조명변화 하에서 복합–피라미드 기반 영상운동 추정 알고리즘

문영수, 권인소 [o]

한국과학기술원, 자동화 및 설계공학과, [o] 전기 및 전자공학과

A Hybrid Pyramid-based Motion Estimation Algorithm under Illumination Variations

Young- Su Moon and In-So Kweon [o]

Department of Automation and Design Engineering, KAIST

[o] Department of Electrical Engineering, KAIST

Summary

In this paper we present a hybrid hierarchical (pyramid-based) estimation algorithm for robustly estimating dominant motion under illumination variation. This algorithm can be briefly summarized as the selective application of the three dynamic image models for accommodating local brightness changes between two subsequent image frames according to the pyramid level in conjunction with the coarse-to-fine strategy. For this, we first devise two simple and differential dynamic image models from the generalized dynamic image model (GDIM model) previously proposed. The GDIM model has the local parameters of the illumination gain and offset factors, while the devised two differential models contain either only the local illumination gain factor (MDIM model) or contains no illumination factor (NDIM model). Especially, since the NDIM model shows the strong superiority during the initial parameter updating process, starting from the initial zero motion estimate, at the coarse pyramid level, the accurate motion estimate by the NDIM model at the coarse level drives the subsequent motion estimation by the remaining models at the finer pyramid level to be performed successfully. In order to illustrate the validness of the hybrid hierarchical estimation framework, some results through simulation and experiments are provided.

1. Introduction

Robust estimation of the dominant image motion in an image sequence is very important in various computer vision applications, such as image registration, motion segmentation, and structure from motion.

Since most motion estimation algorithms depend on the intensity constancy assumption (ICA) [1], they become useful only under constant or slowly varying illumination condition. However, such an assumption is easily violated in real scene environments. According to the work [2, 3], an intensity value at an image pixel corresponding to a small surface patch in real scene is influenced by the following four photometric factors: light intensity, surface reflection property, response characteristic of image sensor, and geometric configuration between illumination lighting direction, surface normal, and image sensor orientation. So, change of any one among them may invoke significant discrepancy between two subsequent intensity values.

In recent years, several methods for motion estimation under illumination changes have been actively developed. Negahdaripour [4, 5] proposed an image motion constraint equation based on the generalized dynamic image model (GDIM), which expresses the local linear image variation between two subsequent intensity values $I_1(\mathbf{x})$ at time 1 and $I_2(\mathbf{x}+d\mathbf{x})$ at time 2 by the first-order linear equation with illumination gain and offset factors, $a(\mathbf{x})$ and $b(\mathbf{x})$, respectively,

$$I_2(\mathbf{x}+d\mathbf{x}) = a(\mathbf{x})\, I_1(\mathbf{x}) + b(\mathbf{x})\,, \qquad (1)$$

where $d\mathbf{x}$ indicates a small image displacement between two image frames. Assuming the local constancy of the unknown fields of both the optical flow and the image variation fields within a local neighborhood, he solved the unknown fields using the least-squares method. Szeliski and Coughlan [6] suggested the introduction of a global image variation model with constant illumination gain and offset factors to account for global photometric variation in image registration problem. In a recent work [7], the low-order polynomial function has been applied to each of the global illumination gain and offset factors so as to cover the more complicated global illumination variation. Also, there have been some results that solved the illumination variation problem regarding it as a linear parametric combination of some illumination basis images [8, 9]. On the other hand, Black [10] used the filtered input images, which are obtained through the local normalization of an input intensity image using the global sample mean of it, in his open code for robust motion estimation. His method may be valid under constant or globally varying illumination, but it becomes invalid under locally complicated illumination variation. Also, he didn't present any theoretical ground for the introduction of the filtered image and any systematic experimental results.

