Vehicle/Human Metadata Analysis Technology and Its Applications

OAMI Ryoma, HOSOMI Itaru, NAKAJIMA Noboru, HARADA Noriaki

Abstract

The video monitoring system was initially introduced for collection and recording of the video recorded by surveillance cameras. However, when it is combined with an image analysis technology it is expected to become a tool for improving the efficiency and functionality for the safety and security services. By targeting the implementation of a video monitoring system as an IT safety/security tool, NEC is developing a metadata analysis technology that is capable of the efficient extraction of specific images or information from the large amount of video collected by surveillance cameras according to purpose.

This paper overviews the component technologies of the metadata analysis technology, focusing on the characteristics of vehicles and humans, including license plate recognition, super-resolution processing, vehicle features model matching, human face/clothing feature extraction and human feature model matching. It also introduces solutions that are applied to these technologies.

Keywords

video monitoring, metadata analysis, vehicle license plate recognition
gate monitoring, traffic flow monitoring, specific person retrieval

1. Introduction

The dissemination of networked cameras has advanced video monitoring systems into the IT systems domain with applications expanded as tools for supporting safety and security services. Surveillance cameras have previously been used mainly to record and store video but their networking has turned them into sensors for detecting events in the field at the instant they occur. This trend has made it an important issue about how to effectively use the large amount of surveillance video collected via networks by routine safety and security services.

At NEC, we are implementing a system for supporting safety and security personnel in quickly identifying current situations by means of metadata analysis of the video information collected from surveillance cameras. Countermeasures are then enabled by automating functions that used to be done visually, such as vehicle gate monitoring and specific person retrievals.

This paper provides an overview of the element technologies used in the automatic extraction of features and characteristics of vehicles and humans in surveillance video and retrieving scenes by the recognition of vehicle license plates and personal features. It also introduces solutions by applying the technologies discussed above.

2. Vehicle Metadata Analysis Technology

This is a technology for the automatic extraction of vehicle-related information, such as the license plate information, vehicle type, vehicle body color and the type of activity in the frame by analyzing the video. When people describe what they have witnessed in vehicle sighting reports, they often use abstract or ambiguous expressions such as “a red station wagon with a license plate beginning with Shinagawa ‘ *wa* ’ was being driven in the wrong direction on the outbound lane of Expressway XX.” A large amount of labor is required to identify a specific vehicle from the video sent from many surveillance cameras, or to manually match the image with the video archive that is already huge and expanding every minute, by using the above information as a query. This section describes the technologies for automating the above work (**Fig. 1**).

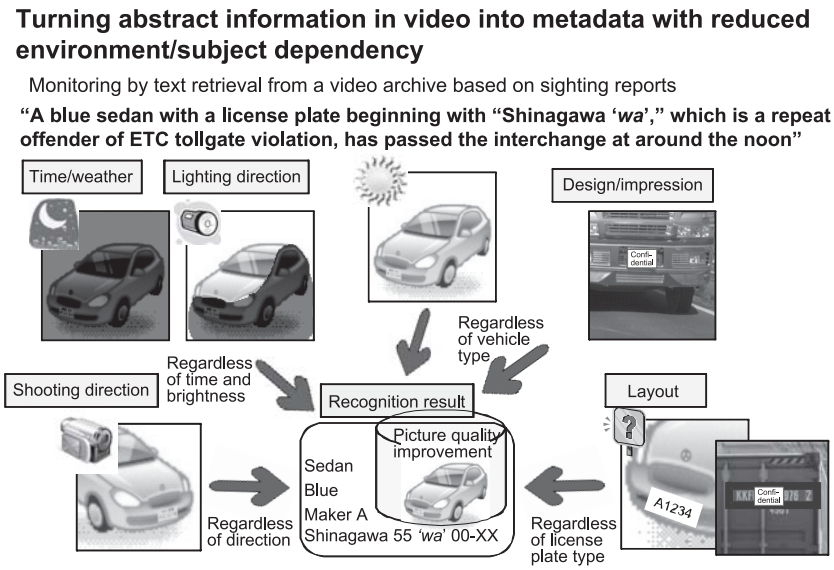


Fig. 1 Vehicle metadata information.

