Identification of the Background from Video Based on Motion Constraints

YU Yongyan[1,a] , SHU Yuqin[2,b]

[1]Department of Computer Engineering,Huaiyin Institute of Technology,Huaian 223003,China

[2]Huaian Qinyin Decoration company,Huaian 223001,China

[a]Shanshan_yyy@163.com , [b]13378402263@qq.com

Abstract

To identity of the background from video is one of the most important topic in 3D reconstruction about dynamic scenes,which would give a lot of information and give the simplest interpretation for the overall scene motion.This paper attempt to lift that limitation and the proposed techniques are based on the fact that the aforementioned motion characteristics of an object are lost not only in the case of wrong relative scales but also in the case of a wrong background identification,so will propose two techniques based on the independence criterion and the non-accidentalness principle.The independence criterion is applicable to any scene, however requires many frames to be statistically valid and needs variation in the motion parameters.

Keywords: relative scale, independence constraint, non-accidentalness constraint,background identification

1 Introduction

Video understanding has a wide range of application within video indexing, robot navigation and human-computer interaction[1]. Mostly, motion features arise from the relative motion between the different objects in the scene and the camera.The assumption that the sensor remains stationary between the incidence of each video frame allows the use of statistical background modeling techniques for the detection of moving objects[2,3].3D reconstruction of dynamic scenes poses various challenges,in which the important and subtle one is to obtain unknown relative scale between the background and the foreground.Therefore,for many applications,background identification is the first step[4,5].

In a special cases that the background is also moving strongly in the video images, the background is often identified on the basis of 2D image related features,such as relative size, spread of texture [6], visibility [7，8], symmetry[9,10].However, the moving objects can almost fill the screen, can move behind the static scene, or can cover an entire image border. It must be noted that the figure/ground problem involves a depth ordering of the objects in the image.However,the objects are not always segmented according to their depth-ordering but possibly according to their 3D rigid motion which is a problem for traditional cues. If 3D analysis of these video shots is possible, it can offer more powerful solutions.The solutions are based on the motion constraints approach,which noted that there is a relative scale ambiguity between the reconstructions of independently moving components of a dynamic scene.It was shown that for relative scale values other than the actual one, the object trajectories lose some of the properties that are quite common in real-life objects.

However in the methods described so far, the background had to be identified beforehand. If not, this adds an additional challenge.We will attempt to lift that limitation,and propose two techniques based on the independence criterion and the non-accidentalness principle.

2 Relative Scale between Background and Foreground

It is known that from an uncalibrated monocular image sequence, we can only come up with a reconstruction up to an unknown overall scale.Lack of information on the relative scales leads to one-parameter families of possible trajectories of the objects with respect to the static background.Consider the example of a video of a moving woman in market. Without incorporating further knowledge about the world, a computer cannot distinguish between a small object hovering in front of the camera and a real persion at a larger distance on the room.

For simplification,we assume that segmentation has been done as a preprocessing step [11], and that the moving objects are rigid, we want to reconstruct the trajectories of the different dynamic parts of the scene with respect to each other. We require the segmentation to be sufficiently precise in order to enable an uncalibrated SfM algorithm to extract robust projection matrices[12].

When the scene contains different rigid parts, moving independently of each other, there is a problem in deciding on the relative scale of the translation, but not on the rotation.The relative rotations are fixed at each time instant and not affected by different scale factors unlike the relative translations. For each different relative scale factor between the background and the independently moving object, a different trajectory for the object relative to the background will result.

