基于欧式距离的关键帧提取

学号：17343124
姓名：伍斌

一、引言

目前针对视频查找的方法依然局限于关键词，而在图片和视频大量传播的今天，使用图片进行匹配查找或许是一种新的展望。这次作业主要是针对老师上课所教的关键帧技术进行拓展。

文档具体包括我所了解到的关键帧提取这项技术的分类、应用和相关的实现算法。以及我自己所做项目的结果展示。

二、背景资料

视频的结构化层次：视频（video）、场景（scene）、镜头（shot）和帧（frame）。

本次项目主要涉及到的层次是帧。

帧：视频可以看做是一个在时间上连续的静态图像序列，而其中每一个静态图像称为一帧。

关键帧：关键帧是指能够描述镜头的主要内容的帧。根据镜头内容的复杂程度，可以从一个镜头中提取一个或多个关键正或者构造一个关键帧（此时，关键帧可能不是镜头中实际存在的帧）。

本次项目主要实现的是在同一个镜头中提取关键帧，所以再补充一下镜头和镜头边界检测的概念。

镜头：一个视频是由许多镜头组成的，而镜头是视频的一个物理单元，每个镜头包括一个帧序列。这些帧被连续地记录且表达了一个在时间上和空间连续的动作。镜头切换最简单的是突变和渐变两种类型。

镜头边界检测：找出视频序列中发生镜头变换的位置，从而将视频分割成独立的镜头片段。典型方法：基于直方图、基于边缘、基于模型、光流检测。

在进行实验之前我还重点研究了一下镜头边界检测的直方图法：

直方图法是使用最多且相对简单的计算帧间差的方法，其抗噪声能力比模板匹配强，用公式表示为：

$$D_{k,k+1}=\sum_{i=0}^{n}\left|H_k(i)-H_{k+1}(i)\right|$$

这种基于直方图的方法能打到 70%~80%的检测准确率，在处理大朗视频数据时具有优势，但是由于直方图描述的是一幅图像像素总体的灰度或颜色分布，忽略了像素的空间信息，所以对小的噪声和灰度不敏感。可能存在两帧的内容完全不同但是图像完全相似的情况。

以下重点介绍所查阅的关键帧提取的算法：

（1）基于镜头边界提取关键帧。将视频分割成镜头后，将每个镜头的首帧或末帧作为关键帧。这种方法实现简单但误差较大，不具代表性。

（2）基于运动分析提取关键帧。利用光流分析来计算镜头中的运动量。该方法计算量大存在误差。

（3）基于图像信息提取关键帧。通过每一帧图像颜色、纹理等视觉信息的改变来提取

关键帧。该方法选取的关键帧不具代表性，且当有物体快速变换使，容易造成关键帧选取冗余。

（4）基于欧式距离法。

这四种方法各有优劣，以下是除了本篇实现的欧氏距离法外，其他三种方法更详细的介绍：

（1） 基于镜头的方法

基于镜头的关键帧提取算法是视频检索领域中最先发展起来，也是目前最为成熟的一种通用方法，该算法的一般实现过程是：先按照某种技术手段把源视频文件按照镜头变化分割，然后在视频每个镜头中选择首、尾两帧作为关键帧。这种方法的优点是实施起来很简单，算法的计算量也很小，但是这种方法存在很大的局限性，当视频中内容变化剧烈、场景非常复杂时，选取镜头中的首、尾两帧并不能代表视频的全部内容变化，所以该方法已经远远不能满足当今社会人们对关键帧提取的标准和要求。

