ESTIMATION AND SEGMENTATION OF DISPLACEMENT FIELD USING MULTIPLE FEATURES

Sanghoon Sull

Mail Stop 210-1
NASA Ames Research Center
Moffet Field, CA 94035-1000
sull@image.arc.nasa.gov

Narendra Ahuja

Beckman Institute
University of Illinois at Urbana-Champaign
Urbana, IL 61801, USA
ahuja@vision.csl.uiuc.edu

ABSTRACT

We present an approach for estimating and segmenting the displacement field between two frames. Our method is based on local affine (first-order) approximation of the displacement field which is derived under the assumption of locally rigid motion. Each distinct motion is represented in the image plane by a distinct set of values of affine coefficients. All sets of values supported by the feature locations in two frames are identified by exhaustive coarse-to-fine search. The integrated use of multiple features (points, regions and lines) increases the probability of finding well-supported sets. The sets of coefficients thus obtained are used to describe the DF. Two experimental results with real images are presented to demonstrate the feasibility of our approach.

1. INTRODUCTION

Estimation and segmentation of the displacement field (DF) between two image frames is one of the most important but difficult steps in the areas of computer vision and image processing, especially for detection and tracking of multiple moving objects, image registration, motion compensation, image data compression, and estimation of three-dimensional (3D) motion and structure of objects. The difficulty comes from 3D structural discontinuities and occlusions, as well as from independently moving objects.

The existing DF estimation methods can be largely divided into three categories which are based on i) block matching, ii) pel-recursive estimation, and iii) discrete feature correspondences, respectively. Block matching techniques can handle large image motion within the range of a given search area. These techniques are useful especially when the displacement is constant for a block of pixels. But, when the displacement is not constant or discontinuous, they do not work well. For small image motion, we can use recursive methods using gradient [1, 2]. However, these methods cannot be applied for large pixel displacement. The DF estimated using recursive methods is usually less reliable than that obtained by block search. Therefore, for those images having large image motion and discontinuities, a feature-based DF estimation approach is a promising choice. An approach based on point correspondences was recently presented for the images whose DF in the whole image plane can be approximated by one set of four coefficients [3].

In this paper, we propose an approach to estimating and segmenting DF between two successive frames. The DF is estimated and segmented by a method that allows the DF in each segment to be represented by one set of six affine coefficients. A segment consists of regions in the image. (A region is defined as a connected set of pixels having similar intensity values in an image.) Our method is based on local affine approximations of the DF which is derived under the assumption of locally rigid motion. Thus, the displacement vector (d_x, d_y) at t_1, which represents the displacement caused by motion of a point located at (x, y), is locally approximated by:

$$\begin{aligned} d_x(\mathbf{c}_x) &= c_0 + c_1 x + c_2 y \\ d_y(\mathbf{c}_y) &= c_3 + c_4 x + c_5 y. \end{aligned} \quad (1)$$

Each distinct motion is represented in the image plane by a distinct set of values of six coefficients $\mathbf{c} \stackrel{\text{def}}{=} \mathbf{c}_x \cup \mathbf{c}_y$ where $\mathbf{c}_x \stackrel{\text{def}}{=} \{c_0, c_1, c_2\}$ and $\mathbf{c}_y \stackrel{\text{def}}{=} \{c_3, c_4, c_5\}$.

To find the values of these distinct set of coefficients, multiple types of discrete features (points, lines and regions) between two successive frames are used. The first-order displacement equations are derived for regions and lines by using Eq. (1). All sets of values supported by the feature locations in two adjacent frames are identified by exhaustive coarse-to-fine search. However, to reduce computational complexity the 6D parameter space is decomposed into two disjoint 3D spaces. The support for any set is computed from the image plane distances between the observed feature locations and those predicted by the coefficient values. The well-supported sets of values thus found establish correspondences between features and group the features into subsets corresponding to locally rigid patches of the moving objects [4]. Note that the integrated use of the multiple features increases the probability of finding well-supported sets. Then, a set of coefficients $\mathbf{c}$ are recomputed linearly by using the matched features in each group since they were obtained from the quantized space.

