

Human Object Recognition Using Colour and Depth Information from an RGB-D Kinect Sensor

Regular Paper

Benjamin John Southwell¹ and Gu Fang^{1,*}¹ School of Computing, Engineering and Mathematics, University of Western Sydney, NSW, Australia

* Corresponding author E-mail: g.fang@uws.edu.au

Received 1 Aug 2012; Accepted 7 Jan 2013

DOI: 10.5772/55717

© 2013 Southwell and Fang; licensee InTech. This is an open access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/3.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Abstract Human object recognition and tracking is important in robotics and automation. The Kinect sensor and its SDK have provided a reliable human tracking solution where a constant line of sight is maintained. However, if the human object is lost from sight during the tracking, the existing method cannot recover and resume tracking the previous object correctly. In this paper, a human recognition method is developed based on colour and depth information that is provided from any RGB-D sensor. In particular, the method firstly introduces a mask based on the depth information of the sensor to segment the shirt from the image (shirt segmentation); it then extracts the colour information of the shirt for recognition (shirt recognition). As the shirt segmentation is only based on depth information, it is light invariant compared to colour-based segmentation methods. The proposed colour recognition method introduces a confidence-based ruling method to classify matches. The proposed shirt segmentation and colour recognition method is tested using a variety of shirts with the tracked human at standstill or moving in varying lighting conditions. Experiments show that the method can recognize shirts of varying colours and patterns robustly.

Keywords RGB-D Camera, Kinect Sensor, Object Segmentation, Colour Recognition

1. Introduction

Human object recognition and tracking is an important task in robotics and automation. For example, to have a robot assist humans by carrying a load, it is important to have the robot to follow the correct person. Many methods have been developed to use computer vision for human motion capturing and tracking [1]. These methods use multiple cameras - either fixed or moving - to obtain 3D information of the moving human objects using stereo vision [2, 3]. However, these methods are either computationally too expensive or structurally too complex to be realized at a reasonable cost for everyday human-robot interaction (HRI). The release of the RGB-D sensors, such as the Microsoft Kinect and Asus Xtion, has provided a low cost access to robust, real-time human tracking solutions.

With an RGB-D sensor, such as the Kinect, the depth information is obtained from an infrared (IR) sensor

independently from the colour information. In [4], a Kinect sensor is used to obtain the spatial information of a tracked human from the depth information captured by the IR sensor in the Kinect. As the method uses only the depth information, it is colour and illumination invariant. It is computationally faster and invariant to effects such as background movement. It also overcomes the colour segmentation difficulties associated with background and object colour similarities [4]. However, recognizing or differentiating between two humans can only make use of inferred [5] spatial information, such as the tracked joint information, and therefore is difficult to achieve using depth information alone.

In vision systems, different classification methods are used to distinguish objects. These methods include shape-based and motion-based identification methods [6]. It is argued in [7] that colour has a larger effect on visual identification and, therefore, is more useful in the tracking of moving objects. As such, it is common to use colour information to distinguish and track moving objects [8]. For reliable identification and tracking, the colour recognition process is required to be illumination invariant. This can be achieved by transforming the images from the RGB pattern space to the HSV pattern space before feature extraction [9]. A feature space is commonly used to reduce the dimension of the image to a manageable size that can result in fast and reliable identification.

In [10] a colour recognition method is developed in the RGB colour space to identify artificial images, such as flags and stamps. The method compares extracted feature vectors from images of stamps and flags against a feature vector database of known flags and stamps. The image is classified as matching the image in the database that has a minimum Euclidean distance between the two feature vectors. However, it is not always possible to have a database of all identifiable objects. This is especially the case in applications such as human identification and tracking, where the human object may not be known beforehand. In addition, the RGB colour space is not suitable for evaluating the similarity of two colours [11]. To deal with this problem, the HSV colour space is used in [12-14] to achieve partial illumination invariance in colour recognition.

