近邻定义为每个关键点的不变描述子矢量之间的最短欧氏距离。

然而，图像中的很多特征与训练数据库可能没有任何正确的匹配，因为它们是从背景聚类中提出的或没有在训练图像中被探测到。因此，有一种方法来丢弃与数据库没有很好地匹配的特征很有用。对最近距离特征的全局限值执行的并不尽如人意，因为一些描述子比其他的要鲜明很多。更有效的方法是使用最近距离与次近距离的比值。如果有同一目标的很多训练图像时，我们定义与第一个来自不同目标的次近距离为最近距离，就像使用含有不同目标的已知图像一样。这个方法执行很好，因为正确匹配需要最近邻显著地接近那些最接近的错误匹配来达到可靠性匹配。对于错误的匹配，由于特征空间的高维度，相似距离内会有很多其他的错误匹配。我们可以把次近距离匹配作为对特征空间的这一部分错误匹配密度的一个估计并同时识别特征不明确的特殊实例。

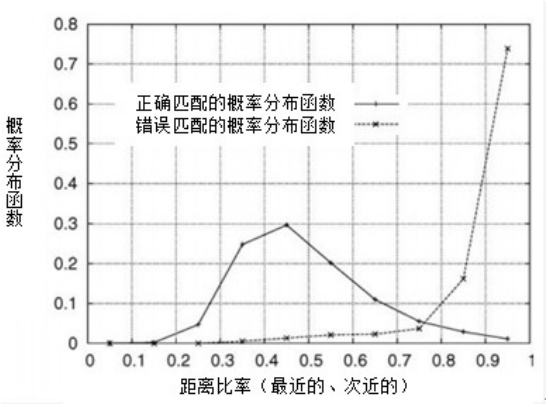

图11. 匹配的正确性可由距离比率决定，即用最近邻距离除以次近邻的距离得到。使用一个有40,000个关键点的数据库，实线显示的为正确匹配距离比率的概率密度函数，而点线为错误的概率密度函数。

图11表明用这种方法对真实图像数据的测量值。正确和不正确匹配的可能性密度函数以每个关键点最近邻与次近邻之比表示。最近邻为正确匹配的概率密度函数的中心比错误匹配的比率低很多。对于我们的目标识别的执行，我们丢弃那些距离比值大于0.8的匹配，这样可以在仅仅丢弃5%的正确匹配的前提下，减少90%的错误匹配。图像是在一个40,000关键点的数据库中，以随机尺度和方向变化下匹配图像生成的，旋转深度为30度，添加了2%的噪声。

7.2 有效的最近邻函数

没有一个现行算法可以在识别高维空间点的准确最近距离时优于穷搜索法（exhaustive search）。我们的关键点描述子有128维的特征矢量，而最好的算法，如k-d树（Friedman等人，1977年）也无法在多于10维的空间中提供比穷搜索法更快速的性能。因此，我们用了一种近似的算法，叫做最优盒优先（BBF）算法（Beis和Lowe，1997）。这是在场景中近似的返回具有最高可能性的最近邻。

BBF算法使用了一种k-d树算法的改进算法，使得特征空间中的箱是以它们在队列位置中最的近距离的顺序被检索。这种优先顺序最早是由Arya和Mount（1993）实践的，他们在Arya等人（1998）的文献中对其计算性能提供了更进一步的研究。这个搜索顺序要求使用一种基于堆优先的队列来实现搜索顺序的高效决策。在检索了特定量的最近箱之后，中断进一步的搜索可以低计算成本的返回一个近似结果。在我们的实现中，我们在检查了最开始的200个最近邻候选值后，中断了搜索。对于一个有100,000个关键点的数据库，它比准确的最近邻搜索加速了约两个数量级，而正确匹配的结果只损失了不到5%。BBF算法运行如此良好的一个原因是我们只考虑了最近邻比次近邻小于0.8以内的这些匹配（如前面章节中提到的），因此无需考虑那些很多近邻距离非常接近的困难情况。

