

一种新的体视对应方法

李象霖 魏永斌 张妙兰

(中国科技大学研究生院 北京 100039)

摘要 选择特征点作为匹配特征,采用视差估计方法,在满足一定匹配准则情况下,可以得到一约束最小化能量函数。模拟实验结果表明,使用模拟退火算法,可使能量函数达到全局最小,从而实现立体匹配。

关键词 体视对应,立体匹配,视差估计,模拟退火

1 引言

基于特征的立体视觉研究的主要内容是特征匹配问题,即寻找同一景物在左、右两幅投影图像上的特征对应关系。在已有的研究分析中,被采用的特征包括特征点^[1]、特征直线^[2]、小平面对景物结构中其他信息以及若干种特征的组合^[3]。在实际研究工作中,多以选择特征点作为匹配特征。

确定特征点匹配关系的步骤一般分为两步:第1步,在左、右图像上选择特征点集合;第2步,寻找和确定两个集合中特征点之间的一一对应关系。由于景物中各物体的空间位置关系复杂,而造成的相互遮挡,不连续以及噪声的影响等,给特征匹配问题带来很大困难。在已有工作的基础上,本文采用视差^[4]估计的方法,选择适当数量的特征点,依据一定的匹配准则,把对应问题归结为一能量函数。于是确定特征点对应问题就变为求能量函数的约束最小化问题。具体实现时,我们采用模拟退火(Simulated Annealing, SA)算法,并与Hopfield神经网络(HNN)方法做了比较。本文最后给出模拟实验结果。

2 横向位移(Epipolar)约束和视差

在立体视觉的研究中,为获得立体图像,一般假设景物静止,摄像机(即观察者)沿一直线运动,而且只沿横向直线运动,没有纵向位移和旋转。摄像机的光轴方向与运动方向垂直。我们把摄像机运动所沿的直线叫做基线(Baseline)。在基线上选择一些观测点,如 O_1, O_2, O_3, \dots ,从这些观测点观测空间静止景物中同一可视点 p ,其空间坐标为 $(x, y,$

1993-08-02 收到, 1994-05-10 定稿

李象霖 男, 1938年生, 副教授, 现从事图像处理、三维运动分析方面的科研与教学工作。

魏永斌 男, 1967年生, 现从事图像处理和模式识别方面的科研开发工作。

张妙兰 女, 1940年生, 副教授, 现从事数字信号处理、图像处理方面的科研与教学工作。

z), P 点与基线上各观测点之间的连线构成一个平面, 被叫做 Epipolar 平面. 这个平面与投影图像平面的交线被叫做 Epipolar 直线. 可以知道, 物体上可视点 P 在各图像上的投影点 P_1, P_2, P_3, \dots , 均在 Epipolar 直线上. 依据空间点与图像平面上投影点之间的几何关系, 可知投影点 P_i , 在第 i 幅投影图像上的横向坐标为 Y_i , 则空间点在左、右图像上投影点的横向位移之差为

$$d = F \frac{D}{z}, \quad (1)$$

其中, D 为左、右两个观测点之间的距离; F 为焦距, 即观测点至图像的距离; d 被称为视差. 我们把摄像机运动只有横向位移方向的这样一种约束, 叫做 Epipolar 约束^[5]. 在这一约束条件下的视差表示物体上同一可视点在左、右图像上成像点的位置差异.

3 对应问题的约束最小化解释

假设在左、右图像上的特征点集合已被选择确定. 在左图像上为 $\{l_i\}, i = 1, 2, \dots, M$, 包括 M 个特征点; 在右图像上为 $\{r_j\}, j = 1, 2, \dots, N$, 包括有 N 个特征点. 如果 $\{l_i\}$ 中的点与 $\{r_j\}$ 中的点是匹配的, 它们之间应该满足以下关系:

- (1) 对应点之间应在一定程度上保持某种定量特征(如灰度)的大致相同;
- (2) 在 $\{l_i\}$ 中的一个特征点不应与 $\{r_j\}$ 中的多个特征点匹配;
- (3) 总体上讲, 视差场的变化是平滑的.

