国防科技大学学报
JOURNAL OF NATIONAL UNIVERSITY OF DEFENSE TECHNOLOGY
第19卷第2期 1997年4月 Vol. 19 No. 2

基于相关度和注意力聚焦的图像内容检索*

曹莉华 李国辉 胡晓峰

(国防科技大学系统工程与数学系 长沙 410073)

摘 要 基于内容的检索中对图像和视频帧的处理是以色彩、形状和纹理等为基本特征,这些特征在图像背景比较单一的情况下容易获得。一旦图像背景为色彩和纹理均很复杂的图像时,图像中目标的颜色和形状等特征就很难获得。若将从计算机视觉中引出的相关度和注意力聚焦两个概念用于从复杂背景图像中定位和选取最有关的信息,则能够有效地检索到最匹配的图像。本文通过对相关度和注意力聚焦两个概念的讨论,介绍了计算机视觉对基于内容检索的作用,尤其是在可视示例查询QVE (Query by Visual Example) 的情况下从大图像数据库中检索图像的情况。

关键词 图像内容,检索,相关度,注意力聚焦

分类号 TP392, TN941.1

Image Content Retrieval Based on Relevance and Focus-of-attention

Cao Lihua Li Guohui Hu Xiaofeng

(Multimedia Research Center, NUDT, Changsha, 410073)

Abstract Color, shape and texture are the essential features for image and video frame processing in content-based retrieval. It is easy to obtain these features when the background of the image is simple. Once it becomes complicated color and texture, it is difficult to get such features. Two concepts, namely: relevance and focus-of-attention quoted from the computer vision can be used to locate and select the most related imformation in the image with a complicated background and the best matched image can be efficiently retrieved. In this paper we discuss these two concepts and discribe their roles in image retrieval, especially in the case of QVE (Query by Visual Example).

Key words content-based retrieval relevance focus-of-attention

* 国防预研基金、湖南省自然科学基金资助项目
1996年6月20日收稿

传统数据库对多媒体内容的检索和查询主要采用基于文本的方法。查询时需指明文本特征。因此要求用户对文本特征的描述具有一定的准确性和描述的规范性。然而，不同的用户对同样的媒体内容可能有着不同的抽象，因此，这种方法不一定获得满意的结果。基于内容的检索直接针对图像或视频帧的图像特征进行处理。这些图像特征有颜色、形状和纹理等，它们是匹配的关键，因此，特征抽取的质量直接影响检索的速度和精确度。对于背景比较单一且目标突出的图像，抽取这些特征已不成问题。然而，很多图像或视频往往具有色彩丰富、纹理复杂的背景，此时要抽取所需的特征就很困难。即使特征抽取出来，里面却包含很多背景信息，使得检索相当困难。例如用户要检索图 1 (a) 所示的图像，该图在背景为纯黑的情况下可得到图 1 (b) 所示的直方图；现图像库中存在与该图一模一样但具有一定背景的图像 (如图 2 (a))，则图 2 (a) 的直方图如图 2 (b) 所示，可见它们的直方图差别是非常大的。如果用基于颜色直方图特征来检索这幅图像，则将会出现很大误差。

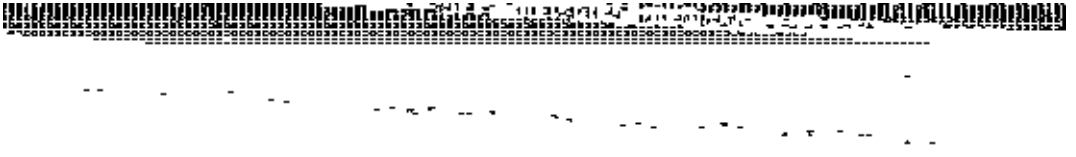

图 1　无背景的图像及其直方图

图 2　有背景的图像及其直方图

计算机视觉旨在分析图像数据以模仿人类视觉系统来理解图像结构和内容。但计算机视觉的困难有许多，如数学上没有很好的定义，难处理的计算问题，抽取信息时受噪声的很大干扰等。寻找一个视觉问题中广义的数学公式始终无法做到，但我们可利用计算机视觉中某些知识来优化一些工作。

