сообщения
объединенного
института
ядерных
исследований
дубна

E10-94-328

S.Baginyan, S.Baranov, A.Glazov, G.Ososkov

APPLICATION OF DEFORMABLE TEMPLATES FOR RECOGNIZING TRACKS DETECTED WITH HIGH PRESSURE DRIFT TUBES

1994

1 Introduction

The experience of Artificial Neural Network (ANN) applications shows their efficiency for solving of various optimization and recognition problems. ANN are, in particular, used for track finding in experiments in elementary particle physics [2, 3, 4, 5].

Special efforts were made to overcome difficulties caused by the presence of background and noise points that leads to considerable worsening of the accuracy of the statistical characteristics of track parameters. These efforts were carried out in different ways. The cellular automaton was designed especially for preliminary data filtering to speed up the Hopfield ANN evolution on the rest of data [6]. In [7] the modified rotor ANN searches for track points avoiding background points due to a carefully elaborated initial configuration of rotors and robust synaptic weights.

One should realize, however, that in the applications listed above ANN's were used for the recognition of track points only. The explicit estimation of track parameter that is, in fact, the aim of physicists is usually to be done at the next step of data handling procedure by one of the robust version of the least square methods (LSM) (see, for example, [9, 10]).

Therefore the deformable template approach [11] (called also the elastic arm method) looks quite promising, since it combines both the ANN evolution and the robust LSM for explicit estimation of the track parameters.

In the given paper we propose our modification of the deformable template method (DTM) application to the problem of track finding and track parameter determination for data detected with high pressure drift tubes (HPDT) in the design of the ATLAS for the muon spectrometer experiment [1].

Our DTM applications consist of two parts, according to two stages of the study.

The first part relates to the stage of HPDT study on the CERN muon beam (BEAM-TEST, 1992-1993 [1]) with the simplest one-prong events without noise signals, where the main obstacle is the left-right ambiguities for each tube. In the second part more complicated HPDT data are to be handled with noise signals and the arbitrary angle between track and Z-axis. It was shown that the suggested DTM development solves the problem of track recognition and track parameter determination for both noiseless and noise data. Results are obtained on the real beam test data and on data simulating the muon spectrometer on the basis of HPDT.

2 Problem formulation

The HPDT spectrometer consists of modules with two superlayers each. The superlayer (see fig. 1 and [1]) is formed by three layers of HPDT's arranged in honeycomb order. All tubes have the same radius R (15mm). In the middle of every tube there is an anode wire with known ZX-coordinates. Any track passing both superlayers at the moment T_0 produces N signals (in the ideal noiseless case $1 \leq N \leq 6$) registered

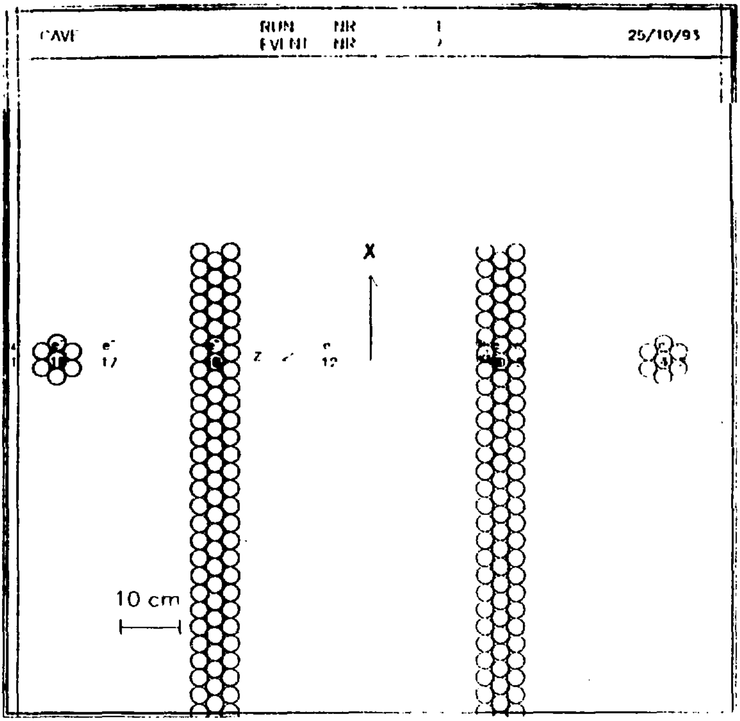

Figure 1.

as drift time t_i, $i = \overline{1, N}$ after T_0. The value of N depends on track parameters and the tube efficiency (about 99%).

