BODY EXPRESSION RECOGNITION FROM ANIMATED 3D SKELETON

Arthur Crenn - Rizwan Ahmed Khan - Alexandre Meyer - Saida Bouakaz

Université de Lyon, CNRS, France
Université Lyon 1, LIRIS, UMR5205, F-69622, France

ABSTRACT

We present a novel and generic framework for the recognition of body expressions using human postures. Motivated by the state of the art from the domain of psychology, our approach recognizes expression by analyzing sequence of pose. Features proposed in this article are computationally simple and intuitive to understand. They are based on visual cues and provide in-depth understanding of body postures required to recognize body expressions. We have evaluated our approach on different databases with heterogeneous movements and body expressions. Our recognition results exceeds state of the art for some database and for others we obtain results at par with state of the art.

***Index Terms*—** Body Expression Recognition, 3D Skeleton, Features Extraction, Animation.

.

1. INTRODUCTION

Many applications would benefit from the ability to understand human emotional state in order to provide more natural interaction e.g video games, video surveillance, human-computer interaction, artistic creation, etc. Emotion is a complex phenomena. Expression of an emotion could be personal and subjective as two persons could perceive and interpret differently the same expression. Its perception changes from one culture to another. Furthermore, human expresses emotional information through several channels, like facial expression, body movement and sound. Several studies from various disciplines have shown that body expressions are as powerful as those of the face to express emotion [16]. Unfortunately, to the best of our knowledge, few publications have focused on this issues.

This paper presents a method to detect and classify expression through a sequence of 3D skeleton-based poses. The challenge is to find common informations between all the movements of the same expression. And this, while the expression is embedded in the action performed by the person, e.g. nervous person walking pattern contains the action of walking and the emotional state to be nervous. With the growth and ease of accessibility of devices that track 3-dimensional body like the Kinect [21, 3] or accelerometer-based motion capture system [23], it is easy to get data. Thus many applications will be benefited from real-time analysis of body movements. There are lots of method proposed in literature for facial expressions [5, 15, 12, 14], however, as mentioned above, there is a lack of research for the analysis of body movement to recognize expressions. Early proposed methods for body movements analysis [13, 18] were limited to specific movements or expressions. Taking an approach proposed by choreographer seeking to describe dance movements, Truong *et al*.[24] proposed descriptors calculated from the movements in order to recognize various gestures and expressions. Inspired by Truong's work and results of psychological research [16], we have looked to quantify the space posture of the body, during an action, by introducing new set of features to characterize an expression. Besides, its computational simplicity allowing the recognition of expression on heterogeneous body movements in real-time. Our method has the advantage to be reusable in many domains as movement synthesis processes for computer graphics and animation. The paper is organized as follows. Section 2 presents the state of the art of the emotion analysis. Section 3 details the set of features we proposed. Section 4 shows the results obtained on different databases and the comparison with other state of the art's method. Finally, Section 5 concludes the paper.

2. RELATED WORK

In recent years, researchers have considerably focused on the automatic facial expression recognition (FER) [5, 15, 12]. However, computer vision domain is lacking research on the detection of emotion based on human's posture. In analogy afterward we use the term "Body Expressions" to refer to emotion expressed by body posture. While, psychological studies show that the human's posture is as powerful as facial expressions in conveying emotions [20]. That's why, psychologists have sought to understand what bodily information is necessary for recognizing the affective state of one person. Psychological studies [2, 10] have investigated the part of the body form versus movement in affect perception. These studies conclude that both form and motion information are important for perceiving emotions from body expressions. Finally, after the form and movement features, psychologists have in-

978-1-5090-5743-6/16/$31.00 ©2016 IEEE

troduced two main levels of bodily details which are high and low-level descriptions [16]. The high level description is often based on the Laban approach which describes body expressions in a global way. The low description is another approach to describe body expression by providing more precise features like distance between joints and angle between body segments. Some techniques related to psychology studies focus on very specific actions. Bernhardt and Robinson [4] proposed a framework to detect implicitly communicated affect of knocking actions. Their method is based on segmentation to divide complex motions into a set of automatically derived motion primitives. Then, they analyzed the parsed motion in terms of dynamic features, i.e. velocity, acceleration and jerk. They obtained a range of 50% to 81% correct classification rate and we are also going to compare our method with their results (refer Section 4 for comparison). Karg *et al.*[13] proposed method for gait patterns recognition. They investigated the capability of gait to reveal a person's affective state. According to their study, speed, cadence and stride length are important factors to correctly discriminate different expressions of a human gait. Paterson *et al.*[22] confirms the role of velocity for the discrimination of affect. They performed visual experiments and concluded that speed plays vital role in the perception of affect.

