

图像语义结构信息的理论分析^{*}

余卫宇, 余英林, 谢胜利, 曹 燕

(华南理工大学 电信学院, 广东 广州 510640)

摘 要: 严格地导出了图像结构信息的表达式, 定义了一个由结构参数组成的结构空间 T 并用心理测度函数的变换函数族把图像从结构空间 T 映射到结构信息空间, 反映人从图像结构中所接收到的结构信息。还用函数来表达人对某个结构参数刺激的特殊响应。最后指出分析图像结构后可以得出结构信息, 而图像人对图像结构了解越清楚, 则人所得的结构信息就越大, 从而人对图像的结构语义的判断就越正确。

关键词: 图像结构信息; 变换函数; 结构空间; 图像语义

中图分类号: TN911.73 **文献标识码:** A **文章编号:** 0529-6579 (2006) 05-0034-05

图像语义早期用于人工智能、知识工程领域中, 1968 年 Quillian^[1] 最早提出了语义网络的概念, 它可用不同符号串表示, 并进行符号运算与逻辑运算、规则提取与推理。后来将语义网络用在自然语言理解, 模式分析领域中, 并在句法模式识别领域中出现了句法模式文法推导等概念。八十年代前后, 在语音处理中和信息论的研究中都提出语义信息的表示与分析问题, 在中文信息处理领域中, 还研究中文语句的语义信息的表达与提取^[2], 这样研究者方便运用知识工程中的语义模型, 并进行句法分析等。国内外关于图像语义的检索已有大量的文献报导^[3-10], 但对语义的理论研究仍处于起步阶段。如图像语义能否用函数表示, 图像结构能否定量分析, 人能否根据图像语义信息评价图像质量, 这也是本文要讨论的问题。

1 图像语义的定义

语义目前没有一个明确的定义, Djemba^[11] 把语义分成两类: 主观语义和客观语义。早期认为图像语义是一种概念。文献 [6-7] 把图像中的颜色、纹理、形状、运动信息等低级特征作为语义特征。文献 [10] 提出了几种图像模型结构。

本文提出了下述的图像结构的几个基本定义。

定义 1 图像结构参数。

灰度结构 T_g 描述图像前景和背景中各个可视物体 (Visual Object) 的主色调, 全图中颜色的种类, 及其各种色调区的共生概率。

几何结构 T_j 描述图像中所含的 VO 数及 VO

的轮廓形状、特征、拓扑性质; 各个 VO 的近似规则纹理结构及全图的主纹理结构; VO 间的空间关系还包括顺序, 方位、度量等。

变化结构 T_v 描述图像各 VO 及主要 VO 的运动速度及方向。对于视频而言, 是主要物体的灰度变化, 即图像进化的速度及方向。

结构密度 T_d 描述在给定的语义描述窗口 D_i 内, ①包含超过门限 T_i 的前述各项结构参数的数量; ②包含的化简后结构产生规则的条数; ③包含的奇异点、边沿线、纹理线的数目。

以上三项有一项或多项的数目, 若超过设定的门限值, 而达到显著值, 则称语义描述窗口内的图像结构是复杂的、繁忙的, 反之是平凡的、简单的。

选择这些结构参数, 其原因主要有: 灰度结构含有大量的表达语义结构的因数, 是一种重要的语义载体, 人们从中能够解读出重要的图像语义信息。例如人们常说的“明亮、阴暗的色调”、“色调配合合理”, 说明灰度结构中含有丰富的语义。图像灰度的不同排列可以表达不同的语义内容。人们从图像中视觉物体的形状轮廓、纹理、运动方向等方面接受到很多关于图像的语义信息, 并可大体知道图像的语义。如人们常用“结构严谨、结构合理”来评价图像。

可见, 上述定义都是低层次物理参数。它们构成图像的结构空间, 根据空间及拓扑的基本性质, 可生成拓扑空间。因为上述结构参数均是正实数, 所以结构空间为正实数多维空间, 而给定图像在图

^{*} 收稿日期: 2006-01-10

基金项目: 国家自然科学基金资助项目 (60372068); 国家杰出青年科学基金资助项目 (60325310)

