Relative Epipolar Motion of Tracked Features for Correspondence in Binocular Stereo

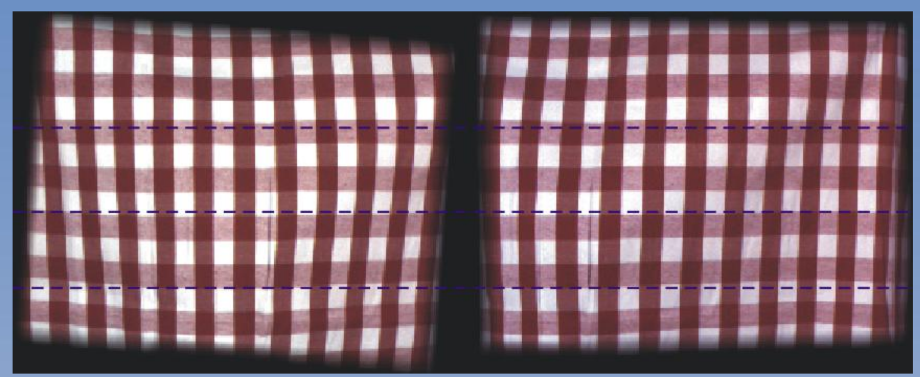
Hao Du, Danping Zou and Yan Qiu Chen

Department of Computer Science and Engineering, Fudan University, Shanghai, China

{duhao | dpzou | chenyq}@fudan.edu.cn

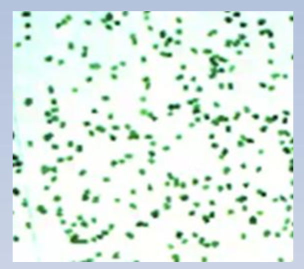
The Problem

For 3D reconstruction using binocular stereo: How do we establish the feature correspondence given dynamic scenes with indistinguishable features?

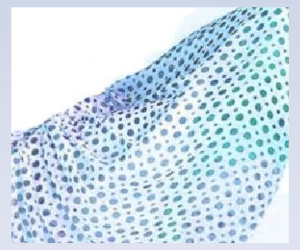


A frame pair captured by stereo cameras: it is difficult to tell the true correspondence even if the image pair is epipolar-rectified.

Applications



Scenes with large amount of undistinguishable drifting particles, e.g. fish schools, bird flocks, microscopic particles.



Repetitively textured surfaces with undistinguishable features.

Textureless surfaces with actively projected particle-like patterns.

The Goal

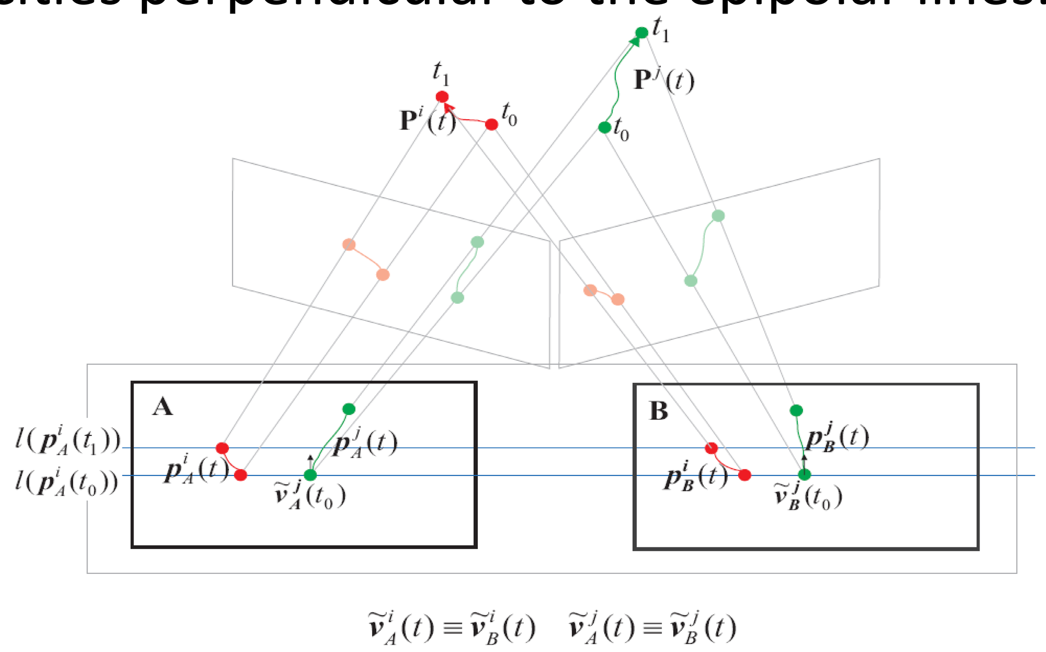
The 3D reconstruction of '*dynamic scenes*' with large amount of '*indistinguishable features*'.

