

基于局部尺度不变特征的目标识别

摘要

我们使用局部图像特征开发了一个目标识别系统。该特征对图像尺度，平移和旋转具有不变性，并且对光照变化和仿射或者三维投影也保持一定的稳定性。这些特征有着和颞下皮层中的神经元（neurons in inferior temporal cortex）相似的性质，用于高级视觉（primate vision）中的目标识别。通过在尺度空间中确定稳定点的分级过滤方法可以有效的探测特征。创建图像键（image keys），允许局部几何变形，通过在多方向多尺度的平面中表示模糊图像梯度。键可以作为最近邻索引的输入，用来确定候选对象匹配。通过寻找一个未知模型参数的低残留最小二乘解（low-residual least-squares solution）完成最终的匹配。实验结果显示在混乱的遮挡图像中可完成稳健的目标识别，并且计算时间在2秒之内。

1、引言

混乱的真实世界场景中的目标识别需要不被混乱和部分遮挡所影响的局部图像特征。特征必须至少对光照，三维投影变换以及常见的目标变化保持部分不变性。另一方面，特征必须足够独特，从而在许多选择中确定具体的目标。目标识别问题的困难很大程度上在于难以成功的找到这些图像特征。然而，近来在稠密图像特征的应用研究（如，Schmid&Mohr【19】）表明可以通过使用在大量的重复位置采样的局部图像描述符完成。

本文展示了一种新的方法，用来生成图像特征，称为尺度不变特征变换（SIFT）。这种方法把一张图像转换为大量的局部特征向量，他们都对平移，尺度以及旋转保持不变性，并且对光照变化和仿射或者三维投影也保持一定的稳定性。先前的部分特征生成的方法缺乏尺度不变性，并且对投影变形以及光照变化更加敏感。SIFT 特征和高级视觉中的（neurons in inferior temporal cortex）相似的性质。本文还介绍索引和模型验证的改进方法。

尺度不变特征可以使用分级滤波方法来有效的确定。第一级确定键在尺度空间中的位置，通过寻找差分高斯函数的最大值或最小值的位置。每个点用来生成一个特征向量，描述相对于尺度空间坐标框架的局部图像区域。特征对局部变化，如仿射或者三维投影保持部分不变性，通过使图像梯度位置模糊。这种方法基于哺乳动物视觉中的大脑皮层的复杂细胞的行为模型。产生的特征向量称为 SIFT 键。在目前的实现中，每张图像大约生成 1000SIFT 键，这个过程需要不到一秒的计算时间。

SIFT 键源自用于最近邻方法的图像，索引以确定候选目标模型。首先通过一个 hough 变换哈希表，发现与潜在模型的构成相一致的键集合，然后通过最小二乘拟合最终估计模型的参数。当至少 3 个键与低残留的模型参数一致时，就构成目标存在的有力证据。由于典型的目标图像中可能有许多 SIFT 键，虽然在图像中可能有大量的遮挡，但仍然保持很高的可靠性。

当前的目标模型表示为可以进行仿射投影的 SIFT 键的二维位置。特征位置可以充分的变化，以考虑远离相机高达 60 度旋转的平面形状或者高达 20 度旋转的三维对象的透视投影。

2、相关研究

目标识别广泛用于机器视觉行业，进行检查，登记和操作。然而，当前的目标识别商业系统几乎完全依赖于基于相关模板匹配的方法。虽然对特定的工程环境非常有效（目标构成和光照被严格控制），但是当目标旋转，尺度，光照和三维构成变化时，模板匹配变得在计算上不可行，当处理部分可见或者大型模型数据库时更是如此。

一种搜索所有图像的匹配位置的选择是从图像中提取特征，这些特征至少对图像形成过程部分不变，并且只与特征进行匹配。许多候选特征类型已经提出并探讨，包括线段【6】，

边集【11, 14】和区域【2】，以及许多其它的提议。虽然这些特征对特定目标类型是行之有效的，但是他们的探测数量往往不够多，对于形成可靠的识别缺乏足够的稳定性。

近来还有一些开发更稠密的图像特征集的研究。一种方法是使用角点探测(更准确地说，在局部图像变化中的峰值探测)来确定重复图像的位置，在这些位置周围可以测量局部图像性质。Zhang 等人【23】使用 Harris 角点探测来确定特征位置，从而校正来自不同视点的图像的核线。这种方法不试图将一张图像中的某区域与第二张图像的所有可能的区域进行相关，而是通过只匹配每张图像中以角点为中心的区域来大大减少计算时间。