The Japanese Institute of Electronics, Information and Communication Engineers (IEICE) pointed out ten key subjects as significant challenges in the field of pattern recognition and media understanding in its journal. The technological domain dealt with in this paper focuses on two of the above subject areas, "fully-automated image structuring" and "recognition of the characteristics of a scene." In the long term, our efforts will also challenge the technology for the "semantic description of images" that is capable of reading more abstract information such as "a local bus at a temporary halt in front of a crowd of people."

The following subsections describe each of the individual elements of these technologies.

2.1 Image Restoration and License Plate Character Detection Unaffected by Outdoor Environmental Factors

In the outdoor environment, video is affected by various external factors due to the impossibility of specifying the lighting conditions and of capturing various objects other than the analysis target in the image frame. As a result, the use of the video archive collected from cameras installed in streets and shops is often hindered by a picture quality that is not high enough for video analysis.

The license plate recognition, which will be discussed later, also requires an image quality high enough for reading the characters in the video clearly, in order to extract and recognize character patterns with high accuracy. In addition, it is also required to find those objects and characters that are required to be recognized in the video. The technologies for solving these issues are described below.

(1) Image Restoration Reducing the Effects of Compensating for Lighting Variations

This technology analyzes the local lighting intensity value distribution in video and by reducing only the lighting variation components it restores the view of objects under uniform lighting by identifying whether each intensity change component is caused by natural lighting in the camera's installation environment, or by an actual texture change on the object surface. As shown in **Fig. 2** , the effects include improvement of visibility when the surroundings are imaged darkly due to the dazzling of headlamps and reduction of shadows cast by other objects.

(2) Super- resolution Processing

This technology inputs low-resolution video and restores high-resolution images by estimating details. In scene images, it is a usual matter that the same subject is captured in several frames or that the statistical properties of the subject shape or object surface pattern indicate resemblances between adjacent partial domains. Such redundancies in time

(a)Upper row: Input images. Lower row: Image restoration results.

(b) Reduction of shadows cast on the license plate characters area
(Left: Input image. Right: Image restoration results.)

Fig. 2 Image restoration resistant to lighting variations.

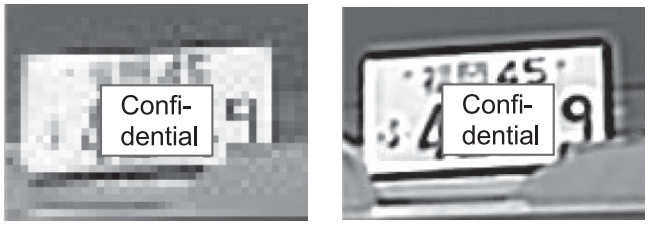


Fig. 3 Result of super-resolution processing (Left: 1 frame among input images. Right: Super-resolution image.)

Fig. 4 Extraction of license plate characters.

or space in the image information can be utilized to restore an image with higher resolution than that of an image sensor device even when a low-resolution camera is used. **Fig. 3** shows the result of application of this technology to an image of a license plate captured with a moving camera. In the case of application of OCR technology for license plate recognition, etc., it is necessary to observe the structures of character details in high definition and the resolution exerts important effects on the accuracy of recognition. This technology is an enhanced application of a technology developed for the processing of documented images [2] .

(3) Detection and Recognition of License Plate Characters from Setting Video

It is a fairly hard task to find characters from a setting video that contains miscellaneous objects. Since the image features to be used as the clues for detecting license plate characters have not been established, it is not easy for a computer to reproduce the operations humans usually execute unconsciously, such as simultaneous and instantaneous character recognition as the means of understanding a scene or the process from character recognition to semantic adaptability evaluation [1] . Therefore, we have adopted a method for determining the positions containing characters, attitudes in terms of 3D characters and categories of characters based on an overall judgment of the following factors;

- Characteristic similarity of local patterns in an image;
- Probability of pattern arrays that seem to be characters;
- Probability of the result of applying character recognition to patterns.
- Probability of character recognition results from the viewpoint of semantic sequence.

We are continuing the character search operation without establishing the character recognition results that can maximize the degrees of probability defined above so that we can eventually determine what kind of character string is captured, in which position and from which direction, with high accuracy (**Fig. 4**).

