Using multiple optical cameras for correspondence identification between objects in the fields of view

Roman Kvyetnyy, Volodymyr Kotsiubynskyi, Oleksandr Kyrylenko, Oleg Kolesnytskyj, Viktoria Dumenko, et al.

Roman Kvyetnyy, Volodymyr Kotsiubynskyi, Oleksandr Kyrylenko, Oleg Kolesnytskyj, Viktoria Dumenko, Andrzej Kotyra, Dinara Mussayeva, Nazerke Abilkaiyr, "Using multiple optical cameras for correspondence identification between objects in the fields of view," Proc. SPIE 12985, Optical Fibers and Their Applications 2023, 129850A (20 December 2023); doi: 10.1117/12.3022791

SPIE.

Event: Optical Fibers and Their Applications 2023, 2023, Lublin, Poland

Using multiple optical cameras for correspondence identification between objects in the fields of view

Roman Kvyetnyy*[a], Volodymyr Kotsiubynskyi[a], Oleksandr Kyrylenko[a], Oleg Kolesnytskyj[a], Viktoria Dumenko[b], Andrzej Kotyra[c], Dinara Mussayeva[d], Nazerke Abilkaiyr[e]
[a]Vinnytsia National Technical University, 95 Khmelnytske shose, Vinnytsia, Ukraine, 21021; [b]Vinnitsa State Pedagogical University named after M. Kotsyubynsky, 32 Ostroz'koho St., Vinnytsia, Ukraine, 21000; [c]Lublin University of Technology, Nadbystrzycka 38a, Lublin, Poland, 20-618; [d]Institute of Economics CS MES RK, 29 Kurmangazy St., Almaty, Kazakhstan, 050000; [e]Al-Farabi Kazakh National University, 71 al-Farabi Ave., Almaty, Kazakhstan, 050040

ABSTRACT

The study aims to explore a method for identifying corresponding objects across multiple camera views, to improve the accuracy of object re-identification. We analyzed various techniques, including contour detection, region of interest extraction, and keypoint extraction. We also examined the challenges of finding object correspondences between multiple camera views. To evaluate the effectiveness of the proposed method, we utilized two human attribute datasets, Market-1501 and DukeMTMC-reID, and performed extensive testing on these datasets.

Keywords: deep learning, person re-identification, SIFT, keypoints, Delaunay triangulation

1. INTRODUCTION

Object re-identification is a fundamental task in automated video surveillance and has been the subject of intensive research in recent years. During this time, numerous re-identification methods have been proposed, which rely on developing specialized object image features capable of accurately characterizing each subject and creating corresponding metrics for comparing the identified features of each object[1,2].

While re-identification is intuitive for humans, as we constantly perform this task effortlessly while detecting, localizing, identifying, and subsequently re-identifying objects and individuals in the real world, it remains a challenging problem in computer vision. Re-identification assumes that a previously detected object is correctly identified in subsequent appearances[3,4].

Examples of re-identification applications include multi-camera tracking (tracking objects across multiple cameras, where the object's identification from one camera needs to be obtained based on information acquired from another camera) and trajectory tracking (if camera locations are known, using re-identification systems, it is possible to track the path of an object's movement from one point to another)[5,6].

Methods based on keypoint extraction involve determining key points in video frames based on a specific criterion, such as local extrema of intensity function. Well-known methods in this group include the Harris detector[11], KLT[20], Kitchena-Rosenfeld[12], SIFT[15], and keypoint tracking methods[24,18,19].

This study aims to investigate a method for finding correspondences between objects in video sequences captured by multiple cameras, aiming to enhance the accuracy of tracking and object re-identification.

2. MODEL EXPERIMENT

The main task of object tracking in distributed video surveillance systems operating in real-time mode is to ensure high accuracy in finding correspondences between detected objects.