The N_S sets of coefficients $\mathbf{c}$ thus obtained are used to describe the DF. For each region extracted at t_1, we select a set of coefficients which best describes the DF in the re-

This work was supported partly by Defense Advanced Research Projects Agency and the National Science Foundation under grant IRI-89-02728.

0-8186-6950-0/94 $4.00 © 1994 IEEE

gion in a sense that it yields the minimum correlation error between the intensity values at t_1 and those predicted by $\mathbf{c}$ from the second frame. Since a region is initially extracted using only intensity values, it can contain different affine motions. In this case, it is split by identifying those pixels whose displacement vectors satisfy the affine transformation given by each set of coefficients.

For image data compression, code words representing the quantized values of affine coefficients and information on the location of segments are generated. Isolated or very small segments can be ignored to make the coding scheme efficient. The method presented does not guarantee that all regions are included into one of N_S segments obtained. (This happens if some regions do not have features enough to yield well-supported sets although the use of multiple features gives a larger number of features in a region.) For these unsegmented regions, we can transmit the average intensity values instead of DF. At the pixels in those regions, it is also possible to compute $\mathbf{c}$ by iteratively minimizing correlation error $E_{R_i}(\mathbf{c})$ which is defined later with respect to $\mathbf{c}$.

Section 2 discusses local affine modeling of the displacement field. Section 3 presents an approach for finding distinct sets of affine coefficients. Section 4 describes how to use the distinct sets of affine coefficients to describe the DF. Section 5 presents the results from two experiments. Section 6 presents conclusion.

2. AFFINE MODEL FOR DISPLACEMENT FIELD

This section presents an affine displacement model to describe locally the image plane motion of an object undergoing a general 3D rigid motion. The X and X' denote variables or labels at t_1 and t_2, respectively.

Consider a point $\vec{X}_0 = [X_0, Y_0, Z_0]^T$ on an object in 3D at t_1. Let $\vec{X}'_0$ be the corresponding point at t_2. Denoting the rotation matrix and translation vector as $\mathbf{R}$ and $\vec{T}$, respectively, the general 3D rigid motion is expressed by

$$\vec{X}'_0 = \mathbf{R}\vec{X}_0 + \vec{T}. \tag{2}$$

Assuming the perspective projection with the focal length equal to one, the image coordinates (x'_0, y'_0) of $\vec{X}'_0$ at t_2 can be expressed as

$$x'_0 = \frac{r_{11}x_0 + r_{12}y_0 + r_{13} + \frac{T_X}{Z_0}}{r_{31}x_0 + r_{32}y_0 + r_{33} + \frac{T_Z}{Z_0}} \tag{3}$$

$$y'_0 = \frac{r_{21}x_0 + r_{22}y_0 + r_{23} + \frac{T_Y}{Z_0}}{r_{31}x_0 + r_{32}y_0 + r_{33} + \frac{T_Z}{Z_0}} \tag{4}$$

where (x_0, y_0) are the image coordinates at t_1, and $r_{11}, \ldots, r_{33}$ are the nine elements of $\mathbf{R}$.

Consider a point $\vec{X} = [X, Y, Z]^T$ on the same object in a neighborhood of $\vec{X}_0$. If we assume that the depth difference is small compared with Z_0 (i.e., $\frac{|Z-Z_0|}{Z_0} \ll 1$), we have

$$Z = Z_0(1 + \frac{Z - Z_0}{Z_0}) \approx Z_0. \tag{5}$$

Let (x, y) and (x', y') be the image coordinates of $\vec{X}$ at t_1 and $\vec{X}'$ at t_2, respectively. Then, the displacement field of the point $\vec{X}$ in the neighborhood of $\vec{X}_0$ is represented by

$$d_x(x,y) \stackrel{\text{def}}{=} x' - x = \frac{r_{11}x + r_{12}y + r_{13} + \frac{T_X}{Z_0}}{r_{31}x + r_{32}y + r_{33} + \frac{T_Z}{Z_0}} - x \tag{6}$$

$$d_y(x,y) \stackrel{\text{def}}{=} y' - y = \frac{r_{21}x + r_{22}y + r_{23} + \frac{T_Y}{Z_0}}{r_{31}x + r_{32}y + r_{33} + \frac{T_Z}{Z_0}} - y. \tag{7}$$

