New Robust Global Motion Estimation Approach Used in MPEG-4*

Qi Wei[1]（祁卫）， Hong-Jiang Zhang[2]（张宏江）, Zhong Yuzhuo[1]（钟玉琢）

[1]Tsinghua University, Department of Computer Science and Technology, Beijing 100084

[2]Hewlett-Packard Labs, 1501 Page Mill Road, Palo Alto, CA 94304

Abstract

This paper presents a new robust global motion estimation method based on pre-analysis of the video content. The novel idea in the proposed method, compared to classical robust statistics-based estimation methods, is to classify the video sequences into 3 classes based on the analysis of scene content before motion estimation. Different motion models and estimation methods are applied to different classes of image sequences. As a result, outliers can be identified and removed from the dominant motion estimate to solve the problem of inaccurate initial descending direction estimates associated with classical global motion estimation methods. The pre-analysis of scene content is based on the STGS (Spatial Temporal Gradient Scale) images derived from the original image sequences. The extra computation time for STGS-image-based pre-analysis of scene content is negligible compared to the overall speed and accuracy improvement achieved with the proposed method. Evaluations based on extensive experiments have shown that the proposed method significantly improves the speed of robust global motion estimation methods (saving about 50% of the execution time of the classical methods).

Keywords dominant motion estimation; robust statistics; STGS image; GN method

The Global Motion Estimation (GME) technique is important for 'sprites' generation and object segmentation tasks involved in MPEG-4. This paper presents a new robust GME approach using the novel idea to pre-analyze scene content before motion estimation. The pre-analysis is based on STGS (Spatial Temporal Gradient Scale) images derived from the original image sequences. As a result, outliers can be removed from the dominant motion estimate at the initial estimation step. Compared with classical robust global motion estimators [1][2][3][4][5] and the current method described in Annex D of the VM (Verification Model), our approach can overcome the problem of inaccurate initial descending direction estimates and local minimums. Consequently, the execution time for global motion estimation is significantly reduced.

* Manuscript received: 1999-04- ; revised: 2000-07-05
Supported by the State High-Tech Developments Plan of China (No. 863-306-03-07)

2. Global motion estimate using robust statistical methods

Before presenting our proposed method, this section will review the robust statistics [6] methods and their application in motion estimation from image sequences. This review identifies the limitations of the robust estimation methods which have motivated development of the proposed method.

2.1 Robust estimation method for global motion

Given two images, their motion transformation can be modeled as

$$I(\boldsymbol{p}, t) = I(\boldsymbol{p} - \boldsymbol{u}(\boldsymbol{p};\ \theta), t-1) \tag{1}$$

where $\boldsymbol{p}$ is the 2D vector of image coordinates and $\boldsymbol{u}(\boldsymbol{p};\ \theta)$ is the displacement vector at $\boldsymbol{p}$ described by the parameter vector $\boldsymbol{\theta}$. The motion transform can be modeled as a simple translational model of two parameters, or a more complex affine model of six parameters. The affine model can describe motions involving camera operations such as zooming and rotation.

In the M-estimation formulation, the unknown parameters are estimated by minimizing an objective function of the residual error. That is,

$$\min_{\theta} \sum_{i} \rho(r_i; \sigma) \tag{2}$$

where r_i is the residual of the i'th image pixel,

$$r_i = I(\boldsymbol{p}_i, t) - I(\boldsymbol{p}_i - \boldsymbol{u}(\boldsymbol{p}_i;\ \theta), t-1) \tag{3}$$

The corresponding Geman-McLure function is then,

$$\rho_{GM} = \frac{r^2}{\sigma^2 + r^2} \tag{4}$$

Hence, the motion estimation task becomes a minimization problem for computing the parameter vector $\boldsymbol{\theta}$. There are many iterative descending algorithms to solve such a minimization problem, such as the Gauss-Newton (G-N) algorithm, SOR algorithm and the L-M algorithm used in current global motion estimation method in MPEG-4 VM. These algorithms first compute a descending direction, then computes the optimal increment step along this direction using a line search method. After some number of iterations, the motion parameters will converge to a set of values.

