FACE DETECTION FOR VIDEO SUMMARY USING ENHANCEMENT BASED FUSION STRATEGY

Richa Mishra[1], Ravi Subban[2]

[1]*Research Scholar, Department of Computer Science, Pondicherry University, Puducherry, India*
[2]*Assistant Professor, Department of Computer Science, Pondicherry University, Puducherry, India*

Abstract

This paper investigates a fusion strategy in the enhancement-based skin color segmentation approach that can improve the performance of face detection algorithm. The algorithm is invariant for the complex background, skin races, indoor and outdoor lighting conditions. The algorithm is split into three phases. The first phase adopts spatial transform techniques in parallel to improve the contrast of the image, transform the enhanced images into YCbCr images, apply skin segmentation technique and yield the binary segmented images. The second phase calculates the weight of accuracy (WoA) of each of the segmented image and fed it into the fusion strategy to get the final skin detected region. The last phase involves face localization. The approach is not limited to only frontal face detection but it is invariant with the different head poses, illumination condition, and size of faces. The experimental result demonstrates the improvement in the accuracy and precision but reduction in FPR as compared to other classifiers.

Keywords: *Image enhancement, Face detection, Fusion strategy, and Skin color detection.*

---***---

1. INTRODUCTION

Computer vision is the artificial implementation of human vision. Human eyes are capable of visualizing wide range of frequencies of the visual spectrum. The objects in the images captured under unpredictable or unconstrained environments can be easily recognized by human eyes but it is one of the major challenging tasks in the computer vision. Face detection – the application of computer vision – plays a vital role in face recognition system, video surveillance, security systems, criminal identification, and human-computer interaction. The intent of any face detection approach is to spot the face regions within the image. The successfulness of real-time face detection techniques depends on the reliability of the facial feature selection under unconstrained environment. The facial feature comprises of the eyes, eye corners, mouth, nostrils, skin-color, texture and many more. The identification of these attributes under controlled environment is the easiest and simplest task, but it becomes complex and difficult under different lightings conditions, facial expressions, head poses, and occlusions.

Skin-color detection is one of the ways used to detect the human faces in color images. It limits the search area for the detection within the image. Instead of the benefits, skin detection has faced major challenges due to some sort of factors like complex background, skin races, indoor and outdoor lighting conditions. In order to improve the accuracy of skin detection, several image processing techniques have been developed. For enhancing the machine vision as like human eyes, the concept of spatial transformation technique has been taken into account. It has been a foremost step in applications like face detection, tracking and recognition. Its basic purpose is to improve the visual quality of digital images that exhibit bright lights and dark shadows. Thus, the improvement of face detection techniques can be based on accuracy of skin-color identification that is transitively based on the enhancement of an image. The spatial transformation can be categorized into intensity and spatial dependent routines like contrast stretching, gamma correction, histogram equalization, smoothing filter, and unsharp masking.

In this paper, an efficient and robust face detection algorithm is presented through the skin-color detection using the fusion strategy of an enhancement-based approach and morphological processing to remove the noise. It is initiated with the enhancement of input image using five enhancement techniques taking all of them in parallel and getting the five enhanced images. The color space of these enhanced images is converted into YCbCr color space and then follows the classification of skin and non-skin pixel using piecewise linear decision boundary algorithm. All of the binary segmented images have been used as input in the fusion phase to yield final skin detected region. Morphological operator is applied to remove the noise. Finally, the segmented components are labeled, their area and centroid are calculated and the bounding box is drawn around the detected human faces. The rest of the paper is organized as follows: Section 2 discusses the related work. Section 3 explains all the phases of the proposed algorithm. Section 4 narrates the experimental results and conclusion is made in Section 5.