Below mentioned are few state of the art articles that proposed heterogeneous body movement analysis. Castellano *et al.*[7] proposed a method for automated video analysis of body movement and gesture expressiveness. They used movement expressiveness to infer emotions. They also present two methods for the classification. The first one is based on a direct classification of time series whereas the second one uses a meta-features approach. A meta-feature is the statistical calculation on a set of features in order to abstract the time. They obtained correct classification accuracy of 61% on their database which contains four expressions (Anger, joy, pleasure and sadness). Kleinsmith *et al.*[18] proposed a system that incrementally learns to recognize the body expression. Their system is based on form features (i.e. distance from one joint to another one) and motion features. They obtained a correct classification rate of 79%, we will compare our method with their results (refer Section 4 for comparison). Truong *et al.*[24] proposed a new set of 3D gesture descriptors based on the laban movement analysis model for gestures expressiveness. They obtained high recognition rates for action recognition (F-Score: 97%) on Microsoft Research Cambridge-12 dataset [9]. They also tested their classification approach on their own proprietary database which contains 882 gestures and achieved best F-Score of 56.9%.

We can take inspiration from the domain of animation in computer graphics as they also analyze body movement and gesture expressiveness. The researchers working in the domain of animation generation proposed different methods for modification of animation using skeleton transformations. This modification of expression can be achieved by changing the timing, speed and the spatial amplitude of the motions of body part joints. [1, 11]. Some researchers [6, 25] have also used methods from signal processing (Fourier transformation, motion wave-shaping, time-warping, etc.) in order to modify the expression of animations. Recently, with the same goal of synthesis expressive animations, Forger and Takala [8] proposed an approach based on the motion signals by using frequency components.

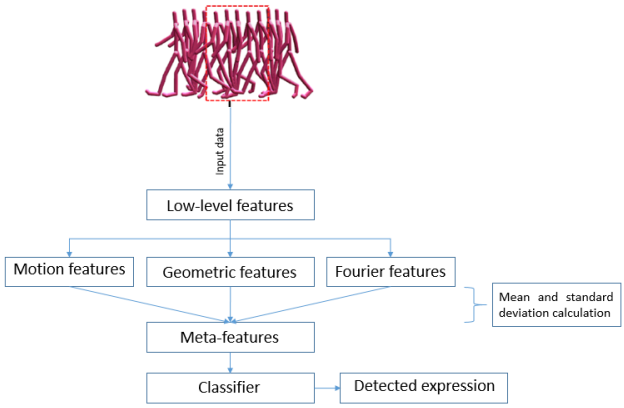


Fig. 1: Overview of our framework.

3. PROPOSED FRAMEWORK

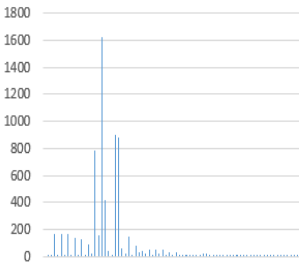
Our motivation was to propose an efficient recognition method of body expressions on heterogeneous movements that rivals state of the art and in real time on a standard computer. Following the foot prints of several previous works [18, 24, 8], we propose a set of novel descriptors based on geometry, motion and frequency-based of body part joints. The overview of our approach is given in Figure 1 and in the following three steps. The rest of this Section is dedicated to the presentation of the features proposed by our method.

1. Our framework computes low-level features for each frame of the motion capture data. We decomposed our features in three types: geometric features, motion features and Fourier features. The details of these features are described below.

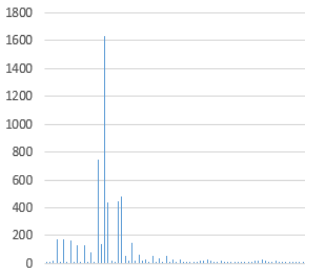
2. Starting from all these low-level features obtained from a time sequence, we compute the meta-features i.e. mean and standard deviations for each feature. Since meta-features are independent of the time, we avoid the computationally complex time-warping step for synchronizing two animations.