* rkvetny@sprava.net

Optical Fibers and Their Applications 2023, edited by Waldemar Wójcik, Zbigniew Omiotek, Andrzej Smolarz, Proc. of SPIE Vol. 12985, 129850A · © 2023 SPIE
0277-786X · doi: 10.1117/12.3022791

Let C_k represent the network of surveillance cameras, where $k = 1, ..., p$ denotes the fields of view of cameras C_i and C_j that do not intersect. The video stream can be represented as a sequence of images.

$$V_n = [f_1, f_2, .., f_N,] \tag{1}$$

where N represents the total number of frames, and f_i represents the current frame. Any object found using the object detection method[4] in a frame can be represented as follows:

$$F_{C_{ij}} = \left\{ I_{f_{ij}}, R_{f_{ij}} \right\}, \tag{2}$$

where $I_{f_{ij}}$ represents the image of the object, $R_{f_{ij}}$ represents the position of the object image, where i is the frame number and j is the object number. R_f can be defined as a set of two points:

$$R_f = \{p_1 = (x_1; y_1), p_2 = (x_2; y_2)\}, \tag{3}$$

where p_1 represents the upper-left point of the rectangular region, p_2 represents the lower-right point of the rectangular region, and x, y are the coordinates of the points.

Let's consider the steps of the method for finding correspondences between objects in multiple non-overlapping cameras (Fig. 1(b)) using the example of two observed objects, X_1 and X_2.

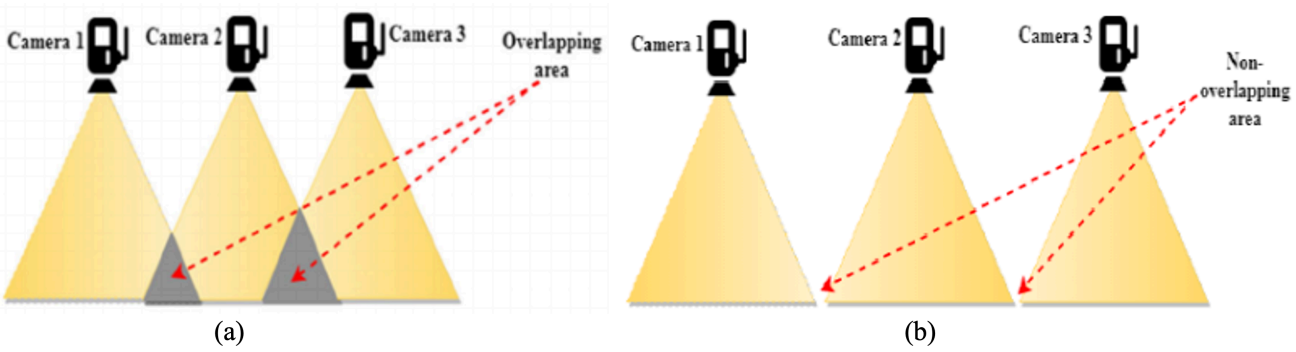

Figure 1. Example multiple cameras (a) overlapping field of view and (b) non-overlapping field of view.

The detection of key points is performed on each image from every camera. Key points are identified by using the SIFT algorithm[15]. These key points represent the unique characteristics of objects in the images. The precise determination of a key point's position is achieved through quadratic Taylor expansion of the difference of Gaussian functions in the scale space, centered at the candidate key point located at the origin. This Taylor expansion is defined by the equation:

$$D(x) = D + \frac{\partial D^T}{\partial x} x + \frac{1}{2x^T \frac{\partial^2 D^T}{\partial x^2} x}, \tag{4}$$

D and its derivative are computed at the candidate point, and $x = (x, y, \sigma)^T$ is the displacement from that point. After detecting key points, the next step is generating a descriptor for each point that describes its unique characteristics. Descriptors can be obtained by describing the properties of the key point using gradient characteristics. To compute the descriptor vector for each key point, a set of histograms of orientations is first created on a 4×4 grid of neighboring pixels, with 8 bins in each histogram. These histograms are computed from the magnitude and orientation values of the elements within a 16×16 region around the key point. The magnitudes are weighted by a Gaussian function with σ, which is equal to half the width of the descriptor window. The descriptor becomes a vector of all the histogram values, resulting in a 128-dimensional vector because there are 16 histograms with 8 bins each. This vector is then normalized to unit length to ensure invariance to affine changes in lighting. To mitigate the effects of non-linear lighting, a threshold of 0.2 is applied, and the vector is normalized again. The thresholding process can improve the matching of results even if no non-linear lighting effects are present. Figure 2 illustrates a portion of an image and the corresponding descriptor obtained[18].