对于目标识别问题，Schmid&Mohr【19】也使用 Harris 角点探测来确定兴趣点，然后在每个兴趣点处创建一个高斯导数图像测量的方向不变向量的局部图像描述符。这些图像描述符用作稳健的目标识别，通过寻找多匹配描述符，满足基于目标的方向和位置约束。这项研究的特点是在大型数据库中的识别速度和处理混乱图像的能力。

这些先前的角点探测方法有一个共同的缺陷，即他们只在单一尺度下检测图像。随着尺度变化的重要性增大，这些检测方法会反映不同的图像点。并且，由于这些检测方法不提供对象尺度的表示，所以有必要创建图像描述符并且试图在大量尺度下进行匹配。本文描述了一种有效的方法来确定不同尺度空间稳定的键位置。这意味着一幅图像的不同尺度不会对选择的键位置集合产生影响。此外，对于每个点确定一个显式的尺度，允许这个点的图像描述向量在每幅图像中相同的测度上进行采样。在每个点处确定一个标准方向，使得匹配可以在一个统一的局部二维坐标框架中进行。这就允许我们使用比 Schmid and Mohr 的旋转不变方法更加独特的图像描述符，并且对该描述符进一步改进，增强它对仿射投影和光照变化的稳定性。

其它基于外观的识别方法包括本征空间匹配 (eigenspace matching)【13】，颜色直方图 (color histograms)【20】，以及感受场直方图 (receptive field histograms)【18】。这些方法在孤立的目标或者预先分割的图像中被证明是成功的，但是由于他们是更全局的特征，这些方法很难扩展到混乱和部分遮挡的图像。Ohba&Ikeuchi【15】成功的把本征空间方法应用于混乱图像中，通过使用许多小的局部本征窗口 (eigen-windows)，但是这种方法需要对新图像中的每个窗口和匹配模板进行大量的搜索。

3、键位置

我们希望确定图像尺度空间中的位置，它对于图像平移，缩放和旋转具有不变性，并且对噪声和小的变形的影响最小。Lindeberg【8】展示了在尺度不变的一般假设下，高斯核以及它的导数是尺度空间分析的唯一可能的平滑核。

为了获得旋转不变性以及高效率，我们在尺度空间中的差分高斯函数的最大值和最小值处选择键位置。这可以有效地计算，通过在每一级之间重采样来建立图像金字塔。此外，我们也在高变化区域和尺度定位键点，使得这些位置在表示图像时非常稳定。Crowley&Parker【4】和 Lindeberg【9】之前基于其它目的，在尺度空间中使用了差分高斯函数。以下我们描述了一种非常有效和稳定的方法，来探测或者表示这个 (差分高斯) 函数的最大值和最小值。

由于二维高斯函数是可分的，它和输入图像的卷积可以通过使用两次一维高斯函数 (水平方向和垂直方向) 来有效地计算：

$$g(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-x^2/2\sigma^2}$$

对于键位置，所有的平滑操作都使用 $\sigma = \sqrt{2}$ ，这可以通过使用包含 7 个采样点的一维核来进行足够精确的估计。

输入图像首先与 $\sigma = \sqrt{2}$ 的高斯函数产生一幅图像 A。然后再使用一次 $\sigma = \sqrt{2}$ 的增量平滑产生一幅新的图像 B，相当于一个 $\sigma = 2$ 的有效平滑。差分高斯函数可以通过在 A 中减去 B 来得到，导致两个高斯（函数）之间的比率为 $2/\sqrt{2} = \sqrt{2}$ 。

为了生成下一个金字塔层，我们对已经平滑的图像 B 进行了双线性内插，在每个方向上使用 1.5 的像素间距。虽然使用 $\sqrt{2}$ 的相对尺度似乎对于重采样更加自然，唯一的约束是采样足够频繁从而检测峰值。1.5 的间距意味着每个新的采样将会是 4 个相邻像素的常线性组合。这样可以有效的进行计算，并且最小化混淆的人工痕迹（改变重采样系数时将会出现）。

