

基于 DNA 计算的多维分配数据关联算法^①

梁 冰^② 冯 林

(大连理工大学创新实验学院 大连 116024)

摘 要 针对多目标跟踪中的多维分配数据关联这一 NP 难题,将 DNA 计算用于数据关联研究,利用 DNA 分子链生化反应的高度并行性,实现了基于 DNA 计算的多维分配数据关联算法。该算法利用包含限制性内切酶 *HaeIII* 回文序列的 DNA 分子链对观测视线进行编码,通过连接和退火操作获得所有目标关联组合,然后利用检测 DNA 发夹结构和凝胶电泳技术排除不满足约束条件的目标关联组合,筛选出多维分配数据关联的最优解。分析结果表明:基于 DNA 计算的多维分配数据关联算法所需的 DNA 分子链数即空间复杂度为 $O(nk!/(c-m)!)$,将多维分配问题的指数级复杂度降低到多项式复杂度 $O(mn)$,其中 n 为传感器数, m 为目标数, k 为目标定位点数。

关键词 DNA 计算, 目标跟踪, 多维分配问题, 数据关联算法

0 引 言

数据关联(data association)是多目标跟踪的核心步骤和难点。数据关联将传感器获得的目标观测集合,按目标进行分类,从而得到目标的航迹^[1,2]实现目标跟踪。Yaakov 等学者将多传感器多目标数据关联描述为多维分配问题(multidimensional assignment problem, MSP)^[3]。多维分配是有效排除虚假定位点的多目标多传感器数据关联算法^[4],该算法通过建立候选分配树,得到所有可能的分配候选关联,确定每一候选关联的代价,然后寻求一种总关联代价最小的分配(关联)方案,其正确关联率高,但随着传感器数和目标数的增加,计算复杂性呈指数急剧增长^[5,6]。因此,多维分配数据关联问题,实质上是在一定约束条件下的最优化问题,属于 NP 难题(np-hard problem, NPP)^[7]。

针对多维分配数据关联的 NP 难题,本文将 DNA 计算用于多维分配数据关联的研究,建立了有效求解多维分配 NP 难题的计算模型,为多传感器多目标数据关联的研究提供了新的方法和理论支持。DNA 计算利用 DNA 分子具有海量的存储能力及生化反应的高度并行性解决 NP 难题^[8]。Adleman 提出的 DNA 计算模式是利用 DNA 碱基互补配

对进行数据信息编码,将数据运算并行地映射成 DNA 分子链的生化过程^[9]。目前关于 DNA 计算的研究集中在三个方向:一是图与组合优化的 DNA 计算模型,解决了如最大团问题^[10]、图同构问题^[11]、图着色问题等 NP 难题^[12];二是通讯领域,杨磊利用 DNA 计算解决了带指定结点约束的路由选择 NP 难题,通过一系列生物操作,筛选出问题的精确解^[13];三是密码学领域, DNA 链的大规模并行性可以同时攻击一个计算问题的不同部分,从而破解数据加密标准的最艰难的组合问题^[14]。由于 DNA 计算的高度并行性,可以通过生化操作高效的获得所有目标关联组合并排除不满足条件的目标关联组合而获得最优解,首先通过包含限制性内切酶 *HaeIII* 回文序列对观测进行编码,利用连接和退火操作获得所有的目标关联组合,然后利用 DNA 发夹结构和凝胶电泳技术排除不满足约束条件的目标关联组合,最后用磁珠分离检测结果。

1 多维分配数据关联问题描述

多维分配问题可做以下数学描述:设有 S 个集合(称为原始集合),第 s 个原始集合包含 n_s 个元素,多维分配的目的在于寻找 S 元的最可能集合(称为关联集合),使得原始集合的每个元素只分配给一

① 国家自然科学基金(61173163)和中央高校基本科研业务费专项资金(DUT11RC(3)54)资助项目。

② 女,1981 年生,博士,工程师;研究方向:数据融合,形式语言与自动机;联系人, E-mail: liangbing@dlut.edu.cn (收稿日期:2011-12-06)

