Detecting Potential Falling Objects by Inferring Human Action and Natural Disturbance

Bo Zheng*[1], Yibiao Zhao*[2], Joey C. Yu[2], Katsushi Ikeuchi[1] and Song-Chun Zhu[2]

***Abstract*— Detecting potential dangers in the environment is a fundamental ability of living beings. In order to endure such ability to a robot, this paper presents an algorithm for detecting potential falling objects, *i.e.* physically unsafe objects, given an input of 3D point clouds captured by the range sensors. We formulate the falling risk as a probability or a potential that an object may fall given human action or certain natural disturbances, such as earthquake and wind. Our approach differs from traditional object detection paradigm, it first infers hidden and situated "causes (disturbance) of the scene, and then introduces intuitive physical mechanics to predict possible "effects (falls) as consequences of the causes. In particular, we infer a disturbance field by making use of motion capture data as a rich source of common human pose movement. We show that, by applying various disturbance fields, our model achieves a human level recognition rate of potential falling objects on a dataset of challenging and realistic indoor scenes.**

I. Introduction

The recent development of consumer-grade range cameras, such as the Kinect camera, has attracted increasing studies in the field of 3D scene understanding [4] [9] [13] [21]. However, most of existing work is focused on locating and naming the object in the scene, and leaves a big gap to answer human-level scene understanding questions, such as: how does a human interact with a scene? how does the scene response to the action? What and where are potential dangers in the environment?

In this paper, we present an potential falling object detection algorithm, which is an essential component of a safety-aware robot. As shown in Fig.1, the algorithm is useful for three main scenarios:

i) *Safety surveillance robots*. Objects have the potential to fall onto or hit people at the construction site as the warning sign shown in Fig.1 (a). To prevent objects from falling freely from one level to another, the safety risk surveillance ensures that objects are being stored where a secure physical barrier provided.

ii) *Human assistant robots* for children, elders and people with disabilities. As the example shown in Fig.1 (b), we can predict a possible action of the child - he is reaching for something, and then infer possible consequences of his action - he might be struck by the falling teapot.

Fig. 1. The detection of potential falling objects is an essential ability of a safety-aware robot: (a) the safety surveillance robot for a construction site, (b) the human assistant robot for the baby proofing, and (c) A building where was crashed by earthquake and tsunami on March 11, 2011, Japan.

iii) *Disaster rescue robots*. The Fig.1 (c) showed post-disaster scene captured by a 3D range sensor. It was a extremely dangerous environment due to the M9.0 earthquake and tsunami in Japan. A robot working in such environments requires to understand the potential risks due to many objects at unstable state.

Related work. The study of falling objects can be traced back to an early work by Kriegman [10] that first proposed an algorithm to calculate the capture regions where a 3D object may fall according to the Morse theory. There is a recent rise of related studies in following four streams:

i). *Safe Motion Planning*. As the planning is a classic problem in robotics, Petti and Fraichard [22], Phillips and Likhachev [23] tackled the problem of safe motion planning in the presence of moving obstacles. They consider the moving obstacles as the real-time constraint inherent to the dynamic environment. However, we first argue that a robot need to be aware of potential dangers even in a static environment due to possible incoming disturbances.

*Bo Zheng and Yibiao Zhao contributed equally to this work.

[1] Bo Zheng and Katsushi Ikeuchi are with the University of Tokyo, Japan {zheng, ki}@cvl.iis.u-tokyo.ac.jp

[2] Yibiao Zhao, Joey C. Yu and Song-Chun Zhu are with the University of California, Los Angeles (UCLA), USA {ybzhao,chengchengyu}@ucla.edu, sczhu@stat.ucla.edu

The project page: http://www.stat.ucla.edu/~ybzhao/research/fallingobjects

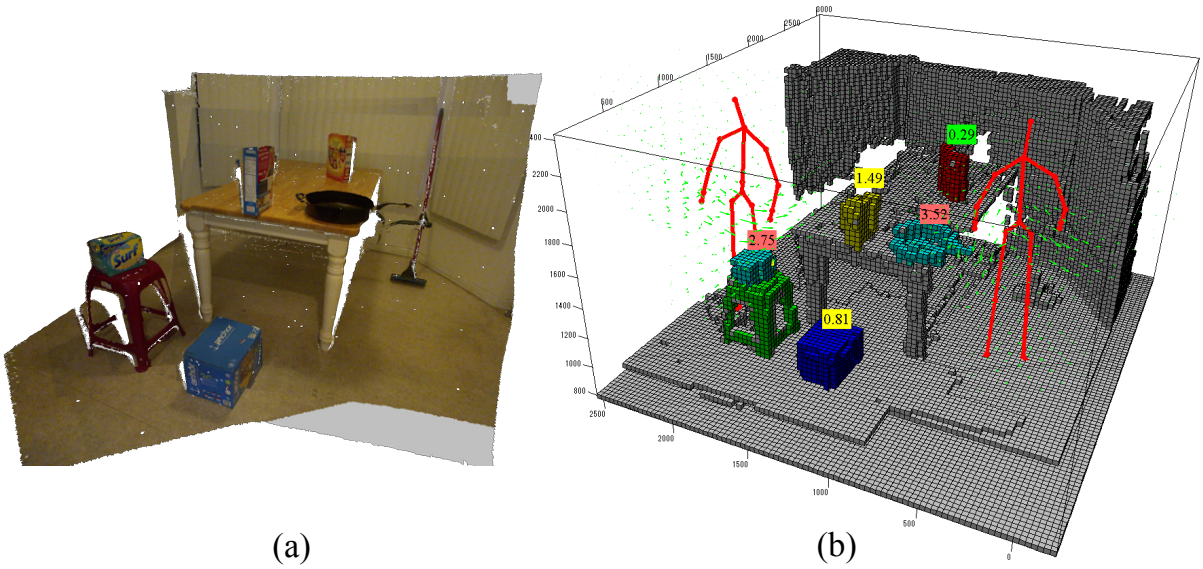

Fig. 2. (a) The input point cloud; (b) "Imagined" human action field and detected potential falling objects with red tags.