这个尺度空间函数的最大值和最小值，通过比较金字塔中的每个像素与相邻像素来确定。首先，一个像素与它所在金字塔的同级的 8 个相邻像素相比较。如果它是这一级的最大值或者最小值，那么最接近的像素位置在金字塔的下一个最底层中计算，考虑 1.5 倍的重采样。如果该像素仍然比这个最接近的像素以及它的 8 个相邻像素高或者低，那么再次测试这一层。由于大部分像素将在几次比较之内消除，探测的代价很小，比建立金字塔的的代价低得多。

如果金字塔的第一级采样速度与输入图像相同，那么最高的空间频率将被忽略。这是因为初始平滑需要提供分离的峰值，从而进行稳健的探测。因此，我们通过一个因子 2 来扩展输入图像，使用双线性插值，在建立金字塔之前。这种方法对于一个典型的 512×512 像素的图像可以获得大约 1000 个键点，而如果没有初始的扩展，则只有四分之一的键点。

3.1 SIFT 键的稳定性

为了在每个键位置表示图像，对金字塔每级的平滑图像 A 进行处理来提取图像梯度和方向。

在每个像素点 A_{ij} 处，图像梯度值 M_{ij} 以及方向 R_{ij} ，使用像素之差来计算：

$$M_{ij} = \sqrt{(A_{ij} - A_{i+1,j})^2 + (A_{ij} - A_{i,j+1})^2}$$

$$R_{ij} = a \tan 2(A_{ij} - A_{i+1,j}, A_{i,j+1} - A_{ij})$$

像素之差可以有效的计算，并且可以提供足够的精度，由于前面的大量的平滑操作。在确定键位置时，有效的半像素位置移动被抵消。

对光照变化的稳健性可以通过将梯度大小的阈值定为最大可能梯度值的 0.1 倍。这就减弱了对三维立体表面光照方向变化的影响，因为光照变化可能导致梯度值发生大的变化，但是对梯度方向的影响较小。

每个键位置被赋予一个标准方向使得图像描述符对旋转具有不变性。为了使这个性质对光照或者反差变化具有足够的稳定性，该方向由局部图像梯度方向的直方图的峰值来表示。方向直方图使用高斯权窗口来创建，其中 σ 为当前平滑尺度的 3 倍。这些权会乘上限定阈值的梯度值，并且在方向 R_{ij} 对应位置的直方图上累积。直方图有 36 段，覆盖了 360 度的旋转范围，并且在选择峰值之前进行平滑。

