学号：15417133

常 州 大 学

毕业设计（论文）外文翻译

（2019 届）

外文题目 OATM:Occlusion Aware Template Matching by Consensus Set Maximization

译文题目 OATM：基于共识集最大化的遮挡识别模板匹配

外文出处 arXiv:1804.02638v1 [cs.CV] 8 Apr

学 生 朱海鹏

学 院 信息数理学院 专 业 班 级 自动化 151

校内指导教师 张继 专业技术职务 讲师

校外指导老师 专业技术职务

二〇一九年三月

OATM:基于共识集最大化的遮挡识别模板匹配

Simon Korman[1;3] Mark Milam[2] Stefano Soatto[1;3]

[1]Weizmann Institute of Science [2]Northrop Grumman [3]UCLA Vision Lab

摘要

我们最新提出了一种有效的模板匹配方法，该方法能够处理部分遮挡，并具有可证明的性能保证。该方法的一个关键部分是将 N 个高维向量中最邻近的搜索问题转换为两组阶$\sqrt{N}$向量之间的邻域所搜问题，利用范围搜索技术就可以有效地找到这一问题。这允许在搜索复杂度上进行二次改进，并使该方法在处理大搜索空间时具有可伸缩性。第二个贡献是基于共识集最大化可处理遮挡的哈希方案。最终所得到的方案可以看作是一个随机假设和测试算法，它保证了迭代次数，以获得一个高概率的最优解。对匹配率的预测验证得知，该算法在速度和鲁棒性两方面都明显有了提高。

1.引言

根据两者之间的关系，将模板 T（小图像）与目标 I（较大的图像）匹配几乎是不可能的。在经典设置中，当 I 是数字图像，T 是它的子集时，这相当于对 N 个离散 2D 平移集的搜索，其中 N 是图像 I 中的像素个数。当 T 和 I 是来自不同视点的同一场景的图像时，它们之间的关系可以用它们的区域的复杂变形来描述，这取决于底层场景的形状，以及它们的范围，也取决于它们的反射率和光照。对于足够小的模板，这种变形近似于区域的仿射变换（“翘曲”）和范围的仿射（“对比度”）变换，除遮挡外：模板的任意部分（包括所有模板）可能被遮挡，因此在目标图像中没有对应的部分。

这给许多低层次的任务带来了一个根本性的问题：要建立本地通信（共见性），模板应该是大的，这样才能区分。但随着面积的增加，目标图像中的对应区域被遮挡的概率增加，如果不明确考虑遮挡现象，就会导致匹配失败。

在本工作中，我们经过建模明确将遮挡作为鲁棒模板匹配的一部分，其中可见区域被假定为模板图像经过仿射变换和添加噪声的结果。我们以一种有效且可证明的方式寻找的转换就是最大限度的共识，即可见集合的大小收敛。

一种简化方法的效率来自于第一种贡献，将目标图像中 d 维模板 T 到 N 个版本的最近邻线性搜索转换为两组向量之间的搜索，每组向量的大小为 O($\sqrt{N}$)（Sect. 2.2）。这使得搜索复杂度从 O(N)降低到 O($\sqrt{N}$)，即使对于非常大的搜索空间(如仿射变换的离散空间)也是实用的。

为了使该方法有效，我们需要一种与遮挡性兼容的哈希方案，该方案是通过修改 Aiger 等人[2]的方案来实现的，这也使我们作出第二项贡献：与其在欧几里德ℓ_2范数下报告近邻，我们更感兴趣的是在最大(可见性)共识集上报告相容的向量对，直到一个阈

值。我们的哈希方案类似于ℓ_∞范数下的随机抽样一致(RANSAC-type)过程（Sect. 2.3）。

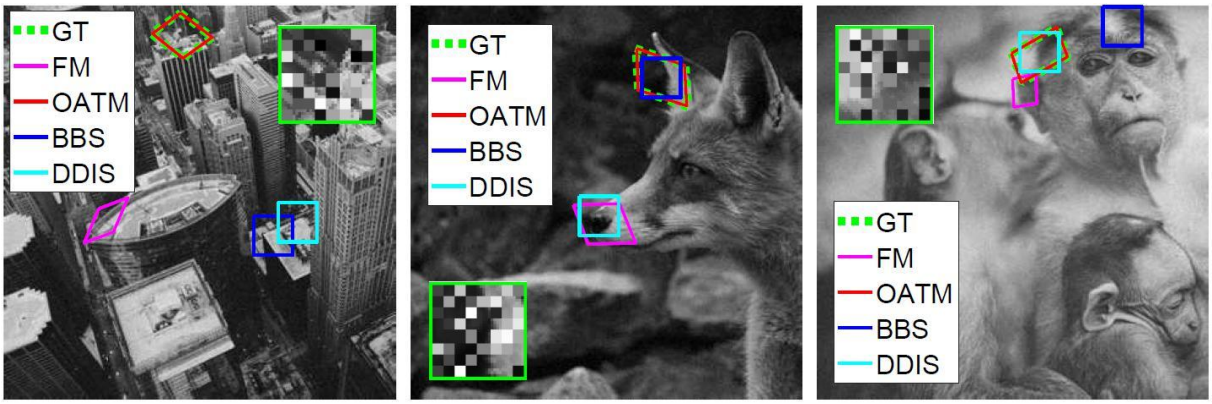


图 1.遮挡实验的实例(Sec. 4.2)在图像中搜索 60%被随机块遮挡的模板(用绿色覆盖)。在处理明显的变形和遮挡时，OATM 显示出最好的效果(缩放作为细节)。

最后，我们的第三个贡献是对算法的分析（Sect. 2.4）特别是关于在一定概率范围内获得最优解所需的候选假设数目的保证，即最大异常率意义上的最优解。

然而对于许多低水平的视觉任务速度，收敛性的保证并不是关键，但在某些应用中，能有授权也是非常重要的，例如卫星机动中的高保证视觉姿态估计。在我们的例子中，我们实现了速度和准确，同时能够处理遮挡，这允许使用更大的，更有区别的模板。

该算法具有一定的通用性，适用于区域空间的一般几何变换，同时给出了对二维平移群和二维仿射群的可能的显式分解。在实验部分，我们的算法在仿射模板匹配方面的效率和鲁棒性方面都优于传统的仿射模板匹配算法[17]。此外，它比一些现代图像描述的最近的 HPatches[4]基准有明显的优势。

1.1.相关工作

模板匹配算法的研究主要集中在效率上，这是视觉系统中低层构件的自然要求。这很大程度上是在有限的二维转换和ℓ_p相似度范围内实现的，在这种情况下，完全搜索等价算法将原本的全搜索方案加速了数量级[22]。与机器人导航和增强现实等实时应用不同，在某些应用中，精度和性能保证是非常重要的，例如对卫星或工业设备等高价值资产的高保证姿态估计。这就需要在多方面扩大研究范围。

