

## Appearance-based re-identification of people in video

**Author:**

Khan, Mohammad Arif

**Publication Date:**

2009

**DOI:**

<https://doi.org/10.26190/unsworks/14889>

**License:**

<https://creativecommons.org/licenses/by-nc-nd/3.0/au/>

Link to license to see what you are allowed to do with this resource.

Downloaded from <http://hdl.handle.net/1959.4/44585> in <https://unsworks.unsw.edu.au> on 2024-04-25



UNSW



>014954664



PLEASE TYPE

THE UNIVERSITY OF NEW SOUTH WALES  
Thesis/Dissertation Sheet

Surname or Family name: Khan

First name: Mohammad

Other name/s: Arif

Abbreviation for degree as given in the University calendar: MPhil

School: Computer Science and Engineering

Faculty: Engineering

Title: Appearance-based Re-identification of People in Video

Abstract 350 words maximum: (PLEASE TYPE)

Re-identification of people in video is of prime importance in many applications. The premise of this task is that given a person if the video how can we re-identify the same person based on appearance. This research investigates colour histograms with ability to integrate spatial distribution of the colour. To describe a person based on the colour of attire, *Colour Context People Descriptor* has been proposed. Three data sets have been used to test and compare different colour histogram methods. Four colour spaces: RGB, HSV, YIQ and XYZ have been used in colour modelling. The reidentification task has been set up as a Content Based Image Retrieval problem. ROC curves for each person in the dataset using different parameters have been obtained. All of these results have been compared within each data set and across all data sets. It has been shown that methods incorporating spatial distribution of colours perform better than colour histograms. Furthermore, *Colour Context People Descriptor* provided the best ROC curves across all methods and data sets.

Declaration relating to disposition of project thesis/dissertation

I hereby grant to the University of New South Wales or its agents the right to archive and to make available my thesis or dissertation in whole or in part in the University libraries in all forms of media, now or here after known, subject to the provisions of the Copyright Act 1968. I retain all property rights, such as patent rights. I also retain the right to use in future works (such as articles or books) all or part of this thesis or dissertation.

I also authorise University Microfilms to use the 350 word abstract of my thesis in Dissertation Abstracts International (this is applicable to doctoral theses only).

Signature

Witness

Date

The University recognises that there may be exceptional circumstances requiring restrictions on copying or conditions on use. Requests for restriction for a period of up to 2 years must be made in writing. Requests for a longer period of restriction may be considered in exceptional circumstances and require the approval of the Dean of Graduate Research.

FOR OFFICE USE ONLY

Date of completion of requirements for Award:

11-09-2009

THIS SHEET IS TO BE GLUED TO THE INSIDE FRONT COVER OF THE THESIS

### ORIGINALITY STATEMENT

'I hereby declare that this submission is my own work and to the best of my knowledge it contains no materials previously published or written by another person, or substantial proportions of material which have been accepted for the award of any other degree or diploma at UNSW or any other educational institution, except where due acknowledgement is made in the thesis. Any contribution made to the research by others, with whom I have worked at UNSW or elsewhere, is explicitly acknowledged in the thesis. I also declare that the intellectual content of this thesis is the product of my own work, except to the extent that assistance from others in the project's design and conception or in style, presentation and linguistic expression is acknowledged.'

Sign

Date 15 OCT 2009



**ORIGINALITY STATEMENT**

'I hereby declare that this submission is my own work and to the best of my knowledge it contains no materials previously published or written by another person, or substantial proportions of material which have been accepted for the award of any other degree or diploma at UNSW or any other educational institution, except where due acknowledgement is made in the thesis. Any contribution made to the research by others, with whom I have worked at UNSW or elsewhere, is explicitly acknowledged in the thesis. I also declare that the intellectual content of this thesis is the product of my own work, except to the extent that assistance from others in the project's design and conception or in style, presentation and linguistic expression is acknowledged.'

....

...

## **COPYRIGHT STATEMENT**

'I hereby grant the University of New South Wales or its agents the right to archive and to make available my thesis or dissertation in whole or part in the University libraries in all forms of media, now or here after known, subject to the provisions of the Copyright Act 1968. I retain all proprietary rights, such as patent rights. I also retain the right to use in future works (such as articles or books) all or part of this thesis or dissertation.

I also authorise University Microfilms to use the 350 word abstract of my thesis in Dissertation Abstract International (this is applicable to doctoral theses only).

I have either used no substantial portions of copyright material in my thesis or I have obtained permission to use copyright material; where permission has not been granted I have applied/will apply for a partial restriction of the digital copy of my thesis or dissertation.'

Signed 

Date

## **AUTHENTICITY STATEMENT**

'I certify that the Library deposit digital copy is a direct equivalent of the final officially approved version of my thesis. No emendation of content has occurred and if there are any minor variations in formatting, they are the result of the conversion to digital format.'

Signed 

Date



# **Appearance-based Re-identification of People in Video**

**Master of Philosophy**

**Mohammad Arif Khan**

September 1 2009

## Abstract

Re-identification of people in video is of prime importance in many applications. The premise of this task is that given a person in the video how can we re-identify the same person based on appearance. This research investigates colour histograms with ability to integrate spatial distribution of the colour. To describe a person based on the colour of attire, *Colour Context People Descriptor* has been proposed. Three data sets have been used to test and compare different colour histogram methods. Four colour spaces: RGB, HSV, YIQ and XYZ have been used in colour modelling. The re-identification task has been set up as a Content Based Image Retrieval problem. ROC curves for each person in the dataset using different parameters have been obtained. All of these results have been compared within each data set and across all data sets. It has been shown that methods incorporating spatial distribution of colours perform better than colour histograms. Furthermore, *Colour Context People Descriptor* provided the best ROC curves across all methods and data sets.



## Acknowledgements

This research has been more strenuous than I thought it would be. Had it not been for the support of people around me, I would not have been able to complete it.

I would like to thank my wife for encouraging me to take up this research opportunity and cheered me up when I was stressed.

I am also grateful to my supervisor, Dr. Jian Zhang. He was very cooperative and patient when things were not moving along as expected and gave me invaluable advice to find the solution to the problems. I really appreciate him making himself available and meeting with me in spite of his busy schedule.

I am especially thankful to my co-supervisor, Dr. Yang Wang for his technical advice during my research. His recommendations helped me concentrate the focus of this research on the current topic.

During the annual reviews when a student needs much support and encouragement, Dr. Getian Ye and Dr. Rod Belcher from NICTA and Dr. Xuemin Ling and Dr. Wei Wang from UNSW were there to assist me and provided precious feedback.

Finally, I am in debt to all the hard working staff at NICTA and UNSW for making a variety of resources available for me and all the other research students.

