Machine Vision and Applications
DOI 10.1007/s00138-013-0535-8

SPECIAL ISSUE PAPER

Context-based person identification framework for smart video surveillance

Liyan Zhang · Dmitri V. Kalashnikov · Sharad Mehrotra · Ronen Vaisenberg

Received: 1 February 2013 / Revised: 4 June 2013 / Accepted: 11 July 2013

Abstract Smart video surveillance (SVS) applications enhance situational awareness by allowing domain analysts to focus on the events of higher priority. SVS approaches operate by trying to extract and interpret higher "semantic" level events that occur in video. One of the key challenges of SVS is that of *person identification* where the task is for each subject that occurs in a video shot to identify the person it corresponds to. The problem of person identification is especially challenging in resource-constrained environments where transmission delay, bandwidth restriction, and packet loss may prevent the capture of high-quality data. Conventional person identification approaches which primarily are based on analyzing facial features are often not sufficient to deal with poor-quality data. To address this challenge, we propose a framework that leverages heterogeneous contextual information together with facial features to handle the problem of person identification for low-quality data. We first investigate the appropriate methods to utilize heterogeneous context features including clothing, activity, human attributes, gait, people co-occurrence, and so on. We then propose a unified approach for person identification that builds on top of our generic *entity resolution* framework called RelDC, which can integrate all these context features to improve the quality of person identification. This work thus links one well-known problem of *person identification* from the computer vision research area (that deals with video/images) with another well-recognized challenge known as *entity resolution* from the database and AI/ML areas (that deals with textual data). We apply the proposed solution to a real-world dataset consisting of several weeks of surveillance videos. The results demonstrate the effectiveness and efficiency of our approach even on low-quality video data.

Keywords Person identification · Context information · Entity resolution · Smart video surveillance

1 Introduction

Advances in sensing, networking, and computational technologies have allowed the possibility of creating sentient pervasive spaces wherein sensors embedded in physical environments are used to monitor its evolving state to improve the quality of our lives. There are numerous physical world domains in which sensors are used to enable new functionalities and/or bring new efficiencies including intelligent transportation systems, reconnaissance, surveillance systems, smart buildings, smart grid, and so on.

In this paper, we focus on smart video surveillance (SVS) systems wherein video cameras are installed within buildings to monitor human activities [17,18,33]. Surveillance system could support variety of tasks: from building security to new applications such as locating/tracking people, inventory, or tasks like analysis of human activity in shared spaces (such as offices) to bring improvements on how the building is used. One of the key challenges in building smart surveillance systems is that of automatically extracting semantic information from the video streams [31,32,35,36]. This semantic information may correspond to human activities, events of interest, and so on that can then be used to create a representation

This work was supported in part by NSF grants CNS-1118114, CNS-1059436, CNS-1063596. It is part of NSF supported project *Sherlock @ UCI* (http://sherlock.ics.uci.edu): a UC Irvine project on Data Quality and Entity Resolution [1].

L. Zhang (✉) · D. V. Kalashnikov · S. Mehrotra · R. Vaisenberg
Department of Computer Science, University of California, Irvine, CA, USA
e-mail: zhangliyan.uci@gmail.com

Published online: 11 August 2013

Fig. 1 Example of surveillance video frames

of the state of the physical world, e.g., a building. This semantic representation, when stored inside a sufficiently powerful spatio-temporal database, can be used to build variety of monitoring and/or analysis applications. Most of the current work in this direction focuses on computer vision techniques. Automatic detection of events from surveillance videos is a difficult challenge and the performance of current techniques often leaves a room for improvement. While event detection consists of multiple challenges, (e.g., activity detection, location determination, and so on), in this paper we focus on a particularly challenging task of *person identification* [3,4].

The challenge of *person identification* (PI) consists of associating each subject that occurs in the video with a real-world person it corresponds to. In the domain of computer vision, the most direct way to identify a person is to perform *face detection* followed by *face recognition*, the accuracy of which is limited even when video data are of high quality, due to the large variation of illumination, pose, expression, and occlusion, etc. Thus, in the resource-constrained environments, where transmission delay, bandwidth restriction, and packet loss may prevent the capture of high quality data, face detection and recognition becomes more complex. We have experimented with Picasa's face detector on our video dataset, [1] and found that it can detect faces in only 7 % of the cases and then among them it can recognize only 4 % of faces.

Figure 1 illustrates the example of frames in our video dataset, where only one face is successfully detected (solid-line rectangle) utilizing the current face detection techniques. Several reasons account for the low detection rate: (1) faces cannot be captured if people walk with their back to the cameras; (2) faces are too small to be detected when people are far away from cameras; (3) the large variation of people's pose and expression brings more challenges to face detection. Thus the traditional face detection and recognition techniques are not sufficient to handle the poor-quality surveillance video data.

To deal with the poor-quality video data and overcome the limitation of current face detection techniques, we shift our research focus to context-based approaches. Contextual data such as time, space, clothing, people co-occurrence, gait, and activities, are able to provide the additional cues for person identification. Consider the frames illustrated in Fig. 1 for example. Although only one face is detected and no recognition results are provided, the identities of all the subjects can be estimated by analyzing the contextual information. First, time and foreground color continuities split the eight frames into two sequences or shots. The first four frames construct the first shot, and the following four frames form the second shot, where subjects within each shot describe the same entities. Furthermore, some other contextual features reveal the high possibility that the two subjects in these shots are the same entity. For instance, they both share the similar clothing (red T-shirt and gray pants), they perform similar activities (walking in front of the same camera, though in opposite directions), and they have the same gaits (walking speed). Thus the context features help to reveal that the subjects in the eight frames very likely refer to the same entity. To identify this person, face recognition process is usually inevitable. However, the activity information can also provide extra cues to recognize people's identity. In the above example, suppose that the first shot in Fig. 1 is the first shot of that day where a person enters the corner office which belongs to "Bob". Then most probably this person is "Bob" because in most cases, the first person entering the office should have the key. Therefore, by analyzing contextual information even without face recognition results, we can predict that the very likely identity of subject in all the eight frames is "Bob". The example demonstrates the essential role that contextual data play in

[1] 704 × 480 resolution per frame.

the *person identification* issue for the low-quality video data. Another significant advantage of context information is its weaker sensitivity to video data quality as compared with that of face recognition. That makes context-driven approaches more robust and reliable when dealing with poor quality data.

In this paper, we extend our previous work [39] to explore a novel approach to leverage contextual information, including time, space, clothing, people co-occurrence, gait, and activities to improve the performance of *person identification*. To exploit contextual information, we connect the problem of person identification with a well-studied problem of *entity resolution* [5,21,26], which typically deals with textual data. *Entity resolution* is a very active research area where many powerful and generic approaches have been proposed, some of which could potentially be applied to the person identification problem. In this paper, we first investigate methods for extracting and processing several different types of context features. We then demonstrate how to apply a relationship-based approach for entity resolution, called RelDC [23], to the person identification problem. RelDC is an algorithmic framework for analyzing object features as well as inter-object relationships, to improve the quality of entity resolution. In this paper we will demonstrate how RelDC framework for *entity resolution* could be leveraged to solve a *person identification* problem that arises when analyzing video streams produced by cameras installed in the CS Department at UC Irvine. Our empirical evaluation demonstrates the advantage of the context-based solution over the traditional techniques, as well as its effectiveness and robustness. The proposed approach shows clear improvements over approaches that only exploit facial features. The improvement is even more pronounced for low-quality data, as it relies on contextual features that are less sensitive to deterioration of data quality.

The rest of this paper is organized as follows: We start by introducing the related work in Sect. 2. Then in Sect. 3, we present the proposed approach for context based person identification. Section 4 demonstrates experiments and results. Finally, we conclude in Sect. 5 by highlighting key points of our work.

2 Related work

2.1 Video-based person identification

The conventional approaches for *person identification* are to first use face detection followed by face recognition. Figure 2 illustrates the basic schema for person identification. Given a face frame, after locating faces via a face detector, the extracted faces are passed to a matcher which leverages the face recognition techniques to measure the similarities between the extracted faces and "gallery faces" (where true identities of people are known) to determine the identities of the extracted faces.