2.2 Object/Color Recognition Based on Model Matching and Kernel Identifier

The technologies for recognizing objects in video can roughly be classified into those for identifying a specific person or object such as biometric technology and those for identifying the category of target with a generic name such as “human,” “truck” or “motorbike.” This paper refers to the former kind of

technology as specific object recognition and to the latter kind as generic object recognition.

With the generic object recognition, we use the kernel GLVQ which is an extension of the GLVQ [3]), an NEC-original identification/learning method. The kernel GLVQ has advantages that are not found with other methods, such as; expectation of simultaneous optimization of hyper parameters dependent on the identification target; ease of implementation of high generalization performance thanks to the inheritance of empirical risk minimization framework from GLVQ, and; the possibility of formulation in a natural form of identification problems with a large number of categories. It can therefore be regarded as a recognition method suitable for multi-class identification problems with which the properties of the identification target cannot be specified in advance, such as for generic object recognition and character recognition.

When the recognition of a vehicle class is in question, the vehicle class dictionary is created by first determining categories (truck, motorbike, person, standard vehicle, etc.), then by preparing the images corresponding to them, generating their feature vectors using digitalization and quantization of image patterns by means of orthogonal transforms, etc., and by applying learning for associating the spatial placement of feature vectors to the above categories. At the time of recognition, the feature vectors of the target vehicle are obtained by a similar feature extraction method to the above and by matching with the dictionary to identify the category that the target vehicle belongs to. The recognition of the vehicle body color also uses a similar learning method in order to enable matching with ambiguous, subjective expressions such as "blackish" or "metallic."

When the recognition of a specific object is in question, object shapes are modeled accurately and the target object image is matched with the models by considering image deformation depending on the shooting angle. When applied to the recognition of vehicle shape, emblem and logo, this method makes it possible to extract metadata information that is much more detailed than the vehicle class information (**Fig. 5**).

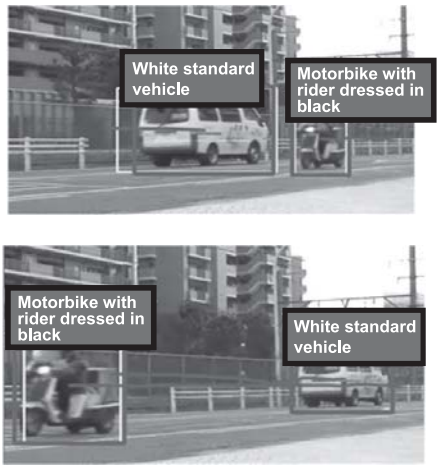


Fig. 5 Recognition and tracking of vehicles (Upper row: Initial. Lower row: Subsequent frame).

2.3 Tracking and Motion Recognition of a Target Object

We are developing the technology for tracking the recognition target in an environment including objects other than the target as well as miscellaneous natural backgrounds that vary every second. To continue tracking successfully even when the recognition target suddenly changes its motion for a sudden start, sudden stop or sudden lane change or when the target is hidden behind another object, we introduce a method that merges the information on the motion of the target object and that of the appearance of the target object and updates the probability of these conditions with a sequential prediction using Monte Carlo approximation.

The history and predicted values of the information on the target object including its displacement speed, acceleration, shape (view) and position are stored internally, so that judgments of vehicle behavior such as stopping, being driven along the lane correctly or being driven at an optimum speed is possible at anytime. In addition, risky driving patterns such as driving in the wrong direction, ignoring traffic lights, driving at an excessive speed, meandering or dropping a load are also detected.

3. Person Retrieval Based on Human Metadata Analysis

3.1 Person Retrieval System Based on Text Input

We also have the technology for retrieving a person from a large amount of video using the clothing or face features of the person. Such retrieval is possible either by input of a text such as keywords or input of sample images. NEC has developed a person retrieval system that can search a person by input of a text describing the color and type of clothing or of a sample image of the face (**Fig. 6**).

