Analyzing Muscle Activity and Force with Skin Shape Captured by Non-contact Visual Sensor

Ryusuke Sagawa[1], Yusuke Yoshiyasu[1], Alexander Alspach[2], Ko Ayusawa[1], Katsu Yamane[2,3], and Adrian Hilton[4]

[1] Intelligent Systems Research Institute,
National Institute of Advanced Industrial Science and Technology
[2] Disney Research, Pittsburgh
[3] The Robotics Institute, Carnegie Mellon University
[4] Centre for Vision, Speech and Signal Processing, University of Surrey

Abstract. Estimating physical information by vision as humans do is useful for the applications with physical interaction in the real world. For example, observing muscle bulging infers how much force a person puts on the muscle to interact with an object or environment. Since the human skin deforms due to muscle activity, it is expected that skin deformation gives information to analyze human motion. This paper demonstrates that biomechanical information can be derived from skin shape by analyzing the relationship between skin deformation, force produced by muscles, and muscle activity. We first obtained the dataset simultaneously acquired by a range sensor, a force sensor, and electromyograph (EMG) sensors. Since recent range sensors based on non-contact visual measurement acquires accurate and dense shape of an object at high frame rate, the deforming skin can be observed. The deformation is calculated by finding the correspondence between a template shape and each range scan. The relationship between skin deformation and other data is learned. In this paper, the following problems are considered: 1) estimating force from skin shape, 2) estimating muscle activity from skin shape, 3) synthesizing skin shape from muscle activity. In the experiments, the database learned from the sensor data can be used for the above problems, and the skin shape gives useful information to explain the muscle activity.

1 Introduction

Geometry of an object is one of the most important information to visually understand a scene. The geometrical analysis by using cameras has been a major topic of computer vision and has been applied to observe and analyze various targets including human motions. Geometrical information by vision is however often not sufficient for effective and safe physical interaction between humans, robots, objects, and the environment. When a robot picks up an object, for example, motion planning and control would be easier if its physical properties such as weight and compliance are known in advance. It would also be beneficial to know the ground surface properties such as friction coefficient before actually

walking on it. For human-robot interaction, knowing a person's muscle tension level is useful for deciding whether the robot should offer physical help. Unfortunately, obtaining such information requires installation of additional force sensors outside the robot, which is not always possible especially in uncontrolled environments.

For humans, it is actually possible to estimate physical quantities only from visual clues in some special cases. For example, we can roughly estimate an object's weight by observing another person holding the object, possibly based on such clues as the person's facial expression, pose, and body shape including muscle bulging, all of which are available through visual sensors. If we can implement a similar inference process, robots would be able to estimate physical quantities without additional sensors.

This paper presents an example of estimating physical quantities from visual information. More specifically, we develop a method for estimating the force applied to the environment by a human subject using muscle geometry data obtained from a projector-camera system. The method is then applied to force estimation of objects dropped onto the subject's hand.

Motion capture (mocap) system is widely used for human motion analysis. The articulated motion of human body is captured by tracking markers attached on the human body. To obtain the physical information, force plates are used to measure the force produced by human body with a mocap system. Since force plates are placed on the floor, the internal forces of a human body cannot be measured directly. To obtain the information of individual muscle, electromyograph (EMG) sensors are often used. The sensors are attached to the skin above the target muscles, and record the electrical activity produced by muscles. Since using these sensors is high cost, it is useful and our future goal if physical information is predicted only by a non-contact visual sensor. In this paper, we use these sensors for training step of the proposed method, and compared to the results in the evaluation.

The idea we adopt in this paper is to observe skin shape based on non-contact visual measurement. If the activity level of a muscle changes, the skin above the muscle deforms according to the muscle shape. It is therefore expected that the skin shape has information to estimate the muscle activity. Since the skin deformation is non-parametric and the difference between individuals is difficult to model, we take an approach to learn the relationship between skin shape, force produced by muscles, and muscle activity with the dataset simultaneously acquired by a range sensor (a.k.a. depth sensor), a force sensor, and EMG sensors.

Range scans are captured frame-by-frame, and it is necessary to calculate the deformation from the acquired data to prepare the dataset. The proposed method consists of four steps: 1) acquire a dataset of various motions with a range sensor, a force sensor, and EMG sensors, 2) extract the feature vector of skin deformation by finding the correspondence between a template shape and range scans, 3) train the database with acquire dataset, and 4) predict force, muscle activity, and skin shape for input variables. The contribution of this paper is summarized as follows.

- It is demonstrated that visual measurement of surface shape can be used to estimate arm muscle activation and force.
- The proposed approach based on learning the relationship succeeds to predict force, muscle activity, and skin shapes each other.

