NDAC/LO/005

1982 Sept 01

WP3100: Preprocessing of IDT Data (Definition)

by L Lindegren

1. Principles and Terminology

The preprocessing of IDT data shall provide instantaneous and simultaneous phase determination of the intensity modulations produced by the several stars observed in a frame. It is achieved by separate harmonic analysis of the photon counts belonging to the different stars, using a unique, assumed attitude motion as reference. Due to the interlacing of observations, it can be assumed that low-order attitude errors (such as a constant offset and a small rate error) will not significantly bias the average modulation parameters. Indeed, if the assumed harmonic signal model is good enough and an efficient estimator used, the process should entail very small information losses but some considerable data compression.

The sequence of photon counts (or equivalent) contained in a frame is called the IDT count record: $\{N_k, k = 1, 2, ..., NSAMP\}$, where NSAMP $\simeq$ 2400 is a fixed number. The non-negative number N_k is the k:th sample. With each sample we may associate a time t_k (reckoned in seconds from TSTART), an object identification number ID_k, and a quality index IQ_k. The last item should tell whether the sample was obtained in analogue or digital (puls counting) mode, and flag suspectedly bad samples (IFOV switching, jet firing, particle event, ...).

Of course the complete arrays $\{t_k\}$, $\{ID_k\}$, $\{IQ_k\}$ are not given on the input tape, but they must be computable from available data, e.g. on the attitude tape (for jet firings) or in a short record accompanying each IDT frame, the frame schedule.

2. Summary of I/O Data

The following table lists the Data Interface Descriptions (DID's) relevant for the IDT Preprocessing:

DID#	SOURCE	DATA INPUT TO IDT PREPROCESSING
9	IDT Data	• frame schedule • (coded) count record
10	Star Catalogue	• astrometric data; B, B-V
11	Ephemerides	• barycentric motion of observer
12	Attitude Reconstitution	• attitude angles
13	LSD Data	• *a priori* large-scale distortion
14	SSD Data	• small-scale distortion table

DID#	DESTINATION	DATA OUTPUT FROM IDT PREPROCESSING
15	IDT Signal Catalogue	• signal parameters and covariances
16	OTF Data	• updates of modulation coefficients and shifts (M_1, M_2, v_2)

3. Process Description

The IDT preprocessing is here broken down into four parts, to be executed in sequence:

(i) collection of input data;
(ii) computation of reference grid coordinates;
(iii) estimation of signal parameters;
(iv) output

3.1. Collection of input data

The process starts by reading one frame from the IDT data tape (although in practice many frames will probably be contained in a single physical record). This consists of the frame schedule and the coded count record. The form of the frame schedule is still undefined and will depend e.g. on the adopted observing strategy, but for illustration let us consider the following format adequate for a simple regular (cyclic) IFOV switching:

IFRAME	= frame identification number	(4 bytes)	
TMID	= frame mid-time	(6 bytes)	
ID(I)	= object identification no.	(4 bytes)	for I = 1 to 10
IFIELD(I)	= field index	(1 byte)	for I = 1 to 10
MODE(I)	= detection mode	(1 byte)	for I = 1 to 10
...	= ...	(1 byte)	for I = 1 to 10
LAST(I)	= last sample of first dwell period on star no. ID(I)	(1 byte)	for I = 1 to 10
	total:	70 bytes	

This assumes that at most 10 stars are observed in a frame, that a fixed record length is required, and that 'bad' samples are not individually flagged. The numbers LAST(I) and ID(I) obviously allow the sequence $\{ID_k\}$ to be reconstructed; similarly we obtain $\{IQ_k\}$, possibly with input also from the attitude tape.

Next, the coded count record (1 byte per sample) is decoded and possibly scaled (for the analogue detection mode) in order to create the count record $\{N_k\}$. This may then be split into separate arrays for the different stars, but the original sample number (k) must be kept track of.

The astrometric and photometric parameters for the stars ID(I) are fetched from the Star Catalogue, which at this stage is identical to the Input Catalogue; from the Ephemerides, the barycentric motion of the observer [ROBS(K), VOBS(K), K = 1 to 3] at three successive frame mid-times (TMID-TFRAME, TMID, TMID+TFRAME); and from the Attitude Reconstitution the attitude angles α_z, δ_z, ω at the same three instances, and the time of any attitude rate discontinuity (jet firing) that may have occurred in the interval.

