

中图分类号:TP391.4 文献标志码:A 文章编号:1006-8961(2011)09-1625-07

论文索引信息:周丽平,孙志峻,赵淳生. 用于目标快速识别的特征选取方法 [J]. 中国图象图形学报,2011,16(9):1625-1631

用于目标快速识别的特征选取方法

周丽平¹⁾,孙志峻¹⁾,赵淳生²⁾

¹⁾(南京航空航天大学机电学院,南京 210016)

²⁾(南京航空航天大学精密驱动研究所,南京 210016)

摘要:现有的基于图像局部特征的目标识别算法,在保证较高识别率的情况下无法满足实时性要求。针对这个问题,考虑到多数局部特征是不稳定、不可靠或与目标无关的,可通过正确匹配的训练图像,对图像局部特征选取一个子集用于目标识别。提出一种在特征包方法基础上,通过无监督地选取鲁棒性强及足够特殊、稳定的局部特征用于目标识别的新方法并应用于目标识别实验。实验结果证实该方法在仅仅使用原图像约4%的局部特征的情况下获得了与使用全部局部特征几乎相近的目标识别率,目标识别时间由秒缩短至几十毫秒,满足了目标识别实时性要求。

关键词:局部特征;特征包;目标识别;实时性

Research on selection of features approach for fast object recognition

Zhou Liping¹⁾, Sun Zhijun¹⁾, Zhao Chunsheng²⁾

¹⁾(College of Mechanical and Electrical Engineering, Nanjing University of Aeronautics & Astronautics, Nanjing 210016 China)

²⁾(Precision Driving Laboratory, Nanjing University of Aeronautics & Astronautics, Nanjing 210016 China)

Abstract: Existing methods based on local features cannot recognize objects in real-time while keeping a high recognition rate. Considering that many local features are unstable, unreliable, or irrelevant, we are able to select a small subset of features used for recognition by correctly matching features in training images. A new, robust, and stable method based on a bag-of-features is proposed in this paper. Distinctive features are selected by an unsupervised preprocessing step. Our experiments demonstrate that this selection approach can reduce the amount of local features and reduce the memory requirements, while allowing an average of 4% of the original features per image to provide matching performance that is as accurate as the full set. The method can meet real-time requirements since the time required for matching has been reduced from seconds to tens of milliseconds.

Keywords: local features; bag-of-features; object recognition; real-time

0 引言

目标识别是目标跟踪、机器人视觉导航与视觉定位、智能监控等相关方向首要研究的课题,是计算机视觉领域中最活跃的研究主题之一。由于

局部特征在部分遮挡的情况下或杂乱的场景中具有很强的不变性,已经被证明可以较好地用于目标匹配和识别等任务^[1-3]。在众多局部特征检测算法中,由于尺度不变特征(SIFT)对于光照变换、视角变换、部分遮挡及仿射变换等具有不变性,该方法在模式识别领域取得了巨大成功^[1]。该方法

收稿日期:2010-07-23;修回日期:2010-11-11

基金项目:国家自然科学基金重点项目(90816003,U0934004)。

第一作者简介:周丽平(1985—),男,南京航空航天大学机械电子工程专业博士研究生,主要研究方向为机器视觉、模式识别。E-mail:zhouliping85@yahoo.com.cn。

在无监督的情况下对目标的识别率为 80% ~ 95%, 在有监督的情况下对目标识别率为 95% ~ 100%^[4]。虽然利用 SIFT 特征进行目标识别能获得一个较高的识别率, 但 SIFT 特征匹配时计算负载大。对于一幅分辨率为 640 × 480 的图像大约能检测出 2 000 个 SIFT 特征点, 两幅图像之间的匹配时间约为 2 s^[1]。所以文献[1]中方法不能满足目标跟踪、机器人视觉导航与视觉定位、智能监控等领域对目标识别的实时性要求, 因此目标识别时间成为了 SIFT 方法在上述领域成功应用的瓶颈, 即对逐个存储与比较每幅图像的每个局部特征的传统方法提出了挑战。