作者简介: 余卫宇 (1972年生), 男, 博士, 讲师; E-mail: yuweiyu@scut.edu.cn

像结构意义下可表示为该结构空间中的一点。

定义 2 图像结构信息。

由上述结构参数作为信息源符号向观察者发出的信息称为该图像的结构信息。本文定义的结构信息是一个确定性图像结构所载荷的信息，是该确定性结构的知识的表达与度量。

定义 3 图像语义。

图像语义是人们对图像确定性结构参数所表达的结构信息的理解，抽象概括出来的一个概念，它能最恰当地表述该图像结构所载荷的语义。

2 图像结构信息的分析式

2.1 结构信息

人作为图像的观察者、接受者，图像结构参量输入到“眼脑心理判断”系统，观察者理解该图像结构并进行抽象得出关于它的语义的表达。我们采用变换函数族，把图像从结构空间映射到结构信息空间，记作：

$$\varphi = \varphi(\varphi_g, \varphi_s, \varphi_m, \varphi_d) \tag{1}$$

则有

$$\varphi: T(T_g, T_s, T_m, T_d) \rightarrow I(I^1, I^2, I^3, I^4) \tag{2}$$

φ 把图像结构空间中的一点投影到结构信息空间中 I ，记结构信息为 I^1 ，则有：

$$I^1 = \varphi_g(T_g) + \varphi_s(T_s) + \varphi_m(T_m) + \varphi_d(T_d) \tag{3}$$

空间 T 与空间 φ 均为 Hilber 空间中的子空间。由于变换函数 φ 已经模拟了“眼脑心理判断”系统的变换，则结构信息就是观察者通过图像结构诸元理解得到的结构信息。

(3) 式只是结构诸元表现的自结构信息 $I^1(T_g, I^1(T_s), I^1(T_m), I^1(T_d))$ 。除了结构元本身的自结构信息外，结构元之间还有关联功能，例如：从 T_g 与 T_s 结合理解可得到一种细的结构信息，记为 I^2 。又进一步，可从 T_g, T_s, T_m 三者结合会得到一种更细的结构信息，记为 I^3 。同理，对于四维实数结构空间，有下列结构信息。分别用 (4)，(5)，(6) 式表示：

$$I^2(T_g, T_j) = \varphi_{ij}(T_g, T_j), \text{ 式中 } (i, j) \in (g, s, m, d) \tag{4}$$

$$I^3(T_m, T_g, T_s) = \varphi_{mgs}(T_m, T_g, T_s) \tag{5}$$

$$I^4(T_m, T_g, T_s, T_d) = \varphi_{gsmd}(T_g, T_s, T_m, T_d) \tag{6}$$

根据拓扑性质有：

$$I^1 \subset I^2 \subset I^3 \subset I^4 \tag{7}$$

I^1 表示最粗的结构信息， I^4 为最细的结构信息。

以上各式可进一步推导如下：由 (2) 式，得

$$I = \varphi(T) = \varphi(T_g, T_s, T_m, T_d) \tag{8}$$

由定义一可知，结构参数 T_g, T_s, T_m, T_d 均是多维向量，它们是结构空间 T 中的子集。

由图像的物理意义看出，函数组 φ 有下列的性质：

变元中只有一个 (或多于 1 个) 为零，则 $\varphi=0$

$$I=0 \text{ 当 } T_g, T_s, T_m, T_d < 0, \quad I = \max, \text{ 当 } T_g, T_s, T_m, T_d \rightarrow \infty$$

随着结构各参数数值的不断增加 (这里指超过门限值以上的值)，即结构参数对 HVS 系统的有效刺激日益增大，观察者接受到的结构信息也随之增加，但当这些刺激增到一定程度时，观察者对图像结构理解就已经充分，所以当结构参数的刺激增加到趋于饱和程度时，观察者能完全理解给定图像的结构，对它的理解没有模糊度，此时结构信息达到最大。由此可见，函数族 φ 随着各坐标值正向逐渐增加时， φ 是单调增函数，到某个刺激的临界值以后， φ 不再增加，趋于饱和，达到极大值 $I_{\max}$ 。

