Detecting Motion Patterns in Dynamic Crowd Scenes

Chongjing Wang, Xu Zhao*, Yi Zou, Yuncai Liu
Department of Automation and Key Laboratory of China MOE for System Control and Information Processing
Shanghai Jiao Tong University
Shanghai, China
{*ccjj,zhaoxu,jbyiiiii,whomliu*}*@sjtu.edu.cn*

***Abstract*—Detecting motion pattern in dynamic crowd scenes is a challenging problem in computer vision field. In this paper, we propose a novel approach to detect the motion patterns from global perspective. To extract the discriminative spatial-temporal features, we introduce the Motion History Image (MHI) into the optical flow algorithm. Motion patterns are then detected by automatic clustering of optical flow vectors through hierarchical clustering. Experiment evaluation on some challenging videos shows reliable detection results and demonstrates the effectiveness of our proposed approach.**

***Keywords*-crowd ; motion pattern; spatial-temporal feature; cluster**

I. Introduction

Visually detecting motion pattern of moving objects is an important research topic in computer vision field. Especially, crowd scene dynamic analysis is attracting more attention recently [7]–[10]. As shown in Fig.1, a crowd scene may include hundreds even thousands of objects. Crowded scenes commonly appear in where of our daily lives, such as subway station, supermarket, large gathering and so forth.Therefore, capturing crowd dynamic is very important to the public security and emergency management as many serious accidents had occurred in crowded scenes. However, there is still a large gap between the real application and current techniques of visual surveillance on the crowd level, due to the complexity and diversity of the crowd scenarios.

A large amount of methods for crowd analysis have been proposed recently. Zhao and Nevatia [1] perform people tracking in crowd scenes by modeling the human shape and appearance as articulated ellipsoids and color histograms respectively. This approach is one of the first algorithms for tracking in crowded scenes. N. Vaswani and R. Chellappa [2] analyze the motion of all the moving objects by learning the temporal deformation through connecting the locations of the objects in consequent frames. X. Wang and Grimson [3] propose a similarity measure for trajectories cluster, and then learn the scene models from the cluster. Z.Khan and F. Dellaert [4] use Markov chain Monte Carlo based particle filter to handle the interactions between targets in a crowded scene. They proposed a notion that the behaviors of objects are influenced by the neighborhood of the objects. All the methods proposed above require that the scenes where moving objects are not densely and the tracking results of objects are available. However, these traditional methods seem to fail in dealing with high density crowd scenes.

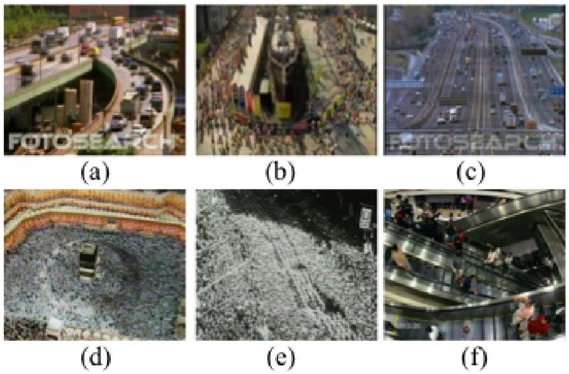

Figure 1. Some samples of crowd scenarios.

The research into crowd scenes is challenging due to many difficulties. 1) The features which can be obtained from a single object reduce drastically because its size and resolution in image is too small to recognize clearly. 2) It's hard to track single objects due to severe occlusions and similar appearance. 3) The motion behaviors of single objects are complicated because of the interaction between objects. To overcome these difficulties, researchers are studying new methods according to specific properties of crowd scenes. Ali and Shah [5] segment coherent crowd flows in video segmentation by using a mathematically exact framework based on Lagrangian Particle Dynamics. This method is incapability to segment incoherent motions. Ali and Shah [6] also track persons in high density of crowd scenes by analyzing floor fields of the scene. This method is suitable for structured scenes, heavily dependents on the physical properties of the scene and does not apply for unstructured scene. Kratz and Nishino [7] propose a novel probabilistic method by extracting the local spatial-temporal features to exploit the inherent varying structured motion pattern. This method is ineffective for significant appearance variations and severe occlusion.X. Wang and Grimson [8] propose a

978-0-7695-4541-7/11 $26.00 © 2011 IEEE
DOI 10.1109/ICIG.2011.114

unsupervised learning framework with hierarchical Bayesian models to model activities and interactions in crowd traffic scenes and train station scenes.This approach avoids tracking single object in scenes, employing only local motion as features. Saleemi and Shah [9] propose to learn dense pixel to pixel transition distributions using tracking trajectories. It is used to detect abnormal events and segment motion foreground from background. This kind of approaches rely on high quality tracking trajectories; however, they are usually very hard to abtain.