7.3 霍夫变换的聚类

对小而高度遮挡的目标识别实现最大化，我们希望以最少的可能的特征匹配数进行目标识别。我们发现在最少使用3个特征的情况下可靠识别是可能的。一个有着2，000个或更多特征的典型图像可能会有很多不同的目标和聚类背景。而第七章中所述的距离比率实验允许我们从聚类背景中丢弃大量的错误匹配，而这并不减少其他有效目标的匹配。通常，我们仍需要从含有99%异常值的匹配中找到那少于1%的正常值识别正确的子集。很多众所周知的稳健地配置（fitting）方法，如RANSAC或最小平方中值，在正常值小于50%时运算结果就会很差。幸运的是，在姿态（pose）空间使用Hough转换（Hough，1962；Ballard，1981；Grimson，1990）的聚类特征可以很好的表现。

霍夫转换通过每个特征与所有目标中特征一致的姿态进行投票通过（vote for）的一致性解译来识别聚类。当发现特征聚类与一个目标投票通过了同一姿态，这种解译正确的可能性比任何单一特征要高很多。我们的每个关键点都有4个参数：二维位置，尺度和方向，而数据库中每个匹配的关键点都有这个关键点与找到的训练图像相关联的记录。因此，我们可以建立一个霍夫变换，由匹配假设输入预计的模型位置，方向和尺度。这个预计有很大的误差界限，因为这四个参数的相似变换只是一个三维目标全六自由度的姿态空间的估计值，并且没有做任何非刚性的变换。因此，我们30度方向的宽箱大小，2因素的尺度以及投影训练图像位置维数（使用预计的尺度）最大值的0.25倍。为了防止边界效应在箱指定中的问题，每个关键点匹配在每个维度中都指定（vote for）了两个最近的箱，这个假设共有16个输入，姿态范围扩宽更多。

在多数霍夫转换的实现中，用多维阵列来表现箱。然而，很多潜在的箱保持为空，由于它们共有的依赖性，很难计算箱值可能的范围（比如说，选择范围上可能的位置离散值的依赖性）。这些问题可通过使用箱值的伪随机散列函数向一维散列表中插入投票（votes），从而可以简单的探测到冲突。

7.4 仿射参数的解决方法

霍夫转换是用来识别箱中至少有三个实体的所有聚类。每一个这样的聚类都要进入一个用最小二乘法来计算与训练图像向新图像转换有关的最佳的仿射投影参数的几何验证程序。

在正射投影下，仿射变换可以正确求解（account for）一个平面的三维旋转，但对于非平面的目标的三维旋转估值就很差了。更普遍的方法是解基础矩阵（Luong和Faugeras，1996；Hartley和Zisserman，2000）。然而，与仿射法只需要3个点匹配相比，一个基础矩阵式要求至少7个，而实际中，为了更好的稳定性，需要更多的匹配。我们希望只用三个特征匹配就完成识别，因此仿射变换就提供了一个很好的起始点，我们可以通过将允许的残差值增大来计算（account for）仿射估计中的误差。想象在目标周围放了一个球形，然后将球形旋转30度，球内的任意点不会移动超过球形投影直径的0.25倍。对于本文中的一个典型三维目标的例子，在我们允许残差不大于目标投影维数的最大值的0.25倍时，仿射方法可以很好地解决问题。Brown和Lowe（2002）提出了一种更普遍的方法，初值由相似变换得到，然后计算已经找到足够匹配数的基础矩阵。

模型点$[x,y]^T$对于图像$[u,v]^T$的仿射变换可以被写为：

$$\begin{bmatrix} u \\ v \end{bmatrix} = \begin{bmatrix} m_1 & m_2 \\ m_3 & m_4 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} + \begin{bmatrix} t_x \\ t_y \end{bmatrix}$$