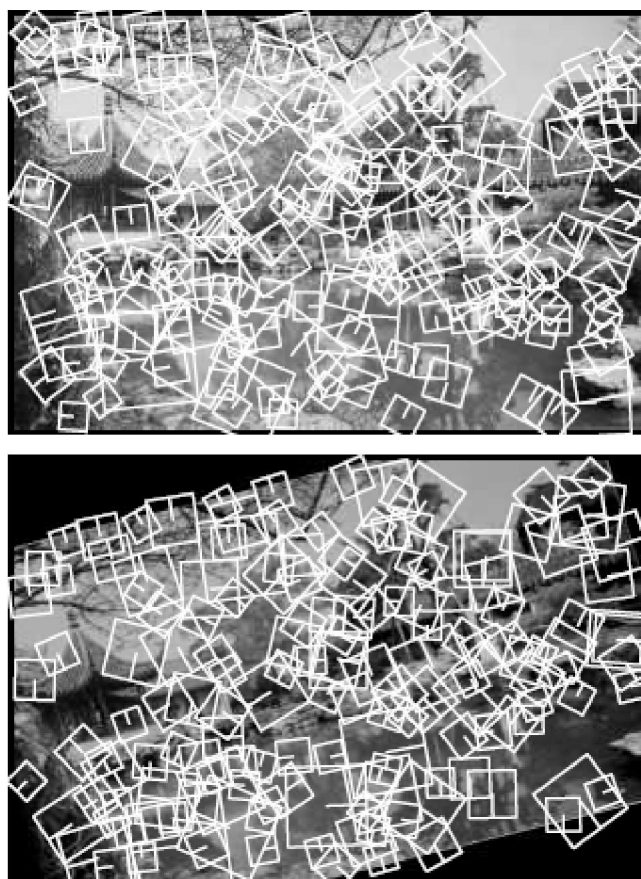


图 1：第二张图像由第一张图像通过旋转，缩放，伸缩，亮度和对比度变化，以及加入噪声来生成。尽管发生了这些变化，第一张图像中 78 的键在第二张图像中都有精密的匹配键。为了减少混乱，这些例子只展示了键的一个子集。

获得的键的稳定性可以通过把自然图像进行仿射变换，对比度和亮度变化，以及加入噪声来测试。在第一副图像中每个键的位置可以在转换图像中通过变换参数的知识来进行预测。这个框架用来选择以上给出的不同的采样和平滑参数，以获得最大的效率同时保持对各种变化的稳定性。

| Image transformation | Match % | Ori % |
|------------------------------|---------|-------|
| A. Increase contrast by 1.2 | 89.0 | 86.6 |
| B. Decrease intensity by 0.2 | 88.5 | 85.9 |
| C. Rotate by 20 degrees | 85.4 | 81.0 |
| D. Scale by 0.7 | 85.1 | 80.3 |
| E. Stretch by 1.2 | 83.5 | 76.1 |
| F. Stretch by 1.5 | 77.7 | 65.0 |
| G. Add 10% pixel noise | 90.3 | 88.4 |
| H. All of A,B,C,D,E,G. | 78.6 | 71.8 |

图 2：对 20 幅图像的一个样本应用的不同的图像变换，这个表给出了在匹配位置和尺度找到的键的百分比（Match），以及还在方向上匹配了的键的百分比（Ori）。

图 1 展示了在 2 个大尺度的倍频程 (octave range) 内的数量相对较少的键 (以避免过度的混乱)。每个键表示为一个正方形, 包含一条由中心到正方形的一条边的直线, 来指示方向。在图的下半部分, 图像旋转 15 度, 缩放了一个 0.9 的因子, 在水平方向上伸缩了一个 1.1 的因子。像素深度, 在 0 到 1 的范围中, 从他们的亮度值中减去了 0.1, 对比度通过乘以 0.9 减少。然后加入随机像素噪声, 每个像素信号加入少于 5bit 的信号。尽管采用了这些变换, 第一副图像中 78 的键在第二幅图像中的预定位置, 尺度和方向上有精密的匹配键。

键对图像变换的总体的稳定性可以从表 2 来判断。这个表的每个条目是用 20 种不同的测试图像的组合结果来生成的, 总结了大约 15000 个键的匹配。表的每一行展示了一个特定的图像变换。图 1 给出了与转换的图像中的键相匹配的键的百分比, 其中位置在 σ 之内 (相对于键的尺度) 以及一个 1.5 的缩放因子。第 2 列给出了满足这些标准以及在预测方向的 20 度范围之内的键的百分比。

4、局部图像描述符

对于每个键, 给定一个稳定的位置, 尺度以及方向, 就可以描述局部图像区域, 它对于这些变换具有不变性。另外, 我们希望这种表示对局部的小的几何变化, 如仿射或者三维投影具有稳健性。一种方法是由视觉皮层的复杂神经元的响应特性启发的, 即一个特征位置允许在小范围内变化但同时保持方向和空间频率的特性。Edelman, Intrator&Poggio【5】进行实验, 模拟了复杂神经元对计算机图形模型的不同的三维视图的响应, 发现复杂细胞的输出比简单的基于相关的匹配提供了更好的辨别力。这可以看出, 比如, 如果一个仿射投影伸缩了一张图像, 在相对于另一种图像的一个方向上, 这就会改变梯度特征的相对位置, 对他们的方向和空间频率有较小的影响。

这种对局部几何变形的稳健性可以通过表示布局图像区域来获得。多张图像对每一张都表示了一些方向 (称为方向平面)。每个方向平面只包含该方向相应的梯度。每个方向平面都经过了模糊和重采样, 以允许梯度位置发生大的变化。