3.2. Computation of reference grid coordinates $\tilde{G}_k$

The subset of the count record referring to a particular star will be Fourier analyzed with respect to the reference grid coordinates $\{\tilde{G}_k\}$ computed from available data on the star's position, the attitude, and the field distortions (including small-scale distortions, SSD). This will eventually give an estimated phase error ($\hat{\beta}_3$) from which an 'observed' mean grid coordinate is computed as $G_\kappa = \tilde{G}^*_\kappa + \hat{\beta}_3/(2\pi)$, where subscript κ = INT((NSAMP+1)/2) refers to the mid-frame sample ($t_\kappa \equiv$ TMID) and the asterisk signifies a smoothed grid coordinate, i.e. excluding the SSD. From this it is clear that a constant error in $\tilde{G}_k$ will have no effect on the final result, and so the computations need not be rigorous as long as the neglected components are constant (to some 0.1 mas) in the frame and absolutely less than about 0.1 as (in order not to contribute unnecessarily to the slit ambiguity problem). Thus, gravitational light deflection, for instance, need not be included in the computation of $\tilde{G}_k$.

The smoothed grid coordinates at the frame mid-times, $\tilde{G}^*_\kappa$, are thus obtained in analogy with G_{cat} in the Set Solution (see NDAC/LO/001: Set Solution - Definition), only with some simplifying approximations. This applies to stars and asteroids alike. Interpolating between the frame mid-times, we then obtain smoothed grid coordinates for each sample ($\tilde{G}^*_k$); finally, applying the small-scale distortions yields the desired reference grid coordinates: $\tilde{G}_k = \tilde{G}^*_k + \delta G(\eta, \zeta)$.

It may be convenient to process some 10 - 20 frames in a batch, since data related to adjacent frames are anyway needed for the interpolation. This will also permit more accurate interpolation of the grid coordinate (or attitude), especially in presence of discontinuities.

3.3. Estimation of signal parameters

In Annex A to the NDAC Proposal, pp. 27 - 30, two different formulations of the signal model were discussed. Of these, the second formulation [using the parameter vector $\underline{\beta}$, Eqn (A.10.11)] is the more general one and adopted in the following. Thus, the expectation of the IDT sample value is written

$$E(N_k) = \beta_1 + \beta_2[\cos(H_k+\beta_3) + \beta_4\cos(2H_k+2\beta_3) + \beta_5\sin(2H_k+2\beta_3)] \quad (1)$$

where $H_k = 2\pi.\mathrm{frac}(\tilde{G}_k)$.

The estimation process follows the Maximum Likelihood (ML) principle, assuming Poissonian count distribution; some mechanism for automatic rejection of grossly deviating samples must however be incorporated in the iterative adjustment. The basic formulae for the ML fit are (A.10.2) - (A.10.5); see also "Further Details on the Scientific Data Processing" (Lindegren, 1980 Febr 22). The initial approximation $\underline{\beta}^{(0)}$ is computed from an unweighted least-squares solution of the five conventional Fourier coefficients. [In the 1980 note it was suggested that 'statistical weighting' $(N_k+1)^{-\frac{1}{2}}$ be used for the initial approximation. The experiments reported in NDAC/LO/004 (Test of IDT Preprocessing Algorithm) indicated however that the unweighted least-squares solution should be preferred, at least for small N_k.]

The ML fit will result in the 5-vector of estimated signal parameters ($\hat{\underline{\beta}}$) and an estimate of its covariance ($\underline{\Phi}^{-1}$). The latter being a symmetric (5,5)-matrix, the result will thus consist of 20 real numbers per star per frame. Provided the signal model (1) is adequate, these will in principle contain all relevant (photometric and astrometric) information. As indication of the model fit we may add e.g. the chi-square and the number of samples accepted and rejected.

