Crowd motion detection in video by combining CNN and integral optical flow

Huafeng Chen
College of Information Science and Technology
Zhejinag Shuren University
Hangzhou, China
eric.hf.chen@hotmail.com

Angelina Pashkevich
Faculty of Mechanics and Mathematics
Belarussion State Universyty
Minsk, Belarus
angelinapd98@gmail.com

Rykhard Bohush
Faculty of Information Technology
Polotsk State University
Novopolotsk, Belarus
r.bogush@psu.by

Sergey Ablameyko
Faculty of Mechanics and Mathematics
Belarussion State Universyty
Minsk, Belarus
ablameyko@bsu.by

***Abstract*— The paper proposes a new approach for crowd motion detection in video by combining CNN and integral optical flow. At first, definitions of crowd motion are given, along with motion parameters that can be used to perform crowd analysis. Secondly, crowd motion features and parameters are defined. Thirdly, an algorithm of crowd behavior analysis using CNN and integral optical flow is proposed. Experimental results show that, with the help of CNN, optical flow can be calculated more accurately and quickly, and by using integral optical flow, the algorithm demonstrates stronger robustness to noise and the ability to get more accurate boundaries of moving objects.**

Keywords — optical flow, CNN, crowd motion analysis, video surveillance

I. INTRODUCTION

Recently, the so-called situational analytics has been developing. Within its framework, non-standard behavior of people is analyzed in order to monitor the behavior of people in crowded places or atypical behavior of a person or groups of people. Crowd is a unique group of individual or something involves community or society. Various tasks for detecting crowd behavior can be defined such as crowd density estimation, crowd behavior identification, crowd motion, crowd tracking. A good review of crowd behavior situations is given in paper [1].

There are two main tools that are used for crowd analysis: optical flow and neural networks. Comparatively, optical flow is a standard and widely used tool for video analysis. Ali et al. [2] presented a framework in which Lagrangian Particle Dynamics is used for the segmentation of high-density crowd flows and detection of flow instabilities. Anees and Kumar [3] identified flow patterns based on stability analysis of the crowd flow using the Jacobian and Hessian matrix analysis along with corresponding eigenvalues. Lalit and Purwar [4] used a feature extraction-based model using contrast, entropy, homogeneity, and uniformity features to determine the threshold on normal and abnormal activity. Nayan et al [5] presented a method based on correlation analysis of the optical flow for crowd anomalous behavior detection. Wang et al [6] used streaklines that calculated by variational optical flow model to acquire the crowd motion trajectory information, then obtained the angular histogram and the regions of interest by calculating and clustering the dasymetric dot maps of the starting and ending points of the trajectory, finally identified specific crowd behaviors in the regions of interest by combining the dasymetric dot map and angular histogram information.

In recent years, there have been great improvements in the study of Convolutional Neural Network (CNN), thus many works used CNN for crowd behavior analysis. Direkoglu [7] used optical flow to generate motion information images which were used to train a convolutional neural network (CNN) for abnormal crowd event detection. Shao et al [8] we propose a method combining multitask cascading CNN (MC-CNN) and multiscale infrared optical flow (MIR-OF) to detect crowd aggregating and crowd escaping. Sharma et al [9] presented a crowd behavior analysis method using motion map and energy level distribution based features to detect the abnormal crowd behavior. Xu et al [10] presented a dual-channel convolutional neural network (DCCNN) for automatic and online crowd anomaly detection. A good survey on crowd behavior analysis using CNN is presented by Tripathi et al [11].

Crowd behavior is determined by the context of the scene, application background and other specific circumstances. Crowd behavior is difficult to have a clear definition and boundary, thus the detection and analysis of abnormal behavior is very challenging. In the process of video surveillance, the abnormal events that need to be paid attention to in the scene generally occupy only a small part of the whole surveillance process. Therefore, in addition to the definition of a specific event, how to define the universal abnormal behavior to make the system have greater adaptability is a very worthy of study and has greater application value. In our paper [12], we used optical flow to detect abnormal behavior of people and defined several crowd anomalies and showed how to detect them.