Table of Contents

1 Introduction ..... 1

1.1 Research Scope & Methodology..... 1

1.2 Proposed Contributions ..... 4

1.3 Thesis Outline..... 4

2 Literature Review ..... 6

3 Appearance Modelling ..... 12

3.1 Colour Spaces..... 13

3.1.1 RGB ..... 13

3.1.2 HSV..... 15

3.1.3 XYZ ..... 16

3.1.4 YIQ..... 17

3.2 Colour Representation..... 17

3.3 Histogram Matching..... 19

3.3.1 Bhattacharyya Coefficient..... 19

3.3.2 Histogram Intersection..... 20

3.3.3 Euclidean Distance ..... 21

3.3.4 Manhattan Distance..... 22

3.3.5 Kullback-Leibler Divergence ..... 23

3.3.6 Jeffrey Divergence..... 23

3.4 Body Histogram ..... 24

3.5 Top-bottom Histogram..... 26

3.6 Colour Context People Descriptor..... 28

3.6.1 Torso Descriptor ..... 29

3.6.2 Lower Body Descriptor ..... 33

3.6.2.1 Proportional Area Back-Projection ..... 36

3.6.2.2 Scanned Area Back-Projection..... 38

3.6.2.3 Segmentation Selection ..... 39



3.6.2.4	Simple Histogram.....	43
3.7	Conclusion.....	43
4	Experiments.....	46
4.1	Parameters.....	46
4.2	Datasets .....	50
4.2.1	CAVIAR Dataset.....	50
4.2.2	NICTA Dataset .....	52
4.2.3	NICTA-CAVIAR Merged Dataset .....	54
4.3	ROC Curves .....	54
4.4	CAVIAR Dataset Benchmark.....	56
4.4.1	Body Histogram.....	56
4.4.2	Top-Bottom Histogram.....	57
4.4.3	Top-Bottom Back-Projection Histogram.....	59
4.4.4	Colour Context People Descriptor.....	60
4.4.5	Hybrid Colour Context People Descriptor .....	61
4.4.6	Analysis.....	62
4.5	NICTA Dataset Benchmark.....	66
4.5.1	Body Histogram.....	66
4.5.2	Top-Bottom Histogram.....	67
4.5.3	Top-Bottom Back-Projection Histogram.....	68
4.5.4	Colour Context People Descriptor.....	69
4.5.5	Hybrid Colour Context People Descriptor .....	70
4.5.6	Analysis.....	71
4.6	NICTA-CAVIAR Benchmark .....	75
4.6.1	Body Histogram.....	76
4.6.2	Top-Bottom Histogram.....	76
4.6.3	Top-Bottom Back-Projection Histogram.....	77
4.6.4	Colour Context People Descriptor.....	78

4.6.5	Hybrid Colour Context People Descriptor .....	79
4.6.6	Analysis.....	80
4.7	Conclusion.....	85
5	Conclusion .....	89
References .....		i
Appendix.....		iii
Best Individual Results – CAVIAR.....		iii
Worst Individual Results - CAVIAR .....		vii
Best Individual Results - NICTA.....		xi
Worst Individual Results - NICTA.....		xiii
Best Individual Results – NICTA-CAVIAR.....		xvi
Worst Individual Results – NICTA-CAVIAR.....		xxii



**List of Figures**

Figure 1-1: Video frame resolution..... 3

Figure 3-1: RGB Cube [20] ..... 14

Figure 3-2: HSV Colour Space [21]..... 15

Figure 3-3: Different Image with Same Colours..... 18

Figure 3-4: Same Histogram of Different Images ..... 18

Figure 3-5: Creating Body Histogram..... 24

Figure 3-6: Extreme Image Rotation ..... 25

Figure 3-7: Realistic Image Rotation ..... 25

Figure 3-8: Different People with Similar Colour Histogram ..... 25

Figure 3-9: Recovering Body Sections..... 26

Figure 3-10: Creating Top-Bottom Colour Histogram ..... 27

Figure 3-11: Different People with Same Torso Colour Distribution..... 30

Figure 3-12: Shape Context..... 30

Figure 3-13: Extending Shape Context to Include Colour..... 31

Figure 3-14: Colour Shape Context – Colour Dimension..... 32

Figure 3-15: Person with Bounding Box..... 32

Figure 3-16: Creating Colour Context Histogram ..... 33

Figure 3-17: Unwanted Background around Legs ..... 33

Figure 3-18: Colour Separation between Torso and Legs ..... 35

Figure 3-19: Background Rejection using Proportional Rectangles..... 37

Figure 3-20: Foreground Segmentation using Proportional Rectangles ..... 37

Figure 3-21: Background Rejection using Scan Search ..... 38

Figure 4-2: CAVIAR - Body Histogram ROC ..... 57

Figure 4-3: CAVIAR - Top-Bottom Histogram ROC..... 58

Figure 4-4: CAVIAR - Top-Bottom Back-Projection Histogram ROC ..... 59

Figure 4-5: CAVIAR - Colour Context People Descriptor ROC..... 60

Figure 4-7: CAVIAR - Average ROC per Method ..... 62

Figure 4-9: CAVIAR - Worst ROC per Method ..... 64

Figure 4-10: CAVIAR - Best Colour Spaces..... 65

Figure 4-11: CAVIAR - Worst Colour Spaces..... 66

Figure 4-12: NICTA – Histogram ROC ..... 67

Figure 4-13: NICTA – Top-Bottom Histogram ROC ..... 68

Figure 4-14: NICTA – Top-Bottom Back-Projection Histogram ROC..... 69

Figure 4-15: NICTA – Colour Context People Descriptor ROC ..... 70

Figure 4-16: NICTA – Hybrid Colour Context People Descriptor ROC ..... 71

Figure 4-17: NICTA – Average ROC per Method ..... 72

Figure 4-19: NICTA – Worst ROC per Method..... 74

Figure 4-20: NICTA – Best Colour Spaces..... 75

Figure 4-21: NICTA – Worst Colour Spaces ..... 75

Figure 4-22: NICTA–CAVIAR - Histogram ROC..... 76

Figure 4-23: NICTA–CAVIAR – Top-Bottom Histogram ROC ..... 77

Figure 4-24: NICTA–CAVIAR – Top-Bottom Back-Projection Histogram ROC..... 78

Figure 4-25: NICTA–CAVIAR – Colour Context People Descriptor ROC ..... 79

Figure 4-26: NICTA–CAVIAR – Hybrid Colour Context People Descriptor ROC ..... 80

Figure 4-27: NICTA–CAVIAR – Average ROC per Method..... 81

Figure 4-28: NICTA–CAVIAR – Best ROC per Method ..... 82

Figure 4-29: NICTA–CAVIAR – Worst ROC per Method..... 83

Figure 4-30: NICTA–CAVIAR – Best Colour Spaces..... 84

Figure 4-31: NICTA–CAVIAR – Worst Colour Spaces ..... 84

Figure 4-32: Average ROC Per Method ..... 86

Figure 4-34: Worst Colour Spaces..... 87

**List of Tables**

Table 3-1: Location Invariance of Legs ..... 40

Table 3-2: Scale Invariance of Legs..... 41

Table 3-3: Scale Invariance of Legs..... 41

Table 3-4: Finding Maximum A Posteriori ..... 42

Table 4-1: People in CAVIAR dataset..... 51

Table 4-2: People in CAVIAR dataset..... 51

Table 4-3: People in CAVIAR dataset..... 52

Table 4-4: People in NICTA dataset..... 53

Table 4-5: People in NICTA dataset..... 53

List of Equations

Equation 3-1 ..... 20

Equation 3-2 ..... 20

Equation 3-3 ..... 21

Equation 3-4 ..... 21

Equation 3-6 ..... 22

Equation 3-7 ..... 23

Equation 3-8 ..... 24

Equation 3-9 ..... 40

Equation 3-10 ..... 42



# 1 Introduction

---

Re-identification of people in video is an important task. But despite its importance, this subject has not received its fair share of research attention. The purpose of this research is to investigate the re-identification of people in video based on their appearance. Given a video frame, one of the most distinguishing appearance features is the colour of people's clothing. Colour is an important attribute in video and is easily obtainable. Furthermore, colour provides a degree of rotation invariance if front and back colour distribution is similar. But the downside of naively using colour as a dominant feature is that the recognition rate will drop. The reason for this is because different people might be wearing similarly coloured clothing. So to improve the re-identification, it is important to utilise the distribution of the colour. This research has culminated in an appearance-based people descriptor that uses colour distribution to obtain better re-identification results.