于是, 我们可以利用左图像上 M 个特征点 l_i 和右图像上 N 个特征点 r_j , 建立一个 $M \times N$ 的二维状态空间 C , 当 l_i 与 r_j 匹配对应时, $C_{ij} = 1$; 否则, $C_{ij} = 0$. 因此, 我们将两个特征点集合匹配时应该满足的上述关系作为匹配准则. 具体地说, 选择左图像上的点 l_i, l_k , 右图像上的点 r_j, r_l , 则有

准则 1 当 l_i 与 r_j 的灰度值比较接近时, C_{ij} 可以为 1;

准则 2 对每一个 i, l , $C_{ij}(j = 1, 2, \dots, N)$ 和 $C_{kl}(k = 1, 2, \dots, M)$ 只能有一个为 1, 即

$$\sum_{j=1}^N C_{ij} = 1, \quad \sum_{k=1}^M C_{kl} = 1. \quad (2)$$

准则 3 对应特征点对之间应该相互兼容, 即 $H_{iikl} \cdot C_{ij} \cdot C_{kl}$ 尽量大, 这里 H_{iikl} 表示对应点对 (l_i, r_j) 和 (l_k, r_l) 之间的相容程度, H_{iikl} 愈大, 对应点对愈相容.

在立体视觉中的视差是立体视差. 所谓视差平滑是指空间两个可视点在左、右两图像上成像点的位置接近, 视差越接近表示越相容. 比如, 左图像上的点 l_i 与 l_k 之间、右图像上的点 r_j 与 r_l 之间距离较远, 它们之间无所谓相容与否, 也就不考虑它们之间的相容度量. 因此, 我们取下面的公式做为相容度量, 即

$$H_{iikl} = \left[\frac{2}{1 + (\Delta d)^{k_1}} - 1 \right] \cdot \frac{1}{1 + D^{k_2}}, \quad (3)$$

其中 $\Delta d = |d_{ij} - d_{kl}|$, 这里 d_{ij} 是 l_i 与 r_j 之间的视差, d_{kl} 是 l_k 与 r_l 之间的视差, Δd 是视差的差; $D = \max[D_{i_k}, D_{j_l}]$, 这里 D_{i_k} 是左图像上点 l_i 与 l_k 之间的距离, D_{j_l} 是右

图像上点 r_j 与 r_l 之间的距离; k_1 与 k_2 为待定系数, 确定时与图像内容有关, 一般 $k_1, k_2 > 0$ 。从(3)式可以看出, 当特征点之间距离接近, 而且视差相似时, $H_{ijk_l} \rightarrow 1$; 当特征点较近, 而视差不相似时, $H_{ijk_l} \rightarrow -1$; 当特征点较远时, $H_{ijk_l} \rightarrow 0$ 。这正是对准则 3 的描述。因此我们可以用下面公式表示:

$$\sum_{i=1}^M \sum_{j=1}^N \sum_{k=1}^M \sum_{l=1}^N H_{ijk_l} \cdot C_{ij} \cdot C_{kl}, \quad (4)$$

此式最大为 1, 只有当视差平滑、对应点对的灰度相近时成立。

对于准则 2, 我们可以使用下式描述:

$$\sum_{i=1}^M \left(\sum_{j=1}^N C_{ij} - 1 \right)^2 + \sum_{j=1}^N \left(\sum_{i=1}^M C_{ij} - 1 \right)^2, \quad (5)$$

此式只有当 $C_{ij} = 1$ 时, 才为最小, 最小值为零。当 l_i 与 r_j 的灰度值接近时, C_{ij} 可以为 1, (5)式达到最小。因此这正是对准则 2 的具体体现。

根据以上对三个准则的描述, 我们可以把对应问题归结为用一个约束能量函数 E 表示, 即

$$E = -\frac{A}{2} \sum_{i=1}^M \sum_{j=1}^N \sum_{k=1}^M \sum_{l=1}^N H_{ijk_l} \cdot C_{ij} \cdot C_{kl} + \frac{B}{2} \sum_{i=1}^M \left(\sum_{j=1}^N C_{ij} - 1 \right)^2 + \frac{B}{2} \sum_{j=1}^N \left(\sum_{i=1}^M C_{ij} - 1 \right)^2, \quad (6)$$