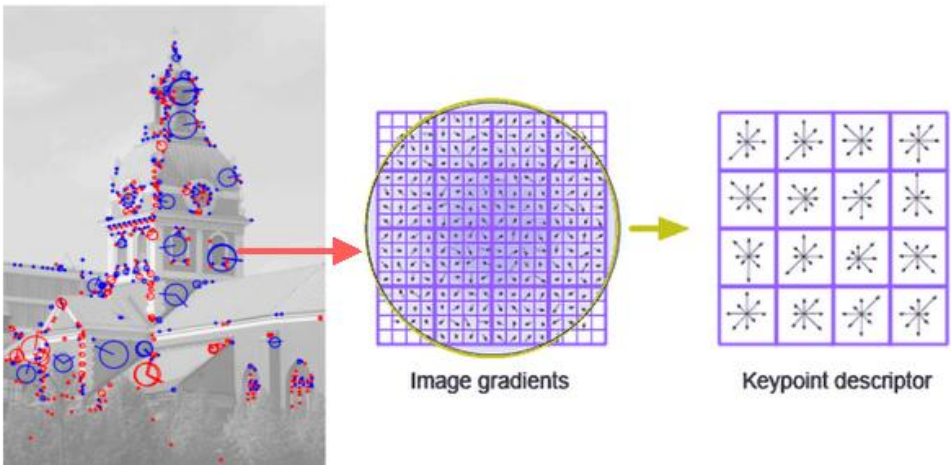

Figure 2. An example array of 4x4 descriptors obtained from a 16×16 sample.

After obtaining descriptors for key points on each image, the next step is to match descriptors between pairs of images captured by different cameras. Comparing a pair of images is primarily done using a distance-based matching method, which involves computing distances between all possible pairs of descriptors, denoted as $\rho(d_i, d'_j)$, d represents the descriptor of the first image with feature vector α_k, d' represents the descriptor of the second image with feature vector α'_k, $\forall d_i \in D, \forall d_i \in D'_j \in D', i = 1 \dots |D|, j = 1 \dots |D'|$, the dimensionality of the feature vector |K| is determined based on the specific point description method being employed[28,29,30].

To compute the distances, the Euclidean metric is commonly used:

$$\rho(d_i, d'_j) = \sqrt{\sum_{k=0}^{|K|} |\alpha_k - \alpha'_k|^2}, \tag{5}$$

The descriptors are compared using a known operation called k-nearest neighbors (k-NN) search, which involves finding the k elements that are most similar to a given query descriptor. The k-NN search computes the distance between descriptors of the input image and images in the collection, returning k pairs with the smallest distances to the classified object[31,32].

Next, for each descriptor d_i, the two nearest descriptors d'_j are selected, and vice versa. If a selected d already has two corresponding descriptors, it is skipped, and the search continues. As a result, each descriptor d_i will have at most two mutually nearest descriptors from D'. The parameter of relative length u is used for comparing two descriptors[7].

$$u = \frac{\rho_{i1}}{\rho_{i2}} (\rho_{i1} < \rho_{i2}), \tag{6}$$

where ρ_{i1} and ρ_{i2} are the distances between possible pairs of descriptors. Based on this parameter, descriptors that do not meet the required level of determinacy are filtered out. If u exceeds a given threshold u_{max}, d_i is not further considered. Otherwise, is associated with the descriptor d_j with a distance of ρ_{i1}.

After matching the descriptors and establishing correspondences between points in images from different cameras, the Delaunay triangulation method is applied. The main idea of Delaunay triangulation is to surround each object point that needs to be triangulated with triangles in such a way that they satisfy the Delaunay criterion. The Delaunay criterion requires that no point falls within the circumcircle of any triangle in the triangulation[25].