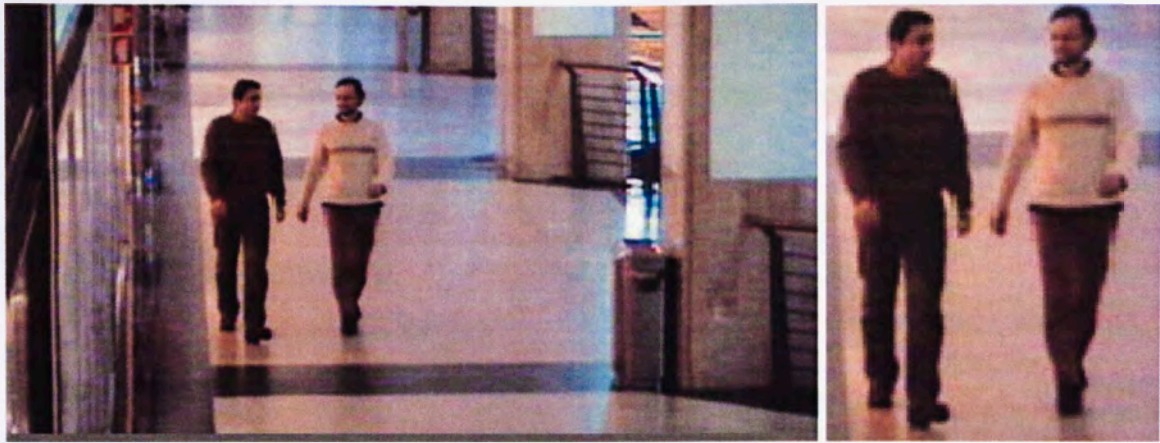
## 1.1 Research Scope & Methodology

The research deals with re-identification of people in video. This is the main focus of this research and it does not encompass detection of people in video. Detection of people in video is an active research theme of its own. There are a few assumptions underlying the research. Firstly, it requires that the people are in upright pose and are fully visible. This is not a very restricting assumption as most of the videos will have people standing or walking upright and mostly visible. So, the research has not used videos where people are sitting or laying down such as in sports videos. However,

having said that, it does not mean that the people descriptor will not work in these situations as it has not been tested for these scenarios. Secondly, it assumes that the people in the video are from the same session. In other words, their clothing has not changed. Which again is a valid hypothesis considering the state of the art in this research area. There is no research work to claim cross-session re-identification of people. Lastly, the creation of the people descriptor requires a bounding box already defined around the people. Again, this is a reasonable and non-restricting assumption because a bounding box can easily be placed around a person by using background subtraction techniques, tracking or human recognition. Furthermore, this assumption disconnects the application of this research from the way the bounding box is placed. So to summarise, the following are the assumptions:

- People are in upright pose with full visibility
- People are from the same session
- A bounding box is already placed around people

Often the video images lack the resolution to leverage research from other areas. For example, it is not possible to use face recognition algorithm to complement the re-identification algorithm because of low quality of video images. As shown in the figure below, the head and face area of the people is grainy leaving us the torso and legs area to work with.



**Figure 1-1: Video frame resolution**

In order to use algorithms that require minute details such as facial features, we would require high resolution cameras with zooming capabilities. The main problem with using such cameras is the cost. Implementation of a new network of such cameras will be more costly than using an average colour camera. This would mean compromising the network design such that less number of cameras is used. This in turn will reduce the effectiveness of the camera network. Similarly, there is a lot of investment involved in existing camera networks. This investment cannot be thrown away to install a new network of high resolution zooming cameras.

Furthermore, the algorithms that require minute details, such as face recognition, might not always work because of the angle of the camera or the pose of the person. So using high resolution zooming cameras is not the answer to improving re-identification of people. We need a technique that can use the existing network of cameras to give the maximum result.

This research investigates the modelling of a person for re-identification purposes. The aim is to improve the histogram model to create a more robust descriptor of people

which can enhance the re-identification results. The descriptor is in the form of a histogram which can be used for matching. The advantage of using the histogram format is to leverage the use of existing histogram matching techniques.

Three datasets of images containing people were created. Different features were calculated for the people. For each dataset, ROC curves were computed. All the results were compiled and analysed.

## **1.2 Proposed Contributions**

The following are the proposed contributions of this thesis:

- Experimental evaluation of full body colour histograms and spatial colour histograms for re-identification of people.
- Improvements to full body colour histograms by separating torso and legs histograms.
- Enhancement and simplification of Shape Context descriptor to utilise colour information.
- A spatial colour descriptor for people called Colour Context People Descriptor.

## **1.3 Thesis Outline**

The rest of the thesis contains literature review, colour modelling, appearance modelling and experiments.

Section 2 focuses on the relevant research of appearance-based people re-identification. It will be seen in this section that the research carried out in this area is rather limited. It will be seen that very few appearance features are used. Especially when it comes to using colour, histograms are the only technique used.

Section 3 describes the four colour spaces, RGB, HSV, YIQ and XYZ, considered for the research. Different colour spaces describe the colour in different ways. It is hence important to consider more than one colour space to study the effects of choosing a colour space on the accuracy of re-identification. Colour histograms are also discussed in this section as it is the primary tool of colour modelling used in this research. The advantages and disadvantages of using histograms are highlighted in this section. Furthermore, some techniques for matching histograms are examined. This section also details the four techniques used in this research for people modelling. It includes a simple colour histogram of the body; a top-bottom histogram; a top-bottom histogram with background rejection and *Colour Context People Descriptor*.

The last section 4 includes all the experimentations. This section provides the information on different parameters used, the data sets and the results of extensive experiments.



# 2 Literature Review

---

A review of the literature shows that a limited amount of work has been undertaken in the field of video re-identification. Within this limited research, the majority of research has concentrated on the matching aspects of the re-identification instead of modelling such as using classifiers. It can be shown in the review that most of the work has been based around simple histograms for modelling and no serious attempt has been made to come up with a methodology to create a descriptor of a person. Furthermore, it can be seen in the review that some of the techniques require learning from data which restricts these methods to cases where learning data is available. It should be noted that systems requiring learning from existing data have restricted applicability. This is because whenever a learned parameter changes or is added, retraining is required. At the same time the importance of using colour histograms is also highlighted in this review.

It can be seen from the literature that primarily three attributes of a person have been used for re-identification:

- Colour

- Height
- Gait

Some researchers have looked at texture but largely it is the combination of colour and some other feature that has been used for re-identification. All of these features can be obtained without the use of specialised cameras.

However there is no methodology described in the literature to create an appearance-based descriptor for people which incorporates spatial colour scattering. Rather a histogram is used to model the colour or an emphasis is placed on the matching algorithm instead of the modelling.

Goldmann Et Al. (2006) used colour histograms, colour structure descriptor and co-occurrence matrix [3]. The histograms were constructed by using average RGB colour space, HMMD colour space and pixel intensities. But the recognition was done using Gaussian Mixture Model (GMM), KNN and SVM and hence training data was required. Their work reports recognition using single and combination of features. The best recognition rate of 94.7% was reported for colour histogram with KNN.

Gheissari Et Al. (2006) combined the colour histogram with edgel histogram to create a signature for a person [13]. The salient edgels were recovered using a spatiotemporal segmentation algorithm. Matching was done using interest point operator approach and model based approach. However the colour histogram did not use any spatial information which could increase the re-identification.

Javed Et Al. (2005) also used brightness adjusted colour histogram for tracking people between two cameras [8]. Brightness is adjusted by getting the Brightness Colour Transform between the two cameras. However, learning the colour transform required training data for learning. This limits the application of such approach. This is because each time a camera is added to the network, the transform needs to be learned. Furthermore, it uses a simple colour histogram ignoring any spatial colour distribution.