Consider an image sequence of a scene which is static except for one rigid, independently moving object. Additional moving objects can be dealt with similarly.Suppose we can compute the camera's orientation $\mathbf{R}_c^i$, $\mathbf{R}_x^i$ and position $\mathbf{t}_c^i$, $\mathbf{t}_x^i$ relative to the static part of the scene and with respect to the segmented moving object for every frame i of the sequence.What we would like to find is the rotation $\mathbf{R}_o^i$ and the translation $\mathbf{t}_o^i$ which represent the motion of the object with respect to the background for every frame i.The relation among them can be written as:

$$\begin{bmatrix} \mathbf{R}_x^T & -\mathbf{R}_x^T\mathbf{t}_x \\ \mathbf{0} & 1 \end{bmatrix} = \begin{bmatrix} \mathbf{R}_c^T & -\mathbf{R}_c^T\mathbf{t}_c \\ \mathbf{0} & 1 \end{bmatrix} \begin{bmatrix} \mathbf{R}_o & \mathbf{t}_o \\ \mathbf{0} & 1 \end{bmatrix} \tag{1}$$

Obviously,if we already know $\mathbf{R}_x$ and $\mathbf{R}_c$ in addition to the exact $\mathbf{t}_x$ and $\mathbf{t}_c$, we can extract $\mathbf{R}_o$ and $\mathbf{t}_o$.

Unfortunately,uncalibrated SfM cannot extract the camera motion with respect to the background and the object at an absolute scale due to a scale ambiguity in the translation components.We are free to fix the scale for one, say the background, but there still remains the relative scale to deal with.As we do not know which scale to apply, each incorrect scale $s \neq 1$ applied to $\mathbf{t}_x$ will yield a different object trajectory $\mathbf{t}_{os}$:

$$s(-\mathbf{R}_x^T\mathbf{t}_x) = \mathbf{R}_c^T\mathbf{t}_{os} - \mathbf{R}_c^T\mathbf{t}_c \tag{2}$$

Merging Eqs. (1) and (2) yields the following relation between the actual trajectory of the object $\mathbf{t}_o$ and the computed trajectory $\mathbf{t}_{os}$ of the object when an incorrect relative scale factor $s \neq 1$ is used:

$$s(\mathbf{R}_c^T\mathbf{t}_o - \mathbf{R}_c^T\mathbf{t}_c) = \mathbf{R}_c^T\mathbf{t}_{os} - \mathbf{R}_c^T\mathbf{t}_c \tag{3}$$

Multiplying both sides with Rc leads to

$$\mathbf{t}_{os} = s\mathbf{t}_o + (1-s)\mathbf{t}_c \tag{4}$$

Hence, the object translation $\mathbf{t}_{os}$ found for the relative scale s is a linear combination of the true object translation $\mathbf{t}_o$ and the camera translation $\mathbf{t}_c$ .When $s = 1$, i.e. at the correct scale, $\mathbf{t}_{os}$ equals $\mathbf{t}_o$. For values of s other than 1, $\mathbf{t}_{os}$ will always be contaminated with the camera translation. When $s \to 0$, $\mathbf{t}_{os}$ evolves towards the camera path.

3.Background Identification via Independence Constraint

Eq.(4) states that the reconstructed object trajectory $\mathbf{t}_{os}$ is a mixture of the original object trajectory $\mathbf{t}_o$ and the camera trajectory $\mathbf{t}_c$ .As to the independence criterion, we try to find the relative scale $m = 1/s$ which makes the resulting object trajectory statistically the most independent of the camera's trajectory.Such properties are not only lost when a wrong relative scale is chosen but also when a wrong scene element is used as the 'background'.If the true object and camera motion are not linearly dependent, a linear dependence will only appear for the wrong relative scales.

To give an intuitive feeling,consider a scenario where a camera is moving slowly on a linear path and an object is moving randomly in front of the camera. The camera path and the object path would look quite dissimilar.However, if we consider the moving object as the static background, the actual background would look as if it moves randomly and the camera path would also have this motion in addition to its own linear path.

