基于关键帧检测的视频相似度检验及查重的数学模型

摘要　　本文针对视频相似度及查重的检验方法，建立了基于关键帧检测的数学模型，并结合巴氏系数、SURF 算法、SIFT 算法，给出了检测视频相似度和检验片段是否抄袭的解决方案。并且客观准确地对两种算法进行了比较。

针对问题(1)，基于自定义的关键帧，使用巴氏系数来进行关键帧的提取，并将提取出来的关键帧作为评判视频是否抄袭的基准。

针对问题(2)，基于 SURF 算法，计算 Hessian 矩阵、构建尺度空间、精确定位特征点、确定主方向并生成特征点描述子来匹配特征点来测量图像相似度。

针对问题(3)，基于 SIFT 算法，检测尺度空间极值、定位关键点并描述关键点来匹配特征点来测量图像相似度。

最后，本文对附录中的三组视频基于上述模型进行集中检测，并给出了正确的图片形式的结论。

关键词　　视频查重；　视频相似度；　巴氏系数；　SURF 算法；　SIFT 算法

一、问题的重述

视频抄袭或盗用是指给定若干查询的视频片段，在视频数据库中进行查找，检测在数据库中是否存在相应的视频片段与查询视频片段内容相同；如果存在，查询视频片段就被称为视频抄袭或盗用片段。视频直接或经过编辑后被人盗用会造成侵权及商业纠纷，多数情况下会被编辑。编辑方法基本包括：缩放、剪切、镜像、画中画、改变亮度、增加字幕、增加噪声等。随着数字视频的迅猛发展，当今社会对视频版权保护产生了迫切的需求。请利用相关知识，根据所给视频片段，建立如下模型：

1、自定义一种关键帧，从源视频（每组中的视频1）中提取该关键帧，并解释该关键帧方法在后续检测视频是否被抄袭算法中所起的作用。

2、基于该关键帧提取的结果，试提出相应的数学模型，检测经过亮度处理（如每组中的视频2）、增加字幕（每组中的视频3）的编辑方法后视频中是否有相似的关键帧，并以此为依据推断源视频是否被盗用。

3、试提出数学模型，检测增加噪声后的视频（每组中的视频4）中是否有相似的关键帧，并以此为依据推断源视频是否被盗用。

4、综合运用并改进上述基于关键帧检测的数学模型，从所给三组数据集中将被盗用的视频片段检测出来，分析并评估所使用模型的性能。

二、问题的分析

对于问题(1)，要求自定义一种关键帧并说明在之后算法中的作用。

首先明确关键帧的概念：在视频领域，电影、电视、数字视频等可视为随时间连续变换的许多张画面，帧则指这些画面当中的每一张。在动画软件的时间轴上帧表现为一格或一个标记。角色或者物体运动或变化中的关键动作所处的那一帧即为关键帧。为便于处理，转换为8位的灰度图像后，每一帧的所有像素点分布在256种灰度之中。

自定义一种关键帧，即先以第一帧为基准帧，从第二帧开始的每一帧都与第一帧做某种对比，在这里我们引入图像相似度的概念，以及巴氏距离(巴塔恰里雅距离 / Bhattacharya Distance，用于测量两离散概率分布的可分离性来量化两图相似度的度量)。图像相似度计算主要用于对于两幅图像之间内容的相似程度进行打分，根据分数的高低来判断图像内容的相近程度。利用直方图匹配的算法，基于简单的数学上的向量之间的差异来进行图像相似程度的度量，即分别计算视频中前后两帧的直方图，Hist1，Hist2，然后计算两个直方图的巴氏距离，直方图相交距离，若两帧之间相似度较大则选取下一帧来与基准帧作对比，直到某一帧与基准帧的相似度小于某一阈值 k ，则定义该帧为一关键帧。再将之后的每一帧按上述方法与其之前已确定的关键帧作图像相似度的对比，继续找出后续关键帧，依次类推直至视频的末尾。

检测视频是否被抄袭时，只需对比两视频中关键帧的图像相似度即可。在后续算法中，将待检测视频的关键帧全部提取出来，提取的时候适当改变阈值 k 来使其与原始视频关键帧的数量大致相同，以便进行之后对两视频的关键帧一一对应的比较：若相似度大于既定阈值n，即认定为了雷同；若关键帧的相似度均值大于阈值n，即可认定为抄袭。由于关键帧与其前后的非关键帧差距不大，其可代表一段片断，这大大减少了查重时的工作量。

对于问题(2)，要求对经过亮度处理及增加字幕的视频查重。

根据对附件中所给三组视频的1、2、3进行分析，亮度处理的关键帧与源视频的关键帧相比直方图中的数据普遍增大或减小，直方图的几何曲线进行水平移动，其中以波峰移动的距离为例便于计算。对于亮度处理基于 SURF 算法计算两关键帧的匹配特征点的个数来比较两组关键帧的相似度，若关键帧匹配值均值大于既定阈值 n 即可认定为抄袭；同理，对于增加字幕的视频也同样分析其与源视频的函数型，由于添加了字幕，两关键帧匹配特征点通常会适当减少，故比较两组关键帧的相似度时适当降低度量阈值 n 即可。

SURF 算法首先要构建 Hessian 矩阵，Hessian 矩阵是 SURF 算法的核心，之后构建尺度空间，精确定位特征点，丢弃所有取值小于预设极值的点，并确定主方向，生成特征点描述子。

对于问题(3)，要求对增加噪声后的视频进行查重。

根据对附件中所给三组视频的1、4进行分析，由于噪声干扰，使得直接比较两视频关键帧相似度的可行度降低，并且在处理关键帧的时候可以观察到显著的画中画的现象，即长宽尺寸和纵横比与源视频不同，故本题基于 SIFT 算法对增加噪声的视频与源视频的关键帧进行相似度比较。

首先对视频4进行尺度空间极值检测：搜索所有尺度上的图像位置，通过二维高斯滤波函数来识别潜在的对于尺度不变和对噪声稳定的兴趣点。进而进行关键点定位，在每个候选的位置上，通过一个拟合精细的模型来确定位置和尺度，关键点的选择依据于它们的稳定程度。所有后面的对图像数据的操作都相对于关键点的尺度和位置进行变换，从而提供对于这些变换的不变性。关键点描述在每个关键点周围的邻域内，在选定的尺度上测量图像局部的梯度。这些梯度被变换成一种表示，这种表示允许比较大的局部形状的变形。进行两关键帧基于 SIFT 算法的特征点的比较，检测是否为抄袭。

