ESTIMATION OF GROUND TRUTH FOR EVALUATION OF CRATER DETECTION ALGORITHMS.

G. Salamunićcar[1] and S. Lončarić[2], [1]AVL-AST d.o.o., Av. Dubrovnik 10/II, HR-10020 Zagreb-Novi Zagreb, Croatia, gsc@ieee.org, [2]Faculty of Electrical Engineering and Computing, University of Zagreb, Unska 3, HR-10000 Zagreb, Croatia, sven.loncaric@fer.hr.

Introduction: Catalogue of 17582 craters was assembled, wherein each crater is aligned with MOLA topography and confirmed by three independent sources. It can be used as ground truth (GT) in future evaluations of crater detection algorithms (CDAs).

Proposed Evaluation Environment: CDA applications range from dating of a planetary surface [1], to autonomous landing to planets [2] and asteroids [3], and advanced statistical analysis [4]. For evaluation of CDAs, MOLA data in 1/64° resolution can be used as single input data-set. For CDAs that cannot use it as an input, visual images in different projections and with different shadowing were generated from the above file as shown in Fig. 1. Another prerequisite for CDA evaluation (e.g. the Receiver Operating Characteristics) is GT (locations and sizes of known craters).

Different GT Definitions: As found in the study of automatic detection of volcanoes on Venus [5], for the same image different geologists will produce different labelings and even the same geologist may produce different labelings at different times. Similar problems are with craters' catalogues, which is also work in progress [6, 7]. Different definitions of GT lead to different evaluation results, which are not or are hardly comparable to each other. Accordingly, it is to be expected that work toward standardization of GT can contribute to future evaluations of CDAs.

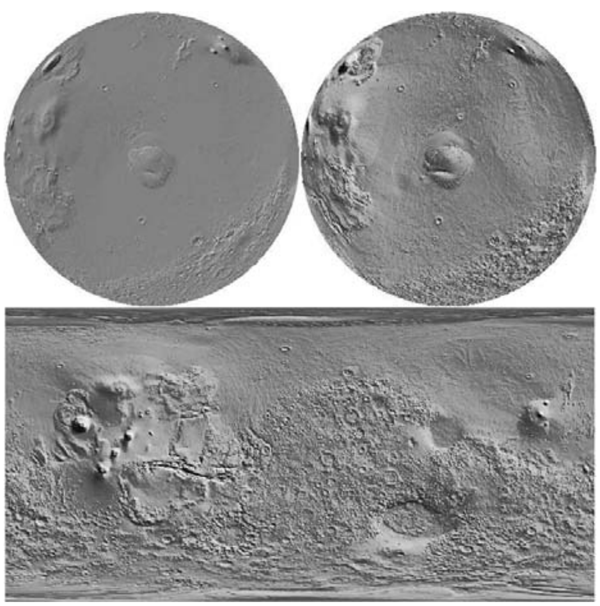


Figure 1: Global topography available for CDAs' evaluations in different resolutions and projections and with different shadowing.

Craters Difference Factor (CDF): Because of precision limitations, it should not be expected that different geologists or CDAs will assign identical coordinates and radius to the same crater. Accordingly, it is necessary to define measure of differences in position and size between two craters. For this purpose, a CDF is defined, call it *f*: when $r_1 \geq r_2$, $r_1 < (1+f) \cdot r_2$ and $d < f \cdot r_2$, two records are considered to belong to the same crater, where *d* is distance between crater centers. The rightmost column of Table 1 shows examples for *f*=1, *f*=0.5, and *f*=0.25. A problem with too large *f* is that different craters can be considered as duplicate records, as shown in Fig. 2, while the problem with a very small *f* is that it requires unrealistic precision as can be seen from Table 1.

Used Craters Catalogues: Three independent sources have been used: (1) catalogue from N. G. Barlow et al. [8]; (2) catalogue from J. F. Rodionova et al. [9]; and (3) revised version of catalogue used in previous work [10]. The first one will be referenced as *B* (contains 42284 craters, in downloaded file was one more than reported), the second one will be referenced as *R* (contains 19308 craters), while the third one will be referenced as *S* (initial version contains 22044 craters). To each crater from *B*,

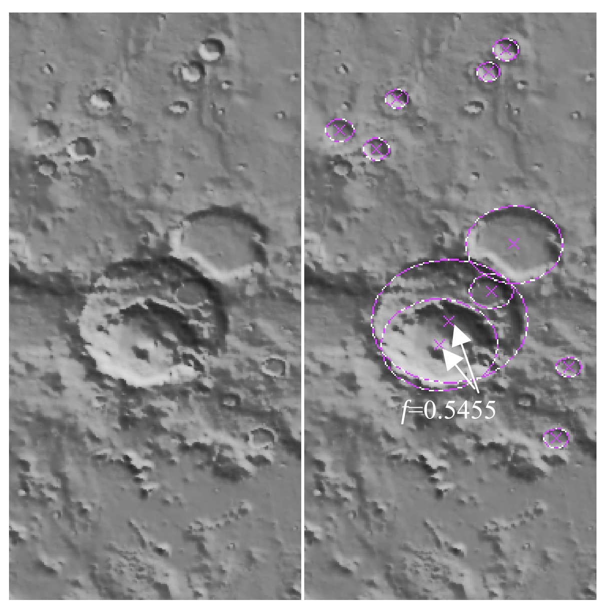
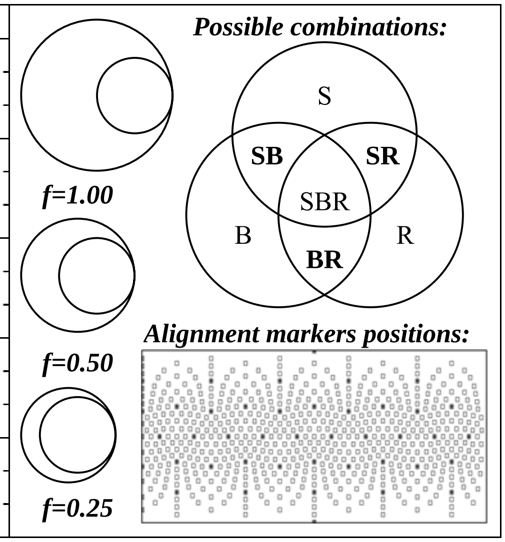


Figure 2: Topography before/after registration of craters (GT) wherein two craters with the smallest *f* in overall catalogue are outlined.