个 S 元关联集合,并且每个关联集合最多只包含来自每个原始集合的一个元素,多维分配问题在维数 $S \geq 3$ 时是一个 NP 难题。

在多传感器目标跟踪系统中,通过不同传感器提供的方位角测量值和传感器自身位置信息作观测视线组合得到目标定位点,多维分配的数据关联方法利用多维分配的思想,找出最优的一种观测视线组合即目标关联组合分配方案,有效地完成多目标数据关联。

1.1 基于形式语言的多维分配数据关联问题描述方法

通过形式语言描述多维分配数据关联问题是建立多维分配数据关联的 DNA 计算模型的基础。DNA 分子可以描述为字母表 $D = \{A, T, G, C\}$ 上的字符串,将 DNA 分子在限制性内切酶和连接酶作用下的重组行为,抽象为在一定规则控制下的字符串运算。因此,本文采用形式语言描述多维分配数据关联问题,通过字符表示目标定位点,目标定位点集合上的字符串(语言)表示数据关联结果,如下所示:

(1) m 个传感器对 n 个目标进行观测,获得 m 个目标观测集合 $\sum_i (i = 1, 2, \dots, m)$, 其中 $\sum_i = \{x_{i1}, x_{i2}, x_{in}\}$ 表示传感器 i 获得的 n 个目标观测的集合。

(2) m 元组 $(x_1, x_2, \dots, x_m)$ 表示 m 条观测视线交叉获得的目标定位点 $s = (x_1, x_2, \dots, x_m)$, 其中 m 元组 $s \in \sum_1 \times \sum_2 \times \dots \times \sum_m$, “ $\times$ ”表示笛卡儿乘积, $\sum_1 \times \sum_2$ 也可简记为 $\sum_1 \sum_2$ 。如果传感器数目为 m , 那么由 m 条观测视线交叉获得目标定位点 s 表示可能的目标。

(3) $\Pi = \{s_1, s_2, \dots, s_k\}$ 表示 m 个传感器获得的所有可能目标定位点的集合,字符 s_k 是 m 元组。

(4) $L(\Pi) = \{\omega \in \Pi^*\}$ 是 Π 上的语言。

字符 $s \in \Pi$ 表示目标对应一个视线组合,字符串 $\omega \in \Pi^*$ 表示多个目标的关联结果。获得正确的数据关联结果,即找到最优的视线组合的字符串。

1.2 多维分配数据关联算法描述

多维分配数据关联算法的目标是在 $L(\Pi)$ 中找到最优字符串 $\omega \in \Pi^*$, 即表示目标关联组合的最优视线组合。最优的视线组合必须满足以下 3 个约束条件:

(1) 同一个目标(视线组合)的观测视线必须来自不同的传感器,同一个目标的两条或两条以上的

观测视线不能来自同一传感器。因为在一个关联周期内,一个传感器对同一目标不会有多条观测视线,即 $\omega = s_1 s_2 \dots s_n = (x_{11}, x_{21}, \dots, x_{m1})(x_{12}, x_{22}, \dots, x_{m2})(x_{1n}, x_{2n}, \dots, x_{mn})$, $s_1 s_2 \dots s_n$ 表示 $s_1, s_2, \dots, s_n$ 之间的笛卡儿乘积。

(2) 一条观测视线源自一个目标,同一视线不能来自多个目标。在进行视线的组合分配时,同一观测视线不能重复分配给不同组合,对于字符串 $\omega = s_1 s_2 \dots s_n$, 如果 $x_{ij} \in s_k$, 那么 $x_{ij} \notin s_1, s_2, \dots, s_{k-1}, s_{k+1}, \dots, s_n$ 。

(3) 观测视线在上述两个约束条件下进行组合分配,得到多个候选观测视线组合集合。能够得到最大目标数目且代价最小的组合即为最优组合。满足最大目标数目的字符串长度为 mn 。

实例 1: 为不失一般性,设有 3 个传感器对 4 个目标进行观测,每个传感器得到 4 条观测视线, $\sum_1 = \{x_{11}, x_{12}, x_{13}, x_{14}\}$, $\sum_2 = \{x_{21}, x_{22}, x_{23}, x_{24}\}$ 和 $\sum_3 = \{x_{31}, x_{32}, x_{33}, x_{34}\}$ 。如图 1 所示,当传感器为 3 个(或更多)时,由于测量误差的存在,对同一目标的 3 条观测视线一般不交于一点,而是两两相交得到 3 个交叉点,对同一目标的观测视线,3 个交叉点相距较近,如视线组合 (x_{11}, x_{21}, x_{31}) 的 3 个交叉点;若 3 条视线不是来自同一目标的观测,3 个交叉点相距较远,如视线组合 (x_{11}, x_{21}, x_{32}) 的 2 个交叉点;从而排除一些错误的视线组合。