ii). *Physics based model*. Gupta *et al.* [7] revisited the block world model and worked on labeling of the 2D image by reasoning the physical force based on a block representation of 2D image segments. Lee *et al.* [11], Zhao and Zhu [15], has made promising progress on volumetric reasoning of 2D indoor scene. Recently, Zheng *et al.* [17] and Jia *et al.* [24] proposed very interesting approaches to segment point clouds and detect 3d objects by incorporating the physics stability as a prior.

iii). *Human in the loop*. This stream of research emphasizes a human-centric representation, differing from the classic feature-classifier paradigm of object recognition. Some recent work utilized the notion of "affordance". Grabner *et al.* [5] recognizes chairs by imagining an "sitting" actor interacting with the scene. Gupta *et al.* [6] predicts the "workspace" of a human given a estimated 3D scene geometry. Fouhey and Delaitre *et al.* [19][20] demonstrate that observing people performing different actions can significantly improve estimates of scene geometry and scene semantics. Jiang [25] [26] proposed scene labeling algorithms by considering humans as the hidden context.

iv). *Cognitive studies*. Psychology studies suggested that approximate Newtonian principles underlie human judgments about dynamics and stability [3] [14]. Hamrick *et al.* [8] showed that knowledge of Newtonian principles and probabilistic representations are generally applied for human physical reasoning, and the intuitive physics model is an important perspective for human-level complex scene understanding.

Overview of our approach. We address the problem of detecting potential falling objects by inferring hidden "causes" (disturbance) and reasoning possible "effect" (falling) using intuitive mechanics. Taking a 3D point cloud as the input as shown in Fig.2 (a), our method first segments the point cloud and recovers volumetric 3D objects in the scene following a recent approach by Zheng *et al.* [17], and predicts the walkable area by hallucinating the human actions [5] [6] [25]. Given the scene geometry and walkable area, we detect the potential falling objects by calculating its expected falling risk given a disturbance field in Fig.2 (b).

i) We infer the disturbance field caused by earthquake or wind, as well as the human activities. A *disturbance field* representing the possible physical work applied to each position in the 3D space. We use the motion capture data of human actions, as the red stick figures in Fig.2 (b), and situate it to the 3D scene (walkable areas) to estimate the statistical distribution of human disturbance. In order to generate a meaningful human action field, we first predict a primary motions on the 2D ground plan which recodes the visiting frequency and walking direction for each walkable position, and add detailed secondary body part motions in 3D space on top. We estimate the distribution of primary motions by synthesizing human walking trajectories following two simple observations: (a) A rational agent mostly walks along a shortest path with minimal effort; (b) A agent has a basic need to travel between any two walkable positions in the scene. As a result, a convex corner, like the table corners in Fig.2 (b), has a high probability to be visited, and the pan on the corner of the table are less safe than others. Similarly, the box on the chair is easy to be knocked off the stool by a swinging hand as well.

ii) We then reason "effects" (falling) of each possible disturbance (an accidental collision) by intuitive mechanism. We first decompose the velocity of input disturbance according to the directions of rotational movement (rolling) and translational movement (sliding) by a parallelogram rule. And we calculate the initial kinetic energy of object after a collision as an input work to the system. According to two principles: conservation of kinetic energy and conservation of momentum, we can infer that the velocity of the object after the collision. We then calculate the minimum kinetic energy to move an entity from one stable point to a local

maximum, *i.e.* knocking it off equilibrium, and then we further calculate the risk of releasing the energy in reaching a deeper minimum.

In experiments, we quantitatively evaluated the accuracy of potential falling object detection, as well as the ranking of falling risk w.r.t. human judgements on a challenging dataset.

II. Definition of the Falling Risk

We measure the risk of a potential falling object as illustrated in Fig.3. The curve represents the change of potential energy in terms of different positions. At the beginning, an object a stays in the position $\mathbf{x}_0$ which is a stable equilibrium. When a work W applies to the object, it start to move upward towards the position of unstable equilibrium $\widetilde{\mathbf{x}}$. The total energy needed to go over the unstable equilibrium $\Delta\mathcal{E}(\mathbf{x}_o \rightarrow \widetilde{\mathbf{x}})$ is called "energy barrier". If the work is larger than the energy barrier $W \geq \Delta\mathcal{E}(\mathbf{x}_o \rightarrow \widetilde{\mathbf{x}})$, then the object will fall over the unstable equilibrium. In this way, we define the falling risk as:

Definition 1. *The falling risk $R(a, \mathbf{x}_0, W)$ of an entity a at $\mathbf{x}_0$ in the presence of a disturbance work W is the maximum energy that it can release when it moves out the energy barrier by the work W.*

$$\mathrm{R}(a, \mathbf{x}_0, W) = \delta[W \geq \Delta\mathcal{E}(\mathbf{x}_o \rightarrow \widetilde{\mathbf{x}})]\Delta\mathcal{E}(\widetilde{\mathbf{x}} \rightarrow \mathbf{x}'_o), \quad (1)$$

$\delta()$ is an indicator function and $\delta(z) = 1$ if condition z is satisfied otherwise $\delta(z) = 0$.

