

VECTOR FIELD ANALYSIS FOR MOTION PATTERN IDENTIFICATION IN VIDEO

Nandita M. Nayak, Ahmed Tashrif Kamal, Amit K. Roy-Chowdhury

University of California, Riverside
900 University Ave, Riverside, CA 92521

ABSTRACT

Identification of motion patterns in video is an important problem because it is the first step towards analysis of complex multi-person behaviors to obtain long-term interaction models. In this paper, we will present a flow based technique to identify spatio-temporal motion patterns in a multi-object video. We use the Helmholtz decomposition of optical flow and compute singular points corresponding to component fields. We will show that the optical flow can be used to identify regions which correspond to different moving entities in the video. The singular points in these regions capture the characteristics of the field around them and can be used to identify these regions. This representation would provide us with a framework to analyze activities of individual entities in the scene as well as the global interactions between them. We demonstrate our algorithm on a dataset composed of multi-object videos recorded in a realistic environment.

Index Terms— Optical flow, activity recognition, Helmholtz decomposition

1 Introduction

Complex activity analysis is an area that is actively being explored today. Most of the existing methods on activity recognition focus on the analysis of single person activity sequences. It is difficult to extend these methods to more realistic multi-object environments since the features used in these methods do not contain information about the number of objects or motion patterns in the video. Therefore, the dimension of the search space increases exponentially as the number of objects in the scene increase. This motivates us to propose a method that can distinguish motion patterns in video in addition to yielding a set of features to represent these patterns. Motion pattern analysis can be the first step towards analysis of complex multi-person behaviors to obtain long-term interaction models. We wish to first identify moving objects in the scene and then extract features which can be used to individually analyze motion of these objects. The ability to distinguish between motion patterns and analyze them separately can greatly reduce the search space in the analysis of complex videos.

Researchers have used different approaches for activity recognition in the past. Local approaches like [3][4] view an action as a collection of local features representing small spatio-temporal regions in the video. Global methods [2] infer activities from global properties of the video like histograms. Local approaches preserve the spatial information about the scene unlike global methods which analyze the scene as a whole. We suggest that in the analysis of complex videos with multiple objects, we need local features while also keeping in mind the global motion patterns in the scene. In other words, we should be able to segregate moving objects in a scene and trace their evolution over time. This would help us to infer multiple activities occurring simultaneously in the scene. A possible way of segregating objects in a scene is with a use of a tracker which will return object boundaries. However the objective of this paper is to achieve this without a tracker since tracking in a complex scene is often very difficult.

Optical flow often serves as a good approximation of the true physical motion projected onto the image plane [5]. Optical flow holds information about the spatial as well as temporal evolution of a video. This makes it a useful tool for our purpose. Researchers have explored the use of optical flow in activity recognition in recent years. The authors in [2] modeled simple activities using histograms of optical flow as features in each frame. Optical flow histograms have also been used to analyze action of a player in a soccer match [6]. Helmholtz decomposition has been used to model crowd scenes by streakline computation in [7].

We deal with the problem of identifying and tracing the evolution of moving objects in a scene. We propose an approach which utilizes optical flow to compute spatio-temporal motion patterns in a video. From the overall flow field, we identify regions which could correspond to different moving objects. We use the Helmholtz decomposition to obtain the irrotational and solenoidal component fields. The singular points of the irrotational field correspond to the sources and sinks of the overall flow. We show that these singular points can be used as local features and can be associated over time to form singular paths(SP) [8] of the video. Thereby, we represent a video sequence as set of singular paths, each labeled with an entity in the video. The novelty in this work lies in the use of optical flow in a framework that combines local and global features for motion analysis.

This work was partly supported under the DARPA VIRAT program and NSF grant IIS-0905671.

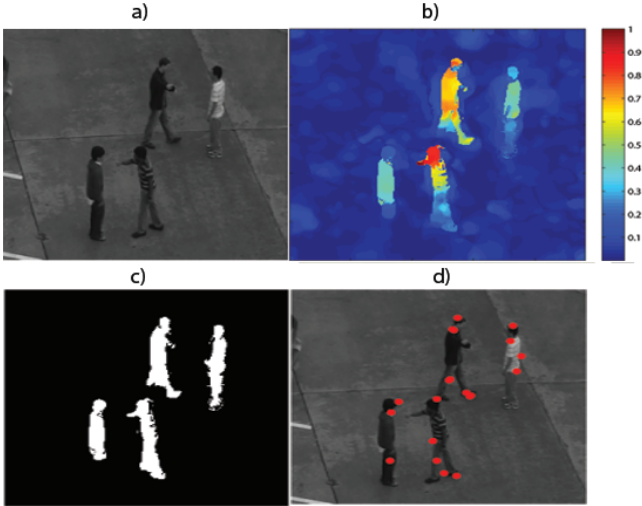


Fig. 1: The figure shows the processing of a sample frame in a video: a) A sample frame, b) Magnitude plot of optical flow, c) Extracted object regions through connected component analysis, d) Singular points of irrotational field

