INFLUENCE OF PROJECTILE MATERIAL ON THE EFFECTIVENESS OF AUTOMATED COMPARISON

Lehi Sudy dos Santos
Palloma Vieira Muterlle
University of Brasilia
Universidade de Brasília, Campus Darcy Ribeiro, Faculdade de Tecnologia - Depto Eng. Mecânica.
Campus Darcy Ribeiro, Bloco SG9 - Lab de materiais, Asa norte Brasília, DF – Brasil. CEP: 70910-900
lehisudy@gmail.com
pmuterlle@gmail.com

Abstract. *When a crime is committed with a gun, forensic experts may have to match fired bullets from the scene with each other, or with a firearm. Although this kind of exam has been done for almost a hundred years, being successfully used in countless cases, it has some limitations. Among them the fact that the analyses are made by the comparison of two specimens at a time in the microscope, which leads to a very cumbersome and time-consuming process, particularly when many specimens are related to the same crime, and that in many situations there is no questioned firearm, which leads to an open case. In order to reduce the time spent in the comparison process and to better deal with open cases, forensic labs all over the world have tried to implement electronic comparison systems.*
To assess the feasibility and accuracy of a future National Ballistic Database, a reference ballistic image database (RBID), using the Ballistic Identification System Evofinder®, has been implemented with test-fired bullets from revolvers on the caliber .38SPL. The hit-list for controlled comparisons, involving projectiles of different materials, were used to derive quantitative effectiveness criterion proposed by Joachim Rahm (2012). The results showed huge differences in system performance between different projectile materials. Brinell hardness tests on the bullets were made to investigate the influence of this parameter on the effectiveness of the system. The results provided a guide to the projectiles that must be used, when colleting samples from a gun, in order to ensure a higher effectiveness criterion and consequently the best correct gun identification probability.

Keywords: *Bullet, Firearm, Automated comparison, Evofinder®, Correlation quality*

1. INTRODUCTION

When a bullet is fired, characteristic marks from the gun barrel are left on the cylindrical section of the bullet (León, 2006). At least two types of marks can be distinguished: class marks and individual marks. Class marks, as number of lands and grooves, their twist direction, width and profile, are useful to identify the caliber, maker and model of the gun. But other marks, as microscopic striated marks, bear individual characteristics from that barrel and are useful to identify the specific gun that fired that bullet. Most of the imperfections of the barrel that generate this marks are produced during the manufacturing process and are totally random in their distribution, shape and size, allowing one to affirm that no two guns, even those consecutively produced, will have the exact same individual marks (Heard, 2008).

To identify the gun used in a crime, test bullets from the suspected gun must be collected and compared, using a comparison microscope, with the bullet recovered from the crime scene, in order to see if both have the same class and individual characteristic marks.

In modern times, the trend in forensic labs has been the implementation of electronic comparison systems, to create databases of images of parts of bullets and cartridges cases and to perform automated comparisons. EVOFINDER® (EV0, 2015a), IBIS (IBIS, 2015a), and LEPUS (Lepus, 2015) are some of the commercial available systems. Nevertheless, despite all advances that these systems had received in the last two decades, there is still not an ultimate solution for the problem of automated comparison.

Recent studies had address some of the problems, including difficulties in differentiating individual marks when dealing with the comparison of thousands of images (Rahm, 2012), necessity of establishment of parameters to properly implement this databases (Kinder, 2002), and dependence of effectiveness of systems regarding ammunition's brand, (Ceuster and Dujardin, 2015). On this paper was investigated another important aspect to implement a future Nacional Ballistic Database on Brazil: the difference of performance of the system regard the projectile material used to collect the test-fired bullets from guns on caliber .38SPL (thirty eight Special) with barrels presenting 6D (six right hand) twist grooves .

2. MATERIALS AND METHODS

Of many commercial systems available, Evofinder® and IBIS® has been the two systems more often adopted for law enforcement agencies in the world (EVO, 2015b and IBIS, 2015b). Recent studies suggest that Evofinder® is one the systems most reliable and that presents the best performance (Ceuster and Dujardin, 2015).