Definition 2. *The falling risk $R(a, \mathbf{x}_0)$ of an entity a at position $\mathbf{x}_0$ in the presence of a disturbance field $p(W, \mathbf{x})$ is the expected risk with respect to the disturbance distribution.*

$$R(a, \mathbf{x}_0) = \int p(W, \mathbf{x}_0) R(a, \mathbf{x}_0, W) dW, \quad (2)$$

The *energy barrier* $\Delta\mathcal{E}(\mathbf{x}_o \rightarrow \widetilde{\mathbf{x}})$ is the minimum energy needed to move from the current state (say a local minimum) $\mathbf{x}_0$ to an unstable equilibrium $\widetilde{\mathbf{x}}$. For example, as shown in Fig. 4, when a cone is currently in stable state B, its energy barrier is the minimum work needed to push it out of the current energy basin. Passing that point $B\prime$, the cone will fall to a new stable state at lower position. Also, for example, when a cup is at the center of the table, its energy barrier is the minimum work needed (to overcome friction) to push it to the edge.

The potential *falling risk* $\Delta\mathcal{E}(\widetilde{\mathbf{x}} \rightarrow \mathbf{x}'_o)$ is the energy released when an entity moves from its unstable equilibrium $\widetilde{\mathbf{x}}$ to a lower minimum $\mathbf{x}'_0$. For example, when the cup falls of from the edge of the table to the ground. The higher the table, the larger the energy risk $\Delta\mathcal{E}(\widetilde{\mathbf{x}} \rightarrow \mathbf{x}'_o)$.

With the definition of the potential falling object, we introduce the inference of the disturbance field in Sect.III and the calculation of potential energy and initial kinetic energy given a disturbance in Sect.IV.

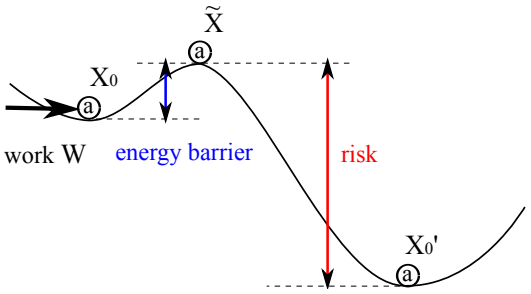

Fig. 3. An illustration of falling risk definition and other basic concepts on the potential energy curve;

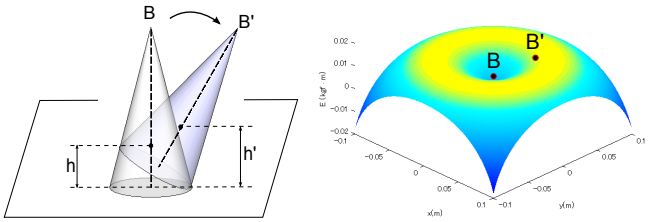

Fig. 4. An simple example that a cone is being knocked down. It is pushed up from the stable equilibrium B, and about to go over the energy barrier B'. The correspondent potential energy map is on the right.

III. Inferring the Disturbance Field

Taking a 3D point cloud as the input as shown in Fig.2 (a), our method first segments the point cloud and recovers volumetric 3d objects the scene following a recent approach by Zheng *et al.* [17], and predict the walkable area and sittable area by hallucinating the human actions [5][6]. The result is shown in Fig.2 (b). In order to approximate arbitrary shape of 3D objects, we discretize the 3D space to voxels, which are the smallest units in the space. So that all the 3D entities are represented by a group of voxels. In such recovered 3D environment, we then estimate disturbance field caused by natural forces and human actions.

A. Natural disturbance field

Despite the gravity applies a constant downward force to all the voxels, other natural disturbances such as earthquakes and winds are also present in a natural scene.

1) **Earthquake** transmits energy by forces of interactions between contacting faces, typically by the frictions in our scenes. Here, we estimate the disturbance field by generating random horizontal forces to the voxels along the contacting surfaces. We use a certain constant to simulate the strength of the earthquake and the work W it generates.

2) **Wind** applies fluid forces to exposed voxels in the space. A precise simulation need to simulate the fluid flow in the space. Here, we simplify it as an uniformly distributed field over the space.

B. Human action disturbance field

In order to generate a meaningful disturbance field of human actions, we decompose the human actions into the primary motions *i.e.* the center of mass movements in Fig.5 and the secondary motions *i.e.* the body parts movements in

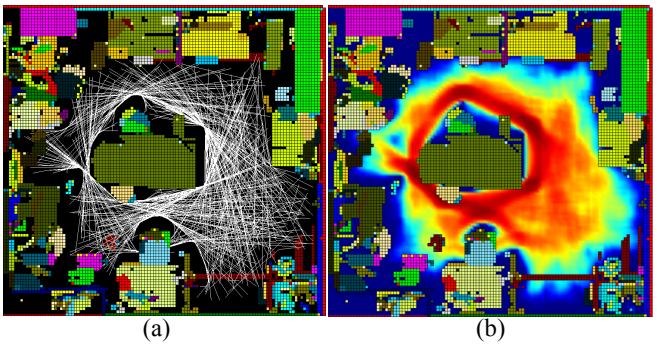

Fig. 5. Primary motion field: (a) The hallucinated human trajectories (white lines); (b) The distribution of the primary motion space. The red represents high probability to be visited.

