# Reliability Assessment of Principal Point Estimates for Forensic Applications

Massimo Iuliani[a,c], Marco Fanfani[b], Carlo Colombo[b], Alessandro Piva[b,c,*]

[a] *Dept. of Mathematics and Computer Science, University of Florence, Firenze, Italy*
[b] *Dept. of Information Engineering, University of Florence, Firenze, Italy*
[c] *FORLAB Forensic Science Laboratory, University of Florence, Prato, Italy*

## Abstract

Although quite recent as a forensic research domain, computer vision analysis of scenes is likely to become more and more important in the near future, thanks to its robustness to image alterations at the signal level, such as image compression and filtering. However, the experimental assessment of vision-based forensic algorithms is a particularly critical task, since they cannot be tested on massive amounts of data, and their performance can heavily depend on user skill. In this paper we investigate on the accuracy and reliability of a vision-based, user-supervised method for the estimation of the camera principal point, to be used in cropping and splicing detection. Results of an extensive experimental evaluation show how the estimation accuracy depends on perspective conditions as well as on the selected image features. Such evidence led us to define a novel visual feature, referred to as Minimum Vanishing Angle, which can be used to assess the reliability of the method.

*Keywords:* Image Forensics, Scene level analysis, Geometric Constraints, Minimum Vanishing Angle, Cropping detection, Splicing detection.

## 1. Introduction

Image Forensics has been proposed as a solution for authenticating the contents of digital images [1, 2, 3]. This technology is based on the observation that each phase of the image history — from the acquisition process, through

*Corresponding author, Email: alessandro.piva@unifi.it, via S. Marta 3, 50139 Firenze Italy.

its storage in a compressed format, to any editing operation — leaves distinctive traces on the data, as a sort of digital fingerprint [4]. It is then possible to determine whether a digital image is authentic or modified, by detecting the presence, the absence or the incongruence of such traces, that are intrinsically tied to the digital content itself. Forensic traces can be found both at "signal level" (invisible footprints introduced in the signal statistics, like demosaicing artifacts [5], sensor noise [6], or compression artifacts [7, 8]) and at "scene level" (inconsistencies in shadows [9], lighting [10, 11], or in perspective and geometry of objects [12, 13]). The former are typically detected by automatic methods, but they often exhibit lower effectiveness when the investigated content has been subjected to an unknown chain of processes (e.g., filtering, resizing, compression) that may partially or completely spoil the traces left by previous operations [14]. The latter usually require particular constraints on the scene (e.g. the presence of Lambertian convex surfaces for lighting estimation [15]) but have the advantage of being robust to common image processing operations, thus appearing suitable even for low resolution images, or when the content has undergone multiple compressions. While in the literature a great effort has been devoted to evaluate the performance of signal-based forensic methods in terms of detection accuracy and reliability, a limited analysis has been carried out until now on scene-based techniques. This is mainly due to the fact that such algorithms are usually tested on small datasets only, since they cannot exclude some human intervention, e.g. image feature selection or analysis supervision.

This paper represents — to the best of our knowledge — the first attempt to analytically evaluate the performance of a scene level trace. In particular, we addressed the problem of estimating the camera principal point (PP) (whose position in the image under analysis is usually detected by exploiting vanishing points related to three mutually orthogonal directions [16]); whose application in a forensic scenario has been proposed in some recent works [17, 18, 19]. For our evaluation, several tests have been performed, on both synthetic and on real images, by varying both the point of view — so as to obtain different perspective conditions — and the number and position of the extracted features. A critical study of the obtained results has led us to define a novel feature, referred to as *Minimum Vanishing Angle* (MVA), allowing us to measure the reliability of the estimated PP. Using the MVA concept, we have also been able to establish a feature selection criterion. Specifically, one should just care about choosing the image lines that provide the widest possible MVA, since the accuracy of PP estimation relies more on MVA amplitude than on the amount of data (i.e. image lines) used.

The paper is organized as follows: in Section 2 the State of the Art is briefly presented, and in Section 3 we briefly review the theory behind the adopted PP estimation method. In Section 4 we introduce the MVA and its relation with the image perspective conditions. Then in Section 5 an in deep analysis of the reliability of the method is given. Section 6 presents two possible forensic applications of the PP: cropping detection — for which we provide a detailed accuracy analysis — and splicing detection. Section 7 concludes the paper and summarizes the contributions in light of the achieved results.

## 2. State of the Art

The estimation of the PP from a single image is a known issue in computer vision and photogrammetry, usually embedded into the camera calibration problem [20, Chapter 2]. In order to calibrate the camera, accurate off-line techniques usually require a known pattern in the scene [21, 22]. Other methods use video sequences or multiple images to self-calibrate the camera while solving the Structure from Motion problem [23]. In addition, other scene elements such as coaxial circles, or Manhattan-World structure [24] can be exploited for calibration tasks [25, 26, 27, 28].

Reported methods assume to use genuine images only, without any malicious modification. This hypothesis allows the authors to impose constraints on the parameters to ease and improve the estimation (for example, the PP is often initialized in the image center). In a forensic application scenario, however, this assumption doesn't hold; Moreover, we have to typically deal with single images already acquired. So, a calibration approach has to exploit useful characteristics of the scene. Given the abundance of images depicting man-made environments, we focus on techniques based on the Manhattan-World assumption.

Given these difficulties, in the forensic literature only a few methods have been presented that try to exploit the camera PP as a clue for tampering detection. In [17], the authors presented a method based on the estimation of the homography mapping a person's eyes to the image plane. Then, the PP is recovered by homography decomposition (supposing focal length is known) and exploited for splicing detection. A similar approach, that exploits circles in the scene to obtain the PP position, is presented in [18]. In [19], the authors notice that asymmetric cropping of an image introduces a correspondent shift of the principal point. Hence, they suggested that the distance between the estimated PP and the image center can be exploited as evidence of cropping. Slightly different, but still related to this topic, is the approach described in [29] where, instead of estimating the PP, tampering detection is based on the direct observation of the vanishing points of different 3D structures (e.g. buildings).

## 3. Principal Point Estimation

The mapping between the 3D world and its 2D images is usually modeled as a central projection of a world point onto the image plane (pinhole model [30], see Fig. 1a). The projection rule can be formally written as $\mathbf{m} = K[I|\mathbf{0}]\mathbf{M}$, where $\mathbf{m} = (x, y, 1)^\top$ and $\mathbf{M} = (X, Y, Z, 1)^\top$ are the homogeneous coordinates of a 2D image point and its corresponding 3D world point respectively, whereas $K$ is the camera matrix, embedding the internal parameters of the acquisition device. $I$ is the identity matrix, and $\mathbf{0}$ a column vector of zeros. Typically, the camera matrix is represented as

$$K = \begin{bmatrix} f & s & p_x \\ 0 & \rho f & p_y \\ 0 & 0 & 1 \end{bmatrix}, \tag{1}$$

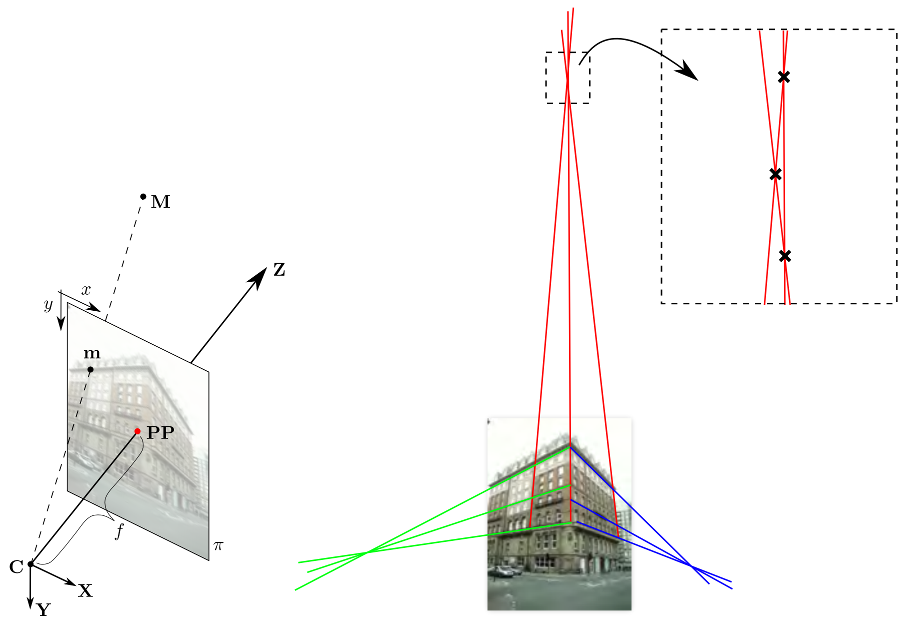
where $f$ is the focal length, while the aspect ratio $\rho$ and skew $s$ take into account the actual shape of a pixel. Lastly, $(p_x, p_y)$ are the coordinates of the PP (see again Fig. 1a). Modern cameras have reached a high level of quality, with unity aspect ratio and zero skew. So, without significant loss of accuracy, the $K$ matrix can be modeled with $\rho = 1$ and $s = 0$, passing from 5 to 3 degrees of freedom [31].