其中，$\begin{bmatrix} t_x \\ t_y \end{bmatrix}$为模型变换，而$m_i$参数表示仿射旋转、缩放和拉伸。

我们希望解出变换参数，因此，上式可以被重写为将未知量变为列向量的形式：

$$\begin{bmatrix} x & y & 0 & 0 & 1 & 0 \\ 0 & 0 & x & y & 0 & 1 \\ & & \cdots & & & \\ & & \cdots & & & \end{bmatrix} \begin{bmatrix} m_1 \\ m_2 \\ m_3 \\ m_4 \\ t_x \\ t_y \end{bmatrix} = \begin{bmatrix} u \\ v \\ . \\ . \end{bmatrix}$$

这个等式表示的是一个单独的匹配，但可以添加后续匹配的数值，每个匹配都要在第一个和最后一个矩阵中加两行。要解这个式子，至少需要三对匹配。

我们可以把这个线性系统写为：

$$Ax = b$$

参数x的最小二乘法可以通过解对应的法方程得到，

$$x = \left[A^T A \right]^{-1} A^T b$$

它为投影模型的位置到图像位置的距离平方和的最小值。这个最小二乘法可以扩展来解决铰接的灵活的目标的三维姿态和内部参数（Lowe，1991）。

通过检查每个图像特征和模型之间的一致度就可以移除异常值。有了更加准确的最小二乘法，我们要求每个匹配要在霍夫转换箱（bin）中的参数的误差一半的范围以内。如果排除异常值后少于三个点，则这次匹配失败。当排除异常值后，要用最小二乘法再次计算留下的点，依次迭代下去。另外，自上而下的进行相位（phase）是为了增加与投影模型位置一致的匹配。可能会在霍夫转换箱时丢失的一些匹配是由于转换的相似性或其它错误。

最后是否接受这个假设取决于之前文章中提到过的精细概率模型（Lowe，2001）。这个方法首先用来在给出模型的投影大小，区域中的特征数和配置（fit）的准确度的情况下，解决模型姿态的错误匹配期望数。贝叶斯分析给出了目标基于实际找到的匹配特征被表现出来的可能性。如果最终正确解译出的可能性大于0.98，我们就接收这个模型。对于投影到图像很小的区域的情况，3个特征就足够进行可靠地识别了。对于纹理较多的大面积目标，错误匹配的期望值较高，可能会需要是个特征匹配这么多。

8. 识别实例

图12所示为一个从包含三维目标的聚类遮挡图像中进行目标识别的例子。如左图所示，测试图像为一个玩具火车和青蛙。中图（大小为600*480像素）为含有被遮蔽的测试目标，并有大量聚类背景的图片。因此，即使是人眼也很难立即发现。右图所示为最后正确识别后叠加在一个降对比度版本下的图像。用于识别的关键点显示为小方形，有一条线来指示方向。方形的大小与用于构建描述子的图像区域对应。在每个识别目标的外围有一个外包的平行四边形，它的边界是测试图像在识别过程中在仿射变换下的投影。

另一个潜在的方法应用是位置识别，这样运动的车辆和设备就可以通过识别熟悉的位置来确定位置了。图13给出了这个应用的一个例子，其中训练图像是从场景中很多地方拍摄的。如左上图所示，这些目标看起来都不太显眼，如木墙或是垃圾桶旁的树。而右上方的测试图像（大小为640*315像素）是将初始图像场景旋转30度后的视点拍摄的，但是，训练图像还是很容易的被找到了。

识别的全部过程都被高效完成，所以图12和图13的所有目标在一台2GHz的奔腾4处理器上的识别总用时少于0.3秒。我们在一台笔记本上用一台摄影照相机实施该算法，并在多种环境下广泛的测试了它们。一般来说，有纹理的平面在被旋转50度以内，在提供做够光亮的或不是过度强烈的几乎任何光照条件下，都可以被可靠地识别。对于三维目标，可靠识别的任意向深度旋转范围只有30度，而光照变化的干扰性也更明显。因此，三维目标的识别最好是从多视角的综合特征入手，如局部特征视角聚类法（Lowe，2002）。