If we consider Taylor's series at the neighbor point of $\vec{X}_0$ whose depth difference is small, we can ignore the higher-order terms. Therefore, if (1) $\frac{|Z-Z_0|}{Z_0} \ll 1$, and (2) the values of the second partial derivatives in the neighborhood of (x_0, y_0) which is not at object or occlusion boundary are not large, we can locally approximate the DF using affine transformations:

$$d_x = x' - x = c_0 + c_1x + c_2y \tag{8}$$

$$d_y = y' - y = c_3 + c_4x + c_5y. \tag{9}$$

Although the normalized image coordinates with the focal length equal to one are used in the above derivation, any coordinate system which is affine-transformed from the normalized image coordinate system, for example, row and column coordinate system, can be used if the camera calibration is modeled by using 4 independent parameters for horizontal and vertical scale factors and image center coordinates. Note that the factor for the focal length (f) is included in horizontal and vertical scale factors. (The purpose of the camera calibration is to establish the projection from the 3D world coordinates to the 2D image coordinates.). Let $\vec{x} \stackrel{\text{def}}{=} [x, y]^t$ and $\vec{u} \stackrel{\text{def}}{=} [u, v]^t$ represent the normalized image plane coordinates and the pixel coordinates (row and column). Then, the calibration of a camera is to find the transformation from $\vec{u}$ to $\vec{x}$. Although the transformation is sometimes represented by six parameters [5], the angle between two retinal axes can be assumed to be perpendicular and therefore the four parameters are to be used to represent the transformation. Denoting the horizontal and vertical scale factors, and the offset for the image center coordinates by s_u and s_v, and $\vec{o} \stackrel{\text{def}}{=} [u_0, v_0]^t$, respectively, we have

$$\vec{x} = \mathbf{S}\vec{u} + \vec{o}, \tag{10}$$

where

$$\mathbf{S} \stackrel{\text{def}}{=} \begin{bmatrix} s_u & 0 \\ 0 & s_v \end{bmatrix}. \tag{11}$$

Writing Eqs. (8) and (9) in a vector form, we have

$$\vec{d}_{x,y} = \mathbf{A}\vec{x} + \vec{b}, \tag{12}$$

where

$$\mathbf{A} \stackrel{\text{def}}{=} \begin{bmatrix} c_1 & c_2 \\ c_4 & c_5 \end{bmatrix}, \tag{13}$$

$\vec{d}_{x,y} \stackrel{\text{def}}{=} [d_x, d_y]^t$, and $\vec{b} \stackrel{\text{def}}{=} [c_0, c_3]^t$. Replacing $\vec{d}_{x,y}$ and $\vec{x}$ with Eq. (10), we have

$$\mathbf{S}\vec{u}' + \vec{o} - \mathbf{S}\vec{u} - \vec{o} = \mathbf{A}(\mathbf{S}\vec{u} + \vec{o}) + \vec{b}. \tag{14}$$

Therefore, we can also represent the DF using a new affine transformation with respect to $\vec{u}$;

$$\vec{d}_{u,v} = \mathbf{A}'\vec{u} + \vec{b}', \tag{15}$$

where $\mathbf{A}' \stackrel{\text{def}}{=} \mathbf{S}^{-1}\mathbf{A}\mathbf{S}$ and $\vec{b}' \stackrel{\text{def}}{=} \mathbf{S}^{-1}(\mathbf{A}\vec{o} + \vec{b})$. In summary, if the DF is modeled by affine transformations, the camera calibration is not necessary and therefore an arbitrary coordinate system can be used. In our experiments, the values of rows and columns are used for the coordinates.

3. COMPUTATION OF AFFINE SETS

This section describes an approach for finding distinct sets of six coefficients which are well-supported by feature locations. We first present first-order displacement models for each type of features (points, regions and lines), and then describe how to obtain distinct sets of coefficients.

