Published in IET Computer Vision
Received on 6th August 2010
Revised on 9th March 2011
doi: 10.1049/iet-cvi.2010.0115

ISSN 1751-9632

Depth space partitioning for omni-stereo object tracking

Z.-H. Xiong[1,2] I. Cheng[2] W. Chen[1] A. Basu[2] M.-J. Zhang[1]

[1]College of Information System and Management, National University of Defense Technology, Changsha 410073, People's Republic of China
[2]University of Alberta, Edmonton, AB, Canada T6G 2E8
E-mail: basu@ualberta.ca

Abstract: Using stereo disparity or depth information to detect and track moving objects is receiving increasing attention in recent years. However, this approach suffers from some difficulties, such as synchronisation between two cameras and doubling of the image-data size. Besides, traditional stereo-imaging systems have a limited field of view (FOV), which means that they need to rotate the cameras when an object moves out of view. In this research, the authors present a depth-space partitioning algorithm for performing object tracking using single-camera omni-stereo imaging system. The proposed method uses a catadioptric omni-directional stereo-imaging system to capture omni-stereo image 'pairs.' This imaging system has 360° FOV, avoiding the need for rotating cameras when tracking a moving object. In order to estimate omni-stereo disparity, the authors present a depth-space partitioning strategy. It partitions three-dimensional depth space with a series of co-axial cylinders, models the disparity estimation as a pixel-labelling problem and establishes an energy minimisation function for solving this problem using graph cuts optimisation. Based on the omni-stereo disparity-estimation results, the authors detect and track-moving objects based on omni-stereo disparity motion vector, which is the difference between two consecutive disparity maps. Experiments on moving car tracking justify the proposed method.

1 Introduction

Object tracking is widely used in applications such as intelligent video surveillance, human–computer interface and intelligent transport systems [1, 2]. During the last two decades, there have been numerous studies on detecting and tracking objects in a sequence of monocular images. Yilmaz *et al.* [3] give a systematic survey on object tracking. Traditional features used in object tracking include colour, edges, optical flow, texture and their combinations [4, 5]. Recently some new features are introduced for object tracking, such as scale-invariant feature transform (SIFT) [6]. However, these features are affected by environmental variations, such as illumination and background changes.

Another feature that is becoming popular in object tracking is disparity or depth in stereo images. Unlike traditional features, such as colour and texture, disparity information is more stable and robust against illumination changes for performing moving-object tracking. Muñoz-Salinas *et al.* [7] combine depth and colour information for object tracking; they also present an approach for multiple people detection and tracking based on stereo vision [8]. The authors in [9] create a multi-object tracking system by combining plan-view projections of depth imagery with feature-based object appearance models. Fiala and Basu [10] use panoramic images to determine a mobile robot's position and orientation, they present the panoramic Hough transform to track line features in the scene.

Bae *et al.* [11] propose a stereo-object-tracking method that is based on disparity motion vectors (DMVs). They extract disparity maps from the captured stereo image-pairs, and then estimate DMV that is defined as the disparity difference between two consecutive disparity maps. Finally, as DMV provides larger disparity difference in moving-target areas than in background areas, therefore the estimated DMV can be used to detect and track a moving object. The main drawbacks of this method include: (i) A small field of view (FOV): when the tracked object moves out of the FOV, it has to rotate both the stereo cameras to 'see' the tracked object. (ii) Uses two cameras to acquire stereo-image pairs for disparity-map estimation: that requires extra synchronisation control between cameras, extra data-transmission channels and hardware costs.

In order to overcome the aforementioned shortcomings, we propose a depth-space partitioning algorithm to estimate disparity in a single-camera omni-stereo imaging system, and then track a moving object based on the estimated omni-stereo disparity information. In this method, we capture a 360° FOV image sequence with a catadioptric omni-stereo system; then we partition the 3D depth space into a series of co-axial cylinders for omni-stereo disparity estimation; finally, a moving object is detected and tracked using omni-stereo DMV. A critical issue in this process is to estimate the omni-stereo disparity, because omni-stereo disparity estimation is totally different from a traditional stereo system, we present a depth-space partitioning method to estimate omni-stereo disparity.

In the first step, we design a single-camera omni-stereo imaging system to capture omni-stereo image sequence for object tracking. This imaging system is composed of one ordinary camera and two vertically-aligned curved mirrors that acquire images as if they are stereo-image 'pairs' by using only one camera. An image is divided into outer and inner parts, corresponding to the upper and lower mirrors, respectively; we can treat these two parts as left and right images in a traditional stereo system. This imaging system has two advantages over traditional stereo systems: (i) instead of using two cameras, it uses only one camera to capture stereo information into one image at the same time; (ii) it has a 360° FOV, making it possible to track-moving objects without rotating cameras.

In the second step, we estimate the omni-stereo disparity that will be used for object tracking. Since the omni-stereo imaging system is totally different from traditional stereo systems (such as Bae *et al.* [11]), the calculation of omni-stereo disparity is also different. To this end, we propose a three-dimensional (3D) depth-space partitioning approach to estimate omni-stereo disparity as follows: (i) partitioning the 3D space with a sequence of co-axial cylinders, with each cylinder representing a depth or disparity; (ii) model the omni-stereo disparity-estimation problem as a pixel-labelling problem [12, 13], the goal is to decide which cylinder each pixel of the omni-stereo image belongs to; (iii) establish an energy minimisation function based on colour difference, piecewise smooth and occlusion constraints for this pixel-labelling problem and solve this function using graph cuts optimisation.

In the final step, we detect and track-moving objects based on the estimated omni-stereo disparity. We subtract two consecutive omni-stereo disparity maps T and $T-1$ to obtain the omni-stereo DMV, and then detect candidate areas of moving objects based on the calculated DMV, and finally decide on the true areas of objects by computing common sizes of the candidate areas in consecutive frames of a DMV; thus, tracking a moving object in every frame.

The main focus of this article is on depth-space partitioning for omni-stereo object tracking. For simplicity we assume that there is one moving object to be tracked in this research. In the experiments, a moving car is driven around the scene covering a wide FOV, and the proposed method tracks it without any need for rotating the camera.

The remainder of the paper is organised as follows: Section 2 designs the single-camera omni-stereo imaging system for object tracking. Omni-stereo disparity estimation based on depth-space partitioning is presented in Section 3. Section 4 demonstrates object tracking using estimated omni-stereo disparity. Experimental results are shown in Section 5. Finally, Section 6 concludes this article.

2 Design of a single-camera omni-stereo system for object tracking

Many computer applications need a wide FOV, such as robot vision [14–18] and visual surveillance [19, 20]. Catadioptric omni-directional imaging that provides a 360° FOV, is useful for this requirement. In our previous work in [21, 22], we present a double-lobed mirror for omni-stereo depth perception. However, the double-lobed mirror has a short baseline, resulting in low depth resolution. In order to improve the depth resolution for object tracking, we design a new single-camera omni-stereo imaging system that has a longer baseline.

Fig. 1 illustrates the design of a single-camera omni-stereo imaging system and the captured omni-stereo image for object tracking. In this design, we extend the conventional catadioptric omni-directional imaging system to one camera plus two separate mirrors, the camera and two mirrors are co-axially placed in a vertical direction.

