Localization of Essential Door Features for Mobile Manipulation

Maciej Stefańczyk and Michał Walęcki

Warsaw University of Technology, Institute of Control and Computation Engineering, Nowowiejska 15/19, 00-665 Warsaw, Poland
{M.Stefanczyk,M.Walecki}@elka.pw.edu.pl

Abstract. In this paper, we present a control system of a robot actively detecting and labeling door features, such as handle, lock or door plate. System is described using agent-based approach, with clear division into control, effector and receptor subsystems, with most important parts deeply explained. Presented approach supplements gap between door detection and opening algorithms, giving a robot the ability to take more actions, like recognizing room number or opening door lock.

Keywords: active vision, door identification, key hole, door handle, fuzzy inference.

1 Introduction

1.1 Motivation

Indoor spaces are natural environments for service robotics. Although there are applications in which a robot operates only in a single room, vast majority of them requires traversing through at least a part of building, where the solution of problem of detecting doors, operating them and passing through is crucial. In this paper we present a method of recognizing essential door features, such as accurate dimensions, opening direction as well as candidate regions for detecting door parts (such as a handle, key hole, text plates) with additional hints.

1.2 Related Work

Door Localization. As a crucial part in many applications, many approaches to door localization exist, used either in mobile robotics or blind people assistant applications. Solutions based on ultrasonic sensors [1] or a pointcloud analysis [2, 3] are of no use in our application due to limited usable range of sonars and 3D sensor used (Kinect). Thus the only reliable data source is color camera.

Some researchers focus on a limited subset of door localization problem, where a camera is mounted rather low, on mobile base [4–6]. This limitation renders their algorithms unusable in the case where lower door part is occluded by some furniture elements. Working only with color information is another limiting factor [7]. Algorithms based on color distribution histograms are prone to lighting

R. Szewczyk, C. Zieliński, and M. Kaliczyńska (eds.), *Recent Advances in Automation, Robotics and Measuring Techniques*, Advances in Intelligent Systems and Computing 267,
DOI: 10.1007/978-3-319-05353-0_46,

condition changes and hard to adapt in new environments, needing additional learning step for every new kind of object.

On the other hand, using only edge information [8] may lead to many false positives, caused by shelves or other rectangular objects in robot's field of view. Complicated algorithms give better results using advanced line segments analysis [9], but it takes additional processing time. Those observations lead to a conclusion, that using many weak classifiers gives better results and takes less time, due to avoiding complicated processing. That's why approaches using for example AdaBoost [5, 6] emerged, integrating simple features like parallel door frame lines, untextured kickplate, gap under door, proper proportions and size, color different from surrounding wall etc.

Handle Recognition. All algorithms presented above return estimated door position, that can be used as a goal for a robot to approach near them and take some action, with opening being the most popular. Trajectory generation for door opening is a problem itself [7, 10], and is out of scope of this article. Localization of door handle, however, is a crucial part in door opening. In this task, as robot is close to the door, both RGB and 3D information can be used. Some researchers focus on lidar scans only [3], looking for specifically for handle in dense pointcloud, rejecting all points except fixed height range. Others [7] use color images only, with some strong assumptions about handle appearance (must contain linear segment perpendicular to door frame and close to it). Eventually, mixed approach is also used [11], in which, after initial detection in RGB image, depth data is used to filter out false positives.

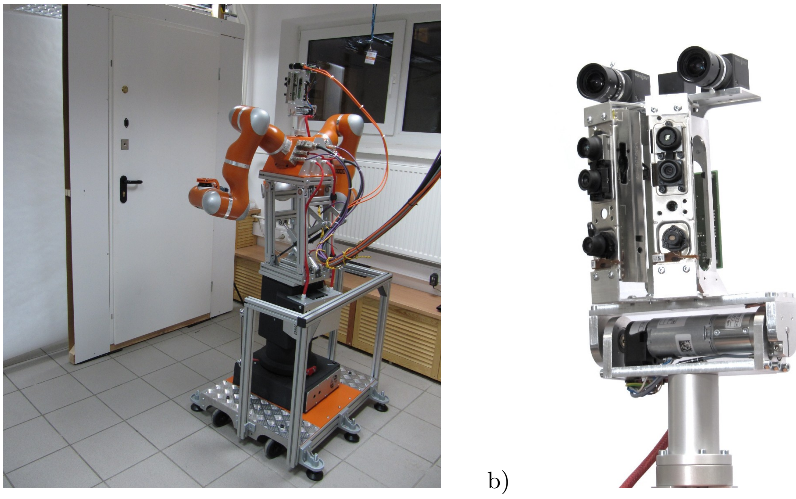
Problem Formulation. Previously presented algorithms solve the problems of locating doors and opening them. However – there is still a gap between those two steps. How to check, whether the door the robot is facing is this the robot wants to open and cross? How to open locked doors? The first problem can be solved by using additional information from map or localization and recognition of door plates with text on them. There are approaches to this problem [8] working successfully. The problem of locked doors is much harder to deal with – there is a wide spectrum of different door locks available with different kind of keys to handle. On doors and around them there are many features that can help robot choose a proper action (whether to open this door or find another, how to open it etc.). Door plates, handles, locks, additional devices – they all carry some useful information.

