Forensic Science International 280 (2017) 213–223

Contents lists available at ScienceDirect

Forensic Science International

journal homepage: www.elsevier.com/locate/forsciint

A convergence algorithm for correlation of breech face images based on the congruent matching cells (CMC) method

Zhe Chen[a,b,*], John Song[b], Wei Chu[b], Johannes A. Soons[b], Xuezeng Zhao[a]

[a] School of Mechatronics Engineering, Harbin Institute of Technology (HIT), Harbin, 150001, China
[b] Physical Measurement Laboratory, National Institute of Standards and Technology (NIST), Gaithersburg, MD 20899, USA

ARTICLE INFO

Article history:
Received 19 April 2017
Received in revised form 11 August 2017
Accepted 31 August 2017
Available online 8 September 2017

Keywords:
Forensics
Ballistics
Firearm evidence identification
Congruent Matching Cells (CMC)
Convergence

ABSTRACT

The Congruent Matching Cells (CMC) method was invented at the National Institute of Standards and Technology (NIST) for accurate firearm evidence identification and error rate estimation. The CMC method is based on the principle of discretization. The toolmark image of the reference sample is divided into correlation cells. Each cell is registered to the cell-sized area of the compared image that has maximum surface topography similarity. For each resulting cell pair, one parameter quantifies the similarity of the cell surface topography and three parameters quantify the pattern congruency of the registration position and orientation. An identification (declared match) requires a significant number of CMCs, that is, cell pairs that meet both similarity and pattern congruency requirements. The use of cell correlations reduces the effects of "invalid regions" in the compared image pairs and increases the correlation accuracy. The identification accuracy of the CMC method can be further improved by considering a feature named "convergence," that is, the tendency of the x–y registration positions of the correlated cell pairs to converge at the correct registration angle when comparing same-source samples at different relative orientations. In this paper, the difference of the convergence feature between known matching (KM) and known non-matching (KNM) image pairs is characterized, based on which an improved algorithm is developed for breech face image correlations using the CMC method. Its advantage is demonstrated by comparison with three existing CMC algorithms using four datasets. The datasets address three different brands of consecutively manufactured pistol slides, with significant differences in the distribution overlap of cell pair topography similarity for KM and KNM image pairs. For the same CMC threshold values, the convergence algorithm demonstrates noticeably improved results by reducing the number of false-positive or false-negative CMCs in a comparison.

Published by Elsevier Ireland Ltd.

1. Introduction

The parts of a firearm that make forcible contact with a cartridge case, such as firing pin, breech face and ejector, create toolmarks on the surface of the cartridge case [1], in firearm identification, these toolmarks are compared to assess whether two cartridge cases were fired from the same firearm. One challenge for the identification of breech face impressions is that, depending on the contact situation during firing, not all the regions of the samples may have been well impressed by the firearm breech face. The overall similarity of the compared surface images, and the accuracy of their registration, may be reduced by these "invalid" correlation regions [2].

In 2012, a new method for ballistics identifications, named Congruent Matching Cells (CMC), was invented at the National Institute of Standards and Technology (NIST) [3]. The CMC method is based on the principle of discretization. The surface image of the reference sample is divided into correlation cells. Each cell is registered to the cell-sized area of the compared image that has maximum surface topography similarity. For each resulting cell pair, one parameter quantifies the similarity of the cell surface topography and three parameters quantify the pattern congruency of the registration position and orientation. The similarity metric of the CMC method is the number of congruent matching cell pairs, i.e., the number of cell pairs that have both a sufficiently similar topography and a congruent registration pattern. Similar to the manual comparison process by a trained examiner, this approach reduces the confounding of "valid" and "invalid" regions, providing

* Corresponding author at: School of Mechatronics Engineering, Harbin Institute of Technology (HIT), Harbin, 150001, China.
E-mail addresses: zhe.chen@nist.gov (Z. Chen), jun-feng.song@nist.gov (J. Song), wei.chu@nist.gov (W. Chu), johannes.soons@nist.gov (J.A. Soons), zhaoxz@hit.edu.cn (X. Zhao).

http://dx.doi.org/10.1016/j.forsciint.2017.08.033
0379-0738/Published by Elsevier Ireland Ltd.

opportunities for improved discrimination of comparison results for same-source and different-source samples. Initial tests were performed using both 3D topography images [4] and 2D reflectance microscopy images [5] of breech face impressions from a set of 40 cartridge cases ejected from pistols with 10 consecutively manufactured slides. The tests showed no overlap between the CMC distributions for known matching (KM) and known non-matching (KNM) comparisons [4,5].

The accuracy of the CMC method can be further improved by considering a feature named "convergence", which means the tendency of the x–y registration positions of the correlated cell pairs to converge at the correct registration angle when comparing same-source samples at different relative orientations. In this paper, the concept and algorithms of the CMC method are reviewed in Section 2. In Section 3, we analyze the difference of the convergence feature between KM and KNM image correlations and develop the convergence algorithm. Validation tests using four datasets are conducted in Section 4, followed by conclusions in Section 5.

2. Drawbacks of current algorithms for the CMC method

2.1. Basic concept of the CMC method

The impression of a firearm's breech face on the primer of a cartridge case is affected by the firing conditions, primer material, pre-fire primer marks, contaminants, corrosion, and wear of the gun slide. As a result, both "valid" and "invalid" correlation regions may exist on the breech face impressions [3,6]. The "valid correlation region" represents complete contact between the cartridge case and the firearm parts in the firing process, and a transfer of "individual characteristics" [1] of the breech face which can be used effectively for identification. Conversely, due to incomplete contact, a lack of individualizing features on the firearm parts, contaminants, and other random factors, some regions may not contain individual characteristics and are termed the "invalid correlation regions" [3,6], which should be eliminated from correlations.

In the CMC method, the surface image of the reference sample is dividing into correlation cells. Each cell is registered to the cell-sized area on the image of the compared sample that has maximum similarity. For two topographies A and B, originating from the same firearm, the cell pairs located in the common valid correlation region of A and B can be characterized by [3,6]:

1) High topography similarity, quantified by, e.g., the normalized cross correlation function maximum CCF_{max},
2) Similar registration angles θ for the registered cells in the compared image, and
3) Similar x–y spatial distribution patterns for the cell arrays in both images.

The congruent matching cell (CMC) pairs can be determined by four identification parameters CCF_{max}, θ, x and y with the corresponding thresholds: T_{CCF}, T_θ, T_x and T_y. If the correlated cell pairs come from the common valid correlation regions of A and B, then their normalized correlation value CCF_{max} should exceed a chosen threshold T_{CCF} and their registration angles θ and x–y registration positions should fall within the chosen thresholds T_θ, T_x and T_y. In this paper, we define the registration position of a cell as the position of the center of the compared image that yields the highest normalized CCF value for the compared image area that coincides with the evaluated cell on the reference image. Cell pairs that meet all four threshold conditions are regarded as congruent matching cell pairs or CMCs. Inspired by the numerical identification criterion of the Consecutively Matching Striae (CMS) method developed by Biasotti and Murdock [7], the initial numerical identification criterion for the CMC method is chosen as $C = 6$, i.e., two images are concluded to be a "match" when the number of CMC pairs between them is at least 6. This initial criterion is being refined by experimental studies aimed at estimating the CMC score distributions for KM and KNM sample comparisons of different firearm and ammunition brands. For example, the KNM distribution can be used to choose a criterion value that yields an acceptable estimated false positive error rate. Alternatively, the distributions can be used to generate a quantitative expression for the weight of the evidence associated with the observed number of CMCs, such as a score-based likelihood ratio.

The application of tolerances in determining the CMC score, i.e., the number of CMCs, yields a similarity metric that has a simple conceptual interpretation, similar to the number of Consecutively Matching Striae for striated marks or the number of matching minutiae in fingerprint analysis. The tolerances are set using the distribution of the cell similarity and registration parameters for KM and KNM sample comparisons [4]. The downside of this approach is that the contribution of each evaluated cell to the similarity metric is either one or zero. An alternative approach, that we have not yet analyzed, is to define the CMC metric as the average of individual cell pair scores, each determined by some combination of cell pair similarity, closeness to the congruent location, and the number of pixels in a cell with measurement values.