2   Related work

The analysis of skin deformation is studied in the field of computer graphics. Skin deformation according to body pose is important factor to generate a realistic model of human body. The methods of modeling muscles is classified to three approaches: geometrically-based, physically-based, and data-driven approaches [8]. The first one models the animation effects of muscle contraction and succeeded to model simple muscles. The second approach involves muscle dynamics and tissue properties to model complex scene of muscles. The third approach directly models the skin shape from the data captured by a mocap system or a range sensor. Skin deformation between individuals are modeled in [1]. The deformation of body shape according to poses are modeled in [3]. The muscle deformation is model from the range data obtained by a depth camera in [18]. In [17], the skin deformation is learned with respect to the pose and acceleration of body parts by kernel regression. The acceleration is used as the external force for each body part. In [14], the skin shape of a moving arm with a barbell in the hand is captured, which is used as external force. The relationship between body pose, body shape and external force is learned by kernel ridge regression. The shape is parameterized base on [20]. The relationship learned from the dataset is used for synthesizing skin shape for a new pose and external force. The co-contraction of multiple muscles is not considered in the methods. The problem considered in their paper is an inverse problem of estimating force from skin shape, which is tackled in this paper.

To capture dynamic objects by non-contact visual sensors, various methods have been proposed, which are roughly classified to passive and active methods. From the viewpoint of accuracy and robustness, active methods are easy to apply. In active methods, Time-of-Flight (TOF) cameras and triangulation based methods are widely used. TOF cameras project temporally-modulated light patterns and acquire a depth image at once by capturing the reflections [13]. In triangulation-based methods, two approaches have proposed to capture dynamic object: temporal-encoding [23, 27] or spatial-encoding method [22, 12, 19]. While temporal-encoding methods need to project multiple patterns, a spatial-encoding method is suitable to capture dynamic scenes because a single image of a fixed projector pattern is sufficient for reconstruction.

Since skin deforms nonrigidly, nonrigid surface registration techniques are required to find the correspondence between a template shape and range scans. The surface registration techniques can be divided into three categories in terms of regularizations that they use: smoothness regularization, isometric regularization and conformal regularization. Early approaches [24, 25, 1, 2] are based on smoothness regularization. These techniques are very flexible in that they can
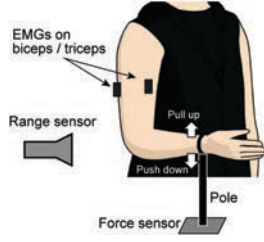
Fig. 1. The setup of the experiment: The skin shape of the arm of a subject is observed by using a range sensor.

(a)   (b)

Fig. 2. (a) an input image of the projector-camera system that casts the static wave-grid pattern on the target. (b) the shape computed from the image (a).

change the shape of the template quite largely; for example, they can deform a sphere into a tooth, if adequate landmarks are given. However, they are poor at preserving template details and mesh structures, which means that they need many landmarks to work properly. In contrast, the isometric (as-rigid-as possible) regularization can preserve original template details and are commonly used in automatic registration techniques [9, 7, 21]. The drawback of these methods are that they are incapable of handling models with different sizes or those which undergo large local stretching. Recently, the techniques based on conformal (as-similar-as possible) regularization [10, 16, 26] are proposed to achieve both flexibility in changing shapes and preservation of mesh structure. They are based on angle-preserving deformation.

3   Data acquisition

In this paper, we analyze the relationship between the skin shape and the muscle activity of a person. However, the shape of skin can be affected by other various conditions. For example, the skin shape depends on the angle of the joint even if the muscle is relaxed, and the muscle mass is different between individuals. Therefore, we simplify the condition considered in this paper, and focus on the relationship between skin shape and muscle activity. Fig.1 shows the setup of the experiment we analyze in this paper. We observe the skin shape of the arm of a subject from the side by using a range sensor. The elbow joint is almost fixed so that the angle is 90 degrees by strapping the wrist to a pole. When the subject pushes down or pulls up the wrist by trying to straighten/bend the elbow, the force transmitted from the pole is measured by the force sensor. In this situation, the force is mainly produced by biceps and triceps muscles. The muscle activity is measured by two EMG sensors placed on the skin above the muscles. Our challenge is to estimate the muscle activity and the produced force from the skin shape, which is the first approach to the best of our knowledge to derive muscle activation from visual observations even in the limited situation.