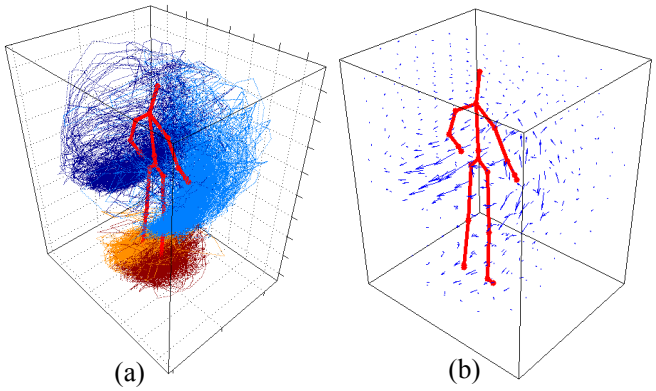

Fig. 6. Secondary motion field: (a) Secondary motion trajectories of hands and feet from motion capture data; (b) Distribution of the secondary motion field. Long vectors represent large velocity of body movement.

Fig.6. We first predict a human primary motion field on the 2D ground plan, and add detailed secondary motions in 3D space on top. The disturbance field is characterized by the moving frequency and moving velocity for each quantized voxel.

Primary motion field captures the movement of human body as a particle. We estimate the distribution of primary human motion space by synthesizing human motion trajectories following two simple observations:

1) A rational agent mostly walks along a shortest path with minimal effort;

2) A agent has a basic need to travel between any two walkable positions in the scene.

Therefore, we randomly pick 500 pairs of positions in the walkable space, we calculate the shortest path that connecting these two positions as shown in Fig.5 (a). And we calculate the walking frequency as well as walking directions based on the synthesized trajectories. Fig.5 (b) demonstrates a distribution of walkable space, the red color means the position has high probability to be visited, and the length of the small arrows shows the probability of moving directions.

In the Fig.5 (b), we can see some more details that convex corners, e.g. table corners, are more likely to be visited, and objects in these busy area may have higher risk than the ones in a concave corners. A hallway connecting two walkable area is also frequently visited, and objects in the hallway are less safe too. It is worth noting that the distribution of moving direction is also very distinctive, it help us to locate human body move in the right direction to generate the human disturbance field.

Secondary motion field is the movement thats not part of the main action *e.g.* arms swinging while walking. But secondary motion is important to capture the random disturbance, for example, people may push objects off the edge of the table by hand or kick objects on the ground by foot. We also use the Kinect camera to collect human motion capture data Fig.6 (a), and then calculate the distribution of moving velocity as shown in Fig.6 (b).

The primary motion field further convolves with secondary motion field, thus generate a dense disturbance field that capturing the distribution of motion velocity for each voxel in the space. The disturbance field is then represented by a probability distribution over the entire space for the velocities along different directions and frequencies that they occur. For example, a cup in the middle of a large table will not be reachable by a walking person and thus the distribution of velocity above the table center, or any unreachable points, is zero. Five typical cases in the integrated field is demonstrate in Fig.7.

IV. Calculating the Physical Energy

Given the disturbance field, in this section, we present a feasible way for calculating input work (energy) that might lead to object falling. However, building sophisticated physical engineering models is not feasible, as it becomes intractable if we consider complex object shapes and material properties, *e.g.* , to detect a cup falling off from a table, huge amount of action need to be simulated until meeting the case that human body acting on the cup. The relation between intuitive physical model and human psychology was discussed by recent cognitive study [8].

In this paper, to obtain a simple intuitive physical model we make following assumptions.

1. All the objects in the scene are rigid.

2. All the objects are made from same material, such as wood (friction coefficient: 0.6, uniform density: $700kg/m^3$).

3. A scene is a dissipative mechanical system that total mechanical energy along any trajectory is always decreasing caused by friction, while kinetic and potential energy may be traded off at different states due to elastic collision.

A. Initial kinetic energy after an elastic collision

We now calculate the initial kinetic energy, which is considered as the input work in Fig. 3 after an elastic collision. Here, we simplify objects as mass points to illustrate the simple idea, we will extend the model to more general rigid bodies with arbitrary shapes and arbitrary collision points in the next sub-section.

A head-on elastic collision between two bodies can be represented by velocities in one dimension along a line

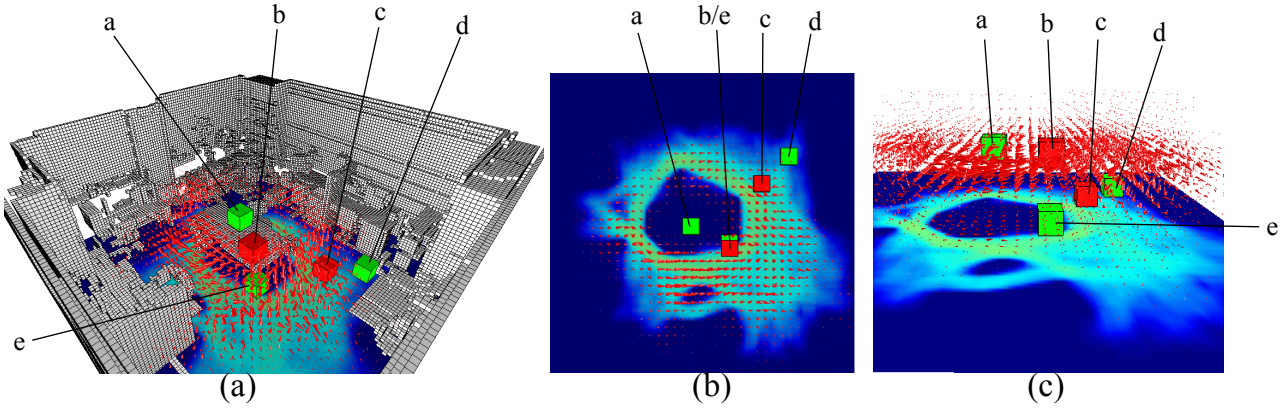

Fig. 7. The integrated human action field by convolving primary motions with secondary motions. The objects **a-e** are five typical cases in the disturbance field: the object **b** on edge of table and the object **c** along the passway exhibit more disturbances (accidental collisions) than other objects such as **a** in the center of the table, **e** below the table and **d** on a concave corner of space.

