Journal of Imaging Science and Technology® 53(1): 010502–010502-6, 2009.

Corresponding Metadata-Based Stereo Object Tracking System Using Disparity-Motion Estimation

Kyung-Hoon Bae▲, Changhan Park and Jong-Eon Lee
Samsung Thales Co., LTD, Chang-Li 304, Namsa-Myun, Cheoin-gu, Yongin-City, Gyeonggi-Do, Korea 449–885
E-mail: khbae.bae@samsung.com

Abstract. *In this article, a corresponding metadata-based stereo object tracking system using disparity-motion estimation (DME) is proposed. In the proposed method, object tracking is performed based on robust characteristics of an object, such as its motion, shape, and color. Elements of metadata and the optimal tracking method are selected at each frame. Because the characteristics of an object are affected by the camera position and by environmental conditions, a tracking method between individual cameras is selected by means of the optimal metadata. Additionally, the time-sequential disparity motion vector (DMV) can be estimated from the feature-based DME, which is extracted from the sequence of the stereo images, and metadata sequences. Using the DMV, the area where the target object is located and its location coordinate are then detected from the stereo images. In an experiment, the proposed system is shown to be robust for nonrigid objects and demonstrates its ability to track a target object adaptively with an average low error ratio of approximately 2.37% between the detected and actual location coordinates of the target object. © 2009 Society for Imaging Science and Technology.*
[DOI: 10.2352/J.ImagingSci.Technol.(2009)53:1(010502)]

INTRODUCTION

The study of tracking and recognition through analysis of a moving object in a video has gained an increasing amount of attention in many areas, including three-dimensional (3-D) broadcasting, virtual reality, special effects, image composition, human computer interaction (HCI), video surveillance,[1] and autonomous mobile robots. Object tracking is one of the most popular applications of computer vision and image processing because it serves as a basic framework for video analysis applications such as intelligent transportation system, public safety, and robot vision, object-based video coding, among others.[2,3] There have been various studies related to video-based object detection and tracking. Haritaoglu et al. proposed a combination of shape analysis and the tracking of body parts to create models of human appearance and related solutions through interactions such as occlusions.[4] Wren et al. proposed a real-time blob tracking algorithm using a human body model.[5] For a more robust analysis of shape-based object tracking, Baumberg proposed the *a priori* shape-based tracking of an object of interest, in which a trained shape is projected onto the closest shape in a given frame.[6] In a typical video surveillance system, video object tracking using a fixed camera has a limited field-of-view (FOV).[7] Therefore, the object of interest can easily be lost if it leaves a particular camera view. This problem necessitates various modifications and extensions of existing surveillance techniques.

This article presents a metadata-based continuous video tracking system using stereo cameras.[8] In the proposed method, a number of simple statistical features are extracted from images that correspond to "metadata" that indicate the apparent position of an object.[9] These features represent features such as the shape, motion, color, size, and ratio of the object. The proposed system has five steps. The first step involves the creation of the best metadata with which the proposed video tracking can continuously track an object over a wide area using metadata. The second step involves defining a disparity-motion vector using feature-based-disparity estimation and corresponding metadata sequences, the third step involves disparity-motion vector (DMV) regularization, the fourth step involves detection of a moving object and its location coordinates, and the fifth step involves the pan/tilt control of the camera for object tracking, which always centralizes the moving object on the FOV of the camera. In general, the basic task of stereo image processing is to find a target object through the 3-D information of the stereo image. That is, if 3-D information can be extracted from the stereo image, it can be used not only for the detection of a target object, but also to obtain its location coordinates. Therefore, from this position information, the moving distance of the target object between consecutive image frames is calculated and the target object is finally tracked by controlling the pan/tilt embedded in the stereo camera system using this moving distance signal. A number of experiments with a regularized DMV-based object tracking system (OTS) using nine frames of metadata were carried out. Additionally, a performance analysis is presented.