2.2 Problems of the robust estimation method in global motion estimation

Robust estimation methods may encounter outliers which are detected only using the absolute values of the residuals. However, in the initial estimation step, the *useful* feature information of the background that helps to correctly estimate the global motion may have nearly the same absolute value as that of the outliers in the foreground. Therefore, the simple absolute value-based assumption may mix the useful background information with that of the outliers. Consequently, the initial estimated descending direction could be very poor which will require a larger number of iterations to correct. The worst case is that the iteration could be trapped in a local minimum, which leads to failure of the algorithm. Some examples of these problems of the robust estimation methods are shown in Section 4.

To avoid the problems with the robust estimation methods, the outliers should be identified from the useful background objects in the initial estimation step when the descending direction is estimated. It is also desirable to obtain some knowledge of the scene structure of an image sequence, so that the most suitable motion model can be selected for the estimation. These two steps can significantly improve the performance of the robust estimation methods. The STGS-image-based method presented in the following section was developed to address these needs.

3. A new motion estimation method based on STGS-image analysis

Figure 1 illustrates the entire process for the proposed STGS image-based robust global motion estimation method. The difference between the proposed method and other robust motion estimation methods is that a STGS-image analysis is performed to obtain information on the scene structure of the input image sequence before performing the robust motion estimate. The scene structure is analyzed to identify the outliers in the initial estimate and to select the most suitable motion model for the estimate.

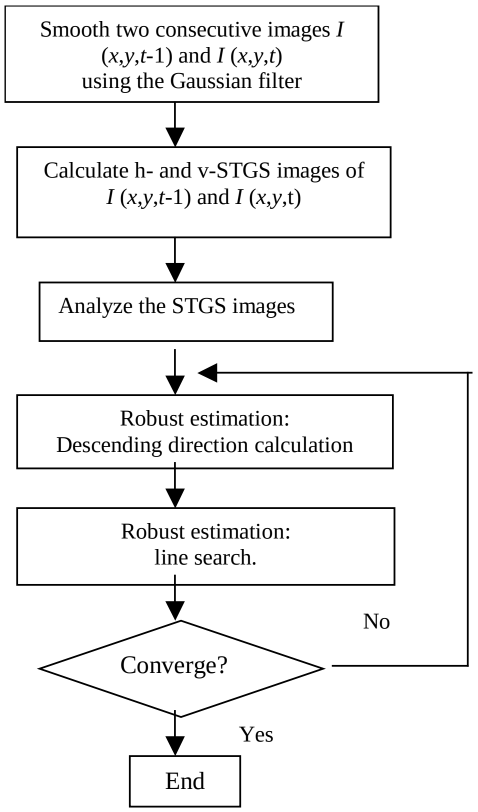


Fig. 1 Framework for the STGS Image-based Robust Global Motion Estimation method

3.1 Definition of STGS-images

STGS-images have two basic elements: the horizontal image (referred as the h-STGS-image) and the vertical image (v-STGS-image). For two consecutive images $I(x, y, t-1)$ and $I(x, y, t)$, the two component images of STGS are defined as:

$$S_{\mathrm{h}}(x, y) = \frac{I_{\mathrm{t}}(x, y)}{I_{\mathrm{x}}(x, y)} \tag{5.1}$$

$$S_{\mathrm{v}}(x, y) = \frac{I_{\mathrm{t}}(x, y)}{I_{\mathrm{y}}(x, y)} \tag{5.2}$$

where $I_{\mathrm{t}}(x, y)$ is the temporal gradient of the pixel (x, y) between the two original images, and $I_{\mathrm{x}}(x, y)$ and $I_{\mathrm{y}}(x, y)$ are the horizontal and the vertical spatial gradients of the pixel (x, y) in the image $I(x, y, t)$. The name STGS-image comes from the fact that they are defined by the scales of the temporal and spatial gradients of a pair of images.