3. The resulting values of these meta-features are fed to the classifier which provides the expression classification of the input motion capture data.

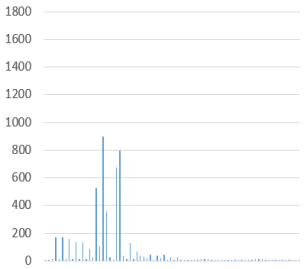
The low-level features we propose are based on distances between body joints, area of triangle defined by specific joints, angles, velocity, acceleration and frequency of movement of different joints. Table 1 describes the set of all our basic features. We have a total of 68 features. Then, for each feature we compute the meta-features. Thus, our approach leads to a total number of 136 features in order to recognize the body expression. Figure 2 shows the set of meta-features for three expressions. For visual purpose, we are only showing first 76 features from the total of 136 dimensional feature vector. It can be observed in the referred figure that our features have a discriminative trend for these expressions. This discriminative ability is used in the proposed framework to robustly classify body expressions.



(a) Histogram of our features extracted from happy knocking



(b) Histogram of our features extracted from sad knocking



(c) Histogram of our features extracted from angry knocking

Fig. 2: Features used for different expressions. These histograms show the discriminative property of our features.

According to the studies of [13] and [4], we have decided to use the following joints for feature extraction: head, pelvis, elbows, shoulders and hands. Experimental results obtained by [13] showed that head, pelvis, elbow and shoulder joints embed most of the emotion of a movement. Results in [4] proved that hands are also important in conveying expressions for several action. Thus, analysis of these five joints is most important to recognize body expressions. All the features extracted for body expression analysis are presented in Table 1 and are described below. We used many features to correctly fit to aspects of body expression described by psychological studies, i.e., form and movement, high and low-level descriptions of bodily detail.

The first feature *V*, is the global space occupy by the skeleton. We use the size of the bounding box of the skeleton in the three directions. The second feature θ is the angle defined between the vertical axis y and the axis binding the center of the hip to the neck. According to studies in psychology, a person tends to extend his body for positive expression whereas for negative expressions a person takes a more compact and forward tilt posture.

The following features are dedicated to correctly capture the configuration of each body part. Distance $\mathcal{D}$ between two joints is important feature to analyze body expression. For instance, we use the distance between hand to the shoulder on the same side. This distance gives indirect information about the elbow. We use distances between hands to hips, hands to shoulders and elbows to hips. Furthermore, a new idea introduced by our approach is the use of triangles *A* to correctly discriminate multiple expressions. The body being symmetric, we take joints on each side of the body, right and left side. The last point of the triangle is chosen on the axis of the body, i.e. either neck or hips. Since the majority of movements is anti-symmetric, the triangle gives lot of information about the expression. For each triangle, we compute its area and the three angles formed by these different triangles. With these four values, we extract information related to shape of the body. For instance one of these triangle-based feature is the area and the angles of the triangle defined by the neck and shoulders. Figure 3 shows the variation of triangles for two different body expressions taken at the same timing of walking. We compute the area and the angle of triangles defined by the hands and the neck; the shoulders and the neck; the elbows and the neck; the hands and the hips.

As mentioned in the state of the art, psychologists have showed that motion is an important characteristic to discriminate expressions for different gestures. Thus, we add to the feature vector the motion features: the velocity $\vec{v}$ and the acceleration $\vec{a}$ of the hands, shoulders, hips, head and elbows joints. The velocity is the first derivative of the position of the current joint. Acceleration being the second derivative of this position.

Finally, we use the fast Fourier transformation $\mathcal{F}$ to obtained the frequency component of our selected joints. We compute the discrete Fourier transform with the fast Fourier transform algorithm on the signal of the joint angle. Section 4 compares the results by using only geometric features, only motions feature or only the Fourier features on the different databases.