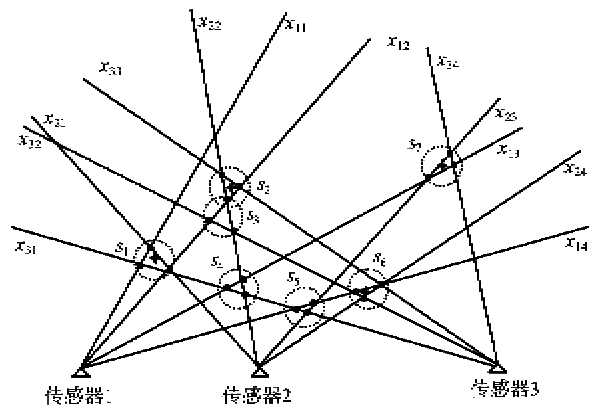


图 1 3 个传感器对 4 个目标观测视线组合

可能视线组合如图 1 中用虚线圆圈标出的 7 个: $s_1(x_{11}, x_{21}, x_{31})$, $s_2(x_{12}, x_{22}, x_{33})$, $s_3(x_{12}, x_{22}, x_{32})$, $s_4(x_{13}, x_{22}, x_{31})$, $s_5(x_{14}, x_{23}, x_{31})$, $s_6(x_{14}, x_{24}, x_{32})$, $s_7(x_{13}, x_{23}, x_{34})$, 每个视线都对应了一个可能的目标,将这些目标在上述的约束条件(1),(2)下进行分配组合,得到用字符串所示的分配结果:

$$\begin{aligned}\omega_1 &= s_1 s_2 s_6 s_7 = (x_{11}, x_{21}, x_{31})(x_{12}, x_{22}, x_{33}) \\ &\quad (x_{13}, x_{23}, x_{34})(x_{14}, x_{24}, x_{32}); \\ \omega_2 &= s_1 s_3 s_7 = (x_{11}, x_{21}, x_{31})(x_{12}, x_{22}, x_{32}) \\ &\quad (x_{13}, x_{23}, x_{34}); \\ \omega_3 &= s_2 s_5 = (x_{12}, x_{22}, x_{33})(x_{14}, x_{23}, x_{31}); \\ \omega_4 &= s_3 s_5 = (x_{12}, x_{22}, x_{32})(x_{14}, x_{23}, x_{31}); \\ \omega_5 &= s_4 s_6 = (x_{14}, x_{24}, x_{32})(x_{13}, x_{22}, x_{31}).\end{aligned}$$

由分配结果可以看出, $\omega_1, \omega_2, \omega_3, \omega_4, \omega_5$ 获得的描述, 目标数分别为 4, 3, 2, 2, 2。只有 ω_1 能够满足约束条件(3), 即能够得到最大的目标数目, 所以分配方案 1 即为数据关联结果。

2 DNA 编码方案

对传感器的每条观测视线 x_{ij} 编码, 每条观测视线 x_{ij} 编码包含一段回文序列, 两侧包含一段鸟嘌呤片段, 使用鸟嘌呤作为前缀和后缀的目的是尽量避免编码的自身退火。回文的中间片段包含限制性内切酶 *Hae*III 的识别位点: 5'-GG|CC-3'; 3'-CC|GG-5', 其中切割点用竖线表示。表 1 为实例 1 中观测视线的 DNA 编码, 每条观测视线 x_{ij} 编码的一般形式为: 5'-GGGN₁N₂...N₆GGCCN₆...N₂N₁GGG-3', 其中,