The single-camera omni-stereo imaging system has rotational symmetry property; therefore it is sufficient to consider the radial cross-section for simplicity. As shown in Fig. 1*a*, we define the coordinate origin O to be at the centre of an image plane, the X-axis lies in the image plane,

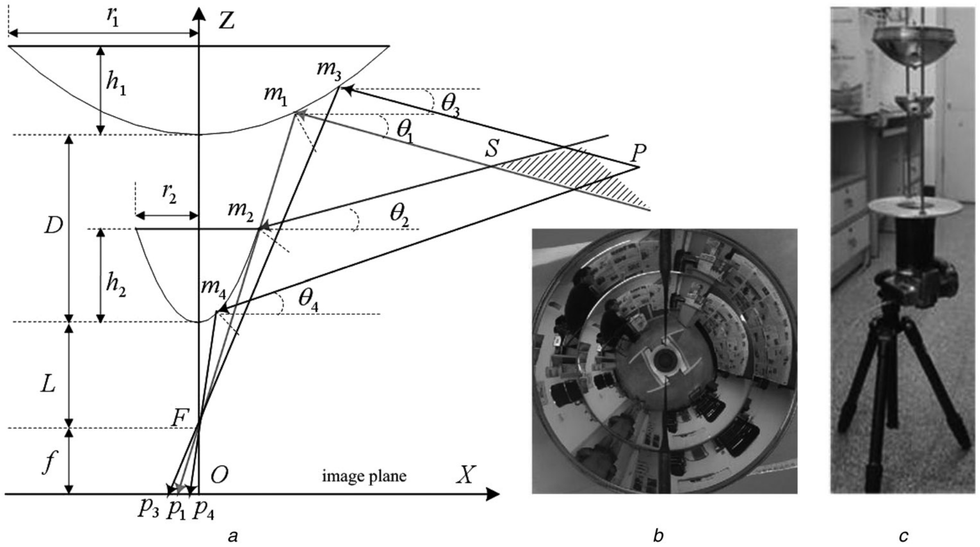


Fig. 1 *Single-camera omni-stereo system and omni-stereo image*

a Single-camera omni-stereo system design
b Omni-stereo image
c Practical system

and the Z-axis is the same as the camera and mirror axis. The radius and height of the upper mirror are r_1 and h_1, the radius and height of the lower mirror are r_2 and h_2. The camera focal length is f, the distance from the focal point to the bottom of the lower mirror is L and the omni-stereo baseline is D (the distance between bottom points of the two mirrors). The upper mirror equation is $z = a_1x^2 + f + L + D$, and the lower mirror equation is $z = a_2x^2 + f + L$.

Suppose P is a 3D point in the common view field of both mirrors, m_3, m_4 are intersection points of P on the upper and lower mirrors, $p_3 = (x_{p3}, 0)$, $p_4 = (x_{p4}, 0)$ are images of P reflected from the upper and lower mirrors. The lines $\overrightarrow{m_3P}$, $\overrightarrow{m_4P}$ are

$$\begin{cases} z - z_{m3} = \tan\theta_3 \times (x - x_{m3}) \\ z - z_{m4} = \tan\theta_4 \times (x - x_{m4}) \end{cases} \tag{1}$$

where

$$\begin{cases} x_{m3} = (f - \sqrt{f^2 - 4a_1|x_{p3}|^2(L + D)})/(2a_1|x_{p3}|) \\ z_{m3} = a_1x_{m3}^2 + f + L + D \\ x_{m4} = (f - \sqrt{f^2 - 4a_2|x_{p4}|^2(L)})/(2a_2|x_{p4}|) \\ z_{m4} = a_2x_{m4}^2 + f + L \\ \theta_3 = 2\arctan(2a_1x_{m3}) - \arctan(f/|x_{p3}|) \\ \theta_4 = 2\arctan(2a_2x_{m4}) - \arctan(f/|x_{p4}|) \end{cases} \tag{2}$$

On the other hand, if the coordinates of point P are (m, n), then the image point $p_3 = (x_{p3}, 0)$ of P reflected by the upper mirror can be determined by

$$\begin{cases} \dfrac{n - z}{m - x} = \dfrac{4a_1(-x_{p3})x + f - 4fa_1^2x^2}{(-x_{p3}) - 4(-x_{p3})a_1^2x^2 - 4a_1fx} \\ z = a_1x^2 + f + D + L \\ \dfrac{z - f}{x} = -\dfrac{f}{(-x_{p3})} \end{cases} \tag{3}$$

Here (x, y) are the coordinates of point m_3, the intersection point of P on the upper mirror. The coordinates of image point $p_4 = (x_{m3}, 0)$ can be determined accordingly.

Fig. 1*b* shows an omni-stereo image acquired with the single-camera omni-stereo imaging system. There are two ‘images’ of a 3D object (such as the person) in the omni-stereo image, reflected by the upper mirror and lower mirror separately. Thus, we divide the omni-stereo image into two parts, namely outer and inner parts, conceptually corresponding to the left and right images in traditional stereo vision.

Fig. 1*c* shows a practical omni-stereo system for object tracking, where the upper and lower mirror radii are $r_1 = 5$ and $r_2 = 1.8$, their 3D shapes are $x^2 + y^2 = 7z$ and $x^2 + y^2 = 2.4z$, the distance between bottom points of the two mirrors is $D = 10$ inches. Finally, the camera focal length f and the distance from focus to the bottom point of the lower mirror L are partially dependent on the camera used.

3 Omni-stereo disparity estimation based on depth-space partitioning

In order to use the disparity information for object tracking, we need to perform disparity estimation from the omni-stereo image. Traditional stereo-disparity estimation is based on pinhole cameras, where perspective projection is applied. However, the single-camera omni-stereo imaging system does not fulfil the perspective-projection conditions, therefore traditional methods for estimating stereo disparity can no longer be used to estimate the disparity in a single-camera omni-stereo imaging system.

In this research, we estimate omni-stereo disparity based on depth space partitioning and energy minimisation. First of all, we partition the depth space with a sequence of virtual co-axial cylinders, these cylinders share the same axis as the omni-stereo camera axis; each cylinder represents all the 3D space points that have the same depth with respect to the axis, and the radius of the cylinder is the depth value. Then, we model the disparity-estimation problem as a pixel-labelling problem [12, 13], where label stands for the cylinder, and the goal is to find a labelling that assigns each pixel in the outer and inner parts of an omni-stereo image a label, ensuring that this labelling fulfils omni-stereo depth constraints and observations. Therefore in the next step, we formulate the omni-stereo disparity-estimation problem as energy minimisation of colour difference, piecewise smooth and occlusion constraints. Finally, we solve the energy minimisation using graph cuts optimisation.

3.1 Scene-space partitioning with co-axial cylinders

In the single-camera omni-stereo, we define the depth of a 3D scene point as the shortest distance from this point to the axis of a single-camera omni-stereo system. Therefore if we create a virtual cylinder and make its axis coincide with the axis of a single-camera omni-stereo system, then all the 3D space points on the cylinder surface have the same depth, and the depth value is simply the radius of this cylinder.

Based on this observation, we partition the 3D scene with a sequence of co-axial cylinders, and we use the cylinders to represent the depth and disparity in the single-camera omni-stereo. Thus, the problem of disparity estimation is converted to deciding which cylinder each pixel of the omni-stereo image belongs to, and this concept is illustrated in Fig. 2. For a 3D point P on cylinder ℓ_p, there is one image in both outer part and inner part of the omni-stereo image, respectively, denoted by p_o and p_i.

The number of co-axial cylinders depends on two major factors: (i) Visible scene depth from the omni-stereo system; (ii) the interval between two adjacent cylinders. The omni-stereo system ‘sees’ interesting objects (such as persons, cars) within about 50 m around in outside scenario. Also, we set the interval between two adjacent cylinders as 0.2 m uniformly, which is sufficient for object detection and tracking, because the width of tracked objects is usually larger than 0.2 m. Therefore 250 co-axial cylinders are used in this work.

3.2 Modelling the omni-stereo disparity-estimation problem

In this section, we model the omni-stereo disparity-estimation problem as a pixel-labelling problem by assigning a virtual cylinder to each pixel in the outer and inner parts of an omni-stereo image. We define the following notation for this model:

- N, total number of co-axial virtual cylinders.
- ℓ_i, an individual virtual cylinder (called a label).