对于问题(4)，基于上述模型对三组视频进行检测，评估并改进上述基于关键帧检测的模型。

利用 MATLAB 编写代码对各组视频分别进行关键帧的提取及图像相似度的比较，检测是否为抄袭并将被盗用的片段检测出来，并在实施过程中寻找改进方法。

三、模型的假设与符号说明

1、模型的假设

(1) 假设源视频的所有帧都不完全相同。

(2) 假设对视频的编辑方式为题中所给编辑方式。

(3) 假设字幕颜色控制在一定范围之内。

2、符号说明

BC 表示 Bhattacharya 系数，度量巴氏距离;

i 表示灰度值 $(i=0,1,2\cdots 255)$;

a_i 表示源视频第一帧中灰度值为 i 的像素点个数;

b_i 表示源视频第二帧中灰度值为 i 的像素点个数;

k 表示检测是否为抄袭的相似度的度量界限;

n 表示检测是否为抄袭的匹配值的度量界限;

L 表示尺度空间;

σ 为尺度空间因子;

其余部分符号的含义在使用时具体给出。

四、模型的建立与求解

4.1 问题(1)的数学模型建立及求解

依据问题(1)的分析，读取视频时已知 a_i $(i=0,1,2\cdots 255)$，b_i $(i=0,1,2\cdots 255)$，先对源视频的前两帧求 BC，即

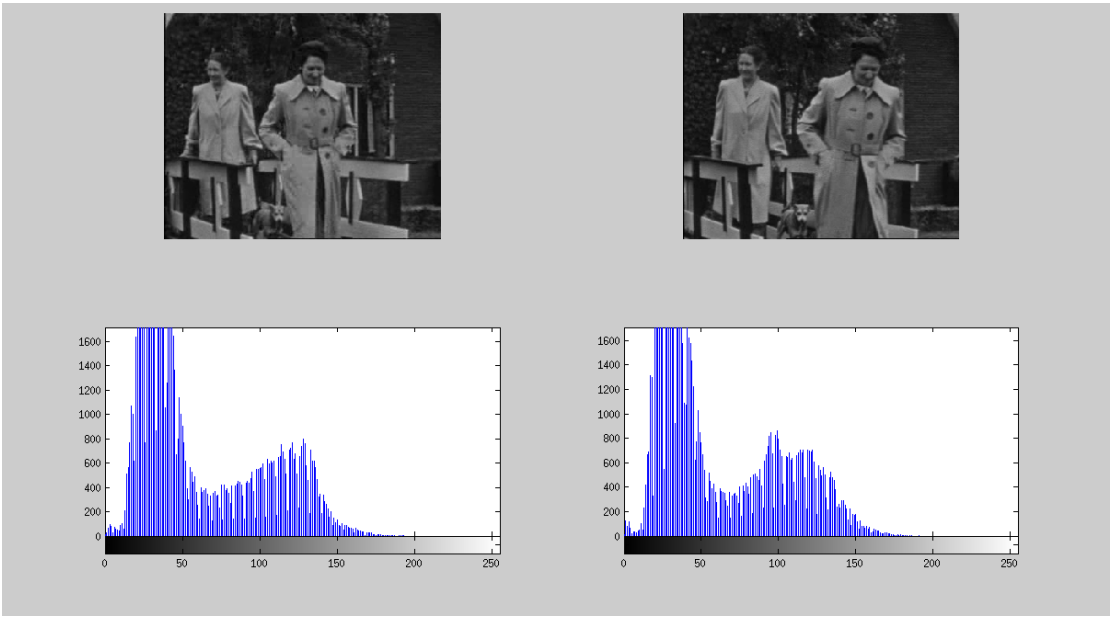
$$BC=1-\sqrt{1-\frac{\sum_{i=1}^{256}\sqrt{a_i\bullet b_i}}{\sum_{i=1}^{256}\frac{a_i+b_i}{2}}}\quad (0\le BC\le 1)$$