2.2. Three algorithms for the CMC method

In this section, we describe the approach and challenges of three existing CMC algorithms. The algorithm properties are summarized in Table 1.

2.2.1. The initial CMC algorithm

In the initial CMC algorithm, the reference image A is divided into an array of cells. Cells containing insufficient data points (in

Table 1
Methods and challenges of existing CMC algorithms.

Algorithm	Method	Challenges
Initial CMC	Evaluate the congruency of the cell locations where cells yield the highest similarity	Missing CMC problem: a cell pair may be classified as a non-CMC because its maximum CCF value is attained at a location outside the CMC tolerance zone, even though it may have an acceptable CCF value within this zone
Total CMC	Evaluate the similarity of cells at locations that meet the congruency criteria	A higher burden on the choice of the similarity threshold (T_{CCF}) to reduce the number of CMCs for KNM image correlations, i.e., the number of false positive CMCs
High CMC	Use the initial CMC algorithm with a reduced angular search domain for image pairs that achieve a similar maximum number of CMCs at similar angles for forward and backward image correlations	Cannot distinguish KM and KNM image pairs thoroughly when the similarity difference between KM and KNM image pairs is small

our test, cells with less than 15% of their pixels representing measured data) are excluded from the analysis. At each discretized rotation angle of the compared image B, an effective cell in the reference image A is cross-correlated with the rotated image B to find the cell position on that image with the highest CCF value. In the initial algorithm, the CCF_{max}, x, y and θ parameters are only recorded at the position and angle that yield the maximum CCF value. This process is repeated for the other effective reference cells. The x, y and θ parameters of all the cell pairs are referenced to their median values x_{ref}, y_{ref} and θ_{ref} respectively. Then the correlation parameters of each cell pair are compared with the thresholds T_{CCF}, T_θ, T_x and T_y to determine whether the cell pair is a CMC [4,5].

Ideally, for the pairwise cell correlations of KM topographies, the maximum correlation value, CCF_{max}, should always be found within the angular and position thresholds T_θ, T_x, and T_y around the virtual reference locations θ_{ref}, x_{ref}, and y_{ref}, and the CCF values obtained outside these thresholds should be relatively small and therefore be excluded. However, case studies have shown that a cell pair may be classified as a non-CMC because its maximum CCF value is attained at a location outside the CMC tolerance zone, even though it may have an acceptable, but smaller, CCF value within this zone [8]. This effect can, for example, occur for a long smooth scratch where any cell location along the scratch may yield similar CCF values. An indication of this effect is the overlap of the cell CCF_{max} distributions of KM and KNM image pairs (to be discussed in Section 4). A large overlap indicates that there are many random cell locations that may meet a practical similarity threshold. Furthermore, cell pairs yielding the highest CCF but located at incongruent locations may also cause a deviation in the location of the virtual reference, which may impair the identification accuracy.

2.2.2. The total CMC algorithm

A modification to the CMC method was developed to address the challenges caused by the initial algorithm. In this modification, the search domain of the cells is constrained to locations that meet the congruency criteria. The CMC number is then obtained as the number of cells that meet the similarity criterion (T_{CCF}). This modification overcomes the "missing CMC" problem by evaluating the similarity of cells at locations that meet the congruency criteria, instead of evaluating the congruency of the cell locations where cells yield the highest similarity. Therefore, the algorithm always finds the maximum number of cell pairs that meet both the similarity and congruency criteria. As a result, the congruency thresholds T_θ, T_x and T_y can typically be set to much lower values without significantly affecting the number of CMCs for KM comparisons.

However, by constraining the evaluated cell locations to congruent locations, this modification places a higher burden on the choice of the similarity threshold (T_{CCF}) to reduce the number of CMCs for KNM image correlations, i.e., the number of false positive CMCs. For the same similarity threshold, this modification typically increases the number of CMCs for both KM and KNM image correlations (see Section 4). In this paper, we use a hybrid version of this algorithm, where the reference angle θ_{ref} is determined in the same way as in the initial algorithm, after which the search domain of the cells is restricted to locations that meet the position congruency criteria within the threshold range T_θ around θ_{ref}.

2.2.3. High CMC algorithm

In another algorithm, namely the high CMC algorithm [8], an additional criterion was introduced for the preliminary identification of matching and non-matching image pairs. The algorithm performs both "forward" and "backward" correlations at each rotation angle and evaluates the number of CMCs for both correlation directions as a function of the rotation angle. Compared to a forward correlation, a backward correlation reverses the role of the reference and comparison image. Especially for KNM comparisons, the reversal can yield different results as the contents of the reference cells and the searched compared surface are changed. For matching image pairs, the forward and backward correlations typically show a similar pattern for the number of CMCs as a function of the rotation angle, with peaks close to the correct registration angle. But for most non-matching image pairs, this pattern is quite different between the two correlation directions [8]. The algorithm employs this characteristic to declare a possible match if the range of the angles with CMC numbers at least equal to the "high CMC" criterion is within the angular congruency tolerance. Here the high CMC criterion is defined as ($CMC_{max} - \tau$), where τ is an empirical constant ($\tau = 1$ in Ref. [8]). If a possible match is identified, the angular search domain is reduced to the angular congruency tolerance to improve the respective CMC results.

However, our case studies show that when the difference in similarity between KM and KNM image pairs is small, the high CMC criterion cannot pre-distinguish KM and KNM image pairs thoroughly, which results in a higher number of false positive CMCs for KNM image correlations (see Section 4).

3. Methods

Ideally, we would like to identify as many CMCs for KM image correlations as the total CMC algorithm, while avoiding the additional false positive CMCs identified by this algorithm for KNM image correlations. To meet this goal, we use an additional criterion for the preliminary identification of possible matching or non-matching image pairs. With such a procedure, the initial and total CMC algorithms can be combined, with the advantages of both for identifying KM and KNM image pairs.

After reviewing KM image correlations using the CMC method, we observed a convergence of the x–y registration positions of the correlated cell pairs to a common center when the matching angle is reached, and conversely, a lack of convergence for the cells' x–y registration positions of KNM image correlations. Fig. 1 demonstrates how the cells' x–y registration positions change with the rotation angle for a KM image correlation. The blue circles in each

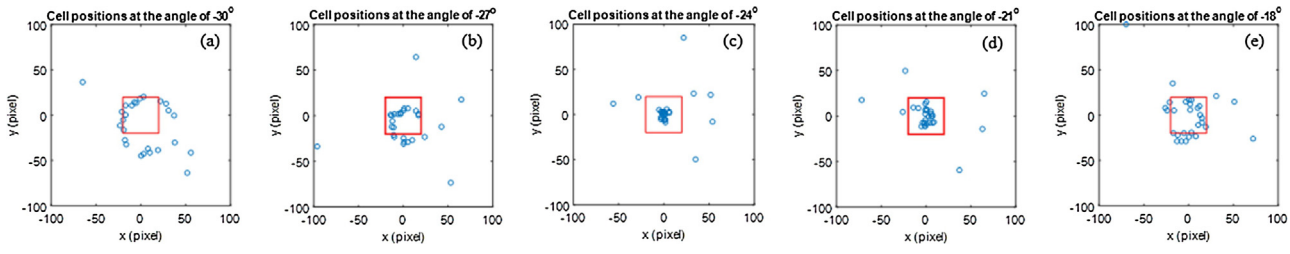


Fig. 1. The convergence of spatial registration positions in a KM image correlation (Fadul dataset [9], sample #01 vs. #02) as a function of rotation angle (a) −30°, (b) −27° to (c) −24°, then dispersion from (c) −24° to (d) −21°, (e) −18° (for interpretation of the references to colour in the text, the reader is referred to the web version of this article).

sub-graph of Fig. 1 represent the registration positions of all the cells at a rotation angle, which are referenced to their median position at that angle. The red square represents the x–y congruency thresholds T_x and T_y = 20 pixels, or 125 μm.

