The Matching of Infrared Markers for Tracking Objects Using Stereo Pairs

R. Sh. Zeynalov, A. A. Yakubenko, and A. S. Konushin

Faculty of Computational Mathematics and Cybernetics, Moscow State University, Moscow, Russia
e-mail: ramiz.zeynaIov@gmaiI.com

Abstract—This paper presents a new algorithm for obtaining inter-frame and inter-view (inter-camera) correspondences to solve the problem of tracking an object labeled with infrared markers using a stereo pair taken simultaneously in the infrared region. In practice it is often necessary to track an object when it is impossible to have contact with it, for example, the tracking of facial movements using the motion capture technique (Motion Capture, [9, 10]) to create realistic animation or the tracking of object movements when interacting with the augmented reality. In such cases contactless object tracking methods are used. In the classic version of this problem, two or more cameras are used to capture the object of interest. In order to restore three-dimensional coordinates of object points, it is necessary to triangulate the received projections of the points. In the case of the visible range, the problem of finding and matching points on the object can be solved using interest point descriptors [2]. However, there are situations in which it is impossible to use the visible range data, for example, a uniformly colored or regularly textured object, which makes it senseless to use interest point descriptors. Thus, the use of interest point descriptors significantly limits the scope of application of algorithms because of the imposition of severe restrictions on the class of tracked objects, objects should have uneven texture. In turn, the motion capture method implies the tracking of an object of a predetermined shape, when it is not always possible to establish its shape. In this work, an alternative approach is proposed, i.e., the use of infrared markers and cameras that capture frames in the infrared range which makes the task of finding critical points irrelevant. On the other hand, infrared markers on cameras that operate in the infrared range are indistinguishable from each other. Therefore, there is the problem of finding correspondences which in the case of interest point descriptors is solved by the very nature of the descriptors. In this paper, we describe algorithms that make it possible to restore point correspondences by a sequence of stereo pair images (Fig. 1) and its calibrations. In this case, epipolar constraints and a voting scheme based on the greedy algorithm are used.

Keywords: tracking, matching, infrared markers, epipolar constraints, triangulation.

DOI: 10.1134/S1054661813040196

INTRODUCTION

The task of tracking of an object in space is found in many practical applications. For example, in order to create a three-dimensional animation model, for human interaction with the virtual reality to create augmented reality, and for detecting the motion of objects during the field tests and physical experiments for the subsequent objective evaluation of experimental results, the motion capture method (Motion Capture [9, 10]) is used. This usually requires a quite accurate evaluation of motion characteristics such as linear and angular velocity and acceleration. Moreover, for information gathering only contactless devices such as cameras can be used. At present, there are many different approaches to solving this problem. Solutions differ in both the hardware used and the algorithms. Recently, the methods of computer vision have become popular in which camera is used as a data acquisition unit. It is known that under certain conditions, by knowing the coordinates of projections of the point on two cameras, the parameters of the devices, and their mutual arrangement, it is possible to restore the three-dimensional coordinates of the point using the so-called triangulation [1]. In cases with many pixels, there is the problem of restoring the correspondence of point projections on different cameras. For this purpose, interest point descriptors are usually used, e.g., SIFT and SURF [2]. In order to use them effectively, it is necessary that the object be unevenly textured. In cases where the object is homogeneous, it is impossible to find interest points on it or guarantee the accuracy of inter-view matches because there are too many similar points that will be erroneously matched.

The uniformly colored object problem is solved by self-identifying markers [3], i.e., sufficiently large plane images in which some unique identifier is encoded. These identifiers can easily and accurately be detected in the case of different spatial orientations of the marker relative to the camera. However, this method is not suitable in cases where the object moves too enough and/or at a long distance from the camera. As a result, such a marker is blurred (a motion-blur effect).

There is another method to track an object using cameras. It consists in the use of infrared markers

Received June 1, 2012

together with cameras that shoot solely in the IR range, which makes it possible to track spatial movements of individual point markers (Fig. 2). In this case, when the observed object moves at a sufficiently high speed, the blur on marker images (Fig. 3) does not greatly affect the recognition of markers and calculation of their coordinates in camera images. At the same time, the problem of matching points is solved algorithmically using epipolar constraints [4]. In this paper, we propose a new method for measuring the characteristics of the object motion.