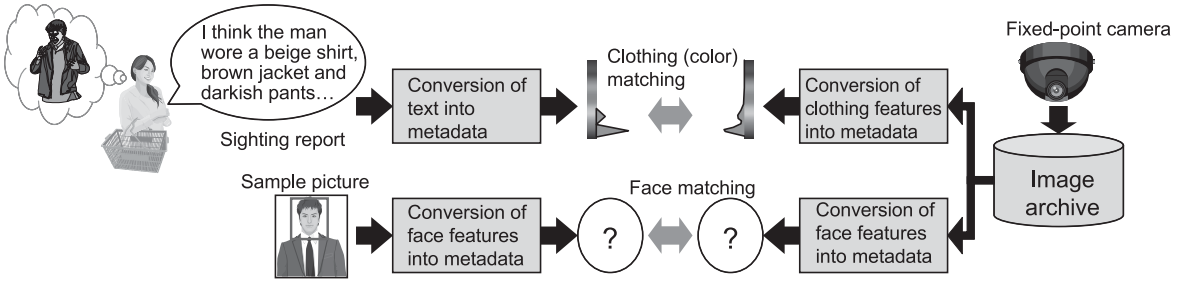


Fig. 6 Outline of operations of the person retrieval system.

Person retrieval by the features of the clothing is used for example when seeking a suspicious or missing person according to an eyewitness testimony. Most of such testimonies are very sketchy, for example "in a blue zip-up jacket on a white shirt," because they are based on the memory of the witness. The system is required to compare such testimonies with the features of a large number of people in the video recorded with surveillance cameras, etc.

Our person retrieval system extracts the information on the type and color of clothing from the input text. If the clothing is expressed in several colors e.g. ("blue on white background," etc.), the share of each color is estimated for the entire clothing according to the way that the colors are expressed. In addition, the system also judges the colors of different body parts from the type of clothing, for example the color of jeans becomes the color of the lower body and the color of a shirt becomes the color of the upper body. This judgment method utilizes the ontology (dictionary) of clothing configuration knowledge that uses a hierarchical structure. If a person wears a "blue jacket and a white shirt," the system can make a judgment that the color distribution of the front of the upper body is mostly or partially blue or that of the rear of the upper body is completely blue. While the expression of "navy-blue clothing" does not permit a judgment of whether the person is entirely blue or only the upper or lower body is blue, the expression of "a yellow shirt and a navy-blue clothing" can lead to the judgment that only the upper body is navy blue. In addition, ambiguous expressions such as "reddish" or "dark green" are used to adjust the color values.

The person retrieval by color of clothing is performed by converting the results of the processing operations described above into metadata on colors and matching it with the metadata of persons extracted from the video archive. Specifically, the system generates the color distribution for each of the front, non-front, upper body and lower body as the clothing features values.

3.2 Extraction and Matching of Clothing Features

Extraction of the features of clothing begins with extraction of the person domain from the video and then advances to the extraction of the image features of the domain corresponding to the clothing. Since the color is regarded as the most important clothing feature in the sighting report, we extract the feature values of the colors. Specifically, our system calculates the distribution of colors in each domain and uses the results as the feature values.

Since the clothing feature used in person retrieval often distinguishes the upper and lower body as discussed in the previous section, the person retrieval system is required to be capable of specifying features of the upper and lower bodies separately. In addition, the view of clothing varies depending on the orientation. Therefore, the clothing feature extraction process involves the following operations:

(1) Auto separation of the clothing feature into the upper and lower body.

(2) Auto determination of the orientation of the person.

With auto separation of the clothing feature into upper and lower body features, the function obtained by mapping the pixel values of the clothing on a vertical axis is calculated as shown in **Fig. 7** . Next, the point where the value of the function changes suddenly is detected and interpreted as the boundary between the upper and lower body. In the case of clothing with inseparable upper and lower parts such as a one-piece suit, this boundary is not obtained because the clothing feature does not change between the upper and lower body. If the clothing feature does not involve a separation position, as in this case, the clothing is separated on the centerline.

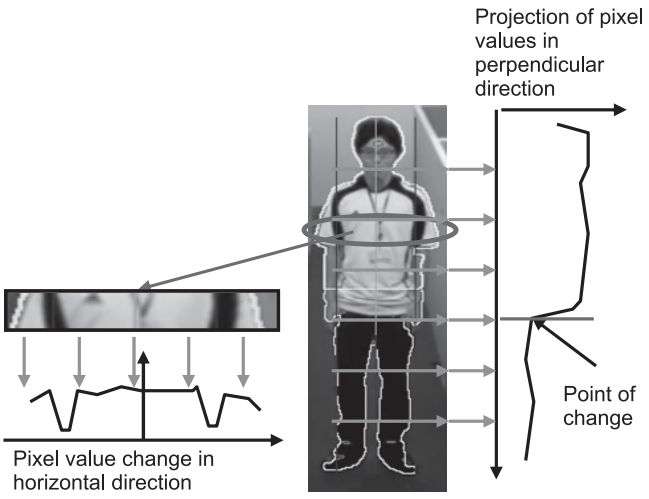


Fig. 7 Upper/lower body separation and symmetry judgment based on changes in pixel values of clothes.