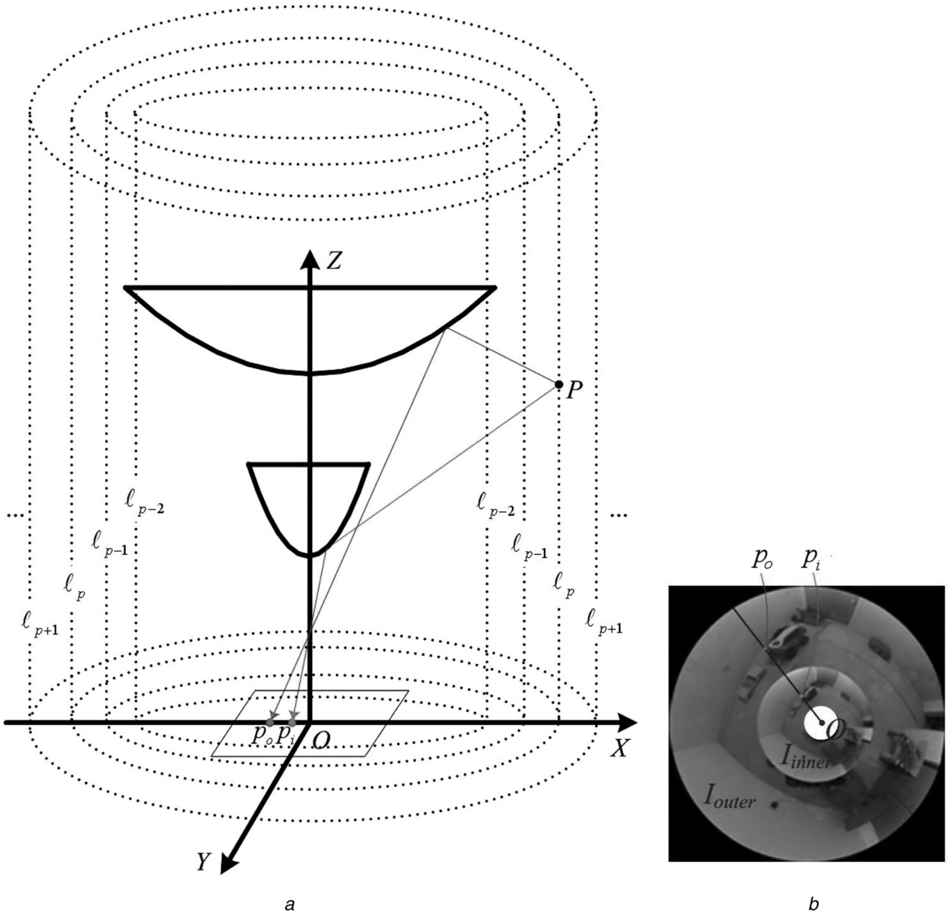


Fig. 2 *Depth space partitioning with co-axial cylinders*

a Partitioning 3D depth space with co-axial virtual cylinders
b Two images of a 3D space point

- φ, the set of co-axial virtual cylinders ℓ_i, where $\varphi = \{\ell_i | i = 1, 2, \ldots, N\}$.
- I_{outer}, all the pixels in the outer part of a single-camera omni-stereo image.
- I_{inner}, all the pixels in the inner part of a single-camera omni-stereo image.

Based on the above notation, the goal of the disparity-estimation problem is to find a labelling f that assigns each pixel $p \in I_{\text{outer}} \cup I_{\text{inner}}$ with a cylinder $\ell_p \in \varphi$, where f satisfies the omni-stereo depth constraints and observations. The constraints and observations on an omni-stereo image include: (i) piecewise smooth depth values, (ii) occlusion constraint and (iii) minimum colour difference between corresponding points. We describe these constraints and observations in the next section.

3.3 Energy minimisation of omni-stereo disparity

After modelling the omni-stereo disparity-estimation problem as a pixel-labelling problem, we solve it through energy minimisation with either local or global methods. The authors of [13, 23] prove that global optimisation methods gives better results than local ones, so we use global methods to solve omni-stereo disparity-estimation problem. A simple but useful case of smoothness constraint is the Potts model [21, 24], which assumes that the majority of nearby pixels have the same label. Compared with other energy-minimisation methods, for example, convergent tree reweighted message passing [25] and belief propagation [26], the graph-cuts based optimisation performs best in terms of speed and accuracy. Therefore we select graph-cuts optimisation to solve the omni-stereo disparity-estimation problem in this research. A critical part of graph-cuts optimisation is constructing an energy-minimisation function based on the particular constraints and observations of the problem that are discussed in the following sub-sections.

3.3.1 Piecewise smooth constraint: Disparity and depth tend to be piecewise smooth; they vary smoothly on the surface of an object, but change dramatically at object boundaries. As omni-stereo has different imaging properties in the radial direction and tangential direction, we define different piecewise smooth constraints accordingly. Suppose $E_{\text{smooth}}(f)$ is the piecewise smooth energy for an omni-stereo depth-label mapping, $E_{\text{smooth\_}r}(f)$ and $E_{\text{smooth\_}t}(f)$ are the piecewise smooth energy in the radial and tangential directions, respectively, we then have

$$E_{\text{smooth}}(f) = E_{\text{smooth\_}r}(f) + E_{\text{smooth\_}t}(f) \tag{4}$$

Because the two mirrors in a single-camera omni-stereo imaging system are placed vertically, the pixels in radial directions correspond to vertical 3D space points, so the vertical points in 3D space have the same disparity or depth. This situation is similar to traditional stereo imaging, so we use Potts model [24] to describe the smooth energy

doi: 10.1049/iet-cvi.2010.0115

in the radial direction as

$$E_{\text{smooth\_r}}(f) = \sum_{\{p,q\} \in N_{\text{r}}} (K_{\text{r}} * T_{\text{r}}(f_p \neq f_q)) \tag{5}$$

where p and q are adjacent pixels in the radial direction; K_{r} is a user-defined constant for radial smoothness penalty; $T_{\text{r}}(\cdot)$ is 1 if its argument is true and 0 otherwise.

In the tangential direction, the omni-stereo imaging system has gradual depth change between tangentially adjacent pixels, we should allow depth labelling difference in this direction, so $E_{\text{smooth\_t}}(f)$ is defined as

$$E_{\text{smooth\_t}}(f) = \sum_{\{p,q,s\} \in N_{\text{t}}} (K_t * \min(T_{\text{t}}, g(|f_p + f_s - 2f_q| + |f_p - f_s|)) \tag{6}$$

where p, q and s are adjacent pixels in the tangential direction; K_{t} is a user defined constant for tangential smoothness penalty; $T_{\text{t}}(\cdot)$ is 1 if its argument is true and 0 otherwise; $g(\cdot)$ is a symmetrical convex function; for example, we set $g(\cdot)$ as an absolute value function in this work.

3.3.2 Occlusion constraint: Fig. 3 demonstrates the occlusion in a single-camera omni-stereo imaging system. In this figure, there are two 3D points (p, ℓ_p) and (q, ℓ_q), where (q, ℓ_q) is occluded by (p, ℓ_p) in the omni-stereo image. It is obvious that (q, ℓ_q) has no matching point in the outer part of an omni-stereo image. Therefore if a 3D space point (q_i, ℓ_{qi}) is an occluded point, then there exists another 3D space point (p_j, ℓ_{pj}) satisfying $\ell_{pj} < \ell_{qi}$ and $(p_j, \ell_{qi}) = (q_i, \ell_{qi})$. Therefore the occlusion energy function can be defined as

$$E_{\text{occ}}(f) = \sum_{\{(p,\ell_p),(q,\ell_q)\} \in R_{\text{occ}}} K_{\text{occ}} \tag{7}$$

where K_{occ} is a user-defined constant; R_{occ} is occlusion relationships, and items $\{(p, \ell_p), (q, \ell_q)\}$ satisfy

1. $p \in I_{\text{outer}}$, $q \in I_{\text{inner}}$, and p, q have the same radial angle in an omni-stereo image.
2. $\ell_p = f(p)$ also denoted as f_p; $\ell_q = f(q)$, also denoted as f_q; and $\ell_p \neq \ell_q$.
3. if $\ell_p > \ell_q$, then $(p, \ell_p) = (q, \ell_p)$; if $\ell_p < \ell_q$, then $(p, \ell_q) = (q, \ell_q)$.