为源视频的前两帧的相似度，再与阈值 k 进行比较（k =0.975）。

若 BC 大于 k，则第二帧不是关键帧，需要选取下一帧与第一帧（基准帧）进行图像相似度的对比，直至选取的帧与基准帧的 BC 小于 k。

若 BC 小于 k，则第二帧为关键帧，根据对问题(1)的分析可知，用第二帧替换第一帧的位置作为基准，取后一帧与第二帧进行比较，重复上述操作，直至最后一帧。

基于上述模型，利用 MATLAB 对第二组中视频1取关键帧可得结果如下：



其中上图为连续两关键帧，下图为对应关键帧的灰度直方图。

在 Linux 系统运行下所取关键帧数目为41，将这组关键帧作为后续检测是否抄袭的基准。

4.2 问题(2)的 SURF 算法模型建立及求解

依据对问题(2)的分析，如图1、图2所示：

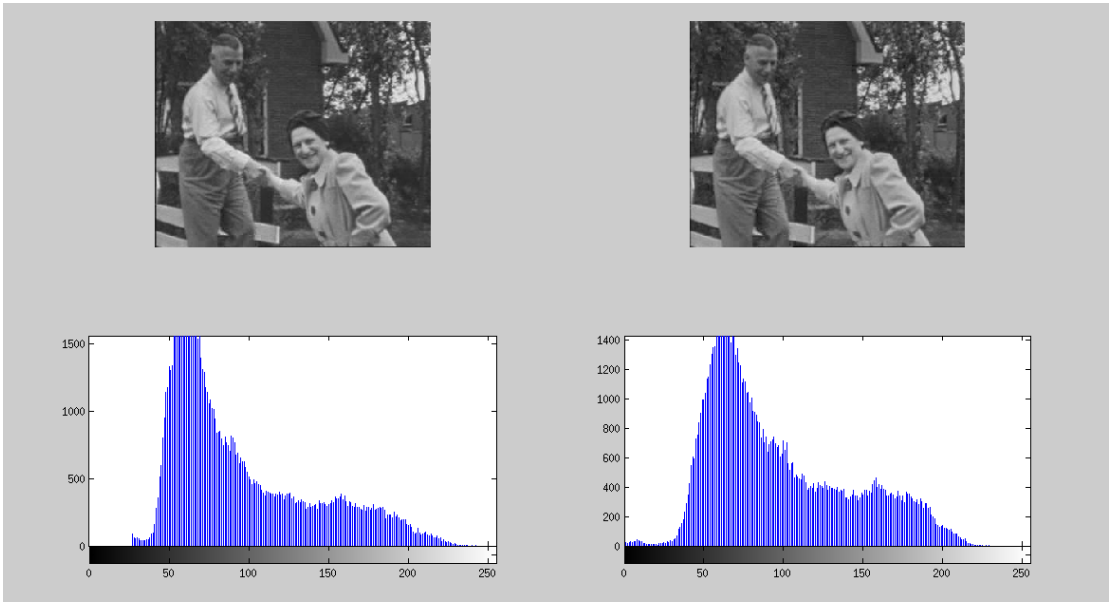


图1

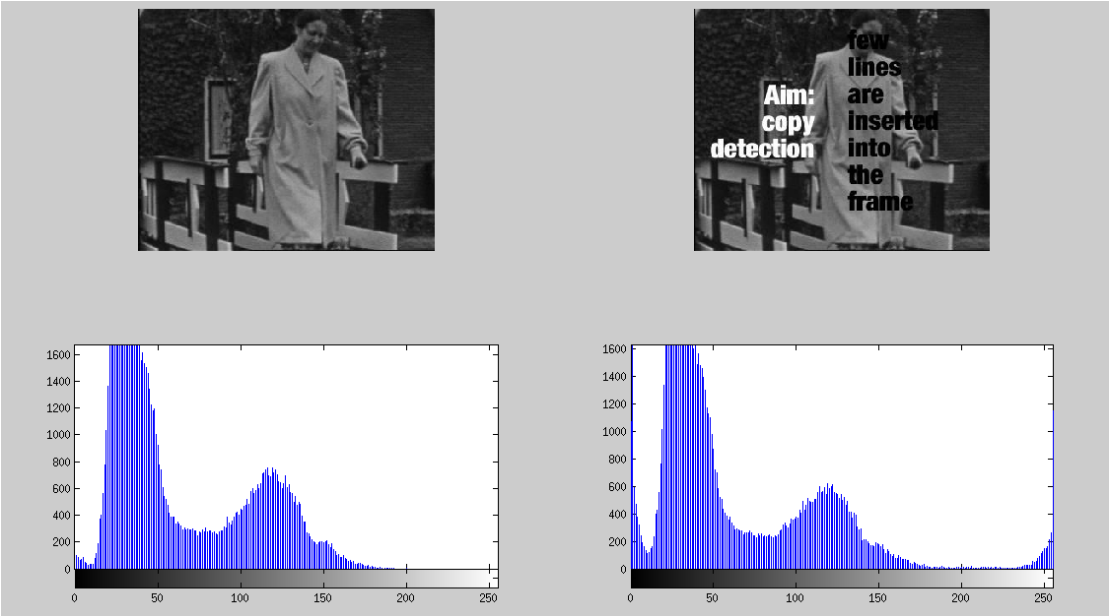


图2

图1为亮度处理与源视频的关键帧相比较，图2为增加字幕与源视频的关键帧相比较。从图1的灰度直方图可以看出，由于亮度的变化，波峰明显向右平移；图2由于增加了字幕，在灰度值小于16和大于220的像素点明显增多。