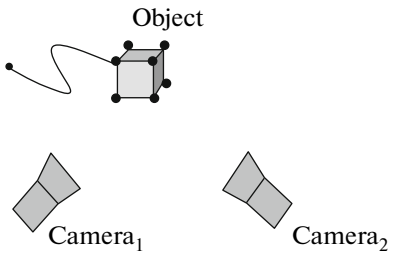

Fig. 1 An object and a stereo pair.

Fig. 2. A fragment and data.

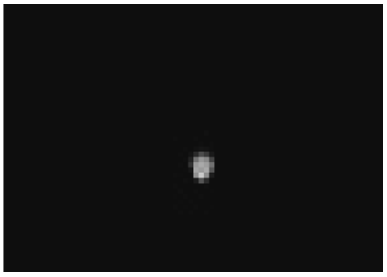

Fig. 3. A marker and a fragment.

CALCULATION OF INTER-FRAME CORRESPONDENCES

At the input two sequences of frames are given that contain sets of projection points and coordinates on the image. The corresponding pairs of frames of sequences with the same number were taken at one point in time; i.e., the sequences are synchronized. This is ensured by special equipment that controls the operation of a stereo pair. For one and the same point in space, there is no correspondence between its projections on different frames, while these correspondences are necessary for the trajectory estimation. Thus, it is necessary to restore these correspondences. More formally the statement of the problem is as follows.

Given

(1) the calibration of a stereo pair (internal parameters of the cameras K_c, lens distortion D_c, and their position T_c) and orientation

$$C_c = \langle K_c; D_c; R_c; T_c \rangle, \quad c = 1, 2, \tag{1}$$

(2) two sequences of frames (one per camera) consisting of a plurality of projections of points (**c** is the number of the camera, n_i^c is the number of projections $q_l^{c,i}$ at the frame i of the camera c, $l_{i,j}$ is the number of the jth point on the ith frame, F_i^c is the frame, S_c is the frame sequence, N is the number of frames)

$$F_i^c = \{p_{l_{i,j}}^{c,i}, j = 1 \ldots n_i^c\}, \tag{2}$$

$$S^c = \{F_i^c\}, \quad i = 1 \ldots N, \quad c = 1 \ldots 2, \tag{3}$$

it is required

(1) to determine the inter-frame correspondences between the points (M^c is the number of projections of the points corresponding to different points in space, $p_j^{c,i}$ are projections of points sorted according to inter-frame correspondences, $l(i, j, c)$ is the number of the q projection corresponding to the point j on the frame i of the camera c)

$$C_F^c: \; q_{l(i,j,c)}^{c,i} \to p_j^{c,i}, \tag{4}$$

and

(2) to determine the inter-view correspondences of point projections (the set of pairs of indexes j_1 and j_2 for which the equivalence relation that determines the correspondence of all the projections of a camera for one point to all projections of one point on another camera)

$$C_V^{1,2}: \; p_{j_1}^{1,i} \times p_{j_2}^{2,i} \to \{True, False\}. \tag{5}$$

The conducted experiments show that in order to obtain inter-frame correspondences it is sufficient to use a simple algorithm based on the inter-frame displacement. Situations in which the algorithm can malfunction are extremely rare in practice, especially with a small number of markers.

Indeed, markers are diodes that emit in the infrared range. The radiation angle is limited (Fig. 4). As a rule the half angle is less than 30°–45° (in the context of measuring the diode radiation angle, the half angle is usually meant). Thus, assuming that one of the cameras observes two diodes in one ray, we obtain that the ray is parallel to the plane of the object on which the diodes are located. Thus, the angle between the direction of the diode (plane normal) and the direction of the optical center of the camera is right. It is signifi-

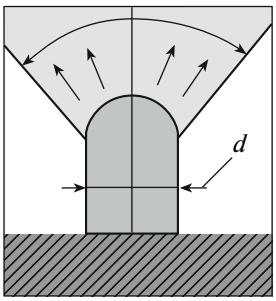

Fig. 4. A marker model.

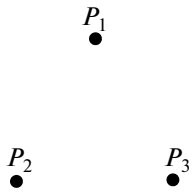

Fig. 5. The first frame.

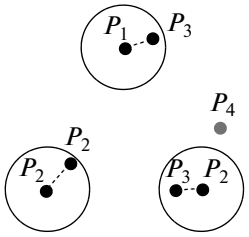

Fig. 6. The second frame.