Sivic 和 Zisserman 提出的基于特征包方法是一种针对大量场景及物体识别的方法。该方法通过矢量化局部图像视觉特征, 对关键点特征进行聚类, 以生成视觉词汇表, 进而对图像进行描述。近年来的许多研究工作已经表明, 利用特征包表示的局部特征能够获得较好的场景及物体分类性能^[5-7]。与单独存储和匹配每一个特征的传统方法相比, 特征包方法减少了存储空间的使用。但基于特征包方法仅仅是减少了一个数量级的内存使用, 而并没有减少图像局部特征的数量, 因此, 特征包方法对目标识别时间并无改善。

注意到许多局部特征是不稳定或与目标无关

的, 在不影响目标识别率的情况下对图像局部特征选取一个子集用于目标识别则可缩短目标识别的时间。基于此, 利用 SIFT 特征的优点, 用目标自身的 SIFT 特征点来描述目标, 在特征包方法的基础上, 在目标全部的 SIFT 特征中通过无监督选取鲁棒性强、足够特殊及稳定的特征用于目标识别, 在获得一个较高的识别率的同时进一步减少了存储每张图像的特征所需的空间及目标特征数量。那些鲁棒性强、足够特殊及稳定的特征本文称为有效特征。图 1 是本文所做的 3 组目标图像(基准图像)对比实验(所涉及图像分辨率皆为 640 × 480), 图中实心红圈为提取的图像 SIFT 特征。图 1 上行图像中红圈是根据文献[1]中方法所提取的图像 SIFT 特征, 每幅图像包含约 2 000 个特征点; 图 1 下行图像中红圈是依据所需识别的目标及本文的方法所选取的图像的有效特征(从左至右分别为教学楼、逸夫科学馆及丰田轿车)。由图 1 上下行对比可看出, 本文方法大大减少了目标图像中大量的前景、背景及目标自身不稳定的特征点, 所保留的有效特征数量约占原图像局部特征总数的 4%。实验结果表明, 应用上述有效特征进行目标识别能够获得与使用全部图像局部特征几乎相近的识别率, 且由于减少了大量图像局部特征而大大缩短了目标识别所需的时间。

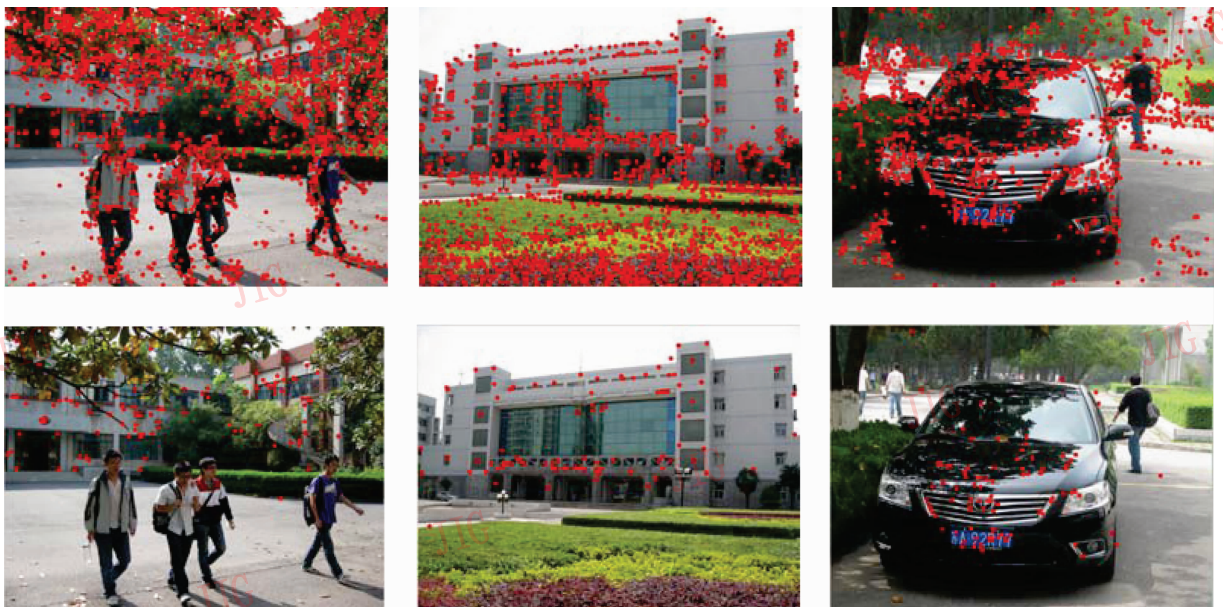


图 1 图像全部特征与有效特征

Fig. 1 Compare of all image features and useful image features

1 特征包

特征包方法主要可分为图像特征提取及描述和将图像特征描述子量化并进行聚类以构建视觉词汇表两个步骤^[8]。

1.1 图像特征提取及描述

采用 SIFT 特征来描述图像局部特征, SIFT 特征对图像的尺度、旋转、光照改变等保持一定的稳定性,在匹配时可解决部分遮挡、杂乱或噪声等问题^[1,8-9],因此 SIFT 特征可以很好地满足目标识别