In the example of Fig. 1, the cell positions converge at a matching angle θ_0 of about −24° (Fig. 1c). This is because for cartridge cases ejected from the same firearm, the similarity of the individual characteristics on the breech face impression is represented not only by the high cell topography similarity, CCF_{max}, but also by the congruent distribution pattern of the registered cell locations. As the rotation angle approaches the common matching angle θ_0, the congruency produces a convergence of the x–y registration positions for the matching cell pairs. Furthermore, the cell positions should converge at the same rotation angle for both forward and backward correlations. The convergence angle can then be used as a new reference angle θ_{ref}' (equal to −24° here), which is more accurate than that calculated in the three algorithms previously described.

In contrast with Fig. 1, Fig. 2 shows how the cells' x–y registration positions change randomly with the rotation angle and fail to converge for a KNM image correlation. For cartridge cases ejected from different firearms, the dissimilarity of individual characteristics on the breech face impressions is represented by both the topography dissimilarity (low CCF_{max}) and non-congruent distribution patterns of the correlated cell pairs. Although at each rotation angle there may exist a few random cell pairs that meet the congruency requirements, KNM image pairs do not show the same degree of convergence as KM image pairs. As a result, there is often no clear convergence angle that can be found like that shown in KM image correlations. This failure to converge reflects the dissimilarity in the location of the individual characteristics on KNM image pairs. Furthermore, in most cases, rotation angles at the optimum convergence show large differences between the forward and backward correlations of KNM image pairs.

Based on the difference in the convergence feature between KM and KNM image correlations, a convergence algorithm for the CMC method is developed to distinguish KM image pairs from KNM image pairs. At each discretized rotation angle of the compared image, we use cross-correlation to obtain for each reference cell the x, y position of a cell-sized area on the compared image that yields the highest CCF value. We then calculate at each rotation angle the median of the thus obtained x–y registration positions of all the cells. Next, we calculate the median of the distances of the cell positions to the median position at each rotation angle. In this process, some outlier distance values are removed when they are more than three times the median absolute deviation (MAD) from the median distance [10], after which the median distance is recalculated. Fig. 3 shows how the median distance of the cells' registration positions changes with the rotation angle in both forward and backward correlations of KM and KNM image pairs. The blue circles indicate the median distance value of all the cell pairs, after the outlier removal, at each rotation angle. The red dashed lines indicate the minimum median distance and the corresponding "convergence angle".

Fig. 3 demonstrates two differences between KM and KNM image correlations:

a) The minimum median distance of a KM image pair is very small for both forward (see Fig. 3a) and backward (see Fig. 3c) correlations, while that of KNM image pairs (Fig. 3b and d) is relatively large; and
b) The convergence angle (reference angle) of KM image pairs is similar for forward and backward correlations, while those of KNM image pairs locate randomly.

Based on the above two differences, an additional criterion is proposed for correlations using the CMC method, defined by conditions (I) and (II):

(I) The minimum median distances of forward and backward correlations are both smaller than the threshold T_x or T_y; and
(II) The range of the reference angles between the forward and backward correlations, determined by the angles where the minimum median distance occurs, does not exceed the rotation step of the correlations.

In general, each image comparison is conducted using the following procedure:

1) Pre-process the image data (see Section 4) to isolate the breech face impression (segmentation) and highlight the individual characteristics (filtering).
2) At each rotation angle, conduct both forward and backward correlations and record the registration information x, y and CCF_{max} for each cell pair.
3) At each rotation angle, calculate the median distance of the cells' registration positions to the respective median registration position for both forward and backward correlations, as illustrated in Fig. 3.
4) If the minima in the median distances and the respective reference angles meet both the conditions (I) and (II), then this correlation is possibly a KM image correlation. For both forward and backward correlations, the CMCs will be obtained within the range $\pm T_\theta$ around the reference angle θ_{ref}' (convergence angle) using the total CMC algorithm.
5) If the median distance pattern does not meet conditions (I) and (II), then this correlation is possibly a KNM image correlation. For both forward and backward correlations, the CMCs will be obtained using the initial CMC algorithm with the respective reference angle θ_{ref}' (convergence angle) [4].
6) The number of CMCs is returned as the smallest of the CMC number of each correlation direction.

4. Validation test and results

Validation tests were conducted on four sets of cartridge case's breech face impressions, namely the Fadul [9], Weller [11], Hamby

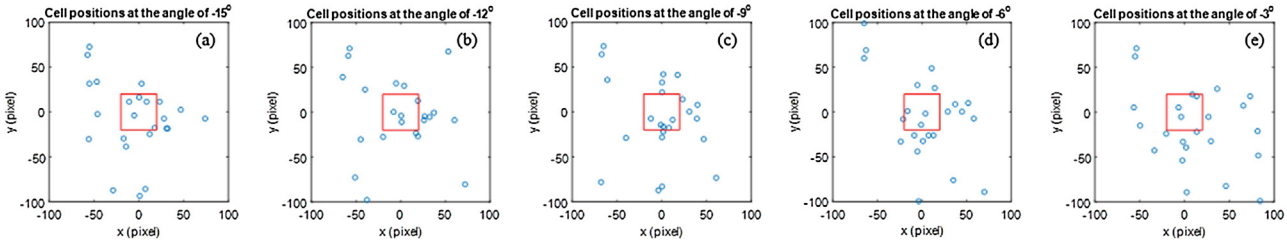
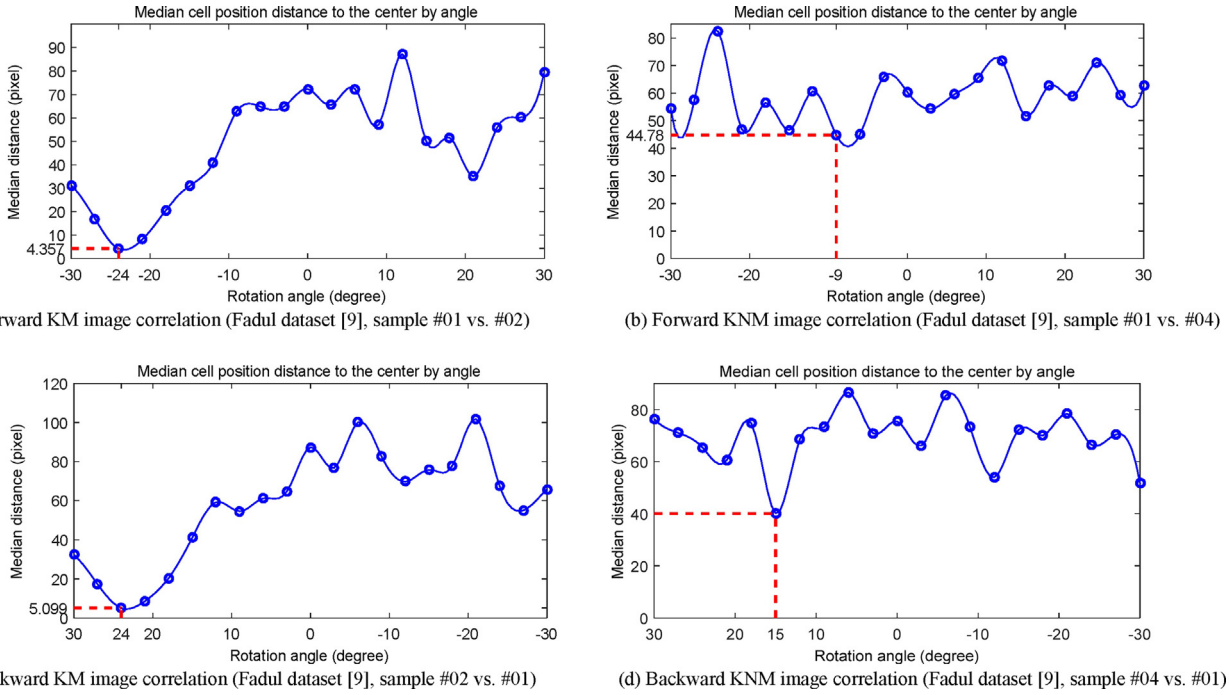


Fig. 2. KNM image correlation (Fadul dataset [9], sample #01 vs. #04) fails to show convergence of the cells' x,y positions at any rotation angle over a range from −15° to −3.



