Crime Scene Sketches Classification Based on CNN

Kaixuan Wang, Houlu Zhang, and Yunqi Tang(✉)

School of Criminal Investigation and Forensic Science, People's Public Security University of China, Beijing, China
wasd4119@163.com, zhanghoulu1996@gmail.com, tangyunqi@ppsuc.edu.cn

Abstract. Crime scene sketch plays a significant role in criminal investigation. In China, the crime scene sketches of all criminal cases should be uploaded to the National Criminal Scene Investigation Information System (NCSIIS). However, there are wrong images and low quality sketches frequently being uploaded to NCSIIS, which would make crime scene sketches unable to undertake their tasks. Yet, checking the sketches uploaded to NCSIIS still reamins as a manual work by the police officers. In this paper, we focus on a new problem of crime scene sketches classification. Firstly, a crime scene sketches database was constructed, sampled from NCSIIS. Secondly, an automatic crime scene sketches classification method is proposed based on CNN. A new architecture, namely Crime Scene Sketch Net (CSS-Net) is designed for high accuracy. Experiments are conducted on the database constructed. The experimental results show that the method proposed by this paper is of good performance.

Keywords: Criminal investigation · Crime scene sketch · Image classification · Convolutional Neural Network

1 Introduction

In criminal investigation domain, crime scene sketch establishes a permanent record of items, conditions, and position relationships [14]. It is an effective way to document a crime scene. Sketches can provide an in-depth understanding of the circumstances of crime scene beyond the level of comprehension that can be attained solely by reading a written report or studying photographs [6]. Compared with notes, sketches are more vivid than words. They also have some unique advantages over photos. They can show some details better, such as track of criminal walking. Sketches can eliminate some unnecessary details and pay more attention to the important items which are more relative to the crime [15].

This paper is a student paper

This work is supported by the National Key Research and Development Program (Grant No. 2017YFC0803506), the Ministry of Public Security Technical Research Project (Grant No. 2018JSYJC20), the Opening Project of Shanghai Key Laboratory of Crime Scene Evidence (Grant No. 2017XCWZK18).

Z. Lin et al. (Eds.): PRCV 2019, LNCS 11859, pp. 149–160, 2019.
https://doi.org/10.1007/978-3-030-31726-3_13

And in court, sketches would complement other documenting methods, such as photo, video and report. In addition to these advantages, sketches also play an important role in restoring and reconstructing crime scene.

Due to the importance of crime scene sketch, the Ministry of Public Security of People's Republic of China has established an information system, named National Crime Scene Investigation Information System (NCSIIS), to store and manage these records. And according to regulations, criminal investigators are supposed to draw and upload two types of sketches, one shows the location of crime scene, the other shows details in crime scene. And the result of spot check on the qualification of uploaded sketches shows a pretty pessimistic phenomenon. Lacking of sketches occurred in many cases, would lead to many serious consequences, such as being unable to reconstruct the crime scene and some legal problems in court. In some cases, investigator did upload two images, nonetheless the two images belong to the same class. And plus that there is no such a function which can recognize different types of sketches in the existing information system. Therefore, in a bid to check these uploaded records, there is no other way except manual verification. However, this method is time consuming and labour-intensive. Besides, manual verification cannot verify all records as numerous of sketches flock into the system. In conclusion, solving the problem how to perform automated classification of crime scene sketches is urgently required.

In 2012, the AlexNet which was proposed on ImageNet Large Scale Visual Recognition Challenge (ILSVRC), showed the dawning of Convolutional Neural Network (CNN) [4]. In recent years, it has become a mainstream on ILSVRCs [10]. And the champions that teams with algorithm based on CNN had won, proved that Convolutional Neural Network have the capacity to solve the problem of image classification, especially on the large-scale data sets. This led to a booming of convolutional neural network application study [1,8]. In recent years, Big Data and Artificial Intelligence started to be employed to Public Security, and they made a great success. The application of Biometric Technology has also provided great help cracking criminal cases. However, to the best of author's knowledge, this is first hand-shaking between Convolutional Neural Network and criminal investigation records management. In addition, peculiarity of crime scene sketch makes it very difficult to build a large-scale data set, let alone a publicly available large-scale image datasets. These all make this classification problem challenging.