In this paper, we propose a formalization of the problem of crowd motion detection and identification in video sequences. We defined all main types of crowd behavior and introduced main parameters of this behavior. We detected a crowd behavior in real word videos by using CNN and optical flow approaches. We showed advantages of both approaches and show what crowd parameters can be defined.

II. CROWD DETECTION IN VIDEO: MAIN NOTIONS

A. *Crowd types*

A video sequence or video stream is a sequence of digital images (frames) $V = \{F_k\}$, k - the number of images in the sequence. The object in the image (Ob) is a local area that differs from the surrounding background and displays some of the features of the real-world object.

Crowd is a large group of people, sometimes with severe occlusions. Individual persons in a crowd are considered "connected", i.e., they are close to each other, thus a crowd can be seen as a connected component. A crowd can be classified as two types, a static crowd, or a dynamic crowd. A static crowd stay at place in some period, whereas a dynamic crowd keep moving.

There may exist several crowds in one frame, where they are separated:

$$CR_k = \left\{Cr_k^{[c]}\right\}, c = 1,\ldots,N^c, \qquad (1)$$

where $Cr_k^{[c]} = \left\{Sg_k^{[c,s]}\right\}, s = 1,\ldots,N^s$, N^s is the number of sub-groups that compose crowd $Cr_k^{[c]}$, and $Sg_k^{[c,s]} = \left\{Pe_k^{[c,s,p]}\right\}, p = 1,\ldots,N^p$, N^p is the number of people that compose sub-group $Sg_k^{[c,s]}$. One thing deserves to be mentioned is that crowd does not keep its composition through time, as mentioned in above examples, it can split, or join together with other crowds.

Whether it is a crowd, a sub-group, or a person, it can be assigned to two main classes:

a stationary object is described by a set of features $\left(Ft^{[idx]}\right)$ and its coordinates $\left(x^{[idx]}, y^{[idx]}\right)$, which do not change during a time interval (t). Such an object can be represented by a formal model:

$$So^{[idx]} = \left(Ft^{[idx]}, x^{[idx]}, y^{[idx]}, Ns_k^{[idx]}\right), \qquad (1)$$

where $\left(Ft^{[idx]}, x^{[idx]}, y^{[idx]}\right) = const \forall F_k, k \in t$, $Ns_k^{[idx]}$ - the set of possible noise effects on the object.

a moving object is characterized by a change in one or more basic parameters: shape, size, and coordinates over a time interval (t). The transformation of the shape and/or size of an object leads to a change in its features in the frames $Ft_k^{[idx]}$. Such an object can be represented by a formal model:

$$Mo^{[idx]} = \left(Ft_k^{[idx]}, x_k^{[idx]}, y_k^{[idx]}, Ns_k^{[idx]}\right), \qquad (2)$$

where $x_k^{[idx]}, y_k^{[idx]}$ - object coordinates; $Ft^{[idx]}$- a set of features of moving object, $Ft_k^{[idx]} \subseteq Ft^{[idx]}, \forall k \in t$. Then $Ft_k^{[idx]} \cap Ft_{k+i}^{[idx]}$, that is, for the same moving object on a sequence of frames, a change in its features is characteristic.

B. *Crowd motion detection*

Because of severe occlusions, single person can hardly be detected or tracked. One common way is to treat a crowd or each sub-group of it as a single entity and consider imaginary particles occupy the crowd/sub-group area. Along with particles moving, crowd will reshape or regroup. It is possible to track for one crowd in a certain frame where its sub-groups go in next frames:

$$Cr_k^{[c]} = \left\{Sg_{k+i}^{[c,s]}\right\}, s = 1,\ldots,N_{k+i}^s, \qquad (4)$$

where $Sg_{k+i}^{[c,s]}$ is a sub-group of $Cr_k^{[c]}$ in F_{k+i}. It is also possible to track for one crowd in a certain frame where its sub-groups came from:

$$Cr_k^{[c]} = \left\{Sg_{k-i}^{[c,s]}\right\}, s = 1,\ldots,N_{k-i}^s, \qquad (5)$$

where $Sg_{k-i}^{[c,s]}$ is a sub-group of $Cr_c^{F_k}$ in F_{k-t}.