(a) Forward KM image correlation (Fadul dataset [9], sample #01 vs. #02)

(b) Forward KNM image correlation (Fadul dataset [9], sample #01 vs. #04)

(c) Backward KM image correlation (Fadul dataset [9], sample #02 vs. #01)

(d) Backward KNM image correlation (Fadul dataset [9], sample #04 vs. #01)

Fig. 3. Median distance of the cells' registration positions to the median registration position as a function of the rotation angle for forward and backward correlations of KM (a and c) and KNM (b and d) image pairs (for interpretation of the references to colour in the text, the reader is referred to the web version of this article).

[12] and Lightstone [13] datasets. The breech face impression topographies were measured by a disk scanning confocal microscope [14]. The original images have a nominal image area of about 3.8 mm × 3.8 mm, or approximately 1200 × 1200 pixels, with a nominal pixel spacing of 3.125 μm. The images were first trimmed to remove impression edge areas with strong roll-off, firing pin impression, firing pin drag marks and firing pin aperture shear. Then the images were downsampled to a pixel spacing of 6.25 μm to increase the speed of correlation. Finally, the images were band pass filtered using a 2nd order robust Gaussian regression filter to reduce noise with short wavelengths and to attenuate waviness and form with long wavelengths, thus highlighting individual characteristics. The short wavelength cut-off of the filter was 16 μm, and the long wavelength cut-off was 250 μm. For the CMC method, the cell size was chosen as 75 × 75 pixels (or 468.75 μm × 468.75 μm), and the angular rotation range was ±30° with 3° increments.

The Fadul set of cartridge cases was created at the Miami-Dade Crime Laboratory for a study of visual firearm identifications by ballistics examiners [9]. It contains 40 cartridge cases ejected from Ruger P95PR15[1] pistols with 10 consecutively manufactured slides. The Weller set [11] consists of 95 cartridge cases obtained from another set of 11 Ruger P95DC firearm slides, the first 10 of which were consecutively manufactured: 9 cartridge cases from each of the consecutively manufactured slides and 5 cartridge cases from the additional slide. The Hamby set consists of 30 cartridge cases from 10 consecutively finished Hi-Point C9 pistol slides [12], where each slide was used to generate 3 cartridge cases. The respective firings were conducted by NIST. The Lightstone set consists of 30 cartridge cases from 10 consecutively manufactured SW40VE Smith and Wesson Sigma pistol slides [13], where each slide was used to generate 3 cartridge cases. Some details of the four datasets are shown in Table 2. Fig. 4 shows typical impression topographies for the four datasets after trimming and filtering, which can be roughly characterized as: granular for Fadul/Weller, striated for Hamby, and striated/granular for Lightstone.

Each topography image was divided into an array of 8 × 8 cells, so the nominal number (N_{nom}) of the correlated cell pairs for each topography correlation was 64. The actual number of cells N involved in each image pair correlation is less than N_{nom}, because a cell was not used unless at least 15% of its pixels represented measured data. The thresholds T_θ, T_x and T_y were set as $T_\theta = 3°$, T_x and $T_y = 20$ pixels (125 μm). The similarity threshold T_{CCF} was set as $T_{CCF} = 40\%$. Fig. 5 shows the observed CCF_{max} distribution of pairwise cell correlations of KM and KNM image pairs for the four datasets, obtained using the initial CMC algorithm. The overlap of the two distributions indicates the closeness between the cell topography similarity of KM and KNM image pairs when the cell registration location is not constrained by congruency requirements. In other words, it describes the closeness between the best topography of each cell sized area of KM and KNM samples. A low degree of overlap, combined with a suitable choice for the cell similarity threshold T_{CCF}, reduces the burden for the congruency evaluation to correctly identify CMCs. For a given threshold T_{CCF}, this challenge can be quantified by two distribution parameters: the false cell similarity identification frequency $f_{1(CCF)}$ and the false cell similarity exclusion frequency $f_{2(CCF)}$, calculated as [6]:

$$f_{1(CCF)} = \frac{\text{Number of KNM cell pairs whose } CCF_{max} \geq T_{CCF}}{\text{Total number of evaluated KNM cell pairs}} \quad (1)$$

$$f_{2(CCF)} = \frac{\text{Number of KM cell pairs whose } CCF_{max} < T_{CCF}}{\text{Total number of evaluated KM cell pairs}} \quad (2)$$

These estimated cell frequencies only address one of the requirements for a cell to be classified as a CMC, namely a sufficient surface topography similarity. The additional congruency requirement will result in an average false cell identification probability

[1] Certain commercial equipment, instruments, or materials are identified in this paper in order to specify the experimental procedure adequately. Such identification is not intended to imply recommendation or endorsement by the National Institute of Standards and Technology, nor is it intended to imply that the materials or equipment identified are necessarily the best available for the purpose.

Table 2
Details about the Fadul, Weller, Hamby and Lightstone sets.

Name	Topography characteristic	Number of cartridge cases	Number of KM image pairs	Number of KNM image pairs	$f_{1(CCF)}$	$f_{2(CCF)}$
Fadul	Granular	40	63	717	34.92%	12.34%
Weller	Granular	95	370	4095	16.60%	8.89%
Hamby	Striated	30	30	405	75.32%	8.40%
Lightstone	Granular/striated	30	30	405	29.98%	31.11%

for KNM comparisons that is equal to or lower than $f_{1(CCF)}$. On the other hand, the average false cell exclusion probability for KM comparisons will be equal to or higher than $f_{2(CCF)}$. Therefore, low values for $f_{1(CCF)}$ and $f_{2(CCF)}$ provide improved conditions for the congruency evaluation to achieve low number of false positive or false negative CMCs. For each dataset, the estimated $f_{1(CCF)}$ and $f_{2(CCF)}$ are listed in Table 2. For Fadul and Weller datasets (Fig. 5a and b), both frequencies are relatively small: $f_{1(CCF)} = 34.92\%$, $f_{2(CCF)} = 12.34\%$ for the Fadul dataset; and $f_{1(CCF)} = 16.60\%$, $f_{2(CCF)} = 8.89\%$ for the Weller dataset. Compared to the Hamby and Lightstone datasets, the respective KM and KNM distributions (Fig. 5) show a lower degree of overlap. This may due to their granular surface topographies obtained by a sand or bead blasting finishing process. The resulting random pattern of pits provides fewer opportunities for KNM comparisons to achieve a high cell topography similarity, provided the correlation cells are large enough to contain multiple pits. For the Hamby and Lightstone datasets, the distribution overlap, and the $f_{1(CCF)}$ and $f_{2(CCF)}$ values, are relatively large. For the Hamby dataset (Fig. 5c), $f_{1(CCF)} = 75.32\%$ which represents the highest $f_{1(CCF)}$ among the four datasets. The large difference between the $f_{1(CCF)}$ and $f_{2(CCF)}$ as shown in Fig. 5c indicates that $T_{CCF} = 40\%$ is selected too low for the Hamby dataset. The high $f_{1(CCF)}$ value indicates a higher burden for the congruency evaluation to avoid false positive identifications. But such a low T_{CCF} and high $f_{1(CCF)}$ can be used as a challenge to the convergence algorithm to see whether it can correctly identify a low number of false positive CMCs in KNM image pairs. For the Lightstone dataset (Fig. 5d), $f_{2(CCF)} = 31.11\%$ which represents the highest $f_{2(CCF)}$ among the four datasets. The high $f_{2(CCF)}$ is another challenge to the convergence algorithm to see whether it can still achieve a significant number of CMCs in KM image correlations. The larger distribution overlap and higher $f_{1(CCF)}$ and $f_{2(CCF)}$ values for the Hamby and Lightstone