In this paper, our works are summarized as follows:

- We proposed a new application problem of Convolutional Neural Network. To solve it, we built a crime scene sketches data set with a training and validation set of 53,324 images, and a test set of 10,897 images. These data are collected from the real-life records stored in NCSIIS with manual labeled precisely.
- To solve our problem, we designed our convolutional neural network based on AlexNet. Then we compared its performance with two classic architectures.
- Finally, we measured the performance of our CSS-Net, to indicates the capability of recognizing each categories.

2 Related Works

Because crime scene sketch is confidential, the related work is less than nothing. However the heart of automated classifying crime scene sketch problem lies a classic challenge, image classification. It has been a hot issue in pattern recognition and computer vision for a long time.

Back to 1990s, Yann Lecun et al. firstly used CNN on handwriting digit recognition and achieved a great success on MNIST dataset. Then, due to defects of CNN, the flourish of traditional pattern recognition methods such as Gaussian mixture model, K-means and support vector machine came to image classification domain. They did a good job on solving some simple problems, but in most cases, they were not competent [9]. Then, with the exploding growth of computing power, some disadvantages of deep neural network were compensated, while their benefits were magnified. CNN has a simple network topology, they can spend less training time by sharing weights, and one architecture of CNN could solve more than one classification problems, which made CNN more popular in image classification. Besides, in resent years, many excellent ideas of deep neural network has emerged, like VGG [11], GoogLeNet [13], ResNet [2] and deepID [12]. Hence, a trend of CNN application started.

In general, convolutional neural network has the ability to solve our problem. On account of small number of categories, there is no need to use a deep network structure. Some classic CNNs could competent this work.

3 Crime Scene Sketch

In this section, we focus on two things, what a qualified sketch should be and how to distinguish the different types of sketches. To figure them out, we viewed plenty of official documents and books related to criminal investigation. In China, there is no such a national or professional standard, but rules of drawing and classifying sketches exist in working specifications and textbooks about criminal investigation.

3.1 Sketch Taxonomy

Graph Performance Range. According to performance range of images, crime scene sketches can be divided into three categories. Location Sketch shows the location and surrounding of crime scene, which covers the biggest range. Crime Scene Overview Sketch describes the overall crime scene and it shows the result of criminal investigation including various items, evidence, traces, etc. The third one is Key Parts Sketch, highlight matters and crucial site, which is strongly related to crime. It is the one that covers the smallest range.

Representation of Image. This classification is similar to the one in United States which divide sketches into floor plan (or bird's view) sketch, elevation sketch and the cross-projection (or exploded view) [7]. But in China, there are two more categories in this sorting mode, Stereo View Sketch and Cutaway View Sketch.

- Floor plan is a horizontal top view drawn on the principle of parallel projection. It is for important evidence and objects distributed on the horizontal surface of the scene.
- Elevation sketch shows the vertical projection of the crime scene.
- Cross-projection is a combination of the first two types of scene maps. It shows other façades or tops of the crime scene, on the basis of floor plan sketch.
- Stereo View Sketch can represent the object's shape in three directions (eg. front, top and side) on one projection map, using the methods of angular parallel or center projection.
- Cutaway View Sketch is a special form of Stereo View Sketch. It removes part of object's surface and reflects the internal state of the object

Plotting Scale. In this sorting mode, sketches are divided into two categories, one with a scale, the other without scale. Sketch with scale should be made to scale, but sketch without scale could not. Therefore, only in some serious cases, investigators draw sketch with scale.

In this study, to meet the needs of criminal investigation, sketches are divided into Crime Scene Overview Sketch and Location Sketch, as these two types can cover the compulsory information of crime scene.

3.2 Rules of Uploading Sketch

What should a qualified crime scene sketch look like? In China, a qualified sketch should meet requirements both in format and content. The Ministry of Public Security revised Public Security Crime Scene Investigation Regulations and Sample of Crime Scene Investigation Records dated in October 2015, so as to meet the requirement of revised "Criminal Procedure Law" for crime scene investigation and further standardize the documenting work. In these two documents, rules are that drawing sketches should meet the following requirements:

- Mark the identifier, discovering time and location of the case.
- The location sketch should exactly reflect the location and scope of the scene.
- Crime Scene Overview Sketch should precisely reflect the main objects related to criminal activities, indicating the specific location of the body, traces, physical evidence, and tools for committing crimes.
- Text description should be concise and accurate
- Proper layout with highlighted priority.
- Clear presentation with standardized signs.
- Sketch should indicate the direction, legend (key), drawing date, cartographer and his (or her) organization.