Theoretical analysis of a STGS-image comes directly from the basic constraint equation of optical flow

$$I_{\mathrm{x}}u + I_{\mathrm{y}}v + I_{\mathrm{t}} = 0 \tag{6}$$

which can be rewritten as

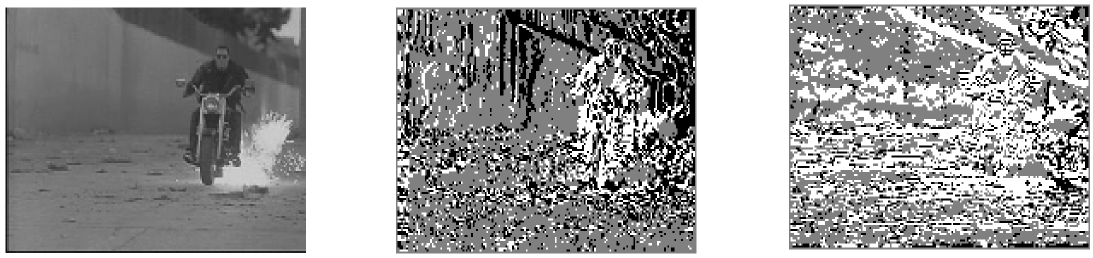
$$u = -\frac{I_{\mathrm{t}}}{I_{\mathrm{x}}} - \frac{I_{\mathrm{y}}}{I_{\mathrm{x}}}v \tag{7}$$

When the two consecutive images have only horizontal motion, the v component in the equation is zero, so the horizontal motion component u is:

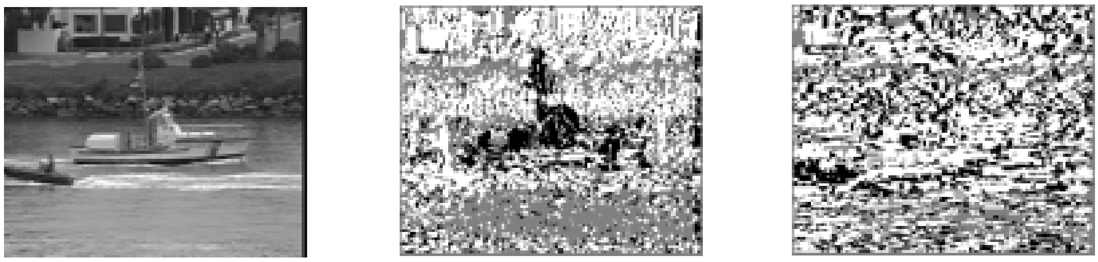
$$u = -\frac{I_{\mathrm{t}}}{I_{\mathrm{x}}} \tag{8}$$

which is just the negative of the (x,y) pixel value of the h-STGS image in (5.1). A similar relationship can be derived between the vertical motion component v and the v-STGS image. Although STGS-images can accurately represent the motion only when the motion between two images is purely one-component translation and the image function has the ideal monotonic form in a local area, they also carry some information of motions between two images that is useful in motion estimation. Therefore, STGS-images can be used as a qualitative measure for pre-analysis of the video content, especially for classification of global motions in typical background-foreground structured scenes. Compared with the optical flow-based pre-analysis approach [7], an important advantage of using STGS-images is that the computational cost to obtain a STGS image is very low.

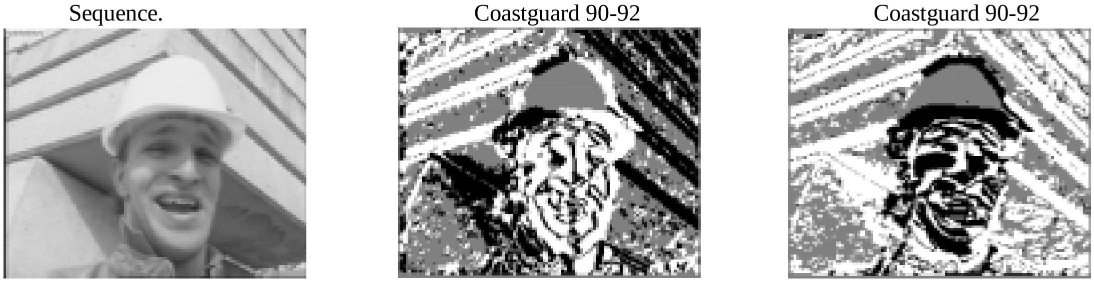
Figure 2 show some examples of STGS-images derived from image pairs with global motion. The image pairs were selected from the MPEG-4 test sequences Bike (frame 100 and 102), Coastguard (frame 90 and 92) and Foreman (frame 15 and 17) in QCIF format. All these images have the typical background-foreground structure and show global motion. Most of the global motion could be represented by a 2-parameter translational model (u,v).