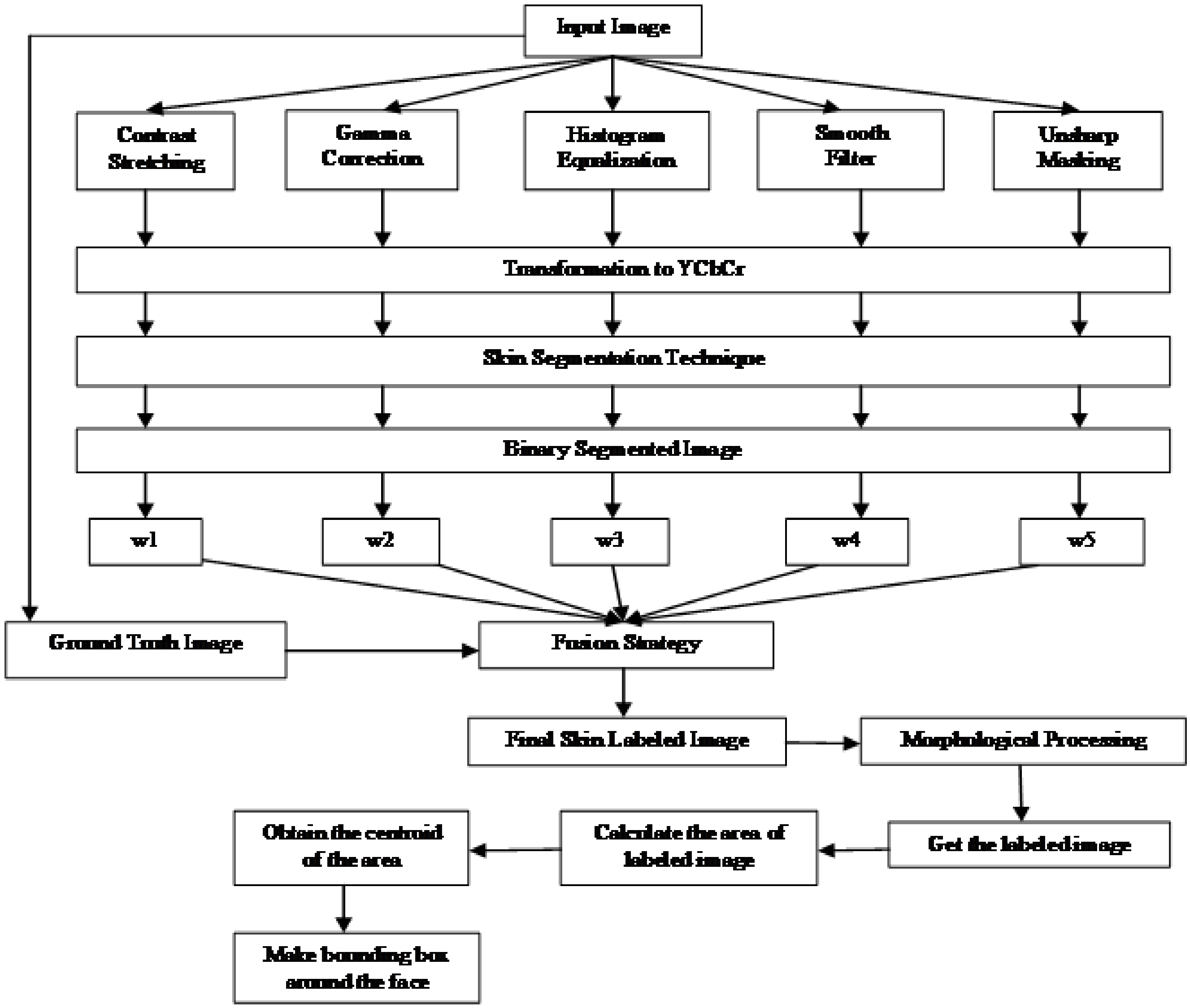

Fig -1: Flow diagram of the proposed system

2. RELATED WORK

In order to improve the performance of real-time detection and recognition systems, improvement in image enhancement and skin detection techniques should be addressed. Tao et al. [3] proposed an image enhancement algorithm by considering the dynamic range compression – neighborhood dependent intensity transformation – followed by local contrast enhancement to enhance the luminance in dark shadows. They compared their algorithm to histogram equalization and multi scale Retinex with color restoration to improve the face detection in complex lighting environment. They experimented on FRGC database containing up-right frontal human face captured under normal visual quality, dim brightness, dark face and bright background, and dark images of very low brightness. The same work has been done by Unaldi et al. [2] using different approach. They proposed wavelet-based image enhancement algorithm considering dynamic range compression and low contrast enhancement algorithm. Lim and Ibrahim [4] proposed two new enhancement techniques – CHEGIC and block based contrast enhancement method. They compared their result with the three conventional enhancement techniques namely, gamma intensity correction, histogram equalization, and homomorphicun sharp masking. Combination of color space with Gaussian model and piecewise linear decision boundary approach respectively proposed by Ravi et al. [10-11] for the segmentation of skin region. They had measured accuracy of their algorithm based on the count of number of images producing correct result – image-wise – instead of the number of pixels producing the correct result in an image – pixel-wise.

They had tested the normalized RGB color space with piecewise linear decision boundary in [12-13]. Lin [5] developed an effective and robust face detection scheme that can significantly decrease the execution time of the algorithm for those images with complex background. They used color and triangle-based approach for feature extraction. Their skin segmentation approach did not tolerate the skin-like objects. They didn't consider the head pose and occlusions in their sample sets. Yun et al. [6] used the illumination-compensation and morphological processing for face detection. They proposed an empirical modified face ratio from golden ratio. They used centroid and area for evaluating the starting and ending points of the face.

3. PROPOSED WORK

Figure 1 shows the flow diagram of the proposed approach. Face detection starts with the selection of attributes that it is going to be considered as features. The attributes that we have selected are skin-color, area of the facial region, and its centroid. The following subsections give the detailed description of the proposed approach.