Although the information of a single object is difficult to obtain from visual observation, but on the contrary, the objects in a group show similar and salient characteristics. Based on this observation, in this paper, we propose a novel approach to detect motion pattern from global perspective. We first introduce the MHI into the optical flow algorithm to extract the spatial-temporal features. Then, we employ a hierarchical automatic clustering method based on graph link analysis techniques to detect motion pattern in crowd scenes. Ignoring the behaviors and properties of individuals, our approach considers the whole characteristics of the crowd. The experimental evaluations on diverse crowd scenarios demonstrate that our proposed method works well in detecting motion pattern.

II. Motion Pattern Representation

According to the gestalt theory of human visual perception, the main factors used in grouping are proximity, similarity, closure, simplicity and common fate (elements with same moving direction are seen as a unit) [10]. In terms of this definition, 'Motion pattern'in our research means the spatial-temporal segmentation in a video. Within the segmentation, the local speed is similar and the direction of motion is proximal or changed smoothly. We employ the Motion History Image (MHI) for the first time in the research into crowded scenes. Considering MHI as temporal features and optical flow as the spatial features, we combine both features to representation the motion patterns.

A. Motion History Image

Motion History Image is a real-time motion template that temporally layers consecutive image differences into a static image template [11], [12]. MHI is calculated as a scalar-valued image, the pixel intensity is a function of the temporal motion history at that point, where more recently moving pixels have brighter values. Motion history images reflect more object moving temporal information over a video sequence robustly and efficiently.

A MHI is calculated by adding simple replacement and decay term as in paper [11]. At the time instant t and the location (x, y) , the MHI is generated as following:

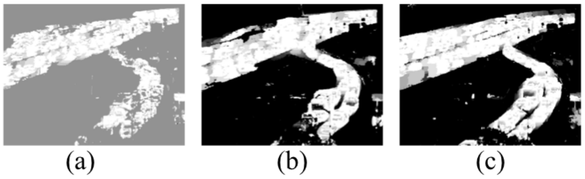

Figure 2. MHI at different time instant.

$$MHI_\tau(x,y,t) = \begin{cases} \tau & \text{if } D(x,y,t)=1 \\ \max(0, MHI_\tau(x,y,t-1)-d) & \text{otherwise} \end{cases} \tag{1}$$

where $D(x, y, t)$ is a binary image of frame difference and τ is the timestamp of motion. The MHI at time t is determined by the previous MHI and current motion image. The constant motion generated by moving objects is highlighted while the old object motion fades away due to the decay term. MHI can encode a wide range of movement, and represents how or where the image is moving. Fig.2 describe the MHI of the video in Fig.1 (a) at time $t-10, t, t+10$. From Fig.2 (a) - (c), the persistent traffic motion is highlighted with the increase of time.

B. Optical Flow Estimation

The field of optical flow estimation is making steady progress as evidenced by the increasing accuracy of current methods on the Middlebury optical flow benchmark [13]. D. Sun and M. J. Black [14] develop a method which ranks at the top of the Middlebury benchmark.

The classical method of optical flow in spatially discrete describes as following:

$$\begin{aligned} E(u,v) = \sum_{i,j} \{ & \rho_D(I_1(i,j) - I_2(i+u_{i,j}, j+v_{i,j})) \\ & + \lambda[\rho_S(u_{i,j} - u_{i+1,j}) + \rho_S(u_{i,j} - u_{i,j+1}) \\ & + \rho_S(v_{i,j} - v_{i+1,j}) + \rho_S(v_{i,j} - v_{i,j+1})]\} \end{aligned} \tag{2}$$

Where $I(i, j)$ is the pixel in the image, u and v are the horizontal and vertical speed in optical flow field from image I_1 and I_2, λ is scale factor. ρ_D and ρ_S are the penalty functions of data space and spatial field respectively.