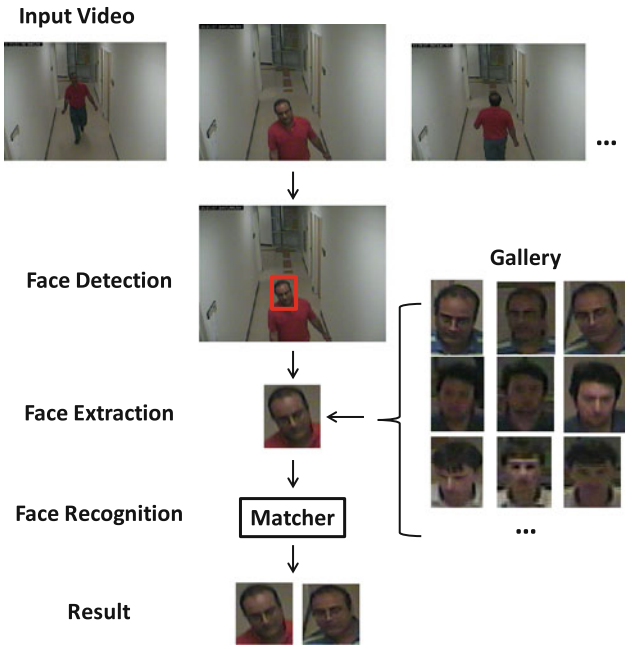

Fig. 2 Example of basic person identification process

In general, face detection is the first and essential component used in person identification. However, in our test dataset, only a small proportion of faces (7 %) could be detected using the current face detection techniques, and out of them very few (4 %) could be recognized, due to the poor quality of video data in our surveillance setting. The failure of detection for most faces makes it impossible to apply the subsequent face recognition process in 93 % of the cases. Hence, the task of achieving high-quality person identification becomes a challenge for video of poor quality.

Face recognition is another active topic of research that has attracted significant attention in the past two decades. Most of the research efforts have focused on techniques for still images, especially face representation methods. Recently, descriptor-based face representation approaches have been proposed and proven to be effective. They include Local Binary Pattern (LBP) [2] describing the microstructure of faces, SIFT and Histogram of Oriented Gradients (HOG) [10], and so on. These face recognition techniques are able to achieve good performance in controlled situations, but tend to suffer when dealing with uncontrolled conditions where faces are captured with a large variation in pose, illumination, expression, scale, motion blur, occlusion, etc. These nuisance factors may cause the differences in appearance between distinct shots of the same person to be greater than those between two people viewed under similar conditions. Thus leveraging context features could bring significant improvement on top of techniques that rely on low-level

visual features only, especially in the context of surveillance videos.

Compared with still images, videos often have more useful features and additional context information that can aid in face recognition. For example, a video sequence would often contain several images of the same entity, which potentially shows the entity's appearance under different conditions. Surveillance videos usually have temporal and spatial information available, which still images do not always have. In addition, video frames are capable of storing the objects in different angles, which contain 3-D geometric information. To better leverage these properties, some face recognition algorithms have been proposed to operate on video data. They include using temporal voting to improve the identification rates, extracting 2-D or 3-D face structures from video sequences [12–14]. However, these methods do not fully exploit the context information and very few of them address the problem of integration of heterogeneous context features.

In this paper, we propose to leverage heterogeneous contextual information to improve the performance of video-based face recognition. To integrate the heterogeneous contextual features together, we connect the problem of *person identification* with the well-studied *entity resolution* problem and apply our entity resolution RelDC framework to construct a relationship graph to resolve the corresponding person identification problem.

2.2 Entity resolution

High quality of data is a fundamental requirement for effective data analysis which is used by many scientific and decision-support applications to learn about the real-world and its phenomena [15,16,24]. However, many Information Quality (IQ) problems such as errors, duplicates, incompleteness, etc., exist in most real-world datasets. Among these IQ problems, *Entity Resolution* (also known as deduplication or record linkage) is among the most challenging and well-studied problem. It arises especially when dealing with raw textual data, or integrating multiple data sources to create a single unified database. The essence of ER problem is that the same real-world entities are usually referred to in different ways in multiple data sources, leading to ambiguity. For instances, the real-world person name 'John Smith' might be represented as 'J. Smith', or misspelled as 'John Smitx'. Besides, two distinct individuals may be referred as the same representation, e.g., both 'John Smith' and 'Jane Smith' referred to as 'J. Smith'. Therefore, the goal of ER is to resolve these entities by identifying the records representing the same entity.

There are two main instances for ER problem: *Lookup* [5,21] and *Grouping* [5,26]. *Lookup* is a classification problem, with the goal of identifying the object that each reference refers to. *Grouping* is a clustering problem, whose goal is to correctly group the representations that refer to the same object. We primarily will be interested in an instance of the lookup problem. Our research group at the University of California, Irvine, has also contributed significantly to the area of ER in the context of Project Sherlock@UCI, e.g., [8,19,20,28–30,38]. The most related work of our group is summarized next.

2.2.1 Relationship-based data cleaning (RelDC)

To address the entity resolution problem, we have developed a powerful disambiguation engine called the *Relationship-based Data Cleaning* (RelDC) [6,7,21–23,27]. RelDC is based on the observation that many real-world datasets are relational[2] in nature, as they contain information not only about entities and their attributes, but also *relationships* among them as well as *attributes* associated with the relationships. RelDC provides a principled domain-independent methodology to exploit these relationships for disambiguation, significantly improving data quality.

Relationship-based data cleaning (RelDC) works by representing and analyzing each dataset in the form of entity-relationship graph. In this graph, entities are represented as nodes and edges correspond to relationships among entities. The graph is augmented further to represent ambiguity in data. The augmented graph is then analyzed to discover interconnections, including indirect and long connections, between entities which are then used to make disambiguation decisions to distinguish between same/similar representations of different entities as well as to learn different representations of the same entity. RelDC is based on a simple principle that entities tend to cluster and form multiple relationships among themselves.

After the construction of entity-relationship graphs, the algorithm computes the *connection strengths* between each uncertain reference and each of the reference's potential "options" that entities it could refer to. For instance, reference 'J. Smith' might have two options: 'John Smith' and 'Jane Smith'. The reference will be resolved to the option that has the strongest combination of the connection strength and the traditional feature-based similarity. Logically, the computation of the connection strength can be divided into two parts: first finding the connections which correspond to paths in the graph and then measuring the strength in the discovered connections. In general, many connections between a pair of nodes may exist. For efficiency, only the important paths are considered, e.g., L-short simple paths. The strength of the discovered connections is measured by employing one of the connection strength models [21]. For instance, one model

[2] We use the standard definition of relational datasets as used in the database literature.

computes the connection strength of a path as the probability of following the path in the graph via a random walk.

After the connection strength is computed, this problem is transformed into an optimization problem of determining the weights between each reference and each of reference's option nodes. Once the weights are computed by solving the optimization problem, RelDC resolves the ambiguous reference to the option with the largest weight. Finally, the outcome of the disambiguation is used to create a regular (cleaned) database.

3 Context-based framework for person identification

3.1 Problem definition

Let $\mathcal{D}$ be the surveillance video dataset. The dataset contains K video frames $F = \{f_1, f_2, \ldots, f_K\}$ wherein motion has been detected. Let t_i denote the time stamp of each frame f_i. When a frame f_i contains just one subject, we will refer to the subject as x_i, or as x^{f_i}. Let $P = \{p_1, p_2, \ldots, p_{|P|}\}$ be the set of (known) people of interest that appear in our dataset. Then the goal of person identification is for each subject x_i to compute w_{ij} which denotes the probability that x_i is person p_j, and correctly identify the person $p_k \in P$ that subject x_i corresponds to. If the subject is not in P, then the algorithm should output $x_i = \texttt{other}$. Table 1 summarizes some of the notations throughout this paper.