DMV FOR OBJECT TRACKING USING METADATA

In summary, the proposed feature-based disparity-motion estimation (DME) based OTS using corresponding metadata, consists of five steps: metadata creation, disparity estimation, disparity regularization, detection of a moving object, and its location coordinates and object tracking.

▲IS&T Member.

Received Jun. 16, 2008; accepted for publication Dec. 6, 2008; published online Feb. 5, 2009.
1062-3701/2009/53(1)/010502/6/$20.00.

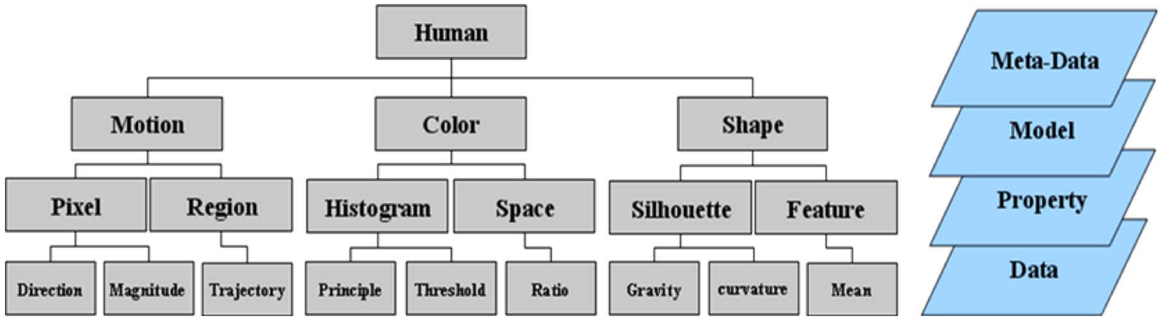

Figure 1. Hierarchical classification of object's descriptors.

```
xml.SetDoc( "<?xml version=₩"1.0₩" encoding=₩"ISO-8859-1₩"?>₩r₩n" );
<ObjectSegment id=" Obj1" >
<VideoTime><numOfFrame>122</numOfFrame></VideoTime>
<MotionDescriptor Property= "Pixel"  type=" Dir" >
<Dir>-1.245663 0.8578824 </Dir>
</MotionDescriptor>
<ColorDescriptor Property= "Histogram"  type= "HSI" >
<Hue>18.44 19.74 ··· 35.06 </Hue>
</ColorDescriptor>
<ShapeDescriptor Property= "Silhouette"  type=" SValue" >
<SValue>120 25 120 117 165 25 165 117</SValue>
</ShapeDescriptor>
</ObjectSegment>
```

Figure 2. Generation of metadata using XML.

Step 1: The Corresponding Metadata Creation

The three main elements of architectural description include the image data, model, parameters, and metadata of the object. Metadata can be considered simply as data about data. This section presents a metadata creation method using intensity, motion, color, and shape. Metadata consists of descriptors, an attribute, and a value. Descriptors define a tracking model, the attribute represents a feature in tracking data, and the value also represents tracking data in term of tracking computation. Figure 1 summarizes a hierarchical classification of the description of an object at four different levels: (i) metadata, (ii) model, (iii) property, and (iv) data.

Generalization is a mechanism that is used to factor out the common features of an object and defining these features in separate, shared metadata. The data defining the common features are typically termed a model, and a model inheriting these features is commonly known as metadata. Metadata can be found at different levels in an object representation of the environment. At the lowest level, object tracking data may be found by describing the special property of a single database. At the next level, an object tracking model is used to describe the different databases within the system. Tracking performance both in the real world and in the model is not promising, as it undesirably predefines the values and properties of the object. Most metadata-based approaches use the term "meta" to describe the level of a model and its elements. The use of the term "meta" can be justified in an absolute sense because all elements at a given level have the same number of meta prefixes. In other words, one can expect all elements of a meta-model to be meta-elements. However, this is not a viable example of providing by attributes.[10] Figure 2 show the metadata generation process using an XML tag.

