

Automatic Face Recognition and Identification Tools in the Forensic Science Domain

Angelo Salici^(✉) and Claudio Ciampini

Raggruppamento Carabinieri Investigazioni Scientifiche, RIS di Messina,
S.S.114 Km 6,400, 98128 Messina, Italy
{angelo.salici,claudio.ciampini}@carabinieri.it

Abstract. This paper describes an experimental work executed by Carabinieri Forensic Investigation Department (Italy) to explore the performance of the automatic face recognition systems in forensic domain. The main goal of the research is to survey the recognition ability and identification performance of these tools. The experiments are carried out using three commercial automatic facial recognition platforms. In our work we compare the difference between the forensic experts' way of manual facial comparison with the machine outcome in two different scenarios; the first is a training and certification environment, a facial image comparisons proficiency test to verify the recognition capabilities. The second is a daily forensic caseworks scenario, formed by 130 real cases successfully investigated by forensic experts, to analyze the identification achievement.

Keywords: Face recognition · Automatic · Identification · Forensic · Likelihood ratio

1 Introduction

In the forensic science world, the identification of the subject from the trace evidence he/she leaves is the main goal. People can be identified by their fingerprints, by traces of their DNA from blood, skin, saliva, hair, by their teeth, by their walk with a gait analysis and by their voice. Anyway, frequently these data does not exist or does not suffice to guarantee a successful identification. In these cases facial recognition from photo or video recording can provide many benefits.

Images or video recordings are often available for investigation and also for these reasons face recognition systems have recently received significant attention. Image or video may come from a witness camera or surveillance system and can show the face of a perpetrator. If we had not a suspect to compare, then we need to provide a manual search in a police database to solve and correlate the crime. This represents a long and difficult job like in other cases where it is necessary to analyze hundreds of hours of video to search the excerpts in which faces are visible.

For these reasons in the last period, thanks to more than 40 years of research that produced feasible technologies, we have seen a wide range of commercial and law enforcement applications to realize an automatic facial identification.

In our paper, we focused attention on the forensic aspects of automatic face recognition and describe a small scale experimental work carried out by Carabinieri Forensic Investigation Department (Italy) which explores the performance of three commercial automatic face recognition system in forensic domain. We compare the difference between the forensic experts' way of manual facial comparison which the machine results in two different scenarios; the first is a training and certification environment, the ENFSI-DIWG 2013 facial image comparisons proficiency test (FIC test 2013), and the second is a daily forensic caseworks scenario, formed by 130 real cases successfully investigated by forensic experts. In Sect. 2 we introduce some related works. In Sect. 3 we describe the principle and the methodology of our approach. In Sect. 4 we present our experimental results. Conclusions are presented in last section.

2 Related Works

In this section we briefly survey the existing literature on face recognition and describe the current state-of-the-art in Automatic Forensic Face recognition systems analyzing the main existing platform and evaluating their main features.

The history of the face recognition dates back to the advent of photography and the researches in automatic face recognition started in the 1960s. In [1] the authors carried out a review in the field of facial recognition and automatic human face detection describing the significant progress that has been achieved in the years. In [2] W. Zhao et al. provide a critical survey on face recognition categorizing recognition techniques and presents detailed descriptions of representative methods within each category. Face Recognition has become a very popular application in several fields and the interest on it is broad interdisciplinary ranging from biometrics and security to psychology and neuroscience. An overview on the wide range of face recognition practical applications is described in [3]. Face recognition has long been a goal of computer vision and a lot of work has been done on this subject to report the new developing techniques and methodologies or to tackle biometrics problem. On the other hand there are very few published works which describe automatic face recognition in forensic domain. The matching score from a biometric face recognition system is not directly suitable for forensic applications where a fundamental requirement is to evaluate the weight of the recognition evidence in a scientific framework. With this goal in mind Aitken and Taroni in [4] explore the use of statistical and probabilistic approach in forensic science to allow evaluation and interpretation, according to the evidence where there is an element of uncertainty. In [5] Tauseef et al. review famous works in the forensic face recognition, report on attempts to use automatic face recognition in forensic context and develop a framework to use automatic face recognition in the forensic setting. In [6] Peacock and Goode describe a pilot study carried out by the Forensic Science Service (UK) through a specific software package (Image Metrics Optasia™ [7]) which explores how reliable and under what conditions digital facial images can be presented as evidence. This work lays a foundation for how face recognition systems can be evaluated in a forensic framework.

