Multi-View Images Fusion Model

M M Zhdanova¹,², V V Voronin¹,², R A Sizyakin¹, M S Minkin¹ and A A Zelensky²

¹Lab. "Mathematical methods of image processing and intelligent computer vision systems" Don State Technical University, Gagarin sq. 1, Rostov-on-Don 344000, Russia
²Moscow State University of Technology “STANKIN”, Vadkovsky line 1, Moscow 127055, Russia

E-mail: mpismenskova@mail.ru

Abstract. The tasks of recognition actions and classification objects are fundamental in computer vision systems. Even subtasks, such as recognition of atomic motion and single objects form the basis for understanding the situation in the work area and the scene in general. This is especially important in video surveillance systems designed to ensure security. Thus, the effectiveness of recognition and classification methods is one of the primary tasks of computer vision. But the visual methods implemented in similar video surveillance systems, encounter some difficulties, such as inhomogeneous background, uncontrolled operating environments, irregular illumination, etc. To address these drawbacks, the paper presents a model for combining visible range images and depth images. This model allows to improve the quality of recognized images, provides the construction of a more informative descriptor, which also positively affects the recognition efficiency. Our results show that it has good performance in fusion visible image and depth map.

1. Introduction

Through the high-tech development of modern society, computer vision is implemented in such systems as security systems, monitoring and control systems, authentication systems, automation of various production processes, quality control, and inspection systems, etc. Recognition of objects and classification actions are fundamental in solving almost all of the above problems. Automated systems for recognizing objects and actions encounter a number of difficulties when implemented in real conditions, such as a different background, uncontrolled operating environments, irregular illumination, etc. In this regard, the topical task is to combine information from cameras in the visible range and depth sensors to obtain informative signs of objects located in the frame.

In a number of previously published papers [1–4], it has been shown that, the fusion of vision and depth cameras or infrared sensors data improves the accuracy of recognition. The focus is on depth cameras and inertial sensors as these two types of sensors are cost-effective, commercially available, and more significantly they both provide 3D human action data and scene structure.

In paper [1] presented thermal infrared and visible data fused for face recognition. Thermal face recognition technology has achieved high resistance to changes in lighting, and therefore, scientific and business circles are studying its application [4]. An infrared thermal image of a face was first introduced as a reliable face signature in [5], because the infrared “appearance” consists of a
temperature picture of the face and branches of a blood vein; therefore, it is considered a reliable and distinguishable physiological biometric feature [1].

In [2, 3, 6] presented techniques for human action recognition based on depth sequences. It was shown that fusing depth and vision information leads to more robust recognition.

In paper [7] presented a new method for object recognition from RGB-D data. In particular, authors focused on making recognition robust to imperfect sensor data. The proposed architecture consists of two convolutional network streams operating on color and depth information. The network automatically learns to fuse these two streams information.

As literature shows, Infrared images provide a picture of the heat of the human body or scene affected by ambient temperature, airflow conditions, exercise, illness, and more [2]. In its turn, depth sensor technology an exhibit less sensitivity to lighting changes and background clutter.

In the paper presents a model for combining visible range and depth images.

### 2. Fusion model

The most notable benefit of the joint use of depth and visible sensors is the complementary nature of different modalities that provide the depth and visible-light information of the scene. And the complementarity of information captured by the different modalities can increase the reliability and robustness of recognition or surveillance [8]. As a prototype of our model in this paper, a new image aware HVS based multi-view images fusion technique that makes use of the Parameterized Logarithmic Image Processing (PLIP) model is proposed [1]. The PLIP model [9], a generalized version of LIP model, views images in terms of their gray-tone functions and processes them using new arithmetic operators replacing the standard arithmetical operators. The gray-tone function \( \Psi \), for a given image, is generated using:

\[
\Psi = \alpha - 1,
\]

where \( \alpha \) is a model parameter. Value \( \alpha \) is the maximum intensity value in the image \( I \).