Step 2: Disparity-Motion Estimation

In this article, the bi-directional matching algorithm is used to determine a correct matching point. The corresponding point (FV) that does satisfy Eq. (1) at the given threshold Th_N is set to the process of bi-directional matching. However, the corresponding point that does not satisfy Eq. (1) at the given threshold Th_N is set to the occluded region.

$$|d_l(fv) + d_r(\overline{fv})| < Th_N. \tag{1}$$

Via the bi-directional extraction of the FV algorithm, the FV can be detected from the edge information of an object, which is extracted from the input stereo images. Here, the FV is a discontinuous point that is displayed using the value of the pixel brightness depending on the characteristics of the input stereo image. In this paper, Canny edge-detection,[11] which is comparatively less sensitive to noises than other algorithms, is used to extract the FV from stereo input images pair. In the Canny edge-detection method, the outputs of two Gaussian derivative masks are used.

In this paper, feature-based disparity estimation (DE) is used as a DE method; once one of the metadata pairs is divided by a pixel unit, each pixel in this image (left image) undergoes a matching process with that of the other image (right image) by means of a cost function. Here, the minimum mean square error (MSE) function is used as a cost function rather than as the minimum mean absolute difference (MAD) to obtain a more accurate disparity vector (DV), although this takes some extra computation time.[12] Equation (2) shows the MSE function as used for disparity estimation in this paper, where m, n denotes the size of the block and $I_L(m,n)$ and $I_R(m+i,n+j)$ denote the coordinates of the left image and its corresponding coordinates of the right image, respectively.

$$\mathrm{MSE}(i,j) = \frac{1}{m \times n} \sum_{m=1}^{m} \sum_{n=1}^{n} |I_L(m,n) - I_R(m+i,n+j)|^2. \tag{2}$$

Essentially, once the left image is divided by the block with $N_x \times N_y$ pixels, for each block the cost function of Eq. (2) is calculated one at a time together with the corresponding block of the right image, as the block is shifting within the given search range. Finally, a block that minimizes the cost function among the possible values is determined as the corresponding matching block of the right image to the selected block of the left image. The position difference between the blocks in the right and left images is then assigned as the disparity value. These disparity values for each block form a DV, which is also known as a disparity map.

Using the motion vector (MV)s extracted from *Step 1*, the metadata-based DMV, which is known as the motion difference between the time-sequential metadata sequences, can also be extracted. Here, the dynamic relationship between the MVs of the T-1 and T frames is highly similar to that between the images of the T-1 and T frames in the conventional 2-D video sequence. In other words, some moving information of the target object can be obtained through the motion difference between the time-sequential metadata sequences. Using this metadata-based DMV, the difference between the MVs of the T-1 and T frames, the potential area where the moving object might be located in each frame, can be detected. From the resulting metadata-based DMV, it was found that the regions where a target object was located before moving and where a target object is located after moving have a relatively large change of motion values between the T-1 and T frames, whereas the background region mostly shows no change in the motion value between the T-1 and T frames.

Step 3: Disparity Regularization

Some regions in each image have similar DMVs; however, but in the matching process, though vertical vector similarity exists, only the similarity in the horizontal direction is considered. Thus, considering the vertical vector similarity, it is possible to remove the false DMVs. Therefore, the DMVs of these regions are substituted with the mean values of the DMVs of the nearby regions. The best means of disparity regularization is vector smoothing inside the segmented object in the image.[13] However, image segmentation is not highly reliable and requires a heavy computational load; therefore, the spatial smoothing filter in the vertical direction is used. In this article, a disparity estimation algorithm employing a neighborhood averaging-based regularization scheme is used to alleviate the problems of matching window overlapping and misallocation that can occur in conventional disparity estimation.[14,15] Here, the regularized pixel $g(x,y)$ is defined by Eq. (3):

$$g(x,y) = \frac{1}{M} \sum_{(i,j) \in s} f(i,j). \tag{3}$$

In this equation, M and S are the number and set of neighborhood pixels $(i \times j)$, respectively.