A methodology and experimental results, carried out through a system based on AdaBoost algorithm [8] and Cognitec FaceVacs [9] commercial face recognition

system, for evidence evaluation in the context of forensic face recognition are presented in [10]. The proposed approach presented in [10] is in accordance with the Bayesian framework where the duty of a forensic scientist is to compute Likelihood ratio (LR) from biometric evidence which is then incorporated with prior knowledge of the case by the judge or jury.

3 Experimental Materials and Methods

The first scenario analysed in our work was the ENFSI-DIWG 2013 Facial Image Comparison Proficiency Test [11]. FIC 2013 Test was set-up and created by Salima SBIA of the Central unit of Belgian Federal Judicial Police – Forensic Science Directorate in Belgium [12]. The SKL (Statens Kriminaltekniska Laboratorium) [13] acted as organiser and have also written the final report. The test has been sent out to the members of the European Network of Forensic Science Institutes (ENFSI) Digital Imaging Working Group (DIWG) [14] as being the annual proficiency test regarding facial image comparisons (FIC). The test contained one to one single image comparisons where the subject sometimes (approximately in twenty image pairs) was wearing some kind of concealment such as a cap, scarf or other type of headwear. Thirty image pairs were of male and ten image pairs were of female subjects. In total forty image pairs were included in the test. The participants were asked to report their answers according to the conclusion scale based on the standards for the formulation of evaluative forensic science expert opinion discussed in [15] and shown in Table 1.

Table 1. Conclusion scale used and LR equivalent.

		Likelihood ratio	Log ₁₀ LR equivalent
+4	The observations extremely strongly support that it is the same person	>10000	>4
+3	The observations strongly support that it is the same person	1000 to 10000	3 to 4
+2	The observations support that it is the same person	100 to 1000	2 to 3
+1	The observations support to some extent that it is the same person	1 to 100	0 to 2
0	The observations support neither that it is the same person nor that it is different persons	1	0
-1	The observations support to some extent that it is not the same persons	1 to 0.01	0 to -2
-2	The observations support that it is not the same persons	0.01 to 0.001	-2 to -3
-3	The observations strongly support that it is not the same persons	0.001 to 0.0001	-3 to -4
-4	The observations extremely strongly support that it is not the same persons	<0.0001	<-4

The second scenario analyzed was a collection of 130 real forensic study caseworks successfully investigated by forensic experts. These daily forensic caseworks contained images of perpetrators extracted from video surveillance footage. Mostly images were in very low resolution and a wide variation exist in pose of face, lighting conditions and facial expression. In these images some kind of “concealment” such as a baseball cap to hide the frontal hairline, glasses and/or scarf was also present. The comparison of these subjects was carried out with a suspect image acquired from a database containing only frontal images with a small variation of pose, light and expression conditions. The time gap between these reference images and the images of perpetrators range from several months to several years.

These reference suspect images and the reference images from ENFSI-DIWG 2013 FIC Test were included in a database of 1829 images, which were used to configure the automatic face recognition systems tested. All images in the database are frontal with only small variations in pose, but variations in order of lighting conditions and facial expression exist.

The automatic facial comparison was conducted with three different commercial software. For every image pairs comparison the automatic face recognition systems provide a similarity score (a numeric value or percentage value) if the images were processed or return a message indicating the failure to process the image. When the images were processed, every system provides also a ranked matching list from the internal database. In Fig. 1 are shown two different examples of ranked matching lists and similarity score values provide by an automatic facial comparison software.