的要求。采用 DOG (difference of Gaussian) 方法计算图像 SIFT 特征。SIFT 描述子是以特征点为中心取 16×16 的图像数据块,如图 2 所示,图 2 左部中心点为当前特征点位置,箭头长度代表梯度模值,箭头方向代表梯度方向,圆圈代表高斯加权的范围(越靠近特征点的像素梯度方向贡献越大);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,计算每个梯度方向的累加值,形成如图 2 右部所示的种子点,最终形成 $4 \times 4 \times 8$ 共 128 维的向量^[1]。

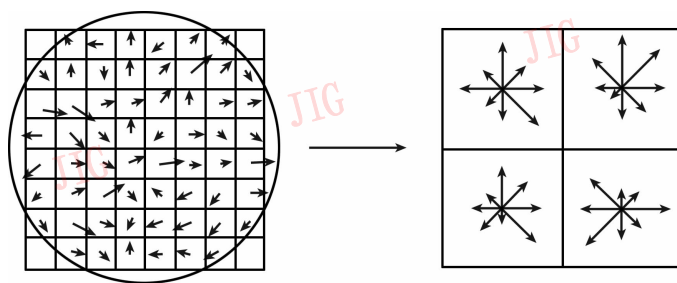


图 2 SIFT 特征点描述方法

Fig. 2 SIFT feature point descriptor

1.2 视觉词汇构建

视觉词汇构建可以看做是特征向量量化及聚类的过程。用 K-means 方法构建所需的视觉词汇表。K-means 方法先随机产生 K 个族中心,然后将特征点逐一的分配至离它们最近的族中心的族中,再重新计算族中心并重复上述步骤。K-means 方法简单易行,且有效,但在实际情况中,我们并不知道如何选取 K 值。 K 值过大则计算负载大; K 值过小则不能保证聚类精度。对此,目前常用的方法是逐步增大 K 值直到获得一个满意的结果^[10],使得最大聚类误差落在区间 $[1\%, 3\%]$ 。所构建的视觉词汇表定义为 $w = \{w_1, w_2, \dots, w_k\}$ 。

2 有效特征检测

2.1 有效特征

在提取图像特征时,许多无效的特征被检测出来。这些无效的特征主要包括不稳定兴趣点特征、无增长信息的特征及前景、背景中的特征。在实际应用中,若能剔除该部分特征将有助于实现目标的实时识别。注意到前景、背景及物体的不稳定的特

征点往往只出现在某一张图像中,而有效特征往往会出现在多张包含同一物体的图像中,基于此,通过检测那些存在于多个视角的目标图像特征来识别有效特征,并剔除不稳定、前景、背景的特征以缩短目标识别的时间。这种方法要求对于要识别的目标至少要有两张不同视角的图像。在实际情况中,这个要求通常是能够满足的。图 3 中是从各个角度拍摄的包含目标的图像,本文称为训练图像,用于训练基准图像目标特征以检测有效特征。为不失一般性,训练图像皆为与基准图像存在仿射变换的图像,或包含杂乱场景的图像,或是目标的部分图像。本文假定所有用来训练的图像都是未标定的。因此,整个有效特征检测是一个无监督的过程。

2.2 有效特征检测原理

有效特征检测过程可以看做是一个对基准图像特征训练的过程。用 N 张包含同一目标的训练图像 $I_j, j=1, 2, \dots, N$, 逐一对基准图像中的特征进行匹配。由于有效特征是对图像视角、光照等因素变化具有鲁棒性、稳定性的特征,因此在对基准图像特征匹配时,基准图像中具有成功匹配的特征点都是候选的有效特征。但简单地将有效特征产生规则定