1.3 Paper Structure

This article is organized as follows. Following section describes robotic system Velma used in our experiments, focusing mainly on the robot's head and its sensory system, which is formally described in section 3. The same section outlines the algorithm used in our works. Experiments results, preceded by experimental system setup, are presented in section 4. The paper ends with summary and an outline of future works.

2 Two-Arm Robotic System Velma

Velma (fig. 1a) is a two-arm robotic system, developed in the Institute of Control and Computation Engineering, Warsaw University of Technology in 2013. It consists of two Kuka LWR-4+ manipulators with BarrettHand grippers, mounted on a 2-dof torso and a head equipped with vision sensors. It is a research platform for the field of service robotics, intended to operate in environment adapted to human needs. Velma is 180cm high and has human-like proportions which allow it to perform typical activities of human, such as opening of doors and kitchen cabinets, unscrewing jars and other manipulation tasks. Kuka LWR-4+ manipulators, combined with custom made torso, form a powerful system capable of manipulation with impedance control [10], useful in complicated, unstable environment and essential when performing tasks with possible contact with human.



a) b)

Fig. 1. Velma robot and a close-up of its head

2.1 Velma's Robotic Head

The head of Velma robot (fig. 1b) itself is a complex robotic device. It is attached to Velma's torso with a fast pan-tilt unit [12]. It gives the head the range of motion $-90°$ to $+90°$ in both horizontal and vertical direction respective to "look ahead" position. Applied drives enable the head to move with speed over $180°/s$, obtaining position measurements with accuracy better than $0.1°$. Main goal of employing a moving head is to extend the sensors' range. Standard cameras mounted in robot's mechatronic head can cover the whole manipulation space and actively explore entire visual field for the information that is essential to the ongoing task. In this particular case head motion is used to scan the whole door while standing next to it.

2.2 Visual Sensory System

Visual information from the environment is gathered using a set of 4 sensors calibrated with each other. On the top of the head there are two PointGrey BlackFly cameras (color, $1280px \times 1024px$), mounted with $9cm$ baseline. Each camera has $8mm$ lens giving $50° \times 38°$ field of view. Beneath them there are two Microsoft Kinect sensors mounted vertically, with 43° angle between them. This setup gives combined field of view of $86° \times 57°$ for both color and depth images from Kinect. Another advantage of depth sensors mounted this way is that their infrared projectors don't interfere with each other giving clean measurements from both units.

3 Robot Control System Specification

In the control system design and implementation an *agent* approach was utilized [13]. According to it, the whole robot with its control system is considered an embodied agent, which is defined as a device or program that has the ability to perceive its surroundings for subsequently influencing the environment state, can communicate with other agents and has an internal imperative to achieve its goal. The agent's goal in this case is to identify and localize features of the door that are essential from the point of view of further task of door opening. Fig. 2 shows internal structure of a generic agent, distinguishing five subsystems. Control subsystem c represents a set of high level algorithms that pilot the task. It communicates with virtual effectors e_i and virtual receptors r_i that form an abstraction layer for real hardware. Virtual effectors and virtual receptors provide convenient software interface to corresponding real effectors E_i and real receptors R_i, which represent agent's corporeal body. Control subsystem, virtual effectors and virtual receptors form a control system that runs on a PC. Each of subsystem owns a set of behaviors, which consist of transfer functions, start condition and terminal condition.