a1 image 100 of Bike Sequence   a2 h-STGS image of Bike 100-102   a3 v-STGS image of Bike 100-102



b1 image 90 of Coastguard Sequence.   b2 h-STGS image of Coastguard 90-92   b3 v-STGS image of Coastguard 90-92



c1 image 15 of Foreman Sequence.   c2 h-STGS image of Foreman 15-17   c3 v-STGS image of Foreman 15-17

Fig.2 First images of three image pairs from the Bike, Coastguard and Foreman sequences and their associated STGS images

The STGS-images shown in Figure 2 are in fact the 'sign' versions of STGS-images as defined by (5). That is, the images shown here were produced based only on the sign of the STGS-image pixels. For the h-STGS-images, if the value $I_t/I_x > 0$, the pixel value is set to 255 (white) to produce the images in Figure 2. If the value is negative, the pixel value is set to 0 (black), while if the value equals zero (including the situation that $I_x = 0$), the pixel value is set to 128 (gray). Therefore, the sign STGS-images represent the direction of motion between two images. According to equation (8), if the motion is towards the right or down, then the signs of the h-STGS-images and the v-STGS-images are negative, if the image coordinate origin is at the up-left corner. Similar, if the motion is towards the left or up, then the signs of the h-STGS-images and the v-STGS-images are positive.

All the image pairs in Figure 2 have two distinct motions. One is the global motion of the background and the other is the motion of the foreground. If the two motions differ in either the horizontal or vertical directions, the rough shape of the foreground can be obtained from the background area by judging only the signs of the pixels of a STGS-image. In cases where the horizontal motion dominates, such as in Coastguard 90-92, the h-STGS-images show clear segmentation of the foreground. Similarly, in cases where the vertical motion dominates, such as in Foreman 15-17, the v-STGS-images also indicate clear segmentation of the foreground. For images with two competitive components of translational motion, such as in Bike 100-102, the h-STGS-image is not very clear; however, one still can observe some foreground shape from the image. These examples and the definition of STGS images show that since we only use the signs of the pixels, the STGS sign image provides a more robust criterion for performing the pre-segmentation of the motion field.

3.2 Using STGS-images to solve the problems of classical robust estimation

As stated in section 2, a major problem of robust estimation methods is that the initial estimation step often confuses the outliers of the foreground and the useful feature information of the background. The outliers are not effectively removed from the global motion estimation process, since the detection of outliers in robust estimation methods is based only on the absolute values of the residuals. Apart from being used as an indication of motion direction, we can also obtain a pre-segmentation mask from the STGS-image by post-processing the original STGS-images by applying median and/or morphological filters. Such pre-segmentation masks can be used to eliminate outliers in motion estimation and improve the performance of motion estimation in terms of both speed and accuracy, as described in the next subsection

The main objective of the analysis of the STGS-image is to improve the performance of the traditional robust global motion estimation, rather than to detect the exact foreground shape. Therefore, the focus is to detect the areas that make strong negative contributions to the descending direction estimate, especially in the initial step. Removing outliers overcomes the inaccurate initial estimation problem in traditional robust estimation methods.

3.3 STGS-image analysis method for global motion estimation

A method is proposed for analyzing STGS-images to build outlier masks based on two important assumptions:

1) Video sequences have typical background-foreground structured scenes where the background is dominant in the sequences and the foreground consists of connected rigid bodies.
2) Background motions are caused only by camera operations.

These two assumptions are realistic for the current motion estimation problem. Based on these two assumptions, pixels that are in the minority in terms of pixel number relative to the size of an image can then marked as outliers. The key idea of the proposed method is to perform the outlier pixel detection by identifying minority pixels from STGS-images and to then classify the images according to the ‘size’ and 'distribution' of the minority pixels. For example, we need only to identify which of the 'black' and 'white' points in a STGS-image shown in Figure 2 is in the minority according to pixel count and then performing the following analysis based on the 'size' and 'distribution' of the minority pixels.