Figure 3 illustrates an example of the person identification problem, where the goal is to determine whom the subject in each video frame refers to: "Bob" or "Alice". We can observe that the entity resolution problem has a very similar goal, that is, to associate each uncertain reference to an object in the database with the real-world object. Hence in this paper we demonstrate how to apply one entity-resolution framework called RelDC to the problem of person identification. The framework will exploit the relationships between contextual features of subjects in the video surveillance to improve the quality of the person identification task.

3.2 General framework

Figure 4 illustrates the general framework of context-based person identification for surveillance videos. Given the stored videos from surveillance cameras, the framework first segments the frames with motion into shots based on temporal information. To facilitate person identification based on person faces the framework performs several preliminary steps such as face detection, extraction, facial representation. and recognition. It then extracts the contextual features including people's clothing, attributes, gait, activities, etc. After the extraction of face and contextual features, the framework constructs the entity-relationship graph and then applies the entity resolution algorithm RelDC on the graph to perform the corresponding person identification task.

In the following we discuss how to extract contextual features from surveillance videos and then leverage RelDC framework to integrate these features together to resolve the person identification problem.

3.3 Contextual feature extraction

Contextual features can provide additional cues to facilitate video-based person identification, especially for poor-quality video data. In the following, we describe how to extract and leverage contextual features, such as people's clothing, attribute, gait, activities, co-occurrence, and so on, to improve the performance of person identification.

Table 1 Notation and description

Notation	Meaning
$\mathcal{D}$	The surveillance video dataset being processed
$F = \{f_1, f_2, \ldots, f_K\}$	The set of detected motioned video frames
$S = \{s_1, s_2, \ldots, s_{\|S\|}\}$	The set of shots after video segmentation
$X = \{x_1, x_2, \ldots, x_{\|X\|}\}$	The set of subjects appearing in the video
$P = \{p_1, p_2, \ldots, p_{\|P\|}\}$	The set of real-world people of interest
w_{ij}	The probability that subject x_i is person p_j
$S^C(x_i, x_j)$	The cloth similarity between subjects x_i and x_j
$S_{ij}^{act}(act_i^m, act_j^n)$	The similarity between activity act_i^m and act_j^n
$\mathbb{P}(x_i = p_j \| act_i, t_k)$	The probability that subject x_i is person p_j based on activity and time
A_i	The attribute vector for subject x_i
$FR(x_i, p_j)$	The probability that subject x_i is person p_j based on facial features
$G = (V, E)$	The entity relationship graph
$cs(x_i, p_j)$	The connection strength measure between subject x_i and person p_j

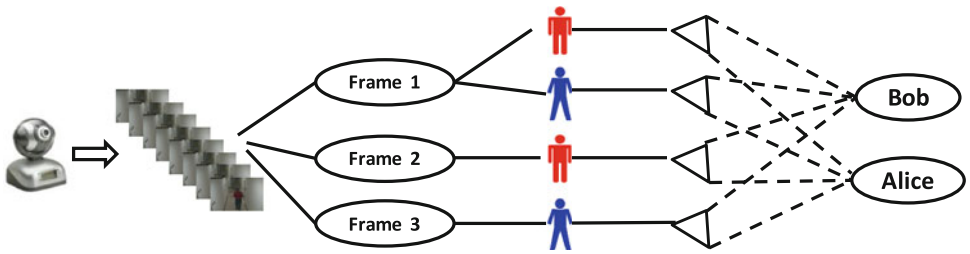

Fig. 3 Example of person identification for surveillance videos

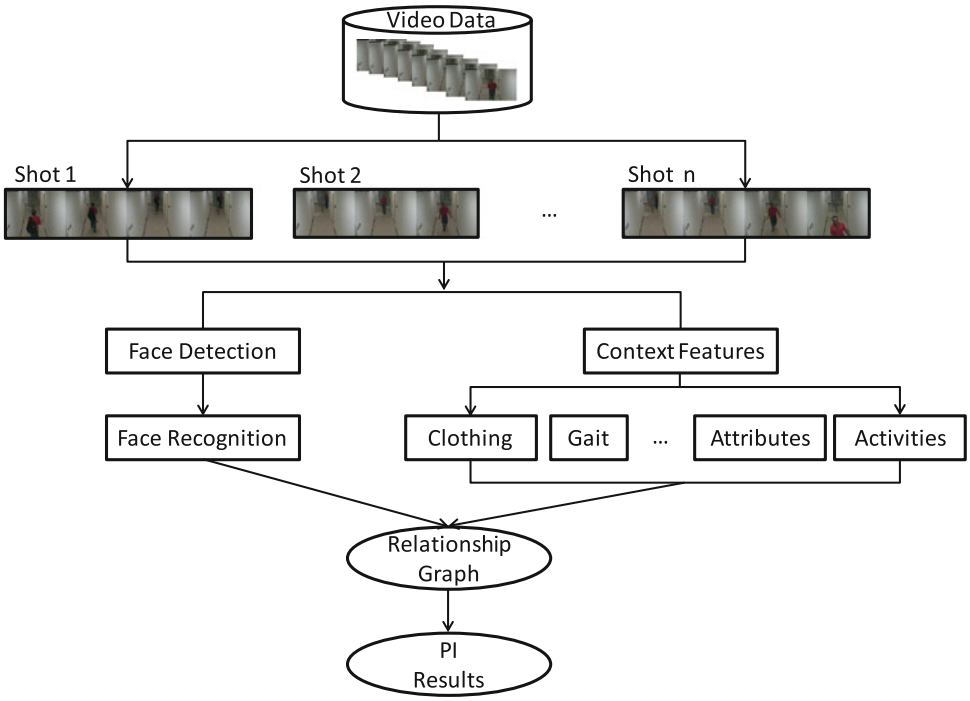

Fig. 4 General framework for context-based approach

3.3.1 Temporal segmentation

We first describe temporal segmentation which is an essential part in video processing. We segment videos into *shots*. Intuitively, subjects appearing in consecutive frames are likely to be the same person. Hence, we initially group frames into shots just based on the time continuity. But time continuity alone cannot guarantee person continuity. If the subjects' color histograms of two consecutive frames are significantly different indicating potentially different people, the shot is split further at such break points.

Suppose that we obtain a set of shots $S = \{s_1, s_2, \ldots, s_{|S|}\}$ after the video segmentation. Most of the time the frames that belong to the same shot describe the same entities. Thus the person identification task reduces from identifying the subjects in an image to identifying the subjects in a shot. We next describe how to extract contextual features for a shot.

3.3.2 Clothing

People's clothing can be a good discriminative feature for distinguishing among people [11,37]. Although people change their clothes across different days, they do not change it too often within shorter period of time, and hence the same clothing in such cases is often strong evidence that two images contain the same person. To accurately capture the clothing information of an individual in an image, we separate the person from the background by applying a background subtraction algorithm [4]. After color extraction processing, the foreground area is represented by a 64-dimensional vector, which consists of a 32-bin hue histogram, a 16-bin saturation histogram, and a 16-bin brightness histogram. Figure 5 shows an example of the extracted foreground image and corresponding color histogram.