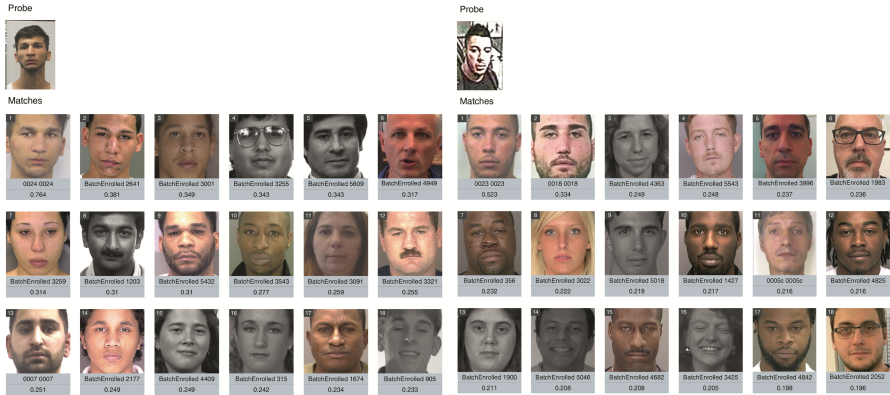


Fig. 1. Examples of ranked matching lists and similarity score values of two different cases.

The score is not directly suitable in the forensic domain. The automatic systems suggest a threshold to decide between two categories, (positive or negative match) but in the forensic world this is unacceptable.

Today, the standard way to report the value of the forensic evidence is based on the Bayesian interpretation framework (or the likelihood ratio framework). A general description of this framework in the forensic context can be found in [16].

4 Results and Observations

Both the two scenarios were engaged with an automatic approach and compared with a typical forensic experts' way of manual facial comparison. The typical forensic face recognition approach is based on the judgement of trained persons and so, in our experiment, we consider the judgement of forensic experts in facial comparisons field. In the first scenario, we analyzed the conclusions of the forensic experts of the 23 European law-enforcement participants at ENFSI-DIWG 2013 FIC Test. In the second scenario, we take into account the conclusions of the forensic Italian Carabinieri experts. The criteria used in the typical investigation from the forensic Carabinieri experts are the same used in the ENFSI-DIWG 2013 FIC Test presented in Table 1. To compare the results of the typical face recognition approach the responses on the upper side (from +1 to +4) of the conclusion scale were considered as a positive match, the responses on the lower side (from -4 to -1) of the conclusion scale were considered as a negative match. To compare the results of the automatic face recognition systems we considered the placement in the candidate rank list and the score provides by every automatic system. About the placement we considered a positive match if the automatic tool evaluated the correct comparison image in the first three positions with a consistent score. If the automatic systems evaluated the correct image in fourth position or on the upper part we considered a negative match. We considered also a negative match if the systems placed the comparison image in the first three positions but with a very low similarity score value. The threshold score value used was dependent by the automatic system tested. Only in the identification context of the second scenario we considered also an inconclusive zone to describe the circumstance where the automatic identification tool finds a match from fourth to fortieth position or not processing the image.

The ENFSI-DIWG 2013 FIC Test results [17] of the forensic experts are summarized in Fig. 2. The results show the sum of corrects (positive and negative matches), sum of zeros (sum of class 0 in Table 1), sum of errors (wrong positive and wrong negative matches).

Forensic Experts Results for FIC test 2013

IMAGE PAIR NO.	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20		
SUM OF CORRECTS	18	11	4	22	23	19	22	12	17	13	19	14	23	19	14	19	18	18	16	20		
SUM OF ZEROS	3	6	7	1	0	3	1	6	4	9	4	7	0	3	4	3	5	3	7	2		
SUM OF ERRORS	2	6	12	0	0	1	0	5	2	1	0	2	0	1	5	1	0	2	0	1		
IMAGE PAIR NO.	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	MEAN	% TOT.
SUM OF CORRECTS	11	19	11	12	23	22	19	19	22	19	23	15	16	22	22	15	23	20	14	30,9	77,17%	
SUM OF ZEROS	7	3	3	6	0	1	3	2	1	3	0	7	7	1	1	1	5	0	2	7	6,0	15,00%
SUM OF ERRORS	5	1	9	5	0	0	1	2	0	1	0	1	0	0	0	0	3	0	1	2	3,1	7,83%

Fig. 2. Results of laboratory participants for ENFSI-DIWG FIC 2013 test.

The results of our experiment on the same test with the three automatic face recognition systems are shown in Figs. 3, 4, 5. For every image pair test comparison, the table value indicates the position on the candidate rank list provided by the

automatic tool. A summarized of the results is shown indicating both the sum and percentage of the correct answers (for positive and negative match) and error made.