We here focus on the problem of robust estimation of large dominant image motion under local illumination variations. Of course, this task can be implemented by applying the previous GDIM dynamic image model to the classical direct model-based hierarchical estimation technique. However, this approach discloses the critical limit in the initial parameter estimation process, which will be discussed in detail later. Therefore, we suggest a hybrid hierarchical estimation strategy for accomplishing accurate motion estimation fast and effectively. For this, we first devise two simple and differential dynamic image models for accommodating local illumination variation from the GDIM model through some simple mathematical manipulations. Then, based on each of three dynamic image models, three algorithms for dominant motion estimation are implemented, and they are compared through some tests. As a result, we arrive at the presented hybrid hierarchical estimation framework, which can be briefly described as follows: the three dynamic image models including the GDIM model and the two devised models are employed selectively according to the pyramid level. Especially, since the second devised model is superior to the remaining models in performing the initial parameter updating process at the coarse level, employment of the NDIM dynamic image model for motion estimation at the coarse pyramid level is the essence of the proposed strategy. Finally, experimental results on synthetic and real images are provided to prove the validness and effectiveness of the presented algorithm.

2. Two differential dynamic image models

Let $m_i(\mathbf{x})$ and $\sigma_i^2(\mathbf{x})$ be the local sample mean and local sample variance of intensity functions, respectively, within a small neighborhood W centered at current image pixel $\mathbf{x}$ at time i. With the assumption of both $\alpha(\mathbf{x})$ and $\beta(\mathbf{x})$ in the GDIM model being constant within W, substituting the definition of $m_i(\mathbf{x})$ and $\sigma_i^2(\mathbf{x})$ into equation (1) gives the following two relations:

$$m_2(\mathbf{x}+\delta\mathbf{x}) = \alpha(\mathbf{x})\, m_1(\mathbf{x}) + \beta(\mathbf{x}) \,, \tag{2}$$

$$\sigma_2^2(\mathbf{x}+\delta\mathbf{x}) = \alpha^2(\mathbf{x})\, \sigma_1^2(\mathbf{x}) \,. \tag{3}$$

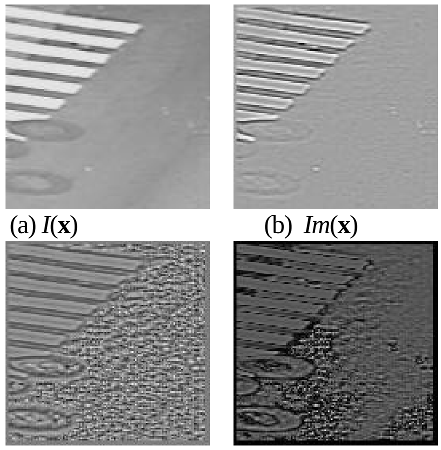
Next, subtraction of $m_i(\mathbf{x})$ in equation (2) from $I_i(\mathbf{x})$ in equation (1) leads to the following local mean difference dynamic image model (MDIM) for the local mean difference image function of $Im_i(\mathbf{x}) \equiv I_i(\mathbf{x}) - m_i(\mathbf{x})$:

$$Im_2(\mathbf{x}+\delta\mathbf{x}) = \alpha(\mathbf{x}) \cdot Im_1(\mathbf{x}) \,. \tag{4}$$

Finally, division of $Im_i(\mathbf{x})$ in equation (4) with $\sigma_i(\mathbf{x})$ in equation (3) provides the following local normalized dynamic image model (NDIM) for the local normalized image function of $Ims_i(\mathbf{x}) \equiv (I_i(\mathbf{x}) - m_i(\mathbf{x})) / \sigma_i(\mathbf{x})$:

$$Ims_2(\mathbf{x}+\delta\mathbf{x}) = Ims_1(\mathbf{x}) \,. \tag{5}$$

Fig. 1 shows a set of differential images including the intensity image, the local mean difference image, the local normalized image, and finally the Black's filtered image for a road scene, which are scaled into the gray range of 0 ~ 255 for visualization. Here, we can see that the latent features well not seen in the road scene are clearly revealed in the local normalized image. This characteristic may increase the amount of the salient image features necessary for accurate motion estimation.



(a) $I(\mathbf{x})$ (b) $Im(\mathbf{x})$ (c) $Ims(\mathbf{x})$ (d) Black's filtered image

Fig. 1 A set of differential images for a road scene

3. Formulation for dominant motion estimation

We here provide some formulations for robust estimation of dominant image motion under illumination variations, based on each dynamic image model. For dominant image motion estimation, direct motion model-based approach is basically implemented.