The extracted clothing features can be used to compute the clothing-based similarity among subjects. For each pair of

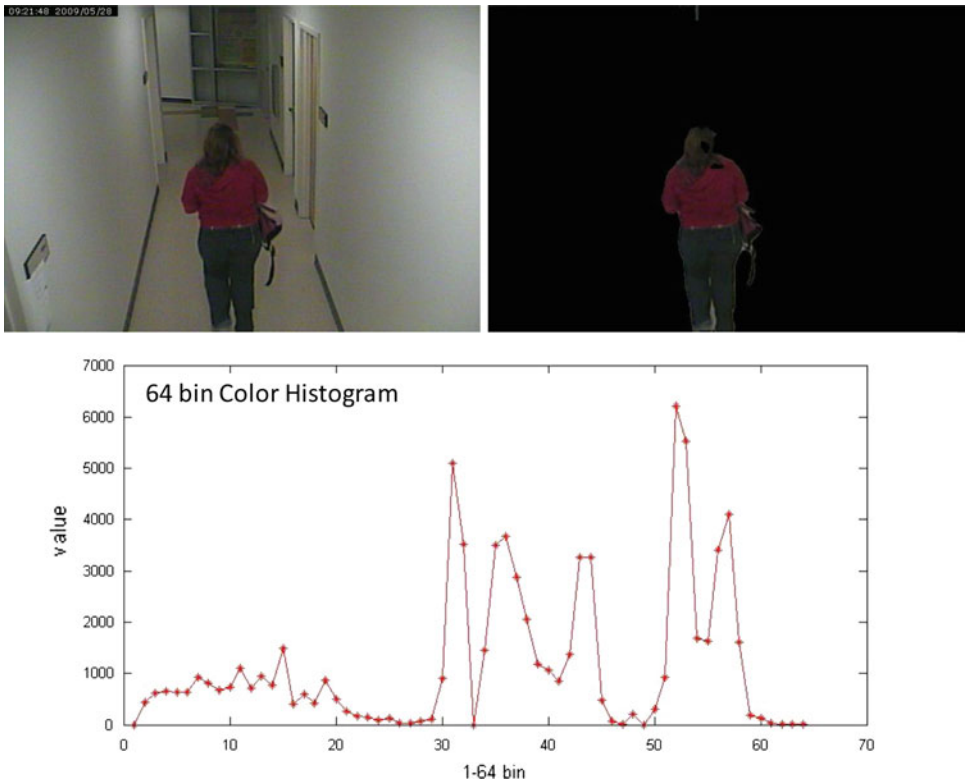

Fig. 5 Example of foreground extraction

subjects x_i and x_j, let C_i and C_j be their clothing histograms and t_i and t_j by the timestamps when x_i and x_j have been captured in video. We can choose an appropriate similarity measure to compute the similarities between them, such as the cosine similarity. For instance, if we assume that people keep the same clothing during the same day, we can define

$$S^C(x_i, x_j) = \begin{cases} \frac{C_i \cdot C_j}{|C_i||C_j|} & \text{if day}(t_i) = \text{day}(t_j) \\ 0 & \text{otherwise} \end{cases}$$

To compute the similarity of subjects from two shots, the algorithm selects a subject from a certain frame in a shot to represent the shot. Usually the algorithm chooses a frame towards the middle, which tends to capture the profile of the person better.

3.3.3 Activity

Activities and events associated with subjects prove to be very relevant to the problem of person identification [9,40]. The trajectory and walking direction can serve as a cue indicating the identity of the individual. For example, the activity of entering an office can provide strong evidence about the identity of the subject entering the office: it is likely to be either (one of the) person(s) who works in this office, or their collaborators and friends. Furthermore, considering the time of the activity in addition to the activity itself can often provide even better disambiguation power. For example, on any given weekday, the person who enters an office first on that day is likely to be the owner of the office. In addition, by analyzing past video data the behavior routines for different people can be extracted, which later can provide clues about the identify of subjects in video. For instance, if we discover that "Bob" is accustomed to entering the coffee room to drink his coffee at about 10 a.m. each weekday, then the subject who enters the coffee room at around 10 a.m. is possibly "Bob". Therefore, subject activities can often provide additional evidence to recognize people. We now discuss how to extract and analyze certain people's activities.

Bounding Box and Centroid Extraction. To track the trajectory of a subject and obtain his activity information, we need to extract bounding box and centroid of the subject. To do that we consider three consecutive frames with the same object. We first compute the differences of the first two frames by subtraction and then compute the differences of the last two frames. By combining the two different parts, we get the location of objects. After obtaining the bounding box, we determine the centroid of subjects by averaging the points of x-axes and y-axes.

Walking Direction. The most common activity in surveillance dataset is walking. The walking direction (towards or away from the camera) is an important factor to predict the subsequent behavior of a person. The walking direction can be obtained automatically by analyzing the changes of the centroid between two consecutive frames in a shot. For example, as illustrated in Fig. 6, by determining that the centroid

Fig. 6 Example of walking direction

Fig. 7 Example for location clustering

of the subject is moving from the bottom to the top in the camera view, we can determine that this person is walking away from the camera.

Activity Detection. We focus on detecting simple regular type of behavior of people, including entering and exiting a room, walking through the corridor, standing still, and so on. These types of behavior can be determined by analyzing the bounding box of a person. For instance, for walking the algorithm focuses on the first and last frame in a shot, which we are called *entrance* and *exit* frames. By analyzing the bounding box (BB) of a subject in the entrance frame, we could predict where the subject has come from. Similarly, the exit frame could tell us where this person is headed to.

If we consider all the BBs in entrance and exit frames, we can find several locations in the camera view, where people are most likely to appear or disappear. These locations, denoted as $L = \{l_1, l_2, \ldots, l_{|L|}\}$, can be automatically computed in an unsupervised way by clustering the centroid of entrance/exit BBs. Based on this analysis, we automatically obtain the entrance and exit point in an image. Figure 7 demonstrates an example of the clustering result of the entrance and exit locations.

After computing the set of entrance and exit locations $L = \{l_1, l_2, \ldots, l_{|L|}\}$, we compute the distance between them and determine the entrance and exit points in each shot. Suppose that in a shot s_m the subject x_i walks from location l_p to l_q. Then we can denote the activity as $act_i^m : \{l_p \rightarrow l_q\}$.

Activity Similarity. For each shot the algorithm extracts the activity information by performing the aforementioned process. We assume that two subjects with similar activities have a certain possibility to describe the same person. Thus based on this assumption, we connect the potentially same subjects through the similar activities. Suppose that for two subject x_i and x_j from shot s_m and shot s_n, respectively, the algorithm extract activity information $act_i^m : \{l_a \rightarrow l_b\}$ and $act_j^n : \{l_c \rightarrow l_d\}$. We can define the activity similarity as follows:

$$S_{ij}^{act}\left(act_i^m, act_j^n\right) = \begin{cases} 1 & \text{if } l_a = l_c(l_d) \text{ and } l_b = l_d(l_c) \\ 0.5 & \text{if } l_a = l_c(l_d) \text{ or } l_b = l_d(l_c) \end{cases}$$

In this equation, activities with the exact opposite entrance/exit points are defined to be equal, for example, the subject x_i with activity $act_i : \{l_a \rightarrow l_b\}$ and the subject x_j with activity $act_j : \{l_b \rightarrow l_a\}$ are considered to share the same activity. Thus the activity similarities can be leveraged to connect the subjects which share the same/similar activities.

Person Estimation Based on Activity. The intuition is that the identity of a person can be estimated by analyzing his activities. In general, given labeled past data we can compute priors such as $\mathbb{P}(x_i = p_m | act_i)$, which correspond to the probability that the observed subject x_i is the real-world person p_m, given that the subject participates in activity act_i, such as entering/exiting a certain location. Similarly, we can compute $\mathbb{P}(x_i = p_m | act_i, t_k)$ which also considers time.

3.3.4 Person gait

Gait is also a good feature to identify a particular person, because different people's gaits are often different. For example, somebody might walk very fast or slow, somebody might walk with swinging arms or head. Thus by analyzing the characteristics of people's gaits, we might be able to better predict the identity of one subject or the sameness of two subjects.

For example, if the walking speed of two subjects differs significantly, then they might not refer to the same entity.

3.3.5 Face-derived human attributes

Face-derived human attributes that could be estimated by analyzing people faces, such as gender, age, ethnicity, facial traits, and so on, are important evidence to identify a person. By considering these attributes, many uncertainties and errors for person identification can be avoided, such as confusing a "men" with a "women", an "adult" with a "child", and so on. To obtain attribute values from a given face, we use the attribute system [25]. It contains 73 types of attributes classifiers, such as "black hair", "big nose", or "wearing eyeglasses". Thus for each subject x_i, the algorithm computes 73-D attribute vector, denoted as A_i. The attribute similarity of two subjects x_i and x_j can be measured as the cosine similarity between A_i and A_j. In addition, if the extracted attribute for x_i is significantly different from that of the real-world person p_m, then x_i is not likely to be p_m.

However, the extraction of reliable attribute values depends on the quality of video data. This limitation usually leads to the failure of attribute extraction on lower quality data.