Base on the classical method, modern optimization and implementation are incorporated in the flow fields to im-

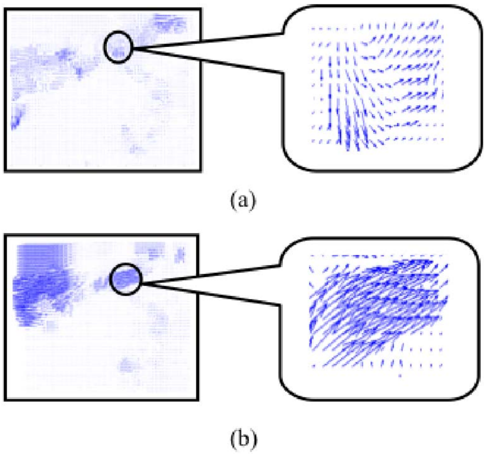

Figure 3. Different methods of optical flow. (a) Traditional optical flow vectors. (b) Optical flow vectors proposed by D. Sun.

prove the accuracy.

$$\begin{aligned} E_A(u,v,\hat{u},\hat{v}) = \sum_{i,j}\{&\rho_D(I_1(i,j) - I_2(i+u_{i,j}, j+v_{i,j})) \\ &+ \lambda[\rho_S(u_{i,j} - u_{i+1,j}) + \rho_S(u_{i,j} - u_{i,j+1}) \\ &+ \rho_S(v_{i,j} - v_{i+1,j}) + \rho_S(v_{i,j} - v_{i,j+1})]\} \\ &+ \lambda_2(\|u-\hat{u}\|^2 + \|v-\hat{v}\|^2) \\ &+ \sum_{i,j} \sum_{(i',j' \in N_{i,j})} \lambda_3(|\hat{u}_{i,j} - \hat{u}_{i',j'}| + |\hat{v}_{i,j} - \hat{v}_{i',j'}|) \end{aligned} \tag{3}$$

where $\hat{u}$ and $\hat{v}$ indicate an auxiliary flow field and $N_{i,j}$ is the neighbors of pixel (i,j), λ_2 and λ_3 denote the scalar weight.

The performance of the optical flow in a large neighborhood is improved by incorporating median filtering of intermediate flow field and alternating optimization at every pyramid level. But in the corner or the narrow structure, it result in oversmoothing phenomenon, ignoring the details. To avoid the fault, a adaptive weight is introduced into the non-local term. The weight is defined according to spatial distance, color-value distance, and occlusion sate.

The method described above is employed for our research. Contrast to the tradition method of optical flow, the accuracy has been highly improved.

As shown in Fig.3, the optical flow vectors calculated by traditional method present multiple and scattered directions. However, the directions of optical flow vectors indicated by D. Sun are uniform, the performance is highly improved.

C. Optical Flow Estimation Based on MHI

In crowd scenes, as the small size and interaction between individuals, optical flow algorithm is not robust, much noise is produced to reduce the accuracy. We first propose the method of optical flow estimation based on MHI, which has reduced the noise efficiently. The global spatial-temporal features of the crowd are extracted to describe the motion pattern.

MHI represents the temporal information of the moving objects, and optical flow vectors reflect the spatial information of the movement. The combination between them reflects the spatial-temporal character. The optical flow vectors are computed based on MHI, not based on original images. The method reduce the noise and help to improve the performance of the following detecting algorithm.

In Fig.4 (a), much noise is generated from the original images because of the complex scene. And optical flow vectors based on MHI in Fig.4 (b) reflect the information of the movement, not dependent on the complex scenes.

III. Motion Pattern Detection

As the definition of 'motion pattern'we have given, the local speed is similar and the direction of motion in a motion is proximal or changed smoothly. Optical flow vectors calculated by the method in section II-C are four-dimensional vectors including the location and velocity. Detecting motion pattern means to cluster the optical flow vectors.