表 1 观测视线的 DNA 编码

观测视线	编码(22bp)	观测视线	编码(22bp)
x_{11}	5'-GGGACTCGAGGCCCTCGAGTGGG-3'	x_{23}	5'-GGGACTCGTGGCCACGAGTGGG-3'
x_{12}	5'-GGGCCTCAAGGCCCTTGAGGGGG-3'	x_{24}	5'-GGGGAGTCCGGCCCGGACTCGGG-3'
x_{13}	5'-GGGGGAACCGGCCCGTTCCGGG-3'	x_{31}	5'-GGGTAAGTGGGCCCCACTTAGGG-3'
x_{14}	5'-GGGATCCGTGGCCACGGATGGG-3'	x_{32}	5'-GGGTACGACGGCCGTCGTAGGG-3'
x_{21}	5'-GGGCCACGAGGCCCTCGGTGGGG-3'	x_{33}	5'-GGGGCGGCTGGCCAGCCGCGGG-3'
x_{22}	5'-GGGTACTGGGGCCCCAGTAGGG-3'	x_{34}	5'-GGGCAGCACGGCCGTGCTGGGG-3'

G 表示鸟嘌呤, N_i 表示任意的核氨酸, $\overline{N_i}$ 表示 N_i 的 Watson-Crick 互补。

3 条观测视线交叉获得目标定位点 (x_{11}, x_{21}, x_{31}) 的双链分子由编码 (x_{11}, x_{21}) 和编码 (x_{21}, x_{31}) 连接反应得到。编码 (x_{11}, x_{21}) 的前半段是编码 x_{11} 后半段的补码, 分子的后边段是编码 x_{21} 前半段的补码, 那么 (x_{11}, x_{21}) 的编码为 3'-GGAGCTCAC-CCCCCGGTGCTCC-5'。编码 (x_{21}, x_{31}) 的前半段是编码 x_{21} 后半段的补码, 分子的后边段是编码 x_{31} 前半段的补码, 那么 (x_{11}, x_{21}) 的编码为 3'-GGAGC-CACCCCCCATTCACCC-5', 因此, 目标定位点 (x_{11}, x_{21}, x_{31}) 的双链分子结构如图 2 所示。

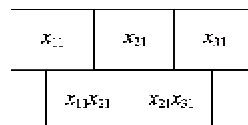


图 2 目标定位点 (x_{11}, x_{21}, x_{31}) 的双链分子结构

对编码目标定位点的所有 DNA 链进行退火和链接, 产生一组表示多目标关联结果的组合链库。通过 DNA 计算可以从组合链库中找出满足多维分配条件的多目标关联最优字符串 $\omega, \omega = (x_{11}, x_{21}, x_{31})(x_{12}, x_{22}, x_{33})(x_{13}, x_{23}, x_{34})(x_{14}, x_{24}, x_{32})$ 的双链分子结构如图 3 所示。

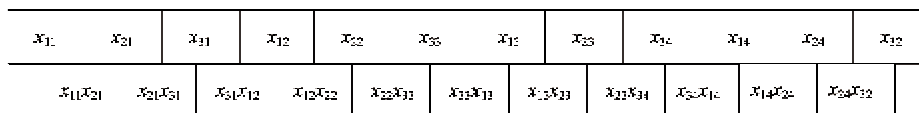


图 3 目标关联组合的 DNA 双链分子编码

3 多维分配数据关联的 DNA 算法

3.1 多维分配数据关联的 DNA 算法实验步骤

多维分配数据关联的 DNA 算法的基本思想是利用包含限制性内切酶 *Hae*III 的 DNA 分子链对观测视线进行编码, 通过退火和连接操作获得目标关

联组合的集合; 然后高温变性操作使其变为 DNA 单链, 含有相同的观测视线的 DNA 单链将高度并行形成发夹结构, 并在生物酶 *Hae*III 的作用下降解生成发夹结构的 DNA 链; 通过凝胶电泳技术获得最大目标数的目标关联组合, 最后通过聚合酶链反应 (PCR) 检测最终结果。多维分配数据关联的 DNA 算法分为 5 步:

- (1) 随机生成表示目标关联组合的字符串。
- (2) 查找表示目标关联组合的字符串中是否包含同一观测视线。
- (3) 删除包含同一观测视线字符串。
- (4) 保留包含 n 个目标定位点的字符串。
- (5) 如果存在字符串, 读出最优的目标关联组合。