By judging the 'size' and 'distribution' features of the minority points, the STGS-images can be classified into three classes: 'good', 'negligible' and 'complex' images. The definition of these three classes of images are based on the size and distribution:

good, if (*size* = *large* and *distribution* = *compact*)

neglect, if (*size* = *small*, and *distribution* = *compact*)

complex, if (*size* = *small, distribution* = *sparse*)

or (*size* = *large, distribution* = *sparse*)

A STGS-image is 'good' if it contains a relatively large and compact area of minority pixels. The large and compact area could correspond to the foreground and could be used as an outlier mask in the successive robust estimation. A.2, b.2 and c.3 in Figure 2 are examples of good STGS-images. A STGS-image is 'negligible' if the foreground area is small or it shows the same motion direction as the background. A negligible STGS-image implies that the foreground areas in the original image sequence will not have any significant negative influence on the global motion estimate; thus, the sequences can be handled by the classical robust estimation method . In other words, this pair of images will generate a negligible STGS-image because the global motion between the image pair has a very strong translational component. In this case, the two-parameter motion model is chosen as the initial model for the robust estimation and an outlier mask is not needed in the initial estimation step. Apart from good and negligible, the rest of the STGS-images all belong to the 'complex' type. If the global motion between an image pair results in a complex STGS-image, then, the motion has strong zooming or rotating components which violate the translational assumption for generating a meaningful STGS-image. Therefore, foreground area detection cannot be achieved directly from a complex STGS-image. For image pairs resulting in complex STGS-images, the classical 6-parameter motion model must be used as the initial model when applying the classical robust estimation method without an outliers mask in the initial step. The assumptions are that typical background-foreground structured scenes will generate good STGS-images. Therefore, our proposed method of applying STGS-image analysis for global motion estimation has real merit.

The STGS-image analysis process consists of the following steps:

1) Count the 'white' and 'black' pixels in both the h- and v-STGS-images. Select the type of pixels which are in the 'minority' in number;
2) Calculate the 'size' and 'distribution' of the minority pixels in both the h- and v-STGS-images;
3) *If* there is a *good* STGS-image, *then*

   use it as the input outlier mask for the robust estimation with the 2-parameter motion model;

   Else, If there is a negligible STGS-image,

   Then, choose the classical robust estimation method with the 2-parameter motion model;

   Else, choose the classical robust estimation method with the 6-parameter motion model.

Since the classification of STGS-images into the three classes relies on the 'size' and 'distribution' of the 'minority' areas, a quantitative measure is needed for these two features. The ratio of the number of pixels belonging to the minority area relative to the overall size of the entire image is the measurement of *size*. To measure the 'distribution', the isolated pixels are removed by applying a median filter and then the variance of the remaining pixels is calculated:

$$vaiance = \frac{\sum_{x,y \in S} \left((X_{\text{center}} - x)^2 + (Y_{\text{center}} - y)^2 \right)}{N} \tag{9}$$

where N is the number of 'minority' pixels after filtering and S is the coordinate set of the 'minority' pixels after filtering. X_{center} and Y_{center} are the coordinates of the center of 'gravity' of the pixels. This is a very simple measure of pixel distribution. However, because of the typical background-foreground structured scene assumption that the foreground is a connected rigid body, the *variance* is sufficient to distinguish a compact distribution from sparse distributions.

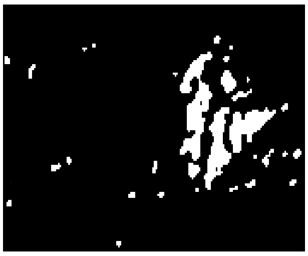
To classify the STGS-images, two thresholds, T_{size} and $T_{\text{distribution}}$, need to be set for the 'size' and 'distribution' measurements. In the classification process, an STGS-image is classified as a negligible image when $size < T_{\text{size}}$ and as a complex image when $variance > \mathrm{T}_{\text{distribution}}$. Hence, a STGS-image is classified as a 'potentially good' image when it does not meet the two conditions. Two threshold values, T_{size}=5%, and $T_{\text{distribution}}$ =2800 have been selected using a simple experimental process, although sophisticated analysis could be performed to obtain the optimal thresholds. If there is only one potentially good image among the two h- and v- STGS-images, then it is identified as the good image. If both the h- and v-STGS-image are classified as potentially good, further selection is needed to determine which one is selected as the outlier mask image. Considering both the size and distribution of minority pixels gives a *ratio* defined as $ratio = variance / size$. The image with the smaller value of *ratio* is selected as the final good image used as the outlier mask.