In the National Institute of Criminalistics of the Federal Police in Brasília, Brazil, an Evofinder® system, version 5.4, has been used since 2011 and has been helpful to solve many criminal cases. The system records images of test-fired bullets and cartridges collected from guns (control samples), and performs automated comparisons of questioned samples against the images of the database. The results are presented in a hit-list in decreasing rank of similarity between images. After that, the expert need to analyze the images or even use the comparison microscope to check and endorse the match.

In order to rate the correlation quality of comparison systems, Joachim Rahm (2012) suggested an effectiveness criterion. In this paper was used a little deviation of this criterion to analyze the behavior of a RBID, built in Evofinder®, by colleting test-fired bullets from guns on the caliber .38SPL, all approximately with the same time of use and wear. By the time the comparisons were performed, the database had approximately 1256 different images of bullets of this caliber.

Initially fourteen test-fired bullets, collected from sixteen different guns, were feed on the system as control samples in a temporary folder. Two control samples from each one of the Tab. 1. In a second step, more five test-fired bullets from each gun were collected and feed on the system as questioned samples (see Tab. 2). All ammunition used were from CBC brand (The CBC Ammo Group consists of four internationally recognized ammunition brand names: Magtech, CBC, Sellier & Bellot and MEN) (HGA, 2015).

In order to analyze the system performance on different configurations of the RBID, on each of the 1120 correlations performed to derivate the results of this paper, only two control sample of the gun were put in the list for comparison at time. This mean that on each comparison the image of one questioned projectile was compared against 1258 images. The position on the hit-list of the only two control samples from the same gun were recorded. This analyzes allowed to investigate the performance of the system regarding the control samples material and the questioned sample material.

Table 1. Projectiles used as control samples (CS).

LRN(1)	SJHP(2)	SJHP(2)	FMJ-ST(3)	JHP-ST(4)	G-JHP(5)	LRN(1)
CS1	CS2	CS3	CS4	CS5	CS6	CS7

(1) Lead Round Nose, (2) Semi-Jacketed Hollow Point, (3) Full Metal Jacket Silver Tip
(4) Jacketed Hollow Point Silver Tip, (5) Gold Jacketed Hollow Point
(1) CS1 and CS7 are of the same material but CS1 was the first to be collected and CS7 was the last.
(2) The differences between CS2 and CS3 were the ammunition propellant, respectively 38SPL and 38SPL+P

Table 2. Projectiles used as questioned samples (QS).

LRN(1)	SJHP(2)	FMJ-ST(3)	JHP-ST(4)	G-JHP(5)
QS1	QS2	QS3	QS4	QS5

In order to explain the observed performance differences, samples of the projectiles were randomly selected and tested for Brinell hardness. One of the reasons to choose this kind of hardness test was because it is an indicative of basic material properties such as strength and hardenability (Hill, *et al*., 1989). Tests were made on not fired bullets as well on fired ones, using a Zwick Roell ZHU250 machine, operating for 20s on each test.

The Brinell hardness tests were made under ABNT NBR NM ISO 6506-1 regulation (ASSOCIAÇÃO…, 2010), and uses a spherical ball of diameter D, under a load F, that produces a well define circular impression on the sample of diameter *d*, measured in two perpendicular directions. The Brinell hardness (HBW) is then determined by Eq. (1).

To keep the Factor Load (F/D²) constant, on LRN samples (CS1 and CS7) was used tungsten indenter of 1mm of diameter under load of 2.5 kgf, and on others bullets tungsten indenter of 2.5 mm with load of 15.625 kgf.

$$\mathrm{HBW} = \frac{0.102 \times F}{\pi . D^2 (1 - \sqrt{1 - \frac{d^2}{D^2}})} \tag{1}$$

3. DETERMINATION OF EFFETIVENESS

To rate the system performance, Joaquim Rahm (2012) proposed to include the results on the hit-list of each comparison up to position 20. Indeed, one of the contributions of that paper was the establishment of a break condition, which for the caliber .38SPL was 11. Nevertheless, the tests performed for this paper were more conservative and the searches for the control samples were carried up until position 20.