In addition, Eq. (4) is used to preserve the edge value in the edge region.

$$g(x,y) = \begin{cases} \dfrac{1}{M} \sum\limits_{(i,j) \in s} f(i,j); & \mathit{if} \left| f(x,y) - \dfrac{1}{M} \sum\limits_{(i,j) \in s} f(i,j) \right| < T \\ f(x,y); & \mathit{otherwise} \end{cases} \tag{4}$$

Mainly, if the difference between the regularized and original pixel value $f(x,y)$ is smaller than a predetermined threshold, the original pixel value is replaced with a regularized pixel. Otherwise, by using the original pixel value, the problems of over regularization can be solved. Here, the threshold value used in this algorithm is determined by finding edges through a Laplacian operation and choosing the points whose disparity gradients are larger than a certain threshold.[16,17]

During the reconstruction process of the right image, a number of occluded regions may emerge in which one of the stereo cameras functions while the other does not. As a DMV is not allocated to this occluded region, the average value of the disparities of nearby regions is assigned through the process of disparity stability. If the viewpoint is not occluded, the disparity is then defined as the distance between the image points in both images. Equation (5) shows the disparity in the horizontal direction and the relationship between the right image I_r and left image I_l:

$$I_r = \begin{bmatrix} i_r \\ j_r \end{bmatrix} = \begin{bmatrix} i_l + DV(i_l,j_l) \\ j_l \end{bmatrix} = I_l + \begin{bmatrix} DV(i_l,j_l) \\ 0 \end{bmatrix}. \tag{5}$$

Step 4: Detection of a Moving Object and its Location Coordinates

Figure 3 shows an overall flowchart of the process of searching the actual area where the object is located from the candidate areas and the location coordinates and moving direction of the object in each metadata frame. In general, the actual target areas and the false areas can be distinguished by calculating the similarity between their disparity data in those areas. Specifically, there is a higher correlation between the disparity data in the actual target areas compared to the other cases. Therefore, two areas having a relatively high correlation value between them are taken as the actual target area in each frame of T-1 and T, and the other two areas are discarded as false areas. To detect the actual target area among the candidate areas, the MSE function is used as a cost function, as used for DE in the previous section. Explicitly, by calculating four cases of the cost functions, as shown in Eqs. (6)–(9), one case to minimize this cost function value is selected. Subsequently, the two candidate areas to form this cost function are determined as the actual target areas in each frame. Once the target areas are detected, the location coordinates of the target object in each frame and the relative moving distance between two consecutive frames can be obtained directly.

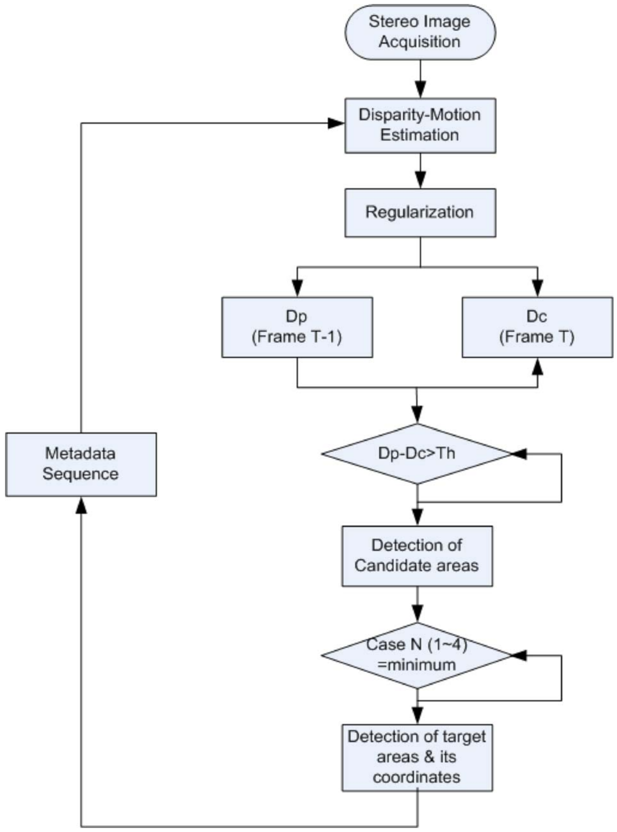

Figure 3. Flowchart for detection of target object and its location coordinates.