Automatic Face Recognition Systems no.1 — Results for FIC test 2013																						
IMAGE PAIR NO.	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20		
POSITIVE MATCH CORRECT				1	1		1		1				1		3		1			1		
NEGATIVE MATCH CORRECT			> 25			> 25		> 25		> 25	> 25			> 25		> 25		> 25				
ZEROS - NOT PROCESSING		NP										NP							NP			
POSITIVE MATCH ERROR																						
NEGATIVE MATCH ERROR	> 25																					
IMAGE PAIR NO.	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	SUM	% TOT.
POSITIVE MATCH CORRECT	2				1		1					1		1		1	1		1		16	40%
NEGATIVE MATCH CORRECT			12	> 25		> 25		> 25		> 25	> 25		> 25					> 25		> 25	17	42,5%
ZEROS - NOT PROCESSING		NP							NP						NP						6	15%
POSITIVE MATCH ERROR																					0	0 %
NEGATIVE MATCH ERROR																					1	2,5%

Fig. 3. Results of the 1st automatic recognition systems tested on ENFSI-DIWG FIC 2013.

Automatic Face Recognition Systems no.2 — Results for FIC test 2013																						
IMAGE PAIR NO.	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20		
POSITIVE MATCH CORRECT	1			1	1		1		1				1				1		1	2		
NEGATIVE MATCH CORRECT		20	13			32		365		463	361	> 524		274		18						
ZEROS - NOT PROCESSING																						
POSITIVE MATCH ERROR																		3				
NEGATIVE MATCH ERROR															8							
IMAGE PAIR NO.	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	SUM	% TOT.
POSITIVE MATCH CORRECT	1	1			1		1		1			1		1	1	1	1		1		20	50%
NEGATIVE MATCH CORRECT				494		342		113		> 524	> 524		322					326		495	17	42,5%
ZEROS - NOT PROCESSING																					0	0%
POSITIVE MATCH ERROR			1																		2	5%
NEGATIVE MATCH ERROR																					1	2,5%

Fig. 4. Results of the 2nd automatic recognition systems tested on ENFSI-DIWG FIC 2013.

Automatic Face Recognition Systems no.3 — Results for FIC test 2013																						
IMAGE PAIR NO.	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20		
POSITIVE MATCH CORRECT	2												1				2			1		
NEGATIVE MATCH CORRECT		> 50	> 50			> 50		> 50		> 50	> 50	> 50		> 50		> 50		> 50				
ZEROS - NOT PROCESSING																						
POSITIVE MATCH ERROR																						
NEGATIVE MATCH ERROR				> 50	21		> 50		> 50						> 50				13			
IMAGE PAIR NO.	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	SUM	% TOT.
POSITIVE MATCH CORRECT							1		2					1					3		8	20%
NEGATIVE MATCH CORRECT			> 50	> 50		> 50		> 50		> 50	> 50		> 50					> 50		> 50	19	47,5%
ZEROS - NOT PROCESSING																					0	0%
POSITIVE MATCH ERROR																					0	0%
NEGATIVE MATCH ERROR	> 50	30			> 50							> 50			> 50	25	10				13	32,5%

Fig. 5. Results of the 3rd automatic recognition systems tested on ENFSI-DIWG FIC 2013.

A summarize table reports the percentage values of the results for forensic experts and automatic systems are shown below in Fig. 6.

No. 1-23 = Laboratory participant FIC Test 2013											I,II,III= Automatic Face Recognition Systems no.1, 2, 3															
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	I	II	III
% EXACTS	92,5	100	87,5	82,5	82,5	92,5	80	67,5	17,5	80	92,5	82,5	82,5	87,5	80	72,5	67,5	52,5	82,5	60	70	77,5	85	82,5	92,5	67,5
% ZEROS	7,5	0	7,5	5	5	0	12,5	25	82,5	10	2,5	12,5	7,5	7,5	5	17,5	25	40	5	37,5	25	2,5	2,5	15	0	0
% ERRORS	0	0	5	12,5	12,5	7,5	7,5	7,5	0	10	5	5	10	5	15	10	7,5	7,5	12,5	2,5	5	20	12,5	2,5	7,5	32,5

Fig. 6. Comparisons between forensic experts and automatic systems results.