First, the first-order Taylor series expansion of the GDIM model (equation (1)) with $\alpha(\mathbf{x})$ and $\beta(\mathbf{x})$ gives the following motion constraint equation

$$\nabla I_2(\mathbf{x})^T \delta\mathbf{x} + I_2(\mathbf{x}) - \alpha(\mathbf{x})\, I_1(\mathbf{x}) - \beta(\mathbf{x}) = 0 \,, \tag{6}$$

where ∇ denotes a spatial gradient operator, and we approximate the small displacement $\delta\mathbf{x}$ as a 2D global dominant translation motion model containing two motion parameters, i.e., $\delta\mathbf{x} \cong \mathbf{a} = (u, v)^T$, for simple and clear analysis. The unknowns of the global motion parameter $\mathbf{a}$ and the local parameters $\alpha(\mathbf{x})$ and $\beta(\mathbf{x})$ at all pixels can be estimated by solving the nonlinear minimization problem of the following global energy function defined over an entire image,

$$E_{GDIM}(\mathbf{a}, \alpha, \beta) = \sum_{\mathbf{x}} \rho\left(r = \nabla I_2^{\,T}\, \mathbf{a} + I_2 - \alpha\, I_1 - \beta; \sigma\right) \tag{7}$$

where ρ indicates a robust loss function of a residual $r(\mathbf{x})$ and a scale factor σ to reduce the bad effect of the outlier components on the resulting motion estimate [12, 13]. To solve the nonlinear minimization problem effectively, after calculating the analytic solution of $\alpha(\mathbf{x})$ and $\beta(\mathbf{x})$ within the local neighborhood W and substituting it into the global energy function, we compute the global estimate $\mathbf{a}$ successively by applying the iterative reweighted least squares (IRLS) technique [4] to the adjusted global energy function. This is similar to the techniques used in [13, 14]. Then, $a(\mathbf{x})$ and $b(\mathbf{x})$ is naturally determined by the analytic solution using the current global motion estimate.

Next, for the MDIM model (equation (4)) with $a(\mathbf{x})$, the corresponding motion constraint equation is

$$\nabla Im_2(\mathbf{x}) \cdot d\mathbf{x} + Im_2(\mathbf{x}) - a(\mathbf{x})\, Im_1(\mathbf{x}) = 0\,, \qquad (8)$$

and the global energy function is defined as follows:

$$E_{MDIM}(\mathbf{a}, a) = \sum_{\mathbf{x}} r\left(r = \nabla Im_2^{\,T}\,\mathbf{a} + Im_2 - a\; Im_1; s\right). \qquad (9)$$

The unknowns of $\mathbf{a}$ and $a(\mathbf{x})$ in this model can be solved similarly to the former algorithm. Finally, in the case of the NDIM model (equation (5)) with only the unknown motion parameter $\mathbf{a}$, the motion constraint equation is determined as

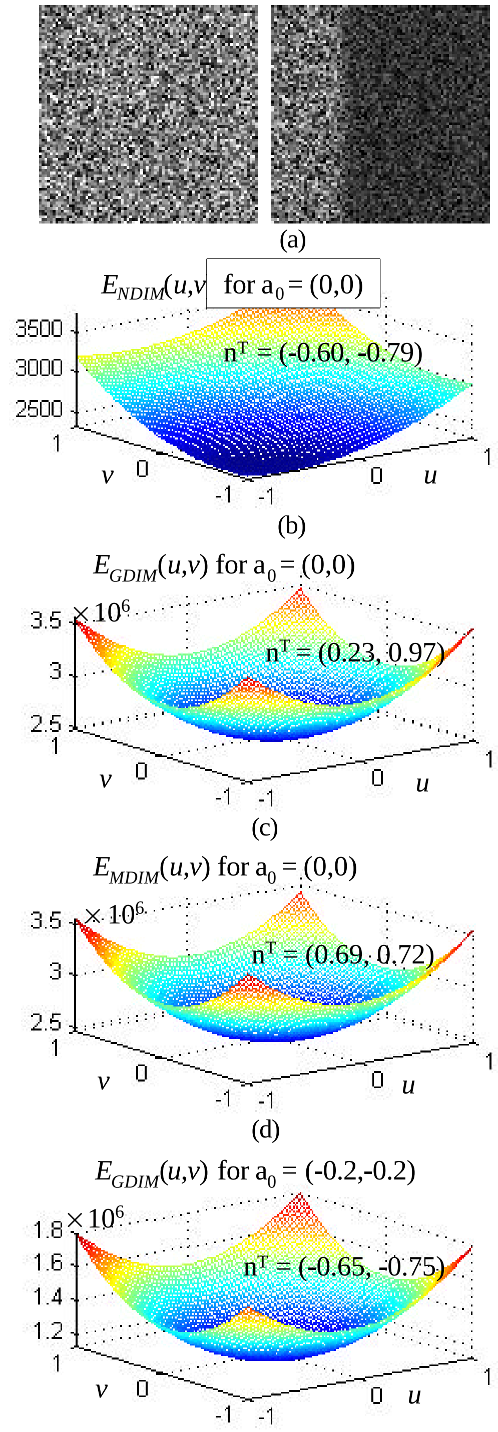
$$\nabla Ims_2(\mathbf{x}) \cdot d\mathbf{x} + Ims_2(\mathbf{x}) - Ims_1(\mathbf{x}) = 0\,, \qquad (10)$$