由上讨论知，函数族 φ 可视为变量 T_g, T_s, T_m, T_d 的连续函数，因为人接收的信息不能突变，所以存在各阶导数，则在 $(T_{g0}, T_{s0}, T_{m0}, T_{d0})$ 附近把函数展成 Taylor 级数，这一点可选择 $\frac{\partial \varphi}{\partial T_g}$

$\frac{\partial \varphi}{\partial T_s}, \frac{\partial \varphi}{\partial T_m}, \frac{\partial \varphi}{\partial T_d}$ 均为极大值处的点，则在超曲面 φ 的切平面的斜率较大的点附近的展式，由此得：在 $T = T_0 = (T_{g0}, T_{s0}, T_{m0}, T_{d0})$ 处 Taylor 级数展式：

$$I = \varphi(T_g, T_s, T_m, T_d) = \varphi(T_0) + I^1 + I^2 + I^3 + \dots + \text{高阶项} \tag{9}$$

$$\begin{aligned} I^1 &= a_{g0}(T_g - T_{g0}) + a_{s0}(T_s - T_{s0}) + a_{m0}(T_m - T_{m0}) + a_{d0}(T_d - T_{d0}) \\ I^2 &= a_{gs0}(T_g - T_{g0})(T_s - T_{s0}) + a_{sm0}(T_s - T_{s0})(T_m - T_{m0}) + a_{md0}(T_m - T_{m0})(T_d - T_{d0}) \\ I^3 &= a_{gsm0}(T_g - T_{g0})(T_s - T_{s0})(T_m - T_{m0}) + a_{gsd0}(T_g - T_{g0})(T_s - T_{s0})(T_d - T_{d0}) + a_{sm d0}(T_s - T_{s0})(T_m - T_{m0})(T_d - T_{d0}) + a_{m d g0}(T_m - T_{m0})(T_d - T_{d0})(T_g - T_{g0}) \end{aligned} \tag{10}$$

显然，(9) 式是 (3) 式的特殊情况，(10) 式是

(4) 式的特殊情况, (10) 式是 (5) 式的特殊情况。

上面已证明, 根据人对图像结构参数的了解而获得的结构信息的表达式 (3), (4), (5) 式可由信息映射函数 φ 的 Taylor 级数展开式来逼近, 而 (3), (4), (5) 式则是更为一般的表达式。其实 (6) 式也可从 Taylor 级数的 4 阶展开式中求得。

下面进一步讨论 (9) 式的意义: $\mathbb{F}^{(1)}$ 是从单个图像结构参数所获得的自信息, 它是最基本的组成部分, 也是结构信息的强度较大的项, 是一种较粗的拓扑结构, 是在某个较大范围, 较粗尺度下得到的结构信息。至于 $\mathbb{F}^{(2)}$ 则是由二个结构参量共同确定的结构信息, 结构信息范围缩小, 细节增加。这可看成多了一些细的结构信息, 所以 $\mathbb{F}^{(2)}$ 的拓扑结构细一些, 但相对强度较弱。 $\mathbb{F}^{(1)}$ 要同时满足三个结构参量的结构信息, 所以其范围会进一步缩小, 拓扑结构更细些, 表示更细的结构信息。

由 (9) 式看出, 高阶结构信息存在于超曲面 φ 的弯曲部分, 即“转折点”附近, 所以得出下列基本结论:

即当结构参数的刺激值超过 φ 的斜率最大点后, 若进一步增加, 则由此靠近超曲面的“转折点”这段刺激值所引起的结构信息, 含有较丰富的高阶结构细节信息。人们所获得的关于该图像的结构信息的范围会更细更具体。进而会得到一个更贴切、更清晰的语义概念来描述该图像结构。从 (9)、(10) 式可知, 它是把 φ 在 T_0 附近小范围展开, 其系数是常数。若把 T_g, T_s, T_m, T_d 及其拓扑 (即二元组合, 三元组合等) 均看成 Hilbert 空间的子空间, 它们仍然是线性空间。从进化模型角度看, 在小范围内系统可用常系数线性微分方程描述。但若在大范围说明变换函数族 φ 曲面的行为必须考虑其非线性特性, 即用非线性函数组成广义线性空间。因而 (10) 式可修改为:

$$\mathbb{F}^{(1)} = w_g \varphi_g(T_g) + w_s \varphi_s(T_s) + w_m \varphi_m(T_m) + w_d \varphi_d(T_d)$$

$$\mathbb{F}^{(2)} = w_{gs} \varphi_{gs}(T_g, T_s) + w_{gm} \varphi_{gm}(T_g, T_m) + w_{gd} \varphi_{gd}(T_g, T_d) + w_{sm} \varphi_{sm}(T_s, T_m) + w_{sd} \varphi_{sd}(T_s, T_d) + w_{md} \varphi_{md}(T_m, T_d)$$

$$\mathbb{F}^{(3)} = w_{gsm} \varphi_{gsm}(T_g, T_s, T_m) + w_{gsd} \varphi_{gsd}(T_g, T_s, T_d) + w_{gmd} \varphi_{gmd}(T_g, T_m, T_d) + w_{msd} \varphi_{msd}(T_m, T_s, T_d)$$

(11)

若把 T_0 点选在 $\mathbb{F}=0$ 处, 即设定 $\varphi(T_0)=0$ 由此得到较一般表达式:

$$\mathbb{F} = \varphi(T_g, T_s, T_m, T_d) = \mathbb{F}^{(1)} + \mathbb{F}^{(2)} + \mathbb{F}^{(3)} + \dots + \text{高阶项} \quad (12)$$

上面 (11)、(12) 式组成较一般的结构信息表达式, 简记为:

$$\mathbb{F} = \varphi \circ T \quad (13)$$

式中,

$$\begin{aligned} \varphi &= (\varphi^{(1)}, \varphi^{(2)}, \varphi^{(3)}, \varphi^{(4)}) \\ T &= (T^{(1)}, T^{(2)}, T^{(3)}, T^{(4)}) \end{aligned} \quad (14)$$

2.2 结构信息映射函数族 φ 的选择

人根据图像结构参数进行图像的模式识别, 这些结构参数作为“眼脑心理判断”的输入刺激, 最后通过心理判断, 得出人接收到关于该图像结构的结构信息。

已知心理测度函数^[8]的幅度是“眼—脑—心理判断”对输入结构参数刺激的正确识别概率。显然, 正确识别概率越高, 即人对图像理解就越正确, 因此人图像获得的图像结构信息也越多。反之, 若心理测度函数的幅度越小, 意味着人对该结构误判较多, 识别误差较大。因而, 人对结构的理解也越不清楚, 越不正确, 所得的结构信息就越小。基于以上分析, 我们可以定义以心理测度函数为基础构成的函数族 φ 。二维心理测度函数的可表示为:

$$p_{\text{new}}(u, v) = \left[\frac{1}{1 + e^{u(-\theta)}} \right] \left[\frac{1}{1 + e^{v(-\theta)}} \right] \quad (15)$$

由此, 我们定义函数族 φ 如下:

$$\varphi_i = \frac{1}{1 + e^{k_i(T_i - \theta_i)}}; \quad i = (g, s, m, d) \quad (16)$$

$$\begin{aligned} \varphi_{ij} &= \left[\frac{1}{1 + e^{k_i(T_i - \theta_i)}} \right] \left[\frac{1}{1 + e^{k_j(T_j - \theta_j)}} \right]; \\ (i, j) &= (gs, gm, gd, sm, sd, md) \end{aligned} \quad (17)$$

$$\varphi_{ijk} = \left[\frac{1}{1 + e^{k_i(T_i - \theta_i)}} \right] \left[\frac{1}{1 + e^{k_j(T_j - \theta_j)}} \right] \left[\frac{1}{1 + e^{k_k(T_k - \theta_k)}} \right] \quad (18)$$

式中: $(i, j, k) = (gsm, gsd, gmd, mds)$

$$\begin{aligned} \varphi_{ijkl} &= \left[\frac{1}{1 + e^{k_i(T_i - \theta_i)}} \right] \left[\frac{1}{1 + e^{k_j(T_j - \theta_j)}} \right] \\ &\quad \left[\frac{1}{1 + e^{k_k(T_k - \theta_k)}} \right] \left[\frac{1}{1 + e^{k_l(T_l - \theta_l)}} \right] \end{aligned} \quad (19)$$