Hence, a linear dependence pops up between the camera path and the background path. To state it more formally, let us write the camera motion and the background motion matrices relative to the moving object. The relative motion of the background is the inverse of the object motion:

$$\mathbf{T}_{bo} = \begin{bmatrix} \mathbf{R}_o^T & -\mathbf{R}_o^T \mathbf{t}_o \\ 0 & 1 \end{bmatrix} \tag{5}$$

and the camera motion relative to the moving object can be derived as:

$$\mathbf{T}_{co} = \begin{bmatrix} \mathbf{R}_x & \mathbf{t}_x \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} \mathbf{R}_o^T \mathbf{R}_c & \mathbf{R}_o^T \mathbf{t}_c - \mathbf{R}_o^T \mathbf{t}_o \\ 0 & 1 \end{bmatrix} \tag{6}$$

Apparently, the translation components of $\mathbf{T}_{bo}$ and $\mathbf{T}_{co}$ become linearly dependent due to the additive components $-\mathbf{R}_o^T \mathbf{t}_o$.

4 Background Identification with Non- accidentalness Constraint

As to the non-accidentalness criterion, we exploit the fact that the additive components from the camera trajectory at the wrong relative scales would cause the object motion to lose special properties which many typical moving objects in real life possess.

Many types of moving objects, such as humans, bikes etc. have a natural frontal side and therefore heading direction.Hence, these heading directions or vectors are usually parallel to the tangent of the object trajectory. The mathematical equation describing this heading constraint is:

$$l^{ij}\mathbf{R}_o^{ij}\mathbf{v}_o^i = \mathbf{v}_o^j \tag{7}$$

where $\mathbf{R}_o^{ij}$ is the rotation of the object from frame i to frame j . l^{ij} is a scale factor due to acceleration. $\mathbf{v}_o^i$ is the tangent to the object's trajectory at frame i which can be approximated by:

$$\mathbf{v}_o^i = g_o^{i+1} - g_o^{i-1} \tag{8}$$

where g_o^i is the position of the centroid of the object at i^{th} frame.This is a valid approximation since we generally use video sequences with relatively high frame rates. A similar expression is also used for the approximation of the camera velocities.

Eq.(7) prescribes that the trajectory tangent vector remains tangent when rigidly attached to the object.It describes a coupling between the object translation and the rotation. We expect such a coupling to vanish in the case of a wrong relative scale due to added camera components.Unlike for the independence criterion,theoretically two frames can be enough to solve for the relative scale and could use a RANSAC[13] scheme to estimate it robustly.

But, the heading constraint is not symmetrically defined.In other words,if an object is moving according to the heading constraint, it does not necessarily mean that the background's relative motion with respect to the object also complies with the heading constraint.We use the asymmetrical nature of the heading constraint for the detection of the background. However, the uncertainty about the relative scale between the different reconstructions of the objects in the scene should be taken into account.

5.Experiment and Result

Fig.1 show one of the four input sequences on which we tested our strategy. A woman is taking a walk in indoor space. During the course of the video clip, the person moves rigidly with the bag so both are reconstructed as a single object. The camera's motion with respect to the static background is mostly backwards although with arbitrary movements. This motion enables us to reconstruct the indoor scenes itself.

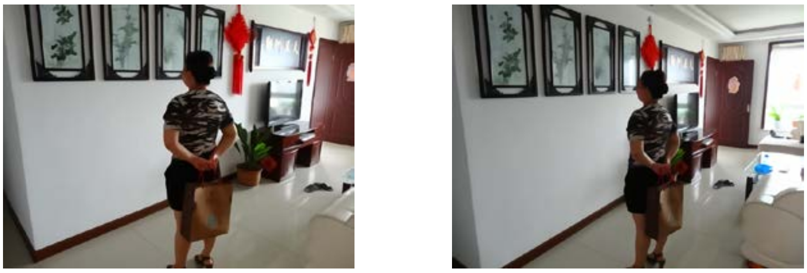

Figure 1: Image samples from the market sequence which contains one moving object except from the background

Instead,in Fig.2, a woman is walking while holding a bag rigidly. The upper torso, the head and the bag are reconstructed as single object. But,the legs are not included since they do not move rigidly.