2 Identification of object regions from flow

The first step of our algorithm is the recognition of moving objects in the scene. This is done by an examination of the optical flow magnitude. The magnitude of optical flow is higher in moving regions of the video as compared to the static regions. The smoothness constraint enforced during the computation of optical flow also ensures that there are no sudden discontinuities. This makes it feasible to threshold the magnitude of flow so as to mark the moving regions of video into a binary image. The threshold is set experimentally. Figure 1c) shows the extracted regions of a scene with four people.

This method is found to work fairly well even in the presence of noise since we can expect the signal magnitude to be much greater than noise. Having obtained the binary image marked with pixels indicating motion, we perform a connected component analysis to identify objects in the scene. Each component corresponds to a possible object and is assigned a temporary label at this stage. We note that the appearance of a new object will increase the number of components by one and the disappearance of objects will similarly decrease the number of components. There could be errors in labeling at this stage. We will demonstrate a method to correct these errors using singular point associations as demonstrated in the following sections.

3 Helmholtz decomposition

We extract features from optical flow using the Helmholtz decomposition. The Helmholtz decomposition theorem states that any arbitrary vector field which is assumed to be differentiable can be decomposed into a curl free (irrotational) component and a divergence free (solenoidal) component[9], i.e., $\mathbf{S} = S_{sol} + S_{irr}$, where S_{sol} represents the solenoidal

component and S_{irr} represents the irrotational component.

Since S_{sol} is divergence free, we have $\nabla \cdot S_{sol} = 0$. Similarly, since S_{irr} is curl free, we have $\nabla \times S_{irr} = 0$. We can also define a scalar potential ϕ and a vector potential \mathbf{A} such that

$$\mathbf{S} = -\nabla\phi + \nabla \times \mathbf{A} \quad (1)$$

This decomposition is particularly interesting for extraction of the features and singularities of a flow. For instance, in 2D fields the curl-free term contains only sources and sinks, while the divergence free term contains only vortices [10].

3.0.1 Solution of Scalar and Vector Potentials (ϕ, \mathbf{A}):

We will follow the technique described in [9] to solve for the scalar and vector potentials. The scalar potential can be obtained by projecting onto the curl-free component and solving the following variational problem:

$$\arg \min_{\phi} \int_{\Lambda} \|\mathbf{S} + \nabla\phi\|^2 dA, \Lambda \subset \mathbb{R}^2 \quad (2)$$

where Λ is the image domain under consideration and A is the area. It can be shown that the solution to ϕ is obtained by solving the following Poisson equation [9]:

$$\nabla \cdot \mathbf{S} = \nabla^2 \phi \quad (3)$$

$$\mathbf{S} + \nabla\phi \cdot \hat{n} = 0 \text{ in } \partial\Lambda \quad (4)$$

where \hat{n} is the unit outward normal to the boundary $\partial\Lambda$. The irrotational component is then given by $\nabla\phi$.

The solenoidal component can be obtained by subtracting the irrotational component from the overall flow.

3.1 Singular Points of a Vector Field

We compute the singular/critical points of the irrotational field at each time instant. These points constitute the local features in our approach. Consider a vector field in 2D given by

$$\mathbf{v}(\mathbf{x}, \mathbf{y}) = \begin{pmatrix} u(x, y) \\ v(x, y) \end{pmatrix}$$

The Jacobian J_v of the field at (x, y) is defined as the determinant of Jacobian matrix at a point (x, y) on the field denoted by

$$J_v = \begin{pmatrix} u_x & u_y \\ v_x & v_y \end{pmatrix}$$

A point $C(x_0, y_0)$ is defined as a singular/critical point of the vector field if $C(x_0, y_0) = (0, 0)^T = 0$ and $C_1(x, y) \neq 0$ for any other point C_1 with coordinates $x \neq x_0, y \neq y_0$ in the neighborhood of (x_0, y_0) . A singular point \mathbf{C} is said to be of first order if its Jacobian $J_C \neq 0$. These points are of interest in the field as they depict the sources and sinks of the flow field. Figure 1 d) shows the singular points marked in red for a single frame in the video.