Furthermore, on the conference of National Criminal Investigation Work on December 24, 2014, the Ministry of Public Security presented a series of regulations of criminal investigation, in order to normalize routine for investigating crime scene. In the regulations, a complete crime scene investigation record should cover two types of sketch describing site layout and location in detail. One is Crime Scene Overview Sketch, and the other is Location Sketch. These sketches uploaded into the NCSIIS system should be finished images and meet the requirements mentioned above.

4 Methodology

Convolutional Neural Network usually works in a common way, containing forward propagation and backward propagation. In the forward propagation phase, when images inputs, CNN begins sampling, down sampling, and finally outputs a loss value in training or scores of each categories in test. The loss value shows the distance between predicted results and label. The scores shows the probability that input image belongs to certain categories. Then it turns to backward propagation. It only exists in training process and works for finding the minimum loss value guided by the gradient [9].

The architecture of Convolutional Neural Network usually contains convolutional layer, pooling layer and fully connected layer. Convolutional layer is responsible for sampling, and pooling layer takes charge of down sampling. These two layers undertake the main tasks of CNN. How these layers work would be illustrated in following parts.

In a convolutional layer, there are more than one kernels available. When an image inputs to this layer, kernels slip on this matrix of image pixels in a fixed stride. In every step, a convolutional calculation exists. And it can be expressed as:

$$f(x) = \sigma(x \times W + b) \tag{1}$$

In Eq. (1), assuming $f(x)$ as output, x represents the input, then x times weight matrix W and plus a bias term b, finally put it to a nonlinear activation function $\sigma(x)$. When kernels traverses the whole image, the convolutional layer outputs numbers of feature maps.

Pooling layer usually plays a role in down sampling feature maps output by convolutional layer. Through this layer, the size of feature maps would shrink, but the number of them would stay. In this study, we used max pooling method. It works by sliding windows walking on the feature maps in a fixed stride. On the area covered by kernel, it works as:

$$f(x) = max(x) \tag{2}$$

In Eq. (2), x means the area covered by pooling kernel and this function output the biggest number in this area. It is a traditional approach where adjacent pooling kernels do not overlap during it sliding on feature map. In this study, however, we used overlapping pooling method. It makes the stride less than the size of kernel. When pooling kernel slides on feature maps, adjacent ones would overlap each other. And it has positive effects on reducing error rates and overfitting.

In the end of convolutional neural network, there would be some fully connected layers. They works for turning the matrix of feature maps into a feature vector and finally get a probability distribution P based on the input. As the Eq. (3) shows, the heart of CNN is to perform the operation of multi-layer filtering, reducing the amount of calculation, so as to obtain a mathematical model of feature expression P.

$$P(j) = P[L = l_j | x_j; (\omega, b)] \tag{3}$$

In Eq. (3), x_j represents the input image, l_j means label of it, and function $L(\omega, b)$ shows distance between label and the predicted result of forward propagation. Ideally, the best trained model reaches the point where value of $L(\omega, b)$ is minimized.

5 Experiment

In this study, our goal is to apply CNN to solve our problem, and this problem is not challenging enough to use some complicated architectures. Therefore, we employed classic architectures, LeNet-5 and AlexNet. To fit our data better, we designed a new architecture of CNN based on AlexNet and called it Crime Scene Sketch Net (CSS-Net). The whole experiment were conducted on Caffe [3], a convolutional architecture.

5.1 Data Set

Due to sensitive information involved in sketch, there is no public dataset to use in this study. Our first task was to build a dataset. We collected 71,839 sketches of 32,409 cases occurred in six different provinces from the database of NCSIIS. These sketches were drawn to documented the investigation of real-life cases and uploaded to the information system by criminal investigators. It might seem an easy job by simply downloading the data to get a dataset, but actually it was time-consuming and labor-intensive.

Manual Inspection. On account of lacking supervision, the images below the mark were mixed with the qualified sketches. We needed to scrutinize the data downloaded from the database. Then, 18,515 unqualified images were removed from the data set.