这种方法可以使用相同的预先计算的用来选择方向的金字塔每级的梯度和方向来有效的实现。对每个键点, 我们在探测到的键所在的金字塔级中使用像素采样。在键位置周围半径为 8 的圆内所有的像素都被插入到方向平面中。方向相对于键的方向来度量, 通过减去键的方向。在我们的实验中, 我们使用 8 个方向平面, 每个在 4×4 的网格中进行采样, 采样间隔为梯度探测使用的像素间隔的 4 倍。模糊是通过分配它的 8 个最邻近像素的梯度, 在方向和两个空间纬度上使用线性内插来完成。这种实现方法比显式模糊和重采样更加有效, 但同时给出了几乎相同的结果。

为了在一个大尺度对图像进行采样, 相同的过程对金字塔高一阶的第二阶重复一次。然而, 这次使用 2×2 而不是 4×4 的采样区域。这意味着相同的图像区域在所有尺度内几乎都会被检查, 使得任何相邻的遮挡不会其它尺度。因此, 在 SIFT 键向量中样本的全部数量为 $8 \times 4 \times 4 + 8 \times 2 \times 2$ 或者 160 个元素, 给定高特性的足够的度量。

5、索引和匹配

对于索引, 我们需要对每个采样图像存储 SIFT 键, 然后从新的图像中确定匹配键。对高维向量确定最相似的键的问题在于得到精确解需要很大的计算复杂度。然而, kd 树算法的一个变体叫做 best-bin-first 搜索方法 (Beis&Lowe【3】), 可以以很大的概率确定最近邻居, 只使用有限的计算。为了进一步改进 best-bin-first 算法的效率, 在大尺度处生成的 SIFT 键样本的权重为小尺度处的两倍。这意味着大尺度实际上能够筛选在小尺度处检查出的最可能的近邻。这种方法还改进了识别性能, 通过给噪声较小的尺度以更大的权重。在我们的试验中, 当在 30000 个键向量中使用概率 best-bin-first 搜索最多 200 个近邻时, 可能会出现一个分离

点，这与找到一个精确解相比几乎没有性能损失。

聚合可靠的模型假设的有效方式是使用 Hough 变换【1】来搜索键，使它符合一个特定的模型。数据库中的每个模型键包含一组相对于模型坐标系统的键参数的记录。因此，我们可以在哈希表中创建一个条目，预测模型位置，方向，以及匹配假设的尺度。我们使用一个桶的 30 度的旋转因子，2 的缩放因子，0.25 倍的最大模型位置维度。这些宽松的尺寸考虑到了聚合，即使在大量的几何变形的表示中，比如由于三维视点的变化（产生的几何变形）。为了避免哈希中的边界效应问题，每个假设哈希到每个纬度的 2 个最近的桶中，对每个假设给定总共 16 个哈希表条目。

6、仿射参数的解

搜索哈希表来确定所有的桶中的至少 3 串条目，桶按大小降序排列。然后，每个这样的串送入核查程序，对模型和图像的仿射投影参数进行最小二乘解。

一个模型点 $[x, y]^T$ 到图像点 $[u, v]^T$ 的仿射变换可以表示为

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

其中模型平移参数为 $[t_x, t_y]^T$ ，仿射，旋转，缩放，伸缩用参数 m_i 表示。



图 3：二维对象的模型图像在最上面。识别结果显示了模型轮廓和匹配使用的图像键。
我们希望解出变换参数，所以上面的方程可以转化为

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

这个方程展示了一个单一的匹配，但是可以进一步加入许多匹配，每个匹配对第一个和最后

一个矩阵贡献两行。需要至少 3 个匹配才能提供一个解。

我们可以把这个线性系统表示为

$$Ax = b$$

参数 x 的最小二乘解可以通过解相应的正规方程 (normal equations) 来确定，

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

使得投影模型位置到相应图像位置的距离的平方和最小。这个最小二乘方法可以轻易的扩展，以解出三维姿态和铰接的易弯曲的对象的内部参数【12】。



图 4：最上方显示了三维对象的模型图像，包含由背景分割找到的轮廓。最下方的图像展示了三维对象的识别结果，包含模型轮廓和匹配使用的图像键

现在奇异值可以通过在给定参数解的情况下，检查每个图像特征和模型的一致性来消

除。每个匹配必须在 15 度方向变化， $\sqrt{2}$ 的尺度变化，0.2 倍的最大模型位置尺寸值之内一致。如果在消除其一点之后只剩下不到 3 个点，匹配将被拒绝。如果任何匹配都被消除，那么最小二乘解可以使用剩余的点得到。