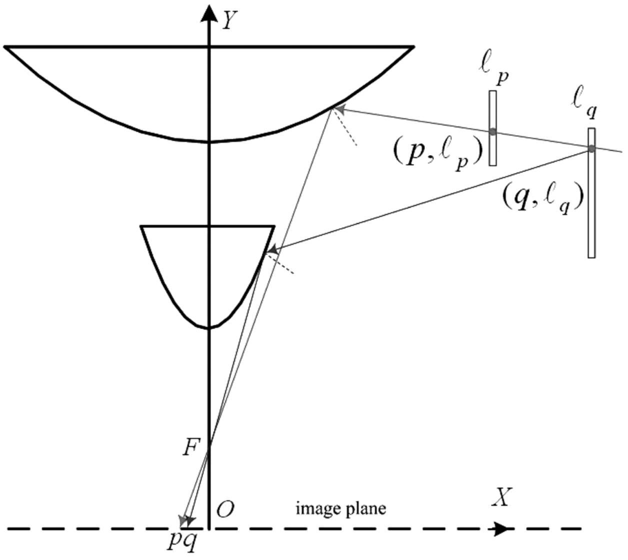


Fig. 3 *Occlusions in a single-camera omni-stereo imaging system*

Because p, q are discrete pixel points in the occlusion model, their corresponding 3D space points (p, ℓ_p) and (q, ℓ_p) are also discrete points. Thus, we define $(p, \ell_p) = (q, \ell_p)$ if and only if their Euler distance satisfies $\text{Dis}((p, \ell_p), (q, \ell_q)) < \varepsilon$, where ε is a user-defined constant.

3.3.3 Minimum colour difference constraint: In graph-cuts based optimisation, the disagreement between labelling f and the observed data is measured by $E_{\text{data}}(f)$. Many different energy functions have been proposed in the literature. The form of $E_{\text{data}}(f)$ is typically

$$E_{\text{data}}(f) = \sum_{p \in P} D_p(f_p) \tag{8}$$

where D_p measures how well label f_p fits pixel p given the observed data. Colour difference is often used to measure D_p in a stereo disparity estimation field. We also use colour difference to define $E_{\text{data}}(f)$ is this research as

$$E_{\text{data}}(f) = \sum_{\{(p,\ell_p),(q,\ell_q)\} \in R_{\text{corres}}} D(p, q) \tag{9}$$

where R_{corres} is a correspondence relationship, and $\{(p, \ell_p), (q, \ell_q)\}$ is: $p \in I_{\text{outer}}$, $q \in I_{\text{inner}}$, $\ell_p = f_p$ and $\ell_p = \ell_q$ and $(p, \ell_p) = (q, \ell_q)$. $D(p, q)$ is the colour difference between p and q, and $D(p, q) = \max\{(\text{colour}_R(p) - \text{colour}_R(q), (\text{colour}_G(p) - \text{colour}_G(q), (\text{colour}_B(p) - \text{colour}_B(q)\}$.

3.3.4 Integrated graph-cut energy-minimisation function: Based on the piecewise smooth constraint, occlusion constraint and minimum colour difference constraint, we define the final integrated graph-cuts energy minimisation function as

$$\min E(f) = \min[E_{\text{smooth}}(f) + E_{\text{occ}}(f) + E_{\text{data}}(f)] \tag{10}$$

where $E_{\text{smooth}}(f)$, $E_{\text{occ}}(f)$ and $E_{\text{data}}(f)$ are the smooth, occlusion and data colour constraints, respectively, as defined in (4)–(9). Equation (10) is a global minimum-optimisation problem, and we use the min-cut/max-flow algorithm introduced by Boykov *et al.* [12] and Ahuja *et al.* [27] to solve this equation.

4 Object tracking based on omni-stereo DMV

4.1 DMV extraction

After we estimate omni-stereo disparity with space-depth partitioning, we detect and track a moving object using omni-stereo DMV. In this study, omni-stereo DMV is defined as a difference between two consecutive omni-stereo disparity maps frames $T - 1$ and T. DMV has a relatively large change of disparity values in the regions where a target object is located before and after motion, whereas it has almost no change of disparity values in regions where there is no movement, such as in the background. Thus, we can use this disparity difference to detect and track moving objects.

Fig. 4*a* shows the overall process of extracting omni-stereo DMVs from an omni-stereo image sequence, and detecting candidate object areas. Each omni-image contains outer and inner parts, corresponding to the 'left' and 'right' images in a conventional stereo system. The disparity map sequence is estimated using the proposed depth-space partitioning method, and it is based on the outer part of an omni-stereo image, as shown in Fig. 4*a* (so there are no inner parts for disparity maps). Suppose D_T is the disparity map of a single-camera omni-stereo image frame *T*, and DMV_T is the DMV between disparity maps D_T and D_{T-1}. Then DMV_T can be calculated as

$$\mathrm{DMV}_T = |D_T - D_{T-1}| \tag{11}$$

From each of these DMV maps, we can find areas that have a large change of disparity values. These areas are created by target movements, and therefore they are potential locations for moving objects being tracked.

4.2 Moving object detection and tracking

In this study, we focus on depth-space partitioning for omni-stereo object tracking, so we assume that there is one moving object to be tracked for simplicity. Under this assumption, there are at most two areas that have a relatively large change of disparity values in each DMV map, and these two areas are the object's areas in the two consecutive frames, respectively. For example, in the candidate area detection step of Fig. 4*a*, there are two areas detected in DMV_T, the area labelled as 1 is the object's area in omni-stereo image frame $T-1$, and the other area labelled as 2 is the object's area in omni-stereo image frame *T*.

Now, the problem is how to decide which one is the object area of omni-image frames $T-1$ and *T*? This can be resolved through a common area-computation process. Fig. 4*b* shows the area correspondence for a moving object in omni-stereo images and DMV maps. In this figure, the car in omni-stereo image frame *T* has one candidate area in both DMV_T and DMV_{T+1}, they are Area 2 in DMV_T and Area $1'$ in DMV_{T+1}. Therefore these two areas share almost the same coordinate areas. This characteristic can be used to decide which candidate area in DMV_T is the true object location of omni-stereo frame *T*. The main steps of the algorithm are:

Step 1: For two consecutive DMV frames DMV_T and DMV_{T+1}, there are totally four candidate areas detected, we mark the two candidate areas in DMV_T as $\mathrm{Cand1}_T$ and $\mathrm{Cand2}_T$, the two candidate areas in DMV_{T+1} as $\mathrm{Cand1}_{T+1}$ and $\mathrm{Cand2}_{T+1}$. For example, $\mathrm{Cand1}_T$ and $\mathrm{Cand2}_T$ corresponds to the areas labelled as 1 and 2 in DMV_T of Fig. 4*b*, while $\mathrm{Cand1}_{T+1}$ and $\mathrm{Cand2}_{T+1}$ correspond to $1'$ and $2'$ in DMV_{T+1} of Fig. 4*b*.
Step 2: Compute the common areas between the two candidate areas in DMV_T and the two candidate areas in DMV_{T+1}, so that there are four cases to be computed, they