第(1)步中多维分配数据关联算法的计算过程从设计一个初始试管 T 开始, 试管 T 中含有目标定位点编码的双链 DNA 集合。试管中的 DNA 双链通过退火和连接形成表示目标关联组合的字符串。

第(2)步中通过升高温度使双链 DNA 分子变性, 形成单链 DNA 分子。如果单链 DNA 分子中包含相同的观测视线的编码时, 将形成发夹结构(如图 4), 形成的发夹结构中的部分双链区包含 *Hae*III 识别位点。将限制性内切酶 *Hae*III 添加到溶液中, 切割所有构成发夹结构 DNA 链的部分双链片段。例如, 一条观测视线源自一个目标, 同一视线不能来自多个目标。如果在进行视线的组合分配时, 同一观测视线 x_{22} 重复分配给组合 $\omega = (x_{12}, x_{22}, x_{33})(x_{13}, x_{22}, x_{31})$, 将形成发夹结构。

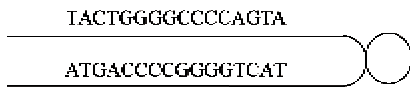


图 4 单链 DNA 构成的发夹结构

第(3)步 将限制性内切酶 *Hae*III 添加到溶液中, 切割所有构成发夹结构的 DNA 链的部分双链片段。剩下未降解的单链 DNA 对应的目标关联组合含有一条观测视线至多一次。

第(4)步 使用凝胶电泳技术将 DNA 分子根据长度加以分离。为了找出最大目标数的目标关联组合的 DNA 链, 需要通过凝胶电泳找出试管中最长的 DNA 链。实例 1 中长度为 $(3 \times 4 \times 22\text{bp}) 264\text{bp}$ 的片段表明是 4 个(最大目标数)目标关联组合, 通过凝胶电泳可测定最长链长度是 264bp , 即是代表最大目标数的关联组合的 DNA 链长度。

第(5)步 输出运算结果。采用磁珠分离检测代表目标关联组合的 DNA 链, 获得 DNA 链由哪 4 个目标定位点构成的目标关联组合, 即多维分配数据关联算法的解, 通过测序可得实例 1 中由 DNA 链表示的多维数据关联算法的最优解 $\omega = s_1 s_2 s_6 s_7 = (x_{11}, x_{21}, x_{31})(x_{12}, x_{22}, x_{33})(x_{13}, x_{23}, x_{34})(x_{14}, x_{24}, x_{32})$ 。

3.2 多维分配数据关联的 DNA 算法操作与实现

实验中用到以下操作:

(1) 复制 $(T, \{T_1, \dots, T_n\})$ ($\text{copy}(T, \{T_1, \dots, T_n\})$)。该操作以试管 T 为基准, 产生试管 T 的 n 个副本。

(2) 删除 $(T_k, x_i = x_j)$ ($\text{remove}(T_k, x_i = x_j)$)。该操作删除 T 中任意一个含有相同字符串 ($x_i = x_j$) 的字符串。

(3) 检测 (T) 操作($\text{detect}(T)$)。若试管 T 中至少包含一个 DNA 分子, 输出“是”; 否则输出“否”。

初始试管 T 由目标定位点集合 Π 上的字符串 $s_1 s_2 \dots s_j$ 集合组成, 该集合表示所有可能的目标关联组合, 多维分配数据关联算法的目标是在初始试管 T 中找到满足 3 条约束条件的最优字符串, 即最优目标关联组合。

算法 MdimAssignDataAssocia (T, Π, m, n)

输入: 输入试管 T , 目标定位点集合 Π , m 是传感器数, n 是目标数

for $k \leftarrow 1$ to n do

$\text{copy}(T, \{T_1, \dots, T_n\})$

end for

$k = 0$

do

{

for $i \leftarrow 1$ to $m \times (n - k) - 1$

for $j \leftarrow i + 1$ to $m \times (n - k)$

$\text{remove}(T_k, x_i = x_j)$ // 在试管 T_k 中删除包含相同的观测视线编码的关联组合

end for

end for

$T_k \leftarrow \text{length-separate}(T_k, m \times (n - k))$ // 查找包含最大目标数的关联组合

$k++$

}

while($\neg \text{detect}(T_k)$) // 如果没有满足当前目标数的关联组合, 那么返回查找比目前目标数少 1 个的关联组合

算法 MdimAssignDataAssocia (T, Π, m, n) 在获得所有可能目标关联组合以及排除不满足约束条件的关联组合时具有极高的并行性, 避免了利用电子计算机时, 当目标数或是传感器数增加时, 计算量和计算时间呈指数级增长。解决多维分配数据关联问题所需的 DNA 分子链数即空间复杂度为 $O(nk!/(c - m)!)$, 由于 DNA 计算中的巨大并行机制, DNA