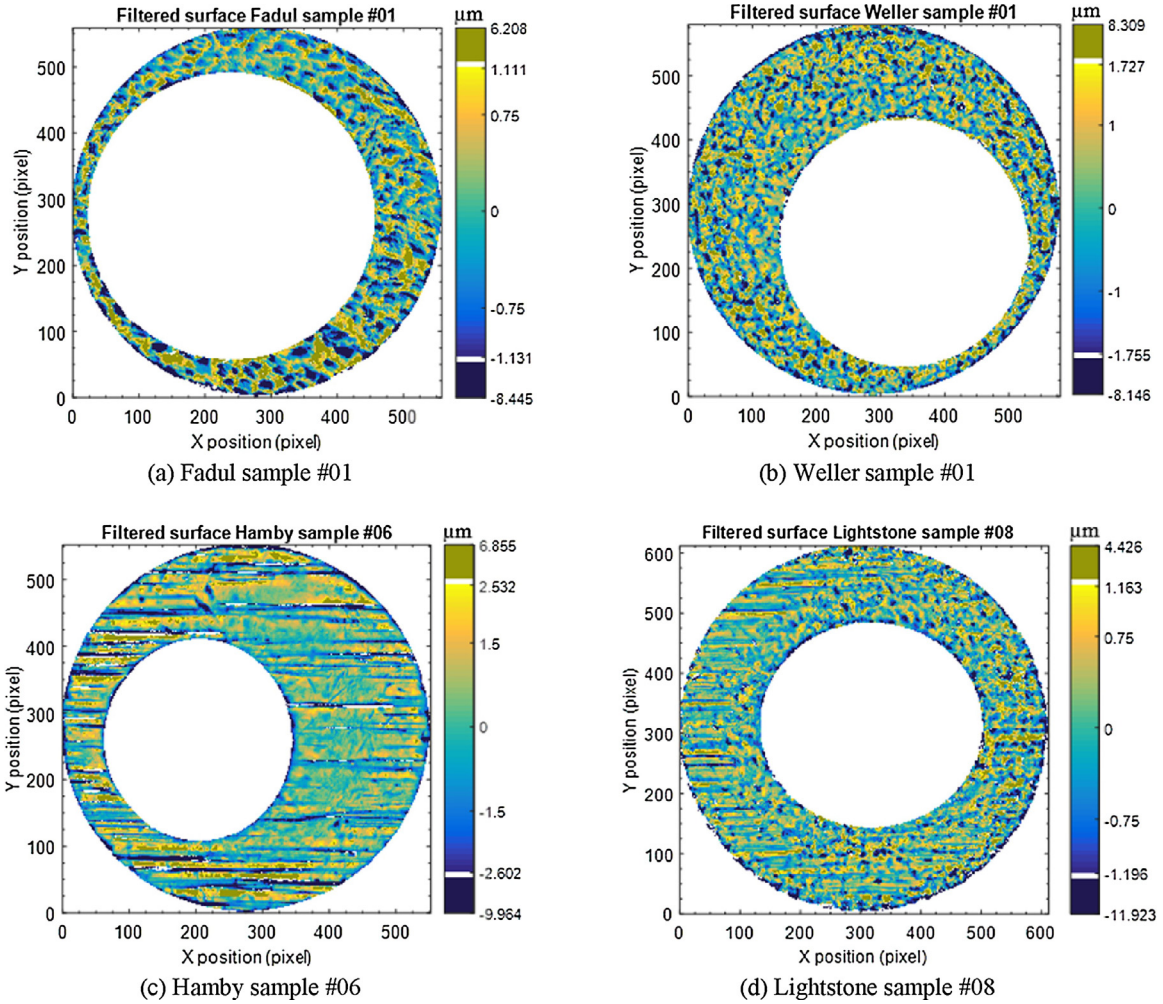


Fig. 4. Topography features of the filtered breech face impression of the four datasets: granular characteristics for Fadul/Weller (a and b), striated characteristics for Hamby (c), and striated/granular characteristics for Lightstone (d).

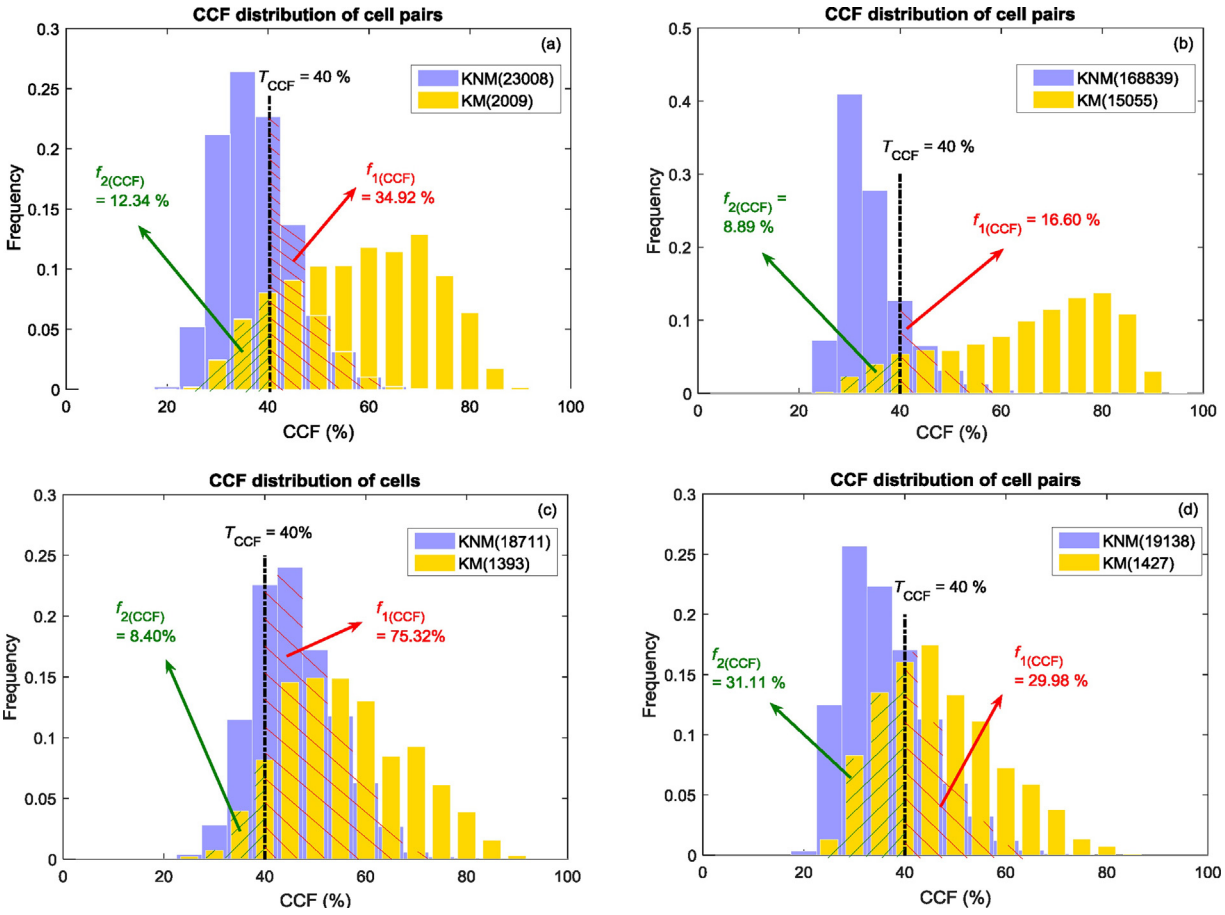


Fig. 5. CCF_{max} distribution of cell pairs in KM and KNM image correlations for the (a) Fadul, (b) Weller, (c) Hamby and (d) Lightstone datasets. The number of cell pairs in each histogram is shown in the legend. The Weller KNM histogram contains the most cell pairs while the Hamby KM histogram contains the least cell pairs.

datasets may relate to their striated surface topographies which show less topography difference between KM and KNM image pairs at the individual (unconstrained) cell level.