这些关键点还被用于解决机器人定位和制图的问题，这个在其他文章中有详细的阐述（Se等人，2001）。在这个应用中，一个三目的立体系统被用来决策关键点定位的三维估计决策。关键点只有同时出现在三幅图像中，并有一致的不同性时才被使用，这样可以保证出现较少的异常值。机器人运动时，它通过与现有的三维地图进行特征匹配来确定自己的位置，然后在更新它们的三维位置时，使用卡尔曼滤波来递增地向地图添加特征。这为机器人在一个位置环境中定位提供了一种具有鲁棒性和准确性的解决方案。这项工作还处理了位置识别的问题，这样，机器人可以在一幅大型地图中被转换到可以识别自己的位置的状态（Se等人，2002），相当于目标识别的三维实现。

图12. 左图所示为两目标的测试图片。它们可以在大量遮挡的聚类图片中被识别出来，见中图。识别结果见右图。每个识别目标周围都画有一个平行四边形来显示初始测试图像的边界，识别过程中解决了仿射变换问题。小方形为用于进行识别的关键点。

图13. 这个例子表明在复杂场景中的位置识别。用于定位的测试图像为左上方640*315像素的图像，是从右上角图像的不同视角拍摄的。识别区域如下图所示，小的方形为识别的关键点，外围的平行四边形为仿射变换后初始测试图像的边界。

9. 结论

本文中所述的SIFT关键点在它的鲜明性方面尤为突出，可以是关键点与大型数据库中的其他关键点进行正确的匹配。这一鲜明性由装配在图像的局部区域内代表图像梯度高维的矢量来实现。关键点对图像旋转具有不变性，对大尺度的仿射变形具有鲁棒性。从典型图像中可以提取大量的关键点，从而使在混杂背景下提取小目标具有更好的鲁棒性。可以从整个尺度范围提取关键点意味着小的局部特征可以与小而高度遮挡的目标进行匹配，而大的关键点则在图像噪音和模糊时具有了更好的表现。它们的计算是高效的，在标配的PC机上，几千个关键点可以被近实时的从典型图像中提取出来。

本文还提出了一种用关键点进行目标识别的方法。这种方法使用了近似的近邻查找，用来识别与目标姿态一致的聚类的Hough变换和最小二乘法进行最后的决策和核查。另一个可能的应用是三维重建、运动跟踪和分割、机器人定位、图像全景集合（assembly）、对极（epipolar）配准和其他需要进行图像间匹配位置识别的视角匹配。

对于图像特征的不变性和鲜明性，未来的研究可以由有很多方向。全三维视点和光照变化数据点集需要进行系统的测试。本文所述特征只使用了单色亮度的图像，因此，进一步的鲜明性可以从光照不变的颜色描述子中得出（Funt and Finlayson, 1995; Brown and Lowe, 2002）。同样，局部纹理测量在人类视觉中也具有重要作用，合并在描述子中后，可以比当前这个从单个空间频率进行研究的描述子更具有普遍的形式。局部特征不变量匹配方法一个吸引人的地方在于这里无需挑选一个特征类型，因为最好的结果往往是使用很多不同特征得到的，因此，本方法可以贡献于获得有用的匹配并提高整体的鲁棒性。

另一个未来的研究方向是研究可以识别的目标分类的特征。这对类属目标尤为重要，分类必须包含所有可能的外形，这是一个巨大的范围。Weber等人的研究（2000）和Fergus等人的研究（2003）显示通过学习小型数据集的适合识别目标类属的局部特征，这种方法有实现的可能性。从长远角度来看，特征集应该包含优先的（prior）和博学的（learned）特征，这些特征将基于对大量目标分来有效的训练数据的数量来使用。

致谢

我要尤其感谢Matthew Brown，他对本文在内容和表述上给了我很多改进的建议，而他本人在特征定位和不变性上的工作也对本方法有贡献。另外，我想谢谢大家宝贵的建议，他们是Stephen Se，Jim Little，Krystian Mikolajczyk，Cordelia Schmid，Tony Lindeberg和Andrew Zisserman。这个研究是由加拿大国家科学工程研究协会（NSERC）、机器人学与智能系统协会（IRIS）和Excellence网络中心支持完成的。

参考文献

Arya, S. and Mount, D.M. 1993. Approximate nearest neighbor queries in fixed dimensions. In *Fourth Annual ACM-SIAM Symposium on Discrete Algorithms (SODA'93),* pp. 271–280.