and its global energy function is defined as

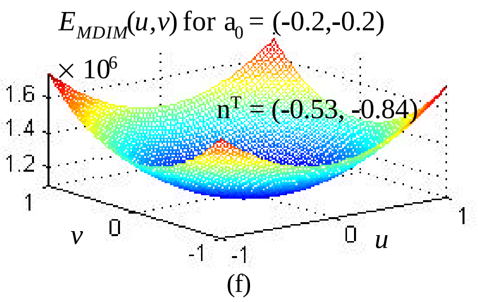
$$E_{NDIM}(\mathbf{a}) = \sum_{\mathbf{x}} r\left(r = \nabla Ims_2^{\,T}\,\mathbf{a} + Ims_2 - Ims_1; s\right). \qquad (11)$$

This can be solved in the same way as in the classical algorithm [4].

To assess the performance of each dynamic image model-based motion estimation algorithm using the formulations described above, they have been applied to a pair of random dot patterns in Fig.2 (a), containing both dominant translation motion of (u,v) = (-1,-1) and synthetic illumination variation. As a result, only the NDIM (model based) algorithm recovered the dominant motion well, while the GDIM and MDIM algorithms failed to recover it. Especially, in the case of the GDIM and MDIM model, the initial parameter updating step from the initial estimate of zero motion becomes unstable easily according to the input image content, namely, by the amount of the salient image features of the dominant image region relative to that of the outlier motion components, the distribution of the frequency components of the image, and etc. For the better understanding of this phenomenon, we constructed the global energy function map on the (u, v) plane, as shown in Fig. 2(b) ~ 2(f), where the descent direction of the energy map at the origin, i.e., at (u, v)= (0,0) informs us which direction in the u-v plane the initial estimate of zero motion will be updated toward. When the initial estimate $\mathbf{a_0}$ is set to zero motion, the NDIM energy map by equation (11) shows the descent direction of $\mathbf{n}^T$ = (-0.60,-0.79) toward the true motion (-1,-1) and the global minimum near the true motion. In the GDIM and MDIM maps, the erroneous descent direction of $\mathbf{n}^T$ = (0.23,0.97) and (0.69,0.72), respectively, heading in the opposite direction of the true motion is observed. However, when using an initial estimate $\mathbf{a_0}$ approximate to the true motion, for example, when using $\mathbf{a_0}$ = (-0.2,-0.2), both GDIM and MDIM maps also show the accurate descent direction of $\mathbf{n}^T$ = (-0.65,-0.75) and (-0.53,-0.84) at the origin, respectively. Note that the actual value for a velocity point in the u-v plane is the sum of its velocity in the map and the used initial estimate, i.e., $(u,v)_{actual} = (u,v)_{in\ the\ map} + \mathbf{a_0}$. Actually, the two algorithms succeed in recovering the dominant motion when using the initial estimate. This superiority of the NDIM algorithm in the initial updating process may be originated from the followings: The conspicuous disclosing of the latent image features in the local normalized image and the relatively low dimension of the unknown parameter space in the NDIM model.



(e)



(f)

Fig. 2 The global energy function maps for a pair of noise patterns

4. Hybrid hierarchical estimation strategy

A coarse-to-fine strategy [15] is required for recovering large image motion. Through a series of spatial filtering and sub-sampling of a pair of input images, image pyramids are constructed. At the coarse level, most of image noise components are significantly smoothed, but, at the full resolution level, the noise components are relatively predominant.