SURF 算法的步骤：

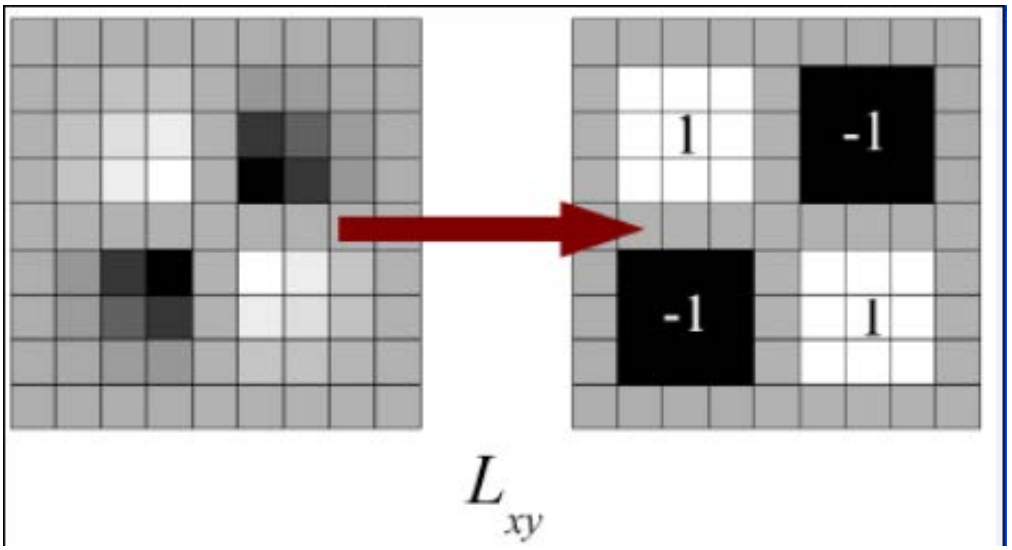
利用 Hessian 矩阵，计算特征值 σ：

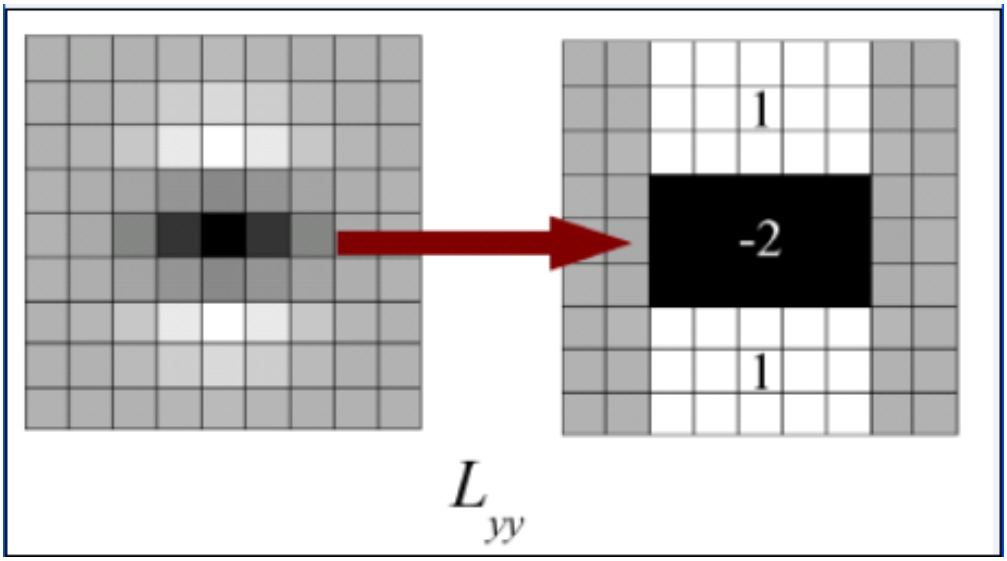
$$H(x,\sigma)=\begin{bmatrix} L_{xx}(x,\sigma) & L_{xy}(x,\sigma) \\ L_{xy}(x,\sigma) & L_{yy}(x,\sigma) \end{bmatrix}$$

其中 $L_{xx}(x,\sigma)$ 是高斯滤波后图像 $g(\sigma)$ 的在 x 方向的二阶导数，其他的 $L_{xy}(x,\sigma)$ $L_{yy}(x,\sigma)$ 都是 $g(\sigma)$ 的二阶导数。之后需求原图像的一个变换图像，因为要在这个变换图像上寻找特征点，然后将其位置反映射到原图中，原图每个像素的 Hessian 矩阵行列式的近似值构成的。其行列式近似公式如下：

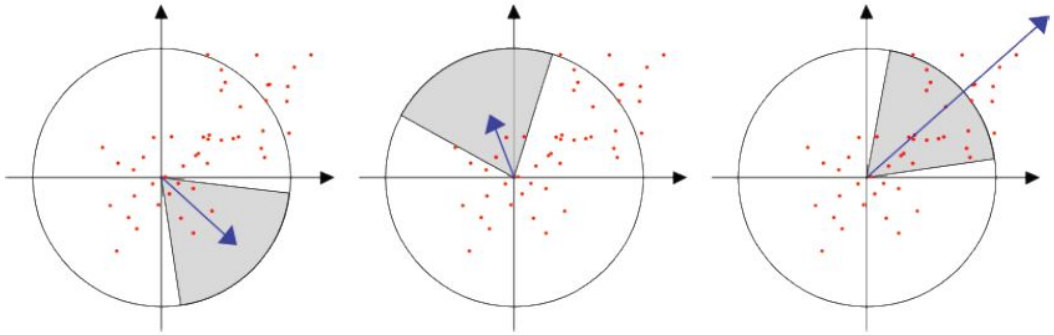
$$\det(H_{approx})=D_{xx}D_{yy}-(0.9D_{xy})^2$$

为加快速度，SURF 在计算过程中采用如图所示的箱式滤波器与积分图像来简化计算，在模糊的基础上将原本的模块近似，利用了积分图像的优势大大减少了计算量。



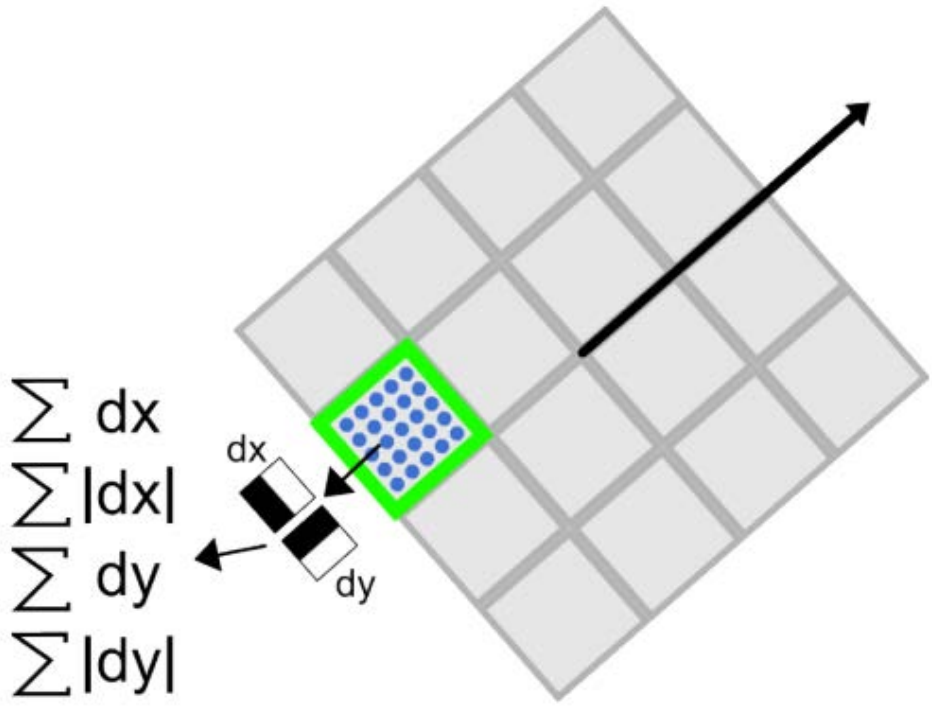


利用非极大值抑制初步确定特征点，将经过 Hessian 矩阵处理过的每个像素点与其 3 维领域的 26 个点进行大小比较，如果它是这 26 个点中的最大值或者最小值，则保留下来，当做初步的特征点。采用 3 维线性插值法得到亚像素级的特征点来精确定位极值点，同时也去掉那些值小于一定阈值的点。之后选取特征点的主方向，统计特征点领域内的 harr 小波特征。即在特征点的领域（比如说，半径为 6s 的圆内，s 为该点所在的尺度）内，统计 60 度扇形内所有点的水平 haar 小波特征和垂直 haar 小波特征总和，haar 小波的尺寸变长为 4s，这样一个扇形得到了一个值。然后 60 度扇形以一定间隔进行旋转，最后将最大值那个扇形的方向作为该特征点的主方向。该过程的示意图如下：