To obtain the PP, we can exploit the relation among three vanishing points, related to mutually orthogonal directions in the 3D space [16]. A vanishing point (VP) is the intersection point of all the projected lines that are mutually parallel in the scene (i.e. they share the same 3D direction). Note that, in a practical scenario, if more than two concurrent image lines are available, their intersection will not be unique (see Fig. 1b) — since noise can perturb the image line detection — and the VP has to be estimated with an optimization algorithm. In our experiments we employ the solution reported in [16], where after initializing the VP by solving a linear least square problem, a non-linear optimization is carried out.

Let $\mathbf{v}_1$ and $\mathbf{v}_2$ be two VPs related to 3D orthogonal directions. Then $\mathbf{v}_1^\top \omega \mathbf{v}_2 = 0$, where $\omega = (KK^T)^{-1}$ is the *image of the absolute conic*, depending on the three camera parameters $f$ and $(p_x, p_y)$. Given three vanishing points corresponding to three orthogonal directions, we can thus define three independent linear constraints on $\omega$, and finally estimate $\omega$ by solving a linear homogeneous system. Eventually $K$ can be obtained using the Cholesky factorization of $\omega$, from which both focal length and principal point can be estimated [16].

The estimation of the PP on a single image can be summarized in three main steps: (1) selection of three groups of concurrent image lines, corresponding to mutually orthogonal directions in the scene; (2) estimation of vanishing points; (3) computation of $\omega$ and recovery of $f$ and $(p_x, p_y)$.

Note that the first step can be done in a manual or automatic way. In the computer vision field, many works have appeared dealing with the problem of line selection and grouping for VP estimation by using Expectation-Maximization approaches [32], the Hough transform [33], or robust estimators,




(a) (b)

Figure 1: (Best viewed in color) (a) Pinhole camera model: Given the camera center $\mathbf{C}$, expressed in the world coordinate system $\{\mathbf{X}, \mathbf{Y}, \mathbf{Z}\}$, and the image plane $\pi$ orthogonal to the $Z$-axis, the principal point $\mathbf{PP}$ is the intersection of the $Z$-axis with $\pi$, while the focal length $f$ is the distance between $\mathbf{C}$ and $\pi$. A 3D point $\mathbf{M}$ is projected in $\mathbf{m}$ on the image plane as $\mathbf{m} = K[I|\mathbf{0}]\mathbf{M}$. (b) In red, green and blue three sets of image lines corresponding to orthogonal 3D directions. Since noise can perturb the line orientations the intersection can be not unique, as shown in the magnified area.

such as the J-Linkage algorithm [34], employed in [35]. If the camera calibration is known, mutually orthogonal line clusters can be selected automatically [36, 37, 38]. On the other hand, with no a priori information about camera calibration (which is our case), it can be extremely hard to check the vanishing point orthogonality without user intervention or by imposing simple heuristics, such as the selection of the most populated clusters. So, in this work we preferred to use a manual line selection scheme. Moreover, notice that also in [29] parallel lines are validated by the user, while in [19] no specific indication is given about the method used to automatically detect orthogonal vanishing points.

## 4. Perspective Analysis

In this Section, we evaluate the performance of the PP estimation algorithm under different perspective conditions, so as to determine if and how its accuracy changes when passing from *weak* to *strong* perspective images. The following two subsections report the results of synthetic and real world tests respectively.

### *4.1. Synthetic tests*

In order to carry out extensive tests, a synthetic dataset featuring 248 representative camera poses was built as follows. A 3D cube with unit length sides was placed in the center of the world coordinate frame with its $X$, $Y$, $Z$ axes aligned with the cube. Then, 248 camera center positions were sampled over a sphere of radius $r$, by varying their azimuth by an angle $\alpha \in (0, \pi/4]$ and their altitude by an angle $\beta \in (0, \pi/2)$ with steps of $\frac{\pi}{32}$ and $\frac{\pi}{64}$ respectively; all other perspective conditions can be deduced by symmetry. Since the VPs are invariant to translation, the camera distance with respect to the world coordinate frame (i.e. the radius $r$) was kept fixed. In the camera coordinate frame, the $z$-axis is the line passing through the camera center and the world coordinate origin. The $x$-axis is perpendicular to the $z$-axis and parallel to the world plane defined by $X$ and $Y$ and, finally, the $y$-axis is obtained from the cross product between the unit vectors of the $z$ and $x$ axes (see Fig. 2).

We excluded extrema positions — i.e. when $\alpha = 0$, $\beta = 0$, $\beta = \pi/2$ — that produce orthographic images of the cube, thus leading to known degeneracies in VP estimation. Likewise, camera roll was not taken into account considering that, as any pure rotation, no parallax effects are induced, thus leaving the perspective appearance of the image unaltered. From each camera pose $P(\alpha, \beta)$, an image of the cube was acquired by using a virtual camera with known PP and focal length. With noise-free measurements (i.e., line points are selected with no error), the PPs were estimated with an Euclidean error with respect to the ground truth lower than $10^{-9}$ pixels in all the positions. The behaviour in the presence of noise was then evaluated by carrying out

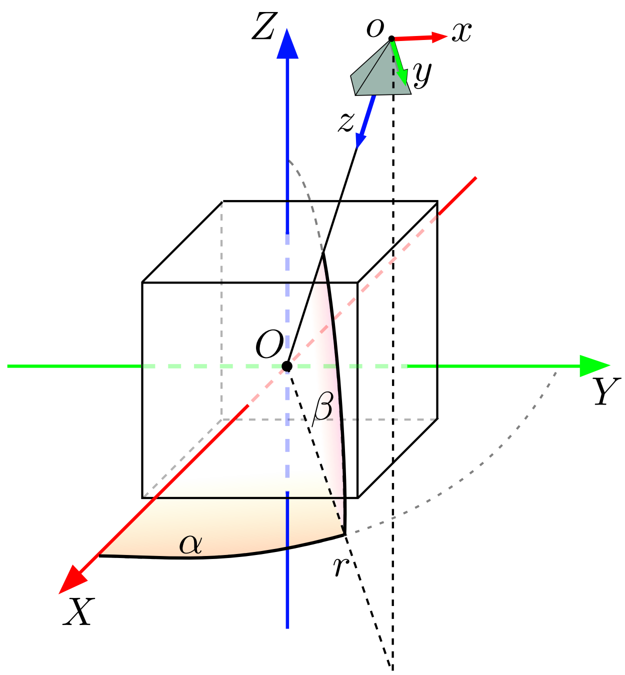
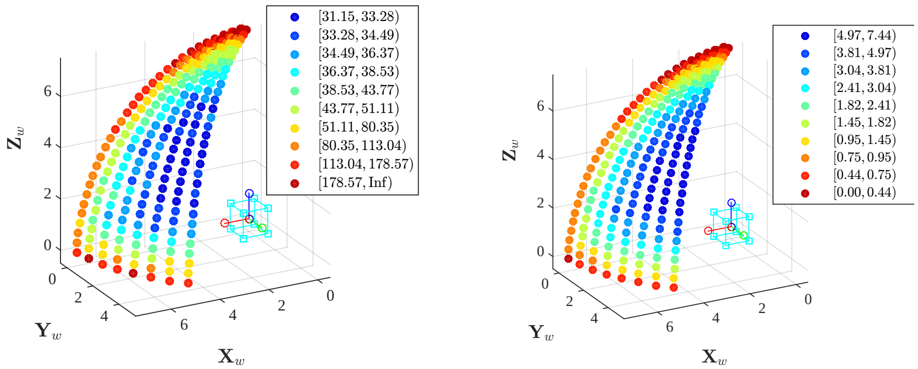