There is a variety of effectors which represent Velma's manipulators and drives of the torso, as well as a variety of receptors, including all the sensors mounted on Velma's head. However, most of them do not play any role in this work and will not be considered.

3.1 Real Effector and Virtual Effector

Effector E_1 represents the pan-tilt unit, moving the robot's head. Its joints are driven with electrical motors powered by custom motor controllers, specially designed for robotic research applications [14], that communicate with PC.

During the task of door recognition only one of behaviours of pan-tilt unit virtual effector e_1 is utilized. The operation of transfer functions corresponding to it proceeds as follows.

Virtual effector e_1 internal memory ${}^{e}e_{dr,h}$ contains a vector of last measured motors positions m_c and a scheduled reference trajectory, written as a vector of

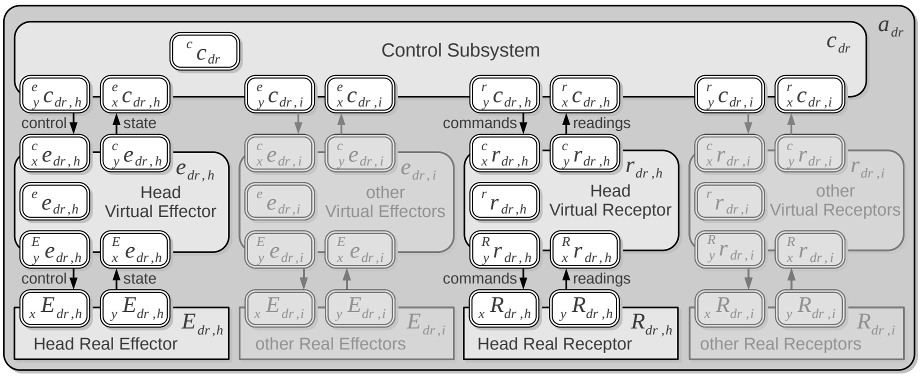


Fig. 2. General agent structure

triplets of joint positions q_d, speeds $\dot{q}_d$ and accelerations $\ddot{q}_d$ to be executed in every subsequent control cycle. Whenever a new target joint position q_t is set by control subsystem through virtual effector's communication buffer ${}^{c}_{x}e_{dr,h}$, a new reference trajectory is computed, beginning from the first scheduled triplet to be executed $q_{d,\iota+1}, \dot{q}_{d,\iota+1}, \ddot{q}_{d,\iota+1}$ and leading to q_d joint position with zero speed and acceleration. When virtual effector receives new readings of motor positions m_c from motor controllers through ${}^{E}_{x}e_{dr,h}$, their values are stored in e_1's internal memory.

Generation of control data for motor controllers is triggered by external *clk* signal. It's period of $10ms$ determines the control cycle duration. In each control cycle a subsequent position $q_{d,\iota+1}$ of joint reference trajectory and current motors position are read from memory and on this basis desired motors positions and speed limits for the next cycle are computed. Together with desired mode of operation they are sent to motor controllers via ${}^{E}_{y}e_{dr,h}$ communication buffer. During the task of door recognition both motor controllers operate in the mode of position regulation with configurable speed limit.

Triggered with the same signal, every control cycle the latest values of motor positions are converted to joint position values and sent to control subsystem through ${}^{c}_{y}e_{dr,h}$ communication buffer.

3.2 Receptor

Gathering Photos. Similarly to effector, receptors $R_{dr,i}$ represent image sensors mounted on robot's head and virtual receptor $r_{dr,r}$ provides a software interface to them, aggregating data and processing it. Although in our system we have more than one real receptor (presented in sec. 2.2), in presented task only one camera is used. To create detailed door image, the robot must be very close to it, which results in acquiring the whole door width at once, but a only limited fraction of its height. Virtual receptor $r_{dr,r}$ has to create a single image from few gathered, so it has different behaviors. One is responsible for gathering