Using the known coordinates, we find the lengths of the triangle sides, and then the centers of the circles inscribed in the triangles. The next step is to determine the radius and directly construct the circle. With these values, we can construct straight contours using the formula:

$$\zeta(\beta) = \left(\rho_1 \beta + i\rho_2 \beta^2\right) e^{ia} + ih, \qquad -1 \le \beta \le 1, \rho_2 = 0 \tag{7}$$

Using the triangulation of triangles, we determine the three-dimensional coordinates of the points (Figure 3) by leveraging information about the cameras and their perspective projections.

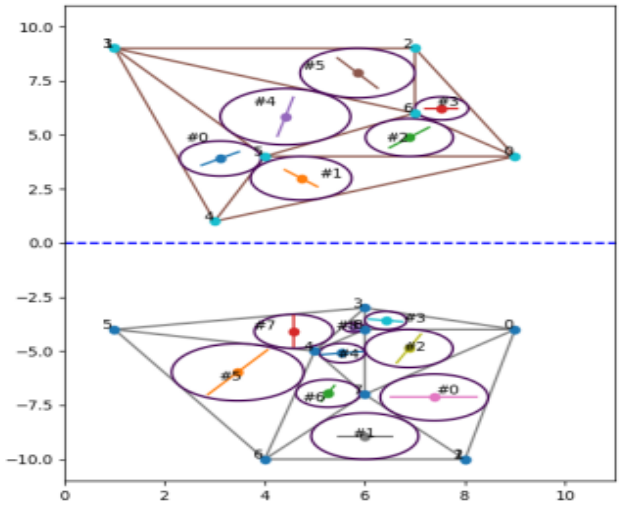

Figure 3. The application of Delaunay triangulation.

Results on the test data are shown in Figure 4.

Figure 4. The method's results on the Market-1501 test data.

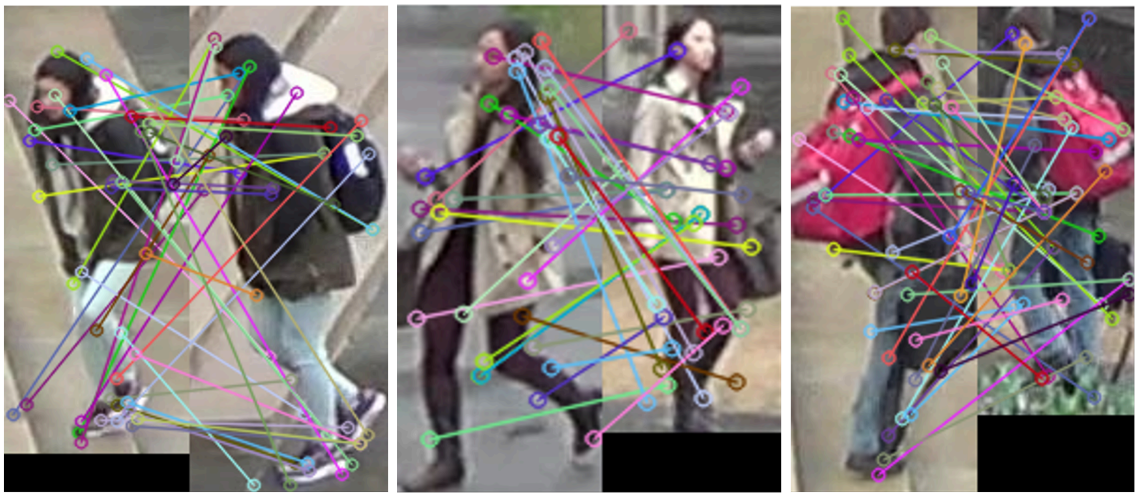

Figure 5. The method's results on the DukeMTMC-reID test data.

The DukeMTMC-reID dataset, collected at Duke University, consists of over 14 hours of video sequences captured by eight cameras. For this dataset, 702 individuals are randomly selected as the training set, while the remaining 702 individuals serve as the test set. In the test set, one query image is chosen for each individual in each camera, while the other images are placed in the gallery. As a result, we have 16,522 training images with 702 individuals, 2,228 query images of other 702 individuals, and 17,661 gallery images[8,26,27]. Examples of the method's results on the test data are shown in Figure 5.