As demonstrated in the previous section, the NDIM model strongly drives the iterative parameter updating process successfully while, in the case of the GDIM and MDIM models, it is easy to fall into a local minimum state (causing the erroneous estimate). This superiority of the NDIM model relative to the remainder may be due to the following: First, since the requirement of the additional estimation for the local illumination variation factors in the GDIM and MDIM models increases significantly the dimension of the unknown parameter space, it can make the behavior of the initial update process even more unstable. Second, the conspicuously disclosing of the latent image features in the local normalized image can increase the amount of the matching features between two consecutive image frames.

Therefore, at the coarse level, where noise components are considerably smoothed, the application of the NDIM dynamic image model is recommended for acquiring a reliable motion estimate, which can harden a more accurate motion estimation in the subsequent iterative parameter updating process at the finer pyramid level. On the other hand, at the fine level, since the noise components appear strongly, especially the application of the NDIM model in the finest level can degrade the accuracy of the motion estimate a little relative to the case of the remaining models. Fortunately, once a reliable estimate has been given as a initial estimate, the GDIM or the MDIM model also can perform the good motion estimation more insensitively to noise components. Therefore, at the fine levels, we recommend the employment of the GDIM or the MDIM model. This is the essential of the proposed hybrid hierarchical estimation framework. Fig. 3 shows the diagram for an example of hybrid hierarchical estimation algorithm with three-level pyramids.

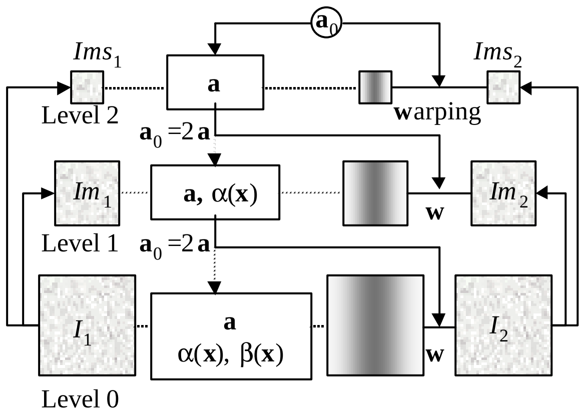
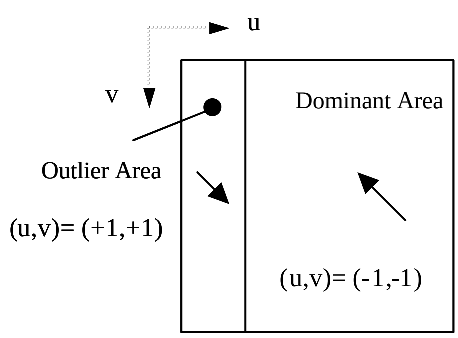


Fig. 3 Diagram of a hybrid hierarchical motion estimation framework with three level pyramids.

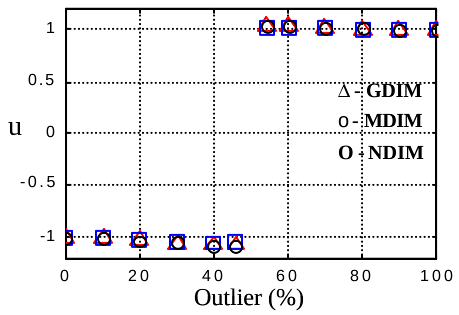
5. Experimental results

To begin with, in order to prove the feasibility of the three model-based algorithms for robust dominant motion estimation under local illumination variations, we have applied them to the synthetic images similar to the noise patterns shown in Fig. 2(a). However, these images contain both the dominant area with motion of (u,v) = (-1,-1) and the outlier area with motion of (u,v) = (+1,+1), as depicted in Fig. 4(a), they were generated synthetically as the percentage of the outlier area being increased from 0% to 100%. The initial estimate in the GDIM and MDIM algorithm was set to (-0.2, -0.2) and in the NDIM algorithm zero motion estimate was used. Fig. 4(b), 4(c), and 4(d) show the estimation results, where we can see that the GDIM, MDIM, and NDIM algorithms all recover the dominant motion successfully. In the case of the GDIM algorithm, the relatively large number of iterations due to the oscillation near the true value is observed, and the NDIM algorithm converges fast relatively to the remaining algorithms.