3.1. Affine Model For Features

3.1.1. Point

Using Eqs. (8) and (9), we define the followings for a pair of point features (P_i, P'_j) at t_1 and t_2 for a given set of $\mathbf{c}$:

$$\delta_{x,P,ij}(\mathbf{c}_x) \stackrel{\text{def}}{=} x'_j - (c_0 + (1+c_1)x_i + c_2 y_i) \tag{16}$$

$$\delta_{y,P,ij}(\mathbf{c}_y) \stackrel{\text{def}}{=} y'_j - (c_3 + c_4 x_i + (1+c_5)y_i). \tag{17}$$

Then, the image error for the pair is defined as

$$\delta_{P,ij}(\mathbf{c}) \stackrel{\text{def}}{=} \sqrt{\delta_{x,P,ij}(\mathbf{c}_x)^2 + \delta_{y,P,ij}(\mathbf{c}_y)^2}. \tag{18}$$

3.1.2. Region

Consider a pair of regions R and R' at t_1 and t_2, respectively. Denoting the coordinates of 2D centroids of the pair of regions by (C_x, C_y) and (C'_x, C'_y), respectively, we can derive the followings for a given $\mathbf{c}$ [4]:

$$\delta_{x,R,ij}(\mathbf{c}_x) \stackrel{\text{def}}{=} C'_{x,j} - (c_0 + (1+c_1)C_{x,i} + c_2 C_{y,i}) \tag{19}$$

$$\delta_{y,R,ij}(\mathbf{c}_y) \stackrel{\text{def}}{=} C'_{y,j} - (c_3 + c_4 C_{x,i} + (1+c_5)C_{y,i}) \tag{20}$$

The image error for the pair of regions is defined as

$$\delta_{R,ij}(\mathbf{c}) \stackrel{\text{def}}{=} \sqrt{\delta_{x,R,ij}(\mathbf{c}_x)^2 + \delta_{y,R,ij}(\mathbf{c}_y)^2}. \tag{21}$$

3.1.3. Line

Consider a pair of lines $L : Ax + By + C = 0$ and $L' : A'x + B'y + C' = 0$ at t_1 and t_2, respectively. Since the perpendicular distance from the predicted image coordinates of an end point (x_p, y_p) of L to L' should be zero, we can derive the following for a given $\mathbf{c}$:

$$\delta_{p,L,ij}(\vec{c}) \stackrel{\text{def}}{=} \frac{A'x'_{p,i} + B'y'_{p,i} + C'}{\sqrt{A'^2 + B'^2}} \tag{22}$$

where $x'_{p,i} = c_0 + (1+c_1)x_{p,i} + c_2 y_{p,i}$ and $y'_{p,i} = c_3 + c_4 x_{p,i} + (1+c_5)y_{p,i}$. The $\delta_{p,L,ij}$ for the other end point (x_q, y_q) of L is derived similarly. Then, the image error for the pair of lines is defined as

$$\delta_{L,ij}(\vec{c}) \stackrel{\text{def}}{=} \sqrt{\frac{\delta_{p,L,ij}(\vec{c})^2 + \delta_{q,L,ij}(\vec{c})^2}{2}}. \tag{23}$$

We note here that Eq. (22) involves both sets of coefficients $\mathbf{c}_x$ and $\mathbf{c}_y$ while only one set appears in Eqs. (16) and (17), and Eqs. (19) and (20).

3.2. Integration Of Multiple Features

The affine transformations for points, regions and lines are given by Eqs. (16) and (17), Eqs. (19) and (20), and Eq. (22), respectively. Note that the error measure of each feature $(\delta_{P,ij}, \delta_{R,ij}, \delta_{L,ij})$ is in terms of the same unit, i.e., the image error, and we can treat them equally when we construct a support function to be maximized. Consider a group of *local* features at t_1 which are close together since the affine transformations are valid locally in the image plane. For each pair of the same type of features at t_1 and t_2, we define a support function F of $\mathbf{c}$ as follows:

$$F(\mathbf{c}) \stackrel{\text{def}}{=} \sum_{i,j} \Psi_\epsilon(\delta_{P,ij}(\mathbf{c})) + \sum_{i,j} \Psi_\epsilon(\delta_{R,ij}(\mathbf{c})) + \sum_{i,j} \Psi_\epsilon(\delta_{L,ij}(\mathbf{c})) \tag{24}$$

where

$$\Psi_\epsilon(x) \stackrel{\text{def}}{=} \begin{cases} 1 - \frac{|x|}{\epsilon} & \text{if } -\epsilon < x < \epsilon \\ 0 & \text{otherwise,} \end{cases} \tag{25}$$

ϵ is a predetermined number. Note that only those pairs of features having similar 2D attributes at two time instants are considered to reduce the computation as well as to increase confidence in the solution.