A. Hierarchical Clustering Method

Minsu Cho and Kyoung Mu Lee [15] propose a novel hierarchical clustering method base on graph link analysis techniques. The method computes the clusters automatically without pre-defined numbers of clusters, and it is available for large amount of data. As the data of optical flow generated by the algorithm above is very large, and in

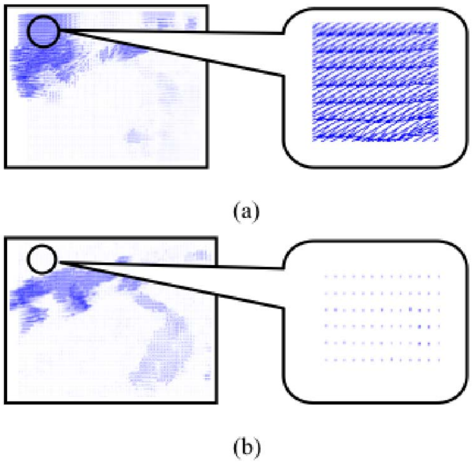

Figure 4. Noise in optical flow vectors. (a) Optical flow vectors generated from the original image. (b) Optical flow vectors based on MHI.

addition the number of motion patterns is unknown, it is difficult to detect the motions of crowd movement. So the scheme proposed by Minsu Cho is employed to solve our problem.

Algorithm 1 Motion Pattern Detection

```
Input: Optical flow vectors based on MHI
Output: Clusters of motion patterns
Compute the distance matrix $D$
according to Equation (4)
Compute the weight matrix $w$
Construct PPR matrix R
Initialize the layers $l$
loop
    $l \leftarrow l + 1$
    Initialize the orders $n$
    loop
        $n \leftarrow n + 1$
        Compute authority nodes
        Determine the clusters by authority
        node traversals
        Propagate PPRs
        if any authority node is shifted then
            Return
        end if
    end loop
    if  all nodes lie in a single cluster then
        Return
    end if
end loop
```

The measure of good clustering is to satisfy the condition of intra-cluster similarity and inter-cluster dissimilarity. The procedure of clustering is seen as the authority seeking on graph, which traversing on each node iteratively by transition matrix until searching high authority scores.

A relational graph $G = (v, \xi, w)$ with nodes v , edges ξ and weights w are given based on a pairwise relation between two components. PageRank vectors $PPR(i)$ are generated, which make a probabilistic landscape of the authority score around node i. The hierarchical scheme is performed to recursively aggregate nodes in each cluster. The authority-shift procedure is performed iteratively for each PPR propagation until the n_{th} order PPR converges.

B. Similarity Measure

Similarity measure is the critical section in clustering. In the method above, weight $\omega \in w$ denotes the similarity or proximity between node i to node j. Euclidian distance function between the point i and point j is the weight function to cluster the point synthetic set in paper [15]. RGB distance function between the pixel i and pixel j is the weight function to segment the image in paper [15]. However, the Euclidian distance or RGB distance are not available for detecting crowded motion pattern. The similarity measure designed in our research should be related to location and velocity. The distance described in paper [10] is applied to our research and the weight matrix W we designed is:

$$W_{i,j} = \exp(\frac{-D(i,j)^2}{\sigma^2})\,\text{for}i \neq j\,\text{and}\,W_{i,j} = 0.$$

Optical flow vector calculated by the method in section II-C is four-dimensional vector $f = (x, y, u, v) = (P, V)$, where $P = (x, y)$ denotes the location and $V = (u, v)$ denotes the speed in horizontal and vertical direction.

The distance $D(m, n)$ between any two optical flow vectors m and n in optical flow field is defined as:

$$D(m, n) = (d_P(m, n)d_S(m, n))^2 \tag{4}$$

where $d_P(m, n)$ is the position distance $d_S(m, n)$ and is the direction distance. The vector m and n present the case of parallel or intersecting in optical flow field. The two cases are computed as following:

1) The vector m and n are parallel:

$$d_P(m, n) = \|P_m - P_n\| \tag{5}$$

$$d_D(m, n) = \left(\frac{2}{1 + \xi + \bar{V}_m\bar{V}_n}\right)^2 \tag{6}$$

Where $\xi = 10^{-6}$ and $\bar{V} = \frac{V}{\|V\|}$.

2) The vector m and n are intersecting:

$$d_P(m, n) = \|P_m + V_m - P_n\| \tag{7}$$

$$d_D(m, n) = \frac{2}{1 + \xi + \cos\theta_m} \cdot \frac{2}{1 + \xi + \cos\theta_n} \tag{8}$$

$$\cos\theta_m = \bar{V}_m \cdot \overline{P_m - P_n} \tag{9}$$

$$\cos\theta_n = \bar{V}_n \cdot \overline{P_m - P_n} \tag{10}$$

where θ denotes the angle between the direction of optical vector and horizontal direction.

The final distance is determined by the minimum of the distance in the above two cases so as to seek the maximum similarity between optical flow vectors. The algorithm of the motion pattern detection is described in Algrithm 1.