最后构造 SURF 特征点描述算子在 SURF 中，也是在特征点周围取一个正方形框，框的边长为 20s（s 是所检测到该特征点所在的尺度）。该框有方向，其方向即第 4 步检测出来的主方向。然后把该框分成 16 个子区域，每个子区域统计 25 个像素的水平方向和垂直方向的 haar 小波特征，这里的水平和垂直方向都是相对主方向而言的。该 haar 小波特征为水平方向值之和，水平方向绝对值之和，垂直方向之和，垂直方向绝对值之和。

该过程的示意图如下所示：



这样每个小区域就有4个值，所以每个特征点就是16×4=64维的向量。

图3

图4

图5

如图3、4、5所示为在 MATLAB 环境下运行的第二组视频1、2的关键帧（左为1、右为2），左图为视频1中的某一关键帧，右图为连续的三张视频2中的关键

帧。测定水平的横线数目，若两图特征点匹配数大于120则认定为雷同，若小于120则不是。例如图4有若干特征点一一对应的，且匹配数大于120，则认定为雷同；而图3图5尽管相似，但由于匹配数小于120，则认定为不同。

图6

图7

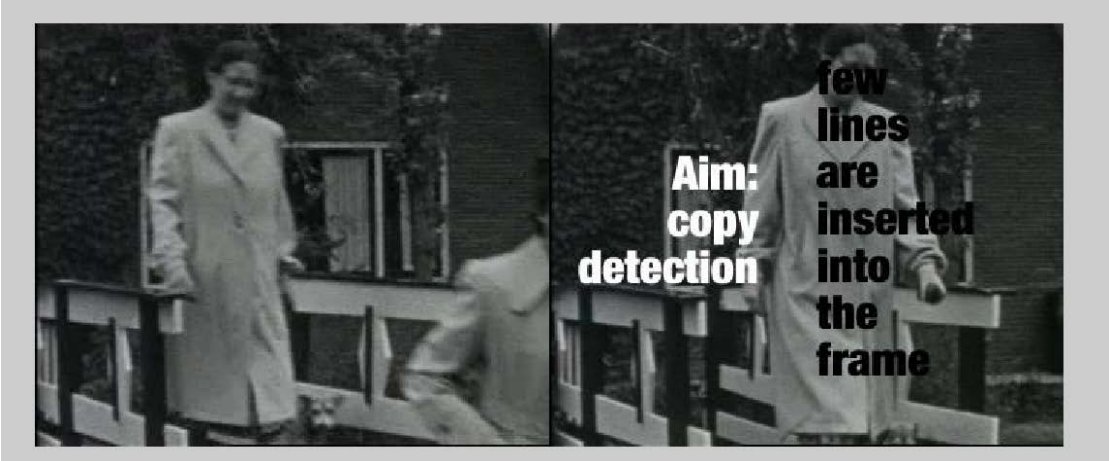


图8

如图6、7、8所示为在 MATLAB 环境下运行的第二组视频1、3的关键帧（左为1、右为3），左图为视频1中的某一关键帧，右图为连续的三张视频3中的关键帧。测定水平的横线数目，由于字幕的影响，应适当下调阈值n，即若两图特征点匹配数大于80则认定为雷同，若小于80则不是。例如图7有若干特征点一一对应的，且匹配数大于80，则认定为雷同；而图6图8尽管相似，但由于匹配数小于80，则认定为不同。

4.3 问题(3)的 SIFT 算法模型建立及求解

依据对问题(3)的分析，根据二维高斯滤波函数，对关键帧进行高斯滤波：

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

一幅图像$I(x\ ,\ y)$，在不同尺度空间下的表示可以由图像与高斯核卷积得到 Gaussian 图像：

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

其中：$(x\ ,\ y)$ 为图像 I 上的点，L 表示尺度空间，σ 为尺度空间因子。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。σ 值越小表示图像被平滑得越大，即分辨率越高。

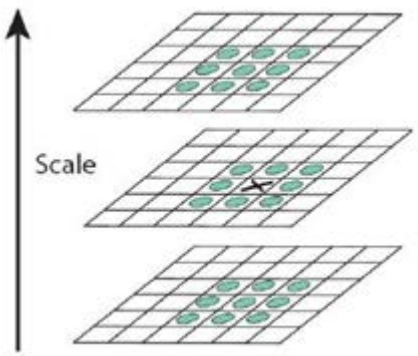
根据尺度函数来建立高斯金字塔，高斯金字塔的第一阶的第一层是原始图像。高斯金字塔有O阶、S层，在同一阶上的两个相邻层之间的尺度比例为m。

在高斯金字塔的基础上，利用同一阶上的两个相邻的两层的尺度空间函数之差得到 DOG 高斯金字塔的一层。

DOG 的表达式定义为：

$$D(x,y,\sigma)=(G(x,y,m\sigma)-G(x,y,\sigma))*I(x,y)=L(x,y,m\sigma)-L(x,y,\sigma)$$

为了检测到 DOG 空间的局部极值点，每一个采样点要和它所有的相邻点比较。如图所示，中间的检测点需要与它同层的8个，上层和下层各9个像素点进行比较，以确保在尺度空间和二维图像控件都检测到极值点。如果该检测点为最大值或者最小值，则该点为图像在该尺



度下的一个候选关键点。

在极值比较的过程中，每一组图像的首末两层是无法进行极值比较的，为了满足尺度变化的连续性，我们在每一组图像的顶层继续用高斯模糊生成了3幅图像，高斯金字塔有每组S+3层图像。DOG 金字塔每组有S+2层图像。