Two of the three automatic systems performed a superior total results than the mean of the forensic experts. The recognition ability of the automatic systems is very high with a good images quality. Also the similarity score is commensurate with the placement, high rank position corresponded with an high similarity score.

The performance of the automatic face recognition tools in the second scenario is shown in Fig. 7. The research has pointed out that the ability of the automatic tools to identify the correct candidates seems to decrease because of the different pose, light conditions and image quality. Instead, all the automatic systems showed steadiness and tolerance with regard to the mutation of facial expression, the concealment and the age variation.

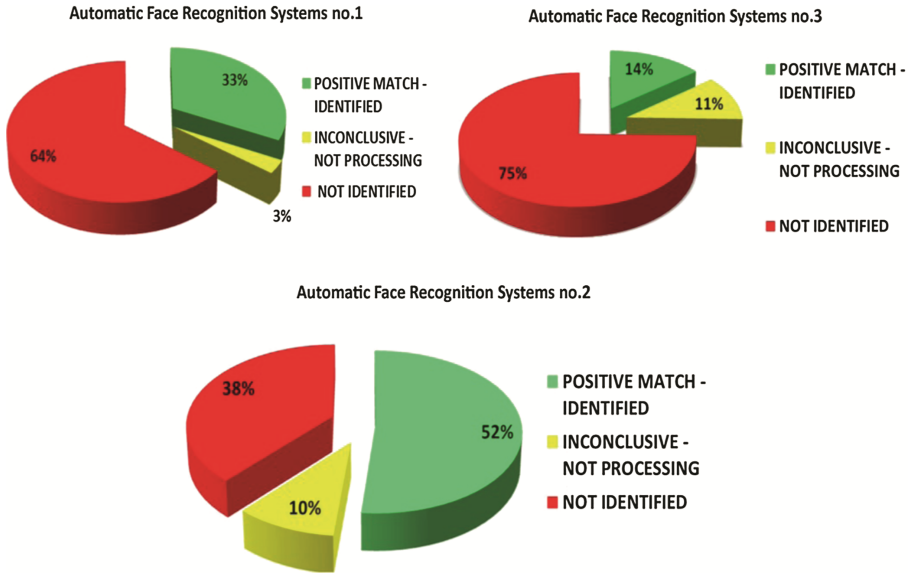


Fig. 7. Identification performance of automatic tools in real forensic cases.

A lot of work has been done to tackle the problem of face recognition across ages. An empirically study on how age differences affect recognition tasks is shown in [18].

Our experiment confirms that the image quality plays a very important role also in automatic facial image comparisons.

The effect of image quality in facial image comparisons have examined in such previous study that suggests that poor image quality leads to more accurate conclusions. How differences in image quality affect the performance of forensic experts and untrained persons is described in [19].

5 Conclusions and Future Works

The present study tested automatic face recognition systems ability to perform a forensic investigation in recognition and identification context. The results show that these systems are ready to support a forensic image comparison laboratory. In face recognition context the study shows a very important feature from a forensic point of view: a very little number of forensic experts, probably with more experience, gave conclusions with less errors and most correct answers. Furthermore the high rate of correct answers of the automatic systems is more than acceptable and makes the automatic tools very important to support the forensic expert decisions. For a complete application in the forensic domain these automatic tools should provide directly a Likelihood Ratio value and not only a similarity score, according to the ENFSI guideline for evaluative reporting in forensic science [20]. In face identification context the experiment shows encouraging and very promising results to adopt these systems in law-enforcement investigation cases where there are not enough suspected candidates. Although our experiment used

real test images, in very low resolution and different pose and light conditions, the results showed the better cases 52% of positive match identification of the unknown perpetrator. This study also suggests that image quality and the choice of databases of facial images plays a very important role. Future research should take into account the Likelihood Ratio computations performance to evaluate a threshold similarity score value and working with larger database of facial data.

Acknowledgements. We would like to express our gratitude towards Salima SBIA, Fredrik Eklöf and Peter Bergström for having set-up, created and organized the ENFSI-DIWG 2013 Facial Image Comparison Proficiency Test. We also want to thank all ENFSI Digital Imaging Working Group and those who have contributed in our experiment.