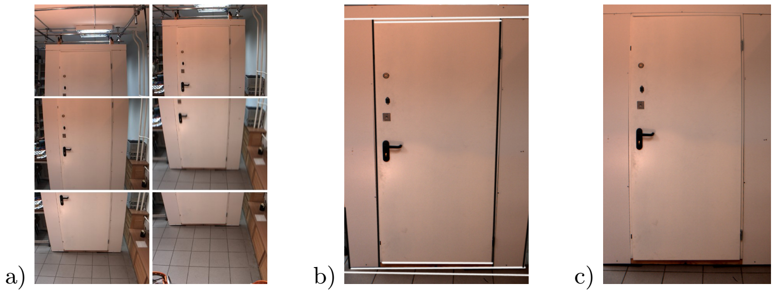


Fig. 3. Input photos (a), stitched, distorted image (b) and corrected image (c)

a series of photos for different head positions – after each trigger from control subsystem (generated after setting a new head position, in 10° steps down from straight-ahead position) a new photo is gathered and stored in internal buffer ${}^{r}r_{dr,h}$ (fig. 3a).

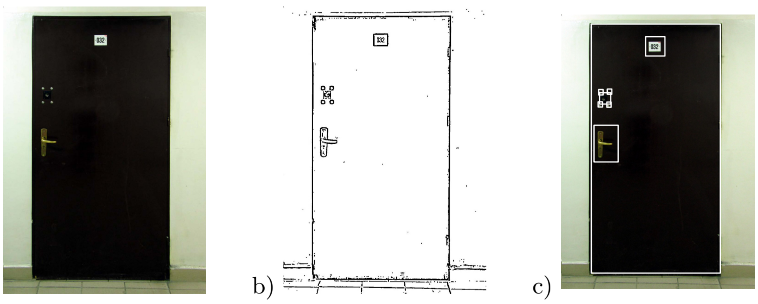
After obtaining all 6 photos, the next behavior is selected, which is responsible for stitching them to create full door view. In general, this task is rather hard to do, but as our scene is mostly flat (doors are only slightly recessed into the wall) it can be done with good results using built-in OpenCV functionality. To retain straight lines in resulting image, remapping must be done using planar warper. Resulting image is almost always perspective distorted (fig. 3b), so, in next step, this distortion is estimated and corrected. Hough transform is used to detect lines in image, which are then analyzed and grouped into horizontal (less than 30°) vertical (greater than 60°) and other (presented on fig. 3b as white, black and gray, respectively). Horizontal and vertical lines are used to estimate vanishing points in image, according to following steps:

1. pick two random lines
2. if angle between them is less than threshold - go to (1)
3. calculate and remember their intersection
4. if there is less than 20 intersection points calculated - go to (1)
5. calculate mean of all points
6. if all points lay in ϵ neighborhood of calculated mean - select it as vanishing point and finish
7. otherwise remove furthest point and go to (5)

Vanishing points are then used to generate remapping from input, distorted image to straight one (fig. 3c).

Regions of Interest Segmentation. Having a straight, complete door image, the next behavior of virtual receptor is responsible for segmenting the door itself and all regions of interest on them and around its frame. Both steps use binary image created by adaptive segmentation and some morphological operations (to remove noise). It makes the algorithm insensitive to lighting conditions, door color and texture (partly), and only slight shadow or deviation in intensity is needed to properly detect edges of door frame or other features (fig. 4b). Line segments, extracted from binary image, are again grouped in horizontal and vertical lines, this time allowed deviation is only 3°, as image is rectified and door frame should be parallel to x and y axis. Door frame segmentation proceeds as follows:

1. pick the longest vertical line
2. find horizontal lines ending close to the vertical line's upper end and pick the longest one
3. if there is no such line - remove the first line and start over
4. find vertical lines close to the other end of a lintel and pick the longest one
5. if there is no such line - remove lintel line and go to (2)
6. find horizontal line connecting both vertical lines bottom ends



a) b) c)

Fig. 4. Input image (a), thresholding (b) and final segmentation (c)

The obtained closed loop is considered as a door frame and acts as a reference border for the next algorithm stages. From all other edges, those with sufficient size and line width are selected and act as regions of interest passed to labeling step. Selected ROIs are presented on fig. 4c.

Regions of Interest Labeling. The regions detected in the previous step can be of many different kinds – door locks, handles, light switches, door plates etc. To make use of it and pick the best algorithm to detect its type, some initial guess is needed about its purpose. In our system, we defined classes for said objects, supplemented with unknown objects. A thing, that is important in our approach is that we don't limit any region to only one class, giving them confidence score for any existing possibility. As different objects can be found in different locations and there are no simple rules to distinguish between them, fuzzy rules were prepared to describe every class, based on few, basic properties – location in a door plane (horizontal and vertical), size and shape regularity. Corresponding fuzzy sets are presented on fig. 5.