此式中的第 1 项为准则 3; 第 2, 3 项为准则 2; A, B 为参数; $C_{ij} = 1$ 的必要条件是 l_i 与 r_j 的灰度相似, 这正是对准则 1 的体现。若(6)式表达的能量函数 E 达到最小, 便说明 $\{l_i\}$ 与 $\{r_j\}$ 两集合中的点匹配。

如果 C_{ij} 的状态发生变化, 所引起的总能量变化为

$$\frac{\partial E}{\partial C_{ijk}} = -\sum_{k=1}^M \sum_{l=1}^N (AH_{ijk_l} - B\epsilon_{ik} - B\delta_{il})V_{kl} - 2B, \quad (7)$$

式中 V_{kl} 表示 k, l 两个特征点的状态。在应用真实图像进行对应时还发现, 即使在基线很短的情况下, 由于可视点的变化, 也会引起景物表面反射角的变化, 使得同一可视点在不同的投影图像上表现为不同的灰度值; 再加上摄像机特性的不稳定, 校正不精确, 特征检测算子不理想以及不可避免的遮挡等原因, 有些特征点根本不存在对应点。为此, 我们把准则 2 修改为: 一个特征点与另一图像中不多于一个的特征点相匹配, 即对每一个 $i, l, C_{ij}(j = 1, 2, \dots, N)$ 和 $C_{kl}(k = 1, 2, \dots, M)$ 最多有一个为 1, 即

$$\left. \begin{aligned} \sum_{i=1}^M \sum_{j=1}^N \sum_{\substack{k=1 \\ k \neq i}}^M V_{ij}V_{ki} &= 0, \\ \sum_{i=1}^M \sum_{j=1}^N \sum_{\substack{l=1 \\ l \neq j}}^N V_{ij}V_{il} &= 0. \end{aligned} \right\} \quad (8)$$

相应的总能量函数为

$$E = -\frac{A}{2} \sum_{i=1}^M \sum_{j=1}^N \sum_{k=1}^M \sum_{l=1}^N H_{ijk_l} \cdot V_{ij} \cdot V_{kl}$$

$$\begin{aligned}
& + \frac{B}{2} \sum_{i=1}^M \sum_{j=1}^N \sum_{\substack{k=1 \\ k \neq i}}^M V_{ij} \cdot V_{kj} + \frac{B}{2} \sum_{i=1}^M \sum_{j=1}^N \sum_{\substack{l=1 \\ l \neq j}}^N V_{ij} \cdot V_{il} \\
& + \frac{C}{2} \left[\sum_{i=1}^M \sum_{j=1}^N V_{ij} - \min(M, N) \right]^2.
\end{aligned} \tag{9}$$

在这种情况下,如果 C_{ij} 的状态发生变化,所引起的总能量变化为

$$\begin{aligned}
\frac{\partial E}{\partial C_{ijk}} = & - \sum_{i=1}^M \sum_{j=1}^N [AH_{ijk} - B\delta_{ik}(1 - \delta_{ij}) - B\delta_{il}(1 - \delta_{jk}) \\
& - C]V_{kj} - C\min(M, N).
\end{aligned} \tag{10}$$

(9)式体现了三个匹配准则。(10)式是根据准则函数得到的。因此只要能量函数达到最小,即被认为达到了匹配,实现了特征点一一对应关系。

对于这样一个约束最小化问题,由于状态空间比较庞大,使用传统的搜索方法几乎难以实现使 E 达到最小。我们采用 SA 算法^[9],使用启发式随机搜索,加入可变大小的随机扰动,可使求解结果趋于全局最小。

4 模拟退火算法

由 Metropolis 提出的 SA 算法是模拟经过加热的物理系统在冷却过程中粒子的迁移行为,随着逐渐降温迭代,粒子逐渐形成低能态的晶格,并使能量达到全局最小。

我们把对应问题归结为一个组合优化问题,使用 SA 算法实现能量函数的最小化。该算法主要涉及两个问题:

(1) 初始温度 T 的选择 T 的选择应使得相应的概率分布与 $T \rightarrow \infty$ 时较为接近。虽然 T 越高越好,但太高了费时间,也无意义。Aarts 等人^[10]建议选择 T_0 ,使其满足一定的接受比例,在状态空间均匀地产生几个扰动。SA 算法的理想情况是当 $T \rightarrow 0$ 时,温度下降要足够慢。为提高算法的运行速度,采用分段修正办法:

当 $T_n > 1$ 时, $T_n = \alpha T_n$;

当 $T_n \leq 1$ 时,采用广为接受的温度下降策略

$$T_n = \frac{C}{\ln(n + n_0 + 1)}, \tag{11}$$

其中 C, n_0 为参数,它们保证温度 $T_n = 1$ 处的连续下降。

(2) 初始状态 由于立体视觉问题中左、右图像上特征点对满足 Epipolar 约束条件,并且对一定的摄像机配置,视差也有一定的范围。另外,候选匹配点的灰度应该一致(准则 1)。因此初始状态考虑所有可能的匹配,对所有满足以上条件的特征点均视为潜在的匹配。对于摄像机沿一直线水平移动这样一种配置,我们选择视差范围为一狭长的水平带 ($W \times H$),视差范围有一定的高度是出于摄像机校正不够准确和量化误差的考虑。

在运行中,我们还对算法的性能做了改进,即在退火过程中跟踪记录能量小的状态,并以此作为最后的结果。

5 模拟实验结果

我们产生了多个模拟立体图像, 并随机抽取特征点。采用 SA 算法进行计算。计算时取 $A = 1.0$, $B = 2.0$, $k_1 = 0.1$, $k_2 = 0.01$ 。

(1) 在一 120×120 的图像上, 随机抽取 50 个特征点, 其分布图见图 1(a)。采用 SA 算法全部正确匹配, 而采用 (HNN) 方法只正确匹配 42 个点, 见表 1。实验用模拟图像的真实视差图见图 1(b), 由实验结果产生的视差图见图 1(c)。实验结果表明, SA 算法比 HNN 算法可靠。

(2) 在一 64×64 大小的图像上, 随机抽取特征点为 $\langle 40, 39 \rangle$, 即左图像上抽取 40 个特征点, 右图像上抽取 39 个, 特征点分布见图 2(a)。采用 SA 算法, 匹配成功 38 点, 而且