Jaffre and Poly (2004) have used colour histogram to model the region below face [12]. In their work of automatic video indexing, they have used a face detector to locate the face of a person in a video. A colour histogram was created for the costume and matching was done using Bhattacharya Coefficient. However their work was restricted to TV videos where face could easily be detected. Hence the technique is not applicable to cases where face information is not available or face detection cannot be performed such as in low resolution videos.

Hahnel Et Al. (2004) used colour features and texture for people recognition [4]. The colour features included RGBL histogram, normalised RG histogram and colour structure descriptor defined in MPEG-7 standard. The recognition was achieved and compared for a Radial Basis Function (RBF) and KNN classifier. However this technique also required training data. Furthermore, the people detection was done using a blue screen background resulting in a perfect segmented person image. Of course, this is not the situation in real world where there is almost certainly some level of noise in the image. Again it should be noted that the best recognition results were obtained by using colour features.

Nakajima Et Al. (2003) also used colour histograms, basic shape features and local shape features using convolution [5]. Simple shape features were extracted into histograms by counting the number of pixels in the segmented image. It is important to not that the best recognition rate was achieved by colour histogram using normalised RGB space. However, matching was done by Support Vector Machine (SVM) and K Nearest Neighbour (KNN) so training was required on hand-labelled data making its application less practical. Hence for each new person to be re-identified, training is required.

Wojtaszek and Laganieri (2002) used a simple colour histogram along with Earth Movers Distance matching for re-identification of people in a tracking system [2]. However, their technique is tightly couple with the background subtraction technique to extract the region below the neck for colour histogram and tracking during which the histogram is created over a few sequences. This means that it will not work in situations where the people are identified using a direct detection technique like histogram of oriented gradients [15]. Furthermore, it ignores other colour information in the legs area which could make the histogram more distinctive. Adding the colour information of legs will add more discriminative power.

For people re-identification purposes, the other two attributes used are height of a person [7] and gait of a person [6]. For gait alone, BenAbdelkader Et Al. (2004) used 4 gait features namely mean of oscillations, amplitude of oscillations, cadence and stride length to recognise people [6]. They used KNN for classification and the best accuracy was 49%. However, since this is based on gait alone, this recognition rate is far less than those reported in the literature. But the research highlights the application of features other than colour for re-identification of people in video.

Similarly, Nixon Et Al. (1999) have reported gait-based recognition [14]. However, the study was conducted on a small data set of 5 people. However, this was only a single feature without incorporating any colour information.

Lastly, height has also been used in conjunction with other features. Madden & Piccardi (2005) have used colour histogram along with height estimation for people re-identification across cameras [7]. They have described the method of obtaining height measurement from the video. However, a simple colour histogram does not retain any spatial information which could improve the results.

One can see from the literature review that colour is a very strong attribute for re-identification and has been used in most of the relevant research. In some cases it has been used exclusively and in others along with other features. Furthermore, [3], [4] and [5] have reported their best results have been obtained using colour histograms. On the other hand using no colour information indicates that ignoring colour information decreases the recognition rates. These studies confirm the effectiveness of colour histogram for people recognition.

It will be shown in this research that extending the histogram to take spatial colour distribution into consideration improves the re-identification rate. The research takes a step-wise approach to incorporation spatial colour distribution. The first step is to use a simple colour histogram and benchmark the performance using ROC curves. The second step is to separate the torso histogram from the bottom histogram. It is proved during the course of the research that this separation increases the performance of re-



identification. In the third step, background rejection is applied to the bottom histogram. It improves the performance in most cases. During the final step, a spatial colour descriptor called *Colour Context People Descriptor* is created. This descriptor outperforms other techniques. So it will be demonstrated at the end of this research that instead of creating a simple histogram model of people, a histogram-based descriptor, that takes into account the spatial distribution of the colour, greatly improves the re-identification of people in video.

# 3 Appearance Modelling

---

We have already seen in the literature review, that using colour histograms as part of appearance modelling increases the performance of re-identification techniques. However, a simple colour histogram for the body does not include any spatial information. Keeping in mind the difficulty of re-identification, where only few features are available, even small improvements count. So in order to improve the body histogram method, an emphasis in this research has been placed on including spatial information of the colour. The appearance modelling takes a simple colour histogram of the body as the beginning point. Afterwards spatial information is added to the scheme to improve the performance of re-identification using colour as the appearance feature.

Colour is an intuitive appearance attribute. It is easily available from the image data. Colour is one of the most popular image attributes used in computer vision. In relation to this research the importance of colour is further highlighted. This is because there are only a few attributes available when re-identifying people in video. For example, shape information cannot be used as the general shape of human body is the same and hence does not provide much discrimination. Of course, this argument does not consider height as a shape property, which can add some discriminative power.

## 3.1 Colour Spaces

The human eye has special receptors to distinguish colours. These receptors are called cone cells and are able to distinguish short, middle and long wavelengths. These three wavelengths describe the colour. Henceforth, tristimulus values describe the amount of three primary colours. These tristimulus values are usually given in Commission Internationale d' Eclairage (CIE) 1931 colour space using X, Y and Z coordinates [22]. A way of associating these tristimulus values with each colour results in a colour model. There are many colour spaces available that represent the colour information of an image. Most of the colours have three components, such as RGB, LAB, YUV etc. Using the correct colour space in image processing can affect the performance of algorithms. The following sections describe the four colour spaces used in this research.

### 3.1.1 RGB

This is a popular colour space used in computer graphics. It has three components: Red, Green and Blue. The values range from 0 to 255. All colours are made up by combining these three components. Hence it is referred to as additive colour space. The figure below shows the RGB colour space visualised as a cube:

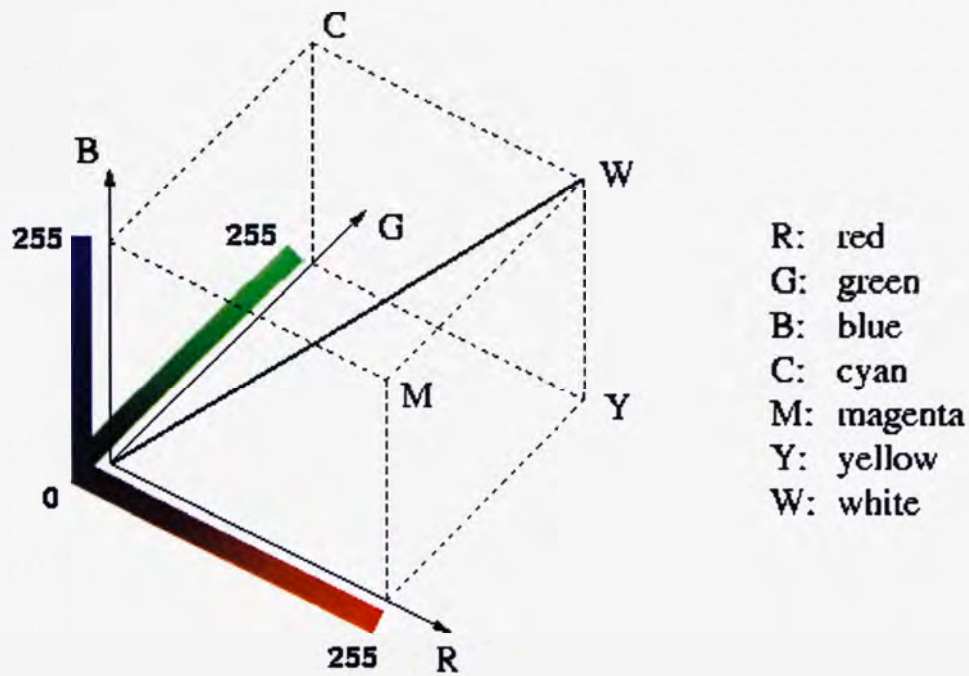


Figure 3-1: RGB Cube [20]

A very popular derivative of RGB colour space is the normalised RGB or rgb colour space. The values of R, G and B are normalised by dividing them by 255. While RGB has its implementation in hardware, this colour space is not perceptually uniform.