相关度通过模仿人类视觉来对图像特征中的基元按其质量好坏进行排序，使得检索时，只对“最好”的基元进行匹配，大大减少了查找空间。

注意力聚焦机制用来定位图像中最感兴趣的“注意区域”，使目标从复杂背景中分离出来，在检索中排除干扰，因此也提高了检索效率。

1　基于内容的检索

基于内容的检索是对传统信息检索的拓广和加深。它是基于一个能体现信息语义的模型，对语义信息的生动表现往往通过可视实例，因此目前常用的基于内容的检索都采用可视示例查询 (Query by Visual Example) 的方法。

可视示例查询可分为两种：全局化的和结构化的。所谓全局化查询主要是基于用户给出的图像的全局信息。典型的有基于色彩或文本，如为表示全局色彩查询，用户可在色彩– 饱和度– 亮度(HSL) 空间上选择一个或多个主色，同时也可以包含有被检索图像的每个颜色的相应百分比。对于文本查询，用户可能要确定主要内容或要检索的文本信息。全局查询还包括基于图像纹理或形状的全局化描述的查询。由于这种查询中用户能

够给出图像的全局信息，相对来说，处理起来比较简单的。处理这样的全局查询有许多传统的图像分析方法[5]、[6]，此处不再赘述。

由于图像本身背景或颜色的复杂性，用户可能很难用全局化的描述来说明查询要求，但是，对于有一定目标的图像，不论其背景多复杂，人类视觉系统都能迅速定位于目标。因此，用户能对目标的结构具有清晰的映像。此时用户能够提供一个（可能是画的）需查找图像的“形状”的粗略轮廓，在这种情况下就没法用全局技术，我们称这种查询为结构化查询。处理结构化查询时直接针对用户给出的所需图像的粗略的结构化描述进行处理。由于每个形状定义了一个必须在图像中空间定位的相关结构，对于形状的每个部分，必须在库中找到相应的形状，这些形状可能就是在图像中的基元以及这些基元间的空间关系，图像匹配通过相应基元的匹配进行。

为了使这种查询模式在实际应用中容易做到，尽量降低匹配的复杂性，在结构化查询处理中作如下三步处理：

1）抽取图像基元并按某种规则排序。

图像基元可通过图像分割方法获得。对于得到的基元，计算各基元的相关度，并按相关度值的大小将基元排序。

2）定位和描述图像中最“感兴趣”的内容。

为从复杂背景中找到感兴趣的目标，必须利用计算机视觉中引出的“注意力聚焦”机制，将图像中感兴趣的部分抽取出来。

3）查询时从图像数据库的已抽取的基元中找到与结构化查询相应的内容。

对所有按相关度排序的基元，通过感兴趣区域的加权得到调整的相关度并作为查询的基本特征存入图像特征数据库中。当用户提出结构化查询要求时，直接提取用户描述结构中的基元并与数据库中的基元特征作匹配，以找到用户要求的内容。

2 相关度和注意力聚焦

2.1 相关度（Relevance）

从一套图像基元中匹配和识别目标是典型的复杂查找问题。用计算机视觉知识来选择出与所需识别对象相关的信息将使查找的复杂性激剧下降。相关度是用来选择相关信息的测度，将其介绍如下。

设考虑的图像（或视频帧）被分成 P 类基元，如线性段、圆弧或区域。不同基元类型的数目由 $p = 1, \cdots, P$ 表示（此处 $P = 3$）. 用一个简单的标记 τ_j^p 表示类型 p 中的某一特定基元，用 M^p 来表示一套从图像中抽取出来的 p 类基元的全部标记图及其属性和空间关系。