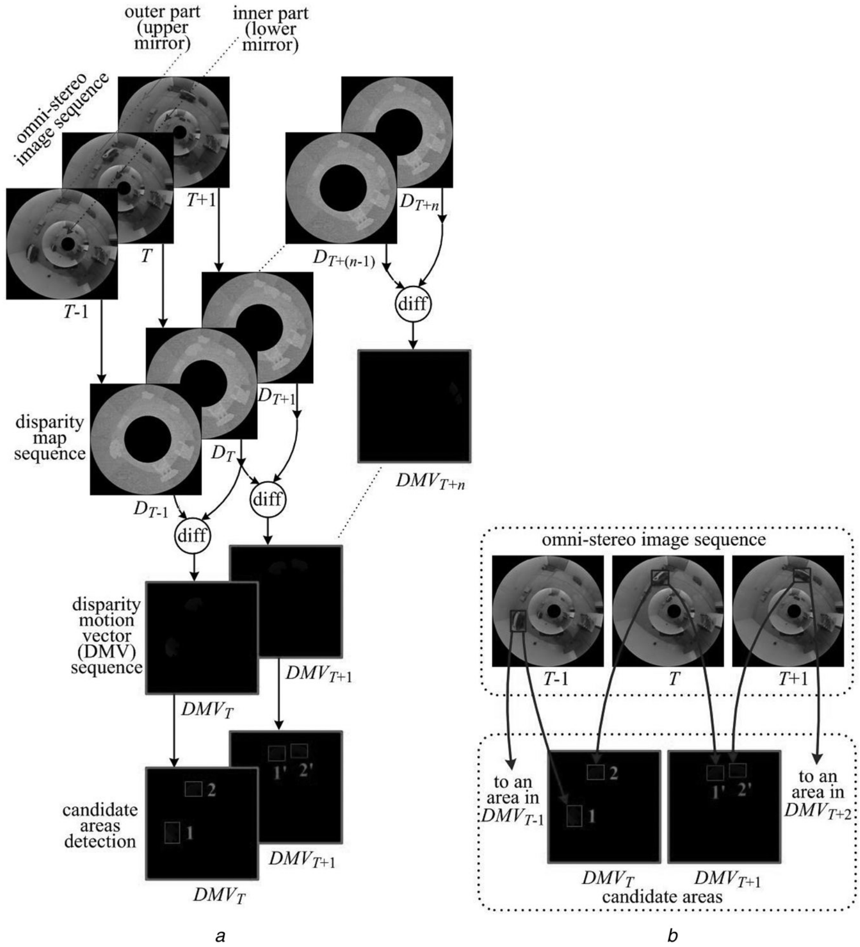

outer part (upper mirror)
inner part (lower mirror)
omni-stereo image sequence
T-1
T
T+1
$D_{T+(n-1)}$
D_{T+n}
diff
DMV_{T+n}
disparity map sequence
D_{T-1}
D_T
D_{T+1}
diff
diff
disparity motion vector (DMV) sequence
DMV_T
DMV_{T+1}
candidate areas detection
2
1
1'
2'
DMV_T
DMV_{T+1}
a
omni-stereo image sequence
T-1
T
T+1
to an area in DMV_{T-1}
2
1
1'
2'
to an area in DMV_{T+2}
DMV_T
DMV_{T+1}
candidate areas
b

Fig. 4 *Object tracking based on omni-stereo DMV*
a DMV extraction and candidate object areas detection
b Correspondence of moving object in omni-stereo images and DMV maps

doi: 10.1049/iet-cvi.2010.0115

are (Cand1_T, Cand1_{T+1}), (Cand1_T, Cand2_{T+1}), (Cand2_T, Cand1_{T+1}), (Cand2_T, Cand2_{T+1}). We define the common area size of two areas ($A1$, $A2$) as the number of pixels that have the same coordinates, which can be denoted by the following formulation

$$\text{Common area size}(A1, A2) = \text{card}\left(\left\{(p1, p2)\middle| \begin{array}{c} p1 \in A1, \quad p2 \in A2 \quad \text{and} \\ p1 \quad \text{and} \quad p2 \text{ have the same} \\ \text{coordinate values} \end{array}\right\}\right) \tag{12}$$

where card(·) is a function for counting the number of items in a set. For example, card ({a, b, d}) is three, meaning that there are three items in the set {a, b, d}.

Step 3: For the four candidate area pairs (Cand1_T, Cand1_{T+1}), ($\text{Cand1}_T, \text{Cand2}_{T+1}$), ($\text{Cand2}_T, \text{Cand1}_{T+1}$), ($\text{Cand2}_T, \text{Cand2}_{T+1}$), compute their common area sizes, respectively, using (12). Then, select the pair that has the maximum common area size and the first candidate area in this pair is just the true object location of omni-stereo image frame T. For example, in Fig. 4b, the candidate area pair (2, 1′) has largest common area size than the other candidate area pairs (1, 1′), (1, 2′) and (2, 2′).

When we carry out the above steps for every omni-stereo image frame, we find the moving object locations for every frame, which means tracking the moving object.

5 Experiments

5.1 Depth-estimation error analysis

The assembly and configuration of a single-camera omni-stereo imaging system will impose depth- and disparity-estimation errors, and finally affect the moving object tracking results. From (1), the depth of a 3D space point P with respect to the omni-stereo imaging system can be denoted as

$$z = (\tan\theta_4 \times x_{m4} + y_{m4} - \tan\theta_3 \times x_{m3} - y_{m3}) / (\tan\theta_4 - \tan\theta_3) \tag{13}$$

where $x_{m3}, y_{m4}, x_{m4}, y_{m4}, \theta_3$ and θ_4 have the same meaning as in (1).

According to (1)–(3) and Fig. 1, we know that the manufacturing error of the two omni-stereo mirrors can be integrated into a_1, a_2 (and they are constant values after manufacturing); the co-axial error of these mirrors can be integrated into x_{m3}, x_{m4}, the focal distance error can be included into D, therefore the depth error of z can be calculated according to D, x_{m3}, x_{m4}. Suppose $z = g(D, x_{m3}, x_{m4}) = (\tan\theta_4 \times x_{m4} + y_{m4} - \tan\ \theta_3 \times x_{m3} - y_{m3})/(\tan\theta_4 - \tan\theta_3)$, the gradient of $g(D, x_{m3}, x_{m4})$ is $\nabla \boldsymbol{g} = (\partial g/\partial D, \partial g/\partial x_{m3}, \partial g/\partial x_{m4})$, then the error of depth z is

$$\begin{aligned} \Delta z &\simeq \nabla \text{g} \cdot (\delta x)^{\text{T}} = \frac{\partial \text{g}}{\partial D} \cdot \delta D + \frac{\partial \text{g}}{\partial x_{m3}} \cdot \delta x_{m3} + \frac{\partial \text{g}}{\partial x_{m4}} \cdot \delta x_{m4} \\ &\leq \left|\frac{\partial \text{g}}{\partial D}\right| \cdot \delta D + \left|\frac{\partial \text{g}}{\partial x_{m3}}\right| \cdot \delta x_{m3} + \left|\frac{\partial \text{g}}{\partial x_{m4}}\right| \cdot \delta x_{m4} \end{aligned} \tag{14}$$

Based on (14), we obtain the depth error results with respect to D, a_1, a_2 and 3D space position (z, w) as shown in Tables 1–4. Note that when we perform the depth-estimation error in these tables, the measurement units of z and D and height are in inches.

Table 1 shows the effect of mirror shape on depth error, parameters a_1, a_2 decide the shape of the two mirrors. Here, we set $D = 10$ in. and the position of P to be (200, 0).

Table 2 shows the effect of omni-stereo baseline length D on depth error, where we set $a_1 = a_2 = 3.5$ and the position of P to be (200, 0).