其中一项工作中需要注意的是由于相机或物体的运动引起的几何变形。早期的工作，如[11 26]扩展滑动窗口方法，以处理旋转和缩放。快速匹配算法[17]被设计用于处理二维仿射变换。它使用分支和定界最小化绝对差之和，提供概率全局保证。[29]采用遗传算法对二维仿射空间进行采样。

为了实现光度不变性，文[13]提出了一种非线性声调映射下的快速匹配方案，而[10]则采用了广义拉普拉斯距离来处理多模态匹配。我们的方法可以提供仿射光度不变性，即，直到全局亮度和对比度的变化。

在本工作中，我们对这些方法的运行复杂度提出了二次改进，它与搜索空间的大小(即维数指数)成线性关系。最近，我们看到了在二维下进行匹配的尝试-使用深度神经网络[9 20]，尽管这些方法没有提供任何保证，而且就像前面提到的方法一样，它们不是设计用来处理部分遮挡的。

最近的两项工作可以通过矩形贴片之间的相似性度量来处理几何变形和部分遮挡：最佳邻近相似度(BBS)参考[8]，基于最大数目的相互最近邻像素对，和变形多样性相似性(DDIS)[25]，它检查最近的邻域之间的不同。DDIS 极大地提高了 BBS 运行时的复杂性，但由于它对大变形的处理有限，所以它的变形范围是有限的。此外，这些方法的缩放性限制了它们所能处理的遮挡范围。虽然 OATM 仅限于处理刚性转换，但可以证明它能够有效地处理高水平的变形和遮挡。

另一个相关并且非常活跃的研究领域是学习图像块的鉴别描述符(自然斑块或特征检测器提取的)，从早期的 SIFT[19]和变体[7 23]到最近的[24 5 12]。我们证明了 OATM 在大变形和遮挡下的匹配能力是优越的。

最后，讨论了计算机视觉的许多其他领域的遮挡处理问题，包括跟踪[31 30 15]，分割[28]，图像匹配[27]，多目标检测[6]，流[14]和识别[21]。

在“X 深度学习”研究的背景下，我们的工作与趋势相反：我们发现在匹配方面提供可证明的保证的需求(尽管与利基应用程序相关)服务不足，而数据驱动的机器学习工具并不是理想的适合这项任务的工具。

2.研究方法

2.1.问题描述

在模板匹配中，假设模板 T 和图像 I 通过域 $F = \{ f : \mathbb{R}^2 \to \mathbb{R}^2 \}$ 的几何变换和范围空间的光度变换相关联。目标是确定域的转换，尽管是范围的转换。这里我们假设 T 和 I 都是离散的、实值的、平方图像，因此我们可以写成 $T : \{1, \dots, n\}^2 \to \mathbb{R}$ (类似 $I: \{1, \dots, n\}^2 \to \mathbb{R}$)其中 T 和 I 分别是 n×n 和 m×m 图像。变换集 F 可以用一组离散的数据 N(可能很大)来逼近，直到达到期望的上限。例如，在标准的二维平移变换中，集合 F 包含在单个像素偏移处模板在图像上的所有可能位置，因此 $N = |F| \approx (m-n)^2$，容差为一个像素。此外，在我们的分析中，我们将假设最近邻插值(四舍五入)，这使我们可以简化讨论，使 $f : \{1, \dots, n\}^2 \to \mathbb{Z}^2$ 形式的完全离散变换。

我们用 $p \in T$ (同样的$p \in I$)表示模板域 $\{1, \dots, n\}^2$中的像素 p 和 T(P)表示它像素值。对于给定的变换 f，通过 $res_f(p) = \left|T(p) - I\left(f(p)\right)\right|$，定义了像素$p \in T$ 处的残差或误差。已知的“亮度调整约束”保证可以通过至少一个变换 f 使残差(在阈值内)变得很小。然而，它只适用于场景的某些部分，这些部分是在恒定光照下看到的，最重要的是：共同可见的部分。

我们现在已经准备好将遮挡感知模板匹配(OATM)作为一个共识集(CSM)问题，在这个问题中，我们搜索一个最大像素数目是可见的转换，即，与一个在 a 阈值内的残

差映射。

定义 1. [遮挡感知模板匹配(OATM)] 对于给定的错误阈值 t，查找以下提供的 $\boldsymbol{f}^*$转换：

$$\boldsymbol{f}^* = \boldsymbol{argmax}\sum_{\boldsymbol{p}\in\boldsymbol{T}}[\boldsymbol{res}_{\boldsymbol{f}}(\boldsymbol{p}) \leq \boldsymbol{t}] \ \textbf{(f} \in \textbf{F)} \qquad \textbf{(1)}$$

其中[·]表示指示函数。

我们对乘积空间的简化广泛地依赖于几何变换之间的距离概念(这取决于源域-模板 T)。

定义 2. [变换之间的距离Δ] 假设 $\boldsymbol{f_1}$，$\boldsymbol{f_2}$ ∈ **F**，我们定义距离：

$$\Delta(\boldsymbol{f_1}, \boldsymbol{f_2}) = \ \boldsymbol{max}_{\boldsymbol{p}\in\boldsymbol{T}}||\boldsymbol{f_1}(\boldsymbol{p}) - \ \boldsymbol{f_2}(\boldsymbol{p})||$$

其中‖ ‖表示图像 I 的(目标)域中的距离。

2.2. 乘积空间的约简

回顾(方程(1)，我们的目的是找到 f* 的一个最优变换，它的残差:

$$\boldsymbol{res}_{\boldsymbol{f}^*}(\boldsymbol{p}) = |\boldsymbol{T}(\boldsymbol{p}) - \boldsymbol{I}(\boldsymbol{f}^*(\boldsymbol{p}))| \qquad \textbf{(2)}$$

尽可能多的像素p ∈ T 处低于阈值 t。为了优化(方程 1)，我们需要比较 T 到 N 个可能的目标向量 I(f(T))(目标图像中所有可能的转换模板)。

这里的主要思想是以不同的方式枚举搜索空间。在源图像边缘，我们定义了由模板 T 的局部扰动获得的一组模板(向量)，而在目标图像边缘，我们定义了一组用于“封面”目标图像 I 的模板，即每个目标模板位置将接近于 V 中的一个。这样，如果模板的副本出现在图像中，则必须有一对类似的模板(向量) u ∈ U和v ∈ V。参见图片 2。

形式上，对于给定的公差 $\epsilon > 0$，设f ∈ F是使 $\Delta(f, f^*) < \epsilon$ 的变换。对于任意的 p′ ∈ T，如果我们假设存在 p ∈ T 使得 f(p) = $f^*(p')$，则通过在方程(2)中替换式子 $p' = f^{*^{-1}}(f(p))$，我们得到如下结果：

$$\boldsymbol{res}_{\boldsymbol{f}^*}(\boldsymbol{p}') = |\boldsymbol{T}\left(\boldsymbol{f}^{*^{-1}}(\boldsymbol{f}(\boldsymbol{p}))\right) - \boldsymbol{I}(\boldsymbol{f}(\boldsymbol{p}))| \qquad \textbf{(3)}$$