图像中感兴趣的目标可能是由一组简单标记构成。一些简单标记的组合称做多元标记 $T_j\{\tau_{j1}^{p1}, \cdots, \tau_{jn}^{pn}\}$，则一个多元标记就是一个由这些基元描述的结构化实体。多元标记在可视示例查询（QVE）中用于目标与模型的定性匹配。各类基元可用图像分割方法得到[2]。

对类型 $p = 1\cdots3$（线性段、圆弧、区域）的每个简单标记 τ_j^p（从输入图像中抽取），相关度 $\rho[\tau_j^p] \in [0, 1]$ 的定义如下[3]：

$$\rho\left(\tau_j^p\right) = r\left(\tau_j^p\right) \quad s\left(\tau_j^p\right) \tag{1}$$

此处 $r\left(\tau_j^p\right)$ 和 $s\left(\tau_j^p\right)$ 分别代表 τ_j^p 的可信度和重要性测度。高可信度代表标记是一个有意义的实体。重要性测度衡量图像标记的突出性; 当 τ_j^p 的属性使它在类中最突出时它的值最大。可信度和重要性测度通过分析计算每个基元的属性时得到。关于可信度和重要性测度的属性详细介绍如下。

用属性 $A_r^p\left(\tau_j^p\right)$ 来计算依赖于所属标记图 M^p 的标记 τ_j^p 的可信度。如对于线性段($p=1$), 有两个属性 $A_r^1\left(\tau_j^1\right)$, $r=1,2$, 分别是长度和对比度; 圆弧的4个属性 $A_r^2, r=1\cdots4$, 是半径长度、弧长度、对比度和误差; 对于区域, 可信度属性 $A_r^3, r=1,\cdots3$, 是面积、平均对比度、色彩分配的标准偏差。

用属性 $A_s^p\left(\tau_j^p\right)$ 来计算依赖于类型的标记的重要性测度; 对线性段, 有两个属性 $A_s^1\left(\tau_j^1\right)$, $s=1,2$, 分别是长度和方位; 圆弧的三个属性 $A_s^2, s=1,2,3$, 是半径、弧长和转动角度; 区域的两个重要性属性 $A_s^3, s=1,2$, 是面积、平均强度。

一个给定标记 τ_j^p 的可信度是定义在基元类型 p 上的所有可信度属性 r(在整个标记图 M^p 上) 的归一化的总和。

$$r\left(\tau_j^p\right)=\sum_r A_r^p\left(\tau_j^p\right)\Big/\sum_r\sum_j A_r^p\left(\tau_j^p\right) \tag{2}$$

重要性测度是通过计算一个标记的属性和所有同类型上的其它标记属性的平方差的总和得到:

$$S\left(\tau_j^p\right)=\sum_s\sum_{j\neq i}\left(A_s^p\left(\tau_j^p\right)-A_s^p\left(\tau_i^p\right)\right)^2 \tag{3}$$

用图3来说明相关度用于估计给定类型标记的各自的“重要性”。图中上列是从输入图象中抽取的基元, 下列为相关性测度的表示(相关值越高的标记使得象素点越黑)。

相关性值由每个基元类型 p 独立计算出。为了得到可与所有基元类型比较的相关性值, 对 ρ 值进行排序, 得到每个基元的相对重要性。

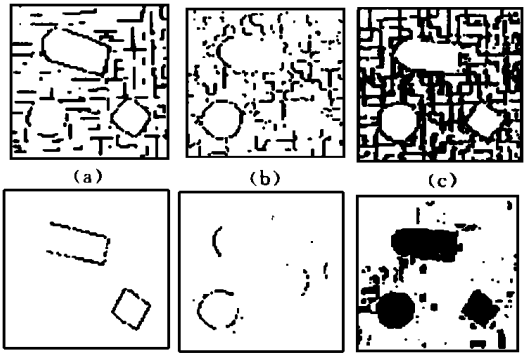

图3 基元和相关度

2.2 注意力聚焦 (Focus-of-attention)