Our goal is to find all sets of the six parameters $\widehat{\mathbf{c}}$ corresponding to the dominant local maxima of F. The support function F is likely to have many small local maxima since the extracted features contain unknown subsets each having a different, unknown motion. Therefore, exhaustive search is one way to obtain all dominant local maxima. However, since the 6D space is too large to search, we decompose this into two disjoint 3D spaces. A similar decomposition of 6D space was also used by Adiv [6] for Hough transform. We define the decomposed support functions F_x and F_y as follows:

$$F_x(\mathbf{c}_x) \stackrel{\text{def}}{=} \sum_{i,j} \Psi_{\frac{\epsilon}{\sqrt{2}}}(\delta_{x,P,ij}(\mathbf{c}_x)) + \sum_{i,j} \Psi_{\frac{\epsilon}{\sqrt{2}}}(\delta_{x,R,ij}(\mathbf{c}_x)) \tag{26}$$

$$F_y(\mathbf{c}_y) \stackrel{\text{def}}{=} \sum_{i,j} \Psi_{\frac{\epsilon}{\sqrt{2}}}(\delta_{y,P,ij}(\mathbf{c}_y)) + \sum_{i,j} \Psi_{\frac{\epsilon}{\sqrt{2}}}(\delta_{y,R,ij}(\mathbf{c}_y)) \tag{27}$$

where the terms representing line pairs are not included since the expressions for $\delta_{p,L,ij}$ and $\delta_{q,L,ij}$ in Eqs. (22) involve all six coefficients.

To obtain all the dominant local maxima of F, we find the global maximum of F one at a time for the largest remaining features as follows: For the largest group of local features, we first find N_3 number of the candidate sets of $\mathbf{c}_x$ and $\mathbf{c}_y$ corresponding to the peak values of F_x and F_y by

searching each quantized 3D space, respectively. Then, for N_3^2 combinations of $\{\mathbf{c}_x, \mathbf{c}_x\}$, we select $\hat{\mathbf{c}}$ corresponding to the maximum of F defined in Eq. (24). These two steps are performed at coarse-to-fine resolution to reduce the computation. Note that ϵ in Eq. (24) is a function of resolution in the search space. The search is separately performed in the 3D parameter spaces for feature points, line points and regions. (A line point is defined as the intersection of lines.) Then, the combination of solution triples which corresponds to the maximum value of F is obtained. Feature correspondences are established by using the set of the six coefficients yielding the maximum value of F and the corresponding matched features comprise a segment. After removing these matched features from further consideration, the above process continues until there is no dominant peak in the search space.

3.3. Linear Estimation Of An Affine Set

Given a segment of feature correspondences, denoted by S_l, a set $\mathbf{c}_l$ of the six affine parameters of S_l is linearly computed, thus yielding more accurate values than $\hat{\mathbf{c}}_l$ obtained by searching the quantized space. This linear computation is used to merge any two segments into one if they satisfy one affine transformation. Let $m_{P,l}$, $m_{L,l}$ and $m_{R,l}$ be the numbers of matched pairs (P_i, P_i'), (R_i, R_i') and (L_i, L_i') of points, regions and lines in S_l, respectively. (Here, features are relabeled in such a way that matched features have the same subscripts.) Then, we define an objective function to be minimized with respect to $\mathbf{c}_l$ as follows:

$$\delta_{S_l}(\mathbf{c}_l)^2 \stackrel{\text{def}}{=} \sum_{i=1}^{m_P} \delta_{P,ii}^2(\mathbf{c}_l) + \sum_{i=1}^{m_R} \delta_{R,ii}^2(\mathbf{c}_l) + \sum_{i=1}^{m_L} \delta_{L,ii}^2(\mathbf{c}_l) \quad (28)$$

where $\delta_{P,ij}$, $\delta_{R,ij}$ and $\delta_{L,ij}$ are defined in Section 3.1. Note that each term in the above objective function has the same unit. This minimization is a standard linear least squares problem which can be easily solved.

