IPSJ Transactions on Computer Vision and Applications Vol.5 55–59 (July 2013)

[DOI: 10.2197/ipsjtcva.5.55]

Express Paper

Mixed-motion Segmentation Using Helmholtz Decomposition

Cuicui Zhang[1,a)] Xuefeng Liang[1,b)] Takashi Matsuyama[1,c)]

Received: March 11, 2013, Accepted: April 24, 2013, Released: July 29, 2013

Abstract: Motion estimation and segmentation poses challenges in dynamic scenarios where multiple motions are mixed up and interdependent. However, existing approaches in 2D motion field usually require the mixed motions to be independent. Algorithms incorporating 3D information have proven to be superior to purely 2D approaches in many studies. Inspired by this idea, we propose a new algorithm for evolving 3D potential surfaces using Helmholtz decomposition to represent 2D motion field. Meanwhile, a surface segmentation scheme is introduced to put different motions onto different layers, so that those interdependent motions can be separated and recovered efficiently. Unlike other approaches, our method does not require the prior knowledge of the motion model. The performance is demonstrated using real data under various complex scenarios.

Keywords: mixed-motion segmentation, potential surface, Helmholtz decomposition, surface segmentation

1. Introduction

Motion segmentation plays a central role in video analysis, such as the content-based retrieval, surveillance, human-computer interaction, action recognition, robot learning, etc [4]. 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. Background motion (global motion) includes two cases: 1. one is produced by a moving camera with unconstrained and a prior unknown motion occupying the entire view [4]; 2. another is represented by a dominant motion which occupies majority of a view. In many applications, the camera/dominant motion is of much less interest, and solely the local object motion is expected.

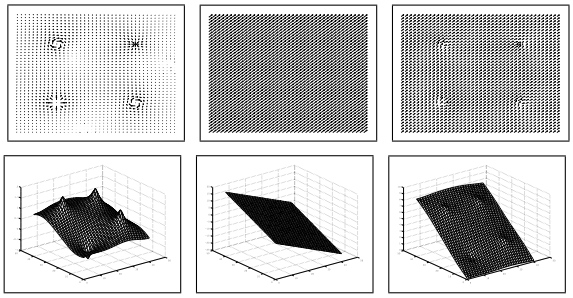
To address this problem, several approaches have been proposed for global motion estimation and motion segmentation. The work in Ref. [10] introduced a parametric form which assumed the global motion model from simple translation to general perspective transformation using different parameters. But this assumption may not always fit the data, especially when the global motion is represented by a dominant motion. In addition, if the global motion is mixed with object motions (the global motion and the object motion are also named as inlier and outlier, respectively), the inlier estimation may suffer from outliers. To solve this problem, a joint global motion estimation and segmentation method was proposed in Ref. [3]. It iteratively updates the inlier model by segmenting the outlier out. A regression scheme, using gradient descent (GD) [10] or least squares (LS) [9], is also applied to refine the inlier model by iteratively excluding the outliers that do not fit the current inlier model. An outlier rejection filter in Ref. [2] explicit filters motion vectors by checking their similarity in a pre-defined window. The window with low similarity is regarded as the outlier, and removed. RANSAC in Ref. [5] is a statistical method which estimates the inlier model by iteratively updating the probability of inlier. All above methods investigate the relations between inlier and outliers in 2D space. They require the multiple motions to be independent. But for interdependent motions, they may fail to deal with.

Unlike existing works, our motion segmentation and recovery method does not analyze motion vector field directly in 2D space, but in 3D space. The main contributions of our method include two aspects: (1) The 2D vector field is transformed into a 3D potential surface which locates different motions onto different layers, no matter if they are dependent or independent, rigid or non-rigid. (2) Based on surface fitting, the global motion is estimated by constructing a smooth surface which approximates the basic shape of the potential surface. Subsequently, local motions are recovered by removing the global motion from the original motion field. Experimental results demonstrate that our method is more efficient and accurate than those working in the 2D space.