Figure 2: (Best viewed in color) Synthetic data setup. A cube is placed at the center of the world coordinate system $O$, with its sides aligned with the axis $X$,$Y$,$Z$. The image is taken from the camera — represented here as a pyramid — with center $o(\alpha, \beta)$ with a relative coordinate system $x, y, z$.

a Monte Carlo simulation: for each pose we collected 1000 principal points $PP(\alpha, \beta) = \{PP_1(\alpha, \beta), \ldots, PP_{1000}(\alpha, \beta)\}$ by perturbing the line points with a noise from a zero mean Gaussian distribution with standard deviation $\sigma = 0.5$ pixel — representing an uncertainty of at most 1.5 pixel radius in points selection. For each test we determined a robust index for the dispersion of the collected $PP(\alpha, \beta)$ as follows: we trimmed the 5% of the points with highest distance from the ground truth PP, then we calculated the standard deviations ($\mathrm{STD}_x$, $\mathrm{STD}_y$) of the remaining points along the $x$ and $y$ axes and we chose their maximum as a dispersion index of the estimated PP for that position.

Results are graphically reported in Fig. 3a, where the synthetic cube is placed in the origin of the coordinate frame aligned with the orthogonal axes, while each point represents a camera position, colored according to the correspondent




(a) (b)

Figure 3: (Best viewed in color) 3D plots representing results obtained with the synthetic data setup: in both figures, the virtual cube is placed in the origin of the coordinate system, aligned with the orthogonal axis. Colored points represent the tested camera positions. In (a) we report the maximum STD (between x and y-axis) of the estimated PP: the PP dispersion is bigger for reddish and, lower for blueish points. In (b) the same camera poses are reported but with color related to the MVA: poses with wider MVA are reported in blue, while poses with narrower MVA are in red. Note that poses with lower STD are characterized by wider MVA, and vice-versa. In both plots, the thresholds used to assign colors are obtained from the deciles (i.e. ten quantile with step of 10%) of the respective distribution (STD and MVA).

estimated dispersion. Notice that the scattering of the estimated PPs is strictly related to the image perspective: Most of the poses have comparable uncertainty, except when marginal $\alpha$ or $\beta$ occurs. In those cases, the computation accuracy of the VPs strongly drops, and the PP estimates become unreliable and virtually useless for forensic purposes.

These results suggest the possibility to define a novel image feature to be used by the forensic analyst to evaluate the expected accuracy. Firstly, given a vanishing point $\mathbf{v}_i$, let $\theta_i$ be the widest angle among those obtained from the pairwise intersection of lines concurrent to $\mathbf{v}_i$ (see Fig. 5). Then, given $\theta_1$, $\theta_2$, and $\theta_3$, related to three mutually orthogonal VPs, we can define the *Minimum Vanishing Angle* (MVA) as

$$\text{MVA} = \min(\theta_1, \theta_2, \theta_3) \tag{2}$$

A visual representation of the MVA values for different camera poses is reported in Fig. 3b. Its comparison with the results in Fig. 3a confirms our intuition


Figure 4: Twelve images, with associated names, from the York Urban Database [39], used in the real test to assess results obtained with the synthetic cube dataset. Top row shows images with strong perspective, with MVAs spanning from 7.52° to 5.53°. Second row includes mid perspective images with MVAs from 3.96° to 2.11°. Finally, the last row shows images with low perspective and MVAs from 1.09° to $\sim 0.00$ °. MVA here reported are the mean value of the MVAs computed on each image during the tests, since any user can select different lines and obtain slightly dissimilar MVA.

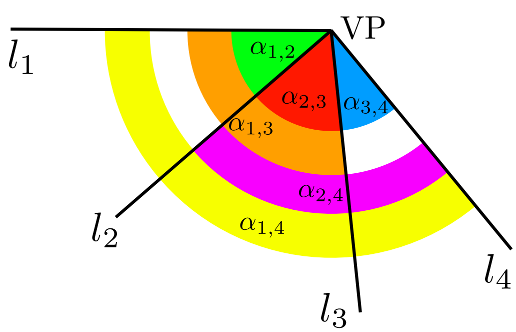



Figure 5: (Best viewed in color) Graphical visualization of angles obtained from the pairwise intersection of lines concurrent to the same VP. In this case $\theta_i$ correspond to $\alpha_{1,4}$ since it is widest angle available.

that the proposed feature is a sensible indicator of PP dispersion. Indeed, small MVAs are associated to marginal poses characterized by a weaker perspective. Notice that also in [29] the authors try to define a way to evaluate the quality of the estimates: They propose to use the distance between the VP and the PP, the latter supposed to be in the image center. However, this criterion may lead to erroneous evaluations in the presence of cropping, since the PP would not be close to the image center. Moreover, relying on a distance-based criterion instead than on an angle-based criterion such as ours, would inevitably introduce a dependency on image resolution.

### *4.2. Tests on real images*

To compare the synthetic data with real experiments we clustered the 248 synthetic poses in three groups according to their correspondent MVAs: Weak Perspective ($MVA < 1.5°$), Mid Perspective ($1.5° \leq MVA < 4°$), and Strong Perspective ($MVA \geq 4°$). Then we considered 12 images from the York Urban Database [39] spanning several MVAs between 0° and 7.52°. For each image 25 different PPs were computed, as described in Section 3, by letting 25 different users to select three lines for each direction. In Figure 4 we reported the name of the selected images, their MVAs estimated by users selection and the perspective group they belong to (Weak, Mid or Strong).

The achieved results are compared in Figure 6. Crosses represent the estimated PPs on real images: in red, green and blue for the images belonging to Weak, Mid and Strong perspective groups respectively. The plotted ellipses represent the 95% confidence ellipses estimated on the corresponding synthetic clusters. Synthetic results show that the estimation is expected to be extremely noisy on the Weak perspective cluster while more accuracy and stability is expected on the Mid and Strong cluster where the MVA is wide enough. Real data confirm the synthetic prediction ($\mathrm{STD}_x$ is 435.69, 38.52 and 29.69 pixels on Weak, Mid and Strong perspective clusters respectively). Looking at the picture, a horizontal dispersion of the real data sticks out. This is due to the fact that the images of the considered dataset are characterized by small altitudes, while the synthetic data is built considering all possible viewing angles.

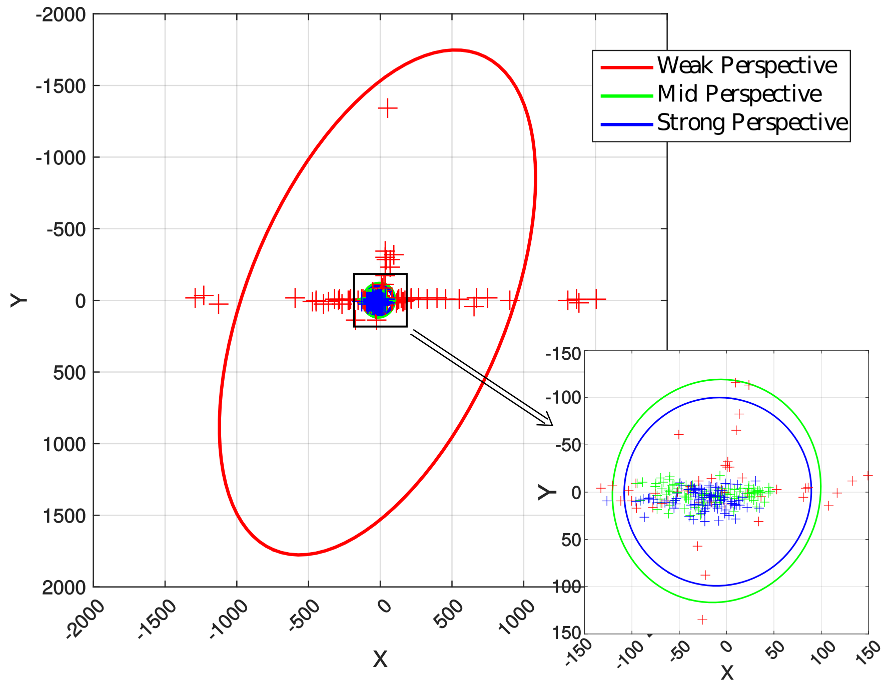



Figure 6: (Best viewed in color) Comparison of results achieved from synthetic and real images. Crosses represent the estimated PPs (red for *subway*, green for *hall*, blue for *building*). Ellipses enclose the PPs distribution obtained in synthetic tests.