The main advantage of using RGB colour space is that most of the cameras use this colour space as default so there is no need to convert the colour to RGB format. This was the case with the cameras used to record NICTA dataset. This saves some computational power that can be utilized to increase the performance of the software. On the other hand, the main disadvantage of this colour space is that the hue i.e. the colour is not independent of intensity and saturation of the colour. Furthermore, the luminosity depends on all three components as it is the sum of all three components.

### 3.1.2 HSV

HSV colour space is more perceptually uniform. This colour space is depicted in form of a cone. It also consists of 3 components: Hue, Saturation and Value. Hue is in form of an angle in the ranging from 0 to 360. This represents the rotational symmetry of the cone. Saturation is in the range of 0 to 1 and its axis perpendicular to the Value axis. Value is also in the range of 0 to 1 with its axis the rotational symmetry axis of the cone. The colour space is shown below:

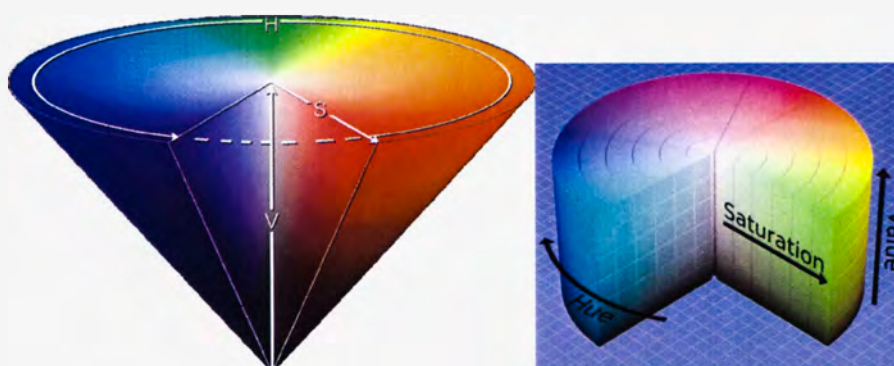


Figure 3-2: HSV Colour Space [21]

RGB values obtained from the hardware needs to be converted to HSV colour space.

The following paragraph describes this conversion process.

Normalise the RGB colour space to rgb. Let  $\max = \text{MAX}(r, g, b)$  and  $\min = \text{MIN}(r, g, b)$ :

$$h = \begin{cases} 0 & \text{if } \max = \min \\ (60^\circ \times \frac{g-b}{\max - \min} + 360^\circ) \bmod 360^\circ, & \text{if } \max = r \\ 60^\circ \times \frac{b-r}{\max - \min} + 120^\circ, & \text{if } \max = g \\ 60^\circ \times \frac{r-g}{\max - \min} + 240^\circ, & \text{if } \max = b \end{cases}$$

$$s = \begin{cases} 0, & \text{if } \max = 0 \\ \frac{\max - \min}{\max} = 1 - \frac{\min}{\max}, & \text{otherwise} \end{cases} \quad v = \max$$

There are three other popular derivatives of HSV colour space. They are HSB, HSI and HSL. Each of them replaces the Value component with brightness, intensity and lightness respectively.

The advantage of this colour space is the separation it provides between hue, saturation and luminosity. But as seen earlier, if the camera does not support providing HSV information, the image colour needs to be converted to HSV which takes up some computing resources.

### 3.1.3 XYZ

XYZ is the base colour developed by CIE and serves as a basis for many other colour spaces. It is founded upon direct measurements of human visual perception. The Y component represents luminance where as X & Z curves have been obtained using experiments using human observers. The transform for converting RGB to XYZ colour space is given below:

$$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = \frac{1}{0.17697} \begin{bmatrix} 0.49 & 0.31 & 0.20 \\ 0.17697 & 0.81240 & 0.01063 \\ 0.00 & 0.01 & 0.99 \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix}$$



The benefit of XYZ colour space is that it is a device independent colour space. So it can display consistent colour information across cameras with different characteristics. This was the reason to include XYZ colour space in this research as the image datasets used for experiments came from different cameras. On the downside, XYZ colour space is perceptually non-uniform and hence is not commonly used in image processing [24].

### 3.1.4 YIQ

YIQ colour space was specifically developed for NTSC colour television broadcast. Y represents brightness information and is only used by black-and-white televisions. I and Q represent chrominance. This colour space also takes advantage of human visual perception. Human eye is more sensitive to orange-blue (I) range as compared to purple-green (Q) range [23]. The transform for converting RGB to YIQ is given below:

$$\begin{bmatrix} Y \\ I \\ Q \end{bmatrix} = \begin{bmatrix} 0.299 & 0.587 & 0.114 \\ 0.595716 & -0.274453 & -0.321263 \\ 0.211456 & -0.522591 & 0.311135 \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix}$$

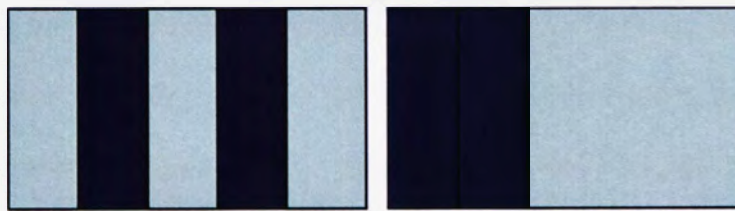
As it incorporates human visual perception in two channels, it was used in the experimentation processes.

## 3.2 Colour Representation

In order to use colour information of an image, we need a way to represent the colour information. Histogram is the primary method of representing colour information. The use of histogram to model colour information was first reported by Swain and Ballard

[1]. The use of histogram to characterise colour information is a well researched in computer vision. It is fast to compute and inherently provides rotation invariance. There are many techniques available to match histograms. Some of these methods are described in section 3.3.

However, there is one major drawback of using colour histograms: It does not retain the spatial colour distribution. This is the rotation invariance property of the histogram which can cause problems in some applications like the re-identification of people. Consider the following images with different spatial colour distribution:



**Figure 3-3: Different Image with Same Colours**

Since the above shapes consist of only two colours, they will both have the same colour histograms as shown below:



**Figure 3-4: Same Histogram of Different Images**

### 3.3 Histogram Matching

Once the image feature has been converted into a histogram, we need to perform matching. The objective of matching is to find how similar or dissimilar the underlying objects are. The matching process establishes correspondence between two histograms. One histogram belongs to the target image and the other to a candidate image. In the target image, a feature, such as colour, is chosen and converted to histogram. Then in the candidate image, a histogram of same feature is created. By matching these histograms, we can determine how closely the two images are related.

In order to match two histograms, we need to use some kind of measure that results in a value indicating the closeness of the underlying images. To find a close match, a threshold is selected. Values that are equal to or above this threshold are considered to be the same as the target object. But in CBIR systems, we return the top matches and hence there is no need to set a threshold. Instead the values returned by matching process are ranked and the results are returned. There are many techniques available for performing histogram matching. Some of these methods are described in this section.