图3 包含目标训练图像

Fig. 3 Training images contain of objects

义为具有 t 次成功匹配的特征点, 则对一般性目标难于确定一个通用的 t 值, 使得在不影响对目标识别率的情况下选取的有效特征数量最优, 其中 $t = 1, 2, \dots, N$ 。经实验验证, 如果 t 值偏大, 检测的有效特征数量将过少, 直接导致对该目标的识别率大大降低; 相反, 如果 t 值偏小, 检测的有效特征将过多, 还可进一步优化以缩短目标识别时间。对一般性目标难于找到一个通用 t 值使得检测的有效特征数量在不影响对目标的识别率情况下最少。

对此, 通过统计基准图像的视觉词汇 w_i 在训练过程中具有的正确匹配次数, 引入 TF-IDF 方法计算基准图像所有视觉词汇 w_i 的权重系数 x_{ib} , 并据此自适应地检测有效特征。具体方法如下:

对训练图像集合 I_j , 其中 $j = 1, 2, 3, \dots, N - 1, N$, 用训练图像特征对基准图像特征进行匹配, 对基准图像中具有正确匹配的视觉词汇 w_i , 使用式(1)重新计算视觉词汇 w_i 在基准图像中出现的次数。

$$m_{ib} = n_{ib} + \sum_{j=1}^N n_{ij} \quad (1)$$

式中, n_{ib} 表示视觉词汇 w_i 在基准图像中出现的次数, n_{ij} 表示视觉词汇 w_i 在训练图像 I_j 中出现的次数, m_{ib} 表示训练完成后新的 n_{ib} 值。

训练完成后使用直方图表达视觉词汇 w_i 的出现次数 n_{ib} , 使用式(2)计算视觉词汇 w_i 的 TF-IDF 权重系数

$$x_{ib} = \frac{n_{ib}}{K} \log \frac{N}{\sum_{i=1}^N n_{ib} \quad \sum_{i=1}^N (n_{ij} > 0)} \quad (2)$$

式中, x_{ib} 表示视觉词汇 w_i 在基准图像中 TF-IDF 的权重系数, N 是用于训练图像的数量, K 是视觉词汇的数量。

保留过多特征点不能满足目标识别实时性要求, 相反将导致目标识别的失败。因此, 本文在不影响目标识别率的情况下, 经过多次实验将最终有效特征数目限定在区间 $[50, 100]$ 。用 F 和 g 分别表示具有正确匹配的视觉词汇 w_i 集合和数目, 有效特征选取规则如下:

- 1) 若 $100 \geq g \geq 50$, 则集合 F 中元素为最终输出的有效特征;
- 2) 若 $g > 100$, 则根据 x_{ib} 在集合 F 元素中选取前 100 个最大值作为有效特征;
- 3) 若 $g < 50$, 则将集合 F 中元素及在所有视觉词汇中根据 x_{ib} 取前 $100 - g$ 个最大值作为有效特征。

2.3 有效特征检测实验

实验中首先对每类目标的基准图像及训练图像提取 SIFT 特征, 用 K-means 方法对所有图像 SIFT 特征进行聚类构建词汇表。聚类误差随 K 值变化曲线如图 4 所示。本文中 K 取值在区间 $[200, 1200]$ 变化, 由图 4 可知, 在 $K = 600$ 增长到 $K = 1200$ 之间, 聚类误差变化缓慢, 且满足 K-means 方法的聚类精度要求, 因此, 本文取 $K = 600$ 以保证聚类精度和计算速度的平衡。

用 N 张训练图像分别对基准图像进行训练, 通过统计基准图像中视觉词汇 w_i 的出现次数、直方图

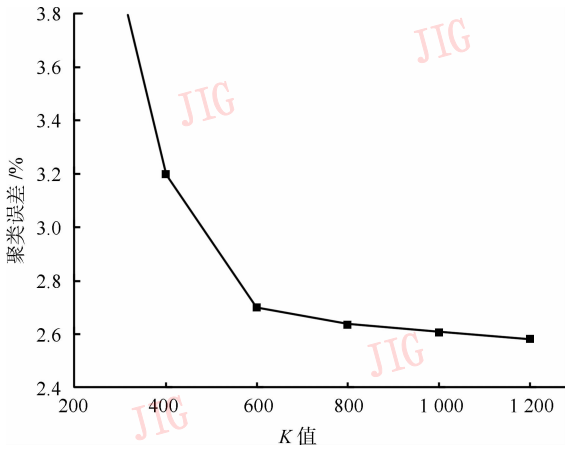


图4 聚类误差随K值变化曲线

Fig. 4 The error rate found of different choices of K

及其 TF-IDF 权重系数来检测有效特征。在分别训练 1,3,5 次情况下,对图 1 中 3 组目标图像的有效特征检测实验结果如图 5 所示。与图 1 对比中可以