## 5. Image Characteristic Analysis

In the previous section we defined the MVA feature, after observing a strong relationship between the amplitude of the vanishing angles and the PP estimation accuracy. In practical cases, the scene may allow the forensic analyst to extract more lines for each direction and possibly forming even wider MVAs. In this section we investigate more deeply the estimation accuracy with reference to the MVA amplitude. For this purpose, we take into account only MVAs with sufficient amplitude able to provide reliable results, and we evaluate how increasing it improves the estimation accuracy.

We also study how the performance is sensitive to an increase in the number of lines intersecting in the same VP: Since VPs are obtained by minimization, we expect an accuracy improvement when more data are available. As for the tests of Section 4, a synthetic image dataset is used first, then tests on real

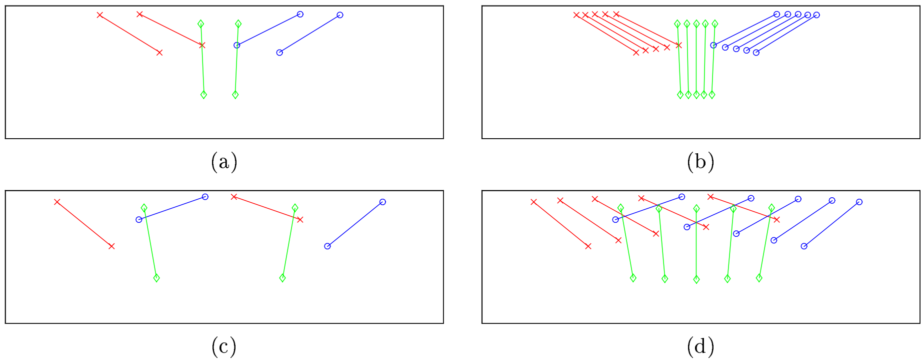



Figure 7: (Best viewed in color) Example images produced to test the PP estimation algorithm with reference to the extracted features. On the left, images with two lines for each VP, with different minimum vanishing angle (i.e. MVA={5,20}); on the right, similar images but with five lines. Lines with the same color converge to the same vanishing point.

images are carried out to corroborate the synthetic results.

### *5.1. Synthetic tests*

We generated different MVAs with different numbers of lines: starting with two lines for each VP, with an angle of incidence of $5°$, we progressively added new lines into the image and increased the angle. More specifically, we used $n = \{2, 3, 4, 5\}$ lines, with a length of 200px, and angles $\theta = \{5°, 10°, 15°, 20°\}$ (see Fig. 7 for some synthetic image examples). Gaussian noise with zero mean and standard deviation $\sigma = 0.5$ pixel was added to the point coordinates, and the evaluation was repeated 1000 times for each image.

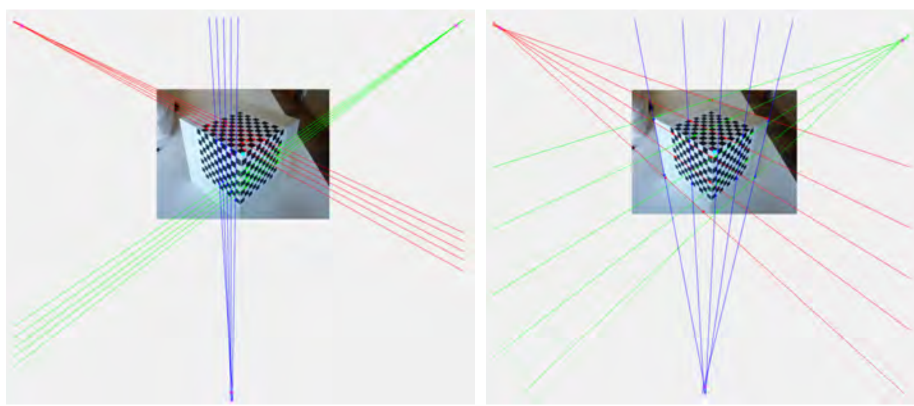
Table 1 shows the maximum STDs (as defined in Section 4.1) for the estimated PPs, along the $x$ and $y$ image directions. As clearly visible, the accuracy is almost stable when adding new lines, while it significantly grows using well spaced lines (i.e., wider MVAs).

### *5.2. Tests on real images*

As before, the results obtained with the synthetic data were validated on real tests with the help of 25 different users, having to select up to five lines per VP, with quasi regular spacing. For this purpose, the image of a cube-like

Table 1: Max STD of estimated PPs between $x$ and $y$ direction

| | | MVAs | | | |
|---|---|---|---|---|---|
| | | **5°** | **10°** | **15°** | **20°** |
| **#Lines** | **2** | 18.55 | 10.15 | 7.09 | 5.98 |
| | **3** | 19.54 | 9.85 | 6.59 | 5.56 |
| | **4** | 18.67 | 9.53 | 6.17 | 5.12 |
| | **5** | 17.03 | 8.74 | 6.12 | 4.79 |



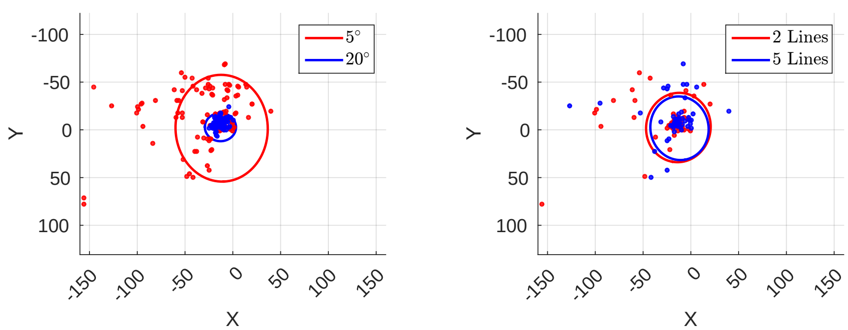
(a) (b)

Figure 8: Examples of lines selected by the user on the real image searching for (a) narrow and (b) wide MVAs.

checkerboard pattern was used. The considered image allows the user to select either narrow or wide MVAs of approximatively 5° and 20° respectively. 25 PPs were collected in both cases — i.e. the narrow (Fig. 8a) and wide (Fig. 8b) selection schemes — and the results were evaluated with respect to MVA amplitude and number of lines.

The PPs estimated on the real images are represented as colored dots in Fig. 9a — in red for angles of 5°, in blue for wider angles (20°). The 95% confidence ellipses of PPs obtained during the synthetic tests (see Section 5.1) are also shown, with the same color coding. In Fig. 9b, a similar plot considering instead the line number is presented. Almost all PPs obtained on the real images fall inside the associated ellipse, confirming that synthetic results are in close agreement with the real ones. Furthermore, these tests corroborate the observation that increasing the MVA clearly improves the estimation stability (Fig. 9a), while adding more lines does not significantly affect the performance (Fig. 9b).

In conclusion, results obtained in Sections 4 and 5 can be summarized in two main outcomes: (i) Images characterized by a narrow MVA should not be used for forensic analysis based on PP; (ii) To improve accuracy, the selection of few well spaced lines is preferable over many, closely spaced lines.




(a) (b)

Figure 9: (Best viewed in color) Results on real images obtained by varying MVA ans and line number. In (a) dots represent estimated PPs, clustered with respect to the MVA, while in (b) PPs are grouped by the line number. Reported ellipses represent the PP dispersion on the synthetic data. The coordinate system is centered in the ground truth PP.