To measure the goodness of the segment S_l and the estimated $\mathbf{c}_l$, we define the average image error for matched features in S_l as follows:

$$\bar{\delta}_{S_l}(\mathbf{c}_l) \stackrel{\text{def}}{=} \sqrt{\frac{\delta_{S_l}(\mathbf{c}_l)^2}{(m_P + m_L + m_R)}}. \quad (29)$$

4. DESCRIPTION OF DISPLACEMENT FIELD USING DISTINCT AFFINE SETS

The $\mathbf{c}_j$ $(j = 1, \ldots, N_S)$ obtained in this way are used to describe the DF. For each region R_i $(i = 1, \ldots, N_R)$ extracted at t_1, we select one set $\mathbf{c}_l$ which best describes the DF in the region. Given image intensity functions I_1 and I_2 between two successive frames, we define an error measure with respect to $\mathbf{c}_j$ for a region R_i at t_1 as follows:

$$E_{R_i}(\mathbf{c}_j) \stackrel{\text{def}}{=} \frac{1}{A} \sum_{(x,y)\in R_i} \phi(I_1(x,y) - I_2(x'(\mathbf{c}_j), y'(\mathbf{c}_j))) \quad (30)$$

where $x'(\mathbf{c}_j) \stackrel{\text{def}}{=} x + d_x(\mathbf{c}_j)$, $y'(\mathbf{c}_j) \stackrel{\text{def}}{=} y + d_y(\mathbf{c}_j)$, and A is the area of R_i. For the noninteger values of (d_x, d_y), the bilinear interpolation is used. The robust error measure $\phi(x)$ [7] is used to reduce the effect of outliers, for example, caused by occlusion in a region:

$$\phi(x) \stackrel{\text{def}}{=} \frac{-1}{1 + (\frac{x}{\Delta})^2} \quad (31)$$

where Δ is a constant scale factor. Therefore, for each region R_i, we select $\mathbf{c}_l$ which yields the minimum value of $E_{R_i}(\mathbf{c}_j)$, $j = 1, \ldots, N_S$.

However, a region can have different affine motions since the regions are initially extracted using only intensity values. If the minimum value $E_{R_i}(\mathbf{c}_l)$ is larger than a given threshold ϵ, R_i is split using matched and grouped sets of discrete features. That is, if R_i includes matched features from different groups, it is segmented by identifying those pixels whose displacement vectors satisfy the affine transformation given by $\mathbf{c}_l$ from each group, respectively.

5. EXPERIMENTAL RESULTS

We first summary the major steps of our algorithm.

Algorithm

1. Extract features (points, lines and regions) independently at t_1 and t_2.
2. Find all sets of the $\mathbf{c}_j$ well-supported by feature locations in two frames by using exhaustive coarse-to-fine search.
3. For each extracted region R_i at t_1, select $\mathbf{c}_l$ which yields the minimum value of $E_{R_i}(\mathbf{c}_j)$, $j = 1, \ldots, N_S$. If the minimum value is smaller than a given threshold ϵ_S, the region P is included into the segment S_l.
4. Split the regions which are not included in any segment in Step 3.

In this paper, we present the experimental results obtained by Steps 1, 2 and 3 of the algorithm. Implementation details for Step 2 in the algorithm is described in [4]. The values of Δ and ϵ_S used were 10 and 0.7 respectively.

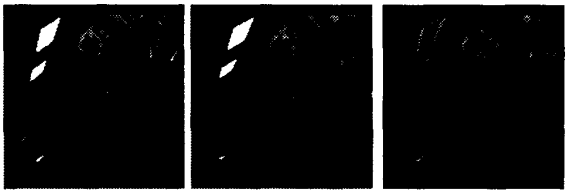
5.1. Indoor Images

Two frames of indoor scenes, of size 512 by 512, are used. The magnitude of the displacement vector in the lower-right part is as large as 37 pixels.