3. CONCLUSIONS

The research aimed to investigate and analyze methods for contour extraction, region of interest detection, and keypoint extraction to address the challenge of object correspondence between multiple camera views to improve the quality of object re-identification.

The study delved into the difficulties of establishing object correspondences across different camera views. The SIFT algorithm was thoroughly examined for its effectiveness in identifying key points of an object in an image, generating descriptive features known as descriptors, and comparing these descriptors using the k-nearest neighbors technique to establish correspondences between points in images captured by different cameras. The Delaunay triangulation method was applied as well.

Two datasets focusing on human attributes, namely Market-1501 and DukeMTMC-reID, were carefully considered, and extensive testing of the method was conducted on these datasets. The practical experiments conducted on the test datasets provided compelling evidence supporting the technique's applicability in systems equipped with up to eight cameras operating at a resolution of 720×576 pixels. Moreover, the research strongly suggests the importance of further investigation into object correspondence between multiple camera views to enhance the reliability of object re-identification and validate the effectiveness of the obtained results using real-world data.

REFERENCES

[1] Holosovker, Ia. E., "Budushchee system vydeonabliudenyia: mnohokamernoe soprovozhdenye, <http://synesis.ru/blog/article/budushhee-sistem-videonablyudeniya:-mnogokamernoesoprovozhdenie> (19 October 2019).

[2] Peleshko, D. D. and Ivanov, Iu. S., "Suprovid rukhomykh ob'iektiv na osnovi obchyslennia zmishchennia yikh obmezhuvalnykh oblastei u videoposlidovnostiakh," Proc. on Intellectual Systems of Decision-making and Problem of Computational Intelligence, KhNTU, 480-482 (2014).

[3] Maslii, R., Kyrylenko, O. and Marushchak, Y., "Analysis of methods of person reidentification in multi camera environment," Norwegian Journal of Development of the International Science, 47, 46-49 (2020).

[4] Kyrylenko, O., Kvyetnyy, R. and Maslii, R., "Research of human attributes for the problem of re-identification", Information Technology and Computer Engineering, 49(3), 4-13 (2020).

[5] Zheng, L., Shen, L., Tian, L., Wang, S., Wang, J. and Tian, Q., "Scalable person re-identification: A benchmark" Proc. IEEE Int. Conf. Comput. Vis., 1116-1124 (2015).

[6] Ristani, E., Solera, F., Zou, R., Cucchiara, R. and Tomasi, C., "Performance measures and a data set for multi-target, multi-camera tracking," Proceedings on European Conference on Computer Vision – ECCV 2016 Workshops, 17-35 (2016).

[7] Chan, T. and Vese, L., "Active contours without edges," IEEE Transactions on Image Processing, 10(3), 266-277 (2001).

[8] Chan, T., Vese, L. and Sandberg, Y., "Active contours without edges for vectorvalued images," Journal of Visual Communications and Image Representation, 11(2), 130-141 (2000).

[9] Christiansen, M. and Martin, H., "Deblurring methods using antireflective boundary conditions," SIAM Journal on Scientific Computing, 30(2), 855-872 (2008).

[10] Cipolla, R., Sebastiano, B., and Giovanni, M. F. (eds.), [Computer Vision: Detection, Recognition and Reconstruction], Springer, Berlin and Heidelberg, 263-280 (2010).

[11] Kitchen, L. and Rosenfeld, A., "Gray Level Corner Detection," Pattern Recognition Letters, 1, 95-102 (1982).

[12] Kass, M., Witkin, A. and Terzopoulos, D., "Snake: Active Contour Models," International Journal of Computer Vision, 1, 321-331 (1988).