如果我们另 h = $f^{*^{-1}}$(f)，则上式可以写成：

$$\boldsymbol{res}_{\boldsymbol{f}^*}(\boldsymbol{p}') = |\boldsymbol{T}(\boldsymbol{h}(\boldsymbol{p})) - \boldsymbol{I}(\boldsymbol{f}(\boldsymbol{p}))| \qquad \textbf{(4)}$$

对于子模板中的像素 $T_h = \{p \in T : h(p) \in T\}$，其中h(p) = $p' \in T$。

关于 h,由于我们知道 $\Delta(f, f^*) < \epsilon$,容易得出 Δ(h, id) < ϵ/s($f^*$),其中 id 是恒等变换，s($f^*$)是$f^*$的最小尺度，由s(f) = $min_{p\in T}||f(p)||/||p||$。

如果我们调用 $\epsilon' = \epsilon/s(f^*)$，我们现在可以在函数空间 F 中定义函数的受限子集(它是半径为ϵ'的球，围绕恒等式)：

$$\boldsymbol{F}_{\boldsymbol{\epsilon}'} = \{\boldsymbol{h} \in \boldsymbol{F} : \ \Delta(\boldsymbol{h}, \boldsymbol{id}) < \boldsymbol{\epsilon}'\} \qquad \textbf{(5)}$$

设 $Net_\epsilon(F)$是空间 F 上关于距离Δ的任意ϵ网。即，对于任何f ∈ F，都存在一定的 $f' \in Net_\epsilon(F)$使得Δ(f, f') < ϵ。

结果表明，我对方程式 1 中f ∈ F的最优搜索进行了分解，在乘积空间$F_{\epsilon'} \times Net_\epsilon(F)$中寻找等价(h,f)的最优对。即，我们可以将 OATM 问题(等式(1))重新设置为：

$$\boldsymbol{f^* = argmax \sum_{p\in T} 1/|T_h|\,[|T(h(p)) - I(f(p))| \le t]} \quad \textbf{(f} \in \mathbf{F}_{\epsilon'} \quad \textbf{f} \in \boldsymbol{Net_\epsilon(F)}\textbf{)} \qquad \textbf{(6)}$$

为了描述和实现的简单性，我们可以使用由所有子模板的交集定义的$\{T_h\}_{h\in F}$，这将导致：

$$\boldsymbol{f^* = argmax \sum_{p\in T'} [|T(h(p)) - I(f(p))| \le t]} \quad \textbf{(f} \in \mathbf{F}_{\epsilon'} \quad \textbf{f} \in \boldsymbol{Net_\epsilon(F)}\textbf{)} \qquad \textbf{(7)}$$

看来，到目前为止，我们一无所获，因为在变换集Net_ϵ(F)的任何合理的条件下，它都认为|F| ≈ $|Net_\epsilon(F)| \cdot |F_{\epsilon'}|$，即搜索空间的大小保持不变。然而，这种分解使得我们可以为两组向量设计更好的方案(h(T')和 f(T')分别适用于$\{h(p)\}_{p\in T'}$和$\{f(p)\}_{p\in T'}$)。

$$\mathbf{U} = \{\boldsymbol{T(h(T'))}\}_{\boldsymbol{h}\in \boldsymbol{F_{\epsilon'}}} \qquad \textbf{(8)}$$

$$\mathbf{V} = \{\boldsymbol{I(f(T'))}\}_{\boldsymbol{f}\in \boldsymbol{Net_\epsilon(F)}} \qquad \textbf{(9)}$$

使对所有(h, f) ∈ $F_{\epsilon'} \times Net_\epsilon$(F)的$|T(h(p)) - I(f(p))|$项的有效搜索成为可能。

效率来源于以集 U 和 V 具有近似相等大小($\sqrt{N}$)的方式设计乘积空间，以及使用一种搜索算法，该算法的复杂性取决于空间大小之和($\sqrt{N}$)，而不是它们的乘积(N)。我们给出了二维平移和二维影射空间的显式表示。

2.3. 基于随机网格的搜索

我们已经将单个向量与 N 个目标向量之间的匹配问题转化为在两组目标向量之间寻找匹配向量的问题。在搜索文献中，高维点集之间的匹配是一个经典问题，显然与在一个单点集中寻找所有近邻的问题有关。我们的方法基于随机网格[1]——一个易于实现的算法，并在实践中得到了很好的应用[2]。

在[1]中，为了哈希 d 维点的集合，通过放置一个随机移动的均匀网格将空间划分为单元（每个单元都是一个边长 c 的平行立方体）。这些点相应地排列在哈希表中，然后检查在哈希表中共享条目的所有对点，报告那些距离低于指定阈值的点。然后，该过程被重复适当的次数，以保证，以很高的概率，所有或大多数对关闭点的报告。

与其他人的工作不同。[1，2]使用ℓ_2范数来度量向量之间的相似性，我们使用绝对差低于阈值的坐标数。此外，我们用随机选择少量的坐标(像素)代替了[2](一个 Johnson 变换)中的约简，以便在匹配下进行匹配。这些变化需要对算法进行不同的分析。有关我们的基本模块的总结，请参阅算法 1。

2.4. 分析

高保证模板的主要结果是保证算法 1 的成功概率。我们将使用以下术语：

$$\mathbf{P}(\boldsymbol{\alpha}, \mathbf{d}, \boldsymbol{\hat{d}}) = \frac{\binom{\boldsymbol{\alpha d}}{\boldsymbol{\hat{d}}}}{\binom{\boldsymbol{d}}{\boldsymbol{\hat{d}}}} = \frac{\boldsymbol{\alpha d \cdot (\alpha d - 1) \cdot \ldots \cdot (\alpha d - \hat{d} + 1)}}{\boldsymbol{d \cdot (d - 1) \cdot \ldots \cdot (d - \hat{d} + 1)}}$$

要求 1. [算法 1 分析] 算法 1 至少具有成功的概率：

$$\mathbf{P}\left(\boldsymbol{\alpha}, \mathbf{d}, \hat{\boldsymbol{d}}\right) \cdot \left(\mathbf{1} - \frac{\boldsymbol{t}}{\boldsymbol{c}}\right)^{\hat{\boldsymbol{d}}} \tag{10}$$

证明，推导是直截了当的，因为算法如果在哈希表中有一对最优匹配向量 u,v，考虑到两个事件的组合，碰撞肯定会发生。首先，$\hat{d}$抽样维度集是αd维度的子集的事件。这发生在概率$\mathrm{P}(\alpha, \mathrm{d}, \hat{d})$上，因为这是一个在 d 的总体中具有$\alpha d$成功项的几乎几何分布，$\hat{d}$样本都必须是成功项。第二，我们需要将碰撞发生的概率乘以网格偏移量中发生碰撞的概率。在这种情况下，$\hat{d}$维$\hat{u}$和$\hat{v}$在每个坐标上最多相差 t, 因此，由于偏移量在坐标之间是一致和独立的，所以$\hat{u}$和$\hat{v}$被映射到相同的单元格(因此在哈希表中)，并且概率至少是$(\frac{c-t}{c})^{\hat{d}} = (1 - \frac{t}{c})^{\hat{d}}$。