In general, the Euclidean distance-based classification methods - such as the one used in [10] - do not fully utilize all the descriptive information in the feature vector. Therefore, it is susceptible to false positives (wrongly classified as a match when it is not), especially with images in an unknown environment with varying lighting conditions. Other classification methods, such as fuzzy logic [15, 16], cluster analysis [17], statistical pattern analysis [17], normalization [18] and neural networks, [19] can provide more robust recognition results.

In this paper, a human object identification method is introduced based on information gathered from an RGB-D sensor, such as the Kinect. This method is to be used for a human assistive robot in following a human after seeing the human and identifying the colour of the shirt the person is wearing. This method includes two phases: firstly, the method introduces a mask based on the depth information of the sensor to segment the shirt from the image (shirt segmentation); secondly, it extracts the colour information of the shirt for recognition and tracking (shirt recognition).

The main contribution of the paper is the introduction of a shirt mask and the confidence-based ruling method to classify matches. This method is more reliable than the conventional Euclidean distance measure used in [10].

The paper contains 5 sections. In Section 2, the depth-based shirt segmentation method is introduced. In Section 3, the confidence-based shirt colour recognition method is introduced. Experimental results of the methods introduced are presented in Section 4. Conclusions are given in Section 5.

2. Shirt Segmentation

The proposed shirt segmentation method involves constructing a number of masks and applying these masks to the depth and RGB images. The final mask - the '*shirt mask*' - is applied to the RGB image to segment the colour information of the shirt of the tracked human. This method makes use of the '*player index*' data that is optionally encapsulated by the Kinect sensor within the depth image [20].

The first mask constructed is the '*depth mask*' and is constructed by iterating through the depth image from the Kinect and evaluating the '*player index*' value of each pixel; if the '*player index*' is non-zero (meaning that it belongs to a tracked human object), the corresponding pixel in the '*depth mask*' is set to 255 (white), otherwise it is set to 0 (black). This mask will result in a black-white image with human objects displayed as white pixels and everything else as black.

The second mask is the '*shirt polygon mask*' and is constructed entirely from the joint spatial data streamed from the tracking algorithm of the Kinect. While the first mask (the depth mask) generalizes all the player pixels by not discriminating between non-zero '*player index*' values, the '*shirt polygon mask*' will apply a '*player*' specific mask. The user can specify which '*player*' is the person of interest and the mask will be applied only to that specified person. The shape of the shirt polygon is depicted in Figure 1. The shirt polygon vertex locations in the image plane uv are obtained from the joint locations returned by the Kinect's tracking algorithm and are then transformed into the RGB image plane co-ordinates uv .

The shirt polygon and the corresponding joints from which the polygon is constructed are shown in Figure 1 and Table 1. The 'shirt polygon mask' is constructed by assigning a value 255 (white) to all pixels inside the shirt polygon (foreground) and a value of 0 (black) to all pixels outside the polygon (the background). This mask will result in a black-white image with a shirt polygon of the specified human object displayed as white pixels and everything else as black. As this method is developed by assuming that the specified person is willing to be tracked, during the 'shirt segmentation' process the person will be generally facing the sensor in an orientation where all 8 joints are visible.

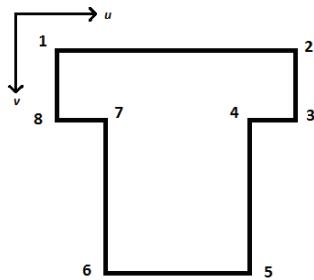


Figure 1. The Shirt Polygon Vertices

| Vertex | <i>u</i> | <i>v</i> |
|--------|----------------|-----------------|
| 1 | Right Elbow | Shoulder Centre |
| 2 | Left Elbow | Shoulder Centre |
| 3 | Left Elbow | Left Shoulder |
| 4 | Left Shoulder | Left Shoulder |
| 5 | Left Shoulder | Left Hip |
| 6 | Right Shoulder | Left Hip |
| 7 | Right Shoulder | Right Shoulder |
| 8 | Right Elbow | Right Shoulder |

Table 1. Shirt Polygon Mask Vertices Points

The final mask, the 'shirt mask', is obtained by the Bitwise AND operation of the 'depth mask' and the 'shirt polygon mask'. The 'shirt mask' is then applied to the RGB image to segment the shirt from the original image for further process. In Figure 2, the original RGB image, a 'depth mask', a 'shirt polygon mask', a 'shirt mask' and the resultant segmented shirt are shown.