Finally, the analysis of the STGS-images is very fast since it involves very few computations. Moreover, both the spatial gradient and the temporal gradients calculated in the STGS-image analysis could be re-used in the subsequent gradient based descending method, which is an important advantage of the proposed method as shown with experimental data in Section 4. The worst penalty of possible misclassification by the proposed method is to fall back to the classical robust estimation method, costing only a little extra time for analysis of the STGS-image. Therefore, applying the proposed method guarantees improved total performance compared to classical robust estimation methods.

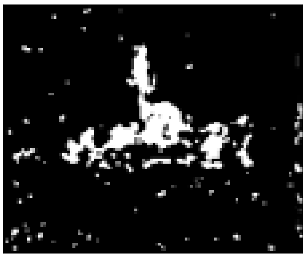
4 Experimental evaluation

The STGS-image based pre-analysis method can improve the initial descending direction estimate of the classical robust estimation methods for motion segmentation. This section presents experimental results showing that, in comparison with the classical Gauss-Newton robust estimation method, the proposed method can speed up the estimation process and reduce the possibility that the robust estimation falls into a local minimum.

For the experiment, the three pairs of consecutive frames selected from Bike (frame 100 and 102), Coastguard (frame 90 and 92) and Foreman (frame 15 and 17) in QCIF format were used as in section 3.1. Figure 3 shows the median filtered minority area for the three pairs of STGS-images shown in figure 2 using the method presented in section 3.3. The filtered h- and v-STGS image have some 'good' images that show the rough shape of the foreground, such as a.1, b.1, and c.2, which could be used as outlier masks in the subsequent motion estimation.

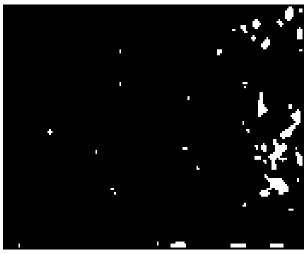


a1 For h-STGS image in figure 2(a.2)



b1 For h-STGS image in figure 2(b.2)

c1 For h-STGS image in figure 2(c.2)



a2 For v-STGS image in figure 2(a.3)

b2 For v-STGS image in figure 2(b.3)

c2 For v-STGS image in figure 2(c.3)

Figure 3 Filtered minority points from the STGS images in Figure 2

Table 1 shows the STGS analysis results for the three image pairs. Using the threshold introduced in section 3.3, the 'good' h- or v-STGS-images were labeled by * in Table 1. All three pairs have strong two-component translational global motion; thus, the outlier mask based 2-parameter model method was chosen.

Table 1 *Variances*, *sizes* and *ratios* of h-STGS and v-STGS for the image pairs from sequence Bike, Coastguard and Foreman.

Sequences	h-STGS			v-STGS		
	variance	*size*(%)	*ratio*	*variance*	*size*(%)	*ratio*
Bike100-102	1116	8	140*	953	2	477
Coastguard 90-92	1455	7	*	3442	3	-
Foreman 15-17	3560	23	-	1263	15	*

Table 2 shows the estimated motion parameters from the two motion estimation methods respectively. Our method can achieve a somewhat higher accuracy than the classical GN robust estimation method.

Table 2 Estimated two-component translational global motion parameters for the GN robust method and our method.