#### 3.3.1 Bhattacharyya Coefficient

Bhattacharyya Coefficient [18] is based on divergence of two distributions. It has its roots in statistics. It has been used extensively in computer vision, such as [16]. Given two populations  $p(i)$  and  $p'(i)$  with  $N$  classes, the Bhattacharyya Coefficient is defined as:

$$\rho(p, p') = \sum_{i=1}^N \sqrt{p(i)p'(i)} \quad \text{Equation 3-1}$$

Where  $p(i)$  and  $p'(i)$  represent the probability distributions of  $N$  classes. Therefore,

we have  $\sum_{i=1}^N p(i) = \sum_{i=1}^N p'(i) = 1$ . Bhattacharyya Coefficient has a geometric

interpretation as the cosine of the angle between  $N$ -dimensional vectors

$[\sqrt{p(1)}, \dots, \sqrt{p(N)}]^T$  and  $[\sqrt{p'(1)}, \dots, \sqrt{p'(N)}]^T$ . Using this geometric analysis, if two distributions are identical, we get a value between 0 and 1 inclusive.

The Bhattacharyya Coefficient of two histograms representing identical images will be 1. As the dissimilarity between two histograms increase, the value of the coefficient will decrease towards 0. Since Bhattacharyya coefficient uses probability distribution, it fits nicely with normalised histograms. Histograms are normalised to provide scale invariance. Normalised histograms are probability distribution as their values add up to 1.

### 3.3.2 Histogram Intersection

Histogram intersection was proposed by Swain and Ballard [1] for colour based image retrieval. Considering two histograms  $I$  and  $M$  with  $n$  bins, histogram intersection is defined as:

$$H(I, M) = \sum_{j=1}^n \min(I_j, M_j) \quad \text{Equation 3-2}$$

To get a value between 0 and 1, the intersection is normalised as follows:

$$H(I, M) = \frac{\sum_{j=1}^n \min(I_j, M_j)}{\sum_{j=1}^n M_j} \quad \text{Equation 3-3}$$

Two histograms representing the same image will result in a Histogram Intersection value of 1. As the difference in the histogram increases, the value will tend towards 0. It should be noted that colours not present in the query image are included in the intersection. Furthermore, as noted in [1], histogram intersection is not symmetric in  $I$  and  $M$  and hence is not a distance metric.

### 3.3.3 Euclidean Distance

Euclidean distance can also be used to find the similarity between two histograms. Euclidean distance is one of the special cases of Minkowski-form distance shown below:

$$M(I_n, M_n) = \left[ \sum_i |I_n - M_n|^p \right]^{\frac{1}{p}} \quad \text{Equation 3-4}$$

Considering two histograms,  $I$  and  $M$  with  $n$  bins, this distance can be found by calculating Euclidean distance between corresponding bins. Euclidean distance between two points in  $n$ -dimensional space is given as:

$$E(I_n, M_n) = \sqrt{(I_1 - M_1)^2 + (I_2 - M_2)^2 + \dots + (I_n - M_n)^2}$$

$$E(I_n, M_n) = \sqrt{\sum_{i=1}^n (I_i - M_i)^2}$$

**Equation 3-5**

The minimum value for Euclidean distance is 0. This shows that the two histograms are identical and hence most likely represent the same underlying image. On the other hand, there is no maximum value for Euclidean distance between two histograms. But the larger the distance, more dissimilar the two histograms are. Hence it can be concluded that the underlying images are also very dissimilar. As opposed to the Histogram Intersection, all histogram bins contributed equally in calculating the Euclidean distance between two histograms.

### 3.3.4 Manhattan Distance

Similar to Euclidean distance, Manhattan distance is also a special case of Minkowski distance. Considering two histograms,  $I$  and  $M$  with  $n$  bins, this distance can be found by calculating Manhattan distance between corresponding bins. Manhattan distance between two points is the absolute difference between them:

$$M(I_n, M_n) = |I_1 - M_1| + |I_2 - M_2| + \dots + |I_n - M_n|$$

$$M(I_n, M_n) = \sum_{i=1}^n |I_i - M_i|$$

**Equation 3-6**

The minimum value for Manhattan distance is also 0. But there is no fixed maximum value for this distance. A distance of 0 indicates a perfect match between two histograms; whereas a larger value will represent a bigger dissimilarity between the histograms. It can be concluded that the underlying images represented by the histograms are similar for smaller Manhattan distances and vice versa.

### 3.3.5 Kullback-Leibler Divergence

Kullback-Leibler [19] measurement comes from probability and information theory. Considering two probability distributions,  $M$  and  $I$ , Kullback-Leibler measures the expected number of extra bits required to code the samples in  $M$  using code based on  $I$ . In other words, it finds the cost of encoding one distribution as another. Given two distributions, Kullback-Leibler is measured as follows:

$$KLD(I_n, M_n) = \sum_i I(i) \log \frac{I(i)}{M(i)} \quad \text{Equation 3-7}$$

It can be seen that for same normalised histograms, Kullback-Leibler divergence will be 0. It should be noted that Kullback-Leibler is not symmetric and so is not a metric but divergence measure. It is sensitive to histogram binning and is numerically not stable.

### 3.3.6 Jeffrey Divergence

Jeffrey divergence is similar to Kullback-Leibler divergence but is numerically more stable. For two probability distributions,  $M$  and  $I$ , Jeffrey divergence can be found using the expression below:

$$JD(I_n, M_n) = \sum \left[ I(i) \log \frac{I(i)}{m} + M(i) \log \frac{M(i)}{m} \right] \quad \text{Equation 3-8}$$

,where  $m = \frac{I(i) + M(i)}{2}$

As with Kullback-Leibler, Jeffrey divergence is 0 for identical distributions. The value of divergence will be smaller and larger for similar histogram and dissimilar histograms respectively.

### 3.4 Body Histogram

Given a person in an image, one way of capturing the colour information is to create a histogram. This has been the process used predominantly in the current research as described in the literature review. So once the person is identified, by using a bounding box, a colour histogram of area is constructed as shown below:



**Figure 3-5: Creating Body Histogram**

The advantage of simple colour histogram is its rotation invariance. So if the object is rotated, it will not affect its histogram. Therefore, the following images will have the same histogram regardless of the orientation of the person:





**Figure 3-6: Extreme Image Rotation**

But in a camera network, one does not encounter such extreme orientation but only slight as shown below:



**Figure 3-7: Realistic Image Rotation**

So one can see that while the rotation invariance of histograms is an important property, it is not of prime importance when looking at images of people taken from camera. A more important consideration is to know the spatial attributes of colours. This will improve true positive rate in situations where someone's top colour is similar to another person's bottom colour and vice versa as shown in the figure below:



**Figure 3-8: Different People with Similar Colour Histogram**

One way to improve the discrimination capability of the colour histogram would be to increase the number of bins. By using more bins, a histogram will throw away less colour information and will be able to capture more details of the colour distribution. But this is not a prudent solution as it will increase the processing times on histogram operations and will defeat the purpose of using histograms for colour modelling. So the body histogram is a normalised colour histogram of the area in the bounding box.

### 3.5 Top-bottom Histogram

The first step in improving a body colour histogram is to separate the histogram of the top body portion from the bottom. There are two ways to find the separation between the torso and legs. The first approach is to use the existing research and the second is to use the camera image to get the information. In this research, the first approach is used by leveraging existing research. In their research work, Park and Aggarwal [10], they have used fixed proportions to identify the three body sections: head, torso and legs. They have described the human body composed of the following proportions:



**Figure 3-9: Recovering Body Sections**

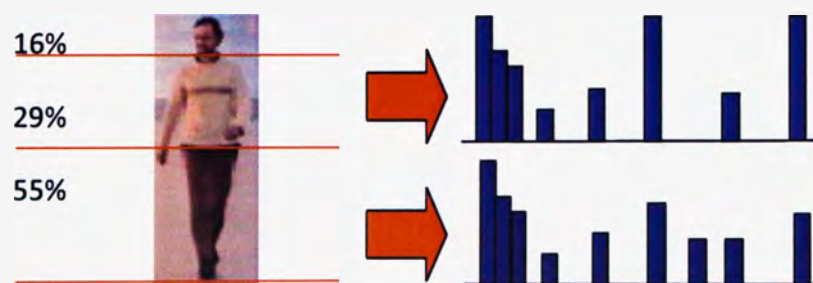
The head, torso and legs consist of 16%, 29% and 55% of the body respectively. The other approach is to find out these proportions from the video image itself. This would normally be the case if the camera is placed at a low angle disturbing the normal



proportions of the human body. Once the torso and legs portions are recovered, then colour information can be retrieved relatively easily.