3.3.6 People co-occurrence

To recognize the identity of a person, people that frequently co-occur/present with that person in the same frames can provide vital evidence. For example, suppose that "Bob" and "Alice" are good friends and usually walk together, then the identity of one person might imply that of the other. Thus given the labeled past video data, we can statistically analyze the people co-occurrence information, and compute the prior probability of one person in the presence of the other. Furthermore, from the co-occurrence/presence of two people in one frame we can derive that the two subjects are different people. This observation can help to differentiate subjects.

3.4 Face detection and recognition

Face detection and recognition is a direct way to identify a person. However, it does not perform well in our dataset due to several reasons. First, the surveillance cameras used are of low quality and also the resolution of each frame is not very high: 704×480. Second, people may actually walk away from cameras, in which case the cameras only capture their backs and not faces. Because of that, the best face detection algorithms we have tried could only detect faces in about 7 % of frames and recognize 1 or 2 faces for a frequently appearing person out of all of his/her images in the dataset. Although the result is not ideal, we could still leverage it for further processing. We define a function $FR(x_i, p_j)$ which reflects the result obtained by the face recognition. If x_i and p_j are the same according to face recognition, we set $FR(x_i, p_j) = 1$, and otherwise $FR(x_i, p_j) = 0$.

3.5 Solving the person identification problem with RelDC

In the previous sections we have described how to extract contextual features including the people's clothing, face-derived attributes, gait, activities, co-occurrence, etc, and obtain the face recognition results. In this section we show how to represent the person identification problem as an entity resolution problem to be solved by our graph-based RelDC entity resolution framework.

Relationship-based data cleaning (RelDC) performs entity resolution by analyzing object features as well as inter-object relationships to improve the data quality. To analyze relationships, RelDC leverages the entity-relationship graph of the dataset. The proposed framework will utilize inherent and contextual features, as well as the relationships, to improve the quality of person identification.

Figure 8 shows an example of the person identification process that employs both the inherent and contextual features. The simple person identification task in the example is to discover whether the subject in the given frames is "Bob" or someone else. The example shows that, by using face recognition, only one face (marked in the red rectangle) can be detected and recognized to be "Bob", whereas the remaining subjects cannot be identified. On the other hand, by leveraging the context information, the identity of all the subjects can be recognized. Context information such as activity, clothing, gait, face-derived attributes can be extracted from the both the probe frames (the ones to be disambiguated) and gallery frames (the references frames where the labels/indentities are known). First, based on the time continuity, the frames are segmented into two shots, where in each shot the frames describe the same person. Thus, the four subjects in Shot 1 all refer to "Bob". For Shot 2, although no face-based features can is computed (since the person is walking with his back towards the camera), the subjects in Shot 2 can also be connected to "Bob" through contextual features. One such connection is the similar contextual features between Shot 2 and Shot 1 that we now know refers to "Bob". Another connection is the special activity of Shot 2 which illustrates that the subject is the first person entering "Bob" offices on that day. Therefore, by constructing an entity-relationship graph which considers both inherent and contextual feature, the identity of subjects in all the probe frames can be resolved.

3.5.1 Entity-relationship graph

In order to apply RelDC, the algorithm first constructs an entity-relationship graph $G = (V, E)$ to represent the given

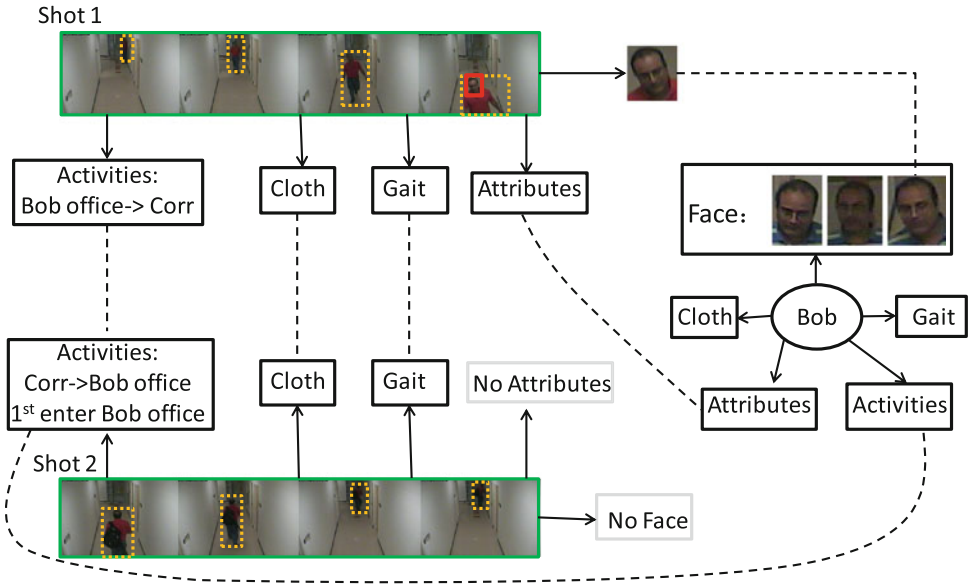

Fig. 8 Example of context-based person identification

person identification task, where V is the set of nodes and E is the set of edges. Each node corresponds to an entity and each edge to a relationship. The graph will contain several different *types* of nodes: *shot*, *subject*, *person*, *clothing*, *attribute*, *gait*, and *activity*. The edges linking these nodes correspond to the relationships. For instance, the edge between a shot node and a subject node corresponds to the "appears in" relationship.

In graph G, edges have weights where a weight is a real number in [0,1] that reflects the degree of confidence in the relationship. For example, if there is an edge with weight 0.8 between a subject node and a person node, this implies the algorithm has 80 % confidence that this subject and person are the same. The edge weight between two color histogram nodes denotes their similarity.

Figure 9 illustrates an example of an entity-relationship graph. It shows a case where the set of people of interest consists of just two persons: Alice and Bob. It considers three shots s_1, s_2, s_3, where s_1 captures two subjects x_{11} and x_{12}, shot s_2 captures x_2, and s_3 has x_3. The graph only shows the clothing and activity contextual features; the other contextual features are not shown for clarity. The goal is to match people with shots.

Subject x_{11}, x_{12}, x_2, x_3 in the graph are connected with their corresponding clothing color histograms C_{11}, C_{12}, C_2, C_3. An edge between two color histogram nodes represents the similarity between them. For instance, the similarity of C_2 and C_3 is 0.8. In addition, subjects are connected to the corresponding activities, which could be indicative of who these subjects are. For example, if the past labeled data are available, from the fact that subject s_3 is connected to activity act_3, we can get the prior probability of 0.7 that s_3 is Bob. The graph also shows that according to face recognition subject x_2 in shot s_2 is Bob.

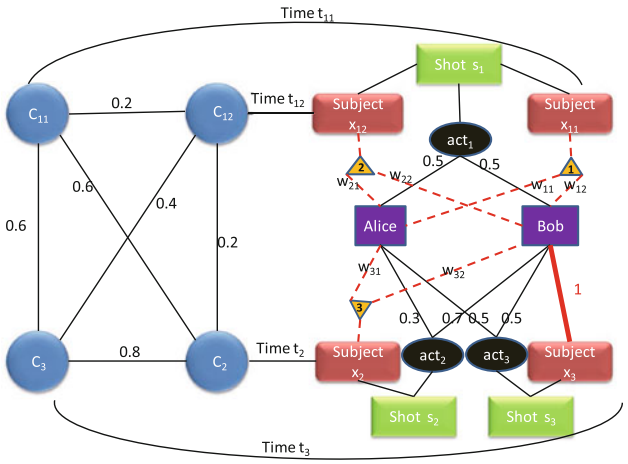

Fig. 9 Example of entity-relationship graph

The main goal is to analyze the relationships between the subject nodes and person nodes and compute the weight w_{ij} that each subject x_i associating with person p_j. Notice, weights w_{ij} are the only variables in the graph, whereas all other edge-weights are fixed constants. After constructing the graph, RelDC will compute the value of those w_{ij} weights based on the notion connection strength discussed next. After computing the weights, RelDC will use them to resolve each subject to the person that has the highest weight.