3. Shirt Recognition

The shirt recognition method is performed using the colour information of the segmented shirt. Firstly, the method transforms the segmented shirt from its RGB colour space representation into an HSV colour space representation to enable the method to handle varying lighting conditions by partially decoupling the chromatic and achromatic information [11]. Secondly, in the HSV colour space a feature vector is constructed for different colours. Using the feature vector, a confidence measure is then introduced to recognize the shirt. The details of these steps are described as follows.

3.1 Colour space conversion

As the colour picture obtained from the Kinect sensor is presented in the RGB colour space. This colour space is known to have shortfalls in the colour recognition of different objects [11]. A colour space that is more robust to different lighting conditions is the HSV colour space. In this paper, once the shirt is segmented from the image, the RGB image of the shirt is converted into the HSV colour space using the OpenCV transformation [21].

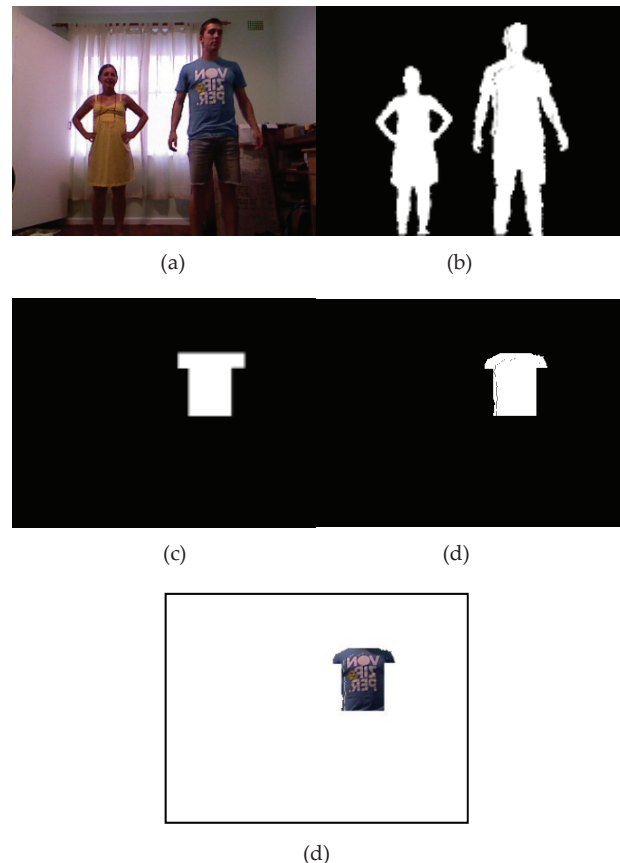


Figure 2. (a) The Original RGB Image (b) A 'Depth Mask' (c) A 'Shirt Polygon Mask'. (d) A 'Shirt Mask'. (e) The Segmented Shirt

3.2 Feature vector

A feature vector is used to measure the location of a particular feature in a feature space that is constructed to reduce the dimensionality of the information to simplify the classification process [22]. The dimensionality of the feature space should be reduced to the lowest size practical while still retaining enough information to achieve robust recognition [22]. In this paper, a feature space derived from the HSV colour space that consists of nine dimensions is introduced. The nine dimensions used are to represent the primary (red, green and blue), secondary (yellow, cyan and magenta) and achromatic (white, black and grey) colours. The tertiary colours are excluded in the feature space, as it is found that the nine dimensional feature space provides enough resolution to

achieve differentiation between non-similar colours while not being so high that the feature space would become sensitive to varying lighting conditions.