$$\text{Case 1: } Cand1 - Cand1' = \frac{1}{m \times n} \sum_{m=1}^{m} \sum_{n=1}^{n} |D_P(xm,yn) - D_C(xm,yn)|^2, \quad (6)$$

$$\text{Case 2: } Cand1 - Cand2' = \frac{1}{m \times n} \sum_{m=1}^{m} \sum_{n=1}^{n} |D_P(xm,yn) - D_C(xm+i,yn+j)|^2, \quad (7)$$

$$\text{Case 3: } Cand2 - Cand1' = \frac{1}{m \times n} \sum_{m=1}^{m} \sum_{n=1}^{n} |D_P(xm+i,yn+j) - D_C(xm,yn)|^2, \quad (8)$$

$$\text{Case 4: } Cand2 - Cand2' = \frac{1}{m \times n} \sum_{m=1}^{m} \sum_{n=1}^{n} |D_P(xm+i,yn+j) - D_C(xm+i,yn+j)|^2. \quad (9)$$

In addition, the moving direction of the target object can be also determined from the polarity of the difference between the center coordinates of the target areas in each frame. That is, if its polarity is positive, the moving direction of the object is determined to be the right and upper direction. In the opposite case, this is determined as the left and lower direction. Finally, the relative moving distance values in the x and y directions and the moving direction are transferred to *Step 5*.

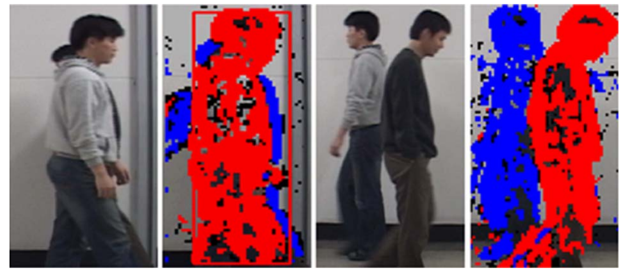

Figure 4. Moving object occluded by a large object.

Step 5: Object Tracking

The moving distance values and the moving direction are directly used to control the pan/tilt embedded in the stereo OTS. This allows a real-time convergence angle control and stereo tracking of a moving object to be achieved. For example, if Eq. (4) has the minimum difference value, the $Cand1$ and the $Cand2'$ in each metadata frame of T-1 and T are determined to be the actual target areas in each metadata frame. From the obtained target areas and the moving direction, the starting coordinates of $D_P(x1,y1)$ in the area of $Cand1$ and the last coordinates of $D_C(x1+i,y1+j)$ in the area of $Cand2'$ can be extracted. The position difference between them is then calculated and sent to the stereo OTS for the convergence control and the pan/tilt control. This process of *Step 1–Step 5* is repeated continuously until the moving object is under tracking. Additionally, if this disparity is transmitted to a receiver system with one of the stereo images in each frame, a stereoscopic 3-D display of the target object is also possible. Accordingly, the proposed method achieves not only stereo tracking but also a stereoscopic 3-D display of a moving object using only the disparity information extracted from the sequence of the input stereo image.

EXPERIMENTAL RESULT

In the experiment, nine frames of the stereo input image pair with 256×256 pixels were used as a test image sequence. A product of Dong Kyoung Electronics Inc. (CS-82393BS) and a product of Hanwool Robotics Co. (HWR-PT1) were used as the stereo cameras with a pan/tilt control system. The proposed algorithm was tested for object tracking an object using metadata. A stereo image acquires 320×240 sequences using IMB-40FT color CCD cameras. The experimental results of the proposed OTS are presented using metadata. The bounding box information of an object outlined by four-direction motion is compared with other metadata. Therefore, at least one object moving can be classified certainly, and noisy motion can be removed. However, the moving object is occluded by a large object, as shown in Figure 4.