要求 2. [算法 1-更强版本的分析] 假设存在一对$\mathbf{u}, \mathbf{v} \in \mathbf{U} \times \mathbf{V}$，它们在其坐标的一个$\boldsymbol{\sigma}$分数处具有标准差$\boldsymbol{\alpha}$的零均值高斯噪声。这样算法 1 至少具有成功的概率：

$$\mathbf{P}\left(\boldsymbol{\alpha}, \mathbf{d}, \hat{\boldsymbol{d}}\right) \cdot \left(\int_0^{\boldsymbol{c}} \left(\mathbf{1} - \frac{\boldsymbol{x}}{\boldsymbol{c}}\right) \cdot \frac{\sqrt{\mathbf{2}}}{\boldsymbol{\sigma}\sqrt{\boldsymbol{\pi}}} \cdot \boldsymbol{exp}\left(-\frac{\boldsymbol{x}^{\mathbf{2}}}{\mathbf{2}\boldsymbol{\sigma}^{\mathbf{2}}}\right) \boldsymbol{dx}\right)^{\hat{\boldsymbol{d}}} \tag{11}$$

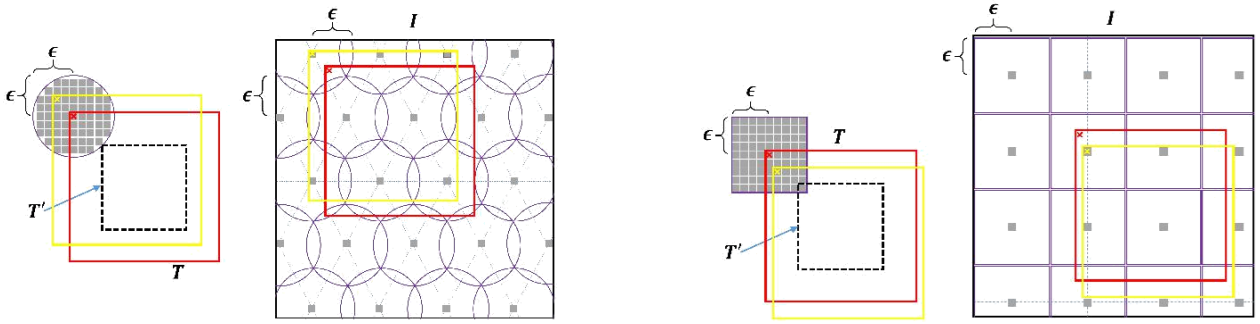
证明，与以前的说法相比，这里唯一的区别是坐标向量落入单个单元的概率。不同之处在于，在这里，我们不仅假定每个坐标上 t 的绝对值相差最大，而且我们还提出了一个更强(但更现实)的假设，即坐标系下的向量仅由已知标准差的高斯噪声而不同。在这种情况下，每个坐标的绝对差遵循折叠式高斯分布(例如[18])，因此我们在[0，c]范围内对可能的绝对差 x 进行积分。

算法 1：向量集中的共识集。

输入：在$\mathbb{R}^d$中设置向量 U 和 V；阈值 t；
输出：最大共识集的向量对 $(\mathrm{u}, \mathrm{v}) \in \mathrm{U} \times \mathrm{V}$
参数：样品维数$\hat{d}$；胞维数 c；

---

1. 从 1,…,d 中选取$\hat{d}$随机维数。
2. 设$\hat{U}$和$\hat{V}$是向量集 U 和 V 约化为$\hat{d}$随机维数。
3. 在[0，c]$^{\hat{d}}$中生成随机$\hat{d}$维偏移向量 o.
4. 根据$\mathrm{Map}(\hat{v}) = [(\hat{v} + o)/c]$，将$\hat{U}$和$\hat{V}$中的每个向量映射为$\hat{d}$维整数。
5. 使用从$\mathbb{N}^{\hat{d}}$到{1,…,|U|}的任何哈希函数将得到的结果安排到哈希表中。
6. 扫描哈希表，其中对于共享哈希值的每一对向量$\hat{u}$和$\hat{v}$，计算$\mathrm{i} \in \{1, \ldots, \mathrm{d}\}$中的坐标数(其中$|\mathrm{u(i)} - \mathrm{v(i)}| \leq \mathrm{t}$)
7. 用最大发现率返回一对 u,v



（a）$\epsilon-$net解释　　（b）简单解释

图 2.说明两种可能的二维变换。在(A)和(B)中的每一组中，采样向量(模板)U(来自 T)和 V(来自 I)的集合由灰度像素表示，灰度像素是采样模板的左上角位置。如果模板（红色）出现在目标图像中，则在 U 和 V（黄色显示）中会有相应的一对匹配样本。该参数使得两边的样本数(灰度平方数)大致相等（都近似为$\sqrt{N}$）。

2.5. 遮挡感知模板匹配

给定算法 1 及其性能保证，我们现在可以指定完整的模板匹配算法。模板将运行算法 1 一定次数，并返回目标位置在图像中，这对应于找到的整体最佳向量对。作为提醒，算法 1 返回一对$\{T(h(p))\}_{p\in T'}$和$\{I(f(p))\}_{p\in T'}$形式的向量，表示这一对变换(h,f)作为候选解，从中提取出单个变换$f^*=f\circ h^{-1}$。

直接评价速率$P^*=1/|T|\sum_{p\in T}[|T(p)-I(f^*(p))|\le t]$的原因有两个，而不是代理$1/|T'|\sum_{p\in T'}[|T(h(p))-I(f(p))|\le t]$。一种方法是通过直接应用级联变换$f^*=f\circ h^{-1}$来避免内插错误。第二和更重要的一个是所检测的速率仅反映 T 的子模板中的T′的像素。

算法 2: OATM：遮挡感知模板匹配。

输入：模板图像 T 和源图像 I；阈值 t；变换族 F(大小为 N)；
输出：具有最大共识的 f ∈ F (方程 1)

---

1. 将 F 转化为乘积$F_{\epsilon'}\times Net_\epsilon(F)$，选择一个ϵ 使得$|F_{\epsilon'}|\approx|Net_\epsilon(\mathrm{F})|\approx\sqrt{N}$
2. 构造向量集 U 和 V。(方程(8)-(9))。
3. 重复 k 次算法 1 的(用相同的 U，V 和 t)，以获得转换值$\{f_i\}_{i=1}^k$。
4. 返回具有最大共识集的转换f_i(方程 1)。

算法 2 总结了遮挡感知模板匹配（OATM）。它由 k 迭代的运行算法 1 组成。如果我们用根据权重要求 2 的等式(11)给出的算法 1 的成功概率表示算法 1 的成功概率，则算法 2 的成功概率至少为：

$$\mathbf{1}-(\mathbf{1}-\boldsymbol{P_\alpha})^{\boldsymbol{k}} \qquad \textbf{(12)}$$