The time t_i measured by the i-th tube surrounding the wire with coordinates (Z_i, X_i) is proportional to the radius $R_i \leq R$, which determines the circle on the ZX-plane. All R_i are independent and normally distributed with mean square deviations σ_i.

Thus the mathematical formulation of the problem is to draw the track line tangentially, to all these circles (see fig. 6, 7). Therefore we do not consider cases when $N = 1$ or $N \geq 2$, but signals are registered from the only one (left or right) superlayer.

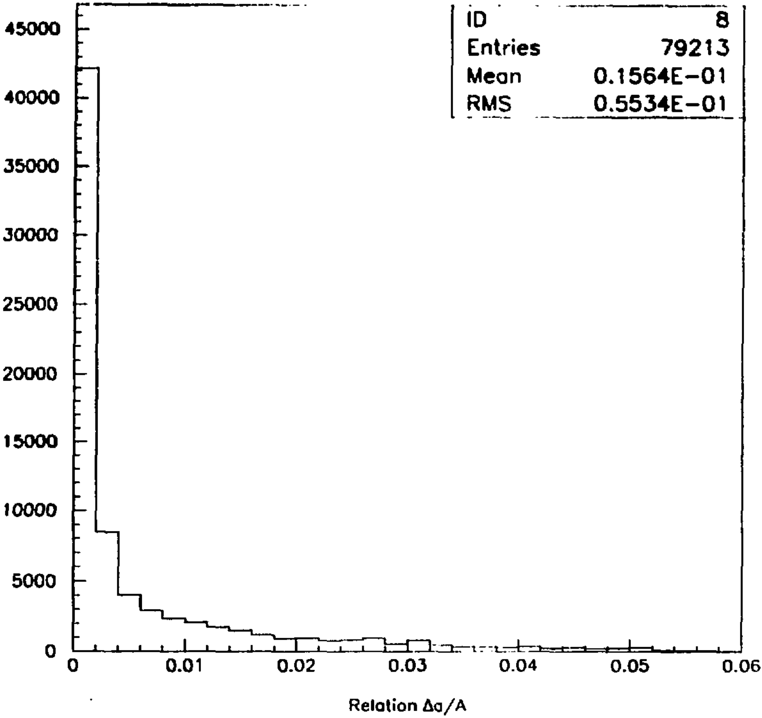

Figure 2.

In the BEAM TEST stage of our study we have to find the track parameters under the following conditions:

- one-prong events only;
- tracks go in plane perpendicular to the tube anode wires (ZX-plane);
- tracks are straight lines, no multiple scattering is present;
- the track slope to the Z-axis is small ($< 2^{\circ}$);

- the time measured by a tube is proportional to the minimal distance from the track to the tube wire in the ZX-plane, errors have gaussian distribution, they are independent in different tubes.

Under these conditions our problem of track recognition and its parameter estimating can be solved by minimizing the *LSM* functional

$$L_1 = \frac{1}{4}\sum\left\{\frac{(D_i(a,b) - R_i)^2}{\sigma_i^2}(S_i + 1)^2 + \frac{(D_i(a,b) + R_i)^2}{\sigma_i^2}(S_i - 1)^2\right\}. \qquad (1)$$