Table 3 gives the effect of 3D space point depth z on depth estimation error, where $D = 10$ in. and $a_1 = a_2 = 3.5$. In this table, as the depth z increases from 50 to 200 in, the depth error Δz increases from 0.172 to 2.387 in., indicating that the depth error Δz increases faster than the increase in depth z. Hence, the depth resolution is non-uniform across the omni-stereo image, the farther away a 3D space point from the omni-stereo imaging system, the lower the depth resolution.

Table 4 shows the effect of 3D space point height w on depth error, where $D = 10$ in and $a_1 = a_2 = 3.5$.

Based on the analytical results in Tables 1–4, we conclude that: (i) depth error decreases as a_1, a_2 increase, however, the omni-stereo FOV decreases; (ii) the longer the omni-stereo baseline D is, the smaller the depth-error results; (iii) when other parameters do not change, the depth error increases

Table 1 Depth error with respect to mirrors shape

a_1	3.5	3.5	3.5	4
a_2	1	1.2	1.5	1.5
x_{m3}	3.300	3.300	3.300	3.767
x_{m4}	0.998	1.196	1.494	1.494
Δz, in	4.661	4.105	3.555	3.343
$\Delta z/z^*$ 1000	23.307	20.526	17.776	16.713

Table 2 Depth error with respect to omni-stereo baseline length

D, in.	5	10	20
x_{m3}	3.384	3.300	3.140
x_{m4}	3.470	3.470	3.470
Δz, in	4.711	2.387	1.227
$\Delta z/z^*$ 1000	23.557	11.935	6.136

Table 3 Depth error with respect to 3D space point depth

z, in.	50	100	150	200
x_{m3}	2.773	3.113	3.237	3.300
x_{m4}	3.3800	3.439	3.459	3.470
Δz, in	0.172	0.624	1.363	2.387
$\Delta z/z^*$ 1000	3.232	6.243	9.085	11.935

Table 4 Depth error with respect to 3D space point height

w, in	−20	0	20	40
x_{m3}	2.988	3.300	3.647	4.029
x_{m4}	3.140	3.470	3.834	4.233
Δz, in	2.657	2.387	2.165	1.985
$\Delta z/z^*$ 1000	13.284	11.935	10.827	9.924

when a 3D space point goes deeper, and the depth error decreases when a 3D space point obtains higher; and (iv) depth estimation error is within 2% in all the above cases.

The depth error is also related to the number of co-axial cylinders used. Generally, if more cylinders are used higher depth resolution can be achieved, and lower depth error can be expected. However, using more cylinders can result in higher computational cost. There are 250 co-axial cylinders used in this work.

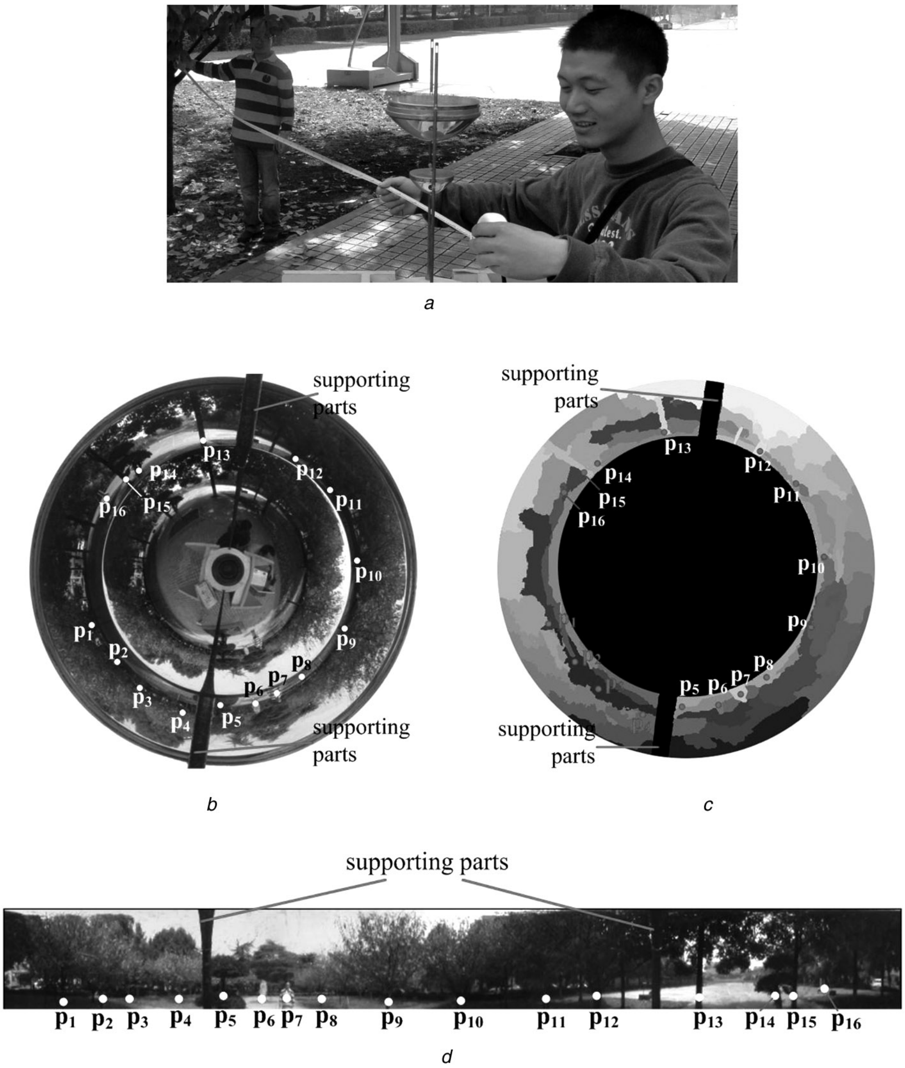
5.2 Depth estimation compared with ground truth

According to depth-estimation error analysis results with respect to mirror shapes, omni-stereo system baseline length, 3D space point depth and 3D space point height, as shown in Tables 1–4, we set $a_1 = 3.5$, $a_2 = 1.5$ and omni-stereo baseline length $D = 10$. Then, we manufacture and assemble the omni-stereo system using these parameters. Fig. 1*c* shows the physical omni-stereo system.

Fig. 5*a* illustrates the procedure for measuring ground truth depth data; in this work, depth is defined as the distance from the middle of the omni-stereo system to the point whose depth is to be measured. Figs. 5*b*–*d* show the 16 representative ground truth points in an omni-stereo image, depth map image and the corresponding unwarped panoramic image, respectively.

In the experiments, we select 16 representative points (Fig. 5*b*), and measure their depths as ground truth data. All the representative points selected are on an object such as a tree or a person. Then, we estimate the depth map of the omni-stereo image (Fig. 5*b*), the estimated depth map image is shown in Fig. 5*c*. In order to make it easier to watch the omni-stereo image, we unwarp the outer part of the omni-stereo image in Fig. 5*b*, and the resulting unwarped panoramic image is shown in Fig. 5*d*.