## 6. Forensic Case Studies

In [19] the distance between the PP and the image center is exploited to identify asymmetrically cropped images (see Fig. 10). Once computed, the image and the PP are normalized in the interval $[-1, 1]$. Then a cropping threshold (CT) — i.e. the radius of a circle centered in the estimated PP — is defined, and the image is labeled as cropped if the distance of the PP from the image center exceeds CT. In the following tests we show how the achieved results can support the analyst in assessing the cropping detection performance:

- *Perspective-based Test*: we verify that the MVA amplitude can suggest whether the cropping detection is applicable on a query image. The test is

Figure 10: (Best viewed in color) In a pristine image (surrounded by a red border) the image center (red cross) falls near the PP (purple dot). On the other hand, if an upper-right cut (green area) is performed, the image center (green cross) shifts falling away from the PP, that remains fixed. The green area is related to the cropping percentage (CP). Blue and cyan circles, centered on the PP, represent instead two cropping thresholds (CT): note that in this example, using the smaller CT (blue circle) the cropping will be successfully detected, since the center of the cropped image center (green cross) fall outside the circle. On the other hand, using the bigger threshold (cyan circle), the image will be erroneously labeled as pristine. Note that in this figure we changed the aspect ratio of the original image (Fig. 4(P1030004)) so to visualize the normalization process in [-1,1].

performed on the synthetic and real data defined in Section 4 and confirms that the technique cannot be applied on images with a narrow MVA;

- *Characteristic-based Test*: we assess the performance variations when more lines and wider MVAs are available on the image. The test is performed on the synthetic and real data defined in Section 5;
- *Robustness Test*: we verify the robustness of the cropping detection to image compression and resizing. We consider a practical case where the image has been exchanged through Facebook at low quality, thus having been resized and compressed.

In our experiments we consider both cropping percentage (CP) — i.e. the size of the cut — and CT from 0% to 50% of the image size, with steps of 5%. Results are reported for an upper-left cropping only, where both dimensions of the image have been cut with the same percentage, thus leaving unchanged the image aspect ratio. However, tests were performed on all the other eleven cases of asymmetric cropping too (upper, left, right, bottom, upper-left, upper-right, bottom-left, bottom-right, left-upper-right, upper-right-bottom, right-bottom-left, bottom-left-upper). These results are summarized in the supplemental material where is shown that performances significantly increase between Weak and Mid perspective in all the cropping cases, confirming that the proposed feature allows the analyst to decide whether the cropping detection can possibly be applied to a query image.

When useful, the performance was evaluated using the Receiver Operating Characteristic (ROC) curve, where each point corresponds to True Positive (TP) and False Alarm (FA) rates for a given CT. The Area Under Curve (AUC) is used to compare the overall performance under different conditions: the more the AUC is close to one, the better is the detector accuracy. In some cases the mean accuracy was also reported (computed as the average of TP and TN rates on all considered cropping percentages). For the sake of presentation, results have been grouped into two clusters, corresponding to slightly cropped (lower than 25% of the image) or strongly cropped (between 25% and 50%) images.

### *6.1. Perspective-based Test*

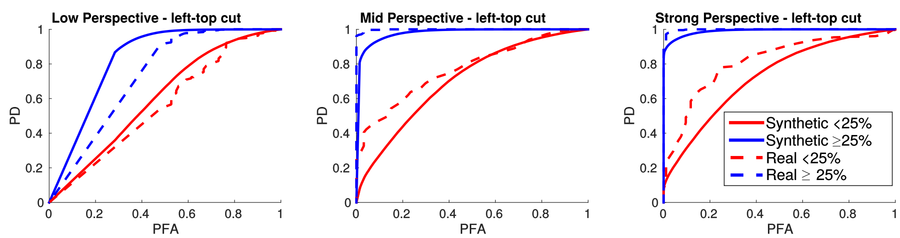
In this test we assess the performance of the cropping detection with reference to perspective conditions. We considered both synthetic and real PPs acquired in subsections 4.1 and 4.2 respectively. The cropping detection performance was evaluated separately on the three clusters (Weak, Mid and Strong Perspective) for both synthetic and real PPs. In Figure 11 we reported the ROC curves considering slightly and strongly cropped images, while in Table 2 we reported the AUC values. In Table 3 we summarize the cropping detection performance on the three clusters for different CTs, namely: FA rate, TP rate for both slight and strong cropping, and the mean accuracy. Note that we only report results considering the CTs in $[0.05, 0.25]$, since we noticed a progressive performance drop for higher CTs.

These results suggest that, given a threshold, the false alarm rate may strongly depend on the MVA. For instance, a false alarm of 0.03 on the Mid

Table 2: AUC for Perspective based Test on synthetic and real data

| Synthetic Data | | | | Real Data | | | |
|---|---|---|---|---|---|---|---|
| CP | Weak | Mid | Strong | CP | Weak | Mid | Strong |
| <25% | 0.60 | 0.70 | 0.72 | <25% | 0.56 | 0.77 | 0.81 |
| 25%-50% | 0.82 | 0.97 | 0.99 | 25%-50% | 0.73 | 1.00 | 1.00 |

perspective cluster (real data) corresponds to a threshold of 0.25 of the image. However, the same threshold on the Weak perspective cluster corresponds to a false alarm of 0.73. Both synthetic and real results confirm that the cropping detection can hardly be applied on Weak perspective images and a threshold on the MVA can be chosen to discern unusable images (AUC passes from 0.73 to 1 from Weak to Mid perspective on real images). Furthermore we notice that, on images characterized by decent perspective ($MVA > 1.5$), the technique is extremely effective when the applied cropping is greater than 25% of the image.




(a) (b) (c)

Figure 11: (Best viewed in color) ROC curves of the cropping detection for synthetic and real data. The results are reported for (a) Weak, (b) Mid and (c) Strong cluster separately.

### 6.2. Characteristic-based Test

In this test we assess the performance of the cropping detection with reference to the number of lines and their MVAs. We tested the cropping detection on the synthetic PPs acquired in Sections 5.1 (for angles of 5° or 20°, and with 2 or 5 lines) and on the real data acquired in Section 5.2. Firstly, we compared the results obtained when the VPs are estimated from 5° and 20° MVAs; the performances are shown through the ROC curves in Fig. 12a and 12b. Secondly, we compared the results achieved using 2 or 5 lines to detect each vanishing point; the corresponding ROC curves are reported in Fig. 12c and 12d. In Table 4 the

Table 3: Cropping detection on both synthetic and real data, considering Weak (a,b), Mid (c,d), and Strong perspective (e,f)

(a)

| Synthetic Weak Perspective | | | | |
|---|---|---|---|---|
| CT | FA | TP (<25%) | TP (25%-50%) | Mean Accuracy |
| 0.05 | 0.96 | 0.99 | 1.00 | 0.52 |
| 0.10 | 0.86 | 0.96 | 1.00 | 0.56 |
| 0.15 | 0.73 | 0.90 | 1.00 | 0.61 |
| 0.20 | 0.62 | 0.81 | 1.00 | 0.65 |
| 0.25 | 0.53 | 0.71 | 0.99 | 0.67 |

(b)

| Real Weak Perspective | | | | |
|---|---|---|---|---|
| CT | FA | TP (<25%) | TP (25%-50%) | Mean Accuracy |
| 0.05 | 0.97 | 0.99 | 1.00 | 0.52 |
| 0.10 | 0.90 | 0.96 | 1.00 | 0.56 |
| 0.15 | 0.80 | 0.90 | 1.00 | 0.61 |
| 0.20 | 0.75 | 0.81 | 1.00 | 0.65 |
| 0.25 | 0.73 | 0.71 | 0.99 | 0.67 |

(c)

| Synthetic Mid Perspective | | | | |
|---|---|---|---|---|
| CT | FA | TP (<25%) | TP (25%-50%) | Mean Accuracy |
| 0.05 | 0.92 | 0.98 | 1.00 | 0.54 |
| 0.10 | 0.72 | 0.93 | 1.00 | 0.62 |
| 0.15 | 0.53 | 0.81 | 1.00 | 0.70 |
| 0.20 | 0.37 | 0.67 | 1.00 | 0.75 |
| 0.25 | 0.25 | 0.51 | 0.99 | 0.77 |