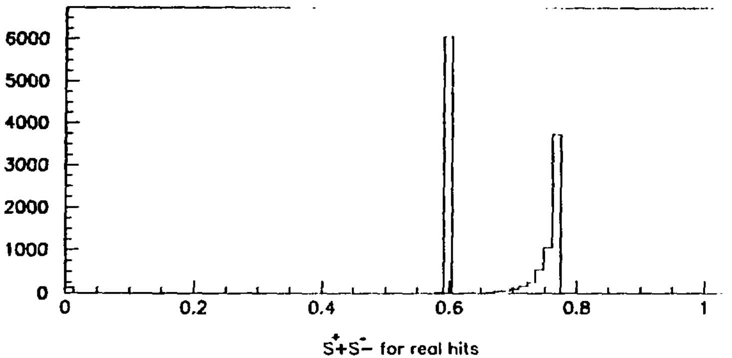

Figure 3.

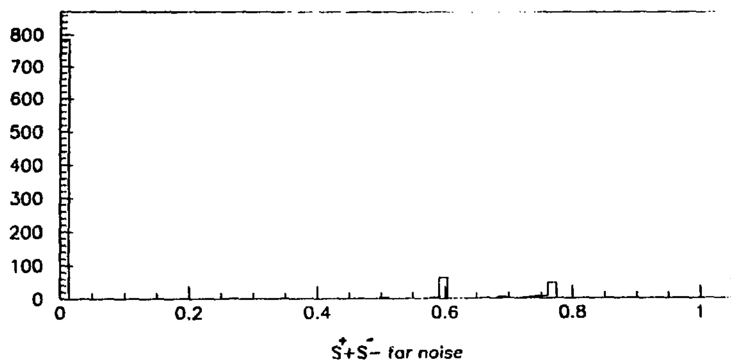

Figure 4.

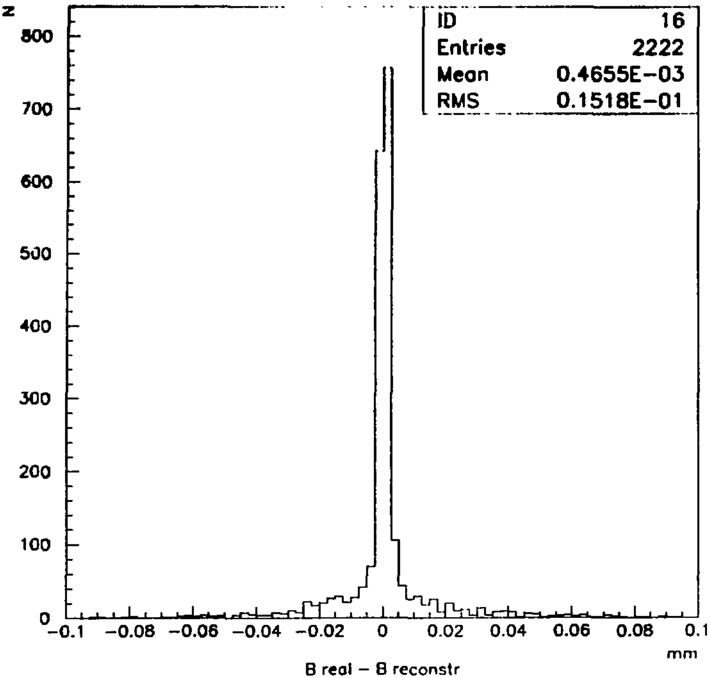

Figure 5.

Here we denote as a, b track parameters, i.e., $X = aZ + b$; $D_i(a, b)$ — distance from the i-th tube anode to the track, for straight tracks it is given by the formula

$$D_i(a,b) = \frac{X_i - aZ_i - b}{\sqrt{a^2+1}}, \tag{2}$$

where Z_i, X_i — coordinates of the wire; R_i — distance measured by the tube; S_i is equal to -1 if the wire lies under the track, or to 1 otherwise (up and down identifications); σ_i — errors of tube measurements. The sum in equation (1) is taken over all triggered tubes.

We want to minimize L_1 with respect to $\{S_i\}$ and a, b.