Table 5 compares the estimated depth with ground-truth depth for the selected 16 representative points. According to Table 5, the average depth-estimation error ratio with respect to ground truth data is 7.51%. We can see that the average depth-error ratio in Table 5 is higher than the theoretical results in Tables 1–4; this is because there are some other errors in a practical omni-stereo system, such as mirror assembly error, and correspondence matching error.



a

b

c

d

Fig. 5 *Ground truth and estimated data of representative points*

a Depth ground truth measurement procedure
b Ground truth points in omni-stereo image
c Ground truth points in depth map image
d Ground truth points in unwarped panoramic image

Table 5 Depth estimation results compared with ground truth depth data

Representative points	*p*1	*p*2	*p*3	*p*4	*p*5	*p*6	*p*7	*p*8
ground truth, m	9.65	7.34	8.23	5.51	4.12	7.53	3.12	5.69
estimated depth,	10.34	7.93	7.45	6.03	3.82	7.38	3.42	5.11
depth error ratio, %	7.15	8.04	9.48	9.44	7.28	1.99	9.62	10.19
Representative points	p9	p10	p11	p12	p13	p14	p15	p16
ground truth, m	8.54	7.13	4.76	5.42	3.52	3.89	3.54	8.92
estimated depth, m	8.97	7.59	5.07	4.91	3.77	4.08	3.16	9.53
depth error ratio, %	5.04	6.45	6.51	9.41	7.10	4.88	10.73	6.84

For an omni-stereo image with 500 × 500 resolution, the average time for performing depth estimation is about 0.35 s. There are two steps to ensure that the proposed method can be used in object detection and tracking: (i) carrying out pre-processing on the omni-stereo image sequence, finding the regions where possible moving objects may exist, (ii) because these regions occupy only a portion of the overall omni-stereo image, estimating the depth of these regions improves the efficiency of depth estimation for object tracking.

5.3 Moving object tracking

Fig. 1*a* shows the single-camera omni-stereo imaging system used in the experiments. Five frames of omni-stereo images with 500 × 500 resolution are used as the test omni-stereo images, in which a toy car is moving from one corner to the opposite corner along the two walls in a room. Fig. 6 illustrates the overall moving object tracking procedure based on a single-camera omni-stereo DMV. In Fig. 6*a*, five frames of omni-stereo images are illustrated. We can

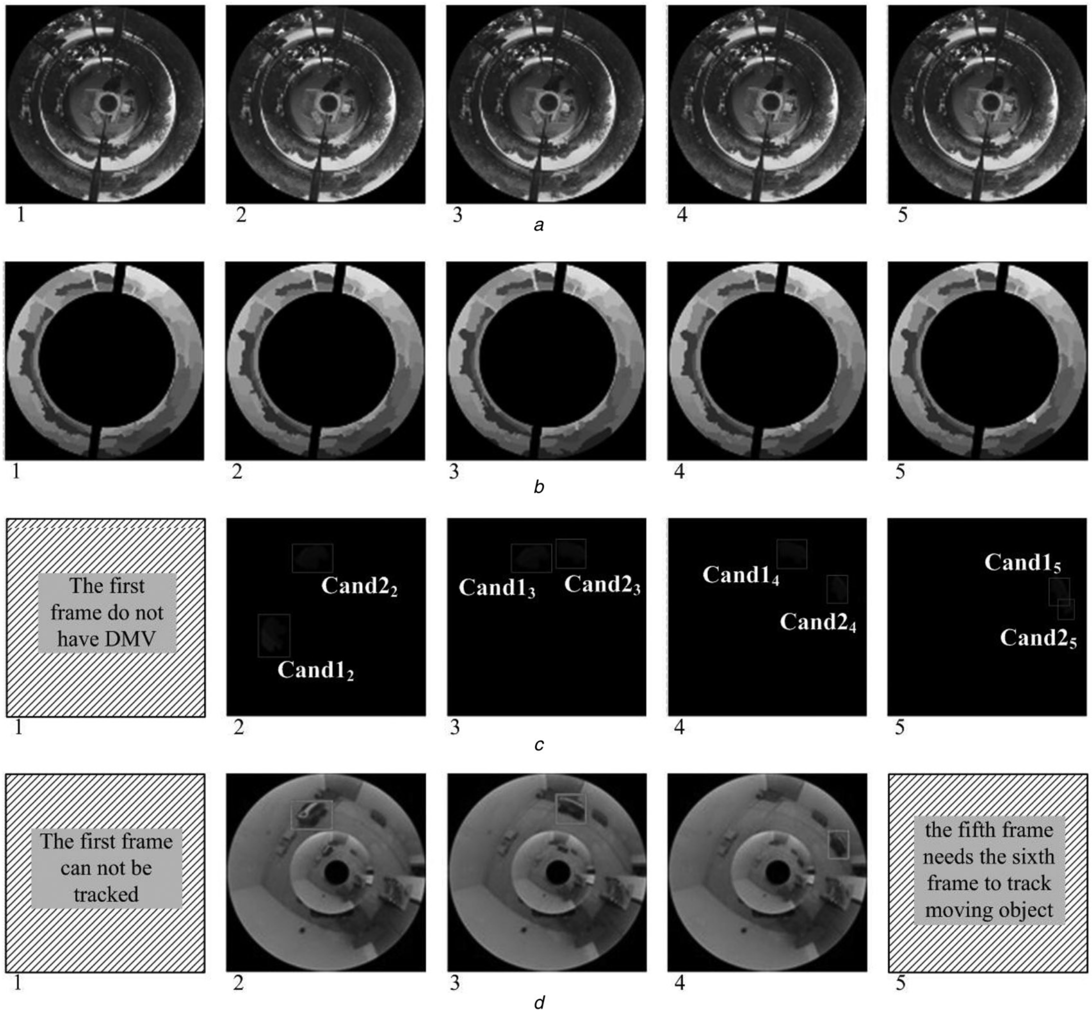

1
2
3
4
5
a
1
2
3
4
5
b
The first frame do not have DMV
Cand2$_2$
Cand1$_2$
Cand1$_3$
Cand2$_3$
Cand1$_4$
Cand2$_4$
Cand1$_5$
Cand2$_5$
1
2
3
4
5
c
The first frame can not be tracked
the fifth frame needs the sixth frame to track moving object
1
2
3
4
5
d

Fig. 6 *Single-camera omni-stereo tracking based on DMV*

a Omni-stereo image sequence
b Disparity map sequence
c DMV sequence
d Tracking the moving object

Table 6 Candidate areas' coordinates

DMV no.	Area	Start(*x*, *y*)	End(*x*, *y*)
2	Cand1_2	(74, 239)	(151, 346)
	Cand2_2	(162, 59)	(260, 130)
3	Cand1_3	(158, 61)	(255, 134)
	Cand2_3	(272, 48)	(348, 120)
4	Cand1_4	(275, 45)	(354, 115)
	Cand2_4	(401, 140)	(451, 208)
5	Cand1_5	(394, 135)	(453, 204)
	Cand2_5	(426, 200)	(465, 249)

Table 7 Common area size of candidate area pairs

DMV no.	Candidate area pair	Common area size (pixels)
2	(Cand1_2, Cand1_3)	0
	(Cand1_2, Cand2_3)	0
	(Cand2_2, Cand1_3)	6417
	(Cand2_2, Cand2_3)	0
3	(Cand1_3, Cand1_4)	0
	(Cand1_3, Cand2_4)	0
	(Cand2_3, Cand1_4)	4891
	(Cand2_3, Cand2_4)	0
4	(Cand1_4, Cand1_5)	0
	(Cand1_4, Cand2_5)	0
	(Cand2_4, Cand1_5)	3200
	(Cand2_4, Cand2_5)	0

see that there are an outer part and an inner part of the scene in each omni-stereo image. Fig. 6*b* shows the disparity maps estimated from the omni-stereo images, the disparity is estimated based on the outer part of each omni-stereo image; this is because the outer part has better resolution. Then, DMV maps are calculated by computing disparity differences between two consecutive disparity maps and candidate areas of the moving object are detected in Fig. 6*c*. In the first frame, there is no DMV, because this frame has no previous frame to perform the disparity difference. Finally, we decide which candidate area in each DMV is the true object area and mark the moving object in the outer part of the omni-stereo image, as shown in Fig. 6*d*.