IV. EXPERIMENTS

To test our method in detecting motion patterns in dynamic crowd scenes, we conduct experiments on the video clips downloaded from internet.These videos include groups of moving people and vehicles with small size or low resolution of individuals and severe occlusions between each other.

As shown in Fig.5, three status of traffic flows which are formed by high density vehicles are detected correctly. It corresponds to the Fig.1(a). In this video,three traffic flows with high speed run in the highway.It is hard to distinguish the two traffic flows with low resolution and opposite directions while the two lanes are very close and

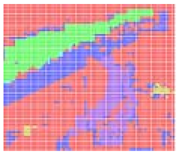

Figure 5. Three motion patterns marked by different colors.

parallel. Our method detect the three motions marked in green,blue and purple colors successfully.

More results are shown in Fig.6 ,Fig.7 and Fig.8. Fig.7(a)shows a crowd scene in supermarket where crowds of shoppers take four elevators up and down. It's a challenging scene because of severe occlusion and complicated background. The result of MHI in Fig.6(b), presents that the temporal information as the persistent moving of elevator is lightened and the static objects, such as several static people sitting around a tetoh and the shadows from the background, faint away with time. And also the optical flow estimation algorithm based on MHI in Fig.6(c) reflects the spatial information, the noise is greatly reduced. It helps to improve the performance for detecting motion patterns.The four motion patterns: two groups going up and two groups going down, are detected correctly and corresponded to colors in red,purple and brown in Fig.8(a). From the result ,since the speed and direction are similar on the middle two adjacent lifts ,the motion patterns merge together at one end of the elevator. However, it is impossible and necessary to discriminate them.

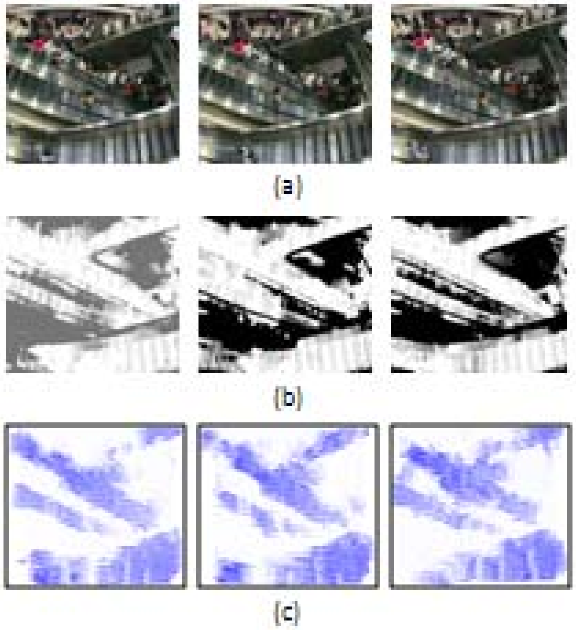

Figure 6. Results of elevator video. (a) Original images. (b) MHI. (c) Optical flow estimation based on MHI.

A marathon video in Fig.8(a) describe the crowd scene which thousands of people are running in a 'U-turn'road. In this scene, it is difficult to detect the motions duo to the great similarity , intricate interaction among the individuals in the team and the variational directions at the position of 'U-turn'road. The constant movement of the marathon team is highlighted and the noise is weaken gradually in Fig.7 (b). And then the optical flow estimation algorithm based on MHI in Fig.7 (c) reflects the available information of movement with less noise. It improve the performance tremendously of subsequent processing. The directions of individuals at the position of 'U-turn'road change smoothly, and it should belong to the same motion pattern according to our definition to 'motion pattern'. Fig.8 (b) demonstrates the strength of our method, the global movement is obtained successfully. In this crowd scene,individuals at the boundary of the team have more freedom than those at the inside of the team, so scattered directions they present at the boundary result in some small motion patterns are also detected.

