Mixed-Motion Segmentation using Helmholtz Decomposition

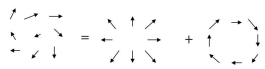
Cuicui ZHANG[1,a)] Xuefeng LIANG[1,b)] Takashi MATSUYAMA[1,c)]

1. Introduction

Motion segmentation plays a central role in video analysis, such as the surveillance, human-computer interaction, action recognition, etc. Extensive studies have been done on the stationary camera scenarios. Recently, more attentions are focusing on dynamic backgrounds with several moving objects in the scene. In many applications, the background motion is of much less interest, and solely the local object motion is expected.

Several approaches have been proposed for global motion estimation and motion segmentation (the global motion and the object motion are also named as inlier and outlier, respectively). The work in [6] introduced a parametric form which assumed the global motion model from simple translation to general perspective transformation using different parameters. A joint global motion estimation and segmentation method was proposed in [2]. It iteratively updates the inlier model by segmenting the outlier out. A regression scheme, using gradient descent (GD) [6] or least squares (LS) [5], is also applied to refine the inlier model. An outlier rejection filter in [1] explicit filters motion vectors by checking their similarity in a pre-defined window. RANSAC in [3] is a statistical method which estimates the inlier model by iteratively updating the probability of inlier. All above methods are 2D based methods. They require the multiple motions to be independent. But for interdependent motions, they may fail to deal with. Considering this, it is better to place different motions on different layers in higher dimensional space. To this end, our method transforms 2D motion field into 3D surfaces. Local extremes on the surface such as peaks, ridges and valleys depict local motions while smoothing places represent the global motion. The 3D surface is calculated using Helmoholtz decomposition. In [4], a particle filter based on Helmholtz decomposition was proposed for flow estimation. Following is its fundamental theory. For an arbitrary flow field $\vec{\xi}$, it is decomposed into two components: curl-free (divergence) component ∇E and divergence-free (curl) component $\nabla \times \vec{W}$, where E and W are what we want to obtain the 3D potential surfaces. The segmentation is performed based on surface fitting. Global motion is estimated by constructing a smooth basic surface while local motions are recovered by removing the global motion from the original motion field. Experimental results demonstrate the efficiency of our method.

$$\vec{\xi} = \nabla E + \nabla \times \vec{W}. \tag{1}$$



2. Main Theory of the Proposed Method

2.1 Potential Surface Calculation

The potential surfaces of the two components, curl and divergence, are defined as: 1. *Vector potential surface* denoted by $\vec{W}$, whose curl operation denotes the curl component $\nabla \times \vec{W}$. 2. *Scalar potential surface* denoted by E, whose gradient is the divergence component ∇E. Where, curl operation is defined as $\nabla \times \vec{W} = (\partial W_v/\partial u) - (\partial W_u/\partial v)$, and gradient is defined as $\nabla E = (\partial E_u/\partial u) + (\partial E_v/\partial v)$. Since the divergence and curl component ∇E are the projection of the original motion field V to the space of the divergence and curl field respectively, the distance between V and the projected ones should be minimal. Therefore, we apply energy minimization to calculate the potential surface:

$$\begin{aligned} D(E) &= \int_\Omega \parallel \nabla E - \vec{V} \parallel^2 d\Omega, \\ G(\vec{W}) &= \int_\Omega \parallel \nabla \times \vec{W} - \vec{V} \parallel^2 d\Omega, \end{aligned} \tag{2}$$

where Ω represents the image domain.

2.2 Global and Local Motion Estimation

Global motion is estimated from both E and $\vec{W}$. Here, we first illustrate the method on E. We assume the global motion is a smooth field. However, when local and global motions are mixed up, local motions present peaks, ridges and valleys named as outliers on E. To estimate the global motion from E, we formulate the problem as construction of a new smooth surface E', which approximates the smooth base of E gradually by applying surface fitting twice (see Eq.(3)). Basically, the first surface fitting plays a role of rejecting outliers. To avoid overfitting, a polynomial of low degree of $d_0 = 5$ is used to produce a surface E_1 in the first surface fitting. Afterwards, the distance $D = ||E - E1||$ serves to locate outliers. If D is higher than a threshold T, then the point is marked as outlier and not be considered in the second round of fitting. To have a better estimation, a polynomial of high degree of $d_1 = 10$ is employed in

1 Graduate School of Informatics, Kyoto University, Kyoto 606-8501, Japan
a) zhang@vision.kuee.kyoto-u.ac.jp
b) xliang@i.kyoto-u.ac.jp
c) tm@i.kyoto-u.ac.jp

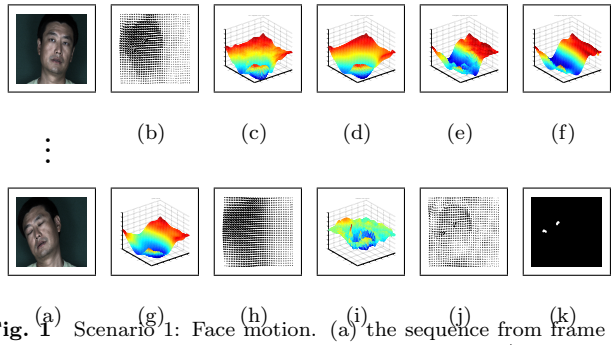


Fig. 1 Scenario 1: Face motion. (a) the sequence from frame 1 to 50. (b) motion field of one frame. (c) $\vec{W}$ with local motions. (d) $\vec{W}$ without local motions. (e) E with local motions. (f) E without local motions. (g) potential surface of estimated global motion. (h) recovered global motion field. (i) potential surface of estimated local motion. (j) recovered local motion field. (k) segmentation result.