The HSV values that correspond to these nine dimensions are presented in Table 2. These values were chosen so that the primary and secondary colours' hue would run directly through the centre of their respective regions. The achromatic colour boundaries were selected by evaluating the HSV of images of achromatic objects. A colour feature vector is introduced as a measure of the ratio of pixels belonging to each of the nine dimensions in the colour space. The feature vector \vec{V} can be expressed by Equation 1.

| Region No. | Region Label | H | S | V |
|------------|--------------|-------------|---------|------------|
| 1 | Red | <22 & >234 | >50 | >39 |
| 2 | Yellow | <64 & >21 | >50 | >39 |
| 3 | Green | <107 & >63 | >50 | >39 |
| 4 | Cyan | <150 & >106 | >50 | >39 |
| 5 | Blue | <192 & >149 | >50 | >39 |
| 6 | Magenta | <235 & >191 | >50 | >39 |
| 7 | White | 0 - 255 | <51 | >180 |
| 8 | Black | 0 - 255 | 0 - 255 | <40 |
| 9 | Grey | 0 - 255 | <51 | >39 & <181 |

Table 2. Proposed HSV Regions for Feature Extraction

$$\vec{V} = [V_{Red} \ V_{Yellow} \ V_{Green} \ V_{Cyan} \ V_{Blue} \ V_{Magenta} \ V_{White} \ V_{Black} \ V_{Grey}] \quad (1)$$

In the feature vector, the i^{th} element in the vector, $\vec{V}[i]$ (where $i = 1, 2, \dots, 9$, corresponds to the colours *Red*, *Yellow*, *Green*, *Cyan*, *Blue*, *Magenta*, *White*, *Black*, *Grey*, respectively), is the ratio of pixels in the shirt image obtained from Section 2 classified to the i^{th} dimension in the HSV colour space, as presented in Table 2. Therefore, all of the feature vectors have a bounding condition, as expressed below:

$$\sum_{i=1}^9 \vec{V}[i] = 1 \quad (2)$$

3.3 Confidence ruling method

Once the feature vector is obtained for an object, object recognition can be achieved by comparing the feature vector of the observed object with a previously identified object's feature vector. Conventional recognition methods are commonly performed by using the Euclidean distance between the two feature vectors [10]. This approach is prone to false classification, as the Euclidean distance is a summative measure of differences in all dimensions in the feature space. In this paper, a confidence measure is introduced to more precisely classify the colour features.

The confidence ruling method proposed in this paper analyses the distribution of colours in the 9 dimensions of the extracted feature vector. To compare two feature vectors \vec{V}_1 and \vec{V}_2 , where \vec{V}_1 is the stored feature vector

that represents a pre-identified object to be recognized later, \vec{V}_2 is the feature vector of an object being considered as a potential match to the pre-identified object. It is important that this allocation is kept the same as the numerical confidence equation proposed does not result in equal confidence ratings when the comparison is reversed, except when the feature vectors are exactly equal. The 'error sum', E_{sum} is the sum of the absolute differences between the two vectors corresponding elements. It is defined in Equation 3:

$$E_{sum} = \sum_{i=1}^9 |\Delta \vec{V}[i]| \quad (3)$$

where:

$$\Delta \vec{V} = [(V_{Red_2} - V_{Red_1}) \ (V_{Yellow_2} - V_{Yellow_1}) \ \dots \ (V_{Grey_2} - V_{Grey_1})] \quad (4)$$

By definition of the feature space (Equation 1) and the bounding condition (Equation 2), the 'error sum' has a range of 0 to 2, inclusive. The confidence measure of \vec{V}_2 matching \vec{V}_1 , Co_{2-1} , is proposed to be:

$$Co_{2-1} = Co_{nondom} + W \times Co_{dom} \quad (5)$$

where Co_{nondom} is the non-dominant colour similarity contribution, Co_{dom} is the dominant colour similarity contribution to confidence and W is a weighting factor selected to suit the application. The weighting factor W allows the designer to alter how the numerical confidence ruling is reliant on dominant and non-dominant colours. In particular, Co_{nondom} and Co_{dom} are calculated as:

$$Co_{nondom} = 2 - E_{sum} \quad (6)$$

$$Co_{dom} = 2 \sum_{i=1}^9 \vec{V}_2[i] (1 - |\Delta \vec{V}[i]|) \quad (7)$$

Clearly, Co_{nondom} has a significant contribution towards the confidence of a match when the scalar differences between the two feature vectors' corresponding elements is small. To make the Co_{dom} component have the same numerical range as that of Co_{nondom} , a multiplication factor of 2 is introduced in Equation 7. This will allow the weighting factor W to be a true weighting between the dominant and non-dominant colours.

The Co_{dom} component is included in the confidence calculation to provide more robust matching and to reduce the possibility of false positive matches. Co_{dom} is calculated based on not only the differences between each element of the two feature vectors but also the colour dominance of that element in the entire feature vector. The purpose of including $\vec{V}_2[i]$ in the Co_{dom} calculation is to add the relevance of each element in the feature vector to the confidence contribution. For example, even if the scalar error between the two vectors' elements is small and the value of $(1 - |\Delta \vec{V}[i]|)$ approaches 1, if there is little or no presence of that colour element ($\vec{V}_2[i]$) in the

feature vector, then the contribution of the error between those elements will become insignificant as it is not relevant.

The confidence rating, Co_{2-1} has a numerical range of:

$$0 \leq Co_{2-1} \leq 2 + 2W \quad (8)$$

3.4 Confidence threshold

After the confidence measure has been calculated using Equation 5, shirt recognition can be performed by using a thresholding function as expressed in the following equation:

$$\text{Match} = \begin{cases} 1 & \text{if } Co_{2-1} > T_m \\ 0 & \text{otherwise} \end{cases} \quad (9)$$

where T_m is the confidence threshold that needs to be obtained for reliable shirt recognition. To obtain T_m , firstly the weighting factor W in Equation 5 needs to be chosen. In this paper, W was chosen to be 2 (giving the confidence range to be 0 to 6). This weighting factor value is more reliant on dominant colours than non-dominant colours. To obtain the T_m for differentiating a match from a non-match, the numerical analysis of sample extracted feature vectors is performed. In this paper, it is done by obtaining 100 feature vectors for each of 8 different shirts in two lighting conditions of approximately 50 and 35 lux indoors with incandescent lighting. Multiple feature vectors for each shirt are obtained and averaged so that a more reliable threshold can be established. All the shirts used are plain and in the following colours: red, green, blue, cyan, magenta, yellow, black and white, as shown in Figure 3.



Figure 3. The shirts and lightings used to obtain the confidence ruling threshold.

Using the averaged feature vectors for the shirts, the confidence was calculated between each shirt and every other shirt using equation 5. From a numerical analysis of the sample confidence ratings, it was found that a confidence threshold of 5 will provide the best results with minimal false positives. The value of 5 for the confidence threshold was chosen to allow the method to recognize the same shirt in different lighting conditions while avoiding false positive matches.

4. Experimental results

To test the effectiveness of the proposed shirt segmentation and recognition method, various

experiments were conducted under different lighting conditions and with different coloured shirts.

4.1 Shirt segmentation.

The shirt segmentation method was tested with a person wearing a white shirt in front of a black backdrop shown in Figure 4.

The clearly contrasted colour setting in this experiment is necessary as it is easier to analyse the shirt and the background using existing thresholding methods. One thousand (1000) images were taken, from which shirts were segmented from each image. These 1000 shirt

images were then converted to greyscale images and the accumulated histogram of these greyscaled images was calculated. The histogram was analysed to determine what clusters belonged to the background, shadows on the shirt and fully illuminated shirt. From analysis of the histogram, a greyscale threshold was obtained so that the shadows on the shirt and the fully illuminated pixels could be counted as good (belonging to the shirt) while the background pixels could be counted as bad during the experiments.