As there were two correct control samples on each comparison, the one in best position on the hit-list was chosen. The number of matches on that position, divided by the number of all matches, establishes a probability to find a match on position n. The cumulative probability, defined by the sum of the probabilities up until position n, was plotted as function of position n, and the effectiveness criterion was defined using a hyperbolic curve that best fit the data as shown bellow.

3.1 Regression Curve

The graphics obtained, so as on the work of Rahm (2012), are best fit for hyperbolic functions as:

$$\mathrm{P}(n) = \frac{a.n}{n+b} + c.n \qquad a\ e\ c\ \in [0,1], and\ \ n\ \in [0, i] \tag{2}$$

Where:

i is the size of the database (in this analyzes *i* was 1258).
P(*n*) is the cumulative probability of a match at the position n.
The boundary conditions are:

$$\mathrm{P}(i) = 1 \tag{3}$$

$$\mathrm{P}(0) = 0 \tag{4}$$

From the boundary condition of Eq. (3) follows that:

$$c = \frac{i.(1-a) + b}{i.(b+a)} \tag{5}$$

Therefore only *a* and *b* are the parameters to be established on the regression curve of each graphic.

3.2 Effectiveness criterion

Consider the cumulative probability of a match as function the position on the hit list (P x *n***)** of the Fig. 1, the effectiveness criterion (Γ_0) was define as:

$$\Gamma_0 = \frac{A2}{A1 + A2} \tag{6}$$

$$\Gamma_0 = \frac{\int_0^i P(n)\, dn}{1.i} \tag{7}$$

$$\Gamma_0 = a + \frac{c}{2}.i + \frac{k}{i} \tag{8}$$

$$with \qquad k = a.b.(\ln b - \ln(i+b)) \tag{9}$$

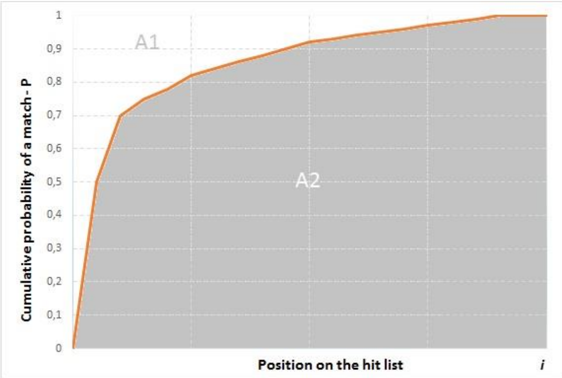

Figure 1. Cumulative probability of a match until up position i of the hit-list

3.3 New effectiveness criterion

Ram (2012) took one step more on Eq. (2) and Eq. (8) by considering the special case where $|a|$, $|b| \leq 1$ and $i \gg 1$, but on this paper this approximation was not used because some results presented $|b| > 1$. Even Table 5 of the paper of the Rahm (2012) shows $|b| > 1$.

Furthermore, especially in cases that the effectiveness was very low, the effectiveness criterion of the Eq. (8) overestimated the system performance. The results of the Fig. 2 represent real data obtained in three different test configurations. Using Eq. (8) the effectiveness criterion (Γ_0) for this data are: to the blue line 0.89, to the yellow line 0.73, and to the green line 0.79.

The reason why Eq. (8) overestimates the effectiveness of the system in data plotted on blue line is because this line presents a great slope at the end of the graphic (15<n<20). Therefore, integration until i became very dependent of this final behavior and it is not a good representation of the better performance we can visualize particularly on data plotted in green line.

To correct that was used a little deviation from the Eq. (8) to rate the performance of a system. Since has been registered results only up until position 20, it is most reasonable to calculate the effectiveness criterion only integrating P(n) from 0 to 20, instead of 0 to i. This new effectiveness criterion (Γ_1) was defined as:

$$\Gamma_1 = \frac{\int_0^{20} P(n)\, dn}{20} \tag{10}$$

$$\Gamma_1 = a + 10.c + \frac{k}{20} \tag{11}$$

$$with \quad k = a.b.(\ln b - \ln(20 + b)) \tag{12}$$

Using Eq. (11) the effectiveness criterions for this data are: to the blue line 0.32, to the yellow line 0.36, and to the green line 0.55; representing more properly the differences of system performances in these three configurations.