For object's horizontal position we have two possibilities – `Edge` and `Center`, which are based on a normalized distance from the door frame. Similarly, vertical

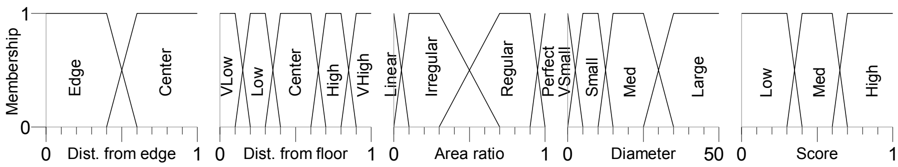


Fig. 5. Fuzzy sets for linguistic variables in presented system

position is based on normalized distance from the floor and possible values are `VeryLow`, `Low`, `Center`, `High` and `VeryHigh`. Shape regularity is computed using ellipse and rectangle circumscribed on segment according to given formula:

$$Ratio = \max\left\{\frac{A_S}{A_R}, \frac{A_S}{A_C}\right\}$$

where A_S, A_R and A_C is area of segment, circumscribed rectangle and ellipse accordingly. A segment can have `Perfect` shape if it's elliptical or rectangular, `Regular`, `Irregular` or `Linear` shape (when it's composed from linear parts only). Size could be `VerySmall`, `Small`, `Medium` and `Large`, depending of diameter of circumscribed circle. All output variables (each class scores) have similar distribution presented as Score on fig. 5, and could be `Low`, `Medium` or `High`.

Another part of our system is a rule database describing relationships between input and output variables. Few rules are presented below:

```
if Size is Med and Vert is High then Plate is High
if Size is Small and Shape is Regular and Vert is High then Plate is High
if Size is VSmall and Horiz is Edge and Vert is High then Lock is High
```

Our system is composed of more than twenty rules at the moment. Operators used were basic `min` and `max` for AND/OR, `min` for implication and `max` for accumulation. Bisection was used as a defuzzifier.

Final decision about object class was made based on a defuzzified result given by system described earlier. If final score for object for all classes was below given value S_{low}, then this object was labeled as unknown. Also, from every segment placed on the wall Lock and Handle classes were removed and from segments placed on the door Switch class was removed.

4 Experiments

To confirm the validity of the approach, tests on a set of different doors were carried. Initially, the robot is placed near the door we're interested in and looks roughly forward at them (fig. 1a). Then series of 6 photos is taken, moving the head (in 10° increments down from straight forward position), which are then merged and analyzed according to the algorithm presented earlier.

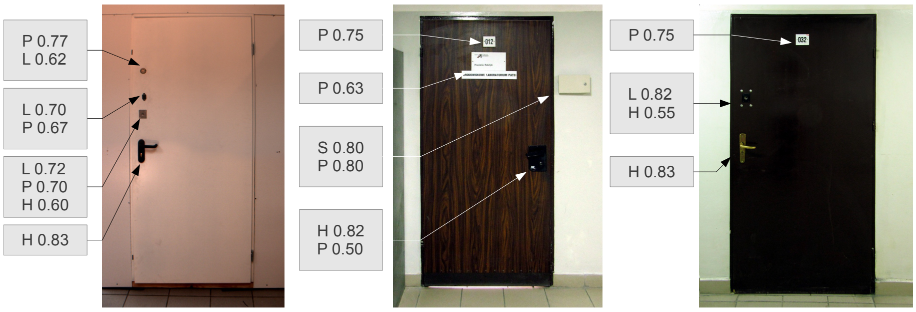


Fig. 6. A sample of test results. Labels shown for classes with membership bigger than 0.4. H - handle, L - lock, P - plate, S - switch

Algorithm was tested on photos of doors varying in color and texture, different handles, door plates and other features. Three of them, with resulting labeling, are presented on fig. 6. For every segment its membership to classes is presented (for score bigger than 0.4).

5 Summary

In this paper we presented a new approach to detection of regions of interest in autonomous door opening task. The analysis of current solutions yielded uncovered area between door detection, handle identification and door opening, which was then analyzed and solution was proposed. Resultant robotic system was presented using agent approach, with clear division into control, effector and receptor subsystems.