4. RESULTS AND ANALYSIS

We have tested our method on three databases (refer table 2 for summary of databases) which are presented below. Two of them are acted databases, while the last database consists

Id.	Type of feature	Description
V	Space of the skeleton	Size of the bounding box of the skeleton
θ	Angle	The three angles induces by the triangle formed by both shoulders and neck Angle between the vertical direction y and the axis binding the center of the hip and the head
$\mathcal{D}$	Distance	Right hand to the hips Left hand to the hips Right hand to the right shoulder Left hand to the left shoulder Right elbow to the hips Left elbow to the hips
A	Area	Triangle defined by both hands and neck Triangle defined by both shoulders and neck Triangle defined by both hands and hips Triangle defined by both elbows and neck
$\vec{v}$	Velocity	Hands Shoulders Hips Head Elbows
$\vec{a}$	Acceleration	Hands Shoulders Hips Head Elbows
$\mathcal{F}$	Frequency	Change of angle with respect to time of the following joints : Hands Shoulders Hips Head Elbows

Table 1: List of set of features that we propose to extract from the human postures for the classification of body expressions.

of synthetic animations generated by the method of Xia *et al*.[26]. We will refer the last database as SIGGRAPH database in the rest of this paper.

1. Biological Motion [19]: this database consists of 1356 motions, especially knocking actions for 4 expressions (angry, neutral, happy, sad).

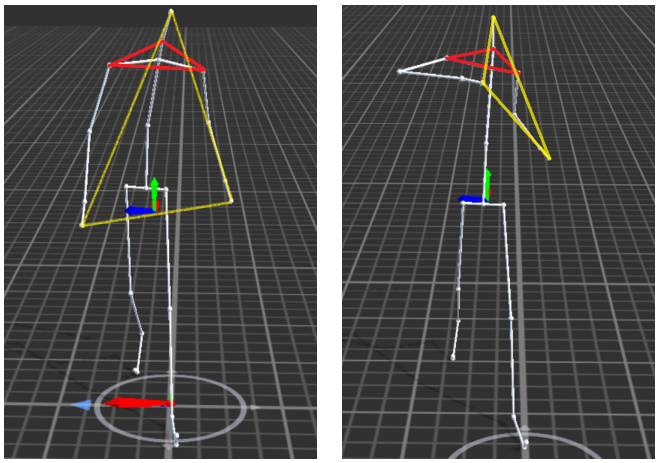
2. UCLIC Affective Body Posture and Motion [17]: this acted database contains 183 acted animations for 4 expressions (fear, sad, happy, angry).

3. SIGGRAPH database [26]: the SIGGRAPH database is synthetic database generated from 11 minutes of motion capture data. It contains 572 animations with 8 expressions or style (angry, childlike, depressed, neutral, old, proud, sexy, strutting). In above mentioned databases, the SIGGRAPH database includes the largest range of movements: jump, run, kick, walk.

DataBase	Number of movements	Number of expressions
UCLIC [17]	183	4
Biological [19]	1356	4
SIGGRAPH [26]	572	8

Table 2: Description of the databases used for the test of our method.

Summary of results obtained by our methods using the different sets of features are presented in Table 3. Our proposed approach can run in real-time as it runs at ˜50 fps (on PC running on i7-4710MQ with 8GB of RAM). Extraction of proposed features, except Fourier features, takes 10ms for each frame while the extraction of Fourier features takes 500ms for a sequence of 27 seconds with 1657 frames. Before the classification, we are scaling each attribute of our features to the range [1,+1] to avoid numerical difficulties during the cal-



(a) First frame of the depressed walking animation.

(b) First frame of the proud walking animation.

Fig. 3: Frames of the same action but with different body expressions. This figure show the variations of the triangle area formed by the both shoulders and the neck. Figure is color coded for the ease of visualization i.e. triangle formed by same joints in two figures are shown in similar color.

culation. Presented results have been obtained using the support vector machine with a radial basis function kernel (SVM - RBF kernel) method for the classification using a 10-fold cross validation technique. The parameters of the classifier were determined empirically. Proposed features (refer Section 3 for the detailed discussion on proposed features) can be categorized in the following sub-categories.

1. Geometric features: distances, area and angles of triangles.
2. Motion features: velocity and accelerations.
3. Fourier features: magnitude of the spectra for different joints.
4. All features: combines all features mentioned above.

The best result is obtained with the combination of the different proposed features (see Table 3). It is interesting to observe the fact that, correct classification accuracy does not degrade significantly when we used only geometric features.