Fig. 6 and Table 3 compare results of the four algorithms for the Fadul dataset [9]. For the 63 KM image pairs, the CMC distribution range is 7–31 for the initial algorithm (Fig. 6a), which increases to 15–32 for the other algorithms (Fig. 6b, c and d). For the 717 KNM image pairs, the CMC distribution ranges are the same for the initial and convergence algorithms (both 0 to 3), smaller than those for the total (0 to 7) and high CMC algorithms (0 to 6). The convergence algorithm also yields fewer KNM image correlation results at the high CMC end (only 1 image pair with 3 CMCs) than the initial algorithm (8 image pairs with 3 CMCs). This is because for potential non-matching comparisons, as determined by conditions I and II, the convergence algorithm evaluates the number of CMCs using the initial algorithm, but with an angular tolerance zone that is now centered at the reference angle (convergence angle) θ_{ref}'.

The results for this dataset show that: (1) The initial algorithm has more false negative CMCs than the other algorithms. E.g., the image pair with 7 CMCs for the initial algorithm (Fig. 6a) has 16 CMCs for the total CMC algorithm (Fig. 6b); (2) The total CMC algorithm (Fig. 6b) yields more false positive CMCs than the initial algorithm (Fig. 6a); (3) The high CMC algorithm cannot distinguish KM and KNM image pairs thoroughly (2 KNM image pairs with 6 CMCs in Fig. 6c); (4) The convergence algorithm not only identifies as many CMCs in KM image correlations as the total algorithm, but also reduces the number of false positive CMCs in KNM image correlations to levels lower than the initial algorithm. This indicates that the additional identification criterion based on the convergence feature improves separation between KM and KNM CMC distributions by achieving the benefits of the total CMC algorithm (high CMC numbers for KM comparisons) while improving the benefits of the initial CMC algorithm (lower CMC numbers for KNM comparisons).

Fig. 7 and Table 4 compare results of the four algorithms for the Weller dataset [11]. For the 370 KM image pairs, the CMC distribution ranges are almost the same for all algorithms: 15–45 for the initial algorithm, 15–46 for the other algorithms. For the 4095 KNM image pairs, the CMC distribution ranges are the same

Table 3
CMC results for the Fadul dataset [9].

Algorithm	KM CMC range	KNM CMC range	No. of KNM image pairs at high CMC end
Initial	7–31	0–3	8
Total CMC	15–32	0–7	1
High CMC	15–32	0–6	2
Convergence	15–32	0–3	1

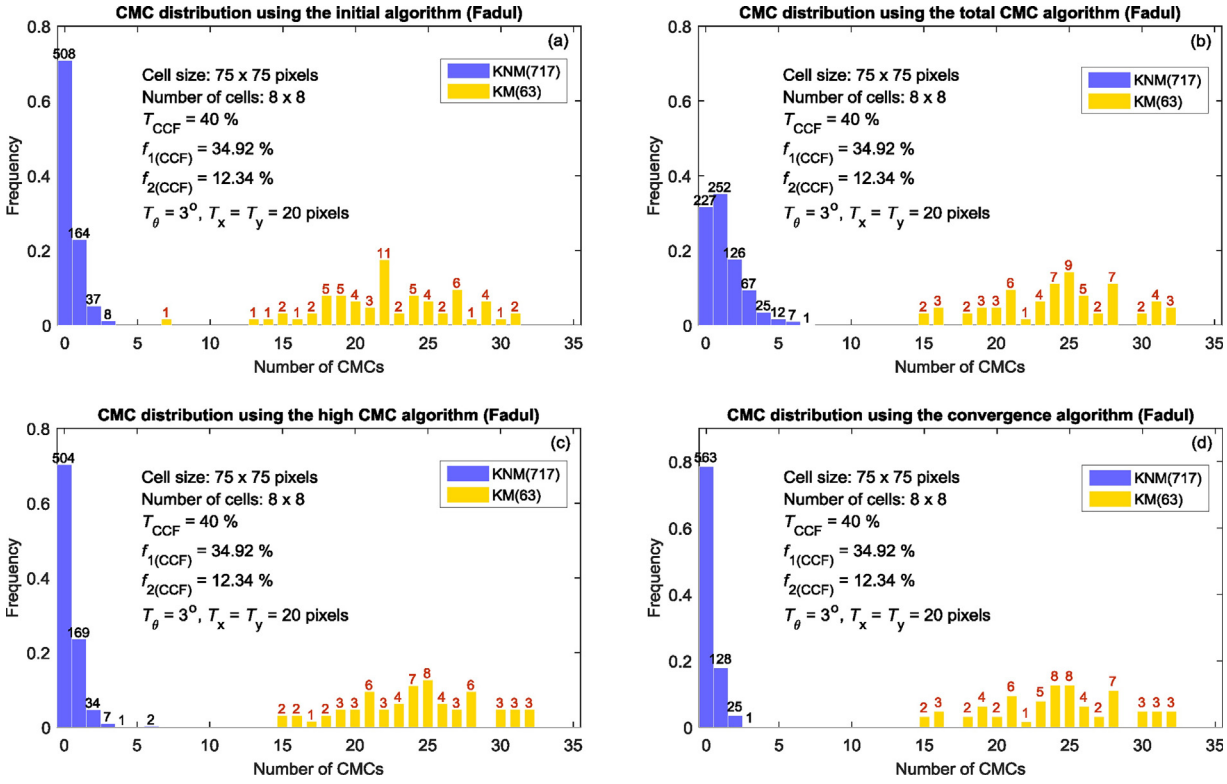


Fig. 6. CMC distribution of the Fadul dataset [9] using (a) the initial algorithm, (b) the total CMC algorithm, (c) the high CMC algorithm and (d) the convergence algorithm.

for the initial and convergence algorithms (0 to 3), which are smaller than those for the total and high CMC algorithms (0 to 5). Results for this dataset indicate that when the cell similarity difference between KM and KNM image pairs is significant, as indicated in Fig. 5b by the relatively small overlap of the cell CCF_{max} distributions which enables low $f_{1(CCF)}$ and $f_{2(CCF)}$ values, all the algorithms can be used to effectively identify the CMCs. The real challenge for the CMC algorithms occurs for datasets with less cell

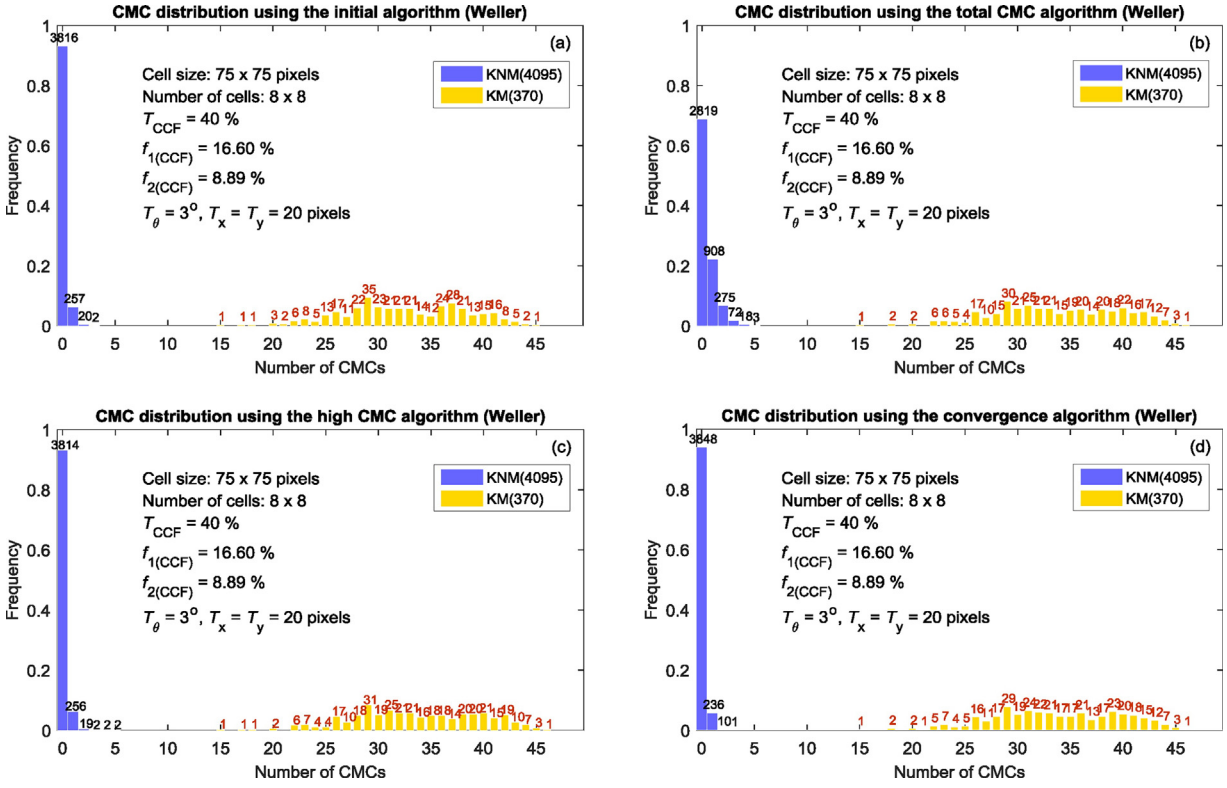


Fig. 7. CMC distribution of the Weller dataset [11] using (a) the initial algorithm, (b) the total CMC algorithm, (c) the high CMC algorithm and (d) the convergence algorithm.