References

1. Patil, C.S., Patil, A.J.: A review paper on facial detection technique using pixel and color segmentation. *Int. J. Comput. Appl.* **62**(1) (2013)
2. Zhao, W., Chellappa, R., Phillips, P.J., Rosenfeld, A.: Face recognition: a literature survey. *ACM Comput. Surv.* **35**(4), 399–458 (2003)
3. Senior, A.W., Bolle, R.M.: Face recognition and its application. In: Zhang, D. (ed.) *Biometric Solutions. The Springer International Series in Engineering and Computer Science*, vol. 697, pp. 83–97. Springer, Boston (2002). [10.1007/978-1-4615-1053-6_4](https://doi.org/10.1007/978-1-4615-1053-6_4)
4. Aitken, C.G.G., Taroni, F.: *Statistics and Evaluation of Evidence for Forensic Scientists*, 2nd edn. Wiley, New York (2004). ISBN: 978-0-470-84367-3
5. Tauseef, A., Raymond, V., Luuk, S.: Forensic face recognition: a survey. In: *Face Recognition: Methods, Applications and Technology. Computer Science, Technology and Applications*, vol. 9. Nova Publishers (2012). ISBN 978-1-61942-663-4
6. Peacock, C., Goode, A., Brett, A.: Automatic forensic face recognition from digital images. *Sci. Justice* **44**(1), 29–34 (2004)
7. Image Metrics Optasia™. www.image-metrics.com
8. Freund, Y., Schapire, R.: A decision-theoretic generalization of on-line learning and an application to boosting. *J. Comput. Syst. Sci.* **55**, 119–139 (1997). Article No. SS971504
9. <http://www.cognitec.com/facevacs-dbscan.html>
10. Ali, T., Spreeuwens, L., Veldhuis, R.: Towards automatic forensic face recognition. In: Abd Manaf, A., Zeki, A., Zamani, M., Chuprat, S., El-Qawasmeh, E. (eds.) *ICIEIS 2011. CCIS*, vol. 252, pp. 47–55. Springer, Heidelberg (2011). doi:[10.1007/978-3-642-25453-6_5](https://doi.org/10.1007/978-3-642-25453-6_5). ISSN 1865-0929
11. <http://forensicscience.eu/fic-test/>
12. http://www.polfed-fedpol.be/org/org_dgj_en.php
13. <http://skl.polisen.se/>
14. <http://www.enfsi.eu/about-enfsi/structure/working-groups/digital-imaging>
15. Association of Forensic Science Providers. Standards for the formulation of evaluative forensic science expert opinion. *Sci. Justice*. **49**(3), 161–164 (2009)
16. Evett, I.: Toward a uniform framework for reporting opinions in forensic science casework. *Sci. Justice* **38**(3), 198–202 (1998)
17. Fredrik, E., Peter, B.: Results and presentation of the ENFSI-DIWG 2013 facial image comparisons proficiency test. SKL Intern rapport. Dokument-och informationsteknikenheten 2014:08. <http://media.forensicscience.eu/2014/05/FIC-test-20131.pdf>

18. Ling, H., Soatto, S., Ramanathan, N., Jacobs, D.W.: A study of face recognition as people age. In: IEEE 11th International Conference on Computer Vision (ICCV 2007), Rio de Janeiro, 14-21 October 2007, pp. 1–8 (2007)
19. Norell, K., et al.: The effect of image quality and forensic expertise in facial image comparisons. *J. Forensic Sci.* **60**(2), 331–340 (2015). doi:[10.1111/1556-4029.12660](https://doi.org/10.1111/1556-4029.12660)
20. ENFSI Guideline for evaluative reporting in forensic science. European Network of Forensic Science Institutes, 8 March 2015 (2015)

<http://www.springer.com/978-3-319-67638-8>

Digital Communication. Towards a Smart and Secure
Future Internet

28th International Tyrrhenian Workshop, TIWDC 2017,
Palermo, Italy, September 18-20, 2017, Proceedings

Piva, A.; Tinnirello, I.; Morosi, S. (Eds.)

2017, X, 263 p. 98 illus., Softcover

ISBN: 978-3-319-67638-8