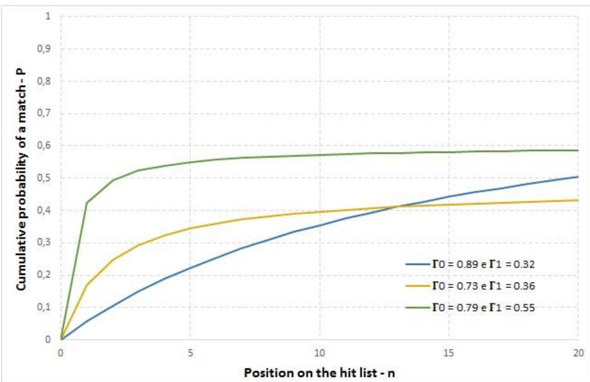

Figure 2. Cumulative probability of a match in three different configurations

4. RESULTS

The Fig. 3 shows the regression curves obtained with the controls samples CS1, and the Fig. 4 with CS6, respectively the worst, and the best, system performances. The regression curves and the corresponding effectiveness criterion (Γ_1) for all seven types of control samples were obtained in the same way. The results are shown in the Tab. 3.

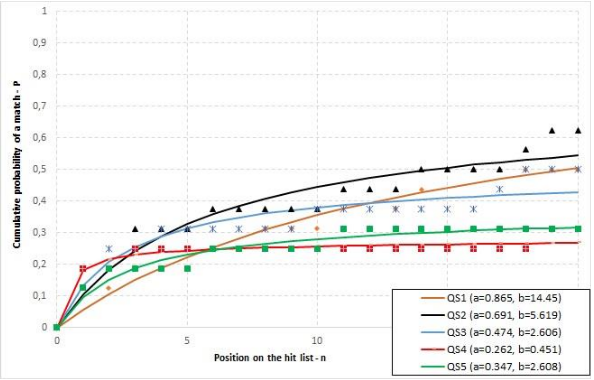

Figure 3. Regression curves of cumulative match probability with control samples CS1 regard all types of questioned samples (QS) (worst performance results)

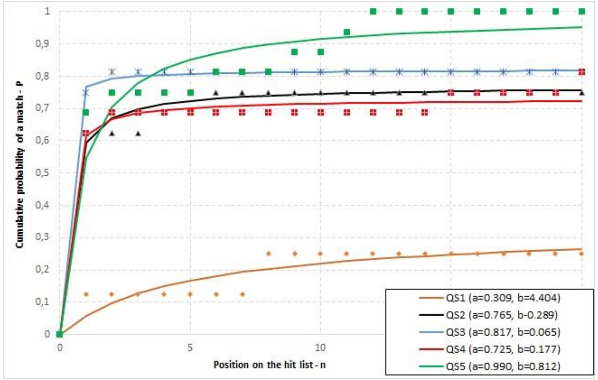

Figure 4. Regression curves of cumulative match probability with control samples CS6 regard all types of questioned samples (QS) (best performance results)

Table 3. Effectiveness criterions of the system (Γ_1) regard control samples (CS) and questioned sample (QS).

	CS1	CS2	CS3	CS4	CS5	CS6	CS7
QS1	0,32	0,30	0,22	0,27	0,16	0,20	0,35
QS2	0,40	0,61	0,91	0,63	0,51	0,72	0,36
QS3	0,34	0,71	0,74	0,90	0,54	0,80	0,48
QS4	0,25	0,44	0,65	0,54	0,56	0,70	0,32
QS5	0,25	0,69	0,70	0,68	0,50	0,86	0,33

The results obtained in the Brinell hardness test showed no meaningful difference between not fired bullets and fired ones, as shown on Fig. 5 of results.