3.2 Computation of Singular Point Associations

The next stage of the algorithm involves a matching of singular points over time to trace their evolution. This is done by mapping them to the inside of a unit circle given by the (γ, r) phase plane as described in [11]. A singular point $\mathbf{C}(\mathbf{x}, \mathbf{y}) = (u(x, y), v(x, y))$ is mapped to $(\gamma(x, y)r(x, y))$ given by

$$\cos \gamma = \frac{u_x + v_y}{\text{sqrt}((u_x + v_y)^2 + (v_x - u_y)^2)} \quad (5)$$

$$\sin \gamma = \frac{v_x - u_y}{\text{sqrt}((u_x + v_y)^2 + (v_x - u_y)^2)} \quad (6)$$

$$r = \frac{1}{2} + \frac{u_x v_y - v_x u_y}{u_x^2 + u_y^2 + v_x^2 + v_y^2} \quad (7)$$

where u_x, u_y, v_x, v_y are elements of the Jacobian of the singular point \mathbf{C} denoted by J_C . The similarity measure between two singular points is given by the Euclidean distance between them in the (γ, r) plane. In order to distinguish between two points resulting from similar motions in different parts of the scene, we also impose a constraint on the distance between the two points in the image plane. The distance function we use to find the association between two points is given by

$$d(C_i, C_j) = w_1 \sqrt{r_1^2 + r_2^2 - 2r_1 r_2 \cos(\gamma_1 - \gamma_2)} + w_2 \|C_i - C_j\| \quad (8)$$

where $w_1 + w_2 = 1$ and $\|C_i - C_j\|$ is normalized to lie in the range $(0, 1)$. We set a threshold on this distance function to declare two points as matched. This threshold is set experimentally. We compare singular points of one time frame to the next to obtain spatio-temporal associations or paths representing motion patterns. These are termed as *singular paths*.

We utilize singular point associations to label objects. Temporary labels are assigned to motion regions in the first step of the algorithm as described in Section 2. However, the labeling may not be accurate. Components might merge or split as objects come close or go apart. These can be identified by an association of singular points from one frame to the next. We point out that a singular path always belongs to a single object. Thereby, by comparing the labels of singular points of the previous frame to the associated singular points in the current frame, we determine if a region belongs to a single object or multiple objects and assign permanent labels to it. Since singular points on a new object cannot be associated with points in the previous frame, the region they belong to is assigned a new label. Figure 2 shows the singular points over a set of frames. They are labeled based on the object. It can be noticed that the singular points have evolved with the motion of the object.

4 Algorithm to compute Motion Patterns in Video

1. Input Video Sequence.

For each frame:

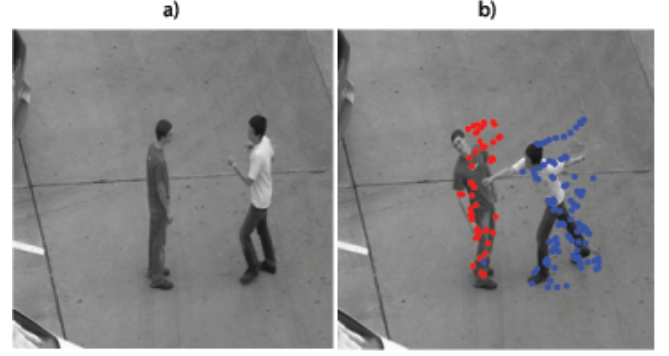


Fig. 2: The figure shows the first frame and singular points marked over an interval of frames for a video from UT Austin dataset of one person punching another. The motion of the head and hands is reflected on the evolution of these points.

2. Threshold magnitude of flow and determine connected components. Assign temporary object labels.
3. Compute Helmholtz decomposition and singular points of irrotational field.
4. For first frame, make temporary labels permanent. Singular points are clustered and labeled based on the connected component to which they belong.
5. For subsequent frames, associate singular points in current frame to those in previous frame. If
 - a. Label of associated singular points match temporary label, make label permanent.
 - b. Label of associated singular points do not match temporary label, merge or split component based on the obtained associations.
 - c. No associations are found and no marked points are found in the vicinity, mark component with new label. Make label permanent.
 - d. No associations are found and there are marked points in the vicinity, mark point with the same label as its neighbors. Make label permanent.
6. Repeat over all frames to obtain motion patterns for the video.

5 Experiments and Results

We illustrate our method by applying it to the problem of tracing the evolution of objects in an outdoor environment in the absence of a tracker. This can be an initial step of a surveillance system before analyzing activities in detail. Our dataset is composed of a 15 minute collection of videos taken from UCR videoweb dataset and UT Austin human interaction data. The UCR dataset is composed of parking lot videos with people and vehicles entering and exiting, and courtyard videos of people grouping and dispersing. The UT Austin data consists of 2-4 people performing some actions like hugging, pushing, shaking hands, etc.