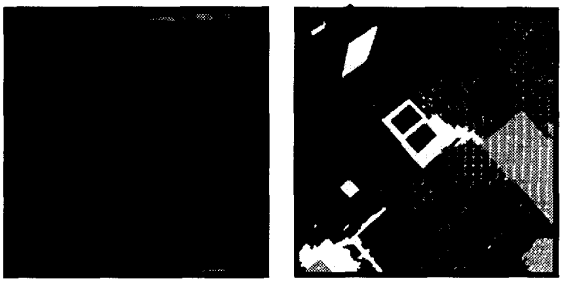
Figures 1(a) and (b) show the images from which 320 and 310 regions were extracted, respectively. Six distinct sets of affine coefficients were obtained by Step 2 of the algorithm. Among 320 regions at t_1, the DF of 295 regions could be represented by one of $\mathbf{c}_j$, $j = 1, \ldots, 6$. The resulting DF is shown in Fig. 1(e) and the six segments are also displayed using different grey values. The comparison of the intensity difference between I_1 and I_2 with the motion compensated intensity difference defined in Eq. (30), which are shown in Figs. 1(c) and (d), demonstrates the feasibility of our approach.

5.2. PUMA Images

Two real images (384 by 500) of PUMA are used. The magnitude of displacement vector at the larger arm is as large as 15 pixels.



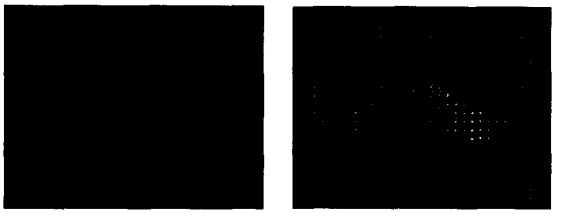
(a) (b) (c)



(d) (e)

Figure 1: Indoor Images: (a) First image I_1. (b) Second image I_2. (c) Intensity difference between I_1 and I_2. (d) Motion compensated intensity difference. (e) Resulting DF in I_1.

(a) (b) (c)



(d) (e)

Figure 2: PUMA Images: (a) First image I_1. (b) Second image I_2. (c) Intensity difference between I_1 and I_2. (d) Motion compensated intensity difference. (e) Resulting DF in I_1.

Figures 2(a) and (b) show the images. The numbers of regions extracted in two frames were 182 and 195, respectively. Three well-supported sets of affine coefficients were found by Step 2 of the algorithm. Among 182 regions, the DF of 173 regions could be represented by one of $\mathbf{C}_j$, $j = 1, \ldots, 3$, showing large amount of reduction of data necessary to represent the DF. The resulting DF is shown in Fig. 2(e) where the three segments are also displayed using different grey values.

Figures 2(c) and (d) show the intensity difference between I_1 and I_2 and motion compensated intensity difference (See Eq. (30).), respectively. Note that the error near the upper edges of both arms is caused by the occlusion.

6. CONCLUSION

In this paper, we have described an approach for estimation and segmentation of DF using distinct sets of first-order coefficients obtained from multiple features.

Our method will be useful especially for those images having large and multiple image motions.

7. REFERENCES

[1] P. P. H. Musmann and H. Grallert, "Advances in picture coding," *Proc. IEEE*, vol. 73, pp. 523–548, April 1985.

[2] D. F. J.L. Barron, "Performance of optical flow techniques," in *Proc. IEEE Conf. Comput. Vision Patt. Recogn.*, (Champaign, IL), pp. 236–242, June 1992.

[3] Q. Zheng and R. Chellappa, "A computational vision approach to image registration," *IEEE Trans. Image Process.*, vol. 2, pp. 311–326, July 1993.

[4] S. Sull and N. Ahuja, "Integrated matching and segmentation of multiple features in two views," *Comput. Vis., Graph., Image Process.: Image Understanding*, under review.

[5] Q. L. O. Faugeras and S. Maybank, "Camera self-calibration: Theory and experiments," in *Proc. European Conf. Comput. Vision*, pp. 321–334, 1992.

[6] G. Adiv, "Determining three-dimensional motion and structure from optical flow generated by several moving objects," *IEEE Trans. Patt. Anal. Mach. Intell.*, vol. PAMI-7, July 1985.

[7] P. R. W. S. F.R. Hampel, E.M Ronchetti, *Robust Statistics: An Approach Based On Influence Functions*. New York: Wiley, 1986.