Sequences	Classical GN Robust Method	Our Method	Ground True
	(u,v)	(u,v)	(u,v)
Bike 100-102	(1.25,-0.91)	(1.38,-0.87)	(1.36,-0.82)
Coastguard 90-92	(-0.75,-0.32)	(-0.82,-0.44)	(-0.82,-0.37)
Foreman 15-17	(-0.24,-0.66)	(-0.25,-0.63)	(-0.25,-0.66)

The relative improvement in computation time is shown by setting the computation time for the classical G-N method to 1 and calculating the relative computational time for our motion estimation method. Note that in our method, we also use the robust G-N descent method after the STGS-image analysis; thus, there is additional computation time spent for the STGS-image analysis process. The execution time for our method is listed in the table 3 for comparison of the two approaches. We can see clearly from Table 3 that the STGS-image analysis only adds on a very small portion of computational load (maximum 6.9% in the three test sequences), and the total time of applying our method is less than 50% of that of applying the classical GN method. The time saving comes from two factors. First, by removing the outliers that make negative contributions to the final solution, the initial estimated descending direction is more accurate than that of the classical method; hence the number of iterations in the motion estimate is reduced. Second, since the outlier areas in each frame are marked out¸ the overall image area in each frame used for the motion estimate in each iteration is reduced, resulting in further reduction in computation time.

Table 3 Computation time of our method relative to that of the classical robust GN method.

Sequences	Time for Robust Estimation	Time for STGS image Analysis	Total Time
Bike 100-102	16.5%	3.4%	20.0%
Coastguard 90-92	28.4%	6.9%	35.3%
Foreman 15-17	43.3%	5.7%	48.9%

Table 4 shows the difference between the initial estimated descending direction resulting from the two approaches. Comparing the data in Table 4 with the final motion estimation data shown in Table 2 shows that the initial estimated descending direction obtained with our method is much closer to the final results of the motion estimate than that obtained with the classical estimation method. For Coastguard 90-92, the motion parameters converged to the final result on the very first iteration. In contrast, the initially estimated descending direction from the classical robust estimation method on the same sequence is far from final converged point. The initial horizontal directions for the Bike and Foreman sequences obtained using the classical method are even reversed from the final result. Obviously, these initial errors will take computational time to correct. As discussed in section 2.2, this problem of the poor initial estimate associated with the classical robust estimation methods stems from the fact that these methods have very poor capability to distinguish outliers from the dominant motion in the initial estimation stage. Solving this problem is the major motivation of the proposed STGS-image-based pre-analysis method. The experimental results show that the STGS-image pre-analysis method is indeed very effective in overcoming this problem associated with the classical robust estimation method.

Table 4 Estimated motion parameters obtained by our method and the classical robust GN method after the first iteration.

Sequences	Classical GN Robust Method	Our Method
	(u,v)	(u,v)
Bike 100-102	(-0.13,-1.07)	(1.09,-1.05)
Coastguard 90-92	(-0.14,-0.43)	(-0.82,-0.44)
Foreman 15-17	(0.02,-0.52)	(-0.05,-0.59)

The worst situation in motion estimate using the classical robust estimation methods is that an inaccurate initial descending direction can lead the iteration to a local minimal. Table 5 shows several cases of motion estimates where the global motion parameters estimated by the classical robust NG estimation method converged to a local minimum that is 'far' from the true motion parameters. In contrast, the results in table 5 show that our proposed method is able to analyze such situations and produce the correct estimate.

Table 5 Converged local minimum global motion parameters obtained by the classical robust GN method, estimated global motion parameters obtained by our method and the "true" global motion parameters.

Seqeunces	Classical GN robust method	Our method	Ground truce
	(u,v)	(u,v)	(u,v)
Bike 130-131	(0.00, 0.00)	(1.73, -0.06)	(1.68, -0.02)
Coastguard 100-104	(-0.03,-0.30)	(-2.89, -0.20)	(-2.84, -0.20)
Foreman 17-19	(-0.03,-0.21)	(-0.57, -0.59)	(-0.54, -0.56)

Finally the speed improvement of our method applied to three entire test sequences is shown in Table 6. The computation time is the ratio of our method to that the traditional robust estimation method. The computational time with our method is much less than that of the robust motion estimation for most of the test image pairs. Overall, the computational time for our method is only 54.7 % of that of the classical robust estimation method. The reduced computational time comes from applying the pre-analysis of the STGS-images before applying the classical robust estimation method. Our global motion estimation method first uses the two-parameter model and then uses the six-parameter affine model to refine the results. Since the two-parameter model can approximate the global motion well, we can get the fine initial value and foreground mask for the subsequent six-parameter model after obtaining a converged two-parameter estimate. Therefore, there is much less work left for the six-parameter model estimate. For the motions such as pure rotation and zoom that can not be approximated by the simple two-parameter model, the six-parameter model is used initially as described in section 3.3.