The difference in view depending on the positioning of the person is used as previous knowledge. If a person standing on a floor surface visible from the monitoring camera can be imaged, the whole body of the person can be captured. However, if the lower body of a person is hidden by an obstacle such as a desk, only the upper body of the person can be captured. In such a case, the separation into the upper and lower body estimates the bottom position of the person domain and reflects the estimation result in the processing of the difference of view.

Determination of the orientation of the person (frontality) uses three kinds of information including the orientation of the face, the motion of the person and the degree of symmetry of the clothing. The orientation of the face is estimated at the same time as face detection. The motion of the person is determined by calculating the motion vector between frames and making a judgment according to its orientation. For the degree of symmetry of the clothing, the changes in the pixel values of the clothing in the horizontal direction at different heights is identified as shown in Fig. 7 and the degree of symmetry with respect to the central axis is calculated. Since the image of the clothing of a person often presents horizontal symmetry when the person is imaged from the front or rear, the degree of symmetry can serve the orientation determination when it is numerically quantized. The three kinds of information described above are integrated in the determination of the orientation (front, rear or other).

After the clothing features are extracted as described above, retrieval is executed by matching the clothing feature values by taking the conditions of extraction into consideration. Basically, the clothing feature values are collected separately for the upper and lower body and for each orientation. For example, when the clothing feature values of both the upper and lower body in the front orientation are obtained, they are matched with the queried frontal feature values of the upper and lower body and the retrieval score is calculated. If only the feature value of the upper body is available, matching is performed using only the feature value of the upper body. If the orientation cannot be determined when the feature values are extracted, collection is performed with both the frontal and non-frontal query feature amount and the result with the higher score is adopted. In case the upper and lower bodies cannot be separated, the feature values of the upper and lower bodies are synthesized into a single feature value before matching. These techniques make possible flexible and accurate collation according to the conditions of the clothing feature value extraction.

3.3 Extraction and Matching of Face Features

With the person retrieval by face, the face image is input as the query and the person with the matching face feature values is searched. For this purpose, the input video is subjected to face detection and the face feature values are extracted from the extracted face domain. The face feature value is associated with the clothing feature value described above before being saved. This procedure enables person retrieval by both clothing and face.

Matching of face feature value involves the following operations:

(1) Matching of face feature value considering the face orientation and the lighting mode.

(2) Matching using partial face domains.

With operation (1), the face feature values of various orientations and lighting conditions are simulated for use in matching. With operation (2), matching is not performed on the whole face but is weighted on a part of the face. These operations contribute to improvement in the accuracy of face matching.

4. Examples of Applications

4.1 Vehicle License Plate Recognition

Among the applications of the metadata analysis technologies, this section introduces cases applying the license plate

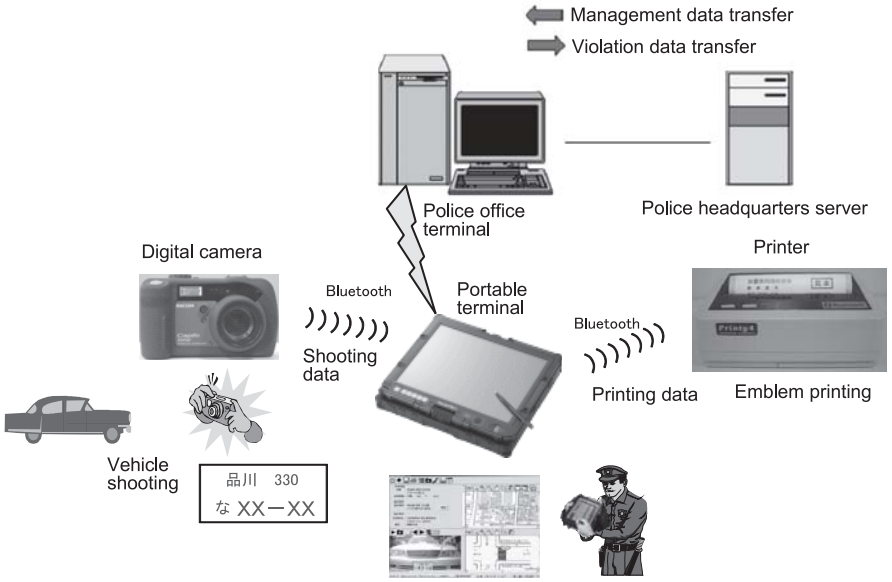


Fig. 8 Image of illegal parking crackdown.