图 5：实例——有遮挡的三维目标的识别。

7、实验

仿射解提供了对二维对象的透视投影的好的近似，所以二维模型提供了该方法的一个好的初始测试。图 3 的最上方展示了对对象的矩形二维面的三个模型图像。该图还展示了包含二维对象的混乱的图像，相同的图像被识别模型所覆盖。显示的模型键是用来识别和最后进行最小二乘解的键。稳健的识别只需要 3 个键，可见解是高度冗余的，并且不受大量遮挡的影响。还展示了模型图像的矩形边界，通过最小二乘解的仿射变换进行投影。这跟图像中的平面区域的真实的边界非常一致，除了由透视投影引入的小的误差。对许多二维对象的图像进行了相似的实验，结果表明，识别在目标相对于相机的旋转度至少 60 度的范围内非常稳健。

虽然模型图像和仿射参数对三维对象在深度上的旋转不适用，他们仍然能够对三维对象进行足够稳健的识别，在每个模型视图的 20 度旋转范围之内。图 4 的最上方展示了三个模型图像的例子。模型在黑色背景中拍摄，对象的轮廓通过在背景区域上进行分割来提取。在该图还展示了一个识别的例子，展示了识别中使用的 SIFT 键。对象轮廓通过仿射参数解来

投影。但是这种情况下一致性没有那么精密，因为这个解对在深度上的旋转不适用。图 5 展示了更多的例子，其中有重要的部分遮挡。

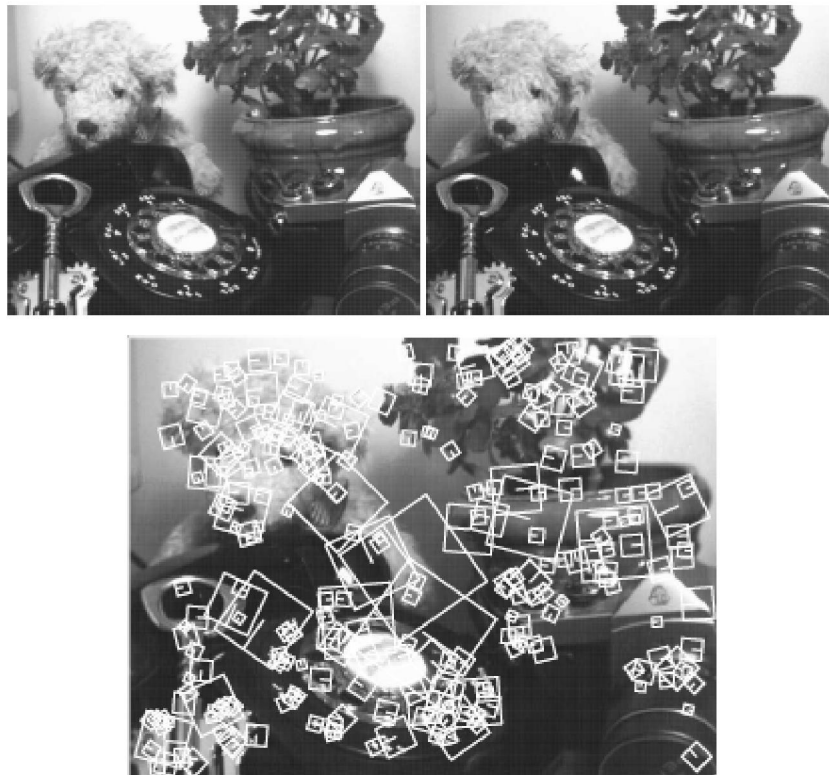


图 6：图像键的稳定性在不同的光照环境中进行测试。第一幅图像从左上方照射，第二幅图像从右上方照射。底部图像的展示的键在前两幅图像的匹配中使用。

这些例子中的图像大小为 384×512 。每张图像中的所有目标识别的计算时间为 1.5 秒，在 SUN 的 10 处理器的工作站上，其中需要 0.9 秒创建尺度空间金字塔以及确定 SIFT 键，然后大约 0.6 秒进行索引和最小二乘验证。这不包括对每个模型图像进行预处理的时间，大约为每张图像 1 秒钟，当然这只需要在模型数据库中创建初始条目时进行一次。