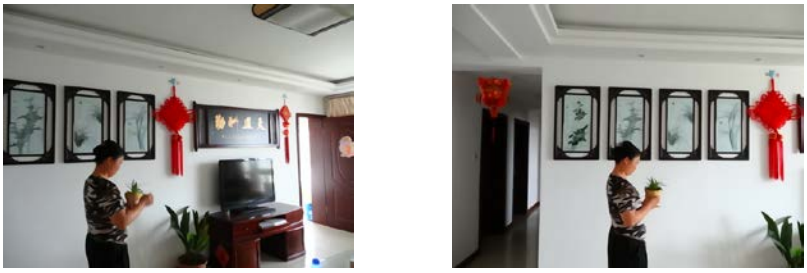

Figure 2: Image samples from the box sequence which contains one moving object except from the background

There are several points to note, the images are segmented beforehand with a semi-manual technique and an iterative perspective SfM algorithm is run over those individual segments.

6. Conclusion

Tthe background could be identified as the object.The independence criterion is applicable to any scene, however requires many frames to be statistically valid and needs variation in the motion parameters.On the other hand,the heading constraint is rather practical,since it requires a small number of frames and many real world objects follow non-holonomic motion. However it has certain degenerate cases, such as when all the objects follow linear paths in the same direction. Although we conducted successful experiments, we are aware that there are still some unexplored phenomena. This work may also pave the way towards a wider rank constraint. The background tends to be the object which results in the smallest overall rank of the object motions in the scene. An optimal method which combines all the proposed methods should be investigated further. An interesting study would be whether human visual system is using such kind of motion simplicity assumptions to detect an object as the background.

Reference

[1] A. Elgammal, D. Harwood, and L. Davis, "Background and Foreground Modeling Using Non-Parametric Kernel Density Estimation for Visual Surveillance," Proc. IEEE, 2002

[2] C. Stauffer and W. Grimson, "Learning Patterns of Activity Using Real-Time Tracking," IEEE Trans. Pattern Analysis and Machine Intelligence, 2000.

[3] C. Wren, A. Azarbayejani, T. Darrel, and A. Pentland, "Pfinder: Real Time Tracking of the Human Body," IEEE Trans. Pattern Analysis and Machine Intelligence, 1997.

[4] Vidal R., Soatto S., Ma Y., Sastry S.: Segmentation of dynamic scenes from the multibody fundamental matrix. In ECCV Workshop on Vision and Modeling of Dynamic Scenes (2002).

[5] Costeira J., Kanade T.: A multi-body factorization method for motion analysis. In International Conference on Computer Vision (1995), pp. 1071–1076.

[6] Vidal R., Soatto S., Ma Y., Sastry S.: A factorization method for 3D multi-body motion estimation and segmentation.
Tech. rep., 2002.

[7] Rubin E.: Visuell wahrgenommene Figuren. Copenhagen:Gyldendals, 1921.

[8] Rubin N., Nakayama K., Shapley R.: Enhanced perception of illusory contours in the lower versus upper visual hemifields. Science 271 (1996), 651–653.

[9] Stricker M., Leonardis A.: Figure-ground segmentation using tabu search. In Proc. of the IEEE Intern. Symposium on Computer Vision (1995), pp. 605–61.

[10] Pao H., Geiger D., Rubin N.: Measuring convexity for figure/ground separation. In International Conference on Computer Vision (1999), pp. 948–955.

[11] Vidal R., Ma Y.: A unified algebraic approach to 2-d and 3-d motion segmentation. In European Conference on Computer Vision (2004), vol. 1, pp.

1–15.

[12] Faugeras O., Luong Q.-T.: The geometry of multiple images.the MIT Press, 2001.

[13] Fischler M., Bolles R.: Random sampling consensus: a paradigm for model fitting with application to image analysis and automated cartography. Communications of the Association for Computing Machinery 24 (1981), 381–395.