看出,通过训练图 5,结果中减少了基准图像中大量的前景、背景等杂乱环境的特征及目标自身不稳定的特征,进一步减少了存储图像特征所需的存储空间。由于图 3 中的训练图像皆为与基准图像存在仿射变换的图像,或包含杂乱场景的图像,或是目标的部分图像且都是未标定的,因此检测出的有效特征具有鲁棒性强、稳定及特殊等特点。有效特征数量约为图像局部特征数量的 4%,应用有效特征可以实现对目标的快速识别。同时注意到,有效特征检测结果随着训练次数的增加而有所不同,这是由于基准图像中视觉词汇 w_i 的权重系数 x_{ib} 在不同训练次数情况下重新排序结果。此外,检测出的有些特征仍然包含少数前景、背景的特征点,这主要是用 K-means 方法对 SIFT 特征进行聚类的误差及前景、背景中存在少量权重系数较大的视觉词汇。随着训练次数的增加,有效特征包含的这类特征数量呈逐步缩小趋势。



图5 有效特征检测结果

Fig. 5 Experimental results of useful features detection

3 目标识别实验

为验证在仅使用有效特征进行目标识别的情况下对目标识别率的影响,对上述目标在不同时间、不同光照、不同角度条件下分别采集了 30 幅图像作为测试图像,用于目标识别实验。目标在测试图像中至少有 25% 是可见的。

对于测试图像 I ,按照上述所描述的方法对其提取 SIFT 特征,并将图像特征量化成视觉词汇。用 $d = \{d_1, d_2, d_3, \dots\}$ 来描述所提取的测试图像 I 的 SIFT 特征;所构建的视觉词汇为 $w = \{w_1, w_2, \dots, w_k\}$,对于描述子 d 到视觉词汇 w 的映射可用最近邻方法计算,即计算 d_j 与 w 中所有视觉词汇的欧氏距离,而后将 d_i 量化成 w 中对应的与 d_j 的欧氏距离最小的视觉词汇 w_i ,其中 d_j ,

w_i 分别属于集合 d, w ,如式(3)所示

$$d_j \rightarrow w_i = \underset{w}{\operatorname{argmind}}(d_j, w) \quad (3)$$

然而随着视觉词汇量 K 的增大,关于描述子 d 到视觉词汇 w 的映射的计算负载会迅速增加,因此,本文用近似最近邻方法来计算描述子 d 到视觉词汇 w 的映射。近似最近邻方法在保证高精度的同时大大减少了计算时间^[11]。对所有测试图像,在完成将图像特征量化成视觉词汇后,用于本文的目标识别实验,根据文献[1]结论,基准图像与测试图像中目标有 3 个或 3 个以上匹配成功的 SIFT 特征即可认为对目标识别成功。表 1 是在使用基准图像全部局部特征与仅使用有效特征的情况下对测试图像的目标识别率的实验结果,且分别列举了在不同训练次数(N 值)情况下对目标的识别率。

表 1 不同训练次数情况下的目标识别率

Tab.1 Object recognition rates with different of N

测试样本图像	识别率/%			
	全部局部特征	$N = 1$	$N = 3$	$N = 5$
教学楼	90.00	86.67	86.67	90.00
逸夫科学馆	93.33	90.00	93.33	93.33
丰田轿车	83.33	76.67	76.67	80.00
平均识别时间/(ms/幅)	1 500		87	