3.1Skin Detection Technique

Detection of skin region approach has been used in the paper for localizing human faces within the sample image. It starts with the selection of color spaces followed by the skin detector model that can label pixels either as skin or non-skin. Commonly, it is based on the theory that the color model can effectively help in modeling the skin-color that follows the segmentation of skin regions [7]. But the overlapping of skin and non-skin pixels limits its accuracy. It has been observed in the literature that the accuracy of the technique can be enhanced by adding additional features that helps in labeling the pixels more accurately. The additional feature that we have considered in the paper is image enhancement techniques as a preprocessing step that are discussed in the succeeding subsections.

3.1.1 Preprocessing Step

This step involves the application of the spatial transformation techniques for improving the visibility of image to machine. It helps in contrast manipulation and sharpening of images. There are five techniques that we have considered for improvising the details of the images namely, contrast stretching (CS), gamma correction (GC), histogram equalization (HE), smoothing filter (SF) and unsharp masking (UM). The input image, I(x,y), is fed to each of the techniques simultaneously and converted into different enhanced images, generated by applying the transformation function of the techniques to each channels of the color space. The mathematical representation for preprocessing the image pixels are given in equation (1).

$$E(x,y) = T[I(x,y)] \quad (1)$$

Here, E(x,y) represents the enhanced image, T represents the transformation function, applied on the input image, I(x,y) to produce the enhanced image. The above equation represents the transformation function applied on the grayscale image. In this paper, color images are considered. Thus, we have modified the equation as shown in the equation (2).

$$[E_R(x,y)E_G(x,y)E_B(x,y)] = T[I_R(x,y)I_G(x,y)I_B(x,y)] \quad (2)$$

$$E_{ki}(x,y) = T_i[I_k(x,y)] \quad (3)$$

Here, k = {R, G, B} channels and i = 1 to 5. The purpose of using the above mentioned techniques is due to their benefits in different situations. Contrast stretching attempts to increase the appearance of large-scale light-dark transitions. Power law transformation (Gamma correction) is used when expansion of dynamic range of image requires. Unsharp mask increases the appearance of small-scale edges. Figure 2 shows the effect on original image after applying the five different transformation functions independently and individually.

3.1.2 Color Space

The rationale behind the detection of human faces in color images is to get more information than a gray level image [1]. This implies that color images are capable of improving the accuracy of detection.