Once sub-groups of a crowd are located in a previous frame or a posterior frame, further analysis of the crowd can be performed to determine whether abnormal crowd behaviour happens. Note that sub-groups of different crowds in a certain frame may form one crowd in a different frame.

C. *Crowd motion parameters*

Certain parameters can help describe crowd motion:

Motion direction indicates a destination where crowd move. We can simply divide $[0,2\pi)$ into several intervals with equal length and count for each interval number of pixels whose motion direction is in that interval. Interval with most pixels shows main motion direction of crowd.

Crowd motion directionality in region r is represented as follow:

$$md_k^{itv}(r) = \frac{n}{\left|\sum_{p\in r}\left(\cos\theta(p), \sin\theta(p)\right)\right|}, \qquad (6)$$

where n is pixel number in r, $\theta(p)$ is motion direction of pixel $F_k(p)$, itv is the time interval considered, $md_k^{itv} \geq 1$.

Crowd motion symmetry could also be described by (8), the bigger md_k^{itv} is, the more symmetrically crowd move.

Motion speed of pixel $F_k(p)$ in the corresponding time period can be calculated as follow:

$$s_k^{itv}(p) = \frac{\left|IOF_k^{itv}(p)\right|}{itv}, \qquad (7)$$

where $IOF_k^{itv}(p)$ is displacement vector of pixel $F_k(p)$.

Thus, motion intensity in region r in the same time period is defined as flow:

$$mi_k^{itv}(r) = \frac{1}{n}\sum_{p\in r} s_k^{itv}(p), \qquad (8)$$

where n is pixel number in r.

III. Features and Parameters of Crowd Behavior

Four features can help detect crowd behavior: trajectory (TR), speed (SP), acceleration (AC), and time of movement (TM). Crowd movement model can generally be described as:

$$M(P_{mov}) = (Tr, SP, AC, TM). \qquad (9)$$

Behaviour of a crowd can be detected through analysing parameter changes. There are some parameters can be used to describe appearance of a crowd, and other parameters to describe movements of a crowd.

Appearance parameters.

- Size. Size is the area of territory occupied by a crowd or a sub-group. Suppose there are N^s sub-groups in a crowd, and their sizes are $s_1, s_2, \ldots, s_{N^s}$, respectively, then $S = \sum_{i=1}^{N^s} s_i$ is the size of crowd itself.

- Density. Density is the number of people per unit area for a crowd or a sub-group. If $d_1, d_2, \ldots, d_{N^s}$ are densities of total N^s sub-groups of a crowd, respectively, D is density of the crowd, then $S \cdot D = \sum_{i=1}^{N^s} s_i \cdot d_i$.

Motion parameters

- Speed. Speed is average displacement vector of particles which occupied crowd or sub-group area per unit time, e.g., one frame.

- Acceleration. Acceleration is the change in velocity of a crowd or a sub-group per unit time. It is an important parameter useful for detecting abnormal behaviour.

IV. ALGORITHM OF CROWD BEHAVIOUR ANALYSIS USING CNN AND INTEGRAL OPTICAL FLOW

The algorithm for estimating the movement of people is based on the calculation of the optical flow and motion vector maps and it is applied to video sequences obtained by stationary surveillance cameras in public places and is as follows.

At the first stage, the optical flow is calculated. The general principle of operation of CNN when calculating the optical flow between frames: (1) Extraction of pyramidal features - converting an image into a pyramid of high-level multi-level features; (2) Deformation of signs - to facilitate inference of high offset flows; (3) Cascade stream output - further refinement of the coarse stream taking place in order to further improve its accuracy; (4) Flow regulation. Integral optical flow makes it possible to reduce the influence of the background and obtain an area of intense movement [12]. Based on the integral optical flow, it is possible to define and build motion maps that allow one to describe the movements of blocks in each position jointly.