Arya, S., Mount, D.M., Netanyahu, N.S., Silverman, R., and Wu, A.Y. 1998. An optimal algorithm for approximate nearest neighbor searching. *Journal of the ACM*, 45:891–923.

Ballard, D.H. 1981. Generalizing the Hough transform to detect arbitrary patterns. *Pattern Recognition*, 13(2):111–122.

Basri, R. and Jacobs, D.W. 1997. Recognition using region correspondences. *International Journal of Computer Vision*, 25(2):145–166.

Baumberg, A. 2000. Reliable feature matching across widely separated views. In *Conference on Computer Vision and Pattern Recognition*, Hilton

Head, South Carolina, pp. 774–781.

Beis, J. and Lowe, D.G. 1997. Shape indexing using approximate nearest-neighbour search in high-dimensional spaces. In *Conference on Computer Vision and Pattern Recognition* , Puerto Rico, pp. 1000–1006.

Brown, M. and Lowe, D.G. 2002. Invariant features from interest point groups. In *British Machine Vision Conference*, Cardiff, Wales, pp. 656–665.

Carneiro, G. and Jepson, A.D. 2002. Phase-based local features. In *European Conference on Computer Vision (ECCV)*, Copenhagen, Denmark, pp. 282–296.

Crowley, J.L. and Parker, A.C. 1984. A representation for shape based on peaks and ridges in the difference of low-pass transform. *IEEE Trans. on Pattern Analysis and Machine Intelligence* , 6(2):156– 170.

Edelman, S., Intrator, N., and Poggio, T. 1997. Complex cells and object recognition. Unpublished manuscript:

http://kybele.psych.cornell.edu/ ~edelman/archive.html

Fergus, R., Perona, P., and Zisserman, A. 2003. Object class recognition by unsupervised scale-invariant learning. In *IEEE Conference on Computer Vision and Pattern Recognition*, Madison, Wisconsin, pp. 264–271.

Friedman, J.H., Bentley, J.L., and Finkel, R.A. 1977. An algorithm for finding best matches in logarithmic expected time. *ACM Transactions on Mathematical Software*, 3(3):209–226.

Funt, B.V. and Finlayson, G.D. 1995. Color constant color indexing. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 17(5):522–529.

Grimson, E. 1990. *Object Recognition by Computer: The Role of Geometric Constraints*, The MIT Press: Cambridge, MA. Harris, C. 1992. Geometry from visual motion. In *Active Vision*, A. Blake and A. Yuille (Eds.), MIT Press, pp. 263–284.

Harris, C. and Stephens, M. 1988. A combined corner and edge detector. In *Fourth Alvey Vision Conference*, Manchester, UK, pp. 147– 151.

Hartley, R. and Zisserman, A. 2000.*Multiple view geometry in computer vision*, Cambridge University Press: Cambridge, UK.

Hough, P.V.C. 1962. Method and means for recognizing complex patterns. U.S. Patent 3069654.

Koenderink, J.J. 1984. The structure of images. *Biological Cybernetics*, 50:363–396.

Lindeberg, T. 1993. Detecting salient blob-like image structures and their scales with a scale-space primal sketch: A method for focus-of-attention. *International Journal of Computer Vision*, 11(3):283–318.

Lindeberg, T. 1994. Scale-space theory: A basic tool for analyzing structures at different scales. *Journal of Applied Statistics*, 21(2):224–270.

Lowe, D.G. 1991. Fitting parameterized three-dimensional models to images. *IEEE Trans. on Pattern Analysis and Machine Intelligence* , 13(5):441–450.