Fig. 9 Long-hour parking crackdown terminal.

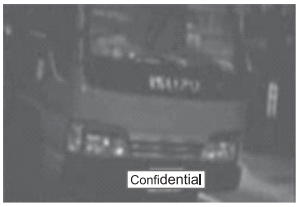
recognition technology, which is the technology featuring the most reliable individual identification.

Japanese Police started outsourcing a crackdown on illegal parking in no-parking zones to the private sector in 2006. At the same time, many districts introduced the license plate recognition function at the portable terminals for use in the crackdown, in order to improve the efficiency of the license plate number input. When the violating vehicle is shot by a camera during the investigation, its license plate number is read, printed on the confirmation emblem (violation sticker) and sent to the center (**Fig. 8** and **Fig. 9**).

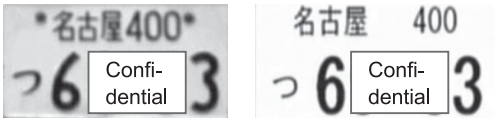
We also apply the license plate recognition technology for tollgate monitoring applications. The system we have developed does not require a specific shooting environment to consider changes in the background scenery or lighting conditions and positioning relationships with target recognition being performed from moment to moment. It achieves high recognition accuracy even at a vehicle speed of more than 100 km/h in the actual driving environment (**Fig. 10**).

An example of a case of introduction of this technology is seen in the immigration/customs gates on the boundary between Hong Kong and China (**Photo**). The system is used to simplify the immigration control and customs clearing operations by reading the vehicle license plate numbers. A partial operation on limited lanes has already been started this year and the actual operation covering all lanes is scheduled for 2011. The system achieves recognition accuracy over 98.5% of vehicles passing through the gates by monitoring using fixed cameras.

For the future, we aim at integrating other metadata analysis technologies with the above procedures including those using car-mounted cameras, in order to offer video monitoring and security solutions that can be applied in a variety of shooting environments.



(a) Input image



(b) Frontal corrected image (left) and recognition result (right)

Fig. 10 License plate reading technology for car gate monitoring system.

Lok Ma Chau Gate, which is one of the four boundary gates introducing the system.

Customs clearance gate.

Photo Hong Kong Immigration/Customs Support System

4.2 Gate/Traffic Flow Monitoring

Accidental trespassing of pedestrians, driving of vehicles in the wrong direction and entrance of vehicles of nonconforming types can cause problems on expressways. Spreading of the ETC system has been increasing the illegal clearing or violation of ETC rules. These trends are raising the need for the monitoring of illegal passage because a huge number of vehicles are thus evading toll payments, including the use of vehicle type disguises, gate bulldozing, high-speed run-through and tailgating. In addition, an increase in the number of smart interchanges, which are unattended interchanges for the exclusive use of ETC users introduced in 2006, has also increased the need for automation of illegal passage detection. Since smart interchanges are often installed directly on the side roads of highways in suburban areas, it is regarded as being essential to provide these unattended interchanges with a function for monitoring and notification of erroneous or intended illegal passages.

(a) Detection of meandering

(b) Recognition of license number, class and color of meandering vehicle

(c) Detection of red light running

(d) Recognition of license number, class and color of red light running vehicle (Desired for vehicles in Singapore)

Fig. 11 Abnormal driving detection systems.

Application of gate/traffic flow monitoring to general highways and expressways is under discussion in many countries worldwide in order to prevent accidents due to traffic violations and risky actions by detecting them automatically and enforcing warnings and performing crackdowns. This technology also has a wide range of potential applications including the monitoring of vehicle entrance/exit in/from public facilities, stores and corporate parking lots, counterterrorism measures, illegal acts surveillance as well as cashless charge

payments, attendance management, VIP/blacklist collation and collection of statistical information on customers (**Fig. 11**).