Table 1: Registration results for craters with *r* larger than 0.078125°. CDFs, possible combinations for different craters catalogues, and positions of 42 and 642 alignment markers over topography were shown in right frame.

Σ_{REG}	S	B	R	SB	SR	BR	SBR	Σ_{ALL}	*f*
	15130	6148	3190	375	239	13473	2248	40803	1
0	17394	11066	7983	101	73	10767	**458**	47842	**0.5**
	17923	16796	13689	22	25	5522	58	54035	0.25
	7465	5835	3348	1324	721	6591	8484	33768	1
42	12299	10645	7967	1173	739	6758	**3815**	43396	**0.5**
	15852	16401	13570	715	442	4263	1019	52262	0.25
	5344	5248	2838	1355	683	5023	10607	31098	1
642	9554	9497	7002	1543	930	5351	**5998**	39875	**0.5**
	13957	15218	12514	1135	736	3845	2200	49605	0.25
	3782	4531	1655	560	347	3385	13901	28161	1
13899	4219	6116	3186	321	151	2068	**13899**	29960	**0.5**
	4528	7570	4554	123	40	812	13899	31526	0.25
	3787	4252	1365	545	341	10	17583	27883	1
17582	4209	4505	1576	317	148	0	**17582**	28337	**0.5**
	4503	4701	1675	121	50	0	17582	28632	0.25



a unique name in form 'BID' was assigned, wherein ID is line number from downloaded file. All craters from *R* already had their ID so only 'R' was added as prefix to construct unique names. All craters from *S* are named only as 'S' because there is no need to reference old ones.

Joint Craters Catalogue (JCC): From *B*, 5 identical craters (identical coordinates and radius) were deleted (B09321, B16179, B21541, B29116 and B41189) leaving 42279 craters. While there were no identical records in *R* and *S*, duplicates ($f < 0.5$) were found in *R* and *S* as well. From *B*, *R* and *S*, JCC is constructed, containing 83631 craters, wherein there are 59742 craters (22405 from *B*, 19308 from *R* and 18029 from *S*) with *r* larger than 5 pixels on 1/64° resolution ($r >$ ~4.622 km). Different combinations are possible for GT, e.g. column *BR* from Table 1 are craters from *B* and *R* that are not in *S* while column *SBR* (that will be used as GT) are craters from *S*, *B* and *R*.

Alignment with Topography: Only alignment of large craters satisfied that $f < 0.5$, while small craters from *S*, *B* and *R* missed underlying topography for distances larger than the craters themselves. For craters to be used as GT, the additional requirement is that they are precisely aligned with topography.

Registration of Craters: Registration is performed as follows: (1) corresponding craters from *S*, *B* and *R* that belong to the same crater from topography are selected; (2) in order to define the coordinates, they are aligned with each other and underlying topography; (3) radius is defined as $r = (r_B \cdot r_R)^{0.5}$, and (4) unique name is assigned, e.g. *S000001B00371R11256Y2005S* is name of the first registered crater where *B00371* and *R11256* were used for registration. For 42 alignment markers over global topography, one crater is registered close to each marker and all other craters were aligned according to the closest registered craters. The procedure was repeated for additional 600 markers, but as shown in Table 1, this was still not sufficient for proper alignment of other craters. The procedure was additionally repeated for all available craters from *S* (13899 registrations). At the end, *S* was revised, additional registrations to be possible (17582 registrations).

Conclusion: Catalogue *B* confirmed itself as the most complete source and *R* also as an excellent choice. Alignment with MOLA topography and confirmation by three independent sources (*B*, *R* and *S*) to be satisfied as well, resulting catalogue with 17582 craters was assembled. Closest craters from this catalogue and repeated calibration of several hundred craters for testing purposes, justified *f*=0.5. The resulting catalogue with images generated from MOLA data is available upon request on DVD. While primary purpose is to be used as GT in future evaluations of CDAs, results show possible contribution to future cataloguing of craters as well.

References: [1] Sawabe Y. et al. (2005) *Adv. Space Res.*, in press. [2] Cheng Y. et al. (2005) *IEEE Conference on Computer Vision and Pattern Recognition*. [3] Leroy B. et al. (2001) *Image and Vision Computing*, *19*, 787-792. [4] Salamunićcar G. (2004) *Adv. Space Res.*, *33*, 2281-2287. [5] Burl M. C. et al. (1998) *Machine Learning*, *30*, 165-195. [6] Barlow N. G. (2003) *6th Int. Conf. on Mars*, Abstract #3073. [7] Barlow N. G. et al. (2003) *ISPRS WG IV/9 Extraterrestrial Mapping Workshop: Advances in Planetary Mapping*. [8] Barlow N. G. et al., *http://webgis.wr.usgs.gov/mars.htm*. [9] Rodionova J. F. et al., *http://selena.sai.msu.ru/Home/Mars_Cat/Mars_Cat.htm*. [10] Salamunićcar G. (2003) *LPS XXXIV*, Abstract #1403.