Fig -2: Enhanced images of the sample images

It has been observed in the literature that YCbCr color space always yields better result as compared to other color space in the discrimination of skin-colors as well as human faces. That's why,YCbCr is used in the paper to convert the color space of enhanced images to YCbCr color space. Equation (4) represents its transformation formula. Figure 3 shows the color converted image of original image after applying gamma corrected enhancement technique. Fig 3(a) shows the original image. Fig 3(b) shows the enhanced image after applying gamma correction. Fig 3(c) shows the change in color space of the enhanced image from RGB to YCbCr.

$$Y = 0.2989 * R + 0.5866 * G + 0.1145 * B$$

$$Cb = 0.1688 * R - (0.3312 * G) + (0.5 * B) \quad (4)$$

$$Cr = (0.5 * R) - (0.4184 * G) - (0.0816 * B)$$

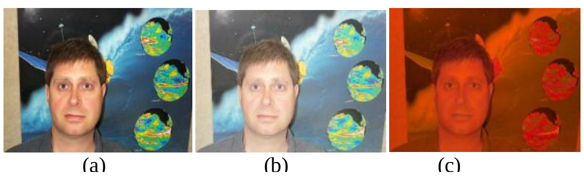

Fig -3: Color converted image of gamma enhanced image

3.1.3 Skin Segmentation

After the completion of color transformation, skin segmentation approach is performed. Piecewise linear decision boundary approach has used for labeling the pixels either as skin or non-skin. The approach takes transformed images one at a time and converts it into binary skin segmented image using multiple thresholds. This intimates that five binary segmented images will be yielded, one for each transformed image and is shown in figure 5.

3.1.4 Fusion Strategy

Fusion strategy implies the fusion of multiple results obtained from different algorithms so that more information can be obtained as compared to single algorithm. It performs either at score level or decision level. Score level makes the final decision on the basis of combination of match scores generated by each classifier to generate a new match score. It involves sum rule, product rule, exponential sum rule, and tan hyperbolic sum rule. Decision level makes the decision either yes or no or in other words, either classifiers output match or no match. Majority voting, weighted voting and "AND" and "OR" rules are the decision level fusion strategies [8]. In biometric recognitions, the sum rule has produced the best performance overall as compared to other methods. The parameter weight is associated with the score level fusion that represents the performance of different detectors. It requires FPR (false positive rate) and FNR (false negative rate) of each classifier to compute their respective weights. Larger the weight, smaller will be the error rate. Thus, we have referred this parameter as weight of accuracy (WoA). Xiao [8] used the sum rule with the fuzzy approach to detect human faces using skin detector, eye detector, and face shape detector. The weight for each classifier is computed using equation (5)

$$w_i = \frac{1-(FPR_i+FNR_i)}{N-\sum_{k=1}^{N}(FPR_k+FNR_k)} \quad (5)$$

$$FPR_i = \frac{FP_i}{FP_i+TN_i} \quad (6)$$

$$FNR_i = \frac{FN_i}{FN_i+TP_i} \quad (7)$$

where, k = 1 to 5, FPRi and FNRi represents the false positive rate and false negative rate of ith classifier, respectively. The manually labeled ground truth of the image has been used for estimating the FPR and FNR of each classifier. After computing the weights, Sf is calculated for each pixel using equation (8).

$$S_f = \sum_{i=1}^{N} w_i s_i \quad (8)$$

where, si represents the ith pixel of binary skin segmented image of each classifier, and N = m x n represents total number of pixels in the input image. This implies that the Sf value is calculated for each and every pixel of the image. If the computed value is greater than some predefined threshold θ, then the pixel is labeled as skin (1) else non-skin (0).

$$y = \begin{cases} 1, & S_f > \theta \\ 0, & \text{otherwise} \end{cases} \quad (9)$$

3.1.5 Morphological Operator

Morphological operator is used to remove the noise that was added during the segmentation process. Removing the noise from the binary image by deleting all the connected components that have fewer than predefined pixels, results in another binary skin segmented image. After that, erosion operation is applied followed by dilation morphological operator to get the final skin segmented image.

3.2 Face Localization

This step involves the localization of face candidate from the skin segmented regions as shown in figure 4. As we have stated, increasing TPR by reducing FPR for improving the localization of human faces is the main objective of our paper. The task is split into two steps: computing the facial area and identifying the face region. First, the regions are labeled using 8-connected components shown in figure 4.