Alternatively, using pre-recorded images from the camera, the proportions of head, torso and legs can be recovered. The area occupied by head, torso and legs can be manually identified. Depending on the camera angle, these proportions will vary.

The top-bottom histogram ignores the information contained in the head section of the image. The primary reason to exclude it is because this colour information does not have the discriminatory power to contribute to the improvement of re-identification process. This is because there are two main sources of colour information: face and hair. But due to the resolution of images, it is hard to distinguish between different hair and face colours. Furthermore, it has been discovered that Hue component of face of most of the people is the same [17]. Therefore, the top-bottom histogram consists of a torso colour histogram and legs colour histogram. Both histograms are normalised to make it scale invariant. The image below shows a top-bottom histogram.



**Figure 3-10: Creating Top-Bottom Colour Histogram**

Experiments have shown that separating using top-bottom histogram greatly enhances the recognition results as compared to a simple histogram. This improvement is due to

the consideration of spatial distribution of colour. A simple colour histogram does not make any distinction between torso and bottom colours; whereas, top-bottom histogram does.

### 3.6 Colour Context People Descriptor

The people descriptor describes the appearance properties of people using colour. The main method used to describe the colour properties is the histogram. But as it has been described in the previous section, it is important to consider the colour distribution of the object being described. Not considering the colour distribution will often result in false positives during classification. This is especially true when it comes to creating a colour-based appearance descriptor of people. For example, a person wearing blue trousers and brown top will match a person wearing brown trousers and blue top if we create a descriptor by using a single colour histogram. Furthermore, it has been observed by looking at the data sets that most people tend to wear similar coloured trousers so the colour of bottom portion of people is not very discriminative. Similarly, it has been noticed that there are more colours in the top portion of torso than the bottom. For example, people tend to wear multi-coloured tops or wearing a jacket on top of the shirt, such that some portion of shirt is visible, increase the number of colours in the top portion of the torso.

Keeping all these observations in mind, if we allow for the descriptor to accommodate them, it will allow in a more robust colour descriptor for people. Henceforth, the people descriptor uses colour histograms to capture the colour and is able to incorporate the above observations. The people descriptor, called *Colour Context People Descriptor*, takes the colour distribution into account at the following two levels:

- It separates colour of lower body from upper body
- It captures torso colour with its spatial distribution

Concatenating lower body colour histogram with the upper body colour distribution to create a single distribution separates the lower body and torso distribution. To incorporate spatial colour distribution of torso, histograms based on shape context scheme are created. The following sections describe the *Colour Context People Descriptor* in detail.

### 3.6.1 Torso Descriptor

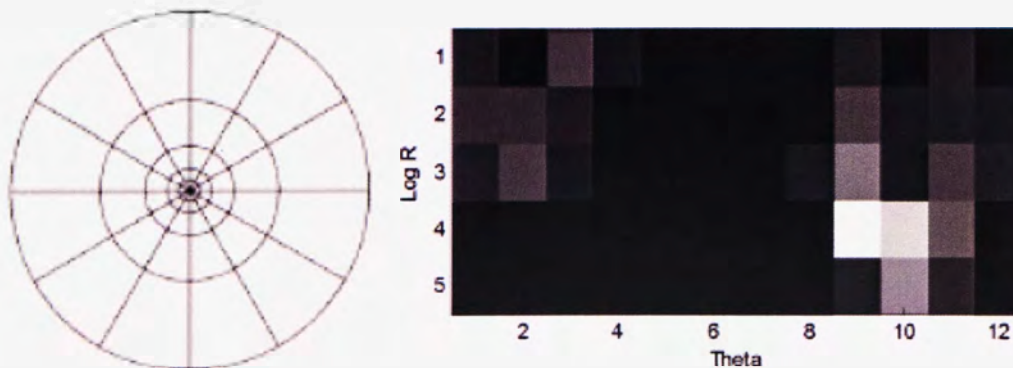
It has been already shown that colour histogram ignores spatial location of colours. This can both be good or bad depending on the application. More often than not, the torso of the people consists of more than one colour. It has been noted during the research that it is usually between 1 – 3 dominant colours. To increase the discriminative power of a histogram, an improved technique is required to incorporate the spatial distribution of the colours. To add spatial distribution to the colour information, this research has focused on creating a new spatial colour structure. Considering the following images, one can see that the torso colour histogram will be very similar however they are two different torsos.



**Figure 3-11: Different People with Same Torso Colour Distribution**

Belongie Et. Al [9] described shape context structure for matching objects based on their shapes. The shape context structure consists of radial and angular distances that can be tuned to get the best results. The shape context results in a 2D structure that captures the relationship between points on the shape contour. The shape context structure is centred on few points around the contours and then matching is performed. This in itself is computation intensive process. The authors have described ways to optimise the calculation of shape context.

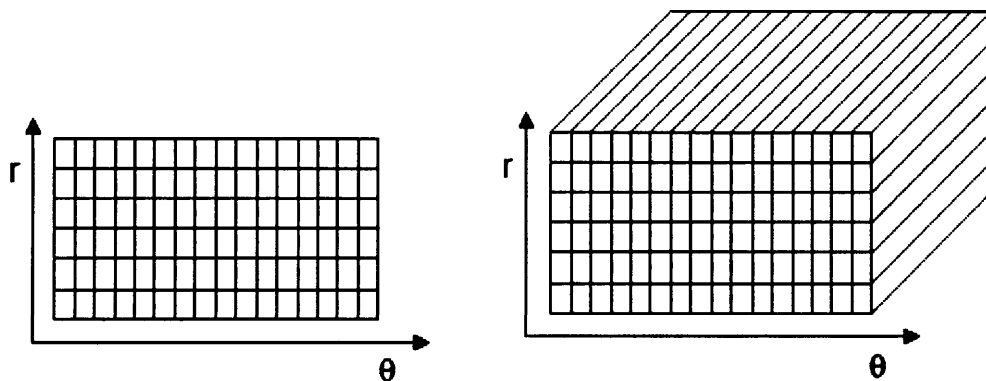
The following figure shows the structure of a shape context with 12 angular and 5 radial distances along with the resulting radial-angular plot of distances:



**Figure 3-12: Shape Context**

The same structure has been adopted for this research to retaining the colour distribution in spatial domain using histogram, hence the name *Colour Context Histogram*. Instead of using the structure at the edges of an object the shape context structure is placed at the centre of the object and histograms are created for each radial-angular cell. Each resulting histogram is concatenated generating a single description of the colour distribution.

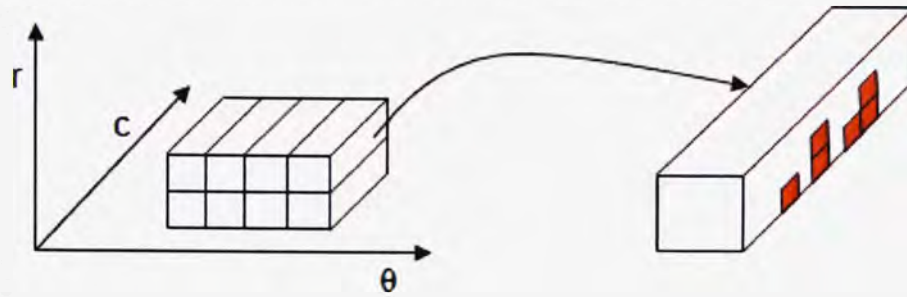
Shape context is 2D structure. To create *Colour Context Histogram*, a third colour dimension is added to the shape context structure. This results in a 3D structure containing radial-angular colour distribution as shown below:



**Figure 3-13: Extending Shape Context to Include Colour**

The third dimension data is simply a normalised colour histogram corresponding to a particular radial-angular region.