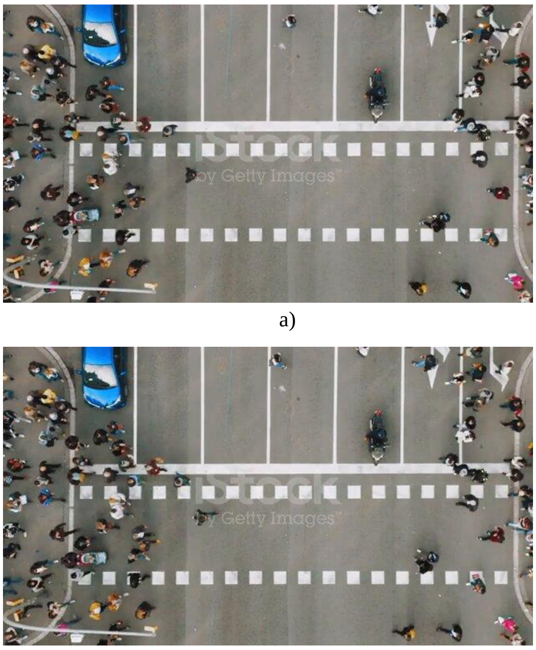
Finally, regional movement indicators are introduced to analyse movement at the level of areas, which consist of moving blocks, to analyse the movement of a group of people or a crowd.

V. EXPERIMENTAL RESULTS

To calculate the optical flow, we used the LiteFlowNet3 CNN. On Fig. 1 video frames, in which two crowds of people move in opposite directions. To form a vector field, a grid is constructed and the displacement vectors of only those pixels that are located at the nodes of the constructed grid are displayed.

As a result of the construction of the LiteFlowNet 3 optical flow and its visualization (Fig. 2), two directions of movement can be clearly distinguished, depicted in red and blue colors, respectively, as well as a sedentary, almost static predominantly light background (road, trees, almost stationary cars, etc.) with individual inclusions of various light shades of various colors. But the boundaries of objects are blurred and difficult to define clearly (Fig. 2a).

When visualizing the integral optical flow between 6 successive frames (Fig. 2b), a white background is unambiguously determined and two crowds are clearly outlined: one moves to the right and its direction is determined in red, the second moves to the left and its direction is highlighted in blue tints.



a)

b)

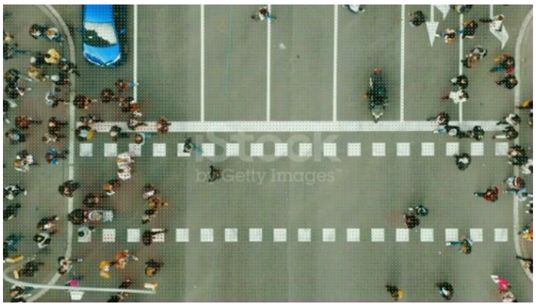
Fig. 1 Frame 1 (a) and frame 2 (b) video with transition.

a)

b)

Fig. 2 Visualization of the optical flow between frames 1 and 2 (a) and frames 1 and 6 (b)

To visualize the directions of the flow branches, a pixel density was chosen - the ratio of grid node pixels to the total number of pixels. High density makes the vectors small and covers a very large percentage of the frame (Fig.3a), whereas small density leads to greater accuracy, but potentially also cause loss of pixel groups and fluctuations (Fig. 3b).

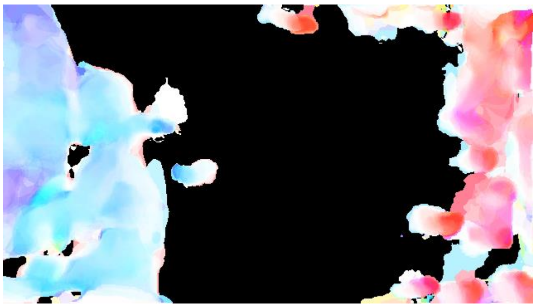


a)

b)

Fig. 3 Visualization of the integrated optical flow by vectors with grid density 25%(a) and 5%(b)

a)



b)

Fig. 4 Visualization of the ICM map (a) and OCM map (b)

With the definition of the behavior of the crowd, the direction of its movement and the ratio of the directions of movement of several crowds, subgroups, movement maps can help. The maps can also confirm the direction of movement of the crowds towards each other. The colors of the vectors on the ICM (Fig. 4a) and OCM (Fig. 4b), respectively, at one point of the ICM are opposite directions, which is also emphasized by the color: where the color is red on the ICM, blue on the OCM and vice versa.