Due to the above conditions $a \ll 1$, so we can neglect the value a^2 in (2) and for the fixed set of values $\{S_i\}$ the optimal values of a, b for L_1 can be obtained as the solution of the normal system of the linear equations produced by differentiating of L_1 by these two parameters.

In the case of non-fixed S_i the minimization of L_1 is fulfilled by our modification of deformable templates. It'll be described later, since L_1 minimizing is a particular case of more general problem considered in the next stage of study.

In the second stage of our study we have to solve an analogous problem, but for tracks going at an arbitrary angle to the axis Z and in the presence of noise signals.

The latter circumstance leads to a serious difficulty in HPDT data processing, which was not surmounted, for instance, in [8]. Noise measurements are not included also in the functional (1).

Let us include noise signals in our formalism. Whenever the i-th tube produces the signal measured as R_i, we attribute it to a noise measurement, if the following condition holds:

$$\frac{(D_i(a,b) - R_i)^2}{\sigma_i^2} > \lambda. \tag{3}$$

Here a, b are parameters of the true track and λ is the constant (we state it by $\sqrt{\lambda} = 3.5$, which corresponds to the 99.9% confidential level under the assumption of the R_i normality with the standard deviation σ_i).

The variable $S_i = \pm 1$ in L_1 points out how the given track is situated (upward-downward) relative to the tube center. It would be quite reasonable to supplement this variable by a special state for cases of the noise measurements.

At this point of our consideration the conception of neuron is needed (in fact, there is a close connection between neural networks and deformable templates [10, 11]).

We define the neuron as two-dimensional vector $\vec{S}_i = (S_i^+, S_i^-)$ with values $(1,0), (0,1), (0,0)$ corresponding to three situations:

- track goes upward to the i-th tube center;
- track goes dounward to the i-th tube center;
- R_i measurement is noise, i. e., (3) holds.

Now we can generalize the functional L for cases of noise signals in the following expression:

$$L = \sum_{i=1}^{N} \left\{ \frac{(D_i(a,b) - R_i)^2}{\sigma_i^2} S_i^- + \frac{(D_i(a,b) + R_i)^2}{\sigma_i^2} S_i^+ \right\} + \lambda \sum_{i=1}^{N} ((S_i^+ + S_i^-) - 1)^2. \tag{4}$$

Thus the problem of the full track recognition, i. e., all track points finding and track parameter estimating, is reduced to minimizing the functional L with the following constraint for the vectors $\vec{S}_i$:

$$S_i^+ + S_i^- \leq 1. \tag{5}$$

One can see that both functionals(1) and(4) have a form similar to the method of deformable templates [10, 11]. So we can apply the minimizing procedure developed for it, so-called EM-method. It consists of the following steps:

1. Find initial approximation for $a, b, \{\vec{S}_i\}$.

2. Iterate through:

   (a) given a, b, find $\vec{S}_i$ in each tube: $\vec{S}_i^{n+1} = S(a, b, \vec{S}_i^n)$

   (b) from given S_i find a^{n+1}, b^{n+1}

3. Stop if the minimum has been achieved, i. e., $\vec{S}_i^{n+1} = \vec{S}_i^n$, $|a^{n+1} - a^n| < \Delta_{astop}$, $|b^{n+1} - b^n| < \Delta_{bstop}$.

The possible version of the renewal procedure for vectors $\vec{S}_i$ in the iteration process could be based on the stepwise function $S(\{\vec{S}_i\}, a, b)$ of the argument $sign(D_i(a, b))$. However this procedure has an essential disadvantage such that with a big probability the functional L could fall into one of its local minima.

From various tricks to avoid sticking to local minima one can try to use combinatorics to search for the best initial approximation (see step 1 above). However, just in our case of false HPDT signals and the great level of noise this way is not effective.