Some parts of the image in this sequence have a slight texture with various colors. Color comparison does not yield satisfactory results for object matching, as the color of the

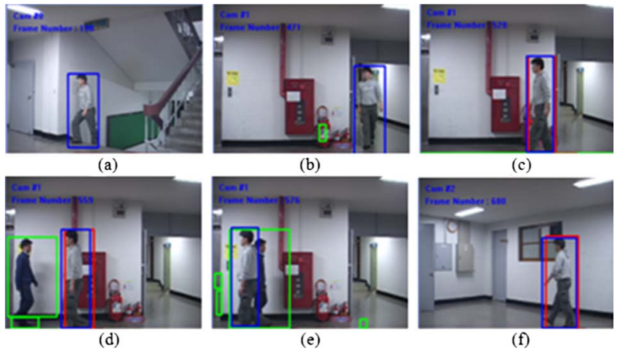

Figure 5. Object tracking results using four cameras under different environments (a) 198th frame, (b) 471th frame, (c) 528th frame, (d) 559th frame, (e) 576th frame, and (f) 680th frame.

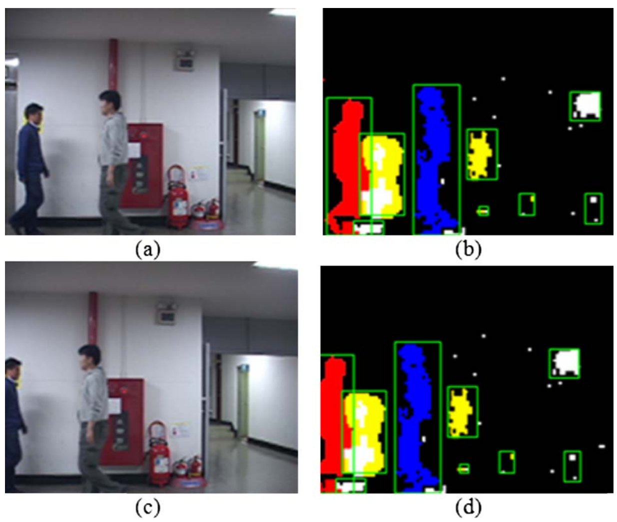

Figure 6. Original images and metadata sequences (a) left original image, (b) metadata of the left image, (c) right original image, and (d) metadata of the right image.

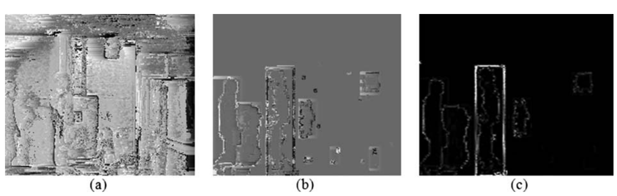

Figure 7. DMV maps (a) feature-based disparity map, (b) metadata-based motion map, and (c) DMV map.

object decreases when the object moves far away. Hence, the proposed method used color matching for an object that was 30×30 or larger in size. In order to demonstrate the performance of the proposed tracking algorithm in a low contrast environment, two experiments were performed, as shown in Figure 5. The object tracking results using the metadata sequences are shown in Figure 6.

Initially, the time-sequential DVs are extracted from the input stereo image and the motion vectors between the metadata frames, the two candidate areas, are detected from each disparity map. Subsequently, from the four detected candidate areas, two actual areas of the location of the object

Table I. Location coordinates of the detected four candidate areas.

	Areas			
Coordinates	*Cand*1	*Cand*1′	*Cand*2	*Cand*2′
Start (x, y)	(61, 125)	(66, 129)	(83, 138)	(90, 142)
End (x, y)	(79, 141)	(86, 147)	(103, 156)	(110, 160)
Center (x, y)	(70, 133)	(76, 138)	(93, 147)	(100, 151)

Table II. Calculated results of the cost function for each candidate area.