passing through the bodies. If the velocities are u_1 and u_2 before the collision and v_1 and v_2 after, the equations expressing conservation of momentum and kinetic energy are:

$$m_1u_1 + m_2u_2 = m_1v_1 + m_2v_2 \tag{3}$$

$$\tfrac{1}{2}m_1u_1^2 + \tfrac{1}{2}m_2u_2^2 = \tfrac{1}{2}m_1v_1^2 + \tfrac{1}{2}m_2v_2^2\,. \tag{4}$$

Considering the case that one hand with m_1 knocked off a cup with m_2, we set the initial velocities of hand as u_1 and the cup is still $u_2 = 0$. The final velocity of the cup is given by

$$v_2 = \left(\frac{2m_1}{m_1 + m_2}\right) u_1\,. \tag{5}$$

If the cup has the same mass as the hand, then the hand that was moving is now stopped and the cup is moving away at speed u_1. However, if the hand collide with a table with much greater mass, then the table will be little affected by a collision while the hand will be rebounded back.

Given the initial velocity of the object, we can easily calculate the initial kinetic energy, which is also the input work in Fig.3:

$$W = E_k = \frac{1}{2}m_2v_2^2 = \frac{2m_1^2m_2}{(m_1 + m_2)^2}u_1^2 \tag{6}$$

B. Decomposition of the force, the velocity and the momentum

Here, we treat the object as a rigid body with arbitrary shape. As shown in Fig.8, the input force V can be decomposed to a force V_t along a line passing through the center of mass and another force V_r perpendicular to V_t. The former force V_t generates an translational movement, while the latter force V_r generates an rotational movement. V_t can further be decomposed as three velocities V_t^x, V_t^y, V_t^z along three axes, and V_r is decomposed as three rotational velocity V_r^x, V_r^y, V_r^z around three axes. The input force or momentum can be decomposed in the same way.

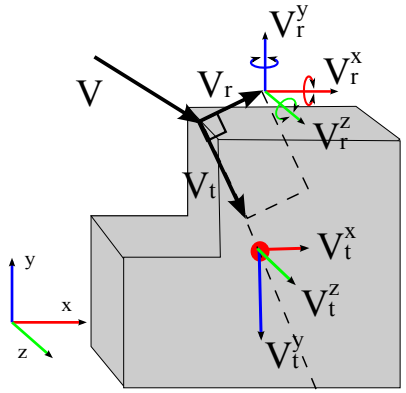

Fig. 8. The decomposition of action velocity. The gray polygon represents an object with its center of mass on the red dot. The action velocity V first decompose as a rotational velocity V_r and translational velocity V_t, and each velocity is further decomposed as three components along three dimensions.

Consider the object supported by a flat surface from the bottom, we can ignore V_t^y because it will be rebounded back along the y axis as we discussed before. We can also ignore the V_r^y because the rotation around the y axis will not change potential energy, and it also suffer a large friction at the time.

C. Potential energy

As we discussed in the Sect.II, we calculate an energy map of potential energy. By comparing the input work with the energy landscapes on potential energy map, we calculate the falling risk according to Eq.1 and Eq.2. In a same spirit of decomposition above, fortunately we can decompose the change of potential energy according to rotation (rolling by itself) movement and translation (position change) movement. By ignoring the translation and rotation along y axis, we calculate the rotational energy map according to two vectors V_r^x, V_r^z, which can be also projected onto spherical coordinate system see [10]; and calculate the translational energy map according to the V_t^x, V_t^z.

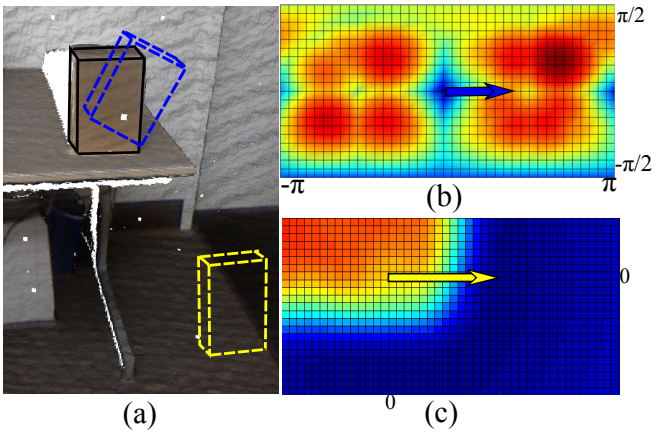

Fig. 9. Potential energy map for (b) the rotational movement and (c) the displacement movement of the box on a table in (a).

Fig.9 shows a simple example, giving energy a book is falling off table. We roughly decompose this process into two sub-steps: 1) it rolls from stable state (in black) to unstable state (in blue); and 2) it falls off to the position (in yellow) as a mass point. Therefore we can draw the state change (along the blue and yellow arrows) on the energy maps shown in Fig. 9 (b) and (c) respectively. In each energy map, red means high potential energy, whereas blue means low potential energy. We can see that the object is initially lying at the energy minimum (stable equilibrium) on both maps, and it need some work to push out of the unstable equilibrium. Once it is pushed into the unstable states, the case in Fig.9 (c) releases much more energy than that in Fig.9 (b).