Lowe, D.G. 1999. Object recognition from local scale-invariant features. In *International Conference on Computer Vision* , Corfu, Greece, pp. 1150–1157.

Lowe, D.G. 2001. Local feature view clustering for 3D object recognition. *IEEE Conference on Computer Vision and Pattern Recognition,* Kauai, Hawaii, pp. 682–688.

Luong, Q.T. and Faugeras, O.D. 1996. The fundamental matrix: Theory, algorithms, and stability analysis. *International Journal of Computer Vision*, 17(1):43–76.

Matas, J., Chum, O., Urban, M., and Pajdla, T. 2002. Robust wide baseline stereo from maximally stable extremal regions. In *British Machine Vision Conference*, Cardiff, Wales, pp. 384–393.

Mikolajczyk, K. 2002. Detection of local features invariant to affine transformations, Ph.D. thesis, Institut National Polytechnique de Grenoble, France.

Mikolajczyk, K. and Schmid, C. 2002. An affine invariant interest point detector. In *European Conference on Computer Vision (ECCV)*, Copenhagen, Denmark, pp. 128–142.

Mikolajczyk, K., Zisserman, A., and Schmid, C. 2003. Shape recognition with edge-based features. In *Proceedings of the British Ma-chine Vision Conference*, Norwich, U.K.

Moravec, H. 1981. Rover visual obstacle avoidance. In *International Joint Conference on Artificial Intelligence* ,Vancouver, Canada, pp. 785–790.

Nelson, R.C. and Selinger, A. 1998. Large-scale tests of a keyed, appearance-based 3-D object recognition system. *Vision Research*, 38(15):2469–2488.

Pope, A.R. and Lowe, D.G. 2000. Probabilistic models of appearance for 3-D object recognition. International Journal of Computer Vision , 40(2):149–167.

Pritchard, D. and Heidrich, W. 2003. Cloth motion capture. *Computer Graphics Forum (Eurographics 2003)*, 22(3):263–271.

Schaffalitzky, F. and Zisserman, A. 2002. Multi-view matching for unordered image sets, or 'How do I organize my holiday snaps?"' In *European Conference on Computer Vision* , Copenhagen, Den-mark, pp. 414–431.

Schiele, B. and Crowley, J.L. 2000. Recognition without correspondence using multidimensional receptive field histograms. *International Journal of Computer Vision* , 36(1):31–50.

Schmid, C. and Mohr, R. 1997. Local gray value invariants for image retrieval. *IEEE Trans. on Pattern Analysis and Machine Intelligence* , 19(5):530–534.

Se, S., Lowe, D.G., and Little, J. 2001. Vision-based mobile robot localization and mapping using scale-invariant features. In *International Conference on Robotics and Automation* , Seoul, Korea, pp. 2051–2058.

Se, S., Lowe, D.G., and Little, J. 2002. Global localization using distinctive visual features. In *International Conference on Intelligent Robots and Systems*, IROS 2002 , Lausanne, Switzerland, pp. 226–231.

Shokoufandeh, A., Marsic, I., and Dickinson, S.J. 1999. View-based object recognition using saliency maps. *Image and Vision Computing* , 17:445–460.

Torr, P. 1995. Motion segmentation and outlier detection, Ph.D. The-sis, Dept. of Engineering Science, University of Oxford, UK.

Tuytelaars, T. and Van Gool, L. 2000. Wide baseline stereo based on local, affinely invariant regions. In British Machine Vision Conference, Bristol, UK, pp. 412–422.

Weber, M., Welling, M., and Perona, P. 2000. Unsupervised learning of models for recognition. In *European Conference on Computer Vision*, Dublin, Ireland, pp. 18–32.

Witkin, A.P. 1983. Scale-space filtering. In *International Joint Conference on Artificial Intelligence* , Karlsruhe, Germany, pp. 1019–1022.

Zhang, Z., Deriche, R., Faugeras, O., and Luong, Q.T. 1995. A robust technique for matching two uncalibrated images through the recovery of the unknown epipolar geometry. *Artificial Intelligence,* 78:87–119.