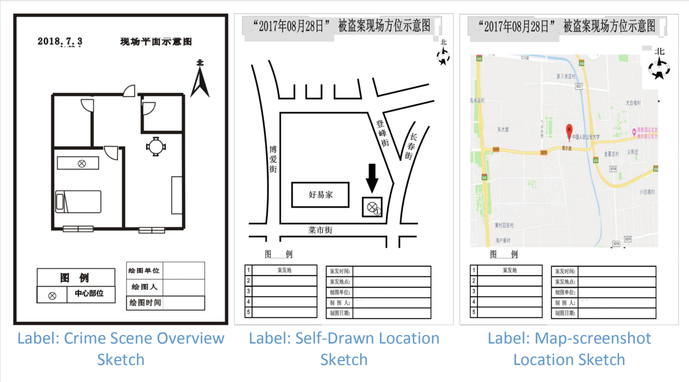


Fig. 1. Labeled crime scene sketches

Manual Label. In this study, these data should be labeled into two classes, crime scene overview sketch and location sketch, based on the range of grapy performance. Finishing labeling, we found that a huge intra-class variation exists in the location sketches. As known to us all, huge intra-class variation can damage the performance of the classifier. To get a better result, we subdivided the location sketches into two classes by drawing method. One is called Map-screenshot Location Sketch. It considers the screenshot of electronic map as main body of the image to show where the crime occurred. The other is called Self-drawn Location Sketch. It is manually drawn using graphics software. This type of sketch uses symbols to represent the buildings, roads, rivers and so on, to indicate the location and scope of crime scene. Finally we labeled these sketches into three categories, as shown in Fig. 1 (Table 1).

Table 1. The composition of dataset

	Train and validation set	Test set
Crime scene overview sketch	16425	4975
Self-drawn location sketch	8876	3647
Map-screenshot location sketch	28023	2275
Total	53324	10897

Train and Validation Set. Then, we got a dataset containing 16,425 Crime Scene Overview Sketch, 8,876 self-drawn location sketches, and 28,023 map-screenshot location sketches. In this dataset, we randomly sampled 80% of each type of sketches as the training set, and the remaining 20% as the validation set.

Test Set. In order to test the robustness of the trained model, we collected 10,897 new sketches from some provinces different with the six provinces. After examined and labeled, these new data were used as test set.

This dataset contains variable-resolution images, while the architecture of CNN needs a constant input dimension. Consequently, we normalized the sketches to a fixed resolution of 256*256. Finally, we got our crime scene sketch dataset.

5.2 Architecture

In this study, we employed two classic architectures, LeNet-5 and AlexNet. The details of them could turn to the Ref [5] and Ref [4]. Although we got a trained model that can apply to solve our problem, it was not good enough. Therefore, we revised the AlexNet and designed our new architecture. In this section, we will mainly focus on our CSS-Net.

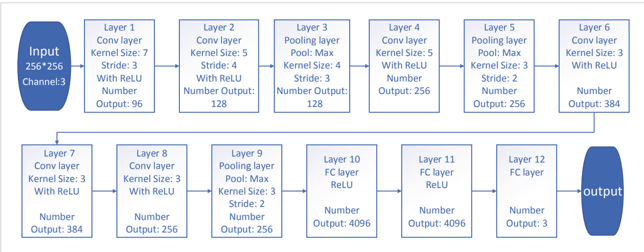


Fig. 2. Architecture of CSS-Net: Conv layer means convolutional layer, FC layer represents fully connected layer

Our CNN has six convolutional layers and three fully connected layers. Different from AlexNet, we added a convolutional layer and set size of input as 256*256 pixels. And for fear of training difficulty growing sharply, we used small convolution kernels to replace a big kernel. The tricks in AlexNet which benefit the training process were retained. However, the data augmentation fell short of lifting accuracy of classification and greatly increased cost of computing resource. Thence, image cropping was fired. Finally, we got a better classifier than classic CNNs. More details are shown as Fig. 2.

6 Results and Analysis

6.1 Compared with Classic CNNs

In this part, we trained our CSS-Net and classic nets of same parameters except the ones mentioned. Details are shown in Table 2.