3.5.2 Connection strength computation

The constructed entity-relationship graph G illustrates the connections and linkages between subjects appearing in the video shots and real-world people. Intuitively, the more paths exist between two entities, the stronger the two entities are related. Thus we introduce the definition of connection

strength $cs(x_l, p_j)$ between each subject node x_l and person node p_j, to reflect how strongly subject x_l and person p_j are related. The value of $cs(x_l, p_j)$ can be computed according to some connection strength model. The computation process logically consists of two parts: finding the connections (paths) between the two nodes and then measuring the strength in of the discovered connections.

Generally, many different paths can exist between two nodes and considering very long paths could be inefficient. Therefore, in our approach, only important connection paths are taken into account, for instance, L-short simple paths (e.g., $L \leq 4$). For example, in Fig. 9 one 4-short simple path between subject x_2 and person "Bob" is "x_2-C_2-C_3-x_3-Bob". We will use $P_L(x_l, p_j)$ to denote the set of all the L-short simple paths between subject node x_l and person node p_j.

To measure the strength of the discovered connections, some connection strength models [21] can be leveraged. For instance, we can compute the connection strength of a path p^a as the probability of following path p^a in graph G via random walks. The connection strength $cs(x_l, p_j)$ can be computed as the sum of the connection strengths of paths in $P_L(x_l, p_j)$.

$$cs(x_l, p_j) = \sum_{p^a \in P_L(x_l, p_j)} c(p^a). \tag{1}$$

3.5.3 Weight computation

After computing the connection strength measures $cs(x_l, p_j)$ for each unresolved subject x_l and real-world person p_j, the next task is to determine the desired weight w_{lj} which should represent the confidence that subject x_l matches person p_j. RelDC computes these weights based on the Context Attraction Principle (CAP) [21] that states that if $c_{r\ell} \geq c_{rj}$ then $w_{r\ell} \geq w_{rj}$, where $c_{r\ell} = c(x_r, p_\ell)$ and $c_{rj} = c(x_r, p_j)$. In other words, the higher weight should be assigned to the better connected person. Therefore, the weights are computed based on the connection strength. In particular, RelDC sets the weight proportional to the corresponding connection strengths: $w_{rj}c_{r\ell} = w_{r\ell}c_{rj}$. Using this strategy and given that $\sum_{j=1}^{N} w_{rj} = 1$ (if each possible "option node", that is, each possible person, are listed), the weight w_{rj}, for $j = 1, 2, \ldots, N$, can be computed as follows:

$$w_{rj} = \begin{cases} \frac{c_{rj}}{\sum_{j=1}^{N} c_{rj}} & \text{if } \sum_{j=1}^{N} c_{rj} > 0; \\ \frac{1}{N} & \text{if } \sum_{j=1}^{N} c_{rj} = 0. \end{cases} \tag{2}$$

Thus, since some paths can go through edges labeled with w_{ij} weight, the desired weight w_{rj} can be defined as a function of other option weights $\mathbf{w}$: $w_{rj} = f_{rj}(\mathbf{w})$.

$$\begin{cases} w_{rj} = f_{rj}(\mathbf{w}) & \text{(for all } r, j) \\ 0 \leq w_{rj} \leq 1 & \text{(for all } r, j) \end{cases} \tag{3}$$

The goal is to solve System (3). System (3) might not have a solution as it can be over-constrained. Thus, a slack is added to it by transforming each equation $w_{rj} = f_{rj}(\mathbf{w})$ into $f_{rj}(\mathbf{w}) - \xi_{rj} \leq w_{rj} \leq f_{rj}(\mathbf{w}) + \xi_{rj}$. Here, ξ_{rj} is a slack variable that can take on any real nonnegative value. The problem transforms into solving the optimization problem, where the objective is to minimize the sum of all ξ_{rj}:

$$\begin{cases} \text{Constraints:} & \\ f_{rj}(\mathbf{w}) - \xi_{rj} \leq w_{rj} \leq f_{rj}(\mathbf{w}) + \xi_{rj} & \text{(for all } r, j) \\ 0 \leq w_{rj} \leq 1 & \text{(for all } r, j) \\ 0 \leq \xi_{rj} & \text{(for all } r, j) \\ \\ \text{Objective: Minimize } \sum_{r,j} \xi_{rj} & \end{cases} \tag{4}$$

System (4) always has a solution and it can be solved by a solver or iteratively. In our scenario, we solve this system in an iterative way [21]. The solution of this system are the values for all w_{rj} weights.

3.5.4 Interpretation procedure

The computed weight w_{rj} reflects the algorithm's confidence that subject x_r is person p_j. The next task is to decide which person to assign to x_r given the weights. The original RelDC chooses the person p_j who has the largest weight w_{rj} among $w_{r1}, w_{r2}, \ldots, w_{r|P|}$, when resolving the references of subject x_r.

The original strategy is meant for the case where each possible person p_j that x_r can refer to is known beforehand. However, in the setting of the person identification problem, this is not the case, as the algorithm is trying to decide if x_r refers to one of the known people p_j of interest or to some "other" person. To handle this new "other" category, we modify the original RelDC algorithm to also check if all of the computed weights are above a certain predefined threshold t. If they are below the threshold, this means the algorithm does not have enough evidence to resolve subject x_r, in which case it assigns x_r to "other". Otherwise, it will pick the person with the largest weight—the same way as the original algorithm.

4 Experiments and results

4.1 Experimental datasets

Our experimental dataset consists of two weeks' surveillance videos from two adjacent cameras located in the second floor of CS Department building at UC Irvine [34]. These cameras are distributed in the corners of a corridor, near the offices of the Information System Group (ISG) members. Activities of graduate students and faculty, such as entering and exiting offices, hallway conversations, walking, and so on, are

captured by these cameras. Frames are collected continuously when motion is detected with the frame rate of 1 frame a second for each camera. The resulting video shots are relatively simple, with one (or, rarely, a few) person(s) performing simple activities. The task is to map the unknown subjects into known people.

To test the performance of the proposed algorithm, we manually labeled four people from the video dataset to assign the ground truth labels. The video collected over 2 weeks contains several (over 50) individuals of which we manually labeled 4. We then have divided the dataset into 2 parts. The first week has been used as training data and the second week as test data. From the training data, we get the faces of the chosen 4 people and train a face recognizer. We also extract activities of people and compute priors based on activities.

4.2 Evaluation metrics

We have applied RelDC (in a limited form with a simplified connection strength model) to identify the four people from the testing dataset. After obtaining the weight w_{rj} for each subject x_r to person p_j, we decide which person that each subject should be assigned to using our strategy. The subject x_r can be assigned to p_j only if two requirements are satisfied: (1) w_{rj} is the largest among $w_{r1}, w_{r2}, \ldots, w_{r|P|}$, (2) $w_{rj} \geq$ threshold. If the weights of a subject for each optional person are almost equal, and none of them is larger than the threshold, then this subject will be considered as "others". By setting different thresholds, we can get different recognition results.

To evaluate the performance of the proposed method, we choose precision and recall as the evaluation metrics. By selecting a particular threshold value, each subject x_r can be assigned a label denoted as $L(x_r)$. The ground truth of identity for each subject x_r is referred as $T(x_r)$. Then as to each person p_j in the person set P, we can compute the corresponding precision and recall based on $L(x_r)$ and $T(x_r)$. Thus the total precision and recall can be obtained by averaging the precision and recall for each targeted person p_j.

$$\text{Precision} = \frac{1}{|P|}\sum_{j=1}^{|P|}\frac{|\{x_r|L(x_r) = p_j \wedge T(x_r) = p_j\}|}{|\{x_r|L(x_r) = p_j\}|} \quad (5)$$

$$\text{Recall} = \frac{1}{|P|}\sum_{j=1}^{|P|}\frac{|\{x_r|L(x_r) = p_j \wedge T(x_r) = p_j\}|}{|\{x_r|T(x_r) = p_j\}|} \quad (6)$$

4.3 Results

Figure 10 illustrates the precision-recall curve achieved by selecting different threshold values. We compare our approach with two conventional approaches.