Now, we discuss how to decide on true object areas. In each DMV of Fig. 6*c*, there are two candidate areas where a moving object may be located, one is the true location of an object in the current frame, and the other is the object location in the previous frame. Table 6 shows the candidate area coordinates in DMV frames 2–5, where Start(*x*, *y*) is the top-left coordinates of the area and End(*x*, *y*) is the bottom-right coordinates. Based on this information, we can calculate the common area sizes for candidate area pairs in consecutive DMVs. The results are shown in Table 7. Candidate pairs (Cand2_2, Cand1_3), (Cand2_3, Cand1_4) and (Cand2_4, Cand1_5) have the largest common area in DMV frames 2, 3 and 4; therefore Cand2_2, Cand2_3 and Cand2_4 are true object locations in omni-stereo image frames 2, 3 and 4.

6 Conclusion

In this article, we proposed using a single-camera omni-stereo imaging system and a depth-space partitioning method for object tracing. The single-camera omni-stereo imaging system captures 360° FOV omni-stereo image sequence, ensuring that there is no need to rotate the camera when tracking a moving object. Since omni-stereo imaging is totally different from traditional perspective-view imaging, and traditional disparity-estimation methods cannot be used for omni-stereo disparity estimation, we propose a depth-space partitioning method to estimate omni-stereo disparity. A moving object is tracked using omni-stereo DMV that is the difference between two consecutive disparity maps. In future work, we will extend the proposed method to track multiple moving objects.

7 Acknowledgments

This research was partially supported by National Natural Science Foundation (NSFC) of China under project numbers 61175015, 61175006, 60705013, 60872150, 60803101 and 60773023; China Postdoctoral Science Foundation special funding with Project No. 200902665, China Postdoctoral Science Foundation with Project No. 20070410977, Natural Science Foundation of Hunan Province in China with Project No. 08JJ4018, iCORE (Alberta Innovates) Alberta and NSERC Canada.

8 References

1 Loza, A., Bull, D., Canagarajah, N., Mihaylova, L.: 'Structural similarity-based object tracking in multimodality surveillance videos', *Mach. Vis. Appl.*, 2009, **20**, (2), pp. 71–83

2 Leibe, B., Schindler, K., Cornelis, N., Van Gool, L.: 'Coupled object detection and tracking from static cameras and moving vehicles', *IEEE Trans. Pattern Anal. Mach. Intell.*, 2008, **30**, (10), pp. 1683–1698

3 Yilmaz, A., Javed, O., Shah, M.: 'Object tracking: a survey', *ACM Comput. Surv.*, 2006, **38**, (13), pp. 1–45

4 Wang, J., Yagi, Y.: 'Integrating color and shape-texture features for adaptive real-time object tracking', *IEEE Trans. Image Process.*, 2008, **17**, (2), pp. 235–240

5 Greminger, M., Nelson, B.: 'A deformable object tracking algorithm based on the boundary element method that is robust to occlusions and spurious edges', *Int. J. Comput. Vis.*, 2008, **78**, (1), pp. 29–45

6 Zhou, H.Y., Yuan, Y., Shi, C.M.: 'Object tracking using SIFT features and mean shift', *Comput. Vis. Image Underst.*, 2009, **113**, (3), pp. 345–352

7 Muñoz-Salinas, R., Aguirre, E., García-Silvente, M., Gonzalez, A.: 'A multiple object tracking approach that combines color and depth information using a confidence measure', *Pattern Recognit. Lett.*, 2008, **29**, (10), pp. 1504–1514

8 Muñoz-Salinas, R., García-Silvente, M., Medina Carnicer, R.: 'Adaptive multi-modal stereo people tracking without background modeling', *J. Vis. Commun. Image Represent.*, 2008, **19**, (2), pp. 75–91

9 Tang, F., Harville, M., Tao, H., Robinson, I.N.: 'Fusion of local appearance with stereo depth for object tracking'. Proc. IEEE Computer Society Conf. on Computer Vision and Pattern Recognition Workshops, Anchorage, Alaska, USA, June 2008, pp. 142–149

10 Fiala, M., Basu, A.: 'Robot navigation using panoramic tracking', *Pattern Recognit.*, 2004, **37**, (11), pp. 2195–2215

11 Bae, K.H., Koo, J.S., Kim, E.S.: 'A new stereo object tracking system using disparity motion vector', *Opt. Commun.*, 2003, **221**, (13), pp. 23–35

12 Boykov, Y., Veksler, O., Zabih, R.: 'Fast approximate energy minimization via graph cuts', *IEEE Trans. Pattern Anal. Mach. Intell.*, 2001, **23**, (11), pp. 1222–1239

13 Liu, X.Q., Veksler, O., Samarabandu, J.: 'Order-preserving moves for graph-cut-based optimization', *IEEE Trans. Pattern Anal. Mach. Intell.*, 2010, **32**, (7), pp. 1182–1196

14 Xiong, Z.H., Zhang, M.J., Wang, Y.L., Li, T., Li, S.K.: 'Fast panorama unrolling of catadioptric omni-directional images for cooperative robot vision system'. Proc. 11th Int. Conf. on Computer Supported Cooperative Work in Design, Melbourne, Australia, April 2007, pp. 1100–1104

15 Andreasson, H., Treptow, A., Duckett, T.: 'Self-localization in non-stationary environments using omni-directional vision', *Robot. Auton. Syst.*, 2007, **55**, (7), pp. 541–551

doi: 10.1049/iet-cvi.2010.0115

16 Wu, C.J., Tsai, W.H.: 'Location estimation for indoor autonomous vehicle navigation by omni-directional vision using circular landmarks on ceilings', *Robot. Auton. Syst.*, 2009, **57**, (5), pp. 546–555

17 Feng, H.M., Chen, C.Y., Horng, J.H.: 'Intelligent omni-directional vision-based mobile robot fuzzy systems design and implementation', *Expert Syst. Appl.*, 2010, **37**, (5), pp. 4009–4019

18 Liu, Y.C., Lin, K.Y., Chen, Y.S.: 'Bird's-eye view vision system for vehicle surrounding monitoring'. Proc. Robot Vision Conf., 2008, (*LNCS* **4931**), pp. 207–218

19 Boult, T.E., Gao, X., Micheals, R.J., Eckmann, M.: 'Omni-directional visual surveillance', *Image Vis. Comput.*, 2004, **22**, (7), pp. 515–534

20 Ng, K.C., Ishiguro, H., Trivedi, M.M., Sogo, T.: 'An integrated surveillance system: human tracking and view synthesis using multiple omni-directional vision sensors', *Image Vis. Comput.*, 2004, **22**, (7), pp. 551–561

21 Southwell, D., Basu, A., Fiala, M., Reyda, J.: 'Panoramic stereo'. Proc. Int. Conf. on Pattern Recognition, Vienna, Austria, 1996, pp. 378–382

22 Fiala, M., Basu, A.: 'Panoramic stereo reconstruction using non-SVP optics'. Proc. Int. Conf. on Pattern Recognition, Quebec City, Canada, August 2002, pp. 11–15

23 Szeliski, R., Zabih, R., Scharstein, D., *et al.*: 'A comparative study of energy minimization methods for Markov random fields'. Proc. European Conf. on Computer Vision, Graz, Austria, May 2006, pp. 16–29

24 Potts, R.: 'Some generalized order-disorder transformation', *Proc. Camb. Philos. Soc.*, 1952, **48**, pp. 106–109

25 Kolmogorov, V.: 'Convergent tree-reweighed message passing for energy minimization', *IEEE Trans. Pattern Anal. Mach. Intell.*, 2006, **28**, (10), pp. 1568–1583

26 Bernier, O., Cheung-Mon-Chan, P., Bouguet, A.: 'Fast nonparametric belief propagation for real-time stereo articulated body tracking', *Comput. Vis. Image Underst.*, 2009, **113**, (1), pp. 29–47

27 Ahuja, R.K., Magnanti, T.L., Orlin, J.B.: 'Network flows: theory, algorithms, and applications' (Prentice-Hall, 1993)