Value	Case 1	Case 2	Case 3	Case 4
Cost function	61	1224	370	65

are located are searched using the metadata-based DMV. The location coordinates are then finally detected. These location coordinate values are sent to the pan/tilt system embedded in the stereo OTS.

Figure 7 shows the DMV map of the stereo image and the metadata. Fig. 7(a) is a feature-based disparity map, Fig. 7(b) is a metadata-based motion map, and Fig. 7(c) is the proposed DMV map.

The candidate areas were then searched from the disparity maps of Fig. 7. The location coordinates of the candidate areas were found through a comparison of the estimated disparity maps of the stereo image and the metadata. Table I shows the four candidate areas detected from the proposed DME method, in which Start (x,y), End (x,y), and Center (x,y) denote the starting, ending, and center coordinates of the candidate areas, respectively.

By applying the cost function, a case that minimizes the value of the cost function is then selected. Generally, the DV values of the object are nearly identical in a sequential input stereo image, which implies that the cost function values become smaller in the areas where the object is located compared to other areas. From this selected case, the two areas of the object location are located and finally detected. Accordingly, each case in Eqs. (5)–(8) is calculated using the local disparity data of the four detected candidate areas, as shown in Table I.

Table II shows these results; here, Case 1 is shown to have the smallest value of the cost function. As a result, *Cand*1 and *Cand*1′, which are the areas that form the function of Case 3, are finally decided as the actual target areas in each of the first and second frame.

As the center coordinates of *Cand*1 and *Cand*1′ are given as (70, 133) and (76, 138), respectively, from Table I, the location difference between them is calculated easily as $(-6,-5)$. This value is sent to the pan/tilt system as a control signal. Specifically, if the difference between the center coordinates of the target areas in the first and second frames is a positive value, the target object is assumed to move to the right in the x coordinate and up in the y coordinate directions. In the opposite case, it is assumed to move to the left

Table III. Moving distances of the target object.

Frames	Actual results: Moving distance	Actual results: Absolute distance	Experimental results: Moving distance	Experimental results: Absolute distance
1	(0, 0)	(0, 0)	(0, 0)	(0, 0)
2	(56, −5)	(56, 5)	(55, −4)	(55, 4)
3	(56, −2)	(112, 3)	(56, −3)	(111, 1)
4	(54, 1)	(166, 4)	(53, 0)	(164, 1)
5	(55, 3)	(221, 7)	(54, 2)	(218, 3)
6	(56, −2)	(277, 5)	(55, −1)	(273, 2)
7	(57, −3)	(334, 2)	(57, −1)	(330, 1)
8	(56, −1)	(390, 1)	(55, 1)	(385, 2)
9	(55, 3)	(445, 4)	(55, 2)	(440, 4)

Table IV. Location error ratio of the target object.

Locations	Frames 2	3	4	5	6	7	8	9
Error ratio (%)	2.51	1,78	2.62	2.57	2.52	3.50	3.99	1.81
Average error ratio (%)	2.37							

in the x coordinate and down in the y coordinate. In the above case, the difference between the center coordinates of $Cand1$ and $Cand1'$ was determined as $(-6,-5)$, which indicates that the pan/tilt moves 6 pixels to the left and 5 pixels down.

Table III shows the actual and detected moving distances of the target object for nine consecutive frames of metadata. According to Table III, only an error of approximately 2 pixels exists between the actual and detected moving distance values of the moving object.

At the second frame, the object moves from the origin to (55, 4) and moves to (111, 1), and (164, 1) at the third and fourth frames, respectively. Table IV shows the tracking error ratio between the actual and detected location coordinates of the target object. It was found to be nearly 2.37% on average, in which case the tracking error ratio is defined of the difference between the actual and detected moving distances divided by the actual moving distance. From these experimental results, it was found that the proposed corresponding metadata-based object tracking using DME method can be applied to the practical implementation of the OTS.