视觉注意力模型模拟生物视觉系统能快速检测和定位静态视网膜图像上“感兴趣”部分的能力, 以便减少目标识别的数据量。人类视觉系统在估计图像中有刺激性的部分的重要性时, 用到了一些准则, 如从下往上或数据驱动方法[4]等。用这些方法, 可通过在每个位置上抽取的信息与剩余图像进行比较来计算具有突出性的部分。

注意力聚焦机制主要分为三个阶段, 标于图4. 第一, 多视网特征网 (Feature Maps F-Maps) F^k 的获得, 其中 $k=1\cdots K$, 它从输入图像抽取, 并反映了一些在视觉皮层计算出的图像性质。将输入图像经过红-绿、蓝-黄色彩对半的滤波器得到代表色彩信息, 其它一些特征图则代表无色的高频率信息, 它们可通过一阶高斯滤波器得到[4]。

注意力聚焦的第二阶段是对关注图 (conspicuity maps C-maps) C^k 的抽取, 对应于每个特征类型 k 的特征图抽取一个关注图。关注图代表在图像每个位置上有兴趣部分的

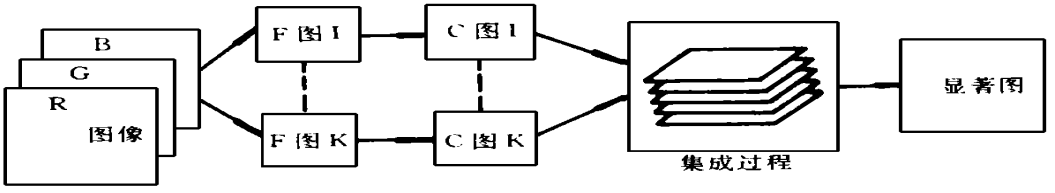

图 4　注意力聚焦机制

K 个从下往上的测度, 它的值在[0, 1] 区间。这些测度可通过特征图与高斯滤波器的多尺度卷积得到。关注图 C^k 则通过计算在每个位置上的测度的平方得到[4]。

在第三阶段, 所有关注图合成为一个显著图, 定义为 C – 图的平均总和。直接从原始 C – 图中计算出的简单的平均总和就能够得到所有显著位置的平均值, 而不用清晰地检测它们。可针对所有 C 图用迭代非线性算法计算得到。还可通过最小化能量测度来减少噪声[1]。

图 5　用注意力聚焦机制提取的图像

图 5 给出了用图 4 方法对不同类型的输入图像处理后得到的结果。即使没有任何感兴趣目标的先验知识, 结果也成功地检测出了图像中 "不规则 "的形状, 它对应于那些从复杂的纹理背景中清晰地显示出来的目标。在图像的检索中, 这个机制可用来确定那些 "感兴趣 "目标和要用来存贮和检索的图像的某些部分。更重要的是, 它能用来确定那些必须在数据库中找到的与查询图像最相关的元素。

3　相关度和注意力聚焦在基于内容检索中的应用

对于大多数背景比较复杂的图像, 仅靠一般的图像处理方法进行基于内容的检索将是困难的。相关度和注意力聚焦机制可优化可视示例查询。

图 6　用注意力聚焦和相关度得到的图像

1) 首先按图像分割方法抽取图像基元, 对每个基元标记 τ_i^q 按公式(2) 和(3) 计算其可信度和重要性测度, 由此可计算出该基元的相关度(公式(1)) 。用相关度对图像基元按相关值的大小排序。这样处理使得在开始匹配时就可以选择最相关的基元, 以大大减少查找空间。

2) 对每幅图像按图 4 所示的注意力聚焦机制定位图像中感兴趣结构的 "注意区域 "。这样定位出的 "注意区域 "实际上就是图像的目标区域, 通过判断基元可能属于的区域来分割图像, 得到与 "注意区域 "最接近的基元。这些基元是下步用于匹配的特征。具体做法如下:

在存取图像或识别目标时, 依照由注意力聚焦机制得到的位置区域来加权初始相关度。可考虑每个基元所属的注意区域的重心, 计算一个近似的测度 $\pi(\tau_i) \in [0, 1]$, 与重心相近的基元其 $\pi(\tau_i)$ 的值就大, 而较远的基元其值就小。

对标记 τ_j 的相关度 $\rho(\tau_j) \in [0,1]$，用 $\pi(\tau_j)$ 进行调整，即 $\rho(\tau_j) \in [0,1]$：

$$\rho(\tau) = (\rho(\tau) + \pi(\tau))/2 \qquad (5)$$

这样就可将有高相关值（$\rho(\tau)$）的特征区域检测出来（如图 6）。

3) 将上步得到的“注意区域”内的基元用于匹配。基元匹配直接针对第 3.1 节中描述的基元的属性 $A_r^p(\tau_j^p)$ 和 $A_s^p(\tau_i^p)$。

总之，基于相关度和注意力聚焦机制的图像检索需作如下工作。首先，在数据库建立阶段，必须对所有图像进行预处理以便以后更好地查找。预处理包括从图像中抽取最相关的基元并组织它们、关键字的自动选择以及建立文本索引。第二，由于基于内容的检索中的高度交互性和受人控制性，要考虑人与媒体的交互。为将可视示例查询转换为可检索的图像基元，此时采用计算机视觉技术。最后，在检索阶段，视觉技术还将用于对图像查询的响应。总之，计算机视觉技术能够提供有效的允许在图像多媒体信息系统中进行基于内容的查询的方法。

4 结束语

相关度和注意力聚焦机制的结合引出一个允许从不同源中融合数据、从背景较复杂的图像中检索图像的方法。相关度允许选择最关联的基元、定量分析并按它们的“质量”排序，注意力聚焦将图像中最引入注意的特征按空间定位，滤出不相关的基元。在多媒体信息系统中，相关和注意力的概念可用来进行图像存储、图像检索和人与媒体的交互；尤其在 QVE 中，大大减少了查找空间。对于图像存储，这些概念允许选择出重要的特征并按它们的关联性排序，定位图像中的感兴趣项，然后就可用这些最重要的项作为查询图片的索引关键。对于图像检索，相关度和注意力聚焦机制作为检测数据库的总体变化的测度，提供了一个能够用于 QVE 的衡量索引和匹配的量。最后，在人与媒体的交互中，相关度用来对从相应查询的数据库中检索图像并排序。总之，计算机视觉提供了一系列用于图像信息处理的技术。相关度和注意力聚焦机制使得基于内容的图像检索更精确、更快捷。

参考文献

1 Thierry P and R Milanese. Computer Vision and Multimedia Information Systems. Proceedings of Multimedia Information Systerm and Hypermedia. Tokyo, Japan. 1995

2 王润生. 图像理解. 长沙：国防科技大学出版社，1995

3 Bost J M, Milanese R and Pun T. Temporal Precedence in asynchronous visual indexing. Proc. 5th Int. Conf. on computer Analysis of Images and Patterns (CAIP 93), budapest, Hungary, Sept, 13- 15. 1993. 468- 475

4 Milanese. R. Wechster. H. Gil. S. bost. J. M. Pun. T. Intergration of bottom-up an top-down cues for visual attention using non-linear relaxation. IEEE-CVPR 94 (Computer Vision and Pattern Recognition). Seattle, Washington. June 20- 23. 1994. 781～785

5 Venkat N, Gudivada and Vijay V. Raghavan. Content-Based Image Retrieval systems. IEEE. Computer. September, 1995

6 Myron Flickner, Harpreet Sawhney, etal. Query by Image and Video Content: The QBIC System. IEEE. Computer. September, 1995

(责任编辑　潘　生)