Table 4
CMC results for the Weller dataset [11].

Algorithm	KM CMC range	KNM CMC range	No. of KNM image pairs at high CMC end
Initial	15–45	0–3	2
Total CMC	15–46	0–5	3
High CMC	15–46	0–5	2
Convergence	15–46	0–3	1

Table 5
CMC results for the Hamby dataset [12].

Algorithm	KM CMC range	KNM CMC range	No. of KNM image pairs at high CMC end
Initial	15–29	0–10	1
Total CMC	23–37	0–18	1
High CMC	23–37	0–18	1
Convergence	22–35	0–10	1

topography difference between KM and KNM image pairs, as shown by the relatively large overlap of the cell CCF_{max} distributions and high $f_{1(CCF)}$ and/or $f_{2(CCF)}$ values for the Hamby and Lightstone datasets (Fig. 5c and d).

Fig. 8 and Table 5 compare results of the four algorithms for the Hamby dataset [12]. For the 30 KM image pairs, the CMC distribution range is 15–29 for the initial algorithm (Fig. 8a), which increases to 23–37 for the total and high CMC algorithms (Fig. 8b and c), then decreases a little to 22–35 for the convergence algorithm (Fig. 8d). The small decrease is mainly due to small differences in the reference angle used by the respective algorithms. For the 405 KNM image pairs, the CMC distribution ranges are the same for the initial and convergence algorithms (0 to 10), which are smaller than those for the total and high CMC algorithms (0 to 18). The number of KNM image pairs with 0 CMCs is 131 for the convergence algorithm, which is the most of all algorithms. Even with a low T_{CCF} and high $f_{1(CCF)}$ at 75.32%, the convergence algorithm can still improve the correlation results by identifying a large number of CMCs for KM image pairs while reducing the number of false positive CMCs for KNM image pairs. The other three algorithms yield either more false negative CMCs for KM image pairs (Fig. 8a) or more false positive CMCs on KNM image pairs (Fig. 8b and c).

Fig. 9 and Table 6 compare results of the four algorithms for the Lightstone dataset [13]. For the 30 KM image pairs, the CMC distribution range is 9–36 for the initial algorithm, which increases to 13–40 for the other three algorithms. The CMC distribution range for the 405 KNM image pairs, from small to large, is 0 to 4 for the convergence algorithm, 0 to 5 for the initial algorithm, 0 to 8 for the high CMC algorithm and 0 to 10 for the total algorithm. The number of KNM image pairs with 0 CMCs is 200 for the convergence algorithm, which is the most of all algorithms.

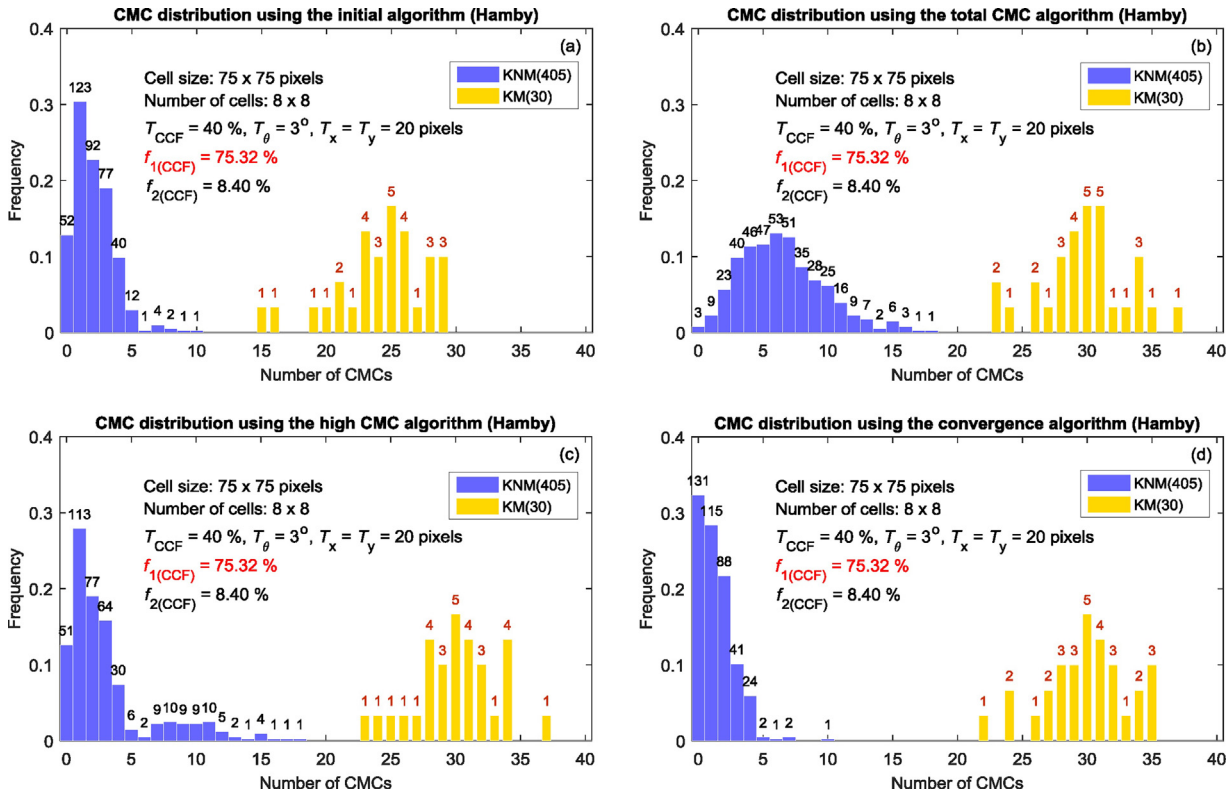


Fig. 8. CMC distribution of Hamby dataset [12] using (a) the initial algorithm, (b) the total CMC algorithm, (c) the high CMC algorithm and (d) the convergence algorithm.