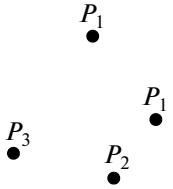

Fig. 7. The first frame.

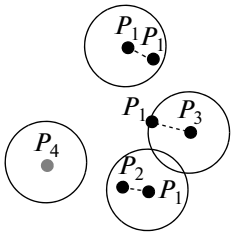

Fig. 8. The second frame.

cantly larger than the diode radiation half angle. It means that such markers will not be detected by the camera because only a tiny fraction of the diode radiation energy will reach the camera matrix.

Let us introduce a threshold that characterizes the allowable displacement of the projection of the point on the image between adjacent frames. Consider the first frame of the sequence.

We assign a unique identifier, a natural number, to each projection in this frame (Fig. 7). The points P_1, P_2, and P_3 get identities 1, 2, and 3, respectively. Then, we process the next frame in a similar manner; i.e., for each projection in the previous frame, we find the closest unmarked projection on the current frame. If the distance between them (in the image plane) is less than the threshold, we mark the projection on the new frame with the same identifier as on the previous frame.

We mark each unmarked projection on a new frame with a new identifier (Fig. 6). If at any time there are no more projections on the current frame for the projection from the previous frame, this point has just appeared (Fig. 6), i.e., the point P_4. If for some point on the previous frame it has not been possible to find a point on the current frame, it means that this point has disappeared (Figs. 7, 8), i.e., the point P_2. As noted above, this algorithm is very fast and the situations in which its result can be false are extremely unlikely.

CALCULATION OF INTER-VIEW CORRESPONDENCES

In order to calculate inter-view correspondences, epipolar constraints (6) are used [4]; i.e., for a point in space its projection p on one camera limits the position of the projection q of the same point on the other camera by the epipolar line l_q (Figs. 9, 10), which is set by the fundamental matrix F and vice versa (7).

The algorithm is divided into two stages, i.e., the calculation of possible inter-view correspondences for a pair of frames and the calculation of inter-view correspondences for the entire sequence of pairs of frames.

$$q^T F p = 0, \tag{6}$$

$$\begin{cases} l_q = Fp \\ l_p = q^T F. \end{cases} \tag{7}$$

The possible inter-view correspondences for a frame are calculated as follows: for each point p_j on each camera, all points of $\{q_l\}$ on the other camera are selected in the vicinity of the corresponding epipolar line. All pairs $\langle p_j, q_l \rangle$ are added to the list of possible correspondences for a given pair of frames. These operations are carried out for all pairs of frames.

Because of errors in individual frames, there can be sets of possible correspondences contradicting each other. In order to avoid them, the voting scheme is used. Let M_1 be all the different points in the sequence of

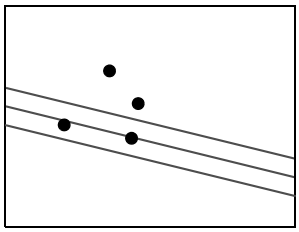

Fig. 9. The left frame.

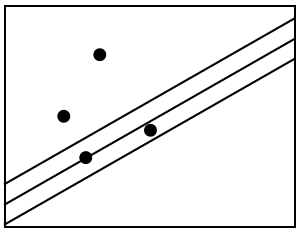

Fig. 10. The right frame.

images from one camera, and let M_2 be all such points from the second camera. We introduce the matrix MC filled with zeros. For each match $\langle p_j^1, p_l^2 \rangle$ of each pair of frames, we increase the element $MC_{j,l}$ of the matrix MC by one. As a result, in each element of the matrix there will be the number of votes for the match $\langle p_j^1, p_l^2 \rangle$. In order to obtain a set of correspondences for the whole sequence of pairs of frames, we will extract the maximum element from the matrix MC while it contains elements with values greater than the threshold. The voting scheme described eliminates impacts of emissions on the condition that there are not many of them.

RESTORATION OF LOST CORRESPONDENCES

Some points could disappear from cameras on some interval and then reappear. However, they will be recorded as new points. This situation is illustrated well by the following example: an object with markers $P_o = \{P_i, i = 1...N_o\}$ is detected by both cameras. At the same time, each camera C_c records its own subsets $P_c \subseteq P_o$, and both cameras detect points from the set $P_x = P_1 \cap P_2$. At the next moment t_0, a point $p_x \in P_x$ disappears from the cameras; i.e., at some time the point $p_x \notin P_x$. Then, at time $t_1 > t_0$, this point re-enters the set P_x. In this case, the inter-frame identification algorithm and thus the inter-view identification algorithm will assign a new ID to this point and the equivalence relation of this point at different times will be lost.