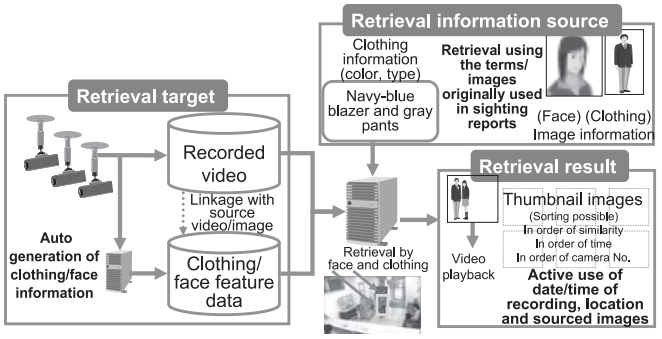
4.3 Suspicious Individual Collation

Installation of surveillance cameras is no longer a special topic of traffic terminals and other facilities gathering visitors. As big cities where international events are often held are extending surveillance cameras in their streets, it is expected that a large amount of video recorded with these surveillance cameras will be accumulated in the future.

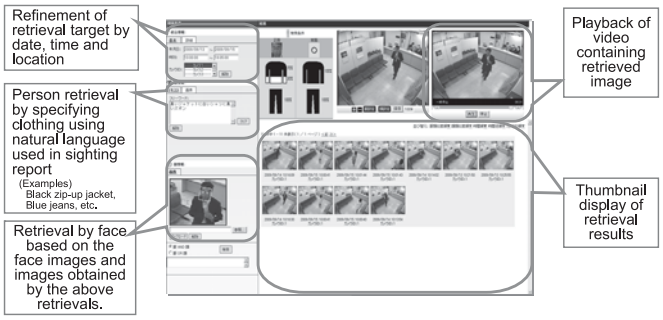
Previously, when an incident or accident occurred, the recorded video was played back assuming that it would be checked visually by an operator. This led to the need for a vast amount of human resources and time, and the extraction of information was dependent on the knowledge and judgments of operators.

At NEC, we noticed the facial features, clothes, age and sex as the features that could characterize individuals and proceeded to develop a system that automatically extracts a video scene that may record the target person based on an ambiguous description of a witness. Previous technologies handled still images only and their application was limited to a search of previously set color or a resembling scene. When the retrieval targets were video data, it has been general practice that the content administrator registers the keywords characterizing each piece of image data in advance.

The system being developed performs semantic analysis of video recorded by surveillance cameras. When the operator inputs the clothing/face features in a natural language, the system can extract and play back the scenes showing the specific person from the stored surveillance video. This can be applied over a wide range including criminal investigation based on sighting reports or face images, search for stray children, tracing of people in care facilities etc. (**Fig. 12**).



(a) Image of the system



(b) Example of system control window

Fig. 12 Image of person retrieval system.

5. Future Perspectives

In the above, we introduced our efforts for the implementation of a metadata analysis technology that focuses on vehicle/personal features information such as the license plate, vehicle class and vehicle color as well as on facial features and clothes.

Applying the metadata analysis technology to a huge video archive of vehicles and persons is expected to significantly improve the efficiency of safety and security services. In the future, we intend to actually accelerate application of the most advanced technologies in the field so that we can thereby contribute to the creation of a safer and more secure society.

References

1) Washimi, et. Al., “PATAAN NINSHIKI, MEDIA RIKAI NO JUDAI CHARENJI TEEMA, SHINGAKUSHI, Vol.92, No.8, pp.665-675, 2009.
2) N. Nakajima, et. al., Video Mosaicing for Document Imaging, Proc. CBDAR, pp.171-178, 2007.
3) A. Sato and K. Yamada, Generalized Learning Vector Quantization, NIPS 8, pp. 423-429, MIT Press, 1996.

Authors' Profiles

OAMI Ryoma
Principal Researcher
Information and Media Processing Laboratories

HOSOMI Itaru
Principal Researcher
Information and Media Processing Laboratories

NAKAJIMA Noboru
Assistant Manager
Ubiquitous Solutions Group
Advanced Technology Solutions Division
NEC Informatec Systems, Ltd.

HARADA Noriaki
Senior Manager
Platform Strategic Marketing Division