由表 1 可知,在仅使用有效特征的情况下获得了与使用图像全部局部特征几乎相近的目标识别率,由于减少了大量的前景、背景及目标自身不稳定的特征点,基准图像与测试图像之间平均匹配时间即对目标识别的时间降至 87 ms(所使用的计算机 CPU 为 P4 3.2 GHz,内存 512 MB),较使用基准图像全部局部特征进行目标识别的计算效率提高了约 17 倍,能够满足目标跟踪、视觉导航与视觉定位、智能监控等领域对目标识别实时性的要求。同时,随着训练次数增加,使用基准图像有效特征对目标识别率呈上升趋势。这是由于根据式(1),基准图像中具有一定鲁棒性、较稳定的特征会随着训练次数增加得到增强,该特征对应的视觉词汇 w_i 的 TF-IDF 权重系数亦因此增大,即随着训练次数增加,基准图像中的鲁棒性强、稳定的特征被检测出的概率越大,有效特征整体鲁棒性更强。

4 结论

针对 SIFT 方法在模式识别领域拥有较高的识别率但无法满足实时性这一问题,考虑到目标图像中多数局部特征是不稳定、不可靠或与目标无关的,提出一种在特征包方法的基础上通过正确匹配的训练图像,在目标图像全部 SIFT 特征中无监督的选取一个鲁棒性强及足够特殊、稳定的特征子集用于目标识别的新方法。实验结果证实,该方法能够减少目标图像中大量的前景、背景及目标自身不稳定的特征点,在特征包方法的基础之上进一步减少了每张图像保存的特征数量及存储每张图像特征所需的空空间,在仅仅使用原图像约 4% 的局部特征的情况下获得了与使用全部局部特征几乎相近的目标识别率,而目标识别速度提高了约 17 倍,目标识别时间缩短至 87 ms。将本文方法用于目标识别既维持了

一个较高的识别率,又满足了目标跟踪、视觉导航与视觉定位、智能监控等领域对目标识别实时性要求。然而,本文的工作要求目标至少要有两张不同视角的图像,对单一的图像本文的方法并不适用;此外,本文的工作并未提高对目标的识别率。如何在基准图像中增加相邻视角图像的部分特征以提高对目标的识别率及在单一图像中选取有效特征将是以后工作着重解决的问题。

参考文献 (References)

- [1] Lowe D G. Distinctive image features from scale-invariant keypoints [J]. International Journal of Computer Vision, 2004, 60(2) :91-110.
- [2] Mikolajczyk K, Schmid C. Scale & affine invariant interest point detectors [J]. International Journal on Computer Vision, 2004, 60(1) :63-86.
- [3] Lowe D G. Local feature view clustering for 3D object recognition [C] // IEEE Conference on Computer Vision and Pattern Recognition, Kauai, Hawaii, USA ; IEEE Press, 2001 :682-688.
- [4] Munich M E, Pirjanian P, Bernardo E D, et al. IFT-ing through features with ViPR: application of visual pattern recognition to robotics and automation [J]. IEEE Robotics and Automation Magazine, 2006, 13(3) :72-77.
- [5] Mikolajczyk K, Schmid C. A performance evaluation of local descriptors [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2005, 27(10) :1615-1630.
- [6] Philbin J, Chum O, Isard M, et al. Object retrieval with large vocabularies and fast spatial matching [C] // IEEE Conference on Computer Vision and Pattern Recognition. Minneapolis, USA ; IEEE Press, 2007 :1-8.
- [7] Philbin J, Chum O, Isard M, et al. Lost in quantization: improving particular object retrieval in large scale image databases [C] // IEEE Conference on Computer Vision and Pattern Recognition. Anchorage, Alaska ; IEEE Press, 2008 :1-8.
- [8] Xiong Ying, Ma Huiming. Extraction and application of 3D object SIFT feature [J]. Journal of Image and Graphics, 2010, 15(5) :814-819. [熊英, 马惠敏. 3 维物体 SIFT 特征的提取与应用 [J]. 中国图象图形学报, 2010, 15(5) :814-819.]
- [9] Zhang Jieyu, Bai Xiaojing, Xu Liyan, et al. A method of correcting SIFT mismatching based on spatial distribution descriptor [J]. Journal of Image and Graphics, 2009, 14(7) :1369-1377. [张洁玉, 白小晶, 徐丽燕, 等. 基于空间分布描述符的 SIFT 误匹配校正方法 [J]. 中国图象图形学报, 2009, 14(7) :1369-1377.]
- [10] Csurka G, Dance C R, Fan Lixin, et al. Visual categorization with bags of keypoints [C] // ECCV International Workshop on Statistical Learning in Computer Vision. Prague, Czech Republic ; ECCV Press, 2004 :1-22.