图 6 中展示了 SIFT 键的光照不变性。两张图像在相同视点的相同场景中，但是第一张图像从左上方照射，第二幅图像从右上方照射。完整的识别系统把第一张图像作为模型来确定第二张图像，第二张图像通过与第一张图像匹配来正确地识别。该图只显示了识别的部分 SIFT 键。有 273 个键被认定为最终的匹配，这意味着在每一种情况下，不仅在相同位置探测到了相同的键，而且它也是与第二张图像中相应的键的最精密的匹配。这些键中任意 3 个对于识别来说都是足够的。虽然在一些显著或者阴影变化区域没有找到匹配点（比如相机的闪亮的顶端），一般来说键对光照变化具有很好的不变性。

8、与生物视觉的联系

人类视觉的性能显然远远超过当前的计算机视觉系统，所以我们可以通过模拟生物过程来极高性能。幸运的是，在过去的几年里，在理解动物和人类进行目标识别的原理方面，已经有了显著的改善。

近来在神经科学方面的研究表明灵长类动物的目标识别使用中间复杂度的特征，它对尺度，位置和光照变化有很大的不变性（Tanaka【21】，Perrett&Oram【16】）。这种在颞下皮层（IT）中发现的中间特征的一些例子是对黑暗五角星，具有突出元素的圆，或者三角边界内的水平纹理区域进行响应的神经元。这些神经元对形状特征保持高度的响应特性，这些形

状特征出现在视场的大部分区域，以及多个尺度范围上（Ito 等人【7】）。许多这些特征的复杂性似乎与当前的 SIFT 特征相同，虽然还有一些神经元响应更复杂的形状，比如面。许多神经元除了响应形状之外，还响应颜色和纹理属性。特征响应取决于先前的对包含特征的特定对象的视觉学习（Logothetis, Pauls&Poggio【10】）。这些产生在大脑中的特征，通过高度计算密集型的并行处理，跟本文中的分阶段滤波方法非常不同。然而，结果大致相同：一幅图像转化为大量的局部特征，每个特征匹配一小部分的潜在对象，仍然对常见的视图变换保持很大的不变性。

我们还知道，人脑中的目标识别取决于一系列从注意力到约束特征到目标解译，确定姿态，从混乱背景中分割目标【22】。这个过程在本文中应用的参数求解和奇异点探测方面，几乎跟验证发挥着相同的作用，因为解译的准确度常常取决于单一视点约束的执行【11】。

9、总结和评论

SIFT 特征改进了先前的方法，因为它对尺度，光照，和仿射变形具有很大的不变性。典型的图像中大量的特征考虑到对混乱图像中的部分遮挡进行稳健的识别。解决仿射模型参数的最后阶段考虑到相对于确定依赖于索引的方法，更准确的验证和姿态的确定。

进一步研究的重要领域是从不同视角建立模型，从而表示对象的三维结构。这就会有更大的优势，多个视角条件下的键可以组合到一个模型中，从而增大了在新的视角找到匹配的可能性。模型可以用基于由运动到结构的方法（structure-from-motion solution）进行真三维表示，或者以自动聚类 and 内插的方式表示外观空间（Pope&Lowe【17】）。后一种方法的优点是它还可以对非刚性进行建模。

识别性能可以通过增加新的 SIFT 特征类型到混合的颜色，纹理和边集，以及变化的特征大小和偏移来进一步地改善。尺度不变的边集可以对局部图面进行判别，在对象边界非常有用，边界处背景的混乱会干扰其它特征。索引和验证框架允许所有类型的尺度和旋转不变特征，合并到单一的模型表示中。通过许多不同的特征类型，依靠索引和聚类在特定图像中选择最有用的特征类型，可以获得最大的稳健性。

10、参考文献