The shape of the skin is deformed according to the muscle activity. We observed the shape of the front and upper arm at 30∼50 frames/second by using a
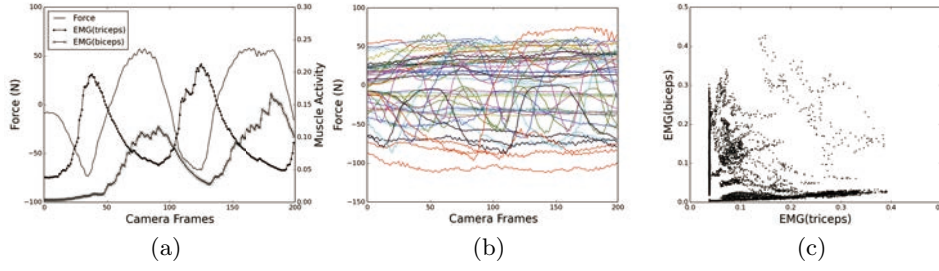
Fig. 3. (a) An example of the force and muscle activity simultaneously acquired in a push-and-pull sequence. (b) The range of force of all sequences is -110∼+75N. The force is measured along the vertical axis, and the upward direction is positive. (c) The distribution of the activity of two muscles. It shows that both muscles are simultaneously active a some sequences.

projector-camera system proposed in [19]. Since the shape of a front arm is affected by the wrist and hand pose, the wrist angle is fixed to be straight as much as possible and the hand stays open and relaxed during the shape acquisition. The shape of the hand is omitted from the analysis to remove the effect in the experiment. Fig.2 shows one of the frame of the range sensor. (a) is the input image of the projector-camera system that casts the static wave-grid pattern on the target. (b) is the shape computed from the image (a). The image acquisition is synchronized with the force sensor and the EMGs.

We captured three types of the sequences: pull-up, push-down, and push-and-pull motions. In the push-and-pull sequences, the subject changed the force direction during the sequence, while the pole was continuously pulled up or pushed down the pole during the pull-up and push-down sequences. 10K frames of range scan are obtained in total. An example of the force and muscle activity is shown in Fig.3(a). The force is measured along the vertical axis, and the upward direction is positive. The range of force of all sequences is -110∼+75N as shown in Fig.3(b). The muscle activity is defined as the integrated EMG signal normalized by the signal of the maximal voluntary contraction (MVC) [11]. If it is close to zero, the muscle is relaxed. Since this sequence is one of the push-and-pull sequences, the co-contraction of the biceps and triceps muscles is observed. Fig.3(c) shows the distribution of the activity of two muscles, and both muscles are simultaneously active in some sequences.

4   Feature extraction of skin deformation

The skin shape deforms according to the muscle activity. We use the deformation of a template shape as the feature vector to explain the state of muscle.The template shape is a range scan with the least muscle activity chosen from all range scans. The deformation of the template shape are calculated for all vertices

of each frame. Since the range scans is captured by the range sensor frame-by-frame, the deformation is calculated by finding the correspondence between the template shape and each range scan by the method explained in Sec.4.2.

4.1   Defining feature vector to explain skin deformation

The template shape is represented as a set of 3D vertices. Now, we assume each vertex $\boldsymbol{v}$ of the template shape corresponds to the point $P(\boldsymbol{v})$ of a range scan. Although we want to extract the skin deformation caused by muscle activity, the difference $P(\boldsymbol{v}) - \boldsymbol{v}$ includes the change of arm pose in addition to the skin deformation. Although the arm is almost fixed in the experiment of this paper, it slightly moves and the elbow angle changes. Additionally, we assume the arm pose does not affect the skin deformation since the arm motion is small. If we know the arm pose, the relationship between $P(\boldsymbol{v})$ and $\boldsymbol{v}$ is written as

$$P(\boldsymbol{v}) = R(\boldsymbol{v} + d(\boldsymbol{v})) + \boldsymbol{t}, \tag{1}$$

where R and $\boldsymbol{t}$ are the rotation and translation calculated by the arm pose for the vertex, respectively. $d(\boldsymbol{v})$ is the vector of skin deformation, which is used as the feature to explain the muscle activity. $d(\boldsymbol{v})$ is calculated from the correspondence by $d(\boldsymbol{v}) = R^T(P(\boldsymbol{v}) - \boldsymbol{t}) - \boldsymbol{v}$.

Once the skin deformation $d(\boldsymbol{v})$ is calculated for each vertex of the template shape, the feature vector D_k for k-th range scan is defined by

$$D_k = [d_k(\boldsymbol{v}_1), d_k(\boldsymbol{v}_2), \dots d_k(\boldsymbol{v}_M)], \tag{2}$$

where M is the number of vertices in the template shape. Since $d_k(\boldsymbol{v})$ is a three-dimensional vector, the length of feature vector D_k is $3M$.

4.2   Finding correspondence between the template shape and each range scan

Calculating the skin deformation $d(\boldsymbol{v})$ is based on finding the correspondence between the template shape and each range scan. The surface of an arm changes according to the arm pose and the muscle activity under the skin. Although the arm pose can be estimated by the registration of articulated model or by using a motion capture system, the residual of the deformation is not explained by an articulated model. Therefore, a method of vertex-wise nonrigid registration between two models is necessary to find the correspondence for each vertex.