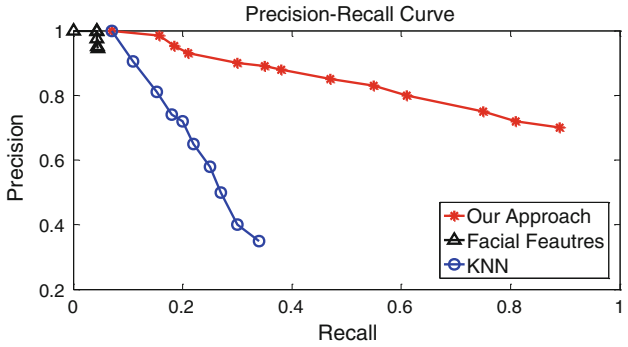

Fig. 10 Precision-recall curve

– Facial features based method. As shown in Fig. 10, if merely leveraging facial visual features, the performance of person identification is very poor. The recall is pretty low because most faces in the dataset cannot be detected due to the low quality of data, and thus the following recognition process is not able to be performed.
– K nearest neighbors method (KNN). To perform KNN, we just simply aggregate all the heterogeneous context features to obtain the overall subject similarities and then label the K nearest neighbors of the resolved subject with the same identity. By introducing context features, this method can achieve better performance than the facial feature based method. However, in this method, the underlying relationships between different context features are not considered.

The comparison with the above two approaches demonstrates the superiority of our approach. The advantages of our approach lie in that we not only leverage heterogeneous context features, but also explore the underlying relationships to integrate heterogeneous context features together to improve the recognition performance.

To test the robustness of our approach, we degrade the resolution and sampling rate of frames in our dataset respectively, and run a series of experiments on such dataset. Our algorithm mainly relies on context features such as activities, which are less sensitive to the deterioration of video quality. Figure 11 indicates that the decrease of frame resolution does not affect the performance of activity detection since the contextual information (such as time and location) is less sensitive to the frame resolution. But the performance of activity detection (suppose the performance with the original resolution and sampling rate is 100 %) drops when sampling rate reduces from 1 frame/sec to 1/2 and 1/3 frame/sec, because many important frames are lost with the decrease of sampling rate. Figure 12 illustrates that person identification result drops with the reduction of resolution and sampling rate, due to the loss of activity and color information. However, person identification result of our algorithm even

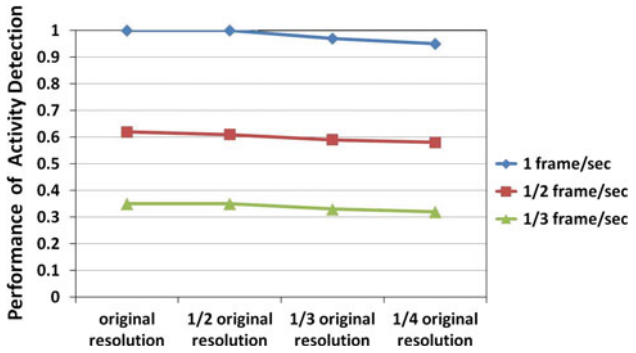

Fig. 11 Activity detection with decreasing of resolution and sampling rate

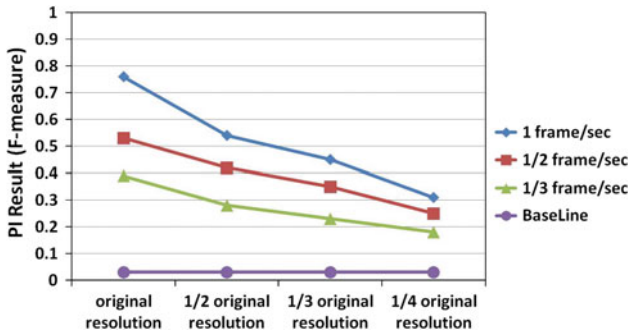

Fig. 12 PI result with decreasing of resolution and sampling rate

with the lowest resolution and sampling rate is much better than the baseline results of Naive Approach (which predicts results just based on the occurrence probability in the training dataset). Consequently, Fig. 12 demonstrates the robustness of our approach with low-quality video data, because our approach leverages contextual data rather than merely relying on the quality of video data.

5 Conclusion

In this paper we considered the task of person identification in the context of Smart Video Surveillance. We have demonstrated how an instance of indoor person identification problem (for video data) can be converted into the problem of entity resolution (which typically deals with textual data). The area of entity resolution has become very active as of recently, with many research groups proposing powerful generic algorithms and frameworks. Thus, establishing a connection between the two problems has the potential to benefit the person identification problem, which could be viewed as a specific instance of ER problem. Our experiments of using a simplified version of RelDC framework for entity resolution have demonstrated the effectiveness of our approach. This paper is, however, only a first step in exploiting ER techniques for video data cleaning tasks. Our current approach has numerous assumptions and limitations: (1) The approach assumes that color of clothing is a strong identifier for a person on a given day; if several people wear similar color clothes and have similar activities, it is hard to distinguish them using the current approach. (2) If several people appear together, it is sometimes hard for the algorithm to correctly separate these subjects, and this negatively affects the result. Our future work will explore how additional features derived from video, as well as additional semantics in the form of context and metadata (e.g., knowledge of building layout, offices, meeting times, etc.) can be used to further improve person identification.

References

1. Project sherlock @ uci. http://sherlock.ics.uci.edu
2. Ahonen, T., Hadid, A., Pietik, M.: Face description with local binary patterns: Application to face recognition. IEEE Trans. Pattern Anal. (2006)
3. An, L., Kafai, M., Bhanu, B.: Dynamic bayesian network for unconstrained face recognition in surveillance camera networks. IEEE J. Emerg. Sel. Top. Circuits Syst. **3**(2), 155–164 (2013)
4. Balcan, M., Blum, A., Choi, P.P., Lafferty, J., Pantano, B., Rwebangira, M.R., Zhu, X.: Person identification in webcam images: An application of semi-supervised learning. In: ICML Workshop on Learning from Partially Classified Training Data (2005)
5. Chaudhuri, S., Ganjam, K., Ganti, V., Kapoor, R., Narasayya, V., Vassilakis, T.: Data cleaning in Microsoft SQL Server 2005. In: ACM SIGMOD Conference (2005)
6. Chen, S., Kalashnikov, D.V., Mehrotra, S.: Adaptive graphical approach to entity resolution. In: Proceedings of ACM IEEE Joint Conference on Digital Libraries (JCDL 2007), Vancouver, British Columbia, Canada, June 17–23 (2007)
7. Chen, Z., Kalashnikov, D.V., Mehrotra, S.: Exploiting relationships for object consolidation. In: Proceedings of International ACM SIGMOD Workshop on Information Quality in Information Systems (ACM IQIS 2005), Baltimore, MD, USA, June 17 (2005)
8. Chen, Z.S., Kalashnikov, D.V., Mehrotra, S.: Exploiting context analysis for combining multiple entity resolution systems. In: Proceedings of ACM SIGMOD International Conference on Management of Data (ACM SIGMOD 2009), Providence, RI, USA, June 29–July 2 (2009)
9. Cristani, M., Bicego, M., Murino, V.: Audio-visual event recognition in surveillance video sequences. IEEE Trans. Multimed. (2007)
10. Dalal, N., Triggs, B.: Histograms of oriented gradients for human detection. In: CVPR (2005)
11. Gallagher, A., Chen, T.: Clothing cosegmentation for recognizing people. In: IEEE CVPR (2008)
12. Gao, Y., Tang, J., Hong, R., Yan, S., Dai, Q., Zhang, N., Chua, T.: Camera constraint-free view-based 3D object retrieval. IEEE Trans. Image Process. **21**(4), 2269–2281 (2012)
13. Gao, Y., Wang, M., Tao, D., Ji, R., Dai, Q.: 3D object retrieval and recognition with hypergraph analysis. IEEE Trans. Image Process. **21**(9), 4290–4303 (2012)
14. Gao, Y., Wang, M., Zha, Z., Shen, J., Li, X., Wu, X.: Visual-textual joint relevance learning for tag-based social image search. IEEE Trans. Image Process. **22**(1), 363–376 (2013)
15. Grossman, R.L., Kamath, C., Kegelmeyer, P., Kumar, V., Namburu, R.R.: Data Mining for Scientific and Engineering Applications. Kluwer, Dordrecht (2001)