V. Experiments

In our experiments, we evaluate our approach by two datasets of large-scale point clouds. The first dataset captured by Microsoft Kinect sensor contains 100 scenes, and each scene is composed by 20-30 rgb-depth images with a powerful SLAM algorithm [12]. Another dataset is captured by a high-end 3D sensor Leica ScanStation C10. It contains 20 large scenes, and each snapshot of the sensor scans 260 rgb-depth images covers a panorama of the scene.

Qualitative evaluation. As shown in Fig. 10, we compare the potential falling objects under three different disturbance fields: 1) The human action field in Fig. 10 (b,e); 2) The wind field (an uniform directional field) in Fig. 10 (c,f) and 3) earthquake (random forces on contacting object surface) in Fig. 10 (d,g). As we can see the cups with red tags are detected potential falling objects, which are very close to human judgements.

In Fig.11, we show four large-scale point clouds in each row, where (a) shows input 3D point clouds with rgb color for reference; (b) illustrates inferred human action fields, the larger and more complex environment like the last scene on the bottom exhibits more sophisticated motion patterns, which beautifully matches with human motion patterns; (c) shows a overview of potential falling objects with their risk scores on yellow tags; and (d) shows the zoom-in details of some typical successful and failure detection examples. Some false positives may caused by highly occlusions.

Quantitative evaluation. We conduct two quantitative evaluations:

Accuracy of potential falling object detection. In this experiment, we first manually labeled 83 potential falling objects from 20 large scale point clouds, some of them are shown in 2 5 7 9 10 Fig.11. The groundtruth come from majority vote ($> 50\%$) of 10 participants. We calculated the ROC curve of potential falling object detection by our proposed approaches in Fig.12 (a). It is shown that our algorithm can reliably detect potential falling objects with 80% true positive rate and keep a 20% false positive rate at the same time.

Ranking of falling risk. The human judgements of potential falling objects can be very subjective, and they may not be reliable ground truths. Instead of calculating the error rate, we compare the ranking of several potential falling objects in a scene with the ranking of human judgement in this experiment. We asked 10 participants to choose a reasonable order of the object according to their falling risk. The results are shown in Fig.12 (b) where the model output fit well with the human judgement, but still keep a certain variance. Then we conducted a similar experiment. We random split the participants into two groups, and evaluate the correlation between these two groups. As shown in Fig.12 (c), the correlation between human judgements keep the same amount of variance as the correlation between model and human. It is also interesting to note that the variance is larger when the risk score is low (lower left corner of Fig.12 (b,c)), or say the falling risk judgement will become less ambiguous when the risk is higher.

The similar judgment correlation between machine and human in Fig.12(b,c) implies the algorithm may pass the Turing test because the judge cannot reliably tell the machine from the actual human according to the answers.

VI. Conclusion and Discussion

This paper presents a novel approach for detecting potential falling objects. We demonstrated that, by applying various disturbance fields, our model achieves a human level recognition rate of potential falling objects on a dataset of challenging and realistic indoor scenes. Differing from traditional object classification paradigm, our approach goes beyond the estimation of 3D scene geometry. The approach is implemented by making use of the ”causal physics”. It first infers hidden and situated ”causes” (disturbance) of the scene, and introduces intuitive mechanics to predict possible ”effects” (falls) as consequences of the causes. Our approach revisits classic physics-based representation, and feeds by the state-of-the-art algorithms. Further studies along this way, including friction, material properties, causal reasoning, can be very interesting dimensions of vision research.

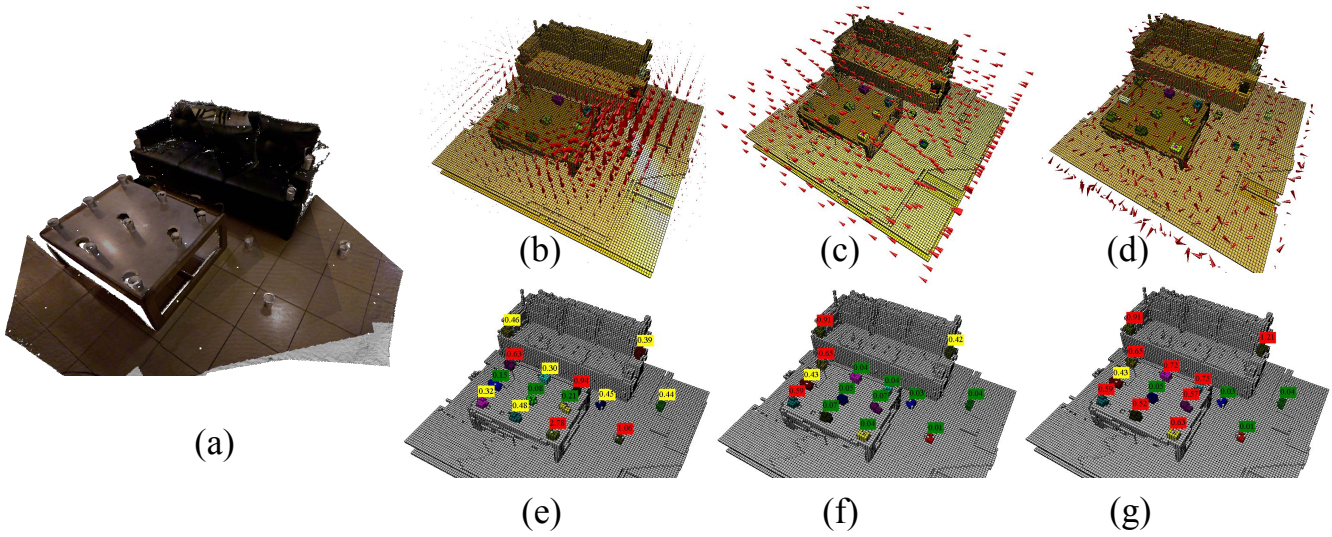

Fig. 10. The potential falling objects (with red tags) under the human action field (b,e), the wind field (c,f) and the earthquake field (d,g) respectively. The results match with human perception: (i) objects around table corner are not safe w.r.t human walking action; (ii) object along the edge of wind direction are not safe w.r.t wind disturbance; and (iii) object along all the edges are not safe w.r.t earthquake disturbance.