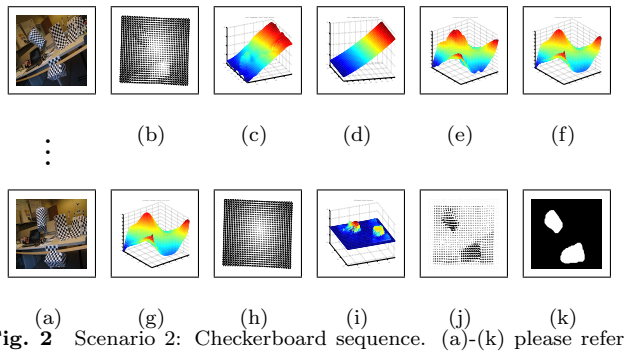


Fig. 2 Scenario 2: Checkerboard sequence. (a)-(k) please refer to the description in Fig.1.

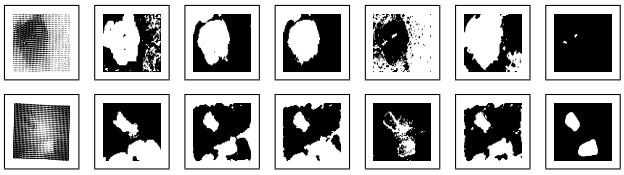


Fig. 3 Segmentation results of five reference methods and ours on two scenarios (a) motion field of one frame, (b) GME method [2], (c) LS [5], (d)GD [6], (e) Filter [1], (f) RANSAC [3], (g) our method.

the second surface fitting. The global motion field of divergence and curl component is calculated by $G_1 = \nabla E'$ and $G_2 = \nabla \times \vec{W}'$, respectively. The final global motion field is estimated by linear combination of G_1 and G_2.

$$z = a_{d0}x^d + a_{0d}y^d + \cdots + a_{ij}x^i y^j + \cdots + a_{00}. \quad (3)$$

After obtaining global motion, local motions are recovered by subtracting the global motion from the original motion field.

3. Experiments

3.1 Testing Data

Face Motion Sequence. This scenario demonstrates a challenging problem in dynamic facial expression analysis: expression is often mixed up with the head motion which is of less interest and expected to be removed. Fig.1.(a) shows a sequence in which the head is rotating, meanwhile, the eyes are blinking. In the optical flow (b), the eye motion shows a misleading direction because of fusing with the head motion. By applying our method, the head motion (h) and the eye motion (j) are recovered. (j) shows the the actual eyes pointing to the lower jaw.

Checkerboard Sequence. This is a benchmark sequence in [7]. The scene involves three motions: a rotating camera view, a rotating basket and a translating box. See Fig.2.(a). The results demonstrate the robustness of our method in this kind of complex scenarios.

3.2 Comparison with the State-of-the-art

We compared our method with five well-known methods including: (1) GME [2], (2) Least Square (LS) [5],(3) gradient descent (GD) [6], (4) Filter [1], and (5) RANSAC [3]. The segmentation results are shown in Fig.3. Where, the local motions are illustrated in white, the global motion and static background are in black. We can see our method outperforms the others obviously. We also did numerical evaluation using the segmentation error, defined by the missegmentation percentage in outliers and inlier. The comparison is performed on the second scenario (see Table 1). We can see a significant superiority of our method.

Table 1 Segmentation error of different methods

Method	GME	LS	GD	Filter	RANSAC	Our
Error	21.07	15.28	24.66	8.17	18.00	2.85

4. Conclusion

The proposed method performs well on a wide range of motions: independent/dependent, rigid/non-rigid, and single/multiple motions. By transforming 2D vector fields into 3D potential surfaces, global motion and local motions are separated onto different layers. Applying surface fitting on the potential surface, global and local motions are recovered accurately. Compared with well-known works, our method performs much better in challenging scenarios where global and local motions are mixed up and interdependent.

Acknowledgement

This work is supported by: Japan Society for the Promotion of Science, Scientific Research-KAKENHI for Grant-in-Aid for Young Scientists (ID:13276232).

References

[1] Y. M. Chen and I. V. Bajic. Motion vector outlier rejection cascade for global motion estimation. *IEEE Signal Process. Lett.*, 17(2):197–200, 2010.

[2] Y. M. Chen and I. V. Bajic. A joint approach to global motion estimation and motion segmentation from a coarsely sampled motion vector field. *IEEE Trans. on Circuits Syst. Video Technol.*, 21(9):1316–1328, 2011.

[3] M. Fischler and R. Bolles. RANSAC random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography. *Communications of the ACM*, 26:381–395, 1981.

[4] N. Kakukou, T. Ogawa, and M. Haseyama. An effective flow estimation method with particle filter based on Helmholtz decomposition theorem. *ICASSP*, pages 949–952, 2009.

[5] A. Smolic, M. Hoeynck, and J.-R. Ohm. Low-complexity global motion estimation from P-frame motion vectors for MPEG-7 application. *ICIP*, pages 271–274, 2000.

[6] Y. Su, M.-T. Sun, and V. Hsu. Global motion estimation from coarsely sampled motion vector field and the applications. *IEEE Trans. on Circuits Syst. Video Technol.*, 15(2):232–242, 2005.

[7] R. Tron and R. Vidal. A benchmark for the comparison of 3D motion segmentation algorithms. *CVPR*, 2007.