Developed vision system is capable of detecting door frame and different types of features (handles, locks, plates, switches) providing clues for succeeding recognition modules. The usage of fuzzy inference gave our system robustness, which resulted in good labeling even for some not typical cases. This system is planned to be extended with some other variables, like texture diversity as input or other object classes as output. As future work we are going also to create a larger database of different doors, with bigger diversity of features. Another thing is determination of exact door parameters (i.e. size) and their opening direction, which involves the usage of three dimensional sensing (using point clouds gathered by a stereo-pair and Kinect sensors), which will also enable the usage of multimodal image segmentation prepared earlier [15].

Acknowledgment. This project was supported by the grant 2012/05/D/-ST6/03097 funded by National Science Centre.

References

1. Carinena, P., Regueiro, C.V., Otero, A., Bugarin, A.J., Barro, S.: Landmark detection in mobile robotics using fuzzy temporal rules. IEEE Transactions on Fuzzy Systems 12(4), 423–435 (2004)
2. Anguelov, D., Koller, D., Parker, E., Thrun, S.: Detecting and modeling doors with mobile robots. In: Proceedings of the 2004 IEEE International Conference on Robotics and Automation, ICRA 2004, vol. 4, pp. 3777–3784. IEEE (2004)
3. Rusu, R.B., Meeussen, W., Chitta, S., Beetz, M.: Laser-based perception for door and handle identification. In: International Conference on Advanced Robotics, ICAR 2009, pp. 1–8. IEEE (2009)
4. Aude, E., Lopes, E., Aguiar, C., Martins, M.: Door crossing and state identification using robotic vision. In: 8th International IFAC Symposium on Robot Control (Syroco 2006), Bologna, Italy (2006)
5. Chen, Z., Birchfield, S.T.: Visual detection of lintel-occluded doors from a single image. In: IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops, CVPRW 2008, pp. 1–8. IEEE (2008)
6. Hensler, J., Blaich, M., Bittel, O.: Real-time door detection based on adaBoost learning algorithm. In: Gottscheber, A., Obdržálek, D., Schmidt, C. (eds.) EUROBOT 2009. CCIS, vol. 82, pp. 61–73. Springer, Heidelberg (2010)
7. Brooks, R., Aryananda, L., Edsinger, A., Fitzpatrick, P., Kemp, C.C., O'Reilly, U.M., Torres-Jara, E., Varshavskaya, P., Weber, J.: Sensing and manipulating built-for-human environments. International Journal of Humanoid Robotics 1(1), 1–28 (2004)
8. Tian, Y., Yang, X., Yi, C., Arditi, A.: Toward a computer vision-based wayfinding aid for blind persons to access unfamiliar indoor environments. Machine Vision and Applications, 1–15 (2013)
9. Yang, X., Tian, Y.: Robust door detection in unfamiliar environments by combining edge and corner features. In: 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), pp. 57–64. IEEE (2010)
10. Winiarski, T., Banachowicz, K.: Opening a door with a redundant impedance controlled robot. In: 9th Workshop on Robot Motion & Control (RoMoCo), pp. 221–226 (2013)
11. Klingbeil, E., Saxena, A., Ng, A.Y.: Learning to open new doors. In: 2010 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), pp. 2751–2757. IEEE (2010)
12. Walcki, M., Stefaczyk, M., Kornuta, T.: Control system of the active head of a service robot exemplified on visual servoing. In: 9th Workshop on Robot Motion and Control (RoMoCo), pp. 48–53 (2013)
13. Kornuta, T., Zieliski, C.: Robot control system design exemplified by multi-camera visual servoing. Journal of Intelligent & Robotic Systems, 1–25 (2013)
14. Walęcki, M., Banachowicz, K., Winiarski, T.: Research oriented motor controllers for robotic applications. In: Kozłowski, K. (ed.) Robot Motion and Control 2011. LNCIS, vol. 422, pp. 193–203. Springer, Heidelberg (2012)
15. Stefańczyk, M., Kasprzak, W.: Multimodal segmentation of dense depth maps and associated color information. In: Bolc, L., Tadeusiewicz, R., Chmielewski, L.J., Wojciechowski, K. (eds.) ICCVG 2012. LNCS, vol. 7594, pp. 626–632. Springer, Heidelberg (2012)