Nonrigid registration is achieved using the conformal (angle-preserving) registration technique [26] that can capture spatially-varying scale changes in order to capture muscle bulges. This technique assigns an affine transformation X_i to each vertex of the template shape and optimizes the transformation to attract the template toward a range scan while preserving the angles of triangle meshes as much as possible. We use the as-similar-as possible formulation that constrains
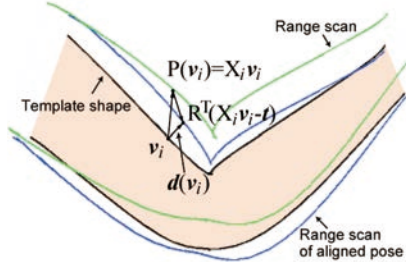
Fig. 4. The template shape and a range scan is represented by black and green curves. The range scan aligned by using the relative pose is represented by the blue curve.
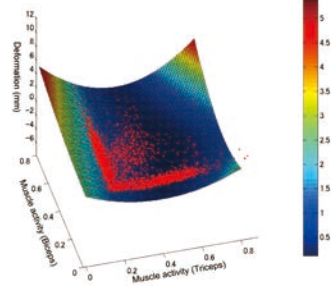

Fig. 5. The curved surface represents the B-spline function fitted to the red points, which are the deformation of captured data.

affine transformations to similarity transformations. Let $X = [X_1 \ldots X_M]^T$ be the affine transformations associated to vertices. We define the cost function

$$E(X) = w_{\text{ASAP}} E_{\text{ASAP}} + w_{\text{reg}} E_{\text{reg}} + w_{\text{Closest}} E_{\text{Closest}}, \qquad (3)$$

where E_{ASAP} constrains deformation as similar as possible, E_{reg} acts as a regularization term to avoid extreme local deformation, E_{Closest} penalizes distances between the closest points of template and target surface. The energy is minimized using the alternating optimization technique where the first step optimizes the vertex positions with fixed transformations and the second step optimizes affine transformations with fixed vertex positions.

Once X_i is estimated, the corresponding point is expressed by $P(\boldsymbol{v}_i) = X_i \boldsymbol{v}_i$. To cancel the effect of the pose, The rotation R and translation $\boldsymbol{t}$ are calculated by assuming locally rigid around each vertex. The parameters are estimated by minimizing the following equation:

$$\arg \min_{R, \boldsymbol{t}} \sum_{\boldsymbol{v}' \in N(\boldsymbol{v})} (X_i \boldsymbol{v}' - R \boldsymbol{v}' - \boldsymbol{t})^2, \qquad (4)$$

where $N(\boldsymbol{v})$ is the set of neighborhood vertices that satisfies $\| \boldsymbol{v}' - \boldsymbol{v} \| < r$. r is a user-defined threshold of the radius around the vertex.

Fig.4 illustrates the calculation of the skin deformation $d(\boldsymbol{v}_i)$. The template shape and a range scan is represented by black and green curves. The corresponding point $X_i \boldsymbol{v}_i$ for a vertex $\boldsymbol{v}_i$ is found by nonrigid registration. The rotation R and translation $\boldsymbol{t}$ for each vertex of the range scan relative to the template shape is estimated by Eq.(4). The range scan aligned by using R and $\boldsymbol{t}$ is represented by blue curve. $d(\boldsymbol{v}_i)$ is calculated as the difference of the corresponding points between the template shape and the range scan of aligned pose.

5   Learning the relationship between skin shape, force, and muscle activity

Although the setup of the experiment performed in this paper is simple, multiple muscles affects the skin shape, and the shape of each muscle and the amount of fat under the skin are unknown. It is therefore difficult to model the skin shape based on the muscle model. Instead, we learn the relationship between skin and muscle from the acquired data. We obtained the skin shape, the force produced by an arm, the muscle activity measured by EMG sensors. We consider three problems in this section.

The first problem is estimating the force from the skin shape. If the force is calculated from the skin shape, which is obtained by non-contact sensor, it gives useful information for biomechanical analysis.

The second one is estimating muscle activity from skin shape. The force produced by an arm is the result of multiple active muscles. If the force can be estimated from the skin shape, it should be able to detect the co-contraction of multiple muscles. We clarify if the skin shape reflects the muscle activity.

The third one is the inverse problem of the second one. If the relationship between skin shape and muscle activity is one-to-one correspondence, the skin shape can be synthesized from a given state of muscle activity, which is useful to generate realistic model of human skin.

Since the skin deformation is non-linear, they are problems to determine the non-linear function between the given data and the estimated result. We find the function by learning-based approach. Since the all values of skin deformation, force, and muscle activity are continuous, they are defined as non-linear regression. We chose Random Forests (RF) [4] as the learning method in this paper, which is applied for regression problems [5, 6, 15].