On the two databases, SIGGRAPH and Biological Motion, geometric features achieved only one percent less correct classification accuracy than the best achieved results. On UCLIC database the recognition rate for the geometric features is 12% less than the combination of all features. This is probably due to the fact that this database is significantly smaller in sample size (183 actions) and presents worst case

DataBase	Set of features used	Results
UCLIC	All features	**78%**
UCLIC	Only geometric features	66%
UCLIC	Only motion features	52%
UCLIC	Only Fourier features	61%
SIGGRAPH	All features	**93%**
SIGGRAPH	Only geometric features	92%
SIGGRAPH	Only motion features	75%
SIGGRAPH	Only Fourier features	90%
Biological	All features	**57%**
Biological	Only geometric features	56%
Biological	Only motion features	48%
Biological	Only Fourier features	46%

Table 3: Our results on the different databases and with different set of features. Classification method is SVM.

scenario for machine learning algorithm. Proposed framework achieved best results on the synthetic database (SIGGRAPH) which is due to the fact that synthetic animations presents exaggerate expressions with high inter-class variations.

Table 4 shows that our method is competitive against state of the art on the UCLIC database with similar recognition rate. In the Biological Motion Database, movements are mainly knocking at a door ($\approx$ 1200 animations out of 1356 animations). The state of the art approach [4] uses this particularity to compute the average movement of knocking at a door and then subtracting this movement before running the recognition in order to emphasis the expression. Their recognition rate for this biased method is 81%. Nevertheless, this trick is possible for very specific movements, when you assume that all movements are similar. They learn a movement's specific bias. Since purpose of our proposed framework is to be robust against heterogeneous movements, we can not apply this assumption. We believe that to evaluate [4] and our approach, their unbiased recognition rate of 50% is to be compared with our approach that obtained correct classification accuracy of 57%. Finally, to the best of our knowledge, no method in literature on body expression analysis has tested the SIGGRAPH database.

DataBase	**Best results from state-of-the-art**	**Our results**
UCLIC	79% [18]	78%
Biological	50% unbias [4]	57%
SIGGRAPH	–	93%

Table 4: Comparison of our methods using all features mentioned in this paper to the state of the art methods.

5. CONCLUSION

We have presented novel approach for automatic recognition of body expressions through 3D skeleton provided by motion capture data. Taking inspiration from psychology domain state of the art, we have proposed simple and representative features to detect body expression from temporal 3D postures, even in complex cases: jump, run, kick etc. We have evaluated our approach on three databases that contain heterogeneous movements and expressions and obtained results that exceeds state of the art. Secondly, our proposed approach runs in real time due to its computation simplicity. Thus, opening up possibilities for human-computer interaction applications. One such application example is new generation of video games that can benefit from real time body expression analysis to adapt its content on run time.

As future work, we will aim at extending the proposed method to use more semantic meta-features reinforcing the recognition of body expressions. For instance, we will seek to fit the low-level features on analytic curves in order to use the curve parameters in the classification. And, we will compute incrementally the low-level features to achieve continuous recognition over time. An improvement of the validation will be to test our approach on multi-simultaneous actions and by cross-validating our method on several databases.

6. REFERENCES

[1] Kenji Amaya, Armin Bruderlin, and Tom Calvert. Emotion from motion. In *Graphics interface*, volume 96, pages 222–229. Toronto, Canada, 1996.

[2] Anthony P. Atkinson, Mary L. Tunstall, and Winand H. Dittrich. Evidence for distinct contributions of form and motion information to the recognition of emotions from body gestures. *Cognition*, 104(1):59 – 72, 2007. 00122.

[3] Stephen W. Bailey and Bobby Bodenheimer. A comparison of motion capture data recorded from a vicon system and a microsoft kinect sensor. In *Proceedings of the ACM Symposium on Applied Perception*, SAP '12, pages 121–121, New York, NY, USA, 2012. ACM.

[4] Daniel Bernhardt and Peter Robinson. Detecting affect from non-stylised body motions. In *Affective Computing and Intelligent Interaction*, pages 59–70. Springer, 2007.

[5] Vinay Bettadapura. Face expression recognition and analysis: the state of the art. *arXiv preprint arXiv:1203.6722*, 2012.

[6] Armin Bruderlin and Lance Williams. Motion signal processing. In *Proceedings of the 22nd annual conference on Computer graphics and interactive techniques*, pages 97–104. ACM, 1995.