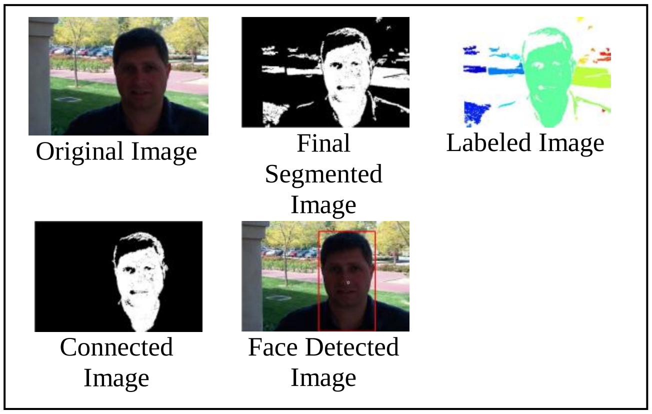

Fig. 4: Face localization in sample image

The area of all the connected components is calculated and compared with the predefined threshold to test whether it is face or non-face. If it is a face, the centroid of the facial area is computed and the coordinates for making the bounding box around is extracted.

4. EXPERIMENTAL RESULTS AND DISCUSSION

The Caltech face dataset [9] is used to compare the performance of the proposed approaches with the other state-of-art methods. Caltech dataset is designed to study face detection results under complex background, indoor and outdoor lighting conditions. It contains 452 images of size 896 x 592. The performance of proposed approach (PA) have been compared with five conventional enhancement techniques (CS, GC, HE, SF and UM). The performance metrics considered for estimating the performance are TPR, FPR, precision, and accuracy. The metrics are obtained by using the following equations:

$$TPR = \frac{TP}{TP+FN} \quad (10)$$

$$Precision = \frac{TP}{TP+FP} \quad (11)$$

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (12)$$

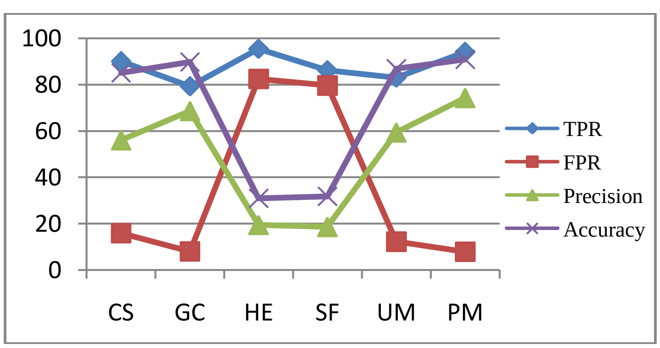

Chart -1: Performance chart of all the techniques

Table -1: Performance evaluation of the proposed approach of the skin segmentation

Performance Metrics (%)	Image Enhancement Techniques					Proposed approach
	CS	GC	HE	SF	UM	
TPR	90.1	79.4	95.5	86.2	83.1	94.1
FPR	15.9	8.1	82.5	79.7	12.3	7.9
Precision	56.1	68.7	19.5	18.6	59.5	74.3
Accuracy	85.2	89.8	31.0	31.8	86.9	90.9

Here, TP (True Positive) represents the number of pixels correctly detected as skin. FP (False Positive) represents the number of non-skin pixels detected as skin. TN (True Negative) represents the number of pixels correctly detected as non-skin. FN (False Negative) represents the number of skin pixels detected as non-skin. The effect of the unconstrained environment that results in incorrect labeling of pixels that transitively results in the incorrect face detection are shown in figure 5.

Fig. 5: Face detection examples

The aim of any skin classifier is to maintain a tradeoff between TPR and FPR. Larger the TPR, smaller should be the FPR. Contrast enhanced segmented image yielded 62.73% TPR with 24.26% FPR for the input image containing human face facing shadow light as compared to its background. It labeled some portion of skin region as non-skin mistakenly as compared to histogram equalized segmented image where FPR is 3.95%. The HE and SF segmented image have higher FPR as compared to other techniques. The qualitative performance of the PA compared with the other techniques is shown in chart 1.