5.1   Estimating force from skin shape

The first problem is to estimate the force f from a given feature vector D of the skin. The function F_f to be determined is defined as $f = F_f(D)$. In this definition of the problem, the muscle activity is not considered explicitly, and we estimate the force directly from the skin shape. EMG sensors are not necessary for this estimation, we use them to check if the co-contraction is occurred or not as shown in Fig.3(a).

In the training phase, we have multiple training examples $(D_k, f_k), k = 1, \ldots, L$, which are obtained at L camera frames during data acquisition. The RF algorithm builds multiple decision trees, of which each leaf node stores the values of f. In the predicting phase, a new feature vector of skin deformation D_{new} is given and applied to the decision trees. The predicted value of f for D_{new} is calculated as the average of multiple leaf nodes.

Each training example is constructed from the data of a single camera frame, and the prediction phase is performed frame-by-frame. No temporal information is used in the current implementation. The number of vertices in the template shape is about 8K∼15K points.

5.2   Estimating muscle activity from skin shape

The second problem is to estimate the muscle activity $\boldsymbol{a}$ from a given feature vector D of the skin. The function F_a to be determined is defined as $\boldsymbol{a} = F_a(D)$. Because we used two EMG sensors in this paper, the activity state $\boldsymbol{a}$ is a two-dimensional vector $\boldsymbol{a} = (a_t, a_b)$, which are activities of triceps and biceps, respectively.

The training and predicting phase is similar to the case of estimating force, but the output of the prediction is a two-dimensional vector $\boldsymbol{a}$. The prediction phase is also performed frame-by-frame in this estimation.

5.3   Synthesizing skin shape from muscle activity

The dataset of the third problem is the same with the second one, but the input and output data are swapped. The feature vector D of the skin is estimated from a given muscle activity $\boldsymbol{a}$. The function F_D to be determined is defined as $D = F_D(\boldsymbol{a})$, which means we assume the skin shape only depends on the activity of muscles that relates to the skin if the joint angle is known.

The captured shape is actually affected by other reasons such as the activity of another muscle and the error of the measurements. We estimate the function by approximating the measured deformation to determine the deformation vector uniquely with respect to muscle activity $\boldsymbol{a}$. Since the motion of muscles is smooth, we estimate a B-spline function defined by

$$F_{D,i}(a_t, a_b) = \sum_{t=1,2,3, b=1,2,3} p_{t,b} w_{t,b} \tag{5}$$

$$[w_{t,b}] = A_t^T A_b, \quad A_t = \boldsymbol{a}_t W, \quad A_b = \boldsymbol{a}_b W,$$

for i-th dimension of the feature vector D, where $p_{t,b}(t = 1, 2, 3, b = 1, 2, 3)$ are 9 control parameters of B-spline function, $\boldsymbol{a}_t = [a_t^2, a_t, 1]$, $\boldsymbol{a}_b = [a_b^2, a_b, 1]$, and a 3×3 matrix $W = \frac{1}{2}[1, -2, 1; -2, 2, 0; 1, 1, 0]$. The muscle activities are normalized to the range $0 \sim 1.0$ by using MVC. The total number of parameters are $27M$ for M vertices of the template shape. $p_{t,b}$ for i-th dimension of D is estimated by minimizing the error $\sum_k (F_{D,i}(a_t, a_b) - D_{k,i})^2$, where $D_{k,i}$ is the i-th dimension of D.

Fig.5 shows an example of the approximating function. The curved surface represents the B-spline function fitted to the red points, which are the deformation of captured data. If an activity state $\boldsymbol{a}$ is given, the deformation vector is calculated by $F_D(\boldsymbol{a})$ for each dimension, and the skin shape for the state is calculated by adding the deformation vector to the template shape. The captured data is expected to cover the space of the input activity $\boldsymbol{a}$ for this approach to work properly.

6   Experiments

In the experiments, we analyze if the skin deformation has sufficient information to explain the force produced by the arm and the muscle activity. First, we
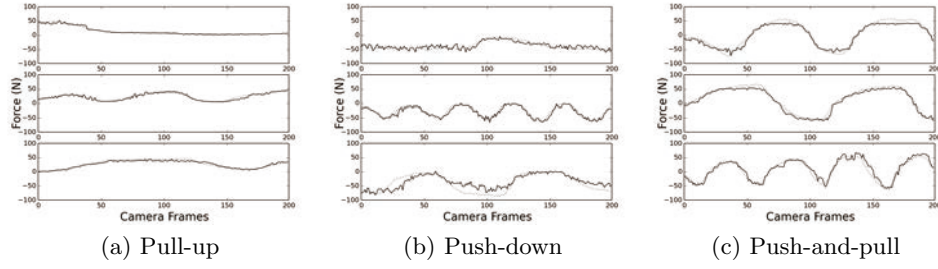
(a) Pull-up   (b) Push-down   (c) Push-and-pull