2. Motivation

Dynamic scenarios are usually composed by multiple motions which are mixed up. Conventionally, the motion field is depicted with millions of vectors on the image plane which like chaos. See **Fig. 1** (c) for illustration. It is rather challenging to ascertain what exact motions they are. To reveal the essence of appearance based multiple motions, it is better to place different motions on different layers. 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

[1] Department of Intelligence Science and Technology, 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

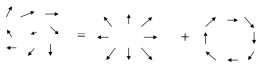


(a) (b) (c)

Fig. 1 (a) 2D motion field containing two vortices, one sink, one source, and its potential surface. (b) a constant 2D motion and its potential surface. (c) the mixture of (a) & (b).

represent the global motion. In our method the 3D surface is calculated using Helmoholtz decomposition, which has been applied in visualization and simulation of computational fluid dynamics (CFD). In Ref. [6], a particle filter based on Helmholtz decomposition was proposed for the 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 (satisfying $\nabla \times (\nabla E) = \vec{0}$) and divergence-free (curl) component $\nabla \times \vec{W}$ (satisfying $\nabla \cdot (\nabla \times \vec{W}) = \vec{0}$), where E and W are what we want to obtain the 3D potential surfaces.

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



3. Main Theory of the Proposed Method

The 3D potential surfaces of curl and divergence component are first calculated using energy minimization. Then, the global motion is approximated based on surface fitting. Finally, local motions are recovered by subtracting the approximated global motion from the original motion field.

3.1 Potential Surface Calculation

The potential surface is a surface whose gradient corresponds to a vector field in a connected region. Figure 1 shows the original vector field and its corresponding potential surface. 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 $\nabla E = (\partial E_u/\partial u) + (\partial E_v/\partial v)$. They have the following relationship: given a 2D vector field $V = (u, v)$, where u and v denote the horizontal and vertical component of the vector field respectively, assume every vector in V is rotated 90° in counter clockwise order, $V^{\perp} = (-v, u)$, then

$$(\nabla \times \vec{W})^{\perp} = \frac{\partial W_u}{\partial u} - \frac{\partial(-W_v)}{\partial v} = \nabla W. \tag{2}$$

Since the divergence component ∇E is the projection of the original motion field V to the space of the divergence field, the distance between V and the projected ∇E should be minimal. Therefore, we apply energy minimization to calculate the potential surface as follows:

$$D(E) = \int_{\Omega} \| \nabla E - \vec{V} \|^2 \, d\Omega, \tag{3}$$

where Ω represents the image domain. Similarly, the curl component $\vec{T} = \nabla \times \vec{W}$ is calculated by minimizing the following energy function:

$$G(\vec{W}) = \int_{\Omega} \| \nabla \times \vec{W} - \vec{V} \|^2 \, d\Omega. \tag{4}$$

According to the definition in Section 2, the curl component does not exist in the divergence component (∇E). And vice versa. Now, we can derive the following criteria:

$$\begin{aligned} \int_{\Omega} \nabla \times (\nabla E) d\Omega &= \int_{\Omega} \nabla \times (V - \nabla \times \vec{W}) d\Omega = 0, \\ \int_{\Omega} \nabla \cdot (\nabla \times \vec{W}) d\Omega &= \int_{\Omega} \nabla \cdot (V - \nabla E) d\Omega = 0. \end{aligned} \tag{5}$$

In the discrete domain, Eq. (5) can be rewritten as:

$$\begin{aligned} \sum_{i \in \Omega} \nabla \times (\nabla \times \vec{W}_i) &= \sum_{i \in \Omega} \nabla \times V_i, \\ \sum_{i \in \Omega} \nabla \cdot (\nabla E_i) &= \sum_{i \in \Omega} \nabla V_i. \end{aligned} \tag{6}$$

Since they are linear functions, we can abbreviate them as:

$$S_1 E = B, \qquad S_2 W = C. \tag{7}$$

where S_1 and S_2 are $N \times N$ sparse element matrix, E and W are the $N \times 1$ vector to be calculated, B and C are vectors that can be calculated from the right hand side of Eq. (7). Once we have the potential surfaces E and W by solving Eq. (7), $\nabla \times \vec{W}$ is subsequently obtained by Eq. (2).