We choose the way recommended by the deformable template theory [10] to replace the stepwise function $S(\{\vec{S}_i\}, a, b)$ by a sigmoid function. According to this method, we suppose that our system is situated in a thermostat with the temperature T, so the probability to have parameters $\{\vec{S}_i\}, a, b$ is given by the canonical Gibbs distribution:

$$P[\{\vec{S}_i\}, a, b] = \frac{1}{H} e^{-\frac{L(\{S_i\}, a, b)}{T}}, \tag{6}$$

where

$$H = \sum_{\{\vec{S}_i\}, a, b} P[\{\vec{S}_i\}, a, b].$$

Let us denote

$$\frac{(D_i(a, b) - R_i)^2}{\sigma_i^2} = d_i^- \quad , \frac{(D_i(a, b) + R_i)^2}{\sigma_i^2} = d_i^+ .$$

Then, we should calculate the marginal probability distribution $P_m[a, b] = \sum_{\{\vec{S}_i\}} P[\{\vec{S}_i\}, a, b]$. This gives

$$P_m[a, b] = \frac{1}{H} \sum_{\{\vec{S}_i\}} e^{-\frac{1}{T} L} = \frac{1}{H} \sum_{\{\vec{S}_i\}} \prod_i e^{-\frac{1}{T}\left(d_i^- S_i^- + d_i^+ S_i^+ + \lambda((S_i^+ + S_i^-) - 1)^2\right)} =$$

$$= \frac{1}{H} \prod_i \left\{ e^{-\frac{\lambda}{T}} + e^{-\frac{d_i^-}{T}} + e^{-\frac{d_i^+}{T}} \right\} = \frac{1}{H} e^{-\frac{1}{T} E_{eff}},$$

where

$$E_{eff} = -T\sum_i \log\left(e^{-\frac{\lambda}{T}} + e^{-\frac{d_i^-}{T}} + e^{-\frac{d_i^+}{T}}\right).$$

Let's find derivations of E_{eff} with respect to parameters a, b. We denote as x any of a, b. Then

$$\frac{\partial E_{eff}}{\partial x} = \sum_i \frac{e^{-\frac{d_i^-}{T}}\frac{\partial d_i^-}{\partial x} + e^{-\frac{d_i^+}{T}}\frac{\partial d_i^+}{\partial x}}{e^{-\frac{\lambda}{T}} + e^{-\frac{d_i^-}{T}} + e^{-\frac{d_i^+}{T}}}. \tag{7}$$

Comparing with

$$\frac{\partial L}{\partial x} = \sum_i S_i^- \frac{\partial d_i^-}{\partial x} + S_i^+ \frac{\partial d_i^-}{\partial x} \tag{8}$$

we obtain the procedure of parameter $\{\vec{S}_i\}$ renewal in the next step of iteration as follows:

$$S_i^- = \frac{1}{1 + e^{\frac{d_i^- - \lambda}{T}} + e^{\frac{d_i^- - d_i^+}{T}}}, \tag{9}$$

$$S_i^+ = \frac{1}{1 + e^{\frac{d_i^+ - \lambda}{T}} + e^{\frac{d_i^+ - d_i^-}{T}}}. \tag{10}$$

From (9–10) one can see the sigmoid shape of the function $S(\{\vec{S}_i\}, a, b)$. The higher the temperature T, the larger uncertainty of the track passage in respect to the i-th tube (i. e., $S_i^+ \approx S_i^-$), but if the temperature T is small one has $S_i^+ \approx 1$, $S_i^+ \approx 0$, or vice versa. Thus, decreasing sequentially the temperature (so-called "simulated annealing" procedure, see, for example, [12]) one can force the functional L to approach its global minimum in the iterational process.

3 Algorithm implementation

3.1 Problem linearization

It is obvious that in the case of the arbitrary track slope to the Z-axis the value a^2 in (2) is no more negligible and the problem of L minimization is not linear. However we found a way to obtain the new values of parameters (a, b) in the second step of the EM-method by solving the equation system linear in respect to a, b.