（2） 基于运动分析的方法

这种方法是一些学者基于物体运动特征的属性提出的一种关键帧提取算法，它一般的实现过程是：在视频镜头中分析物体运动的光流量，每次选择视频镜头中光流移动次数最少的视频帧作为提取到的关键帧。利用光流法计算视频帧的运动量公式如下所示：

$$M(k)=\sum\sum|L_x(i,j,k)|+|L_y(i,j,k)|$$

式中，M(K)表示第 k 帧的运动量,Lx(i,j,k)表示第 k 帧像素点(i,j)处光流 X 的分量，Ly(i,j,k)表示第 k 帧像素点(i,j)处光流 y 的分量。计算完成后，取局部最小值作为所要提取的关键帧。计算公式如下所示：

$$M(k_i)=\min[M(k)]$$

这种方法可以从大部分视频镜头中提取适量的关键帧，提取到的关键帧也可以有效地表达出视频运动的特征。但是，这种方法主要的缺点是：算法本身的鲁棒性较差，因为它不仅依赖于物体运动的局部特征，而且计算过程也较为复杂，算法在时间上的开销代价较大。

（3） 基于图像信息的方法

通过每一帧图像颜色、纹理等视觉信息的改变来提取关键帧。该方法选取的关键帧不具代表性，且当有物体快速变换使，容易造成关键帧选取冗余。

尽管因为时间原因，我没有选取无监督（聚类）的方法进行详细研究，但是我还是查阅学习了相关的算法。

基于视频聚类的方法

这种方法在提取关键帧的过程中是通过聚类的方式来表达视频的主题，通过聚类把视频帧划分为若干个簇，这一过程结束后在每个簇中选取相应的帧作为关键帧。该算法基本思想是：首先，初始化一个聚类中心。其次，通过计算聚类中心与当前帧之间的范围，确定被分为类的参考帧或者作为类的新聚类中心。最后，我们选择离聚类中心最近的视频帧处理成关键帧。

该算法的主要实现步骤如下：

① 输入视频帧数据的集合表示为：X={x1,…xn},其中每个 x 分别代表的是第 i 帧对应的 m 维特征向量，在给定的初始聚类个数 k(k≤n)的前提下，划分聚类的集合个数。

② 基于视频帧的颜色直方图的属性来提取集合 X 中的特征值，根据提取到的颜色特征值划分聚类个数，划分过程可以用聚类模型的最小值 C 来表示，计算公式如下所示：

$$C = \arg\min \sum_{i=1}^{k} \sum_{x_j \in C_i} \| x_j - u_i \|^2$$

式中 C={C1,C2…Cn}，就是聚类的结果，ui 表示聚类 ci 的平均值。

③ 从中将视频帧的第一帧对应的特征向量 x1 归入到第一个类中，并且将第一帧对应的的颜色直方图的特征值作为第一个类的初始质心。

④ 计算视频帧到质心的距离，如果当前比较的视频帧的距离大于给定的初始阈值 T，那么就把该帧归入到新的类中；反之，把当前帧归入到距离它最近的类中，并且更新该类的质心。

⑤ 重复 （4）过程，直到最后一帧对应特征向量的值 nx 归入某一个类中或者其作为一个新的类中心。

⑥ 每次选取距离聚类中心最近的视频帧作为关键帧。利用这种算法提取出的视频关键帧不仅冗余度小，而且关键帧可以很准确的反映出视频中发生的全部内容。但是，基于聚类的方法在划分聚类簇的过程中并没有充分考虑到各帧之间时间的先后变化顺序，并且在聚类之前需要预先设定一定数量的簇，所以该方法的适用性受到一定程度的限制。

三、算法复现及实现效果

以下具体介绍我所实现的欧式距离关键帧提取。

用 eulerdisdiff(i)表示第 i 帧图像的帧差欧式距离，其数学表达式为：

$$eulerdisdff(i) = \sqrt{\sum_{i=1}^{N-2} [(x_{i+2} - x_{i+1}) - (x_{i+1} - x_i)]^2}$$