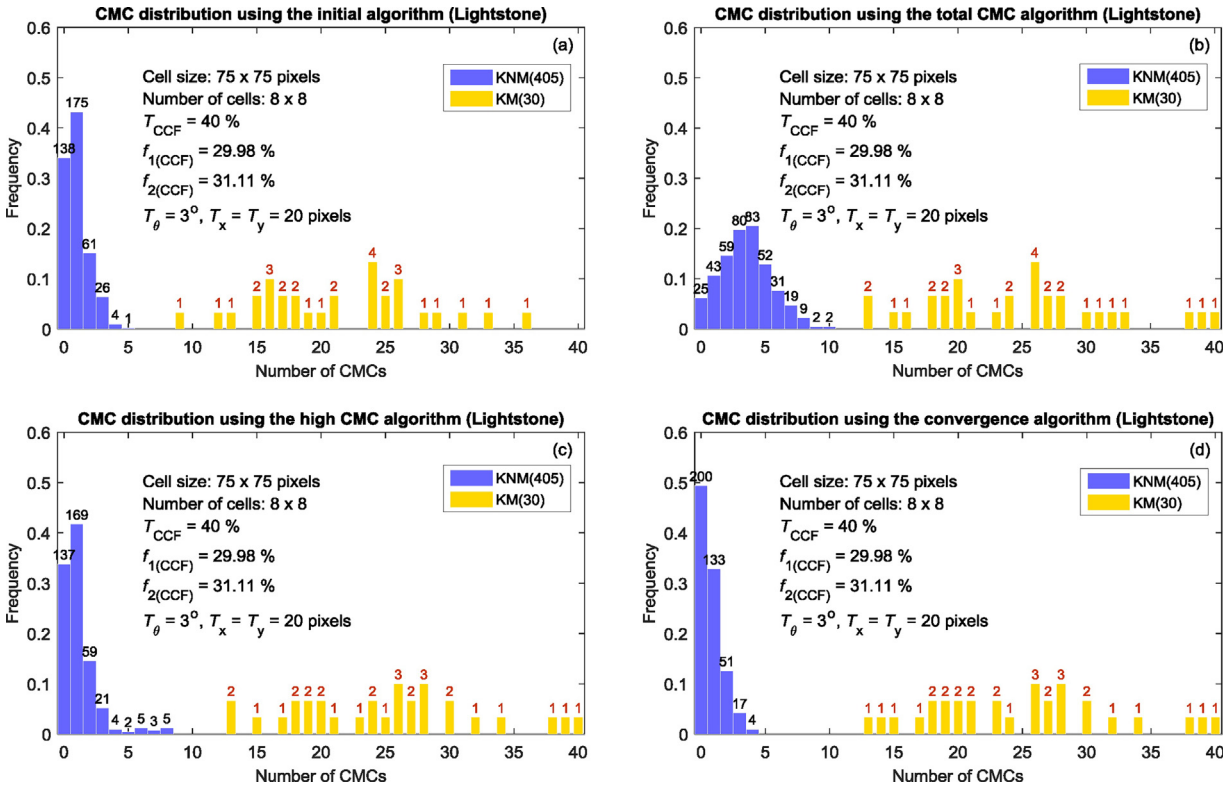


Fig. 9. CMC distribution of the Lightstone dataset [13] using (a) the initial algorithm, (b) the total CMC algorithm, (c) the high CMC algorithm and (d) the convergence algorithm.

Table 6
CMC results of the Lightstone dataset [13].

Algorithm	KM CMC range	KNM CMC range	No. of KNM image pairs at high CMC end
Initial	9–36	0–5	1
Total CMC	13–40	0–10	2
High CMC	13–40	0–8	5
Convergence	13–40	0–4	4

The Hamby and Lightstone datasets have relatively small cell topography similarity differences between KM and KNM image pairs. Compared to the other datasets, the respective correlation results show that the total and high CMC algorithms identify more false positive CMCs in KNM image correlations. For all datasets, the initial CMC algorithm yields more false negative CMCs in the KM image correlations. In contrast, the convergence algorithm significantly improves the correlation results for all datasets, with approximately the same number of CMCs as, respectively, the total CMC algorithm for KM image correlations and the initial CMC algorithm for KNM image correlations.

5. Discussion and conclusion

We developed a convergence algorithm that combines the advantage of the initial CMC algorithm for KNM comparisons with the advantage of the total CMC algorithm for KM comparisons. The new algorithm exploits a feature named “convergence”, that is, the tendency of the *x*–*y* registration positions of the correlated cell pairs to converge at the correct registration angle when comparing KM samples at different relative orientations. For each discretized rotation angle of the compared image, the convergence algorithm records for every reference cell the registration position of the compared image that yields the highest value for the similarity metric. The algorithm then calculates for every angle the median distance of these cell registration positions to their median position at that angle. The median distance as a function of rotation angle shows significant differences between KM and KNM comparisons. For KM image pairs, the minimum median distance of the cell positions is very small and the respective convergence angle is consistent between forward and backward correlations. For KNM image pairs, the minimum median distance of the cell positions is relatively large and the convergence angle is not consistent between forward and backward correlations. Based on two new threshold criteria that address these differences, the convergence algorithm applies the initial CMC algorithm to possible KNM image pairs and the total CMC algorithm to possible KM image pairs, thus combining the best features of these algorithms.

Validation tests show that, for the Fadul and Weller [9,11] datasets with relatively large differences between the cell topography similarity of KM and KNM image pairs, all the four algorithms show significant separation between the number of CMCs obtained for KM and KNM comparisons, with the convergence algorithm performing the best. For the Hamby and Lightstone datasets [12,13], which show relatively small differences between the cell topography similarity of KM and KNM image pairs, the convergence algorithm demonstrates noticeably

improved separation when using the same CMC threshold values. The validation tests indicate that the convergence feature represents a fundamental difference between KM and KNM image correlations. With this additional feature, the capability and accuracy of identification using the CMC method is significantly improved.

Acknowledgements

The funding for this work is provided by the Special Programs Office (SPO) of NIST. The authors are grateful to D. Ott of NIST for assistance with the algorithm and test design; to X. A. Zheng of NIST for providing the topography images; to T. V. Vorburger and J. Libert of NIST for their review and helpful comments on the manuscript; and to R. M. Thompson of NIST for his insights on ballistics identification.

References

[1] Scientific Working Group for Firearm and Toolmarks (SWGGUN), The Foundations of Firearm and Toolmark Identification, (2013) .
[2] J. Song, W. Chu, M. Tong, J. Soons, 3D topography measurements on correlation cells—a new approach to forensic ballistics identifications, Meas. Sci. Technol. 25 (6) (2014) 064005.
[3] J. Song, Proposed NIST ballistics identification system (NBIS) using 3D topography measurements on correlation cells, AFTE J. 45 (2) (2013) 184–194.
[4] W. Chu, M. Tong, J. Song, Validation tests for the congruent matching cells (CMC) method using cartridge cases fired with consecutively manufactured pistol slides, AFTE J. 45 (4) (2013) 361–366.
[5] M. Tong, J. Song, W. Chu, R.M. Thompson, Fired cartridge case identification using optical images and the congruent matching cells (CMC) method, J. Res. Natl. Inst. Stand. Technol. 119 (2014) 575.
[6] J. Song, Proposed "congruent matching cells (CMC)" method for ballistic identification and error rate estimation, AFTE J. 47 (3) (2015) 177–185.
[7] Biasotti, J. Murdock, Firearms and toolmark identification: legal issues and scientific status, in: D. Faigman, D. Kaye, M. Saks, J. Sanders (Eds.), Modern Scientific Evidence: The Law and Science of Expert Testimony, West Publishing Co., St. Paul, 1997, pp. 124–151.
[8] M. Tong, J. Song, W. Chu, An improved algorithm of congruent matching cells (CMC) method for firearm evidence identifications, J. Res. Natl. Inst. Stand. Technol. 120 (2015) 102.
[9] T.G. Fadul Jr., G.A. Hernandez, S. Stoiloff, S. Gulati, An Empirical Study to Improve the Scientific Foundation of Forensic Firearm and Tool Mark Identification Utilizing 10 Consecutively Manufactured Slides, NIJ Report No. 237960, National Institute of Justice, 2012.
[10] C. Leys, et al., Detecting outliers: do not use standard deviation around the mean, use absolute deviation around the median, J. Exp. Soc. Psychol. 49 (4) (2013) 764–766.
[11] T.J. Weller, A. Zheng, R. Thompson, F. Tulleners, Confocal microscopy analysis of breech face marks on fired cartridge cases from 10 consecutively manufactured pistol slides, J. Forensic Sci. 57 (4) (2012) 912–917.
[12] D. La Porte, An empirical validation study of breechface marks on .380 ACP caliber cartridge cases fired from ten consecutively finished hi-point model C9 pistols, AFTE J. 43 (4) (2011) 303–309.
[13] L. Lightstone, The potential for and persistence of subclass characteristics on the breech faces of SW40VE Smith and Wesson sigma pistols, AFTE J. 42 (4) (2010) 308–322.
[14] D.K. Hamilton, T. Wilson, Three-dimensional surface measurement using the confocal scanning microscope, Appl. Phys. 27 (4) (1982) 211–213.