式中 $(i, j, k, l) = (gmsd)$ 。

以 $(T_i)^2$ 项为例, $(T_i)^2 = T_i T_i$, 即两个 T_i 出现的结构信息。而且 T_i 和 T_i 同时出现给予人的刺激效果, 应与一个 T_i 一样, 因此 $(T_i)^2$ 带来的结构信息与 T_i 的一样, 所以 $(T_i)^2$ 从性质上不会带来比更多的结构信息。以上讨论对于 $(T_i)^3$, $(T_i)^4$ 一样成立。

最后, 关于人的“眼脑心理判断”系统对某些刺激的响应可能很大, 这可用两种方式表达, 一

是特别强烈的响应可用 φ 函数的斜率很大来表达；或是用一个离散 δ 函数叠加在 φ 函数的上升沿处表征。因为根据 Taylor 级数的展开式，可以允许存在有限一类间断点，所以如果某人对某个刺激有很强烈的响应，则可在该刺激对应的 φ 函数中叠加一个离散 δ 函数。这样，在获得 I 的中相应多了一个 δ 刺激项，这说明在对应该刺激处，某人可获得极大的结构信息。

2.3 结构参数的表达形式

(1) 关于 T_k 在定义一所述的 T_k 定义中，有一级参量和由此衍生的次级参量，其本身是树结构，所以作为人“眼—脑心理判断”系统的输入刺激应该有强弱之分。由于本文采用一级参量，即彩色强度及彩色种类，作为输入值，所以有：

$$T_k = \sqrt{\left(\frac{I}{I_{max}}\right)^2 + \left(\frac{S}{S_{max}}\right)^2 + \left(\frac{C}{C_{max}}\right)^2} \tag{20}$$

式中： S 、 I 、 C 为彩色体制， S 为色饱和度， I 为强度， C 为彩色种类。

(2) 关于 T_k 按定义一中规定的 T_k ，也有一级参量和次级参量，本身也是树结构。为便于计算，本文采用一级参量，形状可简单分为复杂形状和简单形状，这可由轮廓线中包含的角点数（包含特征点等）来衡量。另一方面，形状可由面积 / 周长比或矩表示。本文采用

$$T_k = \sqrt{\left(\frac{n_i}{n_{ij}}\right)^2 + \left(\frac{r}{r_{max}}\right)^2 + \left(\frac{l_{min}}{l_{max}}\right)^2} \tag{21}$$

式中： n_i 为全图中主 VO 的轮廓线中含有的角点数（即斜率显著突变点）， n_{ij} 为全图各 VO 的角点总数； r 为（面积 / 周长）比， r_{max} 为最大的 r 值； l_{min} 、 l_{max} 分别表示短径与长径。

(3) 关于 T_k 该参量描述视频帧中提取的运动向量，选取其主运动向量的绝对值，对静止图像采用灰度梯度值来表示，故有：

$$T_k = \sqrt{\left(\frac{m_v}{m_{vmax}}\right)^2 + \left(\frac{\nabla_g}{\nabla_{gmax}}\right)^2} \tag{22}$$

式中： m_v 为运动向量绝对值； ∇_g 为主 VO 的主纹理中的梯度值。

(4) 关于 T_k 取在 D_k 窗口内的含有的上述显著值数目，且归一化为小于 1 的参量。有下式：

$$T_k = \frac{(n_k, n_l \in D_k)}{A_{ks}} \tag{23}$$

式中： n_l 为 D_k 窗内的显著结构参量数； A_{ks} 为 D_k 窗的面积。

3 结 语

定义了一个由结构参数组成的结构空间 T 再用心理测度函数的变换函数族把图像从结构空间 T 映射到结构信息空间 I ，图像并用 δ 函数反应人对某个结构参数刺激的特殊响应。本文得出分析图像结构后可以得出结构信息，而人对图像结构了解越清楚，则人所得的结构信息就越大，从而人对图像的结构语义的判断就越正确。