Please visit our project page for the supplementary demo and high-resolution results:
http://www.stat.ucla.edu/~ybzhao/research/fallingobjects.

ACKNOWLEDGMENT

This work is supported by MURI ONR N00014-10-1-0933 and DARPA MSEE FA 8650-11-1-7149, USA; Next-generation Energies for Tohoku Recovery and the 10th Core Project of Microsoft, Japan..

REFERENCES

[1] I. Biederman, R. J. Mezzanotte and J. C. Rabinowitz, Scene perception: Detecting and judging objects undergoing relational violations. *Cognitive Psychology* **14** 143-177, 1982.
[2] M. Brand, Physics-based visual understanding. *Computer Vision and Image Understanding* **65** 192-205, 1996.
[3] R.W. Fleming, M. Barnett-Cowan and H.H. Bulthoff, Perceived object stability is affected by the internal representation of gravity. *Perception* 2010.
[4] A. Anand, H. Koppula, T. Joachims and A. Saxena, Contextually Guided Semantic Labeling and Search for 3D Point Clouds, *International Journal of Robotics Research (IJRR)*, 2012.
[5] H. Grabner, J. Gall and L. Van Gool What Makes a Chair a Chair?, *In CVPR* 2011.
[6] A. Gupta, S. Satkin, A. Efros and M. Hebert, From Scene Geometry to Human Workspace. *In CVPR* 2011.
[7] A. Gupta, A.Efros and M. Hebert, Blocks World Revisited: Image Understanding Using Qualitative Geometry and Mechanics, *In ECCV* 2010.
[8] J. Hamrick, P. Battaglia and J. Tenenbaum, Internal physics models guide probabilistic judgments about object dynamics. *In: Proc. 33rd Ann. Conf. Cognitive Science Society* 2011.
[9] A. Janoch, S. Karayev, Y. Jia, J.T. Barron, M. Fritz, K. Saenko and T. Darrell, A category-level 3-d object dataset: Putting the kinect to work. *In: ICCV Workshop on Consumer Depth Cameras for Computer Vision* 2011.
[10] D.J. Kriegman, Let Them Fall Where They May: Capture Regions of Curved Objects and Polyhedra. *IJCV* 16, 448-472. 1995
[11] D. Lee, A. Gupta, M. Hebert and T. Kanade, Estimating Spatial Layout of Rooms using Volumetric Reasoning about Objects and Surfaces *In NIPS*, pp. 609-616. 2010.
[12] R. Newcombe, S. Izadi, O. Hilliges, D. Molyneaux, D. Kim, A. Davison, P. Kohli, J. Shotton, S. Hodges and A. Fitzgibbon, KinectFusion: Real-Time Dense Surface Mapping and Tracking, *IEEE ISMAR* 2011.
[13] R. Sagawa, K. Nishino and K. Ikeuchi, Adaptively Merging Large-Scale Range Data with Reflectance Properties, *PAMI* 27, 392-405 2005.
[14] M. Zago and F. Lacquaniti, Visual perception and interception of falling objects: a review of evidence for an internal model of gravity. *Journal of Neural Engineering* 2005.
[15] Y. Zhao and S. C. Zhu, Image parsing via stochastic scene grammar. *In NIPS* 2011.
[16] B. Zheng, J. Takamatsu and K. Ikeuchi, An Adaptive and Stable Method for Fitting Implicit Polynomial Curves and Surfaces. *PAMI* **32** 561–568 2010.
[17] B. Zheng , Y. Zhao, J. C. Yu, K. Ikeuchi and S. C. Zhu, Beyond Point Cloud: Scene Understanding by Reasoning Geometry and Physics. *In CVPR* 2013.
[18] Q. Zhou, Random walk over basins of attraction to construct Ising energy landscapes. *Physical Review Letters*, 106: 2011.
[19] V. Delaitre, D. Fouhey, I. Laptev, J. Sivic, A. Gupta and A. Efros. Scene semantics from long-term observation of people, *In ECCV* 2012.
[20] D. Fouhey, V. Delaitre, A. Gupta, A. Efros, I. Laptev and J. Sivic. People Watching: Human Actions as a Cue for Single-View Geometry, *In ECCV* 2012.
[21] J. Xiao and Y. Furukawa, Reconstructing the World's Museums, *In ECCV* 2012.
[22] S. Petti and T. Fraichard, Safe Motion Planning in Dynamic Environments, *In IROS* 2005.
[23] M. Phillips and M. Likhachev, SIPP: Safe Interval Path Planning for Dynamic Environments, *ICRA* 2011.
[24] Z. Jia, A. Gallagher, A. Saxena and T. Chen, 3D-Based Reasoning with Blocks, Support, and Stability, *In CVPR* 2013.
[25] Y. Jiang, H. S. Koppula and A. Saxena, Hallucinated Humans as the Hidden Context for Labeling 3D Scenes, *In CVPR* 2013.
[26] Y Jiang and A. Saxena, Infinite Latent Conditional Random Fields for Modeling Environments through Humans, *In Robotics: Science and Systems (RSS)*, 2013.