Figure 4. The shirt segmentation experiment setup

Three experiments were performed to quantify the accuracy of shirt segmentation with the tracked human moving at different speeds. In each of the three test cases, 1000 images were taken and the accuracies are calculated by averaging over the 1000 segmented shirts. The first

experiment involved the tracked human standing still. In this case, 99.49% of the segmented pixels (shirt mask) belonged to the shirt. The second experiment involved the human moving at a slow walking speed (about 2 km/h). This resulted in 98.30% of the segmented pixels (shirt mask) belonging to the shirt. The third experiment involved the human moving as fast as possible (at a fast pace of around 6km/h) while staying in the field of view. In this case, 94.34% of the segmented pixels (shirt mask) belonged to the shirt.

4.2 Shirt recognition

To test the robustness of the selected confidence threshold T_m in Equation 9, the shirt recognition method was tested in an environment with different lighting conditions to which the confidence threshold was obtained. The environment for the shirt recognition experiment was indoors with fluorescent lighting (Figure 5). 8 new shirts different from the ones used for the threshold calibration were used in this experiment. The new shirts were black with white stripes, bright red with a white logo, aqua with a blue pattern, bright orange with a grey pattern, bright pink with a grey pattern, green with white highlights, blue with yellow highlights and purple with a pink logo (shown in Figure 5). These shirts provided a wide coverage of the colour gamut and possible geometric patterns.



Figure 5. The shirts and lighting used for the recognition experiment

Feature vectors were extracted for each shirt in two different lighting conditions, this time at 421 and 85 lux

(due to the lux sensor being more sensitive to fluorescent light sources). The confidence between each shirt and

every other shirt was calculated and thresholded using $T_m = 5$ to determine whether they matched or not. From the 8 shirts in two different lighting conditions, there are 16 different feature vectors. For the purpose of the experiment it was meaningless to compare the extracted feature vector against itself as the result is always a 100% confidence rating; therefore, there were only 240 confidence comparisons. Correct matches occur between the feature vectors representing the same shirt but in different lighting conditions. Out of the 16 possible correct positive matches, the method achieved all 16 with the lowest confidence being 5.229. However there were 4 false positive matches (1.67%) occurring between the bright orange and bright red shirts in the bright lighting conditions (confidence ranging from 5.059 to 5.098). These false positive matches occurred due to the fact that under the bright light, these two shirts have very similar colour features. It is also due to the fact that the proposed feature space only contains 9 dimensions. This accuracy of 98.33% is also higher than that of 97% in [10] for artificial images.

5. Conclusions

In this paper, a method is introduced to use a RGB-D sensor - the Microsoft Kinect - to achieve human object recognition and tracking using the depth and colour information of the shirt a person is wearing. Firstly, the method introduces a mask based on the depth information of the sensor to segment the shirt from the image (shirt segmentation). It then extracts the colour information of the shirt for recognition and tracking (shirt recognition). It is shown that the shirt segmentation method proved to be very reliable, even when the human object is moving at relatively high speed. The experimental results also show that the shirt recognition method is mostly reliable (with above 98% reliable identifications). The shirt recognition method handled varying colours and patterns in varying lighting conditions robustly. This indicates its suitability for real world applications.

6. Acknowledgements

This work is partially supported by the Australian Research Council under project ID LP0991108 and the Lincoln Electric Company Australia.