参考文献：

[1] QULLIAN Semantic memory semantic information processing M, MIT Press 1968 227—270

[2] 胡俊峰, 俞士汶. 唐宋诗词词汇语义相似度的统计分析及应用[J]. 中文信息学报, 2002(4): 39—44

[3] YONG Rui THOMAS SH SHARAD M Browsing and retrieving video content in a unified framework[C]. Multimedia Signal Processing IEEE Second Workshop on 7—9 Dec 1998 9—14

[4] CHANG Edward Y LI Beiao LI Chen Toward perception—based image retrieval content—based access of image and video libraries 2000 Proceedings[C]. IEEE Workshop on 12 June 2000 101—105

[5] COLOMBO C BMBO A D PAI A P Retrieval of commercials by video semantics Computer Vision and Pattern Recognition 1998 Proceedings[C]. 1998 IEEE Computer Society Conference on 23—25 June 1998 572—577.

[6] DENG Yinling MANJUNATH B S Content-based search of video using color texture and motion[C]. Image Processing 1997. Proceedings International Conference on Vol 2 26—29 Oct 1997 534—537

[7] YU Weiyu YU Yingli XIE Shengli et al Representation on image semantics structure information[C]. The 8th world Multi—Conference on Systemics Cybernetics and informatics Orlando Florida U S A 2004 253—256

[8] STANDY A K Measuring estimating and understanding the psychometric function A commentary[J]. Perception and Psychophysics 2001 63(8): 1421—1455

[9] CORRIGONI J M BMBO A D Film semantic analysis computer architectures for machine perception 1995 Proceedings[C]. CAMP 95 18—20 Sept 1995 202—209

[10] 余卫宇, 余英林, 谢胜利, 等. 几种图像结构模型的分析[J]. 华南理工大学学报, 2005(1): 1—5

[11] DJERABA C Content—based multimedia indexing and retrieval[J]. IEEE Multimedia 2002(2): 18—22

(下转第 41 页)

M icrowave Perm ittivity of A rbon Nanotubes

LU Xiao-lai, ZHAO Dong-lin, TIAN Li-xia

- (1. Department of Physics and Electronic science
2. Key Laboratory of science and Technology of Controllable Reactions of Ministry of Education// Institute of Carbon Fiber and Composites, Beijing University of Chemical Technology, Beijing 100029, China
3. School of Life Science, Sun Yat-sen University, Guangzhou 510275, China)

Abstract: The microwave permittivity of multi-wall carbon nanotubes blended in Paraffin wax has been studied at the frequency range of 2—18 GHz. The dissipation factors of the multi-wall carbon nanotubes are high at the microwave frequencies. The microwave permittivity of the multi-wall carbon nanotubes and paraffin wax (or other dielectric materials) can be tailored by the content of the multi-wall carbon nanotubes. ϵ' and $\tan\delta$ increase with the volume filling factor (v) of the multi-wall carbon nanotubes. The ϵ' can be effectively modeled using second-order polynomials ($\epsilon' = A^2 + Bv + C$). The ϵ' of the multi-wall carbon nanotubes decrease with frequency at the frequency range of 2 ~ 18 GHz. The high dissipation factor $\tan\delta$ (ϵ''/ϵ') of multi-wall carbon nanotubes are due to the dielectric relaxation.

Key words: multi-wall carbon nanotubes; microwave; permittivity

(上接第 37 页)

Theory Analysis of Image Semantic Structure Information

YU Wei-yu, YU Ying-lin, XIE Sheng-li, CAO Yan

- (Institute of Electronic Engineering and Control, South China University of
- Technology, Guangzhou 510640, China)

Abstract: An expression of image structure information is derived strictly. A structure space T is defined and mapped onto structural information space Π by a transformation function set based on psychometric functions. Π denotes the structure information of a given image obtained by an observer who has understood the structures of the image. A function is used to express a special response of an observer to some structure parameter stimulation. It's also pointed out that an observer can obtain the structure information after analyzing the structure of a given image. The more an observer knows about the structure of the image, the more he obtains the structure information, and the more correct decision he can make on the semantics of the structure of the image.

Key words: image structure information; transform function; structure space; image semantic