Depending on the formulation of the Location Estimator, the data set ($\underline{\Phi}^{-1}$, $\hat{\underline{\beta}}$) may not be the best output; some equivalent set may be preferred. E.g. (A.10.14) indicates that the information array (normal equations) ($\underline{\Phi}$, $\underline{\Phi}\hat{\underline{\beta}}$) is more convenient. If the square-root information formalism is adopted, the output would be the data equations $(\underline{\Phi}^{\frac{1}{2}}, \underline{\Phi}^{\frac{1}{2}}\hat{\underline{\beta}})$, where $\underline{\Phi}^{\frac{1}{2}}$ is the upper-triangular Cholesky factor. A priori data, $(\underline{G}^{-\frac{1}{2}}, \underline{G}^{-\frac{1}{2}}\tilde{\underline{\beta}})$, are then incorporated by means of the Partial Triangularization algorithm (Annex A, Appendix C) instead of the analytically equivalent (A.10.14). It may be noted that a different indexing of the signal parameters may then be advantageous: e.g. β_5 should be the most frequently needed parameter [probably the signal phase, β_3 in (1)], since its solution from the data equations requires the least number of operations (one division).

3.4. Output

The main output of the IDT Preprocessing consists of the signal parameters with covariances etc, which are stored chronologically on the IDT Signal Catalogue.

For stars whose data in the Star (Input) Catalogue satisfy certain criteria (e.g. brightness, non-multiplicity, known B-V), the estimated parameters $\beta_4 = (M_2/M_1)\cos(v_2)$ and $\beta_5 = (M_2/M_1)\sin(v_2)$ are stored on a separate file together with their covariance, the star identification number, time of observation (TMID), mean field coordinates $(\tilde{\eta}, \tilde{\zeta}, f)$, and colour (B-V). This file is later processed into a 'calibration table' from which a priori values of $(M_2/M_1)\cos(v_2)$ and $(M_2/M_1)\sin(v_2)$ can be computed as functions of time, field coordinates, and B-V. These values are needed with the Location Estimator to obtain improved signal phases and for the detection of multiple stars.

For the photometry, and also in order to monitor the instrument e.g. with respect to the focus, it is also desirable to estimate and output M_1. This can be done directly for very bright stars for which the background is negligible; otherwise it will be necessary to estimate also the background by comparing stars of different brightness. The procedure for this remains TBD, but we assume that the OTF output will consist of the three parameters M_1, $(M_2/M_1)\cos(v_2)$, $(M_2/M_1)\sin(v_2)$ [= OTF(K), K=1,3] and co-variances.

DID 9	OUTPUT FROM : IDT DATA INPUT TO : IDT PREPROCESSING								PAGE : 1 OF 1 DATE : 82.09.14
DESIGNATION	ANNEX A	EXPLANATION	UNIT	RANGE MIN	RANGE MAX	RANGE DEF	NUM ACC ABS	NUM ACC REL	DIGITS
		1. For each frame:							
IFRAME	-	frame identification number	-	1	9.10^8	-	1	-	9
TMID	t_K (p31)	frame mid-time	s	0	2.10^8	-	10^{-6}	-	15
		1.1. For I = 1 to 10:							
ID(I)	-	object identification number	-	0	10^8	0	1	-	8
IFIELD(I)	$(-1)^f$ (p8)	field index	-	-1	1	0	1	-	1
MODE(I)	-	detection mode etc (TBD)	-	0	255	0	1	-	3
LAST(I)	-	last sample of first dwell period on object ID(I)	-	0	9000	0	1	-	4
		1.2. Once per frame:							
NCCR(K), K = 1 to NSAMP	-	coded count record (N_k)	-	0	255	-	1	-	3