[7] Ginevra Castellano, Santiago D. Villalba, and Antonio Camurri. Recognising human emotions from body movement and gesture dynamics. In *Affective computing and intelligent interaction*, pages 71–82. Springer, 2007.

[8] Klaus Forger and Tapio Takala. Animating with style: defining expressive semantics of motion. *The Visual Computer*, pages 1–13, 2015.

[9] Simon Fothergill, Helena Mentis, Pushmeet Kohli, and Sebastian Nowozin. Instructing people for training gestural interactive systems. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '12, pages 1737–1746, New York, NY, USA, 2012. ACM.

[10] Masahiro Hirai and Kazuo Hiraki. The relative importance of spatial versus temporal structure in the perception of biological motion: An event-related potential study. *Cognition*, 99(1):B15 – B29, 2006. 00030.

[11] Eugene Hsu, Kari Pulli, and Jovan Popovic. Style translation for human motion. *ACM Transactions on Graphics (TOG)*, 24(3):1082–1089, 2005.

[12] Heechul Jung, Sihaeng Lee, Junho Yim, Sunjeong Park, and Junmo Kim. Joint fine-tuning in deep neural networks for facial expression recognition. In *The IEEE International Conference on Computer Vision (ICCV)*, December 2015.

[13] Michelle Karg, Kolja Kuhnlenz, and Martin Buss. Recognition of Affect Based on Gait Patterns. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, 40(4):1050–1061, August 2010.

[14] R. A. Khan, A. Meyer, H. Konik, and S. Bouakaz. Exploring human visual system: study to aid the development of automatic facial expression recognition framework. In *Computer Vision and Pattern Recognition Workshop*, 2012.

[15] Rizwan Ahmed Khan, Alexandre Meyer, Hubert Konik, and Saida Bouakaz. Framework for reliable, real-time facial expression recognition for low resolution images. *Pattern Recognition Letters*, 34(10):1159 – 1168, 2013.

[16] Andrea Kleinsmith and Nadia Bianchi-Berthouze. Affective body expression perception and recognition: A survey. *Affective Computing, IEEE Transactions on*, 4(1):15–33, 2013.

[17] Andrea Kleinsmith, P. Ravindra De Silva, and Nadia Bianchi-Berthouze. Cross-cultural differences in recognizing affect from body posture. *Interacting with Computers*, 18(6):1371–1389, 2006.

[18] Andrea Kleinsmith, Tsuyoshi Fushimi, and Nadia Bianchi-Berthouze. An incremental and interactive affective posture recognition system. In *International Workshop on Adapting the Interaction Style to Affective Factors*, pages 378–387, 2005.

[19] Yingliang Ma, Helena M. Paterson, and Frank E. Pollick. A motion capture library for the study of identity, gender, and emotion perception from biological motion. *Behavior research methods*, 38(1):134–141, 2006.

[20] Albert Mehrabian and John T Friar. Encoding of attitude by a seated communicator via posture and position cues. *Journal of Consulting and Clinical Psychology*, 33(3):330, 1969.

[21] Microsoft. Kinect. https://developer.microsoft.com/en-us/windows/kinect.

[22] Helena M. Patterson, Frank E. Pollick, and Anthony J. Sanford. The role of velocity in affect discrimination. 2001.

[23] Jochen Tautges, Arno Zinke, Björn Krüger, Jan Baumann, Andreas Weber, Thomas Helten, Meinard Müller, Hans-Peter Seidel, and Bernd Eberhardt. Motion reconstruction using sparse accelerometer data. *ACM Trans. Graph.*, 30(3):18:1–18:12, May 2011.

[24] Arthur Truong, Hugo Boujut, and Titus Zaharia. Laban descriptors for gesture recognition and emotional analysis. *The Visual Computer*, 32(1):83–98, January 2016.

[25] Munetoshi Unuma, Ken Anjyo, and Ryozo Takeuchi. Fourier principles for emotion-based human figure animation. In *Proceedings of the 22nd annual conference on Computer graphics and interactive techniques*, pages 91–96. ACM, 1995.

[26] Shihong Xia, Congyi Wang, Jinxiang Chai, and Jessica Hodgins. Realtime style transfer for unlabeled heterogeneous human motion. *ACM Transactions on Graphics (TOG)*, 34(4):119, 2015.