计算实现多维分配数据关联算法,可将指数级的时间复杂度降为 $O(mn)$,其中 n 为传感器数, m 为目标数, k 为目标定位点数。

4 模拟 DNA 计算的实验结果

为了验证多维分配数据关联的 DNA 算法的有效性,本文采用文献[15,16]中的实验方法,给出使

用该算法的求解结果,并与传统关联算法(M-best 多维分配算法、拉格朗日松弛算法)进行时间代价和关联性能的比较。由于多维分配问题的复杂度不但取决于传感器的数目,还取决于目标数,分别模拟在三个到五个传感器条件下,因而对目标数为 3 批、5 批和 8 批的观测数据进行关联。表 2 所示 DNA 算法与 2 个传统的多维分配数据关联算法的性能对比。

表 2 多维分配数据关联算法性能对比

传感器数目	目标数	M-best 多维分配算法		拉格朗日松弛算法		DNA 多维分配算法	
		关联准确率 (%)	时间代价	关联准确率 (%)	时间代价	关联准确率 (%)	时间代价
3 传感器	3 批	89.10	1.35	92.21	0.92	91.73	0.63
	5 批	73.51	8.03	80.43	6.71	87.65	1.32
	8 批	62.92	22.76	76.92	18.23	75.36	2.74
4 传感器	3 批	70.42	7.85	78.40	4.79	83.05	1.24
	5 批	51.81	37.43	63.59	33.65	72.41	2.35
	8 批	45.73	210.81	58.17	135.44	64.54	3.45
5 传感器	3 批	50.64	26.12	54.32	29.28	70.57	5.34
	5 批	43.23	198.31	46.28	123.56	66.35	6.42
	8 批	—	—	—	—	63.29	8.45

根据表 2 所示,拉格朗日松弛算法的性能优于 M-best 多维分配算法,前者虽在时间代价方面较低,但关联准确率高。随着传感器和目标数的增加,以上两种算法在时间代价的增大趋势是相同的,均呈指数增长,而关联准确率也快速下降。当传感器达到 8 个,目标数达到 8 批时,已经超出了计算能力,在时间代价上和目标关联率上,传统拉格朗日松弛算法和 M-best 多维分配算法均不再有效。

在时间代价上与传统多维分配算法对比明显的是,随着传感器和目标数的增加 DNA 多维分配算法求解多维分配问题所需的时间代价并没有太多变化,时间代价都在 10 个单位以内,而传统算法时间代价已经从 1 个单位上升 100 多个单位,直到无法衡量时间;在关联准确率方面,随着 DNA 分子链浓度的增大,DNA 算法也出现了关联性能下降,但较传统的拉格朗日松弛算法在关联准确率上也平均提高 7.67%。DNA 计算过程中,DNA 链的生化反应在几秒钟内获得了关联结果,但生化操作包括分离、输出最优解的时间远远多于生化反应的时间。当传感器数和目标数扩大后,输出结果所花费的时间成为 DNA 计算的瓶颈。

5 结论

本文提出的多维分配数据关联的 DNA 算法,利用 DNA 计算的巨大的并行性和高密度存储能力解决了多维分配 NP 难题,其主要思想是通过形式语言与字符串描述多维分配问题中的目标定位点和目标关联组合,通过回文 DNA 编码获得目标定位点对应的 DNA 链,从而利用 DNA 操作的并行性高效地获得所有表示目标关联组合的 DNA 链,并通过检测 DNA 发夹结构和凝胶电泳技术删除不满足约束条件的 DNA 链,最终获得数据关联的最优解,DNA 分子生物算法将多维分配问题的指数复杂性降低到多项式复杂性。