16. Han, J., Altman, R.B., Kumar, V., Mannila, H., Pregibon, D.: Emerging scientific applications in data mining. Commun. ACM **45**(8), 54–58 (2002)
17. Hong, R., Tang, J., Tan, H.-K., Ngo, C.-W., Yan, S., Chua, T.-S.: Beyond search: event-driven summarization for web videos. TOMCCAP **7**(4), 35 (2011)
18. Hong, R., Wang, M., Li, G., Nie, L., Zha, Z.-J., Chua, T.-S.: Multimedia question answering. IEEE MultiMed. **19**(4), 72–78 (2012)
19. Kalashnikov, D.V.: Super-EGO: fast multi-dimensional similarity join. Int. J. Very Large Data Bases **4**(2), 561–585 (2013)
20. Kalashnikov, D.V., Chen, Z., Mehrotra, S., Nuray, R.: Web people search via connection analysis. IEEE Trans. Knowl. Data Eng. **20**(11) (2008)
21. Kalashnikov, D.V., Mehrotra, S.: Domain-independent data cleaning via analysis of entity-relationship graph. ACM Trans Database Syst. **31**(2), 716–767 (2006)
22. Kalashnikov, D.V., Mehrotra, S., Chen, S., Nuray, R., Ashish, N.: Disambiguation algorithm for people search on the web. In: Proceedings of the IEEE 23rd International Conference on Data Engineering (IEEE ICDE 2007), Istanbul, Turkey, April 16–20 2007 (short publication)
23. Kalashnikov, D.V., Mehrotra, S., Chen, Z.: Exploiting relationships for domain-independent data cleaning. In: SIAM International Conference on Data Mining (SDM 2005), Newport Beach, CA, USA, April 21–23 (2005)
24. Kantardzic, M., Zurada, J.: Next Generation of Data-Mining Applications. Wiley, London (2005)
25. Kumar, N., Berg, A.C., Belhumeur, P.N., Nayar, S.K.: Describable visual attributes for face verification and image search. In: IEEE TPAMI (2011)
26. McCallum, A., Wellner, B.: Object consolidation by graph partitioning with a conditionally-trained distance metric. In: KDD Workshop on Data Cleaning, Record Linkage and Object Consolidation (2003)
27. Nuray-Turan, R., Kalashnikov, D.V., Mehrotra, S.: Self-tuning in graph-based reference disambiguation. In: Proceedings of the 12th International Conference on Database Systems for Advanced Applications (DASFAA 2007), Springer LNCS, Bangkok, Thailand, April 9–12 (2007)
28. Nuray-Turan, R., Kalashnikov, D.V., Mehrotra, S.: Exploiting web querying for web people search. ACM Trans. Database Syst. **37**(1) (2012)
29. Nuray-Turan, R., Kalashnikov, D.V., Mehrotra, S.: Adaptive connection strength models for relationship-based entity resolution. ACM J. Data Inf. Qual. **4**(2) (2013)
30. Nuray-Turan, R., Kalashnikov, D.V., Mehrotra, S., Yu, Y.: Attribute and object selection queries on objects with probabilistic attributes. ACM Trans. Database Syst. **37**(1) (2012)
31. Tang, J., Hong, R., Yan, S., Chua, T.-S., Qi, G.-J., Jain, R.: Image annotation by knn-sparse graph-based label propagation over noisily tagged web images. ACM TIST **2**(2), 14 (2011)
32. Tang, J., Yan, S., Hong, R., Qi, G.-J., Chua, T.-S.: Inferring semantic concepts from community-contributed images and noisy tags. ACM Multimed., pp. 223–232 (2009)
33. Tian, Y., Brown, L.M.G., Hampapur, A., Senior, M.L., Shu, C.: Ibm smart surveillance system (s3): event based video surveillance system with an open and extensible framework. Mach. Vis. Appl.(2008)
34. Vaisenberg, R., Mehrotra, S., Ramanan, D.: Semartcam scheduler: emantics driven eal- ime data collection from indoor amera networks to maximize event detection. J. Real Time Image Process. **5**(4) (2010)
35. Wang, M., Hua, X.-S., Hong, R., Tang, J., Qi, G.-J., Song, Y.: Unified video annotation via multigraph learning. IEEE Trans. Circuits Syst. Video Technol. **19**(5), 733–746 (2009)
36. Wang, M., Li, H., Tao, D., Lu, K., Wu, X.: Multimodal graph-based reranking for web image search. IEEE Trans. Image Process. **21**(11), 4649–4661 (2012)
37. Yang, M., Yu, K.: Real-time clothing recognition in surveillance videos. In: ICIP (2011)
38. Zhang, L., Kalashnikov, D.V., Mehrotra, S.: A unified framework for context assisted face clustering. In: ACM International Conference on Multimedia Retrieval (ACM ICMR 2013), Dallas, Texas, USA, April 16–19 (2013)
39. Zhang, L., Vaisenberg, R., Mehrotra, S., Kalashnikov, D.V.: Video entity resolution: applying er techniques for smart video surveillance. In: IQ2S Workshop in Conjunction with IEEE PERCOM 2011 (2011)
40. Zhong, H., Shi, J., Visontai, M.: Detecting unusual activity in video. In: CVPR (2004)

Author Biographies

Liyan Zhang is currently a Ph.D candidate in Computer Science Department of the University of California, Irvine. Her research interests include information retrieval, multimedia search, computer vision and cyber physical systems. She received the BSc degree in computer science from Hebei University of Science and Technology, China, in 2006 and the MSc degree in software engineering from Tsinghua University, China, in 2009.

Dmitri V. Kalashnikov is an Associate Adjunct Professor of Computer Science at the University of California, Irvine. He received his Ph.D degree in Computer Science from Purdue University in 2003. He received his diploma in Applied Mathematics and Computer Science from Moscow State University, Russia in 1999, graduating summa cum laude. His general research interests include databases and data mining. Currently, he specializes in the areas of entity resolution and data quality and real-time situational awareness. In the past, he has also contributed to the areas of spatial, moving-object, and probabilistic databases. He has received several scholarships, awards, and honors, including an Intel Fellowship and Intel Scholarship. His work is supported by the NSF, DH&S, and DARPA.

Sharad Mehrotra is a Professor in the School of Information and Computer Science at University of California, Irvine and founding Director of the Center for Emergency Response Technologies (CERT) at UCI. Mehrotra's research interests include various aspects of data management, multimedia, and distributed systems. Mehrotra is the recipient of the SIGMOD test of time award in 2012, DASFAA test of time award in 2013, and numerous best paper awards including SIGMOD best paper award in 2004 and ICMR best paper award in 2013. Mehrotra's recent research focuses on data quality, data privacy and sensor driven situational awareness systems.

Ronen Vaisenberg received his B.A degree with distinction, president list of excellence, from the Open University of Israel in 2001. He received his M.Sc degree in Information Systems Engineering from Ben-Gurion University of the Negev in 2005 and his M.Sc in computer sciences in 2008 from the University of California, Irvine. He has graduated with a Ph.D degree from the University of California, Irvine. Under the mentorship of Professor Sharad Mehrotra, Vaisenberg's Ph.D dissertation deals with the issues related to the data management support for sentient systems, motivated by real-world emergency-response application needs. His work has been supported by an IBM Ph.D fellowship, NSF's ITR-Rescue (RESponding to Crisis and Unexpected Events) and DHS's Safire (Situational Awareness for Firefighters). He was a visiting researcher with the Event processing group at IBM-Haifa research lab, Fusion Tables team at Google Research and the Ads backend team at Google. He is currently working on building Google's Knowledge graph—a large-scale semantic representation of the world's entities. Ronen is excited about making the work a better place using information technology.