In order to solve this problem, individual fragments of the object are considered for which it is possible to construct a general coordinate system. Such fragments can be combined with the above points coinciding with a certain accuracy. For the identification of matching points, the threshold of the measurement error of coordinates of points is introduced: if after the recalculation of the coordinates of points in a coordinate system two points get the coordinates that coincide within the selected threshold, then these points are considered one and the same point and are combined. The stereo frame sequence is divided into so-called ranges, i.e., continuous sequences of pairs of frames in which sets P_x are equal, or in other words the stereo sequence is divided into equivalence classes by the relation of equality of sets P_x and the relation of proximity of times. Let the operator $vis\_at(i)$ convert the index of the pair of frames i (range R_k or time $t = i*Fps$, where Fps is the number of frames per second) in a set of visible points P_x at a given time. Then for the ranges $R_1 = (i_1^1 ... i_2^1)$ and $R_2 = (i_1^2 ... i_2^2)$, for which $card(vis\_at(R_1) \cap vis\_at(R_2)) \geq 3$ ($card(A)$ is the cardinality A), i.e., the total number of points is not less than three, it is possible to determine an operator $R_{1,2} = join\_ranges(R_1, R_2)$ that will bring together the two ranges into one (in general, not continuous). In this case, the resulting ranges will have points of $P_u = vis\_at(R_1) U vis\_at(R_2)$ coordinates that can be converted by means of a general coordinate system. In turn, this coordinate system can be constructed by three points which by the initial condition will be in both ranges. This procedure is described in more detail in [15, 16].

EXPERIMENTAL

The experimental verification of the developed algorithms was carried out on real and synthetic data. However, it was problematic to conduct experiments with real data. Thus, these experiments were carried out on a limited basis.

In order to evaluate the match search and reconstruction algorithm, metrics were used, i.e., the reprojection error and the ratio of the number of point projections for which it was possible to find matches to the total number of projections. Real data, pure synthetic data, and synthetic data with noise were used as test data. All data contained the object motion with markers in space combined with its rotation. Tests have shown that the reprojection error is less than 1 pixel on noisy data and does not exceed 0.2 pixel on pure synthetic data. The percentage of reprojection points used for the reconstruction depending on the sequence and noise is in the range of approximately 60–70% for noisy data and up to 80% for pure synthetic data. All the found projections cannot be used because very often the markers are visible on one camera while on the other camera they are not seen.

In order to evaluate the static error, an object with a size of 20 cm at a distance of 5 meters from both cameras was measured (the calibration of the system was carried out for the same distance between the template and the cameras). The calculated value was

found to be 20.09 cm. It corresponds to a relative error of measurement of 0.45% in this experiment.

In order to assess the quality of the algorithms, complex experiments were conducted that consisted in calculating the acceleration of gravity (including air resistance and without it). To do this, a physical model was used for which it the analytical solution was known. The resulting coordinate values were substituted into the analytical equation for the selected model, which made it possible to obtain the coefficients of equations. These coefficients are expressed in terms of the acceleration of gravity and the coefficient of the linear part of air resistance. For example, free fall is represented by (8) and (9) without air resistance and with it.

$$m\frac{\partial^2 x}{\partial t^2} = mg, \tag{8}$$

$$m\frac{\partial^2 x}{\partial t^2} = mg - m\alpha\frac{\partial x}{\partial t}. \tag{9}$$

Experiments have shown that air resistance is negligible and, hence, can be neglected. The acceleration of gravity was calculated at 10.45 m/s^2.

CONCLUSIONS

In this paper, we proposed a new algorithm for tracking the motion of an object in space and contactless calculation of characteristics of its motion. An object with infrared markers installed on it is used as a data source. This object is observed by two cameras that can operate in the infrared range.

In order to solve the problem of matching point projections, a new method is proposed based on the use of epipolar constraints and the voting scheme to improve the robustness of the results.