[13] Matas, O., Chum, O., Urban, M. and Pajdla, T., "Robust Wide Baseline Stereo from maximally Stable Extremal Regions," Image and Vision Computing, 22(10), 761-767 (2002).
[14] Mikolajczyk, K. and Schmid, C., "A Performance Evaluation of Local Descriptors," Computer Vision and Pattern Recognition, Proc. of IEEE Computer Society Conference, Madison, October 2005, 1615-1630 (2005).
[15] Parker, J. R., [Algorithms for Image Processing and Computer Vision], Wiley, Indianapolis, 320-412, (1996).
[16] Peleshko, D., Pelekh, Yu., Rashkevych, M., Ivanov, Yu. And Verbenko, I. "Automatic initial segmentation of speech signal based on symmetric matrix of distances," International Journal of Computer Science and Technology, 13(9), 4783-4790 (2014).
[17] Zhang, Z. et al., "Normalized direction-preserving Adam, " <https://arxiv.org/pdf/1709.04546.pdf> (2017).
[18] Bendersky, E., "The Softmax function and its derivative," <https://eli.thegreenplace.net/2016/the-softmax-function-and-its-derivative>, (2016).
[19] Zhedong, Z., Zheng, N. and Yang, Yi. "Parameter-efficient person re-identification in the 3D space," <https://arxiv.org/pdf/2006.04569>, (2020).
[20] Yu, H. and Zheng, W., "Weakly supervised discriminative feature learning with state information for person identification," Proceedings of the IEEE International Conference on Computer Vision and Pattern Recognition (CVPR), (2020).
[21] Tissainayagam, P. and Suter, D., "Object Tracking in Image Sequences Using Point Features," Pattern Recognition, 38, 105-113 (2005).
[22] Shewchuk, J. R., "Two-dimensional Delaunay triangulations," Lecture notes on Delaunay mesh generation, Berkeley, 25-43 (2012).
[23] Yang, F. and Jiang, T., "Pixton-based image segmentation with Markov random fields," IEEE Transactions on Image Processing, 12, 1552-1559, (2003).
[24] Wójcik, W., Pavlov, S., Kalimoldayev, M., [Information Technology in Medical Diagnostics II] London: Taylor & Francis Group, CRC Press, London, 306-331 (2019).
[25] Avrunin, O. G., Tymkovych, M. Y., Saed, H.F.I., et al., "Application of 3D printing technologies in building patient-specific training systems for computing planning in rhinology," Information Technology in Medical Diagnostics II - Proceedings of the International Scientific Internet Conference on Computer Graphics and Image Processing and 48th International Scientific and Practical Conference on Application of Lasers in Medicine and Biology, 7 (2019).
[26] Selivanova, K. G., Avruninz .O. G., Tymkovychz M. Y. et al., "3D visualization of human body internal structures surface during stereo-endoscopic operations using computer vision techniques," Przeglad Elektrotechniczny, 9, 30-33 (2021).
[27] Tymkovych, M., Gryshkov, O., Avrunin, O., Selivanova, K., et al., "Application of SOFA Framework for Physics-Based Simulation of Deformable Human Anatomy of Nasal Cavity", IFMBE Proceedings, 112 (2021).
[28] Vasilevskyi, O., Kulakov, P., Kompanets, D., Lysenko, O. M., et al., "A new approach to assessing the dynamic uncertainty of measuring devices," Photonics Applications in Astronomy, Communications, Industry, and High-Energy Physics Experiments, 108082E (2018).
[29] Azarova, A. O., Kilymnyk, L. A., "Mathematical model and method of risk level estimation for capital structure by means of Hopfield neural network," Actual Problems of Economics, 1(103), 245–253 (2010).
[30] Vasilevskyi, O., Voznyak, O., Didych, V., Sevastianov, V., Ruchka, O., Rykun, V., "Methods for Constructing High-precision Potentiometric Measuring Instruments of Ion Activity," Proc. IEEE 41st International Conference on Electronics and Nanotechnology (ELNANO), 247-252 (2022).
[31] Romanyuk, O. N., Pavlov S. V., et al., "A function-based approach to real-time visualization using graphics processing units", Proc. SPIE 11581, Photonics Applications in Astronomy, Communications, Industry, and High Energy Physics Experiments, 115810E (2020).
[32] Timchenko, L. I., Kokriatskaia N. I., et al., "Q-processors for real-time image processing", Proc. SPIE 11581, Photonics Applications in Astronomy, Communications, Industry, and High Energy Physics Experiments, 115810F (2020).