DID 10	OUTPUT FROM : STAR CATALOGUE INPUT TO : IDT PREPROCESSING						PAGE : 1 OF 1 DATE : 82.09.14		
DESIGNATION	ANNEX A	EXPLANATION	UNIT	RANGE MIN	RANGE MAX	DEF	NUM ACC ABS	NUM ACC REL	DIG-ITS
		1.1. <u>For each star in frame:</u>							
ID	-	object identification number	-	0	9.10^8	0	1	-	8
RAØ	α_o (p4)	Right Ascension at TEPOCH	rad	$-\pi$	π	-	10^{-11}	-	12
DECØ	δ_o (p4)	Declination at TEPOCH	rad	$-\frac{1}{2}\pi$	$\frac{1}{2}\pi$	-	10^{-11}	-	12
PMRA	$\mu_\alpha \cos\delta_o$ (p4)	Proper Motion in R.A. at TEPOCH (times $\cos\delta_o$)	rad/s	-10^{-11}	10^{-11}	0	10^{-19}	-	9
PMDEC	μ_δ (p4)	Proper Motion in Dec. at TEPOCH	rad/s	-10^{-11}	10^{-11}	0	10^{-19}	-	9
PX	$\tilde{\omega}$ (p4)	trigonometric parallax	rad	-5.10^{-6}	5.10^{-6}	0	10^{-11}	-	6
BMAG	-	blue (B) magnitude	mag	-2	15	99	10^{-2}	-	4
BMV	-	colour index B-V	mag	-1	3	99	10^{-2}	-	4
SDBMV	-	standard deviation of BMV	mag	0	3	99	10^{-2}	-	4

DID 11	OUTPUT FROM : EPHEMERIDES INPUT TO : IDT PREPROCESSING						PAGE : 1 OF 1 DATE : 82.09.14		
DESIGNATION	ANNEX A	EXPLANATION	UNIT	RANGE MIN	RANGE MAX	DEF	NUM ACC ABS	NUM ACC REL	DIG-ITS
		1. For each frame:							
IFRAME	-	frame identification number	-	1	9.10^{8}	-	1	-	9
TMID	t_K (p31)	frame mid-time	s	0	2.10^{8}	-	10^{-6}	-	15
VOBS(K), K=1,3	$\hat{\underline{v}}$ (p5)	velocity of observer (satellite)	m/s	-4.10^{4}	4.10^{4}	-	10^{-2}	-	7
ROBS(K), K=1,3	$-\underline{\rho}_0$ (p5)	barycentric coord. of observer	m	-2.10^{11}	2.10^{11}	-	10	-	11
		1.1. If frame contains asteroid ($ID > 9.10^{7}$):							
IDAST	-	asteroid number = $int(\frac{ID-9.10^{7}}{1000})$	-	1	9000	-	1	-	4
RAST(K), K=1,3	-	barycentric coord. of asteroid at time TMID - $\|RAST - ROBS\|/c$	m	-9.10^{11}	9.10^{11}	-	10	-	11

DID 12	OUTPUT FROM : ATTITUDE RECONSTITUTION INPUT TO : IDT PREPROCESSING	PAGE : 1 OF 1 DATE : 82.09.15

DESIGNATION	ANNEX A	EXPLANATION	UNIT	RANGE MIN	RANGE MAX	DEF	NUM ACC ABS	NUM ACC REL	DIGITS
		1. <u>For each frame</u>:							
IFRAME	–	frame identification number	–	1	9.10^8	–	1	–	9
TMID	t_K (p31)	frame mid-time	s	0	2.10^8	–	10^{-6}	–	15
RAZ	α_z (p6)	Right Ascension of z axis	rad	$-\pi$	π	–	10^{-11}	–	12
DECZ	δ_z (p6)	Declination of z axis	rad	$-\frac{1}{2}\pi$	$\frac{1}{2}\pi$	–	10^{-9}	–	10
ARGZ	ω (p6)	argument of z axis	rad	$-\pi$	π	–	10^{-11}	–	12
KFIRST, KLAST	–	if an attitude discontinuity has occurred in the frame, these give the first and last sample affected	–	0	9000	0	1	–	4

DID 13	OUTPUT FROM : LSD DATA INPUT TO : IDT PREPROCESSING								PAGE : 1 OF 1 DATE : 82.09.15
DESIGNATION	ANNEX A	EXPLANATION	UNIT	RANGE MIN	RANGE MAX	RANGE DEF	NUM ACC ABS	NUM ACC REL	DIGITS
NPAR	-	number of distortion parameters	-	0	99	0	1	-	2
PARA(IPAR)	$g_{k\ell}$, $h_{k\ell}$ (p33)	a priori distortion parameters IPAR = 1 to NPAR	GPER/rad^m	-10^{12}	10^{12}	0	1	10^{-8}	8