The fused image can be obtained using the following,

\[
\text{fusedImage} = \Psi_1 \oplus \Psi_2 \ominus \Psi_3,
\]

where \( \Psi_1 \) is the visible image; \( \Psi_2 \) is the depth image; \( \Psi_3 \) is the maximum-combined image obtained by selecting the maximum intensity value of depth and visible image at each pixel location; \( \oplus \) is the PLIP addition operator and \( \Psi_1 \) is the result of PLIP scalar multiplication as defined below.

\[
\Psi_1 \oplus \Psi_2 = \Psi_1 + \Psi_2 - \left( \frac{\Psi_1 \Psi_2}{\max(\Psi_1, \Psi_2)} \right),
\]

\[
\Psi_1 = (\Omega_1 \Psi_1) = \max(\Psi_1) - \max(\Psi_1) \left( 1 - \frac{\Psi_1}{\max(\Psi_1)} \right)^{\Omega_1}.
\]

The constants \( \Omega_i \) were chosen as represented in [1] that \( \sum \Omega_i = 1 \), \( \Omega_1 = 0.2989 \), \( \Omega_2 = 0.5870 \) and \( \Omega_3 = 0.1141 \).

### 3. Results and discussion

NYU-Depth V2 dataset [10] has been used for experimental protocol design and testing. This dataset consists of 1449 pairs of RGB and depth images.

The results of the depth and visible image fusion using the prosed fusion method are shown in figure 1-5.

![Figure 1. Depth-visible image fusion. (a) visible image; (b) depth image; and (c) fused image](image-url)
Successful image fusion is an important step for the increase in the accuracy of the recognition system. The fused image contains the characteristics of both visible and the depth images combining
the advantages of both the modalities. This model is planned to use in further studies for object recognition and classification of human actions.

4. Conclusions
It has been proposed the model for combining visible range images and depth images. This model allows to improve the quality of recognized images, provides the construction of a more informative descriptor, which also positively affects the recognition efficiency. The fusion approach tested on several databases and showed good results using data in visible range and depth maps.

5. Acknowledgments
This work was supported by Russian Ministry of Education and Science in accordance to the Government Decree № 218 from April 9, 2010 (project number № 074-11-2018-013 from May 31, 2018 (03.G25.31.0284)).

References
[1] Wan Q Rao S P Kaszowska A Voronin V Panetta K Taylor H A Agaian S 2018 Face description using anisotropic gradient: thermal infrared to visible face recognition Mobile Multimedia/Image Processing, Security, and Applications 10668
[2] Chen C Jafari R Kehtarnavaz N 2014 Improving human action recognition using fusion of depth camera and inertial sensors IEEE Transactions on Human-Machine Systems 45 (1) pp 51-61
[3] Chen C Jafari R Kehtarnavaz N 2015 A real-time human action recognition system using depth and inertial sensor fusion IEEE Sensors Journal 16 (3) pp 773-781
[4] Mostafa E Hamoud R Ali A et al 2013 Face recognition in low resolution thermal images Computer Vision and Image Understanding 117(12) pp 1689-1694
[5] Prokoski F J Riedel R B and Coffin J S 1992 Identification of individuals by means of facial thermography pp 120-125
[6] Yang X Zhang C Tian Y L 2012 Recognizing actions using depth motion maps-based histograms of oriented gradients Proceedings of the 20th ACM international conference on Multimedia pp 1057-1060
[7] Eitel A et al 2015 Multimodal deep learning for robust RGB-D object recognition IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS) pp 681-687
[8] Serrano-Cuerda J Fernández-Caballero A and López M T 2014 Selection of a visible-light vs. thermal infrared sensor in dynamic environments based on confidence measures J Applied Sciences 4(3) pp 331-350
[9] Panetta K Wharton E and Agaian S 2007 Parameterization of logarithmic image processing models IEEE Tran. Systems, Man, and Cybernetics, Part A: Systems and Humans
[10] Silberman N et al 2012 Indoor segmentation and support inference from rgbd images. European Conference on Computer Vision pp 746-760