The idea is the following: to use the *a priori* information about the centers of all tubes in order to fit a straight line to the centers of these tubes, producing signals. The parameters a, b of this line remain fixed during the given step of T of simulated annealing iteration while the increments $\Delta a, \Delta b$ are treated as fitted variables. The formula (2) changes as

$$D_i(a,b) = \frac{X_i - (a+\Delta a)Z_i - (b+\Delta b)}{\sqrt{(a+\Delta a)^2+1}}, \tag{11}$$

It seems there is the non-linearity due to Δa in the denominator. However, the ratio $((a+\Delta a)^2+1)/(a^2+1) = 1 + \Delta a(2a+\Delta a)/(a^2+1)$ tends to 1 for vanishing Δa (see also the distribution of $\Delta a(2a+\Delta a)/(a^2+1)$ denoted as $\Delta a/A$ in fig. 2 for the set of modelled events). Therefore, we can neglect Δa in the denominator of (11).

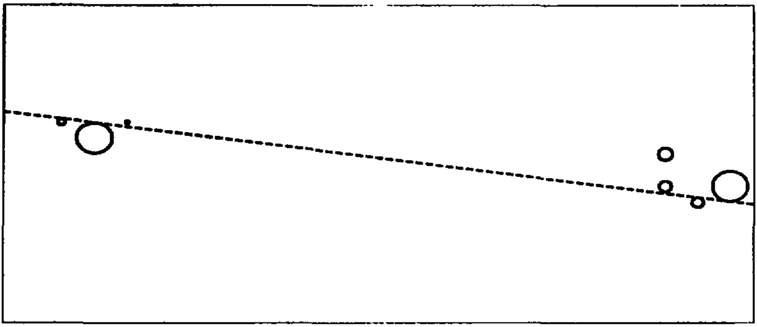

Figure 6.

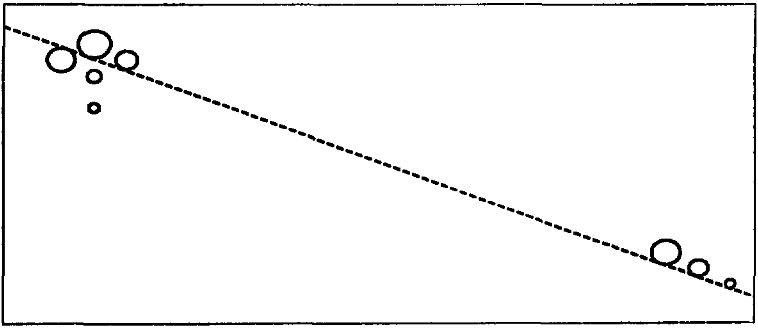

Figure 7.

3.2 The algorithm of the global minimum search

The global minimum search of the functional L is implemented in three steps.

Step 1. The iterational process of the EM-method is going on with the big temperature and the big value of the constant λ in the noise criterium (3).

Step 2. The straight line is refitted with the starting values obtained in step 1. The temperature T and the noise cut-off both are decreased and EM-iterations are repeated till the convergence or exhausting of the iteration limit.

Step 3. Analogous refitting with the temperature close to zero and $\sqrt{\lambda} = 3.5$.

At the end of every step the minimum value of the functional L with fitted parameters $\{\vec{S}_i\}, a, b$ is calculated and compared with L_{cut}. The cut-off L_{cut} was calculated in advance from the distribution of L for simulated events. As soon as $L < L_{cut}$ in some step we finish the whole procedure considering the given event (a track) as the recognized one. Otherwise we go to the next step.

There is no guarantee that the algorithm described above must definitely find the global minimum of the functional L_2. The efficiency of our algorithm varies between 75% and 85% in different series of event simulated by the program GEANT [13].

To increase the efficiency of our algorithm it was completed by auxiliary heuristic procedures elaborated after the careful study of many various situations of mutual dispositions of useful and noise signals in real and simulated events.

- Very reliable is the method of sequential rejecting (MSR) of every measurement R_i from the full set of $\{R_i\}$ assuming R_i to be a noise whenever such an assumption is attainable (senseless to reject the only measurement in one of superlayers). In case of the true assumption the MSR gives a considerable decrease of the L-minimum. However the MSR is too time-consuming , so we apply it in our algorithm not at the beginning, but only when two following heuristic failed.