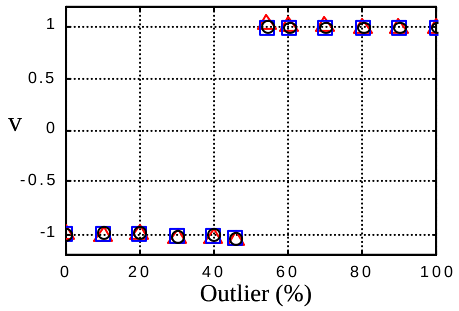
Next, the three algorithms and our hybrid algorithm, implemented in the same manner as depicted in the diagram of Fig. 3, have been tested on a real scene shown in Fig. 5(a), containing the dominant background image motion of (u, v) = (4,4) and the outlying motion of (-3,-3) corresponding to a synthetically added moving square box.



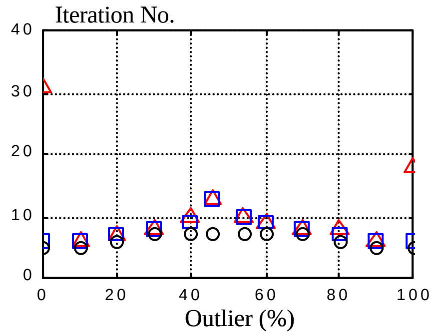
(a)



(b)



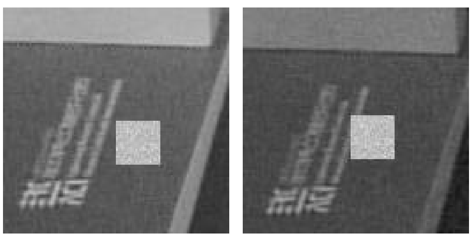
(c)



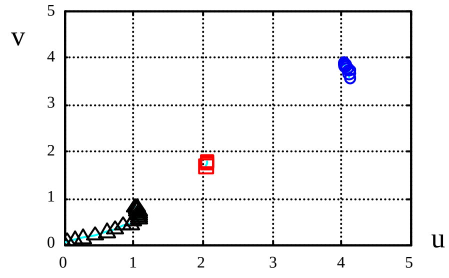
(d)

Fig. 4 The simulation results for robust dominant motion estimation on the synthetic noise patterens as changing the percentage of the outlier component in the images

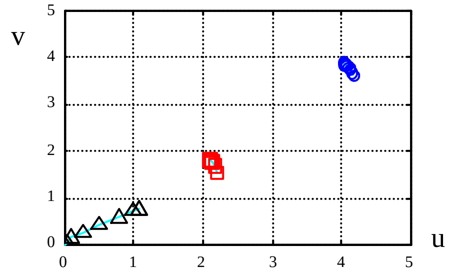
At each pyramid level, the iterative updating process was stopped when the motion estimate was converged or when $|u_{k+1} - u_k| \leq 1.e\text{-}3$ and $|v_{k+1} - v_k| \leq 1.e\text{-}3$, where k denotes the iteration number. The sequentially computed motion estimates by each algorithm are displayed on the *u*-*v* plane, as shown in Fig. 5(b), ~ 5(e), respectively. The resulting figures indicate that, as can see in Table 1, the hybrid algorithm recovers the final good estimate of $\mathbf{a}^T$ = (4.011, 3.922) fast than the others. In the case of the GDIM and MDIM algorithm, the initial motion estimate of (0.1, 0.1) was assigned.



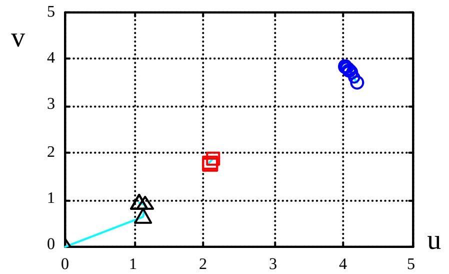
(a)