Table 2. Parameters of training

	CSS-Net	AlexNet		LeNet-5
Crop size	None	227	None	None
Mirror	False	True	False	False
Batch size	128			
lr policy	inv			
base lr	0.0001			

In this table, there are some abbreviations which need to be explained. ‘lr policy’ means learning rate policy which represents the policy of changing learning rate during training phase. ‘base lr’ means base learning rate which is the initial value of it. If ‘Mirror’ was true, data augmentation method of mirror flipping would be used in this net. And during training phase, test on the validation set existed every 1000 iterations. Finally their accuracy made up this following Fig. 3.

As the results show that the rate of convergence on our CSS-Net is faster than classic nets, while LeNet-5 oscillating all the time. Besides, at the end, CSS-Net get a accuracy of 2% better than AlexNet with crop method. Therefore we can figure that CSS-Net performs better on our training set as our net get a higher accuracy and is easy to train.

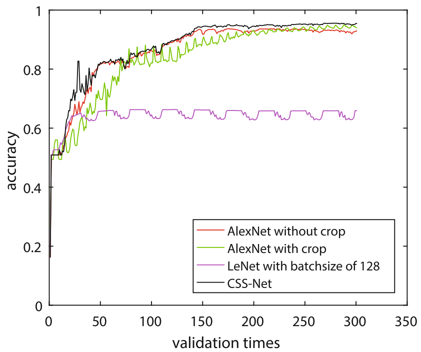


Fig. 3. Accuracy on validation set

To ensure that our trained model was not overfitting, accuracy of test on test set plays a significant role. To figure it out, we tested the models of 250,000 iterations and 300,000 iterations trained on AlexNet and our CSS-Net. The results are shown in following Table 3.

As shown, LeNet-5 lacks the ability to classify crime scene sketches, while AlexNet did a good job on our dataset, and this is the reason for choosing AlexNet to modify. It also shows that the image augmentation methods, image

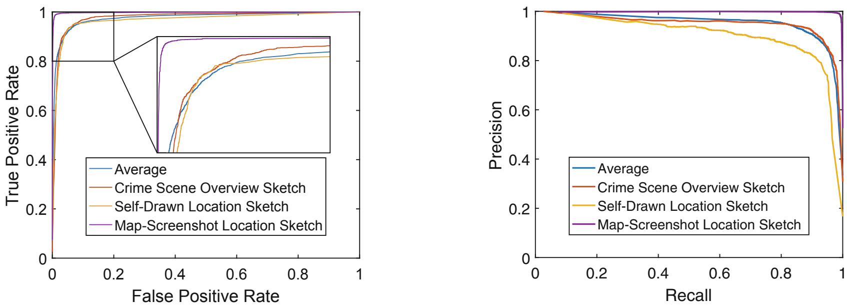
Table 3. Accuracy on test set

	AlexNet with crop	AlexNet without crop	CSS-Net
250,000 iters	86.5229%	84.5688%	89.8073%
300,000 iters	87.8349%	84.1009%	90.1193%

cropping, benefits on our task. This is behind attributes of crime scene sketches. Rotation, deformation, and image noises are hardly seen in crime scene sketches, while image shifting is very common. Image cropping can fix errors caused by image shifting. It, however, makes the net become more difficult to train. This is why we remove it from CSS-Net. To make up for the loss of it and to make our net easier to train, we replaced a convolutional layer using a big kernel with two layers using small ones, and they did a good job. The performance of trained model could partly meet the requirements of application.

6.2 Performance Analysis

In this part, we focus on characterizing the performance of our model trained on CSS-Net. To make it visible, we draw the ROC (Receiver Operator Characteristic) curve and the PR (Precision-Recall) curve. For this study, we plotted curves of each categories and average of them. For example, in the curve of crime scene overview sketch, we define it as positive case, and remaining categories are defined as negative cases. Finally, we averaged the results of each classes and plotted the average curve.