VI. Conclusion

The conducted studies have shown that with the help of motion and flow maps it is possible to distinguish between groups of people and any objects, the directions of their movement or static, as well as the relations of groups, objects of direction: parallel (movement in one direction), opposite (movement across each other or just in opposite directions), perpendicular. Using a neural network allows you to more accurately and quickly calculate the optical flow and more accurately determine the areas of movement of groups of people. The calculation of the integral optical flow allows you to get rid of noise and at the same time get more accurate boundaries of moving objects and visualization of their movement directions.

References

[1] [1] N. A. Sjarif, S. M. Shamsuddin, S. Z. Hashim, “Detection of abnormal behaviors in crowd scene: a review”, International Journal of Advance Software Computing Applications, 4(1), 1-33 (2012)

[2] [2] Ali, S., & Shah, M., “A Lagrangian Particle Dynamics Approach for Crowd Flow Segmentation and Stability Analysis”, In Proceedings of IEEE Computer Vision and Pattern Recognition, 2007, pp.1-6.

[3] [3] V. M. Anees and G. S. Kumar, "Identification of crowd behaviour patterns using stability analysis," (in English), Journal of Intelligent & Fuzzy Systems, Article vol. 42, no. 4, pp. 2829-2843, 2022, doi: 10.3233/jifs-200667.

[4] [4] R. Lalit and R. K. Purwar, "Crowd Abnormality Detection Using Optical Flow and GLCM-Based Texture Features," (in English), Journal of Information Technology Research, Article vol. 15, no. 1, p. 15, 2022, doi: 10.4018/jitr.2022010110.

[5] [5] N. Nayan, S. S. Sahu, and S. Kumar, "Detecting anomalous crowd behavior using correlation analysis of optical flow," (in English), Signal Image and Video Processing, Article vol. 13, no. 6, pp. 1233-1241, Sep 2019, doi: 10.1007/s11760-019-01474-9.

[6] [6] X. F. Wang, Z. S. He, R. Sun, L. You, J. Hu, and J. Zhang, "A Crowd Behavior Identification Method Combining the Streakline With the High-Accurate Variational Optical Flow Model," (in English), Ieee Access, Article vol. 7, pp. 114572-114581, 2019, doi: 10.1109/access.2019.2929200.

[7] [7] C. Direkoglu, "Abnormal Crowd Behavior Detection Using Motion Information Images and Convolutional Neural Networks," (in English), Ieee Access, Article vol. 8, pp. 80408-80416, 2020, doi: 10.1109/access.2020.2990355.

[8] [8] Y. H. Shao, W. F. Li, H. Y. Chu, Z. Y. Chang, X. Q. Zhang, and H. Y. Zhan, "A Multitask Cascading CNN with MultiScale Infrared Optical Flow Feature Fusion-Based Abnormal Crowd Behavior Monitoring UAV dagger," (in English), Sensors, Article vol. 20, no. 19, p. 17, Oct 2020, Art no. 5550, doi: 10.3390/s20195550.

[9] [9] V. Sharma, R. N. Mir, and C. Singh, "Scale-aware CNN for crowd density estimation and crowd behavior analysis," (in English), Computers & Electrical Engineering, Article vol. 106, p. 13, Mar 2023, Art no. 108569, doi: 10.1016/j.compeleceng.2022.108569.

[10] [10] Y. P. Xu, L. Lu, Z. J. Xu, J. He, J. L. Zhou, and C. L. Zhang, "Dual-channel CNN for efficient abnormal behavior identification through crowd feature engineering," (in English), Machine Vision and Applications, Article vol. 30, no. 5, pp. 945-958, Jul 2019, doi: 10.1007/s00138-018-0971-6.

[11] [11] G. Tripathi, K. Singh, and D. K. Vishwakarma, "Convolutional neural networks for crowd behaviour analysis: a survey," (in English), Visual Computer, Article vol. 35, no. 5, pp. 753-776, May 2019, doi: 10.1007/s00371-018-1499-5.

[12] [12] H. Chen, O. Nedzvedz, S. Ye, S. Ablameyko. Crowd Abnormal Behaviour Identification Based on Integral Optical Flow in Video Surveillance Systems. Informatica, 2018, Vol. 29 (2), p. 211-232.