反过来说，若要以预先设定的概率(例如，0.99)获得成功，所需的 k 的数目是：$\log(1-p_0)/\log(1-P_\alpha)$。

必须指出的是，可以根据前几轮的调查结果确定 k 的数量。正如管道中常见的情况一样，每次更新最佳最大共识性(速率)时，所需的重复次数都会相应地减少。

注意，该算法相对于下面的变换空间 F 是通用的，但是它不需要关于如何将其有效分解为产品空间的知识(步骤 1)。接下来，我们描述了二维变换的两个这样的结构，并在补充材料[16]中给出了二维仿射组的构造。

2.6. 二维变换结构

回想一下，在我们算法的基础上，将搜索空间 F 分解成有参数ϵ控制的空间$F_{\epsilon'} \times Net_{\epsilon}(F)$的乘积。根据空间 F(|F| = N)的结构，我们将选取一个ϵ(和ϵ')使$|F_{\epsilon'}| \approx |Net_{\epsilon}(\mathrm{F})| \approx \sqrt{N}$，以最小化依赖于乘积空间大小之和的复杂性。在二维变换的情况下，我们进行了明确的分解。

由于不涉及标度，$s(f^*) = 1$，因此$\epsilon' = \epsilon$。给出了 m×m 和 n×n 维的方形模板 T 和图像 I，如图 2(a)所示，可以用半径ϵ的圆的正方形盖来构造对应的F_{ϵ}和$Net_{\epsilon}(F)$。生成的子空间的大小F_{ϵ}和$Net_{\epsilon}(F)$:$\pi\epsilon^2$以及$(n-m+1)^2/(1.5\sqrt{3}\epsilon^2)$,可以通过调整$\epsilon$使其相等。

然而，由于圆的重叠，该覆盖是由于$1.5\sqrt{3}$的乘法因子而次优的。我们实际上可以得到一个实际的最优分解(而不是严格遵循ϵ-net 定义)，如图 2(b)所示。我们取集合的乘积：$F_{\epsilon} = \{i, j: i, j \in -\epsilon, \ldots, \epsilon\}$ 和 $Net_{\epsilon}(F) = \{i, j: i, j \in \{\epsilon + 2k\epsilon\}$对于 $\mathrm{k} = 1, \ldots, [(\mathrm{n} - \mathrm{m} + 1)/2\epsilon]\}$，其结果是$|F_{\epsilon}| = 4\epsilon^2$和$|Net_{\epsilon}(\mathrm{F})| = (\mathrm{n} - \mathrm{m} + 1)^2/(4\epsilon^2)$，取$\epsilon = 0.5\sqrt{n-m+1}$得$|F_{\epsilon}| = |Net_{\epsilon}(F)| = n - m + 1$。

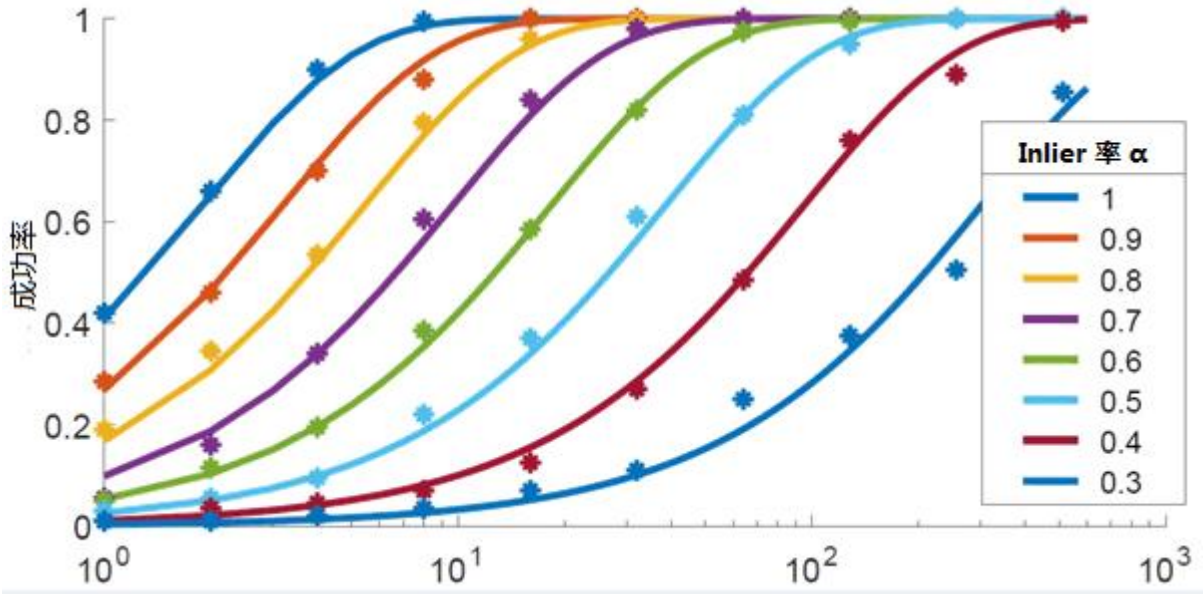


图 3.算法 2 保证的经验验证。对于不同速率下的α(注意对数标度 x 轴)，通过理论上的一致性检验，可以看出与大规模实验中测量的算法成功率(标记值)相匹配的算法成功率(标记值)是由不同速率下的 k(固体曲线)的个数所决定的。

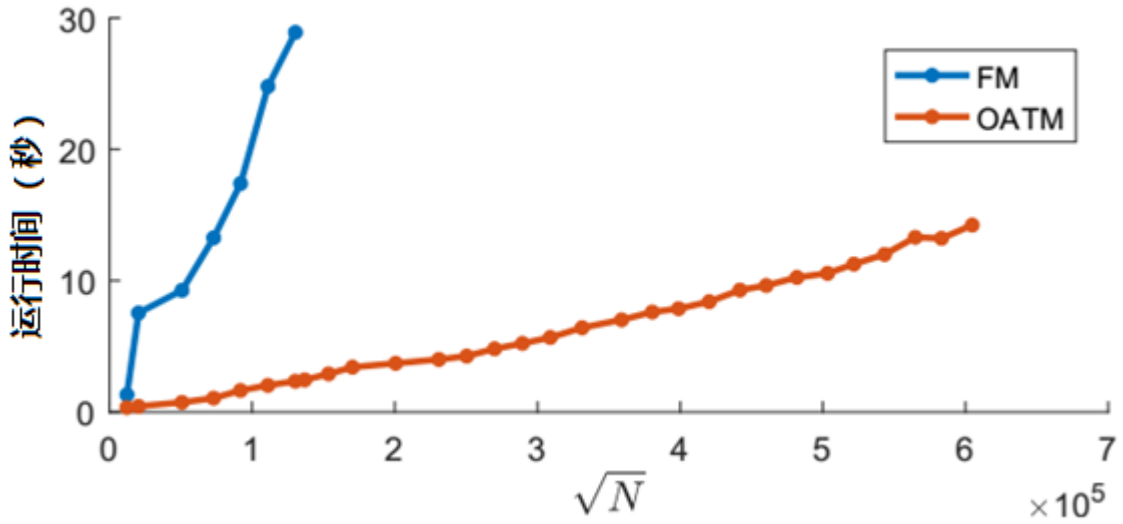


图 4.可扩展性实验。OATM 与 FM 相比较，在大小为 N 的二维近似搜索空间中，如预期：OATM 的运行时间已$\sqrt{N}$增长，FM 的则以 N 线性增长。