DID 14	OUTPUT FROM : SSD DATA INPUT TO : IDT PREPROCESSING									PAGE : 1 OF 1 DATE : 82.09.15
DESIGNATION	ANNEX A	EXPLANATION	UNIT	RANGE MIN	RANGE MAX	DEF	NUM ACC ABS	NUM ACC REL	DIGITS	
NETA, NZETA	-	number of entries along η and ζ	-	1	4000	-	1	-	4	
ETAØ, ZETAØ	-	origin corresponding to I = J = 0	rad	-0.009	0.009	-	10^{-7}	-	5	
DETA, DZETA	-	discretization interval	rad	0	0.009	-	10^{-7}	-	5	
SSD(I,J)	$\delta G(\eta, \zeta)$ (p9)	small-scale distortion at point η = ETAØ+I*DETA, ζ = ZETAØ+J*DZETA	GPER	-10	10	-	10^{-5}	-	6	

DID 15	OUTPUT FROM : IDT PREPROCESSING INPUT TO : IDT SIGNAL CATALOGUE						PAGE : 1 OF 1 DATE : 82.09.15		
DESIGNATION	ANNEX A	EXPLANATION	UNIT	RANGE MIN	RANGE MAX	DEF	NUM ACC ABS	NUM ACC REL	DIGITS
		1. <u>For each frame:</u>							
IFRAME	–	frame identification number	–	1	9.10^{8}	–	1	–	9
TMID	t_K (p31)	frame mid-time	s	0	2.10^{8}	–	10^{-6}	–	15
NSTAR	–	number of objects in frame	–	0	10	0	1	–	2
		1.1. <u>For I = 1 to NSTAR:</u>							
ID(I)	–	object identification number	–	0	10^{8}	0	1	–	8
IFIELD(I)	$(-1)^{f}$ (p8)	field index	–	-1	1	0	1	–	1
ETAA(I),ZETAA(I)	$\tilde{\eta}, \tilde{\zeta}$ (p32)	approximate field coordinates	rad	-9.10^{-3}	9.10^{-3}	–	10^{-7}	–	5
NACC(I), NREJ(I)	–	number of samples accepted and and rejected	–	0	9000	0	1	–	4
GOF(I)	–	goodness of fit	–	-10^{38}	10^{38}	–	–	10^{-6}	6
BETA(K), K=1,5	$\underline{\beta}$ (p29)	estimated signal parameters	–	-10^{38}	10^{38}	–	–	10^{-6}	6
CBETA(K,L), K=1,5; L=K,5	$\underline{\Phi}^{-1}$ (p29)	covariance of BETA(K) (or equivalent, TBD)	–	-10^{38}	10^{38}	–	–	10^{-6}	6

DID 16	OUTPUT FROM : IDT PREPROCESSING INPUT TO : OTF DATA								PAGE : 1 OF 1 DATE : 82.09.15	
DESIGNATION	ANNEX A	EXPLANATION	UNIT	RANGE MIN	RANGE MAX	DEF	NUM ACC ABS	NUM ACC REL	DIG-ITS	
		1. For each frame on certain stars:								
TMID	t_K (p31)	frame mid-time	s	0	2.10^8	-	10^{-6}	-	15	
ID	-	object identification number	-	0	10^8	-	1		8	
BMV	-	colour index B-V	mag	-1	3	99	10^{-2}	-	4	
IFIELD	$(-1)^f$ (p8)	field index	-	-1	1	0	1	-	1	
ETAA, ZETAA	$\tilde{\eta}, \tilde{\zeta}$ (p32)	approximate field coordinates	rad	-9.10^{-3}	9.10^{-3}	-	10^{-7}	-	5	
OTF(K), K=1,3	M_1 etc. (p12)	OTF data, e.g. M_1, $(M_2/M_1)\cos(v_2)$, $(M_2/M_1)\sin(v_2)$	-	-1	1	-	10^{-6}	-	6	
COTF(K,L), K=1,3; L=K,3	-	covariance of OTF(K) (or equiv.)	-	-1	1	-	10^{-6}	-	6	