Table 6 Performance of the proposed method compared with that of the classical robust GN method for three test sequences of Bike, Coastguard and Foreman. Overall average computation time is 54.7 % of the classical method.

Bike	Execution time(%)	Foreman	Execution time(%)	Coastguard	Execution time(%)
90-91	67.0	0-1	45.6	80-81	49.3
91-92	67.0	1-2	37.6	81-82	39.7
92-93	42.6	2-4	69.7	82-83	62.7
93-94	41.4	4-6	69.7	83-84	37.2
94-95	53.5	6-8	60.0	84-85	105.6
95-97	46.5	8-10	76.0	85-86	104.2
97-98	25.5	10-12	66.7	86-87	105.4
98-99	*	12-14	72.2	87-88	*
99-100	28.3	13-15	63.9	88-89	35.4
100-102	20.0	15-17	48.9	89-90	40.8
102-103	26.7	17-19	*	90-91	48.0
103-104	*	19-21	61.3	91-92	48.9
104-105	33.8	21-23	74.2	92-93	41.4
105-106	45.2			93-94	40.8
106-108	*			94-95	40.5
108-109	*			95-96	48.6
109-110	*			96-97	46.9
110-111	*			97-98	43.8
111-113	*			98-99	35.0
113-114	*			99-100	*
114-115	54.3			100-102	58.7
115-116	48.0			102-103	*
116-118	*			103-104	107.3
118-119	*			104-105	105.2
Average	42.8		62.1		59.1

Furthermore, there are a number of cases, labeled by * in Table 6, where the classic method moved into local minimums while our method converged to correct values. This indicates that our method is more robust than the classical method.

There are several cases in the Coastguard sequence, where for some pairs of frames the computation time was over 100%. Most such cases resulted from the fact that the foreground boat had the same motion direction as the background, so that the corresponding STGS were classified as 'negligible' as described in section 3.3. Therefore, the classical motion estimation method and the translational motion model were applied directly without the information about the outliers. The extra computation time over the classical method in these cases was the cost for the STGS-image analysis, which was very small.

5 Conclusions

This paper presents a new global motion estimation method that improves the performance of the popular robust estimation methods for background-foreground structured image sequences. Evaluations based on extensive experiments have shown that the proposed method significantly improves the speed and accuracy of robust global motion estimation methods.

Global motion estimation is an essential step in object segmentation and motion compensation used in object-based coding. Development of a general method to deal with motion estimates for all types of video sequences is not feasible. The work presented in this paper has focused on improving the performance of global motion estimation methods suitable for image sequences with typical background-foreground structured scenes. There are still some limitations in the proposed method which require further research. One of the future tasks is to develop a method to determine the threshold for obtaining the foreground object mask and to refine the object shape in the motion estimation process.

References

[1] Black M J, Anandan P, Robust Dynamic Motion Estimation Over Time. In: Proc. Computer Vision and Pattern Recognition (CPVR 91). Maui: HI, 1991: 296-302.

[2] Black M J. The robust estimation of multiple motions: parametric and piecewise-smooth flow fields. Computer Vision and Image Understanding, 1996, 163(1): 75-104.

[3] Sawhney H, Ayer S. Compact representation of videos through dominant and multiple motion estimation. IEEE Trans. Pattern Analysis and Machine Intelligence, 1996, 18(8): 814-830.

[4] Darrell T, Pentland A. Coopeative Robust Estimation Using Layers of Support. IEEE Trans. Pattern Analysis and Machine Intelligence, 1995, 17(5): 474-487.

[5] Salembier P, Sanson H, Robust Motion Estimation Using Connected Operators. In: Proc. International Image Processing (ICIP 97), 1997: 77-80.

[6] Huber P J. Robust Statistics. New York: Wiley, 1981.

[7] Wang J Y A, Adelson E H. Representing Moving Images with Layers. IEEE Trans. Image Processing, 1994, 13(5): 625-638.