关键点的选取要经过两步：①它必须去除低对比度和对噪声敏感的候选关键点；②去除边缘点。

(1) 去除低对比度的点

对局部极值点进行三维二次函数拟合以精确确定特征点的位置和尺度，尺度空间函数 $D(x,y,\sigma)$ 的泰勒展开式如公式所示：

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

令上式对x的偏导数等于0，可得极限点位置：

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

代入泰勒展开式公式中，可得：

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

若 $\left|D(\hat{x})\right|\geq 0.03$，该特征点就保留下来，否则丢弃。

(2) 去除边缘点

一个定义不好的高斯差分算子的极值在横跨边缘的地方有较大的主曲率，而在垂直边缘的方向有较小的主曲率。主曲率由 Hessian 矩阵求出：

$$H=\begin{bmatrix}D_{xx} & D_{xy}\\ D_{yx} & D_{yy}\end{bmatrix}$$

D的主曲率和H的特征值成正比，令 α 为最大特征值，β 为最小特征值，则：

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

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

令 $\alpha=\alpha\ \beta$，则：

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

如果曲率小于$(r+1)^2/r$，保留该特征点，否则舍弃。

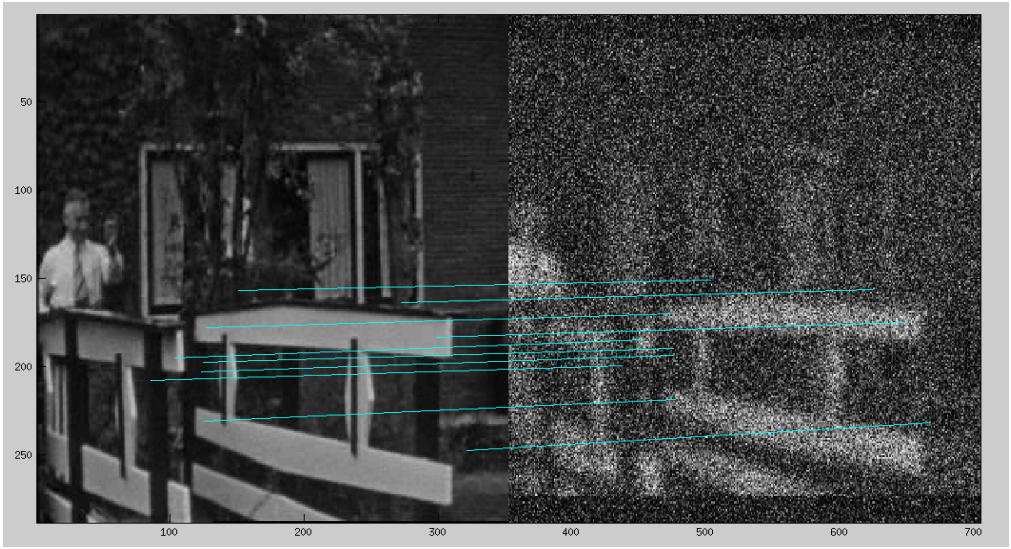

50
100
150
200
250
100
200
300
400
500
600
700

图9

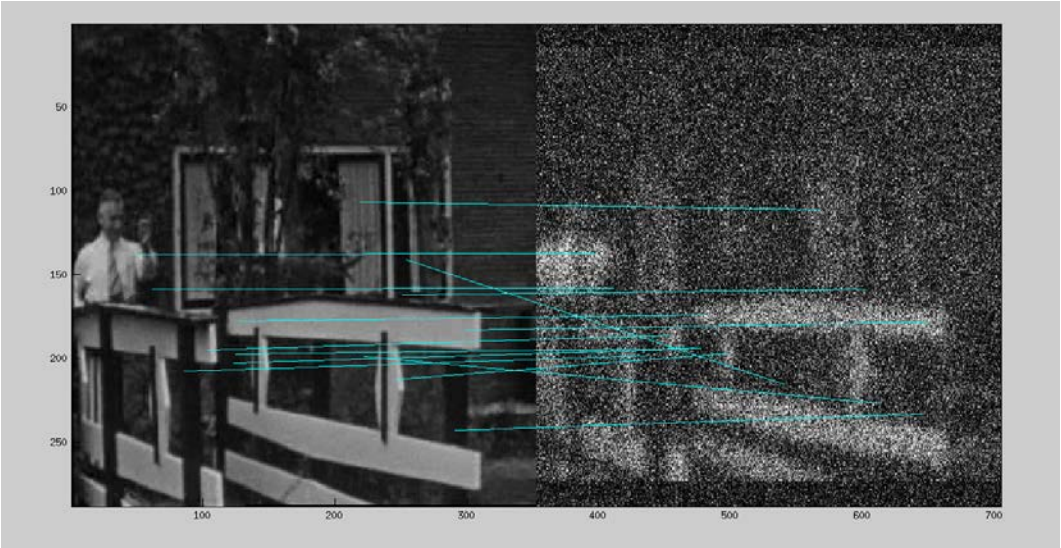

50
100
150
200
250
100
200
300
400
500
600
700

图10

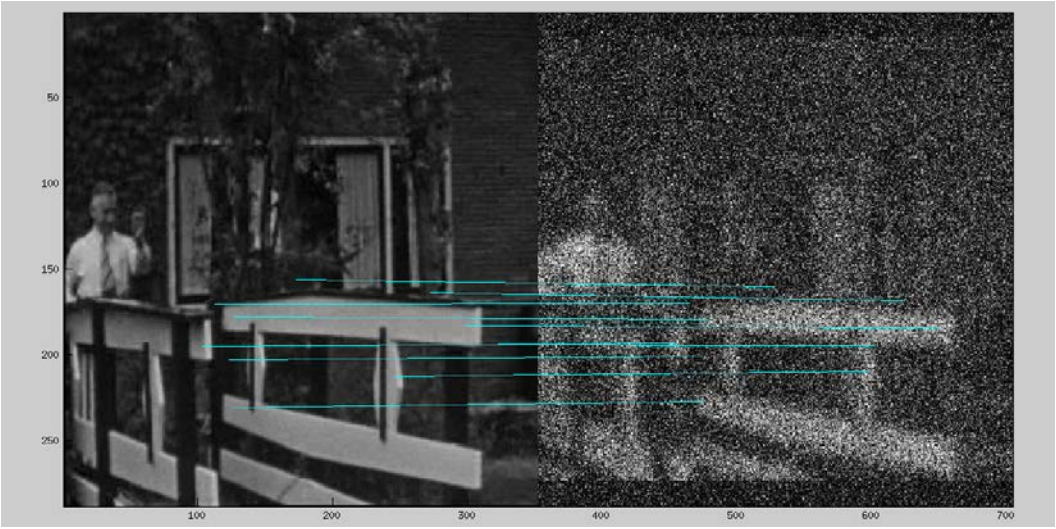

50
100
150
200
250
100
200
300
400
500
600
700

图11

如图9、10、11所示为在 MATLAB 环境下运行的第二组视频1、4的关键帧（左为1、右为4），左图为视频1中的某一关键帧，右图为连续的三张视频4中的关键帧。测定水平的横线数目，若两图特征点匹配数大于10则认定为雷同，若小于10则不是。例如图10有若干特征点一一对应，且匹配数大于10，则认定为雷同；而图9图11尽管相似，但由于匹配数小于10，则认定为不同。

