Cloth Texture Recognition using Vision and Tactile Sensing

Conference Paper · May 2018

CITATIONS
0

READS
94

5 authors, including:

Shan Luo
University of Liverpool
30 PUBLICATIONS **239** CITATIONS
SEE PROFILE

Anthony G. Cohn
University of Leeds
278 PUBLICATIONS **8,411** CITATIONS
SEE PROFILE

Raul Fuentes
University of Leeds
36 PUBLICATIONS **58** CITATIONS
SEE PROFILE

Cloth Texture Recognition using Vision and Tactile Sensing

Shan Luo[1,2,3,4], Wenzhen Yuan[1], Edward Adelson[1], Anthony G. Cohn[2] and Raul Fuentes[3]

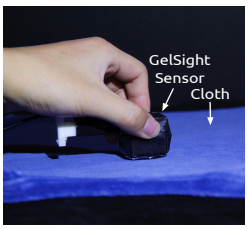
***Abstract*— Vision and touch are two of the important sensing modalities for humans and they offer complementary information for sensing the environment. Robots could also benefit from such multi-modal sensing ability. In this paper, addressing for the first time (to the best of our knowledge) texture recognition from tactile images and vision, we propose a new fusion method named Deep Maximum Covariance Analysis (DMCA) to learn a joint latent space for sharing features through vision and tactile sensing. Results of the algorithm on a newly collected dataset of paired visual and tactile data relating to cloth textures show that a good recognition performance of greater than 90% can be achieved by using the proposed DMCA framework. In addition, we find that the perception performance of either vision or tactile sensing can be improved by employing the shared representation space, compared to learning from unimodal data.**

I. Introduction

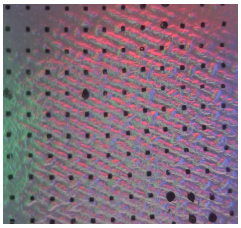
We humans have much experience of "touching to see" and "seeing to feel". For instance, when we intend to grasp an object, we are likely to glimpse it first with our eyes to "feel" its key features, i.e., shapes and textures, and estimate haptic sensations. Such visual features become unobservable after the object is grasped since vision is occluded by the hand and becomes ineffective. In this case, touch sensation distributed in the hand can assist us to "see" corresponding features. By tracking and sharing these clues through vision and tactile sensing, we can "see" or "feel" the object better.

In this paper, we take cloth texture recognition as the test arena to apply this feature sharing mechanism between vision and tactile sensing in robotics: the tactile sensing can perceive very detailed texture such as yarn distribution pattern in the cloth whereas vision can capture similar texture pattern (though sometimes is quite blurry). There are also factors that only exist in one modality that may deteriorate the recognition performance. For instance, color variance of cloth is present in vision but is not demonstrated in tactile sensing. We aim to extract the shared information of both modalities while eliminating these factors. We propose a novel deep fusion framework based on deep neural networks and maximum covariance analysis to learn a joint latent space of vision and tactile sensing. A newly collected dataset of paired visual and tactile data is also introduced.

In the prior works low-resolution tactile sensors (for instance a Weiss tactile sensor of 14×6 taxels) are commonly used to confirm the contacts, instead we use a high-resolution GelSight sensor of (960×720) to capture more detailed textures which is a much harder problem than just confirming the contacts. The GelSight sensor consists of a camera at the bottom and a piece of elastometric gel coated with a reflective membrane on the top. The elastomer deforms to take the surface geometry and texture of the objects that it interacts with. The deformation is then recorded by the camera under illumination from LEDs that project from various directions through light guiding plates towards the membrane. Furthermore, to the best of the authors' knowledge, this is the first work to explore both tactile images and vision data for texture recognition.



(a)



(b)

Fig. 1: (a) GelSight data is collected by pressing the sensor on clothes [1]. (b) The GelSight image collected when the coated membrane is deformed by the cloth texture.

II. ViTac Cloth dataset

We have built a clothing dataset of 100 pieces of everyday clothing of both visual and tactile data, which we call the *ViTac Cloth dataset*. The clothing are of various types and are made of a variety of fabrics with different textures. In contrast to available datasets with only either visual images [2] or tactile readings [3] of surface textures, the data of two modalities, i.e., vision and touch, was collected while the cloth was lying flat. The color images were first taken by a Canon T2i SLR camera, keeping its image plane approximately parallel to the cloth with different in-plane rotations for a total of ten images per cloth. As a result, there are 1,000 digital camera images in the ViTac dataset. The tactile data was collected by a GelSight sensor. As illustrated in Fig. 1a, a human holds the GelSight sensor and presses it on the cloth surface in the normal direction. As the sensor presses the cloth, a sequence of GelSight images of the cloth texture is captured, as shown in Fig. 1b. In total 96,536 GelSight images were collected. All the data is based on the shell fabric of the cloth; any hard ornaments on the clothes were precluded from appearing in the view of GelSight or digital camera. Examples of digital camera images and GelSight data are shown in Fig. 2.

[1]Computer Science and Artificial Intelligence Laboratory, Massachusetts Institute of Technology, 32 Vassar St, Cambridge, MA 02139, USA.
[2]School of Computing, University of Leeds, Leeds, Leeds LS2 9JT, UK.
[3]School of Civil Engineering, University of Leeds, Leeds LS2 9JT, UK.
[4]Department of Computer Science, University of Liverpool, Liverpool L69 3BX, UK.