3.2 Global Motion Estimation

Global motion estimation is one of the vital steps in our method. As decomposed into not just divergence component but also curl component, global motion is estimated from both E and $\vec{W}$. Here, we first illustrate the method on scalar potential surface E. The procedure on $\vec{W}$ is analogous.

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. Basically, the first surface fitting plays a role of rejecting outliers.

$$\begin{aligned} z = a_{d0}x^d + a_{0d}y^d + \cdots + a_{ij}x^i y^j + \cdots + a_{10}x \\ + a_{01}y + a_{00}, \end{aligned} \tag{8}$$

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 the outliers. If D is greater than a threshold T, then the point is marked as outlier and will not be considered in the second surface fitting. Thus, the second surface fitting solely approximates the points beyond outliers. To have a better estimation, a polynomial of higher degree

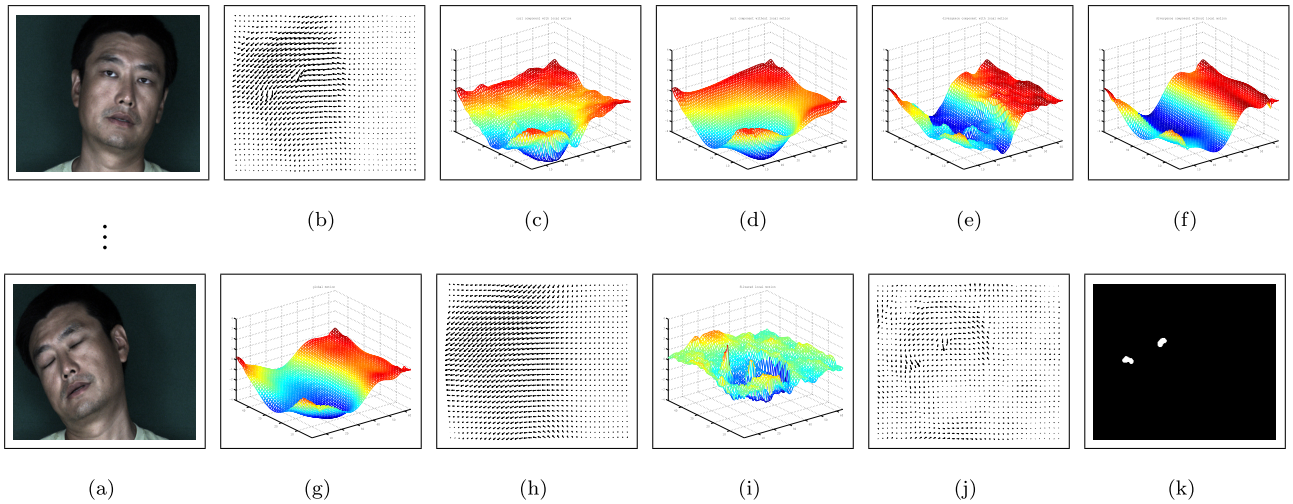


Fig. 2 Scenario 1: Face motion. (a) the sequence from frame 1 to frame 50. (b) motion field of one frame. (c) $\vec{W}$ with local motions. (d) $\vec{W}$ without local motions. (e) E with local motions. (f) E without local motions. (g) potential surface of estimated global motion. (h) recovered global motion field. (i) potential surface of estimated local motion. (j) recovered local motion field. (k) segmentation result.

of $d_1 = 10$ is employed. Finally, we will have a global motion surface E' which best approximates the base of E. The global motion field of divergence component is calculated by $G_1 = \nabla E'$.

Similarly, the global motion field of curl component is computed by $G_2 = \nabla \times \vec{W}'$. The final global motion field is estimated by linear combination of G_1 and G_2.

3.3 Local Motion Recovery

After obtaining global motion, local motions can be recovered directly by subtracting the global motion from the original motion field.