4.4 综合运用基于关键帧检测的模型检测盗用片段

依据对问题(4)的分析，基于上述模型，在 MATLAB 环境下对第一组、第三组做相同操作检测盗用片段。

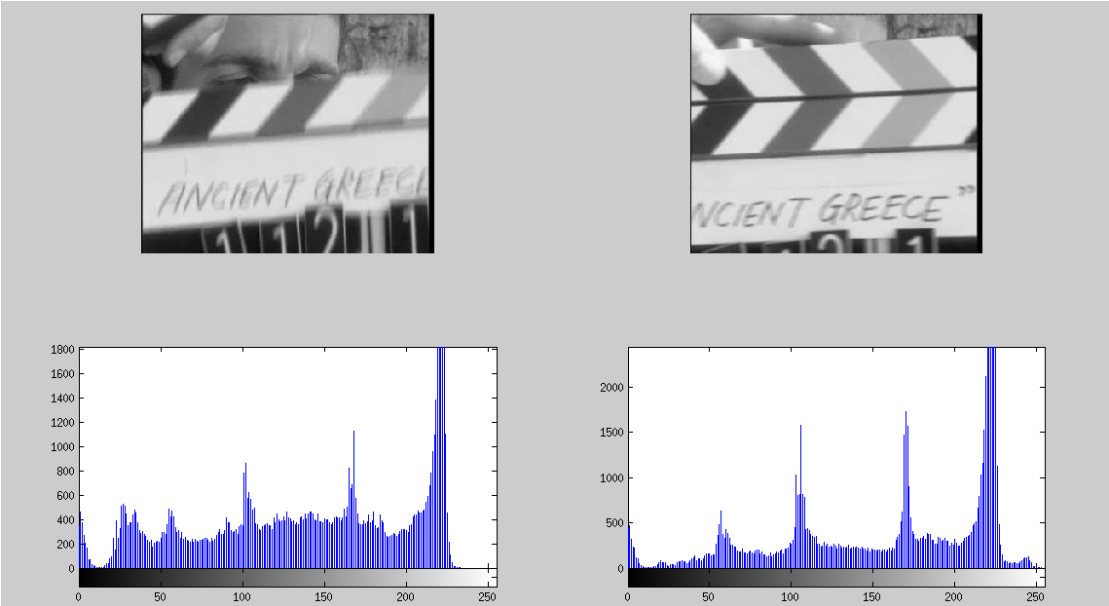


图12

图13

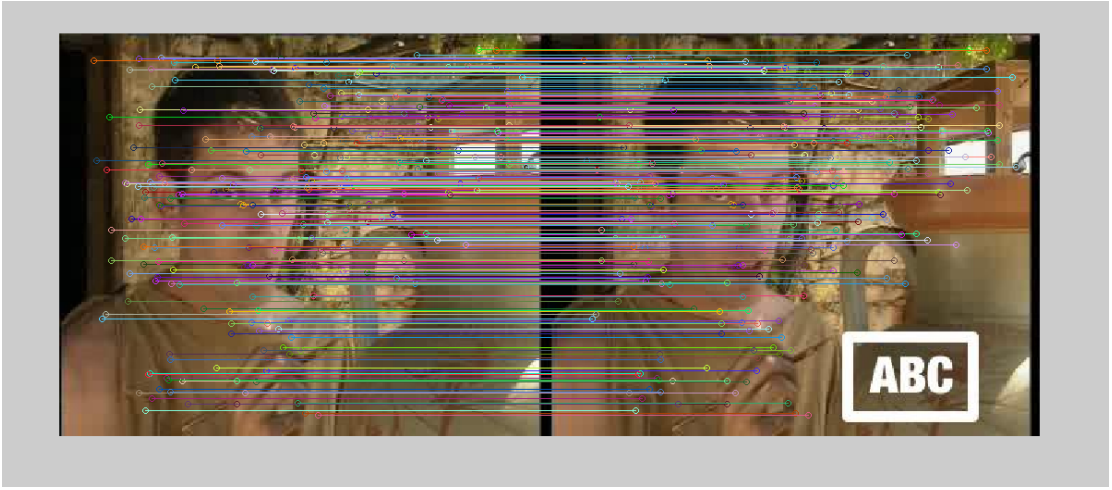

ABC

图14

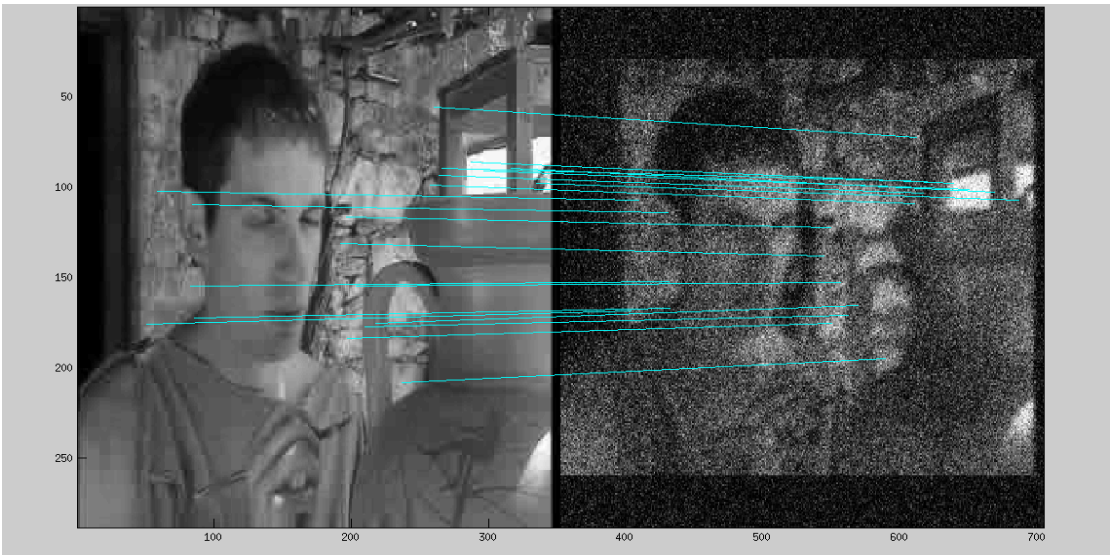

50
100
150
200
250
100
200
300
400
500
600
700

图15

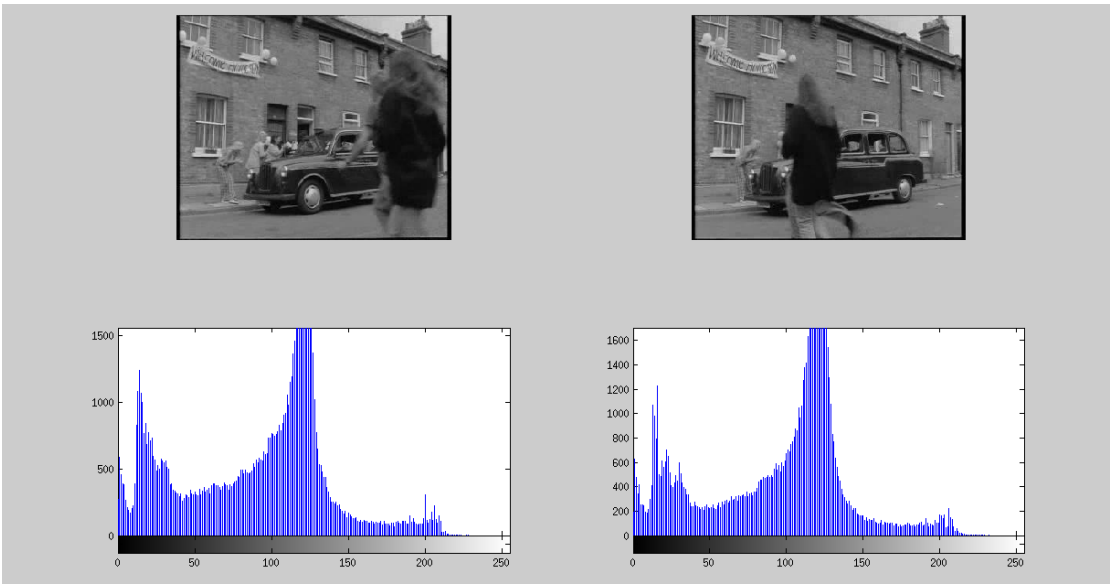

1500
1000
500
0
0
50
100
150
200
250
1600
1400
1200
1000
800
600
400
200
0
0
50
100
150
200
250

图16

图17

图18

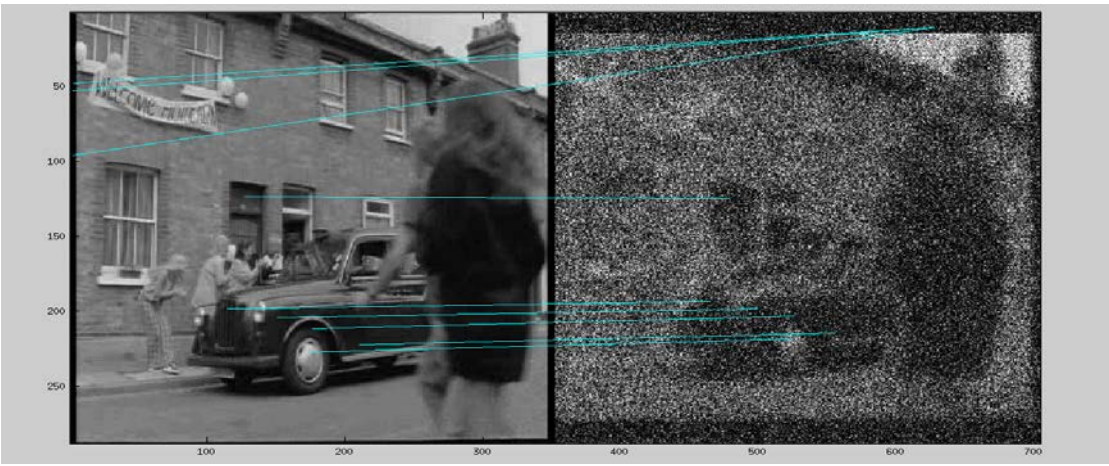


图19

如图12-19为基于上述模型的第一组第三组在 MATLAB 环境下运行的结果，从图中可以轻易地看出算法的可靠性与精确性，将盗用片段迅速地检验出来。

4.5 浅谈 SIFT 和 SURF 算法的比较

SIFT 在尺度和旋转变换的情况下效果最好，SURF 在亮度变化下匹配效果最好，在模糊方面优于 SIFT，而尺度和旋转的变化不及 SIFT，旋转不变上比 SIFT 差很多。速度上看，SURF 约是 SIFT 速度的3倍。

五、模型的评价与改进方向

5.1 模型的评价

(1) 模型从实际情况出发，给出了合理的假设，简化了问题的处理过程，使问题的解决过程更加简单清晰。

(2) 在模型中准确地使用了一些比较成熟的算法，如 SURF 算法、SIFT 算法等，有效地简化了问题的处理过程，使问题的求解更加快速准确。

(3) 本文中所探讨的几种模型均能实现对视频相似度的检验及查重，完全符合题目要求，对今后维护视频版权起到了一定的参考和借鉴价值。

5.2 模型的改进方向

(1) 两组视频若所取对应关键帧之间时间差距较大则匹配值会降低，影响最终结论。应考虑减少两对应关键帧之间的时间差。

(2) 对于 SIFT 算法由于其可进行旋转图形的特征点匹配，故对无旋转编辑方式的视频会有少量无效的匹配（如图10中的斜线），应考虑减少此类无效匹配。

参考文献

[1] Bhattacharya, A. (1943). "On a measure of divergence between two statistical populations defined by their probability distributions". *Bulletin of the Calcutta Mathematical Society* 35: 99–109. MR 0010358.

[2] Bay, H., Tuytelaars, T., Van Gool, L., "SURF: Speeded Up Robust Features", *Proceedings of the ninth European Conference on Computer Vision*, May 2006.

[3] Lowe, David G. (1999). "Object recognition from local scale-invariant features". *Proceedings of the International Conference on Computer Vision* 2. pp. 1150–1157. doi:10.1109/ICCV.1999.790410.