The algorithm does not require any initialization. Two

Scenario	An Object enters scene	An Object exits scene	Two Objects meet	Two Objects separate
Ground Truth	25	25	18	16
Detected Objects	23	24	13	9
Percentage Accuracy	92%	96%	72.2%	56.2%

Table 1: Table showing the accuracy of the algorithm on detecting objects entering and exiting the scene, and objects coming close and going apart. The table shows the ground truth and the number detected by the method.

components merge when when two objects come close or overlap in the image. This is identified by two singular paths landing on the same component. We retain the labels of both objects on the component in such cases. A single object may split into more than one component due to a localized motion. In such cases, we find singular points with the same label in different components. We merge such components. The singular points computed for a sequence from UCR dataset is shown in Figure 3.

There is very little work on these kinds of activities (which is the novelty of this work) and this makes comparisons difficult. Therefore, we use the ground truth data to compute the accuracy of our system in the detection of events like person or vehicle entering a scene, exiting a scene, two people meeting, going apart and multiple objects moving close to each other. The accuracy of this method in detecting objects in the scene was 90.62%. Table 1 shows the detection accuracy of our algorithm in each case. We found that the method could handle a small amount of occlusion in objects. It could also recover from disconnected components created on the same object due to errors in the object identification stage.

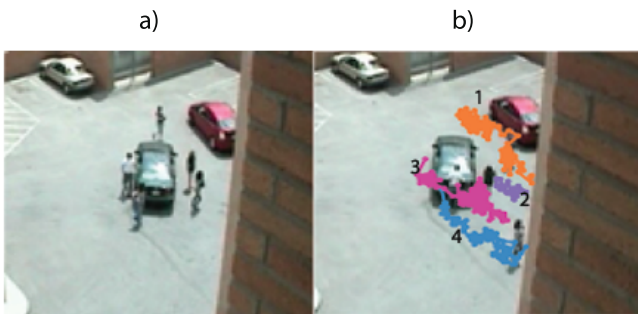


Fig. 3: The figure shows the output of an experiment to identify objects and trace them over time using our algorithm. The paths of individual objects identified in the scene are marked.

6 Conclusion

In this paper, we have illustrated an algorithm to identify motion patterns from optical flow. We have shown that the Helmholtz decomposition is a useful tool in extraction of

motion features. These motion features can be used to map objects from one time instant to the next. The overall flow can be used to separate out objects in the scene. We illustrated an application of our method in tracing the evolution of objects in a scene in the absence of a tracker. Future work would involve using this method for a more detailed analysis of activities in a scene.

References

- [1] I. Laptev and T. Lindeberg, “Space-time interest points,” in *International Conference on Computer Vision*, 2003, pp. 432–439.
- [2] R. Chaudhry, A. Ravichandran, G.D. Hager, and R. Vidal, “Histograms of oriented optical flow and binet-cauchy kernels on nonlinear dynamical systems for the recognition of human actions,” in *Computer Vision and Pattern Recognition*, 2009, pp. 1932–1939.
- [3] I. Laptev and T. Lindeberg, “Local descriptors for spatio-temporal recognition,” in *International Workshop on Spatial Coherence for Visual Motion Analysis*, 2004.
- [4] N. I. Cinbis and S. Sclaroff, “Object, scene and actions: Combining multiple features for human action recognition,” in *European Conference on Computer Vision*, 2010, pp. I: 494–507.
- [5] P. K. Turaga, R. Chellappa, V. S. Subrahmanian, and O. Udrea, “Machine recognition of human activities: A survey,” *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 18, no. 11, pp. 1473–1488, November 2008.
- [6] A. A. Efros, C. B. Alexander, E. C. Berg, G. Mori, and J. Malik, “Recognizing action at a distance,” in *International Conference of Computer Vision*, 2003, pp. 726–733.
- [7] R. Mehran, B. E. Moore, and M. Shah, “A streakline representation of flow in crowded scenes,” in *European Conference on Computer Vision*, 2010, pp. III: 439–452.
- [8] C. Garth, T. Xavier, and G. Scheuermann, “Tracking of vector field singularities in unstructured 3d time-dependent datasets,” in *IEEE Visualization 2004*, 2004, pp. 329–336.
- [9] P. Ghosh, L. Bertelli, B. Sumengen, and B. S. Manjunath, “A non-conservative flow field for robust variational image segmentation,” *IEEE Transactions on Image Processing*, vol. 19, no. 2, pp. 478–490, February 2010.
- [10] Y. Tong, S. Lombeyda, A. N. Hirani, and M. Desbrun, “Discrete multiscale vector field decomposition,” *ACM Transaction on Graphics*, vol. 22, no. 3, pp. 445–452, 2003.
- [11] H. Theisel and T. Weinkauff, “Vector field metrics based on distance measures of first order critical points,” in *International Conferences in Central Europe on Computer Graphics, Visualization and Computer Vision*, 2002, pp. SH–121.