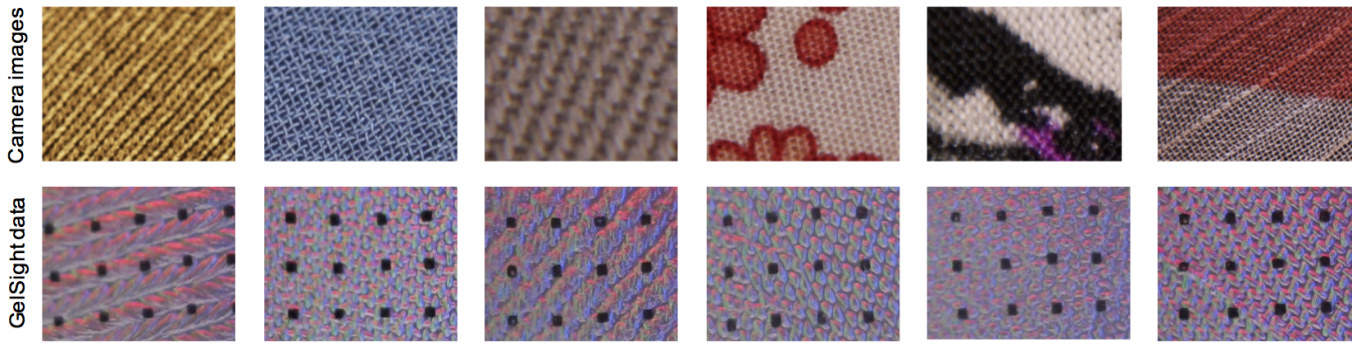


Fig. 2: Example camera images (top row) and corresponding GelSight images (bottom row) from the ViTac Cloth dataset. To make the textures visually distinguishable, the images shown here are enlarged parts of raw camera/GelSight images.

III. Methodology and Results

To match the weakly-paired vision and tactile data, Deep Maximum Covariance Analysis (DMCA) first computes representations of the two modalities by passing them through separate multiple stacked layers of a nonlinear transformation and then learns a joint latent space for two modalities by applying maximum covariance analysis such that the covariance between two representations as high as possible. We evaluate the proposed DMCA method on cloth texture recognition using the ViTac Cloth dataset.

We first perform the classic unimodal recognition task using data of each single modality. When we use the data from the GelSight sensor or digital camera for both training and test set, an accuracy of 83.4% or 85.9% can be achieved for the cloth texture recognition. This shows that the feature representations learned by deep networks enable texture recognition with either modality alone. However, especially for robotics, training data of a particular modality is not always easy to obtain: tactile data is neither commonly available nor easy to collect; also, detailed textures of objects are not always easy to access by digital cameras either. To this end, next we explore the cross-modal cloth texture recognition to train a model using one sensing modality while applying the model on data from the other modality. This is based on the assumption that visually similar textures are more likely to have similar tactile textures, and vice versa.

Perhaps surprisingly, the cross-modal cloth texture recognition performs much worse than the unimodal cases. When we evaluate the test data from GelSight sensor using the model trained on vision data, an accuracy of only 16.7% is achieved. It is even worse the other way around, only an accuracy of 14.8% is obtained. The probable reasons are factors that make the same cloth pattern appear different in the two modalities. In camera vision, scaling, rotation, translation, color variance and illumination are present. For tactile sensing, impressions of cloth patterns change due to different forces applied to the sensor while pressing. These differences mean that the learned features from one modality may not be appropriate for the other. To extract correlated features between vision and tactile sensing and preserve these features for cloth texture recognition while mitigating the differences between two modalities, we explore the proposed DMCA method to achieve a shared representation of textures for both modalities.

In the experiments, we assume that both camera and GelSight data are present during the model learning phase while only GelSight or camera data is used in the later application to new data. The setting can help us to find whether DMCA can acquire low dimensional representations that demonstrate better information embedded in the bimodal data than those learned from unimodal data. We first investigate how the cloth texture classes are classified when only GelSight data is present. The classification performance of DMCA improves as the output dimension becomes larger. As the output dimension continues to increase, the accuracy of DMCA tends to level off and can achieve a classification accuracy of around 90%. A similar performance can be observed when the cloth classes are classified using DMCA when only camera images are available. The classification performance of DMCA enjoys a dramatic increase as the output dimension increases, and then levels off when the output dimension is above 20, achieving a classification accuracy of 92.6%.

Overall, the results show that the proposed DMCA learning scheme performs well on the application of tactile-vision shared representations in either tactile or visual cloth texture recognition. This confirms that MCA is a powerful tool not only for hand-crafted features but also for features learned by deep networks. It has also been demonstrated that inclusion of the other modal data in the learning phase can improve the recognition performance when only one modality is used in the test phase. More details can be found in [4].

References

[1] M. K. Johnson, F. Cole, A. Raj, and E. H. Adelson, "Microgeometry capture using an elastomeric sensor," *ACM Trans. Graph. (TOG)*, vol. 30, no. 4, pp. 46–53, 2011.

[2] T. Ojala, M. Pietikainen, and T. Maenpaa, "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns," *IEEE Trans. Patt. Ana. Mach. Int. (T-PAMI)*, vol. 24, no. 7, pp. 971–987, 2002.

[3] R. Li and E. H. Adelson, "Sensing and recognizing surface textures using a GelSight sensor," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, pp. 1241–1247, 2013.

[4] S. Luo, W. Yuan, E. Adelson, A. G. Cohn, and R. Fuentes, "Vitac: Feature sharing between vision and tactile sensing for cloth texture recognition," *Proc. IEEE Int. Conf. Robot. Autom. (ICRA)*, 2018.