(d)

| Real Mid Perspective | | | | |
|---|---|---|---|---|
| CT | FA | TP (<25%) | TP (25%-50%) | Mean Accuracy |
| 0.05 | 0.82 | 0.98 | 1.00 | 0.54 |
| 0.10 | 0.53 | 0.93 | 1.00 | 0.62 |
| 0.15 | 0.30 | 0.81 | 1.00 | 0.70 |
| 0.20 | 0.15 | 0.67 | 1.00 | 0.77 |
| 0.25 | 0.03 | 0.51 | 0.99 | 0.77 |

(e)

| Synthetic Strong Perspective | | | | |
|---|---|---|---|---|
| CT | FA | TP (<25%) | TP (25%-50%) | Mean Accuracy |
| 0.05 | 0.90 | 0.98 | 1.00 | 0.55 |
| 0.10 | 0.67 | 0.91 | 1.00 | 0.65 |
| 0.15 | 0.45 | 0.78 | 1.00 | 0.73 |
| 0.20 | 0.30 | 0.62 | 1.00 | 0.77 |
| 0.25 | 0.10 | 0.45 | 0.99 | 0.80 |

(f)

| Real Strong Perspective | | | | |
|---|---|---|---|---|
| CT | FA | TP (<25%) | TP (25%-50%) | Mean Accuracy |
| 0.05 | 0.83 | 0.99 | 1.00 | 0.58 |
| 0.10 | 0.45 | 0.88 | 1.00 | 0.75 |
| 0.15 | 0.22 | 0.70 | 1.00 | 0.83 |
| 0.20 | 0.12 | 0.50 | 1.00 | 0.84 |
| 0.25 | 0.07 | 0.31 | 0.97 | 0.82 |

AUCs for the two experiments have been reported to compare the overall performances. To be consistent with the previous test we briefly report in Table 5 the mean accuracy at varying CT for each of the cases. The achieved results show that wider MVAs produce a significant improvement in the detection rate. For instance, with a CT of 0.10, the mean accuracy passes from 0.79 to 0.97 on the synthetic data. This behaviour is confirmed by real data: with the same CT the mean accuracy passes from 0.67 to 0.99. As expected, performances are slightly affected by increasing line numbers. Indeed mean accuracy improvements are always at most 5% for all the synthetic and real cases.

In [19] the authors state that a CT of 0.1 and 0.15 can fit different demands. Anyway this threshold is set regardless of image content. The achieved results suggest instead that a more fitting threshold could be selected according to the

Table 4: AUC for Characteristic based Test on synthetic and real data

(a)

| Synthetic Data | | | | |
|---|---|---|---|---|
| CP | 2 lines | 5 lines | $\sim 5°$ MVA | $\sim 20°$ MVA |
| <25% | 0.87 | 0.96 | 0.87 | 0.99 |
| 25%-50% | 1.00 | 1.00 | 1.00 | 1.00 |

(b)

| Real Data | | | | |
|---|---|---|---|---|
| CP | 2 lines | 5 lines | $\sim 5°$ MVA | $\sim 20°$ MVA |
| <25% | 0.86 | 0.89 | 0.81 | 1.00 |
| 25%-50% | 0.99 | 1.00 | 0.99 | 1.00 |

Table 5: Mean Accuracy for Characteristic based Test on synthetic and real data

(a)

| Synthetic Data | | | | |
|---|---|---|---|---|
| CT | 2 lines | 5 lines | 5° MVA | 20° MVA |
| 0.05 | 0.77 | 0.80 | 0.63 | 0.91 |
| 0.10 | 0.90 | 0.91 | 0.79 | 0.97 |
| 0.15 | 0.91 | 0.91 | 0.86 | 0.93 |
| 0.20 | 0.89 | 0.89 | 0.88 | 0.89 |
| 0.25 | 0.86 | 0.86 | 0.86 | 0.86 |

(b)

| Real Data | | | | |
|---|---|---|---|---|
| CT | 2 lines | 5 lines | $\sim 5°$ MVA | $\sim 20°$ MVA |
| 0.05 | 0.59 | 0.60 | 0.58 | 0.63 |
| 0.10 | 0.82 | 0.85 | 0.67 | 0.99 |
| 0.15 | 0.82 | 0.83 | 0.72 | 0.94 |
| 0.20 | 0.83 | 0.86 | 0.79 | 0.92 |
| 0.25 | 0.82 | 0.87 | 0.84 | 0.87 |

available MVA. Synthetic results show that the best performances are obtained with a CT of 0.20 when a 5° MVA is available on the image. Conversely, with a 20° MVA, a CT of 0.10 should be preferred to achieve the best accuracy. Real data confirmed that two different thresholds should be considered according to MVA amplitude: 0.25 for a 5° MVA and 0.10 for a 20° MVA.

### *6.3. Robustness test*

In this test we assess whether the technique is usable when the image has been resized or compressed. We consider a practical case where the image (considered in the characteristic-based test) was uploaded on Facebook at low quality version and then downloaded: the resolution changes from $2592 \times 1944$ to $972 \times 729$, and its size from 1.4 MB to 80 KB. 25 PPs were collected on the downloaded image (similarly to Section 5.2) and the cropping detection was applied as in the characteristic-based test. In Tables 6 and 7 we report the AUC and the mean accuracy at varying CT: by comparison with the results achieved in the characteristic-based test, we notice that performances are almost unchanged, with the only exception of slightly cropped images, when only narrow MVAs are available, in which case performance drops slightly (AUC passes from 0.81 to 0.66.) This result once more confirms that the MVA amplitude is crucial

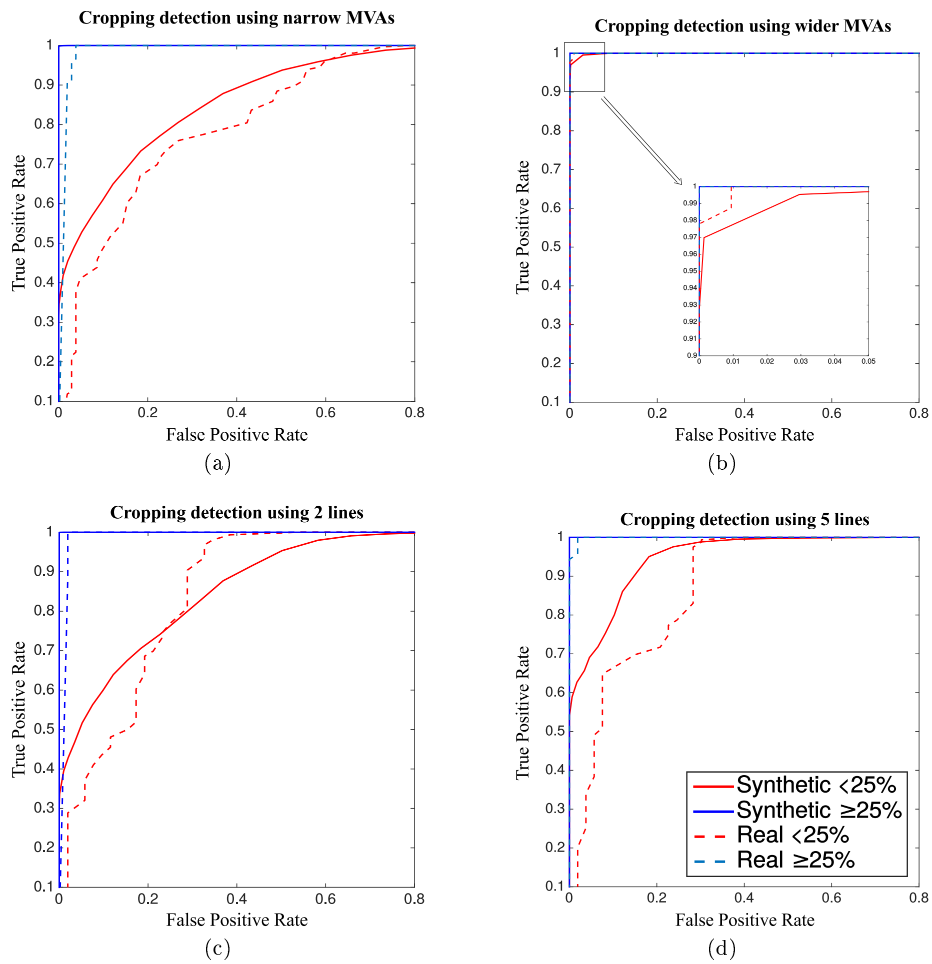



Figure 12: ROC curve on synthetic and real data with different cropping percentage using (a) narrow vanishing angles and (b) wider vanishing angles, and then using (c) 2 lines and (d) 5 lines to detect each vanishing point.