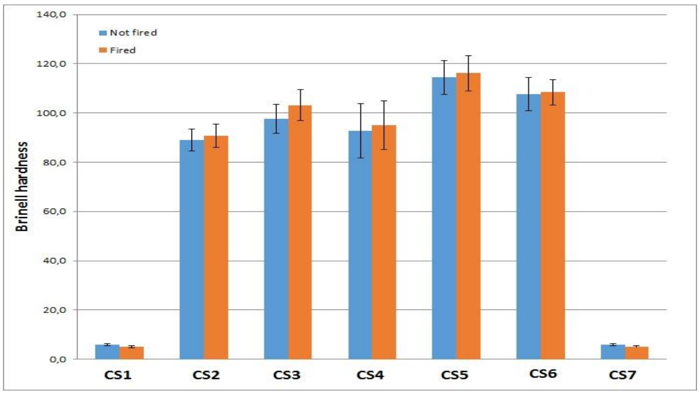

Figure 5. Mean Brinell hardness for all control samples (CS1 to CS7)

5. DISCUSSION

The Fig. 6 shows the mean effectiveness criterion (Γ_1), and the mean Brinell hardness, according to control sample. When operating with CS1 and CS7 the system presented the worst performances, with mean effectiveness equal to 0.31(±0.06) and 0.35(±0.06). This demonstrated no meaningful difference on the effectiveness of the system regarding LRN projectiles collected at the beginning and the end of the test-fires. Comparing these means with LRN bullets against the means with others projectiles showed how difficult was for the system correctly identify LRN projectiles.

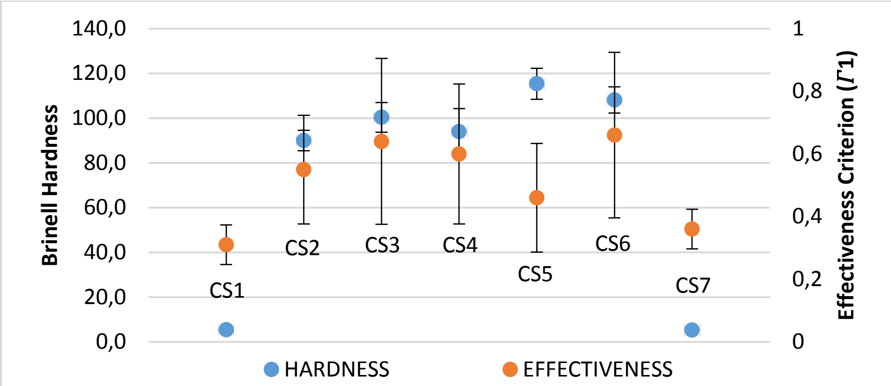

Figure 6. Mean effectiveness criterion (Γ_1) and mean Brinell hardness regarding control sample (CS1 to CS7)

The comparisons of these means effectiveness with the Brinell hardness of the projectiles provide a suitable explanation. The CS1 (and CS7) present a much lower hardness, 5.4 (±0.6) when compared to the mean Brinell hardness of the others projectiles, 101.5 (±9.9). As consequence of that, the number of striated marks generated in LRN bullets fired through the barrel is greater than for others harder bullets. Although more in number, not all these striae represent individual characteristics, because some of them were generated by debris left in the barrel by previous discharges. To further complicate the problem, even small differences in chamber pressure lead to differences in the imprinted barrel imperfections on the LRN bullets. As a result, the system presents poor comparison effectiveness with these bullets.

Figure 7 shows the comparison images of the same land and groove engraved area of two bullets fired from the same gun. The bottom image, captured from a LRN bullet, presents a much greater number of striae than the top image, capture from a G-JHP bullet.

Figure 7. Images of test-fired projectiles from the same gun, at the top of the image a CS6 (G-JHP) and at the bottom a CS1 (LRN).

To analyze this Factorial Design experiment, with two variables, control sample and questioned sample, respectively with 7 (seven) and 5 (five) levels, assuming normality, constant variance and independence of the data, the analysis of variance (ANOVA) is quite appropriate.