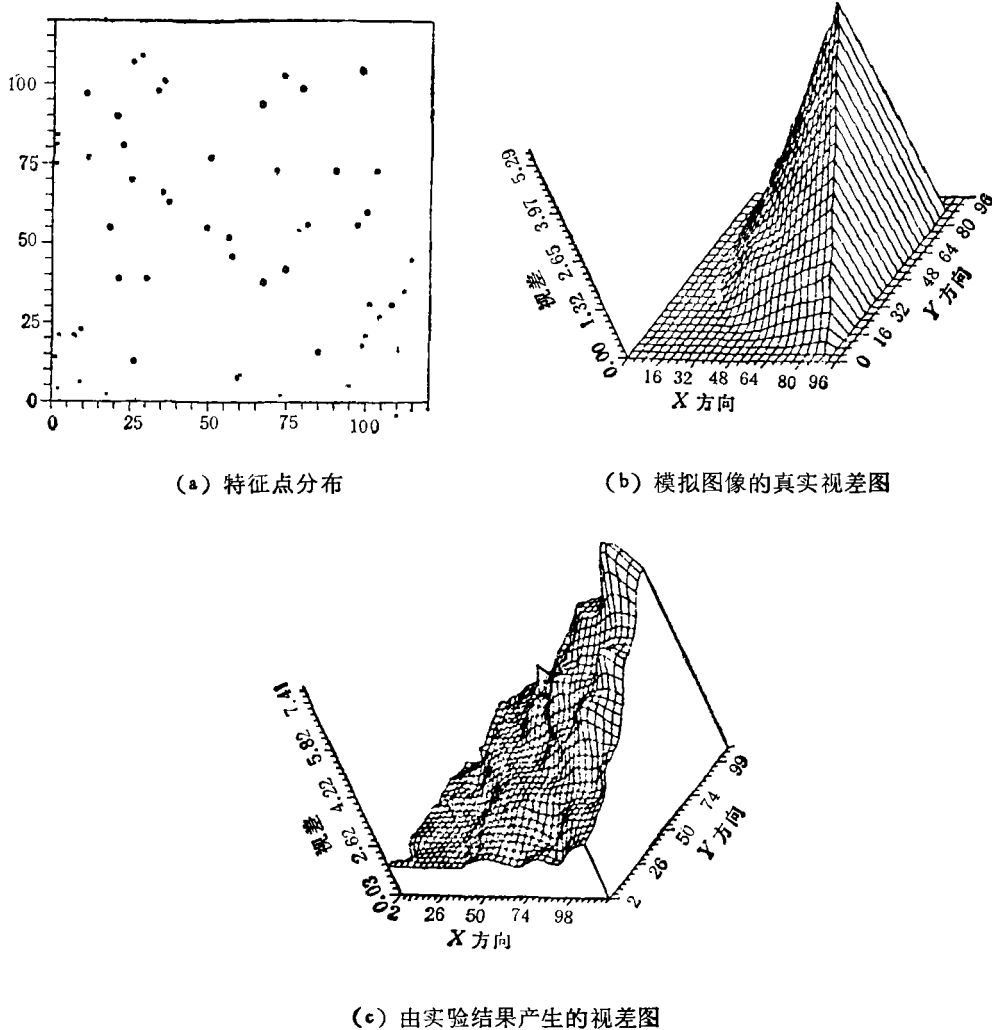


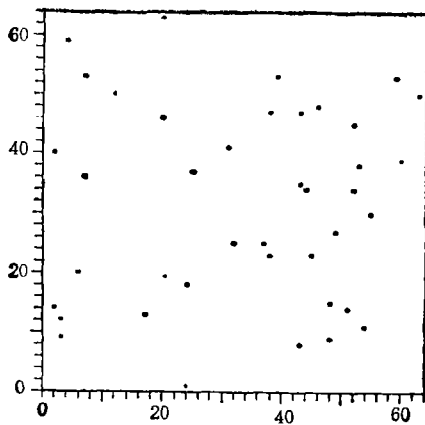
图 1

表 1

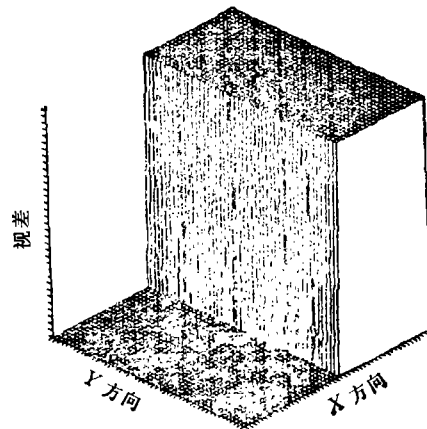
序号	X_1	Y_1	X_r	Y_r	视差	结果	序号	X_1	Y_1	X_r	Y_r	视差	结果
1	2	14	1	14	1	SH	26	54	77	50	77	4	SH
2	2	81	1	81	1	S	27	59	52	56	52	3	SH
3	3	84	1	84	2	S	28	60	46	57	46	3	SH
4	3	75	1	75	2	SH	29	70	38	67	38	3	SH
5	4	4	2	4	2	SH	30	71	94	66	94	5	SH
6	4	21	2	21	2	SH	31	75	2	73	2	2	SH
7	9	21	7	21	2	SH	32	76	73	71	73	5	SH
8	11	6	9	6	2	SH	33	77	42	74	42	3	SH
9	11	23	9	23	2	SH	34	79	103	73	103	6	SH
10	12	97	10	97	2	SH	35	85	56	81	56	4	SH
11	13	77	11	77	2	SH	36	85	99	79	99	6	SH
12	20	55	18	55	2	SH	37	87	16	85	16	2	SH
13	23	39	21	39	2	SH	38	95	73	90	73	5	SH
14	23	90	20	90	3	SH	39	97	5	95	5	2	SH
15	25	81	22	81	3	SH	40	102	18	99	18	3	SH
16	28	13	26	13	2	SH	41	102	56	97	56	5	S
17	28	70	25	70	3	SH	42	103	21	100	21	3	S
18	28	107	25	107	3	S	43	104	31	101	31	3	SH
19	31	109	28	109	3	S	44	105	60	100	68	5	SH
20	32	39	30	39	2	SH	45	106	185	98	105	8	S
21	36	98	33	98	3	SH	46	100	27	104	27	3	SH
22	38	66	35	66	3	SH	47	109	73	103	73	6	SH
23	39	101	35	101	4	SH	48	111	31	108	31	3	SH
24	40	63	37	63	3	SH	49	116	35	112	35	4	SH
25	52	55	49	55	3	SH	50	118	45	114	45	4	S