4. Experiments

4.1 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. In many systems, head motion is of less interest and expected to be taken off. **Figure 2** (a) shows a sequence in which the head is rotating, meanwhile, the eyes are blinking. In this data, the head motion is a dominant motion. In the optical flow (b), the eye motion shows a misleading direction because of fusing with the head motion. In (c) and (e), we can easily see two outliers on the potential surfaces which indicate the eye blinking. 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 practical and pretty challenging case taken by a handheld camera. This scene involves three motions: a rotating camera view, in which there exist a rotating basket and a translating box. See **Fig. 3** (a). This data is a benchmark sequence in Ref. [11]. In their method, several checkerboards are placed in the scene to obtain prominent feature points for motion segmentation. Our method, in contrast, does not rely on these strong key features. Those checkerboards contribute little help for us. In other words, other patterns can replace the checkerboard when using our method. In the optical flow (b), it is tricky to figure out what motions are involved exactly. But, on the potential surfaces (c) and (e), it is not difficult to see two outliers indicating two local motions. After surface fitting, two local motions are cleared depicted in (i). The recovered global motion (h), local motions (j) and segmentation (k) demonstrate the robustness of our method in rather complex scenarios.

4.2 Comparison with the State-of-the-art

We compared our method with five well-known motion segmentation methods including: (1) the joint global motion estimation and segmentation (GME-SEG) in Ref. [3], (2) and (3) are the iterative estimation method based on least-square (LS) [9], and gradient descent (GD) [10], (4) is an outlier rejection filter (Filter) in Ref. [2], and (5) is the RANSAC [5]. The vector field is calculated by optical flow method in Ref. [1] and optimized by [7], [8]. The segmentation results on two scenarios of these five and ours are shown in **Fig. 4**. Where, the local motions are illustrated in white, the global motion and static background are in black. We can see the proposed method outperforms the five reference methods in both two scenarios. Following is the analysis in detail.

GME-SEG [3] defines two types of outliers: the noisy motion (Type 1), and other information that does not fit their predefined inlier model well (Type 2), such as moving objects. This method is robust to detect the outliers. However, its segmentation scheme based on MRF-Bayesian algorithm cannot classify Type 1 and Type 2 outliers well, especially when the noises are close to the moving objects. Figure 4 (b) shows some noises are mis-classified as object motions. The iterative estimation method based on LS and GD segment two data in Fig. 4 (c) and (d), respectively. LS works poorly when the inlier does not occupy the entire view. Thus, it treats head motion as outlier and can not find the eye motions in the first scenario. The result of GD in the first scenario just looks like chaos. GD also faces the same problem in