ANOVA were made to test the hypothesis *h0* that the mean effectiveness criterion regarding the control sample material or the questioned sample material were the same. The alternative hypothesis *h1* is that at least one of the means is statically different from the others. The P-value of the results of the ANOVA of table 4 indicates the probability that *h0* hypothesis is true. High values of F-Fisher when compare to F-Critical indicates that *h1* hypotheses is true on that level of significance.

The ANOVA1 results on Tab. 4 were obtained from all data of Tab. 3, and as expected from the analysis of Fig. 6, indicates differences in the mean effectiveness between control sample materials and between questioned sample materials.

Since the CS1 and CS7 presented the worse mean effectiveness, the ANOVA2 was calculated removing from the analysis all effectiveness for CS1 and CS7 of the Tab. 3, and once more indicates differences in the mean effectiveness between control sample materials as well as questioned sample materials.

Finally the ANOVA3 was performed not including the effectiveness for CS5 (the next worst mean effectiveness criterion), and at this analysis the values obtained indicated differences between questioned sample materials but no more statically differences between control sample materials.

Table 4 – ANOVA with significance level of 5%.

Data	***Variable Source***	***F-Fisher***	***P-value***	***F-Critical***	***Hypothesis confirmed***
ANOVA1 (All data from tab.3)	QS	10,19	5,74823E-05	2,78	*h1*(1)
	CS	5,91	0,00067	2,51	*h1*
ANOVA2 (Without data from CS1 and CS7)	QS	23,53	1,55596E-06	3,01	*h1*
	CS	3,67	0,026368	3,01	*h1*
ANOVA3 (CS2, CS3, CS4 and CS6)	QS	20,70	2,56879E-05	3,26	*h1*
	CS	1,19	0,35472	3,49	*h0*(2)

(1) h1 – at least one of the mean effectiveness criterion are statistically different from the others
(2) h0 – all mean effectiveness criterions are statistically indistinguishable

The above results indicate that the controls samples CS1, CS5 and CS7 are not recommended to be used as test-fired samples in general, and that is worth investigating which projectile materials (control samples) would provide the best results when there are questioned projectile of LRN or JHP-ST. Figure 8 shows the mean effectiveness for these questioned samples regarding the control sample material.

When performing automated comparison with questioned sample QS1 the results suggest that is best to have the same projectile material as control sample (CS1 or CS7) on the database. In the other hand, with questioned sample QS4, the results suggest no gain in effectiveness if there were control samples of the same material (CS5) on the database. Actually to have control samples CS3 or CS6 would be more recommended.

Figure 8. Effectiveness criterions for automated comparisons with questioned samples QS1 and QS4

Notwithstanding the fact that the ANOVA3 (Tab. 4) has showed no meaningful difference in performance of CS2, CS3, CS4 and CS6, the effectiveness of CS3 and CS6 was not only higher in mean, as it also presented higher values regarding all questioned sample materials, excepted for the questioned sample QS1. Figure 9 presents the results of Tab. 3 in a better way to visualize this conclusion, showing higher means and more aggregated results with CS3 and CS6.

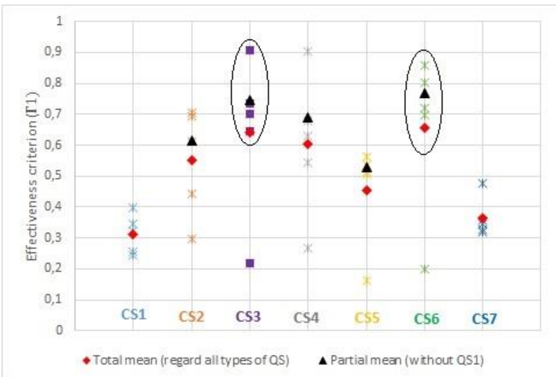

Figure 9. Effectiveness criterion regard control samples (CS) relating to all questioned samples (QS)

6. CONCLUSION

When performing a manual comparison of two bullets, to decide if they came from the same gun, one of the important premises to correct firearm identification is to compare bullets of the same material and profile. Nevertheless, when planning the implementation of a National Ballistic Database this premise cannot be guaranteed; it would be too much expensive and impracticable to collect test-fired bullets of all different projectile materials to register a gun on the system.