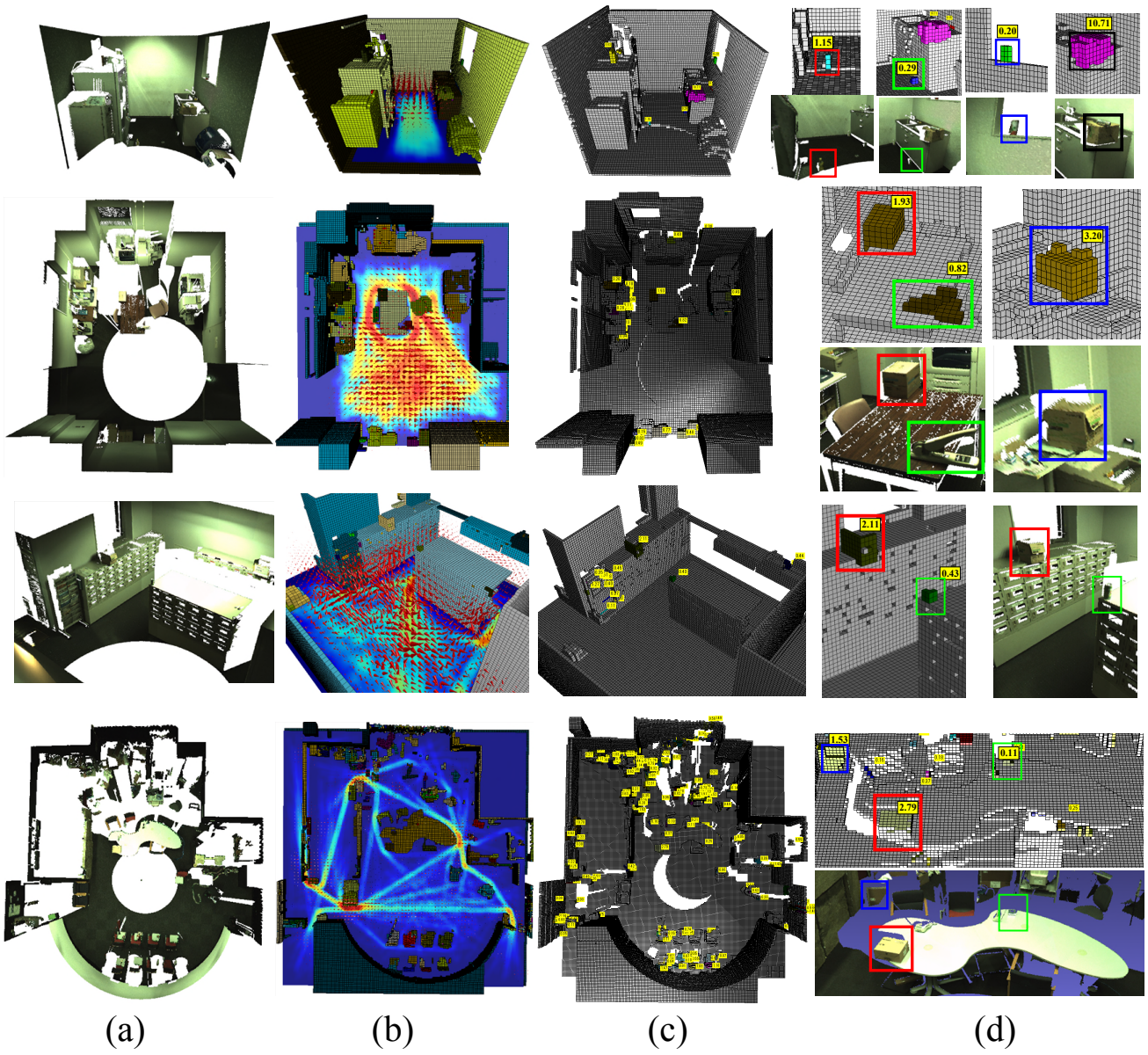

Fig. 11. (a) Input 3D scene point clouds; (b) Inferred human action fields and segmented objects shown in different colors; (c) Detected potential falling objects with their risk scores on the yellow labels; (d) Some zoom-in details of detected potential falling objects. See text for more explanation.

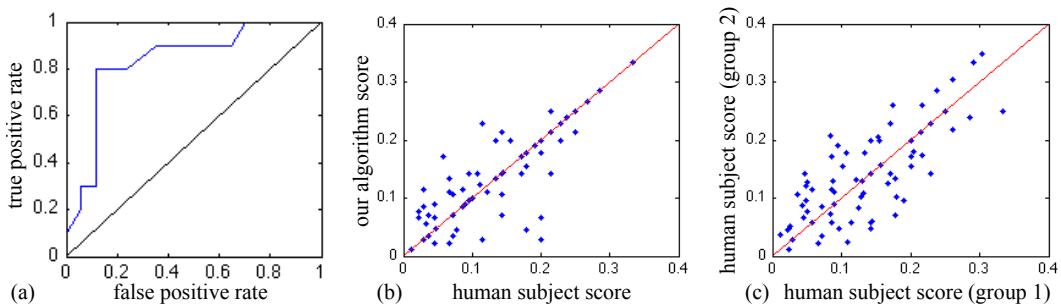

Fig. 12. (a) The ROC curve of the potential falling object detection (b) The correlation between the falling risk ranking by our algorithm and the ranking by human subjects. (c) The correlation between the falling risk ranking by two different random split groups of human subjects.