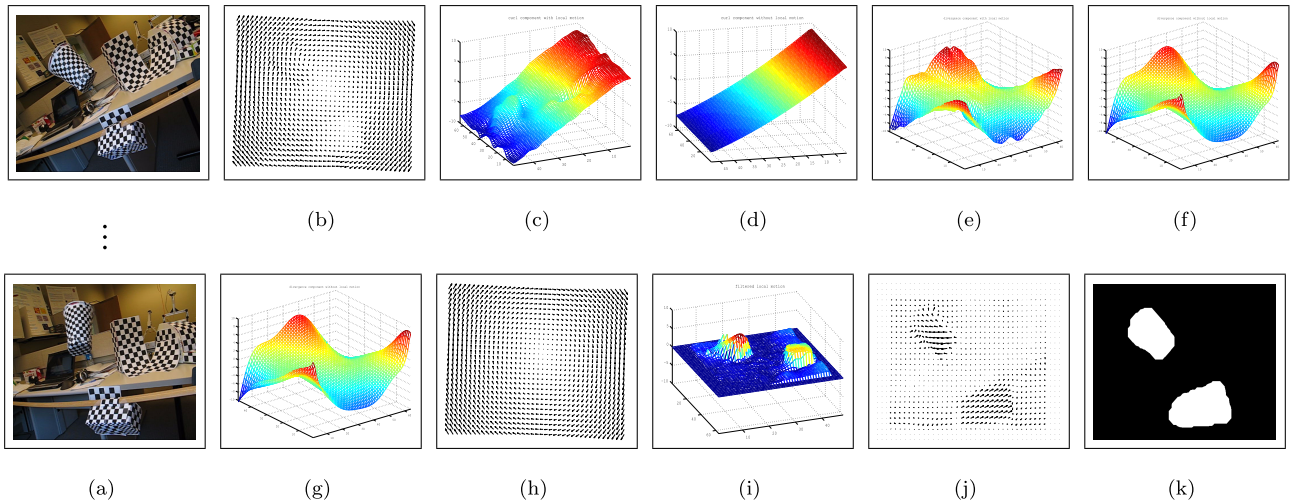


Fig. 3 Scenario 2: Checkerboard sequence. (a) the sequence from frame 1 to frame 27. (b) motion field of one frame. (c) $\vec{W}$ with local motions. (d) $\vec{W}$ without local motions. (e) E with local motions. (f) E without local motions. (g) potential surface of estimated global motion. (h) recovered global motion field. (i) potential surface of estimated local motion. (j) recovered local motion field. (k) segmentation result.

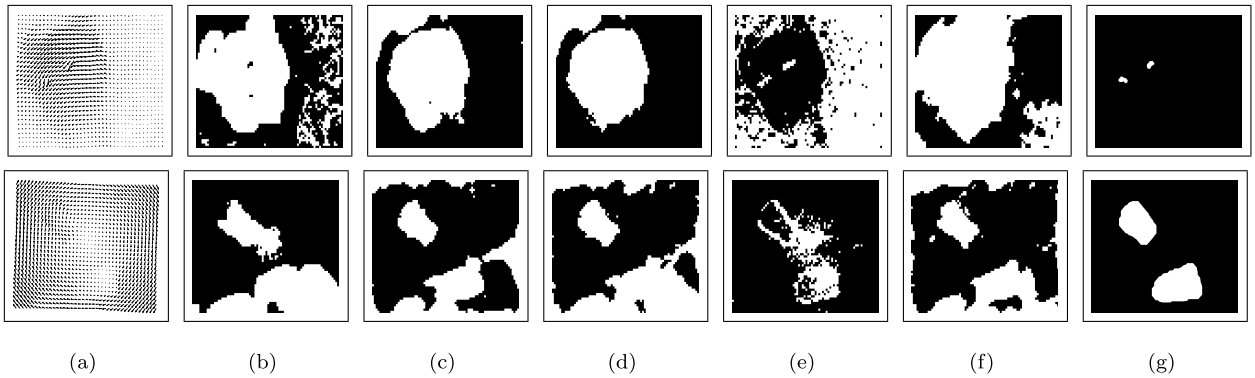


Fig. 4 Segmentation results of five reference methods and ours on two scenarios (a) motion field of one frame, (b) GME-SEG method [3], (c) LS [9], (d) GD [10], (e) Filter [2], (f) RANSAC [5], (g) our method.

these two scenarios as LS. The Filter in Ref. [2] explicitly filters motion vectors by checking their similarity in a pre-defined window. Hence the window size must be well decided, otherwise, too many valuable motion vectors are removed. The results of the second scenario in Fig. 4 (e) show the most inner vectors are excluded due to inappropriate window size. Note that, Filter in Ref. [2] is the only one that can locate eye blinking among five reference methods in the first scenario. The reason might be the default window size is approximately defined for the eye motions. RANSAC in Ref. [5] is a common method for outlier detection, and usually requires many iterations to get the reliable results. It often works worse when the frame has low resolution. Figure 4 (f) shows the segmented results are noisy in both scenarios. Moreover, these five reference methods suffer from a common limitation: they usually perform well on independent motions, but show weak ability on the interdependent motions in a complex scene.