由于 DNA 链的数目随着目标定位点的增加而会呈指数级增长,所以当多传感器获得的目标定位非常大时,DNA 链的数目将会随之剧增,如何降低算法的空间复杂性需要进一步的研究。另外,DNA 计算并行度高,计算速度快,但从信息的输入到获得 DNA 计算的结果的扩增和电泳操作花费的时间却很长,因此提高 DNA 计算的生化操作的效率和可行性也是下一步的研究重点。

参考文献

- [1] Ba H X, Cao L, He X Y. Modified joint probabilistic data association with classification-aided for multitarget tracking. *Journal of Systems Engineering and Electronics*, 2008, 19(3):434-439
- [2] Storms P A, Spieksma P F. An LP-based algorithm for the data association problem in multitarget tracking. *Computers and Operations Research*, 2003, 30(7):1067-1085
- [3] Yaakov B S, Blackman S S, Fitzgerald R J. The dimensionless score function for measurement to track association. *IEEE Transactions on Aerospace and Electronic Systems*, 2007, 43(1):392-400
- [4] Chummun M R, Kirubarajan T, Pattipat R I. Fast data association using multidimensional assignment with clustering. *IEEE Transactions on Aerospace and Electronic Systems*, 2001, 37(3):898-913
- [5] Ali O B, Murat E. Improved assignment with ant colony optimization for multi-target tracking. *Expert Systems with Applications*, 2011, 38(8):9172-9178
- [6] Christina J H, David M, Alex O H. Position representations lag behind targets in multiple object tracking. *Vision Research*, 2011, 51(17):1907-1919
- [7] Sarda W W, Martin D A. Toward multidimensional assignment data association in robot localization and mapping. *IEEE Transactions on robotics*, 2006, 22(2):360-365
- [8] Martin P, Sergio N, Hector C. A survey on parallel ant colony optimization. *Applied Soft Computing*, 2011, 11(8):5181-5197
- [9] Adleman L. Molecular computation of solutions to combinatorial problems. *Science*, 1994, 266(5187):1021-1024
- [10] Marc G A, Daniel M, Alfonso R P. A P system and a constructive membrane-inspired DNA algorithm for solving the Maximum Clique Problem. *Biosystems*, 2007, 90(3):687-697
- [11] Sun Y H, Chao W H, Hsin H C. A DNA-based graph encoding scheme with its applications to graph isomorphism problems. *Applied Mathematics and Computation*, 2008, 203(2):502-512
- [12] Moritz B, Chrystopher L N. Communication and complexity in a GRN-based multicellular system for graph colouring. *Biosystems*, 2008, 94(1):28-33
- [13] 杨磊,黄启鑫,李肯立等. 一种基于 DNA 计算的指定结点路由算法. *计算机学报*, 2009, 32(12):2373-2381
- [14] Andre L, Christoph R, Wolfgang B, et al. Cryptography with DNA binary strands. *Biosystems*, 2000, 57(1):13-22
- [15] Michael H. Fast parallel molecular solutions for DNA-based supercomputing: the subset-product problem. *BioSystems*, 2005, 80:233-250
- [16] Sharif U. A DNA-based computing method for solving control chart pattern recognition problems. *Journal of manufacturing science and technology*, 2010, 3(4):293-303

A multidimensional assignment data association algorithm based on DNA computing

Liang Bing, Feng lin

(School of Innovation Experiment, Dalian University of Technology, Dalian 116024)

Abstract

The DNA computing was applied to investigation of the data association, a NP-hard mathematical problem in multitarget tracking, and a multidimensional assignment data association algorithm was realized through utilizing the super parallelism of DNA computation. The algorithm uses the DNA molecular chains containing the palindromic sequence of the restriction endonuclease *HaeIII* to encode the observed sight lines. All target association combinations are obtained through the operation of connection and annealing. In order to select the optimal multidimensional data association solution, the techniques of DNA hairpin structure detection and gel electrophoresis are used to remove the target association combinations that do not satisfy the constraints. The results of the theoretic analysis show that the DNA computation based multidimensional assignment data association algorithm needs the DNA molecular chain number of $O(nk!/(c-m)!)$ (the space complexity), and decreases the time complexity from exponential to $O(mn)$, where n and m are the sensor number and target number respectively, k is the target localization number.

Key words: DNA computing, target tracking, multidimensional assignment problem, data association algorithm