- Among different reasons of $L > L_{cut}$ one appears often when the starting line is fitted wrongly due to the presence of an outlier in one of superlayer , i. e., the noise tube is situated far from the true track position (see fig. 8). Our trick is to split the ZX-plane into two semiplanes by this starting line and to applay our iterations to both semiplanes separately.

- If in both semiplanes one obtains $L > L_{cut}$ again, we repeat the EM-iterations, but with the starting line previously changed. Namely, we attempt to apply small parallel shifts by changing the parameter b : $b \pm b_{tol}$ or, if necessary, small variations of the parameter a : $a \pm a_{tol}$. Constant a_{tol}, b_{tol} were found experimentally.

4 Results

The solution of the beam track reconstruction in the framework of the BEAM TEST methodical experiment was very successful (100% efficiency) and allowed one to determine the space resolution of HPDT and to study the possibilities of the accurate reconstruction of all anode wire positions.

The fortran program which implements our algorithm for minimizing the functional (4) was tested on different series of events simulated by the famous program GEANT [13]. The program was written in the framework of the CERN management system CMZ [14]. The distribution of data processing time for one event (VAX-8350

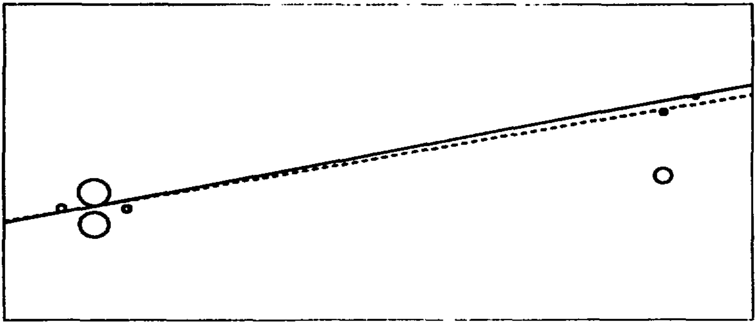

Figure 8.

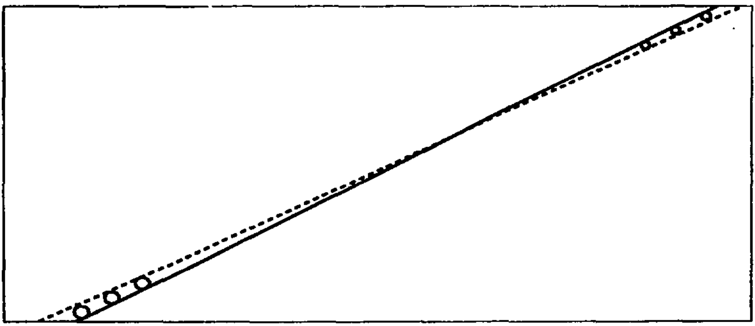

Figure 9.

computer) is given in fig. 10. The number of events in each series varied from 2 to 14 thousands. The efficiency of correctly recognized events after applying the heuristics listed above was kept in the satisfactory range 96%–99%.

From a physicist's point of view the most important track parameter is the absolute term b. The distribution of the difference of the true and estimated value of b is shown in fig. 5.

The distribution of such values of $\{\vec{S}_i\}$ that correspond to true measurements is demonstrated in fig. 3. As one can see, the number of the true measurements qualified by our program as false ones is very small. However, as is seen from fig. 4, where an analogous distribution for the noise measurement is shown, our algorithm cannot suppress all noise.