7. References

- [1] Moeslund TB, Hilton A, Kruger V, (2006) A Survey of Advances in Vision-based Human Motion Capture and Analysis. *Computer Vision and Image Understanding* 104, 90-126.
- [2] Yamazoe H, Utsumi A, Tetsutani N, and Yachida M, (2007) Vision-Based Human Motion Tracking Using Head-Mounted Cameras and Fixed Cameras *Electronics and Communications in Japan, Part 2*, Vol. 90, No. 2, 14-26
- [3] Nga KC, Ishiguro H, Trivedi M, Sogod T, (2004) An integrated surveillance system—human tracking and view synthesis using multiple omni-directional vision sensors. *Image and Vision Computing* 22, 551-561
- [4] Shotton J, et al., (2011) Real Time Human Pose Recognition in Parts from Single Depth Images. *Computer Vision and Pattern Recognition (CVPR)*, 2011 IEEE Conference on. 1297-1304.
- [5] Acero A, (2011) Audio and Video Research in Kinect. Microsoft Research. <http://msrvideo.vo.msecnd.net/rmcvideos/154757/dl/154757.pdf>. Accessed 2012 July 23.
- [6] Hu W, Tan T, Wang L, Maybank S, (2004) A survey on visual surveillance of object motion and behaviors. *IEEE Transactions on Systems, Man, and Cybernetics, Part C: Applications and Reviews*, Volume: 34, Issue: 3, 334-352.
- [7] Vidaković V, Zdravković S, (2009) Color influences identification of the moving objects more than shape. *Psihologija*, Vol. 42, No. 1, 79-93.
- [8] Su F, Fang G (2012) Moving Object Tracking Using an Adaptive Colour Filter. *The 12th International Conference on Control, Automation, Robotics and Vision, (ICARCV 2012)*, Guangzhou, China 5~7 December, 2012, pp. 1048-1052.
- [9] Bow S, (1984) *Pattern Recognition – Applications to Large Data-Set Problems*. Marcel Dekker Inc. New York.
- [10] Stachowicz MS, Lemke D, (2000) Colour Recognition, 22nd Int. Conf. Information Technology Interfaces (ITI), 329-334.
- [11] Cheng HD, Jiang XH, Sun Y, Wang J, (2001) Color image segmentation: advances and prospects. *Pattern Recognition*, Vol. 34, 2259-2281.
- [12] Feng W, Gao S, (2010) A vehicle license plate recognition algorithm in night based on HSV. 3rd International Conference on Advanced Computer Theory and Engineering (ICACTE), Vol. 4, 53-56.
- [13] Flores P, Guilenand L, Prieto O, (2011) Approach of RSOR Algorithm Using HSV Color Model for Nude Detection in Digital Images, *Computer and Information Science* Vol. 4, No. 4, 29-45.
- [14] Patil NK, Yadahalli RM, Pujari J, (2011) Comparison between HSV and YCbCr Color Model Color-Texture based Classification of the Food Grains. *International Journal of Computer Applications*, Vol. 34, No. 4, 51-57.
- [15] Pedrycz W, (1997) Fuzzy sets in pattern recognition: Accomplishments and challenges. *Fuzzy Sets and Systems*, Vol. 90, 171-176.
- [16] Mitchell HB, (2005) Pattern recognition using type-II fuzzy sets. *Information Sciences*, Vol. 170, Iss. 2-4, 409-418.
- [17] Chen CH, Pau LF, Wang PS, (1999) *Handbook of Pattern Recognition and Computer Vision*. 2nd Edition, World Scientific Publishing Co, Singapore.

- [18] Kukuchi T, Umeda K, Ueda R, Jitsukawa Y, Osuymi H, Arai T, (2006) Improvement of Colour Recognition Using Coloured Objects. In: A. Bredenfeld et al. (Eds.). Lecture Notes in Computer Science, Volume 4020, Springer-Verlag Berlin Heidelberg, 537-544.
- [19] Pandya AS, Macy RB, (1995) Pattern Recognition with Neural Networks in C++. CRC Press, Boca Raton – Florida.
- [20] Kinect Sensor, Microsoft Developer Network (MSDN), <http://msdn.microsoft.com/en-us/library/hh438998.aspx>, accessed 2012 July 24.
- [21] OpenCV 2.1 C++ Reference, http://opencv.willowgarage.com/documentation/cpp/imgproc_miscellaneous_image_transformations.html#miscellaneous-image-transformations, accessed 2012 July 24.
- [22] Forsyth DA, Ponce J, (2011) Computer Vision – A Modern Approach, 2nd Edition, Pearson Prentice Hall, Upper Saddle River, New Jersey.