(S:SA 方法匹配成功 H: HNN 方法匹配成功)

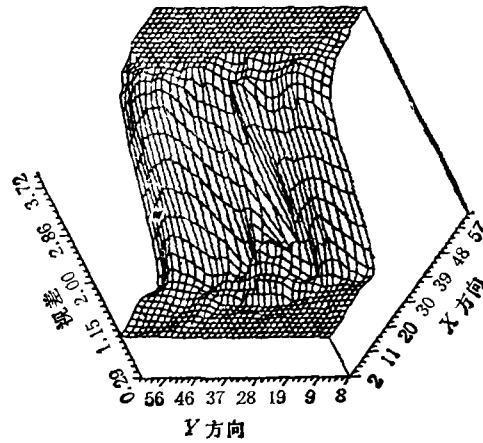
完全正确。采用 HNN 方法有 39 点匹配,其中一个为误匹配,结果见表 2。模拟图像的视差图和由实验结果产生的视差图见图 2(b) 和图 2(c)



(a) 特征点分布



(b) 模拟图像的真实视差图



(c) 重建视差图

图 2

表 2

序号	X_l	Y_l	X_r	Y_r	视差	结果	序号	X_l	Y_l	X_r	Y_r	视差	结果
1	2	14	1	14	1	SH	21	43	8	39	8	4	SH
2	2	40	1	40	1	SH	22	43	35	39	35	4	SH
3	3	9	2	9	1	SH	23	43	47	39	47	4	SH
4	3	12	2	12	1	SH	24	44	34	40	34	4	SH
5	4	59	3	59	1	SH	25	45	23	41	23	4	SH
6	6	20	5	20	1	SH	26	46	48	42	48	4	SH
7	7	0	6	0	1	SH	27	48	9	44	9	4	SH
8	7	36	6	36	1	SH	28	48	15	44	15	4	SH
9	7	53	6	53	1	SH	29	49	27	45	27	4	SH
10	12	50	11	50	1	SH	30	51	14	47	14	4	SH
11	17	13	16	13	1	SH	31	52	34	48	34	4	SH
12	20	46	19	46	1	SH	32	52	45	48	45	4	SH
13	20	63	19	63	1	SH	33	53	38	49	38	4	SH
14	24	1	23	1	1	SH	34	54	11	50	11	4	SH
15	24	18	23	18	1	SH	35	55	38	51	38	4	SH
16	25	37	24	37	1	SH	36	59	53	55	53	4	SH
17	37	25	33	25	4	SH	37	60	39	56	39	4	SH
18	38	23	34	23	4	SH	38	63	58	59	58	4	SH
19	38	47	34	47	4	SH	39	32	25	32	25	0	H
20	39	53	35	53	4	SH							

(S:SA 方法匹配成功, H:HNN 方法匹配成功)

(3) 为检验算法对真实图像的有效性,我们以真实图像——电话机进行了实验。采用文献[8]的方法中特征点检测算子得到特征点,然后选取特征点数为〈47,48〉,其中左图像中特征点分布见图3。采用SA算法正确匹配35个点。匹配对应结果见表3。