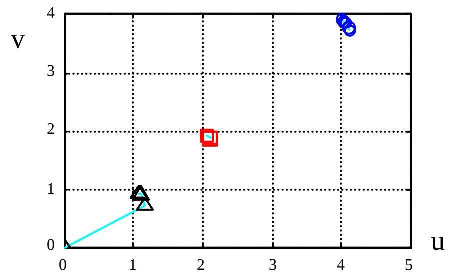
(b) by the GDIM



(c) by the MDIM



(d) by the NDIM



(e) by the Hybrid

Fig. 5 The estimation results on a pair of real images; Δ - at level 2, ❑ - at level 1, O - at level 0

6. Conclusions

We have presented a hybrid hierarchical estimation method for robust estimation of dominant image motion under local illumination variation, where the dynamic image models are employed selectively according to the pyramid level. Especially, since the superiority of the NDIM model in the initial updating process at the coarse level nearly proves the subsequent accurate motion estimation by the GDIM and MDIM model at the fine level, we could accomplish fast and accurate recovery of dominant image motion even under local illumination variation and the presence of the outlier motion. This hybrid hierarchical estimation strategy can be effectively extended to various vision problems, such as image matching, optical flow, and etc.

Table 1 The motion estimate and iteration number obtained by each algorithm at each pyramid level

Pyramid Level.	Algorithm			
	Hybrid	GDIM	MDIM	NDIM
2	(1.066,0.934) #6	(1.022,0.831) #31	(1.095,0.769) #9	(1.066,0.934) #6
1	(2.064,1.908) #3	(2.064,1.797) #5	(2.091,1.809) #7	(2.088,1.761) #3
0	(4.011,3.922) #7	(4.021,3.899) #10	(4.025,3.896) #10	(4.008,3.883) #10

References

[1] B.K.P. Horn, Robot Vision, MIT Press, Cambridge, MA, (1986)

[2] S. K. Nayar, K. Ikeuchi and T. Kanade, "Surface Reflection : Physical and geometrical perspective", IEEE Trans. PAMI , 13(7), July, (1991) 611-634

[3] B.T Phong, Illumination for computer generated pictures, *commun. ACM* **18**, 311-317

[4] S. Negahdaripour, "Revised Definition of Optical Flow: Integration of Radiometric and Geometric Cues for Dynamic Scene Analysis", IEEE Trans. PAMI , VOL. 20, NO. 9, SEPTEMBER, (1998) 961-979

[5] S. Negahdaripour, and C. H. Yu, "A Generalized Brightness Change Model for Computing Optical Flow, Proceedings of International Conference on Computer Vision (ICCV'93), Germany, (1993)

[6] R. Szeliski and J. Coughlan, "Spline-based image registration", IJCV, Vol. 22, No. 3, (1997) pp. 199-218

[7] S.H. Lai and M. Fang, "Robust and Efficient Image Alignment with Spatially Varying Illumination Models", In Proc. CVPR, (1999)

[8] G. Hager and P.N. Belhumeur, "Real-time tracking of image regions with changes in geometry and illumination", In Proc. CVPR, (1996) pp. 403-410

[9] M. J. Black, D. J. Fleet, Y. Yacoob, "A Framework for Modeling Appearance Change in Image Sequences", Sixth International Conf. on Computer Vision, ICCV'98, Mumbai, India, Jan., (1998) pp. 660-667

[10] M.J. Black and P. Anandan, "The robust estimation of multiple motions: Parametric and piecewise-smooth flow fields", CVIU, Vol. 63,No. 1, January, (1996) pp.75-104

[11] A. Papoulis , Probability and Statistics, Prentice Hall, Englewood Cliffs, NJ, 1990

[12] P.J. Rousseeuw and A.M. Leory, "Robust Regressions & Outlier Detection", J. Wiley & Sons, NY, (1987)

[13] H.S. Sawhney, S. Ayer, and M. Gorkani, "Model-based 2D & 3D Dominant Motion Estimation for Mosaicng and Video Representation " Proc. ICCV, (1995) pp. 583-590

[14] K.J. Hanna, "Direct multi-resolution estimation of egomotion and structure from motion", In IEEE Workshop on Visual Motion, Nassau Inn, Princeton, NJ, October, (1991) pp. 156-162

[15] J.R. Bergen, P.Anandan, K.J. Hanna, and J. Hingorani, "Hierarchical model-based motion estimation", In Second European Conference on Computer Vision, Santa Margherita Ligure, Italy, May, (1992) pp. 237-252