Table 6: AUC for on Facebook Data

| Facebook Data | | | | |
|---|---|---|---|---|
| CP | 2 lines | 5 lines | $\sim 5°$ MVA | $\sim 20°$ MVA |
| $<$25% | 0.82 | 0.82 | 0.66 | 1.00 |
| 25%-50% | 0.99 | 1.00 | 0.99 | 1.00 |

to determine the usability of this technique.

Table 7: Mean Accuracy on Facebook Data

| Facebook Data | | | | |
|---|---|---|---|---|
| CT | 2 lines | 5 lines | $\sim 5°$ MVA | $\sim 20°$ MVA |
| 0.05 | 0.61 | 0.59 | 0.51 | 0.71 |
| 0.10 | 0.76 | 0.80 | 0.55 | 1.00 |
| 0.15 | 0.82 | 0.81 | 0.69 | 0.94 |
| 0.20 | 0.84 | 0.81 | 0.73 | 0.92 |
| 0.25 | 0.82 | 0.82 | 0.77 | 0.87 |

### *6.4. A practical example of cropping detection*

We now show how MVA analysis can practically support the forensic analyst to assess whether an image has been cropped. Let us consider the images in Fig. 13a and 13c, downloaded from the web. The analyst estimates the PP on both images selecting lines that intersect with the widest possible angles. As a result he/she obtains that in both cases the normalized distance of the estimated PP from image center is anomalous (0.3875 and 0.2585 respectively). At first glance this fact leads to the conclusion that both images have been cropped. On the other hand, the analyst notices that the MVAs are 4.83 and 1.21 respectively. This means that he can be much more confident with the first result while the PP estimation on Fig. 13c is subjected to strong noise. More specifically, with such a small MVA the estimated PP is unreliable for the purpose. Then the analyst concludes that Fig. 13a is probably cropped while no evidence can be provided on Fig. 13c by this single test.

In figure 13b we report the original version of 13a that can be found on the web, confirming the achieved results.

### *6.5. An Example of Splicing Detection*

In this Section, we provide a simple example of another possible exploitation of the PP for forensics purposes: Splicing detection. In such forgeries, visual contents are inserted into the original image in order to create a plausible composite. Even with careful editing operations, an added object will likely show different perspective deformations with respect to the rest of the image. The PP could then be used to assess if distinct elements into the image have been

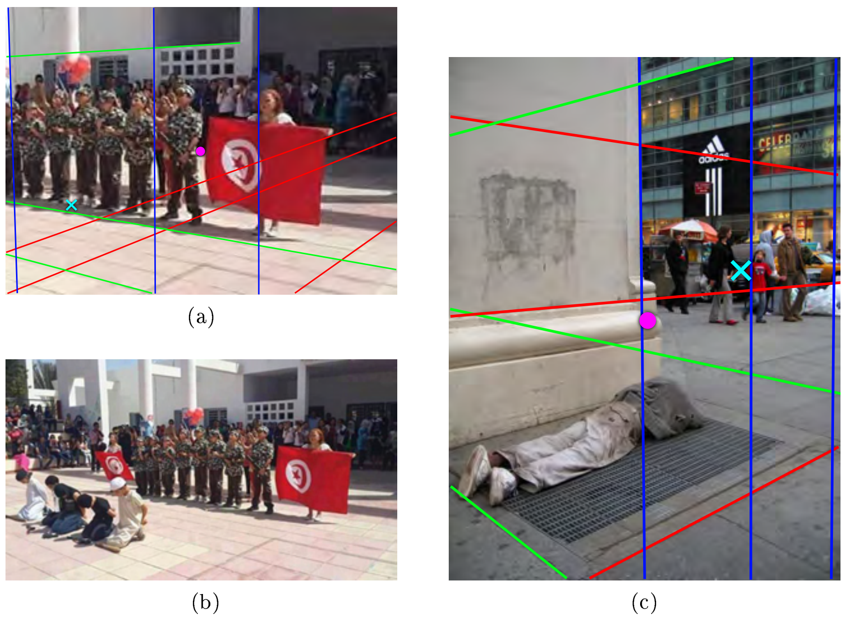



Figure 13: (Best viewed in color) Two examples of cropping detection (a,c), with lines of mutually orthogonal directions in red, green and blue. The purple dot indicates the image center, while the cyan cross shows the estimated position of the PP. In both images the MVA is the angle related to the vertical direction (blue lines): in (a) MVA=4.83, in (c) MVA=1.21. In (b) the original version of (a) is presented

subjected to a different projection, so to judge if the image is pristine or it is the result of a splicing manipulation.

In Fig. 14 a splicing example is reported. Using the image already presented in Fig. 4(P1030004), we manually inserted a blue police cabin and then we extracted lines from both the palace (red, green and blue lines) and the cabin (orange, light green and cyan lines). Then the PPs were estimated independently from the palace and the cabin (purple dots). As can be clearly seen, the computed PPs fall far from each other: This evidence leads to the conclusion that either the palace or the cabin have been maliciously added into the image.

A similar splicing detection approach has been presented in [29], where only a single vanishing direction is used as clue in order to validate the visual content. However, relying only on a single vanishing direction may lead to erroneous conclusions: Observing again Fig. 14, by using the left vanishing direction only

Figure 14: (Best viewed in color) Splicing example: a blue police cabin was added in the left corner of Fig. 4(P1030004). Lines of mutually orthogonal 3D direction have been extracted independently from the palace (red, green and blue lines) and from the cabin (orange, light green and cyan lines) — note that the vertical vanishing points are not reported due to lack of space. Then two PPs are estimated: since they are far from each other, we can assess that the image is manipulated.

(red and orange lines), no splicing evidence is found, since the palace and the cabin share the same vanishing point. On the other hand, by exploiting the PP, we can provide a more reliable evidence.

## 7. Conclusions and Future Work

In this paper we presented for the first an assessment of the reliability of physical-based features for forensic image authentication. In particular we focused on the estimation accuracy of the principal point of an image and its application to the forensic scenario. By observing the principal point estimation accuracy in different perspective conditions, we were able to define a novel feature, the minimum vanishing angle (MVA), strictly related to principal point uncertainty. Then we further investigated the MVA influence on the estimation accuracy by comparing it with respect to the number of detected lines, exploited for the estimation of the PP. Results underlined that the use of wider vanishing angles leads to higher accuracy, while by employing more lines only slight uncertainty reductions are achieved. As shown in the case studies presented in the previous Sections, the application of our criteria to cropping detection allows the analyst to easily exclude an image that is not suitable for the application of this technique. Moreover we verified that on resized and compressed images — as for example pictures downloaded in low quality from Facebook — the performance only slightly decreases, provided that wide MVAs are available. Eventually, we showed how the principal point can be also used for splicing detection.

In future work the proposed MVA will be exploited to analytically compute a likelihood score to provide more than a binary decision on the authenticity of the examined image. Moreover, we are planning to deeply investigate the relation between the MVA and the best cropping threshold to be used, in order to control the false alarm rate. For this purpose, automatic techniques for principal point localization — so as to remove the human-in-the-loop — will be investigated in order to perform tests on a huge amount of real data.

## Acknowledgments

The first author is partially supported by GNSAGA of INdAM. This material is based on research partially sponsored by the Air Force Research Laboratory and the Defense Advanced Research Projects Agency under agreement number FA8750-16-2-0188. The U.S. Government is authorized to reproduce and distribute reprints for Governmental purposes notwithstanding any copyright notation thereon.

## References

[1] H. Farid, A survey of image forgery detection, IEEE Signal Processing Magazine 26(2) (2009) 16–25.

[2] M. Stamm, M. Wu, K. Liu, Information forensics: An overview of the first decade, IEEE Access 1 (2013) 167–200.

[3] G. K. Birajdar, V. H. Mankar, Digital image forgery detection using passive techniques: A survey, Digital Investigation 10 (3) (2013) 226 – 245.