Fig. 6. The results of predicting force are shown for 9 sequences. The solid curves are the results of prediction and the dotted curves are the values measured by the force sensor: (a) the results of three sequences of pull-up motion. (b) the results of three sequences of push-down motion. (c) the results of three push-and-pull motion.

test the force estimation from the skin shape. We evaluate the accuracy of the predicted force by using the trained database. The accuracy is calculated based on leave-one-out cross-validation by using 48 acquired sequences of 200 frames. Fig.6 shows some of the results of predicting force. The solid curves are the results of prediction and the dotted curves are the values measured by the force sensor. The results of three sequences of pull-up motion is shown in (a). The RMS error of pull-up sequences is 5.31N. Since the maximum force in the sequences is about 75N, the error is less than 10% of the range. The results of three sequences of push-down motion is shown in (b), and the RMS error is 11.87N, which is larger than that of the pull-up sequences. The reason is considered that the movement of triceps is smaller than that of biceps and the viewing direction of the range sensor was not the best for observing the triceps. The results of push-and-pull motion is shown in (c), and the RMS error is 13.97N. Although the co-contraction occurs in the push-and-pull sequences as shown in Fig.3(a), the proposed method succeeded to predict the force for each frame.

The next experiment is estimating muscle activity from skin shape. Fig.7 shows the results of predicting muscle activity. The solid curves are the results of prediction and the dotted curves are the values measured by the EMG sensors. The blue curves indicate the values of biceps, and the green curves are those of triceps. The sequences in Fig.7 are the same ones in Fig.6. The RMS error of pull-up sequences shown in (a) is 0.0227. The RMS error of push-down motion shown in (b) is 0.0317. The reason that the error is larger than that of the pull-up sequences is considered the same with the force estimation. The results of push-and-pull motion is shown in (c), and the RMS error is 0.0598. Since the error is sufficiently small compared to MVC, the proposed method succeeded to predict the muscle activity from the skin shape. However, the sequences with co-contraction have larger errors than the other sequences. It is considered because the number of training examples with co-contraction were not large enough to
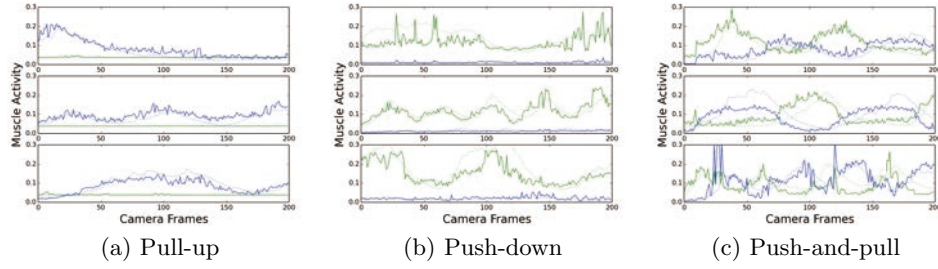
(a) Pull-up   (b) Push-down   (c) Push-and-pull

Fig. 7. The results of predicting muscle activity are shown for the same sequences in Fig.6. The solid curves are the results of prediction and the dotted curves are the values measured by the EMG sensors. The blue curves indicate the values of biceps, and the green curves are those of triceps. The activity values are normalized by maximal voluntary contraction (MVC).

construct the database. To increase the training dataset is one of our future work to improve the result.

The next experiment is synthesizing skin shape from muscle activity. We synthesize the skin shapes based on Eq.(5) for given muscle activity. Fig.8 shows three examples of synthesized shapes by changing the biceps activity a_b from 0.16 to 0.5, while the triceps activity a_t is fixed to 0.08. The muscle around the biceps near the inside of the elbow bulges out according to the activation. 6.5K frames of the captured data is used as the training set for the approximation. The accuracy of the synthesized deformation vectors is estimated by using 4.9K frames that are not in the training set. The RMS error of deformation vector is 1.09mm.

The last experiment is observing a dynamic scene of catching a dropped bottle shown in Fig.9(a). We tried to estimate the force during catching by changing the weight of the bottle. Fig.9(b) is the frame at the moment of catching of a light bottle (0.65kg), and (c) is that of a heavy bottle (1.80kg). While it is difficult to recognize the difference in the input images, the shapes have difference around the biceps. Fig.10 shows the results of the predicted force by observing the arm by the proposed method (solid lines). The force is compared to the one calculated from the motion of the bottle obtained by using a motion capture system (dotted lines). The blue lines are the results of a light bottle, while the red lines are those of a heavy bottle. The frame rate of the camera is 30 frames/second, while that of the motion capture system is 120 frames/second. Although the frame rate of the camera was not sufficient to observe the impact, the proposed method successfully estimated the force in the following frames. Fig.10(a) and (b) are two examples of predicting the force by changing how to catch. The predicted forces clearly have difference between the light and heavy bottles. Since the motion of the wrist is obtained by nonrigid registration, the weight of the bottle can be estimated from the force and the acceleration of the wrist motion. By using frames that the acceleration is small to reduce the noise
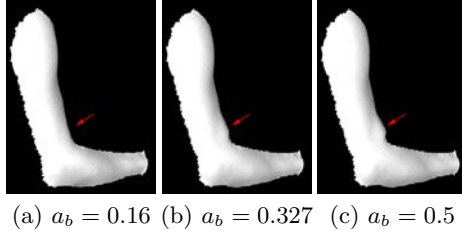
(a) $a_b = 0.16$ (b) $a_b = 0.327$ (c) $a_b = 0.5$