3. 分析的实证验证

算法成功率(二维变换)

我们首先对该算法的理论保证进行大规模验证（为二维变换情况所示），在集合{1,2,4,8,16,32,64,128,256,512}以及其他参数不变的情况下，每个 k 值都保持不变。

我们对每种比率的α进行了 200 次模板匹配试验{0.3,0.4,0.5,0.6,0.7,0.8,0.9,1}，报告的成功率是与之完全匹配的试验的相对数量。对于每一次试验，我们创建了一个模板匹配实例，首先从 500×500 图像 I 中提取一个 100×100 模板 T，然后从数据中随机获取(缩放)。将模板像素的随机α −分数标记为等距像素，并将每个像素 p 的强度 T(p)替换为与其相差 0.5 的强度。这一设置保证了最终的评估率完全是α，而算法只有在对一组纯数据进行采样的情况下才会对其进行调整。最后，我们将图像 I 白色高斯噪声与标注等效为 5 灰度。

结果如图 3 所示，每个α(标记)的经验成功率可以从方程(12)(实体曲线)中看出与理论成功率相匹配。必须指出的是，无论模板和图像内容如何，这些都是确保找到完美匹配的最低成功率，而在实践中，我们经常观察到更好的匹配。

算法可伸缩性(二维仿射)

在这个实验中(如图 4 所示)，我们验证了我们算法的运行时间O($\sqrt{N}$)。这样做的一个简单方法是创建一系列匹配实例(技术细节见 4.1 节中的实验)，在集合中具有不同边长的方形图像中搜索 32 像素的方形模板{100,200,300,…,3200}同时将其他的搜索限制固定在范围内[2/3；3/2]和范围内[−π/4；π/4]。这导致排列尺寸为 N 的序列，其生长为四方形(因此标记在$\sqrt{N}$ x-轴大致线性地分布)。可见，随着$\sqrt{N}$ 的线性增长，可在合理的时间

内处理模板维数与图像维数的比值高达 100 的问题。作为参考，表示在模板匹配中快速匹配(FM)算法[17]的复杂性取决于二维仿射空间(其大小在 N-中线性增长)的标准(详见[17])。可以看出，它不能处理模板-图像边长比超过 20。

4. 结果

在这一部分中，我们通过对实际数据的几个受控和不受控制的实验，说明了该算法的优点。

实施细节 在我们的实现中使用的参数是通过简单的坐标下降在一小部分随机合成实例上选择的(如 4.1 所描述)。对于随机网格，我们使用样本维数$\hat{d} = 9$；单元维数 c = 2.5t；其中我们取阈值 $t = 2\sigma\sqrt{2/\pi}$ (是折叠正态分布×均值的两倍)，给出σ的噪声水平，或者当它未知时 t = 10。该方法通过向量集 U 和 V(在算法 2 的步骤 2 中)获得模板的均值和标准差，从而提供更好的一致性，即全局亮度和对比度变化。

4.1. 模板匹配评价

我们在一个标准的模板匹配评估中测试我们的算法，而不需要再匹配，以便与其他算法进行比较，例如快速匹配(FM)[17]，它代表了当前最先进的模板匹配技术。我们使用模板和图像大小的不同组合进行了大规模的比较(两者之间的差距越大，搜索空间的大小 N 就越大)。模板和图像尺寸采用如下方法：T1 为16×16，T2 为32×32，T3 为64×64。同样的，I1 为160×160，I2 为320×320，I3 为640×640。

对于每个模板-图像大小组合，我们进行了 100 个随机模板匹配试验。每个试验(遵循[17])都涉及到选择一个随机图像(此处，从该数据集)和一个随机转换(在图像中)。在图像中加入 5 种标准的高斯白噪声，生成模板。

对于每一次试验，我们都报告平均重叠错误和重复错误。重叠误差是[0，1]中的一个重叠误差，由 1 减去检测到的目标与真目标的交集和合并的比率。

结果摘要见表 1。OATM 在类似的低误差水平上，通常比 FM 快一个数量级。FM 不能处理设置 T1-I3，由于大量的配置 N(图像边缘长度是模板边缘长度的 40 倍)，而处理更多的$\sqrt{N}$大小。

		模板图像尺寸								
		T1-I1	T1-I2	T1-I3	T2-I1	T2-I2	T2-I3	T3-I1	T3-I2	T3-I3
FM	错误率	0.09	0.13	NA	0.05	0.05	0.09	0.02	0.01	0.03
	时间	12.22	25.37	NA	4.35	7.78	32.07	1.33	1.90	11.61
OATM	错误率	0.07	0.10	0.13	0.02	0.04	0.04	0.01	0.02	0.13
	时间	0.15	0.18	0.39	0.53	0.76	1.73	0.51	0.64	1.01

表 1.用于不同模板图像大小的模板匹配评估，包括平均运行时间(秒)和重叠误差。

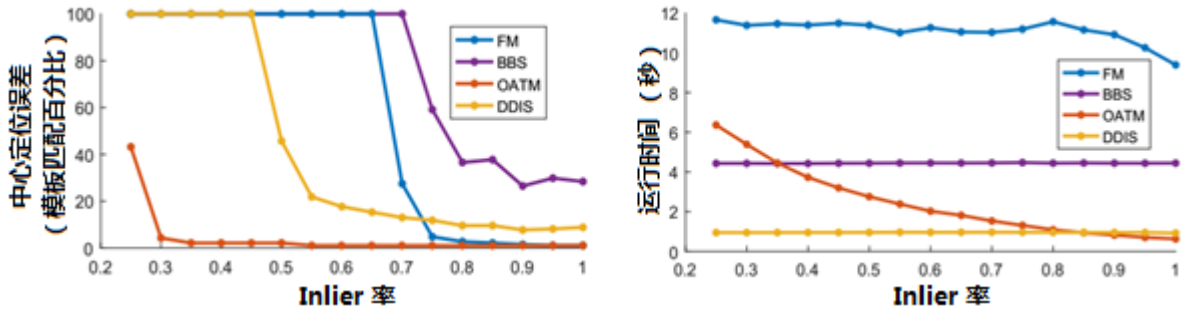


图 5.遮挡实验结果(Sec 4.2)：中间位置误差(左)和平均运行时间(右)。

4.2. 对遮挡的鲁棒性

在这个实验中，我们评估了如何处理遮挡和其他几种方法。我们重复了之前的实验(4.1 节)的协议，除了我们采取一个固定的模板-图像大小(T2-I2)，我们综合引入了一个受控数量的像素。其中一种方法(参见图 1 中的示例)是引入随机 4×4 块。我们用另外两种引入遮挡的方法重复了实验，得到了类似的结果，这是我们在补充材料[16]中提供的。这些结果表明，该方法对遮挡掩码的空间排列具有较强的鲁棒性。