CONCLUSIONS

In this article, corresponding metadata-based stereo object tracking using DME is proposed. The proposed system consists of the five steps of metadata creation, disparity estimation, disparity regularization, detection of moving object, and its location coordinates and objects tracking. In the proposed method, the time-sequential DMV can be estimated from the regularized DMVs that are extracted from the sequence of the metadata pair. Using these disparity motion vectors, the area where the target object is located and its location coordinate are then detected from the metadata, and object tracking is performed based on robust characteristics of the object, such as its motion, shape, and color. In the experiment, the proposed metadata-based OTS was shown to be robust for nonrigid objects. It was able to track a target object adaptively with a low average error ratio of approximately 2.37% between the detected and actual location coordinates of the target object.

REFERENCES

[1] R. T. Collins, A. J. Lipton, and T. Kanade, “Introduction to the Special Section on Video Surveillance”, IEEE Transactions on PAMI **22**, 745–746 (2000).

[2] J. Black, T. Ellis, and P. Rosin, “A Novel Method for Video Tracking Performance Evaluation”, *Proc. of IEEE Workshop on Performance Analysis of Video Surveillance and Tracking* (PETS, Nice, France, 2003) pp. 125–132.

[3] K. Smith, D. Gatica-Perez, J.-M. Odobez, and S. Ba, “Evaluating Multi-Object Tracking”, *Proc. of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition* (IEEE Computer Society Press, Los Alamitos, CA, 2005) Vol. **3**, pp. 36.

[4] I. Haritaoglu, D. Harwood, and L. Davis, “W4:real-time surveillance of people and their activates”, IEEE Trans. Pattern Anal. Mach. Intell. **22**, 809–830 (2000).

[5] C. Wren, A. Azerbayejani, T. Darrel, and A. Pentland, Pfinder, “Real time tracking of the human body”, IEEE Trans. Pattern Anal. Mach. Intell. **19**, 780–785 (1997).

[6] A. Baumberg, Learning deformable models for tracking human motion, Ph.D. dissertation, School of Computer Studies, University of Leeds, UK, 1995.

[7] Q. Zhou and J. K. Aggarwal, “Object tracking in an outdoor environment using fusion of features and cameras”, Image Vision Comput. **24**, 1244–1255 (2006).

[8] I. Howard and B. Rogers, *Binocular Vision and Stereopsis* (Oxford University Press, Oxford, 1995).

[9] B. Prothman, “Meta data”, IEEE Trans. Potentials **19**, 20–23 (2000).

[10] C. Atkinson, “Meta-modeling for distributed object environments”, *Proc. IEEE Conf. Enterprise Distributed Object Computing Workshop* (IEEE. Piscataway, NJ, 1997) pp. 90–101.

[11] J. Canny, “A Computational Approach to Edge Detection”, IEEE Trans. Pattern Anal. Mach. Intell. **8**, 679–698 (1986).

[12] G. J. Battaglia, “Mean square error”, AMP J. Tech. **5**, 31–36 (1996).

[13] A. M. Thompson, J. C. Brown, J. W. Kay, and D. M. Titterington, “A Study of Methods of Choosing the Smoothing Parameter in Image Restoration by Regularization”, IEEE Trans. PAMI **13**, 326–339 (1991).

[14] W. E. L. Grimson, “Computational experiments with a feature based stereo algorithm”, IEEE Trans. PAMI **7**, 17–34 (1985).

[15] K. H. Bae, J. J. Kim, and E. S. Kim, “New disparity estimation scheme based on adaptive matching window for intermediate view reconstruction”, Opt. Eng. **42**, 1778–1786 (2003).

[16] Z. D. Lan and J. Konrad, “Regularized block matching using control points”, Tech. Rep. 99-13, INRS-Telecommunications (1999).

[17] R. March, “A regularization model for stereo vision with controlled continuity”, Pattern Recogn. Lett. **10**, 259–263 (1989).