

## 摘 要

近年来, 随着 Internet 的普及与多媒体技术的迅猛发展, 人们面临着爆炸式增长的视频信息。在这些海量视频之中存在着的大量相似视频, 一方面造成了多媒体数据库的冗余, 要求更加巨大的存储空间, 另一方面也增加了视频检索的难度, 降低了用户的体验质量。因此, 相似视频检测 (Near-Duplicate Video Detection, NDVD) 成为当前国内外的研究热点之一。

相似视频指在视频内容上相似程度很高但却并不完全相同的两段或多段视频。相似视频的成因有多种, 它们既可以是因为拍摄时不同的视角、光照条件造成的, 也可以是由后天经过某些转换模式 (Transformation, 如插入文字、加入噪声、画中画) 人工编辑产生的。由于相似视频的内容信息存在着上述的不确定差异, 相似视频检测问题向传统的视频检索问题提出了更高的要求。本文通过对视频的自相似性分析, 给出了一系列相似视频检测解决方案, 其主要工作包括:

1. 提出了一种基于自相似矩阵的视频特征提取与表达方法。与传统的基于关键帧抽取、通过比对关键帧来匹配相似视频的模式不同, 本文的工作以视频的整体信息为着眼点, 既注重局部特征的抽取, 又立足于视频全局, 将这些特征以自相似矩阵的形式有机组织起来。实验证明, 本文提出的“局部 (特征提取) ——全局 (特征表达)”更加适合相似视频检测问题。由于最终反馈给系统的是视频内容的自相似性, 因此该方法并不局限于帧匹配的正误, 对视觉层面发生了较大变化的相似视频仍然鲁棒。

2. 提出了一种自相似矩阵的描述子, 实现了自相似矩阵的有效距离度量。由于相似视频通常长短不一, 因此其分别产生出的自相似矩阵也具有不同的大小。同时, 注意到自相似矩阵本身的特性 (半定对称阵, 主对角线全 0), 本文提出了一种自相似矩阵描述子, 通过挖掘矩阵主对角线附近区域的模式信息进行两个不同大小的自相似矩阵距离度量, 藉此实现相似视频的精确匹配。实验证明, 与传统的自相似矩阵描述子对比, 我们的 VCS 描述子更加鲁棒, 相似视频检测质量更高。

3. 第一次提出了相似视频转换模式识别问题, 并给出了一种基于多特征的相似视频转换模式识别方法。转换模式是相似视频的成因, 我们将相似视频比作病人, 则转换模式相当于病症。看病时讲究对症下药, 而实验也证明不同转换模式下的相似视频特点不一, 可以通过转换模式的先验知识选择更加有效的“药物” (特征) 进行处理。本文首先提出了转换模式识别问题, 希望通过转换模式识别的研究进一步提高相似视频检测质量。

4. 提出了一种基于转换模式识别的相似视频检测方法。借助上述转换模式识别

的结果，我们选择对该种转换模式更加有效的特征。实验表明，转换模式识别和特征选择模块的加入有效提升了传统的单一特征条件下的相似视频检测质量。同时，相似视频的转换模式作为其自身重要属性也为媒体追踪、视频分类、归档等问题提供了重要信息。

**关键词：**相似视频检测，视频拷贝检测、自相似矩阵，VCS 描述子，转换模式识别

## **Near-Duplicate Video Detection Based on Self-Similarity Matrix and Associate Techniques**

Wu Zhipeng(Computer Application)

Directed by: Huang Qingming(Professor)

Nowadays, with the popularity of Internet and the rapid growth of multimedia technology, people are provided with explosively increasing video information. The existence of large amount of near-duplicate videos in the multimedia database, however, brings redundancy data which usually requires much more storage space, as well as increases the difficulty for video retrieval and degrades user experience. Consequently, Near-Duplicate Video Detection (NDVD) has become a hot topic in the field of multimedia recently.

Near-duplicate video stands for two or more videos which are highly similar in content, but appear differently due to different acquisitions (e.g. different lighting conditions, camera settings, view points) and manually edited transformations (e.g. caption insertion, adding noise, and picture in picture). Note that the near-duplicates are differed from uncertain diversities, NDVD issue raises a high requirement for the traditional content based video retrieval schemes. In this paper, based on the self-similarity analysis of video content, we provide a serious of methods for NDVD. The main contributions of this paper are:

1. Based on Self-Similarity Matrix (SSM), we propose a novel feature extraction and organization method for video content. Comparing with the traditional keyframe selection and matching based NDVD schemes, our approach figures in the entire video pattern and establishes a “local (frame level feature extraction) - global (feature organization by SSM)” video content expression. Based on the self-similarity analysis for videos, the proposed method is robust to extensive visual appearance transformations, and experiments show the effectiveness in NDVD issue.

2. We propose an effective descriptor to measure the distance for SSM. On noticing that the lengths of near-duplicates are always different, the corresponding SSMs are also formed by different size. Moreover, SSM is a symmetric positive semi-definite matrix with zero values along the main diagonal. In this paper, an elaborately designed descriptor called Visual String Character (VCS) is proposed for matching SSM. Based on the analysis of local patterns around the main diagonal area,

the distance between different SSMs can be measured and quantified. Experiments show that comparing with other SSM measurements VCS descriptor is more robust and provides a higher near-duplicate video detection quality.

3. We first discuss the issue of transformation recognition in NDVD, and we present a transformation recognition approach based on multi-features. In our view, near-duplicate video is analogized as patient who wants to be prescribed. Only if the type of disease (transformation) is detected can we use the proper medicine (effective feature) to treat the illness (matching the near-duplicate videos).

4. We propose a near-duplicate video matching scheme based on transformation recognition. According to the transformation recognition result, a more effective feature is selected from the feature pool and used for near-duplicate video matching. The experiments verify the promotion transformation recognition brings to the NDVD issue. Meanwhile, tagging the near-duplicate videos with their transformation also provides useful information which appears to be important in many areas such as media tracking, video classification and archiving.

**Keywords:** Near-Duplicate Video Detection, Video Copy Detection, Self-Similarity Matrix, VCS Descriptor, Transformation Recognition.