除了快速匹配（FM）[17]外，我们还与另外两个模板匹配方法-最佳伙伴相似度（BBS）[8]和可变形多样性相似度（DDIS）[25]进行比较，这两种方法都专门处理复杂的几何变形和高水平的遮挡。为了进行公平的比较，由于 BBS 和 DDB 以滑动窗口的方式与模板匹配(并且在窗口内考虑变形)，我们测量中心位置误差(而不是重叠误差)-目标窗口的中心和真实目标中心位置之间的距离，作为模板尺寸的百分比(在 100％处被修剪)。

图 5 中的曲线总结了实验。可以看到 OATM 在非常宽的 Inlier 速率范围内，从大约 0.25 开始提供最精确的检测。DDIS 可以处理 0.5 以上的自动识别率，但由于其滑动窗口搜索，定位精度稍低。FM 的设计并不是为了显式地处理转帐，但在 0.75 以下的速率下却没有这样做。在这种情况下，不能处理 0.75 以下的再加工率，其定位是处理再加工时的再加工。

就速度而言，DDIS 显然是最有效的。DDIS 和 BBS 对于 Inlier 速度是不可知的，然而，由于 OATM 具有相似的自适应停止准则，其运行时间与 Inlier 速率成反比。

4.3. 遮挡变形匹配的改善

在本实验中，我们使用了最近的[4]数据集，该数据集是为现代局部图像设计的。从 116 条序列(59 条改变视点，57 条改变光照)中提取斑块，每条序列包含 6 幅平面场景图像，由二维给出已知的几何对应关系。从每个序列中的第一图像中提取大约 1300 平方 65×65 个参考贴片(已纠正的最新的仿射检测区域)。然后利用地面真实投影从其他 5 幅序列图像中提取相应的贴片，同时引入 3 级(简单、难、困难)的受控几何扰动(旋转、各向异性缩放和平移)来模拟当前特征检测器的位置。

对于简单/难/困难，这些方法引入了显著的几何变形(例如旋转 10 ℃20 ℃30 ℃)，并增加

了遮挡程度(平均重叠 78% / 63% / 51%)。图 6 显示了在不同难度下提取的参考块及其匹配块的几个示例。

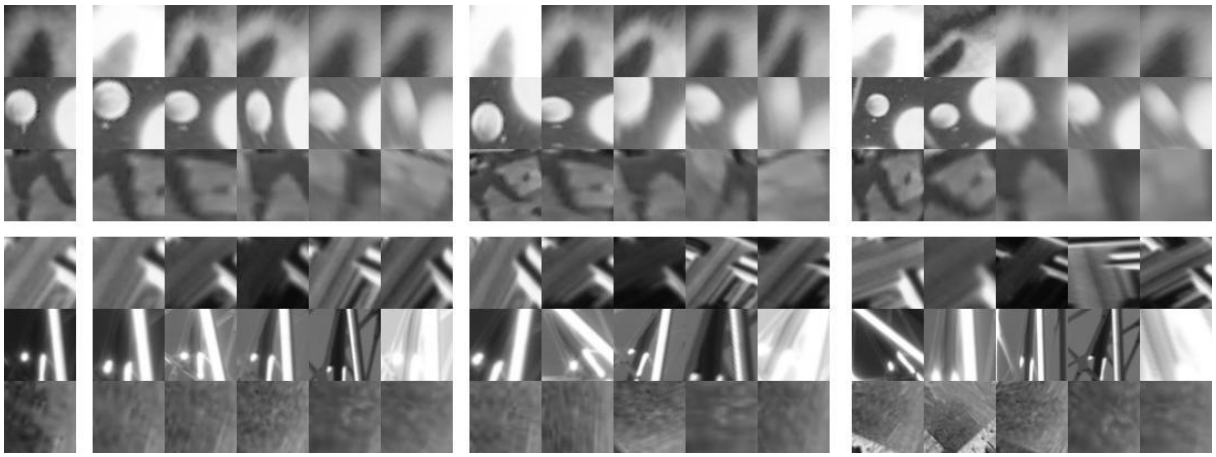


图 6.来自 HPatches[4]数据集的样本。视点序列(第 1-3 行)和照明序列(第 4-6 行)。

这些数据对于显示我们的方法在处理这些挑战方面的能力是有用的，与通过它们匹配特性的一般做法相比是非常有用的。我们关注于提出的“匹配”任务[4]，其中每个参考块需要定位在每个序列图像的每个块之间。模板匹配算法不能严格遵循所建议的任务协议，而任务协议是为匹配补丁定义的。相反，我们将所有(~1300)正方形目标块打包到一个图像中，在该图像中，我们使用 OATM 的光度不变性搜索模板。所选择的目标块是包含该模板块的中心位置的块。对于平均精度(MAP)计算，由于我们的方法只生成一个目标块，所以我们为检测到的目标块分配 1 的权重，对其余的目标块分配 0 的权重。

结果摘要见表 2。参考描述符方法包括[19]及其变换式[3]、二元简写法[7]和[23]法和深二次相似法[24]和重合比法*(TF-R)[5]。对于 SIFT、TF-R、DDESC，给出了较优的和归一化的变型的结果(如[4]中所示)。

	视点集			照明系统		
方法	简单	难	困难	简单	难	困难
BRIEF [7]	25.6	6.9	2.4	20.5	5.9	2.0
ORB [23]	36.4	11.1	3.7	28.9	8.8	3.2
SIFT [19]	59.4	30.6	15.3	52.6	26.1	13.3
TF-R [5]	58.9	35.5	19.0	48.5	28.6	15.6
DDESC [24]	58.6	36.0	20.2	50.7	30.0	17.0
RSIFT [3]	64.0	35.2	18.5	**57.1**	**30.2**	15.9
OATM	**72.7**	**49.2**	**32.1**	43.3	29.3	**19.7**

表 2.在 HPatches[4]图像匹配基准上的结果。结果以平均精度(MAP)为标准，除 OATM 外的所有结果均在[4]中所示。

显然，对于视点序列和光照序列-相对于 OATM 平均精度的方法，更多的是随着几何变形和遮挡程度的增加而增加。虽然技术特征和可能高度接近某些局部几何和不同的变化(因此，在容易照明的情况下，有些优于 OATM)，但它们在处理显著变形和遮挡方面不如 OATM 那么有效，OATM 明确地探讨了仿射变形的空间和遮挡程度的原因。

此外，目前对 OATM 的这一数据的应用表明，可以通过以下方式进一步提高性能：(i) 寻找目标位置的分布，而不是一次检测；(ii) 注意目标图像的块结构；(iii) 使用高级表示而不是相应的描述。尽管如此，与基于描述符的方法不同，OATM 的模板匹配特性肯定不适合大规模匹配，在这种情况下，需要将大量块池与另一个块进行匹配。然而，这里提出的许多想法可能会被修改，例如图像到图像的匹配设置。