Experiments have been conducted to assess the quality of the proposed algorithms and the accuracy of measurements. For the static measurement of distances, an accuracy of about 0.45% was obtained. The measured acceleration of gravity was 10.45 m/s^2.

REFERENCES

1. R. Hartley and P. Sturm, "Triangulation," in *Proc. ARPA Image Understanding Workshop* (Monterey, CA, 1994).
2. H. Bay, T. Tuytelaars, and L. V. Gool, "SURF: speeded up robust features," Comput. Vision Image Understanding **110** (3), 346–359 (2008).
3. M. Fiala and C. Shu, "Self-identifying patterns for plane-based camera calibration," Mach. Vision Appl. **19** (4), 209–216 (2008).
4. R. Hartley and A. Zisserman, *Multiple View Geometry in Computer Vision* (Cambridge, Cambridge Univ. Press, 2000).
5. A. Yilmaz and M. Shah, "Contour-based object tracking with occlusion handling in video acquired using mobile cameras," IEEE Trans. Pattern Anal. Mach. Intellig. **26**, 1531–1536 (2004).
6. Z. Khan, R. Herman, K. Wallen, and T. Balch, "An outdoor 3-D visual tracking system for the study of spatial navigation and memory in rhesus monkeys," Behavior Res. Meth. **37** (2005).
7. S. Smith, "Real-time motion segmentation and shape tracking," in *Proc. 5th Int. Conf. On Computer Vision* (Cambridge, MA, 1995).
8. M. Loaiza, A. Raposo, and M. Gattass, *A Novel Optical Tracking Algorithm for Point-Based Projective Invariant Marker Patterns* (2007).
9. M. Weber, H. B. Amor, and T. Alexander, "Identifying motion capture tracking markers with self-organizing maps," in *Proc. IEEE Virtual Reality Conf., VR'08* (Piscataway, NJ, 2008), pp. 297–298.
10. Y. Zhao, J. Westhues, P. Dietz, J. Barnwell, S. Nayar, M. Inami, M. Nol, V. Branzoi, and E. Bruns, "Lighting aware motion capture using photosensing markers and multiplexed illuminators," ACM TOG **26** (3) (2007).
11. H. Hatze, "High-precision threedimensional photogrammetric calibration and object space reconstruction using a modified DLT-approach," J. Biomech. **21**, 533–538 (1988).
12. G. Welch and G. Bishop, *An Introduction to the Kalman Filter* (Addison-Wesley ACM Press, 1995), pp. 1–16.
13. B. Triggs, Ph. Mclauchlan, R. Hartley, and A. Fitzgibbo, "Bundle adjustment – a modern synthesis," in *Vision Algorithms: Theory and Practice* (Springer Verlag, 2000), pp. 298–375.
14. Ranganathan Ananth, "The Levenberg-Marquardt algorithm 3 LM as a blend of gradient descent and Gauss-Newton itera," Georgia Tech. College of Computing (2004), pp. 1–5.
15. R. Zeynalov, A. Yakubenko, I. Tolkunov, and A. Machikhin, "Object tracking using infrared markers," in *Proc. Graphicon* (Moscow, 2011), pp. 263–266.
16. R. Zeynalov, A. Yakubenko, and A. Konushin, "Infrared marker matching for object tracking in stereo setup," in *Proc. Open German-Russian Workshop at Pattern Recognition and Image Understanding* (Nizni Novgorod, 2011), pp. 369–372.

Translated by O. Pismenov

Ramiz Shakirovich Zeinalov. Born in 1986. In 2004, entered the Faculty of Computational Mathematics and Cybernetics of Moscow State University. In 2009, defended his graduation thesis and continues his studies in graduate school. Scientific interests include computer vision, threedimensional reconstruction, and photogrammetry. Number of publications: 12.

Anton Sergeevich Konushin. Defended Candidate thesis on "Automatic Reconstruction of Three-Dimensional Models by a Sequence of Images" in 2005 at the Institute of Applied Mathematics. Since 2010 he has been a lecturer at the Yandex School of Data Analysis. Scientific interests include computer vision. Number of publications: 43.

Anton Anatol'evich Yakubenko. Graduated from the Faculty of Computational Mathematics and Cybernetics, Moscow State University, in 2007 and continues his studies in graduate school. Scientific interests include three-dimensional modeling of cities, three-dimensional reconstruction, photogrammetry, and processing and analysis of images. Number of publications: 12.