It is clearly observed from the chart that the relation between TPR and FPR is quiet proportional to each other in each classifier except for the proposed approach. From this point of view, we can conclude that the proposed approach is more efficient than the other techniques in the literature. The accuracy and precision of all the classifiers is shown in table 1 along with the TPR and FPR of the average. The proposed approach produces 90.9% accuracy which is better than the other methods used in skin detection.

The performance of the newer approach in face detection is also calculated in terms of precision ratio. The proposed approach is tested on 50 images and got the precision ratio of 90.9% on the average.

5. CONCLUSIONS

Face detection based on geometrical features is proposed in this paper. We have used the enhancement approach along with the fusion strategy to locate the human faces robustly in images. The affected skin segmented region results after enhancement techniques due to the complex background and indoor and outdoor conditions. To resolve the problem, the fusion strategy is considered to reduce the FPR and increase the TPR.

Since, the accuracy of face detection results depends on the accuracy of skin region extraction from an image. Therefore, our algorithm fixes the delocalization problem of bounding box by correctly identifying skin-color, and face area to locate the human face. The benefit of our approach is its simplicity and low complexity. Multiple face detection in various situations will be our future investigation.

ACKNOWLEDGMENTS

This work is supported and funded by the University Grant Commission (UGC) of India under major research project (F.No. 41-648-2012 (SR)) to the Department of Computer Science, Pondicherry University, Puducherry, India.

REFERENCES

[1]. Ing-Sheen Hsieh, Kuo-Chin Fan, and Chiunhsiun Lin, “A statistic approach to the detection of human faces in color nature scene”, Pattern Recognition 35 (2002) 1583-1596, Elsevier.

[2]. NumanUnaldi, Praveen Sankaran, Vijayan K. Asari, and Zia-ur Rahman, “Image enhancement for improving face detection under non-uniform lighting conditions”, In Proceedings of the IEEE, 2008, 1332-1335.

[3]. Li Tao, Ming-Jung Seow, and Vijayan K. Asari, “Nonlinear image enhancement to improve face detection in complex lighting environment”, International Journal of Computational Intelligence Research, vol. 2, no. 4 (2006), 327-336.

[4]. Chiou Shi Lim and Haidi Ibrahim, “Image enhancement for face images using spatial domain processing”, International Journal of Advanced Research in Computer Science and Electronics Engineering (IJARCSEE), vol. 2, iss. 11, Nov 2013, 714-717.

[5]. Chiunhsiun Lin, “Face detection in complicated backgrounds and different illumination conditions by using YCbCr color space and neural network”, Pattern Recognition Letters 28 (2007), 2190-2200.

[6]. Jae-Ung Yun, Hyung-Jin Lee, Anjan Kumar Paul, and Joong-Hwan Baek, “Face detection for video summary using illumination-compensation and morphological processing”, Pattern Recognition Letters 30 (2009), 856-860.

[7]. Michal Kawulok, JolantaKawulok, and Jakub Nalepa, “Spatial-based skin detection using discriminative skin-presence features”, Pattern Recognition Letters (2013), Article in Press.

[8]. Qinghan Xiao, “Using fuzzy adaptive fusion in face detection”, Proceedings of the IEEE, 2011.

[9]. Computational Vision at Caltech, http://www.vision.caltech.edu/html-files/archive.html, March 17, 2005.

[10]. Ravi Subban and Richa Mishra, “Skin segmentation based on Gaussian model by combining color spaces”, International Journal of Emerging Technologies in Computational and Applied Sciences (IJETCAS), 2013, 189-197.

[11]. Ravi Subban and Richa Mishra, “Human skin detection using combination of color spaces in color images”, International Journal of Software and Web Sciences (IJSWS), 2013, 13-21.

[12]. RaviSubban and Richa Mishra, “Rule-based face detection in color images using normalized RGB color space – a comparative study”, Proceedings of the 2012 IEEE International Conference on Computational Intelligence and Computing Research, 2012, 1-5.

[13]. Ravi Subban and Richa Mishra. "Face detection in color images based on explicitly-defined skin color model." In Advances in Computing, Communication, and Control Springer Berlin Heidelberg, 2013 570-582