Motivation & Principle

Motion clues

\# 1. True matching remains on the same epipolar lines no matter how the scene particle moves.

\# 2. True matching possesses the same motion velocities perpendicular to the epipolar lines.



Two scene particles P^i, P^j move from time instance t_0 to t_1.

The Approach

Trajectory Tracking. Features are tracked for getting their trajectories. Tracking ambiguities are removed by cutting trajectories into pieces.

Matching Score. For each pair of trajectories $\mathbf{T}_A^i, \mathbf{T}_B^j$ on image-plane A and B, the matching score s is defined as,

$$s(\mathbf{T}_A^i, \mathbf{T}_B^j) = \begin{cases} \alpha s_1(.;\varepsilon) + \beta s_2(.) + \gamma s_2(.) & \eta(i,j) \neq \emptyset \\ \infty & otherwise \end{cases}$$

where, $\eta(i,j)$ is the common time-span of $\mathbf{T}_A^i, \mathbf{T}_B^j$.

$\mathbf{s_1}$: motion clue #1,

$$s_1(\mathbf{T}_A^i, \mathbf{T}_B^j; \varepsilon) = \begin{cases} e(\mathbf{T}_A^i, \mathbf{T}_B^j) & e(\mathbf{T}_A^i, \mathbf{T}_B^j) < \varepsilon \\ \infty & otherwise \end{cases}$$

$$e(\mathbf{T}_A^i, \mathbf{T}_B^j) = \max_{t \in \eta(i,j)} \{|y_A^i(t) - y_B^j(t)|\}$$

where, ε is to tolerate the calibration inaccurracy, y is the coordinate values of features on the y-axis.

$\mathbf{s_2}$: motion clue #2,

$$s_2(\mathbf{T}_A^i, \mathbf{T}_B^j) = \frac{1}{|\eta(i,j)|} \sum_{t \in \eta(i,j)} |\bar{v}_A^i(t) - \bar{v}_B^j(t)|^2$$

$\mathbf{s_3}$: local feature descriptor (if applicable),

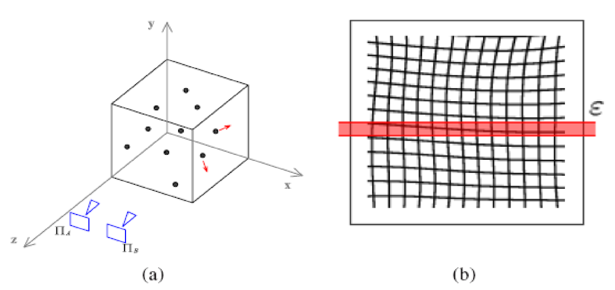
$$s_3(\mathbf{T}_A^i, \mathbf{T}_B^j) = \frac{1}{|\eta(i,j)|} \sum_{t \in \eta(i,j)} C(w_A(q_A^i(t)), w_B(q_B^j(t)))$$

Correspondence Establishment. An instance of the maximum weighted bipartite matching problem.

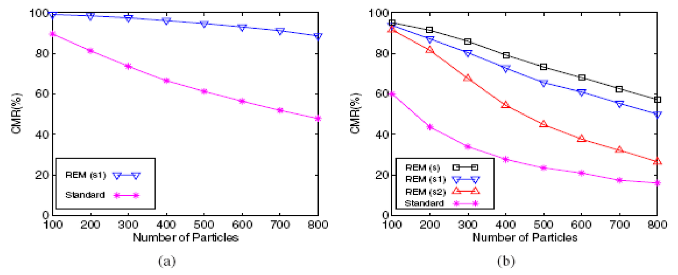
Feature matching can then be established frame by frame based on the trajectory matching result.