其中 N 为视频的一个镜头中的帧图像数目，x_i、x_{i+1}和x_{i+2}分别为第 i、i+1、i+2 帧图像的灰度值。

用帧差欧式距离法进行关键帧提取的步骤：

1）珠帧计算各图像的帧差欧式距离，在 N 帧图像的镜头中总共有 N-2 个帧差欧式距离；

2）计算这 N-2 个帧差欧式距离的极值，以及各极值点对应的函数值；

3）计算各函数值得均值；

4）比较各极值点所对应函数值与均值的大小，取出大于均值的点，其对应的帧图像即为所要选的关键帧图像。

编程是使用了 MATLAB 进行。因为对软件相当不熟悉且时间不足的原因最后没有完成封装。

代码一共分为三部分：main.m，extremum.m 和 EulerDistanceDiff.m。

main.m 是主函数，实现了算法的主要部分；extremum.m 包括 extremum 函数，用于求极大值；EulerDistanceDiff.m 则是算法核心：求帧差欧氏距离。

以下为 EulerDistanceDiff.m 的内容。

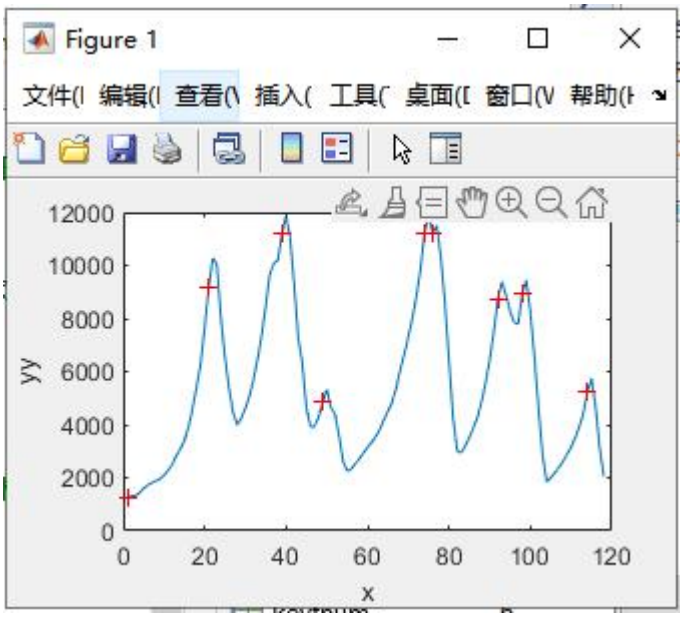
```
function de=EulerDistanceDiff(Mov,n) %求帧差欧氏距离
Xn=rgb2gray(Mov(1,n).cdata);
Xn=double(Xn);
Xnp1=rgb2gray(Mov(1,n+1).cdata);
```

```
Xnp1=double(Xnp1);
Xnq1=rgb2gray(Mov(1,n+2).cdata);
Xnq1=double(Xnq1);
diff = (Xnq1(:)-Xnp1(:))-(Xnp1(:)-Xn(:));
de = sqrt(sum(diff.*diff));
```

实现效果

（1）读取文件夹中的 avi 视频并将其拆分为一系列的帧（一共 120 帧）：

（2）进行相关计算所求出的极大值：



（3）由此程序得出的关键帧及其帧号：

（4）重新编号，方便后续操作：

由实验结果可见，此算法所提取的冗余接近 0，针对此实验视频（道路监控录像片段）有比较好的实现效果。

四、总结

本次项目主要完成了基于欧式距离的关键帧提取算法的复现，完成效果基本符合预期。但是可预见的应用前景相对比较窄，因为目前聚类算法已经发展相当完善，根据所查阅的资料来看，已经取得了相对不错的结果。

但本算法也具有不少优点：不需要数据集，鲁棒性较好，能够大幅度减少其他有监督方法会产生的冗余。

同时提取出来的重新编号的图片按顺序播放可以用来合成一段原视频的视频摘要。虽然丢失了大部分的细节信息但是总体内容的环节和内容还是可表达出来。比如这段道路监控所提取的关键帧（如果原视频更清晰的话），可以用在抓取路过车辆车牌号等管理上。

总的来说，这次的实验项目让我对视频检索和相关应用的发展产生了更多的兴趣。或许以后的检索可以是在搜索框导入一幅图，通过检索相似关键帧给出原视频地址。如今正处在以视频为载体的信息大量流行的时代，这方面相关的研究日益推进，希望我未来也会在此方向上有所涉猎。