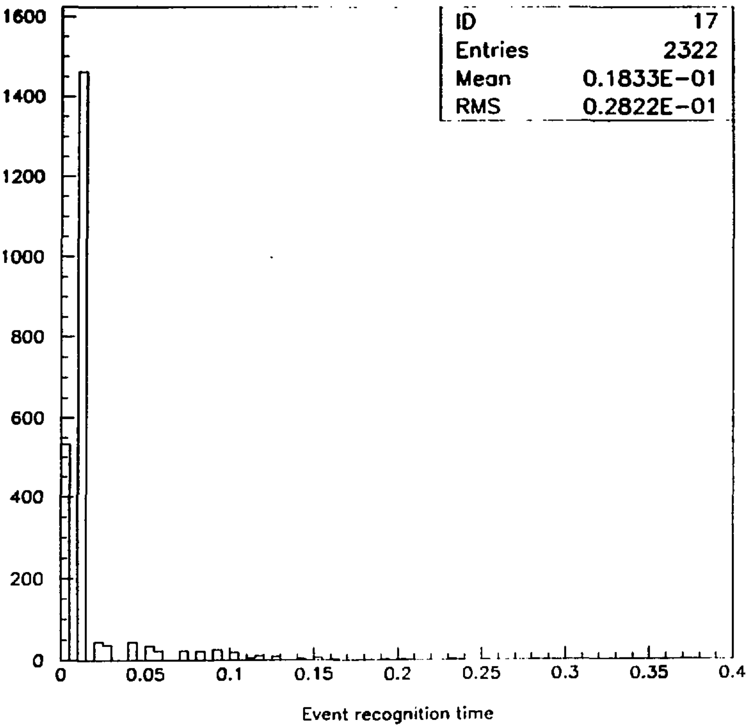

Figure 10.

Examples of the successful recognition by our algorithms are given in figs. 6, 7. The event presented in fig. 6 consists of four true measurements and three noise ones. The event of fig. 7 contains five true measurements and also three noise measurements.

Two examples of events that our algorithm couldn't crunch are demonstrated in figs. 8, 9. Real simulated tracks are denoted by continuous lines, tracks recognized by our program are denoted by dotted lines. In the event in fig. 8 from 7 measurements four belong to noise, so the probability of the correct recognition is vanishing. The event in fig. 9 shows the configuration of measurements with the maximum uncertainty. To handle such events our algorithm should be completed by a special auxiliary procedure (what we are going to do).

This work was completed due to the support of the Russian Foundation for Fundamental Research.

References

[1] ATLAS Internal Note. MUON-NO-029. 28 October 1993.

[2] C. Peterson and B. Söderberg, A New Method for Mapping Optimization Problems onto Neural Networks, IJNS 1,3 (1989).

[3] C. Peterson, Neural Networks and High Energy Physics, Lund preprint LU TP 90-6.

[4] L. Gislen et al., Rotor Neurons — Basic Formalism and Dynamics, LU TP 91-21.

[5] G. Stimpfl-Abele and L. Carrido, "Fast Track Finding with Neural Nets", UAB-LFAE 90-06 (1990).

[6] A. Glazov, I. Kisel, E. Konotopskaya, G. Ososkov, Filtering tracks in discrete detectors using a cellular automaton. Nucl. Instr. and Meth. A329 (1993) 262-268.

[7] S. Baginyan, A. Glazov, I. Kisel, E. Konotopskaya, V. Neskoromnyi, G. Ososkov, *Tracking by modified rotor model of neural network.* Comp. Phys. Commun., 79 (1994) 165–178.

[8] V. Kotov et al., JINR Commun. E10-93-86, Dubna, 1993.

[9] G. Ososkov, Proc. 2nd Intern. Tampere Conf. in Statistics., Tampere Univ. Press (1987), 615.

[10] A. Yuille et al., "Robust Statistics, Transparency and Correspondence", Harvard Robotics Laboratory Technical Report No. 90-7.

[11] M. Ohlsson et al., "Track Finding with Deformable Templates - The Elastic Arms Approach", LU TP 91-27.

[12] И. В. Кисель и др., ЭЧАЯ, т. 24 (1993) в. 6, 1551–1595.

[13] GEANT User's Guide (DD/EE/84-1).

[14] A Source Code Management System CMZ. User's Guide and Reference Manual.

Received by Publishing Department
on August 12, 1994.