The Result

Simulation

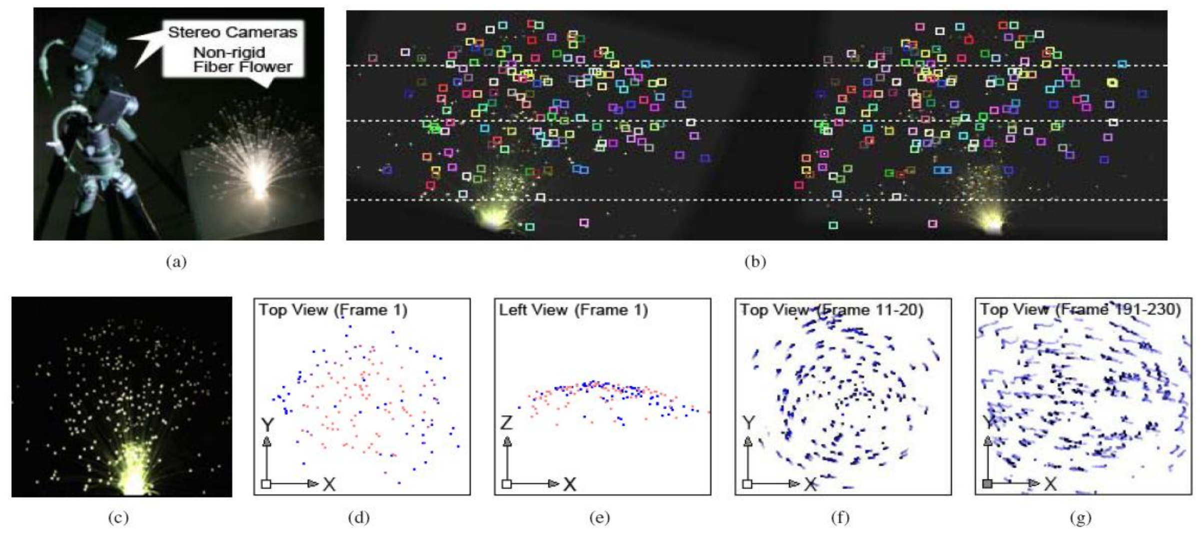


The setting. (a) N particles drifting in a bounding unit cube are captured by a pair of virtual cameras. (b) a tolerance width ε is set to accept the feature detection inaccuracies.



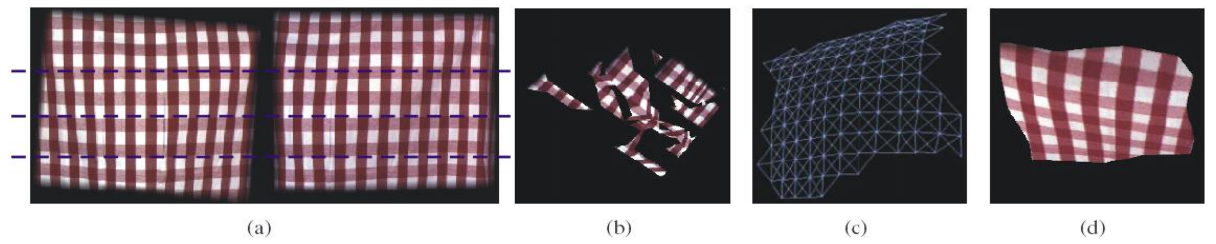
The correct matching rate(CMR). (a) Ideal case with 0 distortion. (b) 5% distortion.

A Non-Rigid Waving Optical Fiber Flower



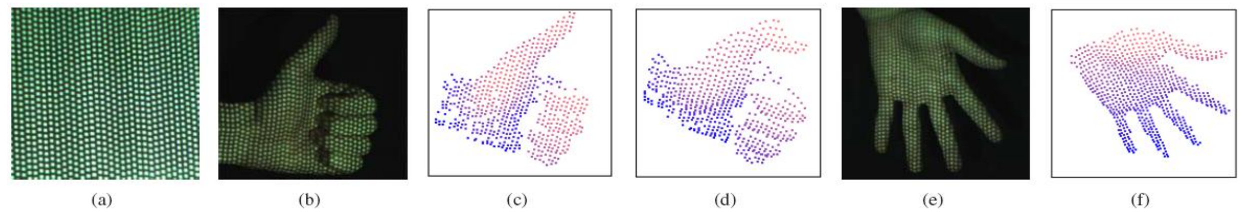
(a) The setting. (c) A captured sample frame, where bright-spots are detectable but with identical features that brings difficulty for conventional stereo matching methods. (b) The matching result using the proposed method. (d-g) The reconstructed depth results.

A Waving Flag



(a) Rectified 1st frame-pair. (b) The reconstruction result using a standard fusion method with dynamic programming. (c,d) The result using the proposed method.

A Gesturing Hand



(a)The active particle pattern in which each column moves vertically and independently. (b-f) the captured sample image and the reconstructed depth map using the proposed method.

Related Methods

The proposed method is free of the limitations that exist in the following related methods.

Single frame, passive matching

Matching errors easily occur when the scene features are undistinguishable, at non-smooth depths and of a large number.

Structured light

Color-coded methods tend to be affected by the surface reflectance properties. Spacetime coded methods require high refresh-rate of the projectors. Also, the added light maybe nuisance or infeasible in some cases.

Structure from motion (SFM)

SFM's can deal with undistinguishable features but are limited to reconstruct rigid scenes or scenes with few number of rigid parts.