[4] A. Piva, An overview on image forensics, ISRN Signal Processing 2013 (2013) Article ID 496701, 22 pages.

[5] P. Ferrara, T. Bianchi, A. De Rosa, A. Piva, Image forgery localization via fine-grained analysis of cfa artifacts, Information Forensics and Security, IEEE Transactions on 7 (5) (2012) 1566–1577. doi:10.1109/TIFS.2012.2202227.

[6] M. Chen, J. Fridrich, M. Goljan, J. Lukas, Determining image origin and integrity using sensor noise, Information Forensics and Security, IEEE Transactions on 3 (1) (2008) 74–90. doi:10.1109/TIFS.2007.916285.

[7] B. Li, T. Ng, X. Li, S. Tan, J. Huang, Revealing the trace of high-quality JPEG compression through quantization noise analysis, Information Forensics and Security, IEEE Transactions on 10 (3) (2015) 558–573. doi:10.1109/TIFS.2015.2389148.

[8] T. Bianchi, A. Piva, Image forgery localization via block-grained analysis of jpeg artifacts, Information Forensics and Security, IEEE Transactions on 7 (3) (2012) 1003–1017. doi:10.1109/TIFS.2012.2187516.

[9] E. Kee, J. F. O'Brien, H. Farid, Exposing photo manipulation with inconsistent shadows, ACM Trans. Graph. 32 (3) (2013) 28:1–28:12. doi:10.1145/2487228.2487236.

[10] T. Carvalho, C. Riess, E. Angelopoulou, H. Pedrini, A. de Rezende Rocha, Exposing digital image forgeries by illumination color classification, IEEE Transactions on Information Forensics and Security (2013) 1182–1194.

[11] M. Johnson, H. Farid, Exposing digital forgeries in complex lighting environments, Information Forensics and Security, IEEE Transactions on 2 (3) (2007) 450 –461. doi:10.1109/TIFS.2007.903848.

[12] H. Yao, S. Wang, Y. Zhao, X. Zhang, Detecting image forgery using perspective constraints, Signal Processing Letters, IEEE 19 (3) (2012) 123–126. doi:10.1109/LSP.2011.2182191.

[13] M. Iuliani, G. Fabbri, A. Piva, Image splicing detection based on general perspective constraints, in: Proceedings of the Information Forensics and Security (WIFS), 2015 IEEE International Workshop, 2015.

[14] M. Zampoglou, S. Papadopoulos, Y. Kompatsiaris, Detecting image splicing in the wild (web), in: Proc. IEEE Int Multimedia & Expo Workshops (ICMEW) Conf, 2015, pp. 1–6.

[15] T. Carvalho, H. Farid, E. Kee, Exposing photo manipulation from user-guided 3d lighting analysis, in: Proc. SPIE, Vol. 9409, 2015, pp. 940902–940902–10.

[16] R. I. Hartley, A. Zisserman, Multiple View Geometry in Computer Vision, 2nd Edition, Cambridge University Press, 2004.

[17] M. K. Johnson, H. Farid, Detecting photographic composites of people., in: Y. Q. Shi, H.-J. Kim, S. K. 0001 (Eds.), IWDW, Vol. 5041 of Lecture Notes in Computer Science, Springer, 2007, pp. 19–33.

[18] J. Hu, Y. Li, S. Niu, X. Meng, Exposing digital image forgeries by detecting inconsistencies in principal point, in: Computer Science and Service System (CSSS), 2011 International Conference on, 2011, pp. 404–407.

[19] X. Meng, S. Niu, R. Yan, Y. Li, Detecting photographic cropping based on vanishing points, Chinese Journal of Electronics 22 (2013) Article ID 496701, 22 pages.

[20] G. Medioni, S. B. Kang, Emerging Topics in Computer Vision, Prentice Hall PTR, Upper Saddle River, NJ, USA, 2004.

[21] B. Caprile, V. Torre, Using vanishing points for camera calibration, Int. J. Comput. Vision 4 (2) (1990) 127–140.

[22] Z. Zhang, A flexible new technique for camera calibration, IEEE Trans. Pattern Anal. Mach. Intell. 22 (11) (2000) 1330–1334.

[23] R. Toldo, R. Gherardi, M. Farenzena, A. Fusiello, Hierarchical structure-and-motion recovery from uncalibrated images, Comput. Vis. Image Underst. 140 (C) (2015) 127–143.

[24] J. M. Coughlan, A. L. Yuille, Manhattan world: compass direction from a single image by bayesian inference, in: Computer Vision, 1999. The Proceedings of the Seventh IEEE International Conference on, Vol. 2, 1999, pp. 941–947 vol.2.

[25] C. Colombo, D. Comanducci, A. Del Bimbo, Camera Calibration with Two Arbitrary Coaxial Circles, in: Computer Vision – ECCV 2006: 9th European Conference on Computer Vision, Graz, Austria, May 7-13, 2006. Proceedings, Part I, Springer Berlin Heidelberg, Berlin, Heidelberg, 2006, pp. 265–276.

[26] E. Guillou, D. Meneveaux, E. Maisel, K. Bouatouch, Using vanishing points for camera calibration and coarse 3d reconstruction from a single image, The Visual Computer 16 (7) (2000) 396–410.

[27] J. Deutscher, M. Isard, J. Maccormick, Automatic camera calibration from a single manhattan image, in: Eur. Conf. on Computer Vision (ECCV, 2002, pp. 175–205.

[28] R. Pflugfelder, H. Bischof, Online auto-calibration in man-made worlds, in: Digital Image Computing: Techniques and Applications (DICTA'05), 2005, pp. 75–75.

[29] Y. Li, Y. Zhou, K. Yuan, Y. Guo, X. Niu, Exposing photo manipulation with inconsistent perspective geometry, The Journal of China Universities of Posts and Telecommunications 21 (4) (2014) 83 – 104.

[30] D. A. Forsyth, J. Ponce, Computer Vision: A Modern Approach, Prentice Hall Professional Technical Reference, 2002.

[31] R. Szeliski, Computer Vision: Algorithms and Applications, 1st Edition, Springer-Verlag New York, Inc., New York, NY, USA, 2010.

[32] J. Košecka, W. Zhang, Efficient computation of vanishing points, in: Robotics and Automation, 2002. Proceedings. ICRA '02. IEEE International Conference on, Vol. 1, 2002, pp. 223–228 vol.1.

[33] T. Tuytelaars, L. V. Gool, M. Proesmans, T. Moons, The cascaded hough transform as an aid in aerial image interpretation, in: Computer Vision, 1998. Sixth International Conference on, 1998, pp. 67–72.

[34] R. Toldo, A. Fusiello, Robust multiple structures estimation with j-linkage, in: Proceedings of the 10th European Conference on Computer Vision: Part I, ECCV '08, Springer-Verlag, Berlin, Heidelberg, 2008, pp. 537–547.

[35] J. P. Tardif, Non-iterative approach for fast and accurate vanishing point detection, in: 2009 IEEE 12th International Conference on Computer Vision, 2009, pp. 1250–1257.

[36] J. C. Bazin, Y. Seo, C. Demonceaux, P. Vasseur, K. Ikeuchi, I. Kweon, M. Pollefeys, Globally optimal line clustering and vanishing point estimation in manhattan world, in: Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on, 2012, pp. 638–645.

[37] C. Rother, A new approach for vanishing point detection in architectural environments, in: In Proc. 11th British Machine Vision Conference, 2000, pp. 382–391.

[38] J. C. Bazin, M. Pollefeys, 3-line ransac for orthogonal vanishing point detection, in: 2012 IEEE/RSJ International Conference on Intelligent Robots and Systems, 2012, pp. 4282–4287.

[39] P. Denis, J. H. Elder, F. J. Estrada, Efficient edge-based methods for estimating manhattan frames in urban imagery, in: D. Forsyth, P. Torr, A. Zisserman (Eds.), Computer Vision – ECCV 2008: 10th European Conference on Computer Vision, Marseille, France, October 12-18, 2008, Proceedings, Part II, Springer Berlin Heidelberg, 2008, pp. 197–210.
URL `http://www.elderlab.yorku.ca/YorkUrbanDB/`