(a) ROC Curve (b) Precision-Recall curve

Fig. 4. ROC and PR curves of our CNN

In Fig. 4(a), the ROC curves show that performance of distinguishing map-screenshot location sketch with the rest of two types almost reaches the peak. But capability of recognizing the two other classes is in a low level, which stuck the improvement of our net's classification ability. In Fig. 4(b), PR curves suggest that our model works well on our data set, while there is still much room for

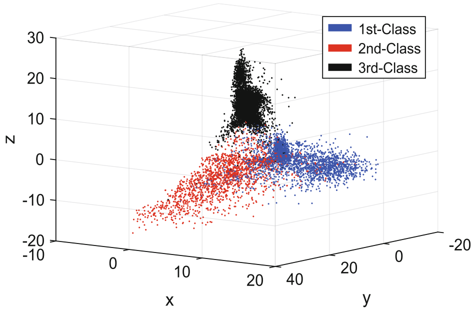


Fig. 5. Distribution of Deep Features: 1st-class represents Crime Scene Overview Sketch, 2nd-class represents Self-Drawn Location Sketch, and 3rd-class represents Map-Screenshot Location Sketch.

improvement, especially for improving capability of recognizing self-drawn location Sketch. Although, in the ROC curve, it is similar to the ability of recognizing crime scene overview sketch.

To show it in a more vivid way, we extracted the deep feature vectors output by the last fully connected layer. Because the number of output equals to number of categories, we can directly plot it on 3-D space. This figure illustrates the spacial distribution of data having experienced forward propagation. From Fig. 5, we can see that the deep features could be distinguished by some decision boundaries, but the degree of classifying difficulty is different. map-screenshot location sketch is the easiest one, while the rest are not discriminative enough. It is behind the results of ROC and PR curves.

7 Conclusion

In this paper, we proposed an application problem of image classification. Then, we built a crime scene sketch dataset and came up with a solution based on convolutional neural network as we called it CSS-Net. Our net performs better than classic ones and can basically meet the requirement of real-life application. But the capability of recognizing self-drawn location sketch and crime scene overview sketch needs to be improved. In this study, our dataset lack a class of negative class to get rid of photos which were usually uploaded as sketches. The number of self-drawn location sketches is not enough. Besides, our architecture is based on AlexNet which is too old. As a result, we will continue this study in two aspects, including expanding the dataset and designing a new architecture with novel technology.

References

1. Acharya, U.R., Oh, S.L., Hagiwara, Y., Tan, J.H., Adeli, H.: Deep convolutional neural network for the automated detection and diagnosis of seizure using EEG signals. Comput. Biol. Med. **100**, 270–278 (2018)

2. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 770–778 (2016)
3. Jia, Y., et al.: Caffe: convolutional architecture for fast feature embedding. In: Proceedings of the 22nd ACM International Conference on Multimedia, pp. 675–678. ACM (2014)
4. Krizhevsky, A., Sutskever, I., Hinton, G.E.: Imagenet classification with deep convolutional neural networks. In: Advances in Neural Information Processing Systems, pp. 1097–1105 (2012)
5. LeCun, Y., Bottou, L., Bengio, Y., Haffner, P., et al.: Gradient-based learning applied to document recognition. Proc. IEEE **86**(11), 2278–2324 (1998)
6. Lyman, M.D.: Criminal Investigation: The Art and the Science. Prentice Hall, Upper Saddle River (2001)
7. Miller, M.T., Massey, P.: The Crime Scene: A Visual Guide. Academic Press, Cambridge (2018)
8. Raghavendra, U., Fujita, H., Bhandary, S.V., Gudigar, A., Tan, J.H., Acharya, U.R.: Deep convolution neural network for accurate diagnosis of glaucoma using digital fundus images. Inf. Sci. **441**, 41–49 (2018)
9. Rumelhart, D.E., Hinton, G.E., Williams, R.J., et al.: Learning representations by back-propagating errors. Cogn. Model. **5**(3), 1 (1988)
10. Russakovsky, O., et al.: Imagenet large scale visual recognition challenge. Int. J. Comput. Vision **115**(3), 211–252 (2015)
11. Simonyan, K., Zisserman, A.: Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556 (2014)
12. Sun, Y., Wang, X., Tang, X.: Deeply learned face representations are sparse, selective, and robust. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 2892–2900 (2015)
13. Szegedy, C., et al.: Going deeper with convolutions. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 1–9 (2015)
14. Wade, C., Trozzi, Y.E., Quantico, V.: Handbook of forensic services. US Department of Justice, Federal Bureau of Investigation, Washington, DC (1999)
15. Weston, P.B., Wells, K.M.: Criminal Investigation: Basic Perspectives. Prentice-Hall, Upper Saddle River (1974)