V. Conclusion

In this paper, we proposed a approach to detect motion patterns in dynamic crowd scenes. By extracting the spatial-temporal features via the optical flow estimation algorithm based on MHI, the motion patterns are detected through the hierarchical clustering method. The experiments are conducted on some challenging videos which are downloaded from different web sources and demonstrate the effectiveness of our proposed approach. Our further research

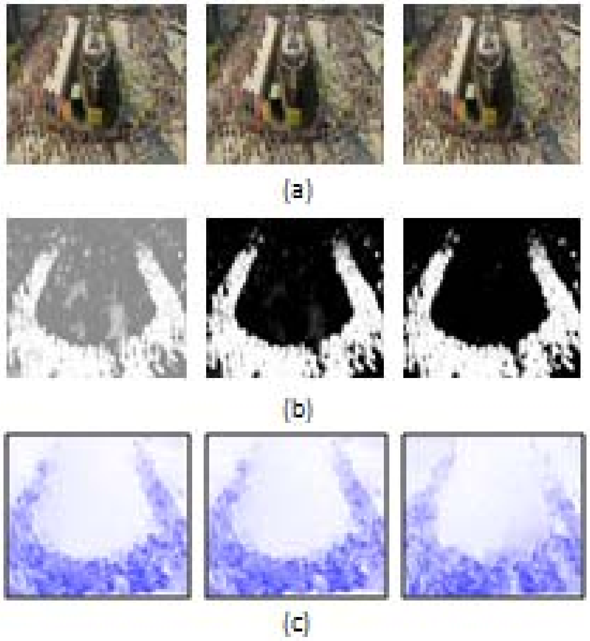

Figure 7. Results of Marathon video. (a) Original images. (b) MHI. (c) Optical flow estimation based on MHI.

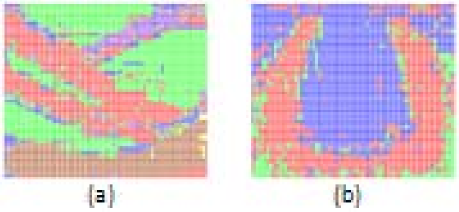

Figure 8. The motion patterns of (a) elevator video and (b) marathon video.

directions include searching for more global characters for crowd scenes to improve the performance of motion pattern detection.

Acknowledgment

This work was supported in part by National Basic Research Program of China (973 Program) 2011CB302203 and the National Science Key Foundation of China under Grant 60833009.

References

[1] T. Zhao, and R.Nevatia, "Tracking multiple humans in crowded environment", In CVPR,2004

[2] N. Vaswani, A. Chowdhury, and R.Chellappa, "Activity recognition using the dynamics of the configuration of interacting objects", In CVPR,2003

[3] X. Wang,K. Tieu, and E. Grimson, "Learning semantic scene models by trajectory analysis", In ECCV,pages110-123,2006

[4] Z. Khan,T. Balch, and F. Dellaert, "An MCMC-based particle filter for tracking multiple interacting targets", In ECCV,pages279-290,2004

[5] S. Ali, and M. Shah, "A lagrangian particle dynamics approach for crowd flow segmentation and stability analysis", In CVPR,pages1-6,2007

[6] S. Ali, and M. Shah, "Floor fields for tracking in high density crowd scenes",In ECCV,pages1-14,2008

[7] L. Kratz, and K. Nishino, "Anomaly detection in extremely crowded scenes using spatio-temporal motion pattern models",In CVPR,2009

[8] X. Wang, X. Ma,and W.E.L. Grimson, "Unsupervised activity perception in crowded and complicated scenes using hierarchical bayesian models", IEEE transactions on pattern analysis and machine intelligence, pages539-555,2009

[9] I. Saleemi, and K. Shafique,and M. Shah, "Probabilistic modeling of scene dynamics for applications in visual surveillance",IEEE transactions on pattern analysis and machine intelligence,pages1472-1485,2008

[10] M. Hu,and S. Ali, and M. Shah, "Learning motion patterns in crowded scenes using motion flow field",In ICPR,2008

[11] A.Bobick, and J.Davis, "The recognition of human movement using temporal templates", IEEE transactions on Pattern Analysis and Machine Intelligence,pages 257-267,2001

[12] J.Davis, "Hierarchical motion history images for recognizing human motion",IEEE Workshop on Detection and Recognition of Events in Video,pages 39-46,2001

[13] S. Baker, D. Scharstein, J. Lewis, S. Roth, M. Black, and R.Szeliski, "A database and evaluation methodology for optical flow",In ICCV,pages 1-8,2007

[14] D. Sun, S. Roth,and M. Black, "Secrets of optical flow estimation and their principles",In CVPR,2010

[15] M. Cho, K.MuLee, "Authority-shift clustering: Hierarchical clustering by authority seeking on graphs",In CVPR,2010