To address this problem this paper has studied the important issue of the effectiveness of an automated comparison system regarding the projectile material used to collected the control samples.

In order to study the influence of projectile material on automated comparison, a RBID was implemented in the system Evofinder®, using test-fired bullets of different projectiles collected from .38SPL caliber guns.

A little deviation of the effectiveness criterion suggested by Joachim Rahm (2012) was proposed and used to analyze thousands of automated comparisons performed controlling two variables, control sample and questioned sample.

The system presented poor effectiveness with LRN bullets. Brinell hardness tests and analysis of the images of land and groove areas of these bullets provided an explanation for this low effectiveness, showing a clear relationship between lower Brinell hardness and poor system effectiveness.

Nevertheless, control samples on LRN bullets was the most effective to find a questioned LRN bullet. Furthermore, when operating with SJHP or G-JHP bullets, the system presented higher effectiveness criterions regarding all questioned sample materials, except when the questioned samples were LRN bullets. Therefore, the combined analysis of these results suggests that, when implementing a ballistic database of bullets from.38SPL caliber guns, it is recommended to collet two test-fired bullets with G-JHP or SJHP bullet and two with LRN bullets from each gun, thus ensuring a higher effectiveness criterion and consequently the best correct gun identification probability.

The relative low effectiveness results obtained, despite all great advances that the commercial system Evofinder® presents in technology to capture and manipulate images, once more supports the idea that a ballistic imaging database of all guns are not yet recommended (Kopel and Burnett (2003), National Research Council (2008) and Ceuster and Dujardin (2015)).

7. REFERENCES

ASSOCIAÇÃO BRASILEIRA DE NORMAS TÉCNICAS. NBR NM ISO 6506-1: Metallic Matterials – Brinell hardness test Part 1: Test method. Rio de Janeiro, 2010. 18p.

Ceuster, J. D. and Dujardin, S., 2015. “The reference ballistic imaging database revisited”. *Forensic Science International*, Vol. 248, pp. 82-87.

National Research Council, 2008. *Ballistic Imaging*. The National Academies Press, Washington, D.C.

EVO, 2015a. “Evofinder”. 22 Mai. 2015 < http://evofinder.com/>.

EVO, 2015b. “Events”. 22 Mai. 2015 < http://evofinder.com/events/>.

Heard, B. J., 2008. *Handbook of firearms and ballistics: examining and interpreting forensic evidence*. John Wiley & Sons, Chichester, 2nd edition.

HGA, 2015. “Hand Gun Ammunition”. 14 Mai. 2015 < http://www.magtechammunition.com/products/handgun-ammunition/>.

Hill, R., Storakersand B., Zdunek A. B., 1989. “A theoretical study of the Brinell hardness test”. *Proceedings of the Royal Society of London. Series A, Mathematical and Physical Sciences*, Vol. 423, pp. 301-330.

IBIS, 2015a. “Ballistic Identification”. 22 Mai. 2015 < http://www.ultra-forensictechnology.com/ibis/>.

IBIS, 2015b. “Solution for a Safety Society”. 22 Mai. 2015 < http://www.ultra-forensictechnology.com/about#ibin >.

Kinder, J. D., 2002. “Ballistic fingerprinting databases”. *Sci. Justice*, Vol. 42, pp. 197-203.

Kopel, D. B., Burnett, H.S., 2003. “Ballistic Imaging: Not Ready for Prime Time”. *Polyce Background nº* 160.

León, F. P. , 2006. “Automated Comparison of firearms bullets”. *Forensic Science International*, Vol. 156, p. 40-50.

Lepus, 2015. “Lepus”. 22 Mai. 2015 < http://www.photonita.com.br/produtos/lepus/>.

Rahm, J., 2012. “Evaluation of an electronic comparison system and implementation of a quantitative effectiveness criterion”. *Forensic Science International*, Vol. 214, p. 173-177.

8. RESPONSIBILITY NOTICE

The authors are the only responsible for the printed material included in this paper.