We also did numerical evaluation using the segmentation error, defined by the mis-segmentation percentage in outliers and inlier (see Eq. (9)). The object region is manually cropped out in each frame to form the ground-truth data. It is easy to implement for the second scenario, because the edges of objects are obvious.

Table 1 Segmentation errors of existing methods and ours.

Method	Segmentation Error (%)
GME-SEG	21.07
LS	15.28
GD	24.66
Filter	8.17
RANSAC	18.00
Ours	2.85

However, it is tricky for the first scenario since the eye region is not well defined. Thus, the comparison is performed only on the second scenario. **Table 1** shows the results. We can see a significant superiority of our method in terms of segmentation error.

$$Error = \frac{misPixelsInOutliers + misPixelsInInlier}{totalPixels} \tag{9}$$

5. Conclusion

This paper proposes a robust motion segmentation and recovery method. It performs well on a wide range of motions, independent and dependent, rigid and non-rigid, single and multiple motions. Because we transform 2D vector fields into 3D potential surfaces using the Helmholtz decomposition, global motion and local motions are separated onto different layers. This makes mo-

tion segmentation be done efficiently. Moreover, applying surface fitting on the potential surface, the global and local motions are recovered accurately. Compared with several well-known works, our method requires no assumption of motion model, and is not sensitive to noises. Results demonstrate that our method performs much better in challenging scenarios where global and local motions are mixed up and interdependent.

Acknowledgments This work is supported by: Japan Society for the Promotion of Science, Scientific Research-KAKENHI for Grant-in-Aid for Young Scientists (ID:13276232). We also would like to thank the following researchers: Shohei Nobuhara, Hiroaki Kawashima, and Tony Tung for their much valuable comments and suggestions on this work.

References

[1] Brox, T., Bruhn, A., Papenberg, N. and Weickert, J.: High accuracy optical flow estimation based on a theory for warping, *ECCV*, pp.25–36 (2004).
[2] Chen, Y.M. and Bajic, I.V.: Motion vector outlier rejection cascade for global motion estimation, *IEEE Signal Process. Lett.*, Vol.17, No.2, pp.197–200 (2010).
[3] Chen, Y.M. and Bajic, I.V.: A joint approach to global motion estimation and motion segmentation from a coarsely sampled motion vector field, *IEEE Trans. Circuits Syst. Video Technol.*, Vol.21, No.9, pp.1316–1328 (2011).
[4] Cucchiara, R., Prati, A. and Vezzani, R.: Real-time motion segmentation from moving cameras, *Real-Time Imaging*, Vol.10, No.3, pp.127–143 (2004).
[5] Fischler, M. and Bolles, R.: RANSAC random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography, *Comm. ACM*, Vol.26, pp.381–395 (1981).
[6] Kakukou, N., Ogawa, T. and Haseyama, M.: An effective flow estimation method with particle filter based on Helmholtz decomposition theorem, *IEEE International Conference on Acoustics, Speech, and Signal Processing 2009* (*ICASSP 2009*), pp.949–952 (2009).
[7] Liang, X., McOwan, P. and Johnston, A.: A biologically inspired framework for spatial and spectral velocity estimations, *Journal of the Optical Society of America A*, Vol.28, No.4, pp.713–723 (2011).
[8] Liang, X., McOwan, P.W. and Johnston, A.: A color neuromorphic approach for motion estimation, *IEEE Winter Vision Meetings, Workshop on Motion and Video Computing*, pp.49–54 (2009).
[9] Smolic, A., Hoeynck, M. and Ohm, J.-R.: Low-complexity global motion estimation from P-frame motion vectors for MPEG-7 application, *ICIP*, pp.271–274 (2000).
[10] Su, Y., Sun, M.-T. and Hsu, V.: Global motion estimation from coarsely sampled motion vector field and the applications, *IEEE Trans. Circuits Syst. Video Technol.*, Vol.15, No.2, pp.232–242 (2005).
[11] Tron, R. and Vidal, R.: A benchmark for the comparison of 3D motion segmentation algorithms, *CVPR* (2007).

(Communicated by *Seiji Hotta*)