5. 结论

本文提出了一种高效的二维仿射模板匹配算法，并对其进行了详细的分析，表明该算法在处理高水平的遮挡和几何变形方面有一定的改进。

Hpatch 数据集的结果提出了基于描述符的匹配是否能够处理由特征检测器引入的定位噪声中固有的几何变形和高遮挡水平的问题。即使在深度学习的出现中，也是这样的情况，并且能够明确地产生变形和遮挡的方法的开发对于改善视觉通信中的现有技术来说似乎是必要的。

鸣谢

收到 ARO W911NF-15-1-0564/66731-CS, ONR N00014-17-1-2072 的技术支持和 Northrop Grumman 的礼物。

参考文献

[1] D. Aiger, H. Kaplan, and M. Sharir. Reporting neighbors in high-dimensional euclidean space. In Proceedings of the Twenty-Fourth Annual ACM-SIAM Symposium on Discrete Algorithms, pages 784–803, 2013. 4

[2] D. Aiger, E. Kokiopoulou, and E. Rivlin. Random grids: Fast approximate nearest neighbors and range searching for image search. In Proceedings of the IEEE International Con-ference on Computer Vision, pages 3471–3478, 2013. 1, 4

[3] R. Arandjelovic ´and A. Zisserman. Three things everyone should know to improve object retrieval. In Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on, pages 2911–2918. IEEE, 2012. 8

[4] V. Balntas, K. Lenc, A. Vedaldi, and K. Mikolajczyk. Hpatches: A benchmark and evaluation of handcrafted and learned local descriptors. In CVPR, 2017. 2, 8

[5] V. Balntas, E. Riba, D. Ponsa, and K. Mikolajczyk. Learning local feature descriptors with triplets and shallow convolu-tional neural networks. In BMVC, volume 1, 2016. 2, 8

[6] P. Baque, F. Fleuret, and P. Fua. Deep occlusion reasoning for multi-camera multi-target detection. In The IEEE Inter-national Conference on Computer Vision (ICCV), Oct 2017. 2

[7] M. Calonder, V. Lepetit, C. Strecha, and P. Fua. Brief: Binary robust independent elementary features. Computer Vision– ECCV 2010, pages 778–792, 2010. 2, 8

[8] T. Dekel, S. Oron, M. Rubinstein, S. Avidan, and W. T. Free-man. Best-buddies similarity for robust template matching. In 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pages 2021–2029. IEEE, 2015. 2, 7

[9] D. DeTone, T. Malisiewicz, and A. Rabinovich. Deep image homography estimation. arXiv preprint arXiv:1606.03798, 2016. 2

[10] E. Elboer, M. Werman, and Y. Hel-Or. The generalized lapla-cian distance and its applications for visual matching. In Pro-ceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pages 2315–2322, 2013. 2

[11] K. Fredriksson. Rotation Invariant Template Matching. PhD thesis, University of Helsinki, 2001. 2

[12] X. Han, T. Leung, Y. Jia, R. Sukthankar, and A. C. Berg. Matchnet: Unifying feature and metric learning for patch-based matching. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pages 3279– 3286, 2015. 2

[13] Y. Hel-Or, H. Hel-Or, and E. David. Matching by tone mapping: Photometric invariant template matching. IEEE transactions on pattern analysis and machine intelligence, 36(2):317–330, 2014. 2

[14]J. Hur and S. Roth. Mirrorflow: Exploiting symmetries in joint optical flow and occlusion estimation. In The IEEE International Conference on Computer Vision (ICCV), Oct 2017. 2

[15]N. Joshi, S. Avidan, W. Matusik, and D. J. Kriegman. Syn-thetic aperture tracking: tracking through occlusions. In 2007 IEEE 11th International Conference on Computer Vi-sion, pages 1–8. IEEE, 2007. 2

[16]S. Korman. Occlusion aware templae matching webpage. http://www.eng.tau.ac.il/~simonk/OATM. 5, 6, 7

[17]S. Korman, D. Reichman, G. Tsur, and S. Avidan. Fast-match: Fast affine template matching. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recogni-tion, pages 2331–2338, 2013. 2, 6, 7

[18]F. Leone, L. Nelson, and R. Nottingham. The folded normal distribution. Technometrics, 3(4):543–550, 1961. 5

[19]D. G. Lowe. Object recognition from local scale-invariant features. In Computer vision, 1999. The proceedings of the seventh IEEE international conference on, volume 2, pages 1150–1157. Ieee, 1999. 2, 8

[20]T. Nguyen, S. W. Chen, S. S. Shivakumar, C. J. Taylor, and V. Kumar. Unsupervised deep homography: A fast and robust homography estimation model. arXiv preprint arXiv:1709.03966, 2017. 2

[21]E. Osherov and M. Lindenbaum. Increasing cnn robustness to occlusions by reducing filter support. In The IEEE Inter-national Conference on Computer Vision (ICCV), 2017. 2

[22]W. Ouyang, F. Tombari, S. Mattoccia, L. Di Stefano, and W.-K.Cham. Performance evaluation of full search equivalent pattern matching algorithms. IEEE transactions on pattern analysis and machine intelligence, 34(1):127–143, 2012. 2

[23]E. Rublee, V. Rabaud, K. Konolige, and G. Bradski. Orb: An efficient alternative to sift or surf. In Computer Vi-sion (ICCV), 2011 IEEE international conference on, pages 2564–2571. IEEE, 2011. 2, 8

[24]E. Simo-Serra, E. Trulls, L. Ferraz, I. Kokkinos, P. Fua, and F. Moreno-Noguer. Discriminative learning of deep convolu-tional feature point descriptors. In Proceedings of the IEEE International Conference on Computer Vision, pages 118– 126, 2015. 2, 8

[25]I. Talmi, R. Mechrez, and L. Zelnik-Manor. Template matching with deformable diversity similarity. In The IEEE Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, 2017. 2, 7

[26]D. Tsai and C. Chiang. Rotation-invariant pattern matching using wavelet decomposition. Pattern Recognition Letters, 23(1):191–201, 2002. 2

[27]Y. Yang, Z. Lu, and G. Sundaramoorthi. Coarse-to-fine re-gion selection and matching. In 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pages 5051–5059. IEEE, 2015. 2

[28]Y. Yang, G. Sundaramoorthi, and S. Soatto. Self-occlusions and disocclusions in causal video object segmentation. In Proceedings of the IEEE International Conference on Computer Vision, pages 4408–4416, 2015. 2

[29]C. Zhang and T. Akashi. Fast affine template matching over galois field. In BMVC, pages 121–1, 2015. 2

[30]T. Zhang, K. Jia, C. Xu, Y. Ma, and N. Ahuja. Partial occlu-sion handling for visual tracking via robust part matching. In 2014 IEEE Conference on Computer Vision and Pattern Recognition, pages 1258–1265. IEEE, 2014. 2

[31]T. Zhang, S. Liu, N. Ahuja, M.-H. Yang, and B. Ghanem. Robust visual tracking via consistent low-rank sparse learn-ing. International Journal of Computer Vision, 111(2):171– 190, 2015. 2