Fig. 8. Three examples of synthesized shapes are shown by changing the biceps activity a_b from 0.16 to 0.5, while the triceps activity a_t is fixed to 0.08. The muscle around the biceps near the inside of the elbow bulges out according to the activation.
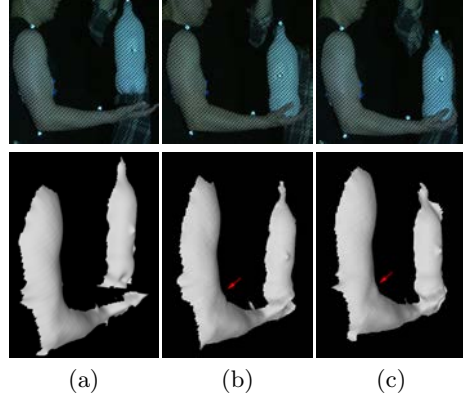

(a)   (b)   (c)

Fig. 9. A dynamic scene of catching a dropped bottle is captured. The weights of the bottle are light (0.65kg) in (b) and heavy (1.80kg) in (c). The shapes of the arm have difference around the biceps.

in calculating the second order derivative of the motion, the average weights estimated for the light bottle are 0.92kg and 0.79kg for these two examples, while those of the heavy bottle are 1.46kg and 1.57kg. The error of the estimated weight is 0.1∼0.4kg. The results indicate that the proposed method can recognize which bottle is heavy from the visual information without using contact sensors at the testing phase.

7   Conclusion

Obtaining physical information by using non-contact visual sensor is important for the applications with physical interaction in the real world. As one of estimating physical information by vision, we proposed a method to analyze the relationship between skin deformation, force produced by muscles, and muscle activity in this paper. We acquired the dataset simultaneously acquired by a range sensor, a force sensor, and EMG sensors. Since the range sensor based on non-contact visual measurement can capture the dense shape of skin, the skin deformation caused by muscle activation is calculated by finding the correspondence between a template shape and each range scan. The database of the relationship between skin deformation, force, and muscle activity is constructed by learning based on Random Forests. The proposed method succeeded to three prediction problems: estimating force from skin shape, estimating muscle activity from skin shape, synthesizing skin shape from muscle activity. The results of experiments shows that skin shape obtained by non-contact visual measurement has useful information to estimate the muscle force and muscle activation.

In our future work, we plan to improve the prediction of force and muscle activation. One idea is to construct the database by increasing the parameters
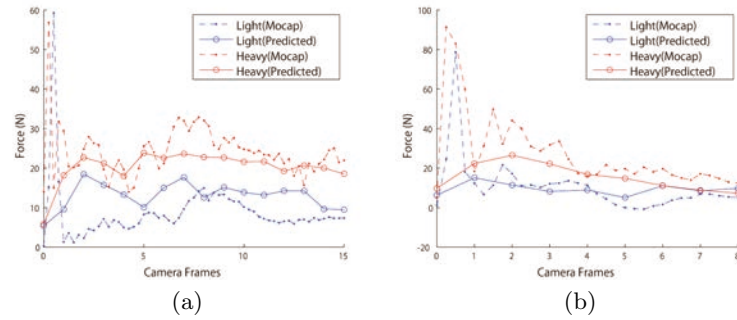
(a)   (b)

Fig. 10. Two examples of predicting force by changing how to catch. The solid lines are the results of the predicted force by observing the arm by the proposed method. The dotted lines are the force calculated from the motion of the bottle obtained by using a motion capture system. The blue lines are the results of a light bottle, while the red lines are those of a heavy bottle. The frame rate of the camera is 30 frames/second, while that of the motion capture system is 120 frames/second.

of the setup, for example, multiple subject, flexible joint motion, and multiple joints of human body. Additionally, we will observe the motion of the skin at the moment of impact by using a high speed camera to analyze dynamic motion more deeply.