表 3

序号	X_l	Y_l	X_r	Y_r	视差	结果	序号	X_l	Y_l	X_r	Y_r	视差	结果
1	12	53	8	53	4	S	19	61	30	58	30	3	S
2	24	68	10	67	14	S	20	66	81	55	80	11	S
3	25	81	15	81	10	S	21	67	75	57	75	10	S
4	28	60	18	60	10	S	22	68	63	58	63	10	S
5	29	86	19	86	10	S	23	68	68	57	68	11	S
6	30	57	26	57	10	S	24	71	108	61	108	10	S
7	43	55	33	55	10	S	25	79	110	68	110	11	S
8	43	109	32	108	11	S	26	80	107	73	108	7	S
9	46	67	36	67	10	S	27	81	183	71	183	10	S
10	46	80	36	81	10	S	28	89	110	79	110	10	S
11	46	108	36	108	10	S	29	90	107	80	107	10	S
12	50	78	40	78	10	S	30	93	102	83	102	10	S
13	51	9	48	9	3	S	31	94	83	84	83	10	S
14	51	64	41	64	10	S	32	129	122	118	122	11	S
15	54	110	44	118	10	S	33	132	112	121	111	11	S
16	56	77	45	77	11	S	34	164	106	156	107	8	S
17	58	74	50	76	8	S	35	61	77	50	77	11	S
18	59	63	49	63	10	S							

(S:SA 方法匹配成功)

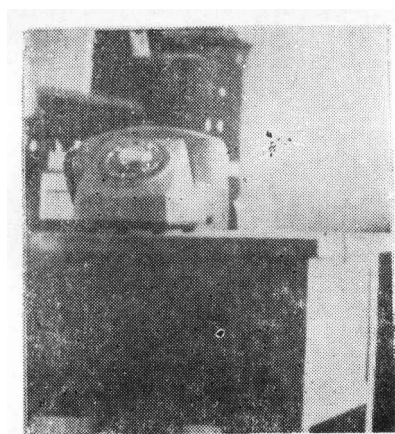


图 3 “电话机”图像特征点分布

6 结 论

从实验结果看出,即使在提取特征时带有噪声,仍能得到比较满意的结果。因此,SA算法大大提高了对噪声的稳健性。另外,由于不连续以及遮挡等原因,能量函数存在许多局部最小,采用SA算法仍能跳出局部最小,趋于全局最小。与HNN方法相比,SA算法稳定可靠,而且没有误匹配。由实验结果重建的视差图与真实视差图之间的误差,是由于特征点稀疏造成的,不是由于匹配引起的。本文实验模拟是在IBM PC/286微机上完

成的,因此比较费时,如果采用工作站,则会大大省时。另外,在参数设定、稠密特征点选择等方面还可做进一步研究。

参 考 文 献

- [1] Marr D, Poggio T. *Science*, 1976, 194:283—287.
- [2] Horaud R, Shordas T. *IEEE Trans. on PAMI*, 1989, PAMI-11(11):, 1168—1180.
- [3] Weng J, Ahuja N, Huang T S. *IEEE Trans. on PAMI*, 1992, PAMI-14(8):806—825.
- [4] Barnard S T, Thompson W B. *IEEE Trans. on PAMI*, 1980, PAMI-2(4):333—340.
- [5] Xu Gang, Tsyji Saburo, Asada Minnorn. *IEEE Trans. on PAMI*, 1987, PAMI-9(2):332—336.
- [6] Kirkpatrick S, Gelatt Jr. C D, Vecchi M P. *Science*, 1983, 220(4598):671—680.
- [7] Aarts E H L, *et al.* *Philips J. Res.*, 1989, 44(11):1168—1180.
- [8] Zhou Y T, Chellappa R. *Stereo Matching Using a Neural Network*, in *Proc. Int. Conf. on ASSP*, 1988, 940—943.

A NEW APPROACH FOR STEREO CORRESPONDENCE

Li Xianglin Wei Yongbin Zhang Miaolan

(*Electrical Engineering Department, Graduate School of USTC, Beijing 100039*)

Abstract The feature points are often extracted as matching feature. And a constraint minimum energy function can be got by using the method of disparity estimation under the condition of certain matching criterion. The results of simulated experiments show that the stereo matching can be obtained. It is possible to escape from local minimum by simulated annealing, and the energy function can reach the global minimum.

Key words Stereo correspondence, Stereo matching, Disparity estimation, Simulated annealing