Acknowledgments

References

1. Allen, B., Curless, B., Popović, Z.: The space of human body shapes: reconstruction and parameterization from range scans. ACM Trans. Graph. 22(3), 587–594 (2003)
2. Amberg, B., Romdhani, S., Vetter, T.: Optimal Step Nonrigid ICP Algorithms for Surface Registration. In: CVPR (2007)
3. Anguelov, D., Srinivasan, P., Koller, D., Thrun, S., Rodgers, J., Davis, J.: Scape: shape completion and animation of people. ACM Trans. Graph. 24(3), 408–416 (2005)
4. Breiman, L.: Random forests. Machine Learning 45(1), 5–32 (2001)
5. Criminisi, A., Shotton, J., Konukoglu, E.: Decision forests for classification, regression, density estimation, manifold learning and semi-supervised learning. Tech. Rep. MSR-TR-2011-114, Microsoft Research (2011)
6. Gall, J., Yao, A., Razavi, N., Van Gool, L., Lempitsky, V.: Hough forests for object detection, tracking, and action recognition. IEEE Transactions on Pattern Analysis and Machine Intelligence 33(11), 2188–2202 (2011)

7. Huang, Q.X., Adams, B., Wicke, M., Guibas, L.J.: Non-rigid registration under iso-metric deformations. In: Proceedings of the Symposium on Geometry Processing. pp. 1449–1457 (2008)
8. Lee, D., Glueck, M., Khan, A., Fiume, E., Jackson, K.: Modeling and simulation of skeletal muscle for computer graphics: A survey. Foundations and Trends in Computer Graphics and Vision 7(4), 229–276 (2012)
9. Li, H., Sumner, R.W., Pauly, M.: Global correspondence optimization for non-rigid registration of depth scans. In: Proceedings of the Symposium on Geometry Processing. pp. 1421–1430 (2008)
10. Liao, M., Zhang, Q., Wang, H., Yang, R., Gong, M.: Modeling deformable objects from a single depth camera. In: ICCV. pp. 167–174 (2009)
11. Merletti, R.: Standards for reporting EMG data. Journal of Electromyography and Kinesiology 9(1), III–IV (February 1999)
12. Microsoft: Xbox 360 Kinect (2010), http://www.xbox.com/en-US/kinect
13. Microsoft: Kinect for Windows (2013), http://www.microsoft.com/en-us/kinectforwindows
14. Neumann, T., Varanasi, K., Hasler, N., Wacker, M., Magnor, M., Theobalt, C.: Capture and statistical modeling of arm-muscle deformations. Computer Graphics Forum (Proc. of Eurographics) 32(2pt3), 285–294 (May 2013)
15. Okada, R.: Discriminative generalized hough transform for object dectection. In: Proc. 2009 IEEE 12th International Conference on Computer Vision. pp. 2000–2005 (2009)
16. Papazov, C., Burschka, D.: Deformable 3D Shape Registration Based on Local Similarity Transforms. Computer Graphics Forum 30 (2011)
17. Park, S., Hodgins, J.: Data-driven modeling of skin and muscle deformation. In: ACM TOG (2008)
18. Robertini, N., Neumann, T., Varanasi, K., Theobalt, C.: Capture of arm-muscle de-formations using a depth-camera. In: Proc. European Conference on Visual Media Production (CVMP). vol. 10 (2013)
19. Sagawa, R., Sakashita, K., Kasuya, N., Kawasaki, H., Furukawa, R., Yagi, Y.: Grid-based active stereo with single-colored wave pattern for dense one-shot 3D scan. In: 3DIMPVT. pp. 363–370 (2012)
20. Sumner, R., Popović, J.: Deformation transfer for triangle meshes. In: SIGGRAPH '04: ACM SIGGRAPH 2004 Papers. pp. 399–405 (2004)
21. Tevs, A., Bokeloh, M., Wand, M., Schilling, A., Seidel, H.P.: Isometric registration of ambiguous and partial data. In: CVPR. pp. 1185–1192 (2009)
22. Vuylsteke, P., Oosterlinck, A.: Range image acquisition with a single binary-encoded light pattern 12(2), 148–164 (1990)
23. Weise, T., Leibe, B., Gool, L.V.: Fast 3D scanning with automatic motion com-pensation. In: CVPR (2007)
24. Weise, T., Li, H., Van Gool, L., Pauly, M.: Face/Off: live facial puppetry. In: Proceedings of the 2009 ACM SIGGRAPH/Eurographics Symposium on Computer Animation. pp. 7–16 (2009)
25. Yeh, I.C., Lin, C.H., Sorkine, O., Lee, T.Y.: Template-based 3D Model Fitting Using Dual-domain Relaxation. IEEE Transactions on Visualization and Computer Graphics 99 (2010)
26. Yoshiyasu, Y., Ma, W.C., Yoshida, E., Kanehiro, F.: As-conformal-as-possible sur-face registration. Computer Graphics Forum 33(5), 257–267 (2014)
27. Zhang, L., Curless, B., Seitz, S.M.: Spacetime stereo: Shape recovery for dynamic scenes. In: IEEE Computer Society Conference on Computer Vision and Pattern Recognition. pp. 367–374 (June 2003)