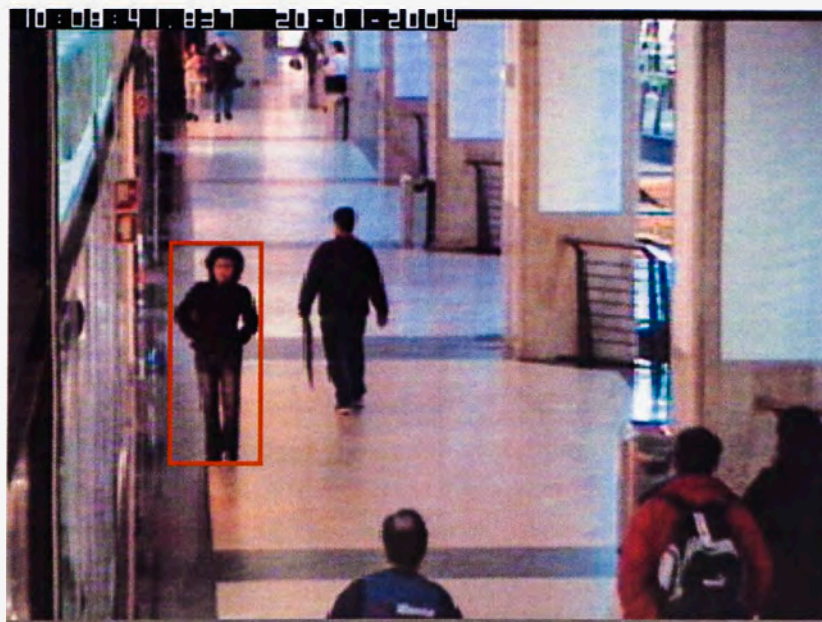




**Figure 3-14: Colour Shape Context – Colour Dimension**

In the implementation, the radial-angular histograms are concatenated to form a single histogram.

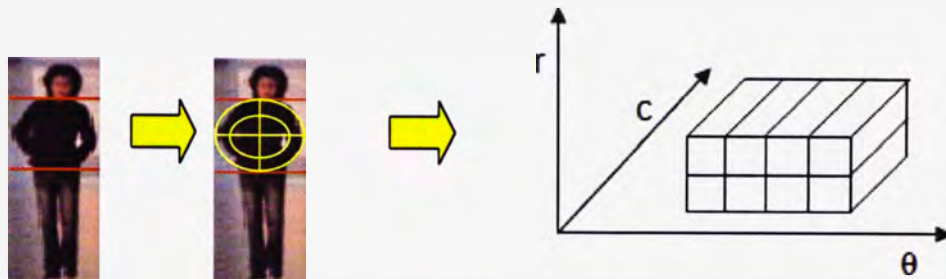
In order to create a descriptor for torso given an image with bounding box, the first step is to identify the torso and legs regions as described earlier.



**Figure 3-15: Person with Bounding Box**



Once the torso has been identified, the shaped context structure is placed on the torso and *Colour Context Histogram* is created as illustrated below:



**Figure 3-16: Creating Colour Context Histogram**

In the above figure, the 8 colour bins correspond to the 8 regions in torso created by using shape context structure. It should be noted that because of the radial structure used in creating *Colour Context Histogram* ignores the regions on the edges. This helps in rejecting any background region in the bounding box and concentrate on the torso region alone.

### 3.6.2 Lower Body Descriptor

More often than not, the bounding box around a person will contain some background area. This is especially true of lower section of the bounding box because of shape and gait of people. Some of the examples are shown below:



**Figure 3-17: Unwanted Background around Legs**

As it can be seen from the examples above, the background tends to occupy a lot of space in the lower section of the bounding box. While including the background information in the descriptor may work in simple scenarios, it will cause lots of errors in more realistic situations. So for example, re-identifying a person in the same scene and in the same vicinity may work, but it will not work where the scene has different backgrounds or the same person appears in another scene where the background is drastically different. Therefore, if the background is not rejected while creating the descriptor, it will generate many errors.

So to summarise, the descriptor for lower section should:

- Reject the background
- Be invariant to pose of the person
- Be invariant to location of person in the bounding box
- Be invariant to the scale

Given that we reject the background, a colour histogram will be invariant to the pose of the person. So for a person with legs together or spread apart, a colour histogram will give us a consistent descriptor. Similarly, a colour histogram will be invariant to the location of a person in the bounding box as long as there are no background pixels. Furthermore, normalising the colour histogram will facilitate scale invariance. Hence if we successfully reject the background, using a colour histogram will meet the conditions given above for the lower section descriptor.

Depending on the pose, the legs may contribute more or less colour information to the histogram. But nevertheless, it is important to recover the colour of legs' pixels in order to make the descriptor more discriminative. The main challenge of modelling the legs area is to correctly identify the valid pixels while ignoring the background. It is important to find the colour boundary where a torso stops and legs begin. But a pre-requisite for using this method is to identify the torso and bottom or legs of people in an image.

There is no way to accurately identify the separation between the torso and legs but we can try to get as close to it as possible. However, it should be noted here that we are not looking for the separation between torso and legs in anatomical sense but only in terms of colour separation. This colour separation will result from the difference between the colour of torso clothing and trousers. But this separation might not be very clear when the colour of torso clothing and trousers are similar. In this case, we need to get a line where the colour dissimilarity is the largest.

The difference in the colour is caused by shirts, jackets, jumpers, coats etc. Some of these cases are shown with the separation between torso and legs marked.



**Figure 3-18: Colour Separation between Torso and Legs**

The main concern in getting the colour information of legs and torso separately is to effectively separate the legs and torso. To find this separation, this research proposes two approaches described in the following sections.

### **3.6.2.1 Proportional Area Back-Projection**

In order to get the legs area, the first step is to get the proportions of torso and legs as described before in section 3.5. Once the lower section of bounding box containing the legs has been identified, the next step is to segment the legs from the background. Given that this research does not require any background frame, the only information available is the colours and shape in the lower bounding box. Shape of the legs depends on the pose and gait of the people and hence is more complex to use. In the first proposed method, colour information is used to identify and segment the legs. Histogram back-projection [1] is used to reveal the legs pixels.

A colour histogram of a portion of legs area is constructed and back-projected on to the lower section of the bounding box to segment the legs' pixels. There is no post processing used after back-projection. The best portion to create the colour histogram from is the centre of the bounding box. The idea is to capture the most likely proportional area of the legs. In the current research, 25% of the width of the bounding box and 25% of the height of the lower section of the bounding box is used. Experiments have shown this to work in most of the cases.

The following figure shows the portion of the bounding box used for creating colour histogram for back-projection:





**Figure 3-19: Background Rejection using Proportional Rectangles**

The white boxes show the areas picked for creating the colour histogram for back-projection. After the back-projection, the resulting segmentation of legs is shown in the image below:



**Figure 3-20: Foreground Segmentation using Proportional Rectangles**

The effectiveness of this method depends on the correct identification of torso and legs area in terms of colour separation. The proportions used for separating torso and legs may not apply to the colour boundary between legs and torso. In the cases where there

the proportions are correct in terms of colour separation or there is little difference between torso and legs colour, this method works fine. But where these criteria fail, the segmented images are noisier because the histogram of legs area will contain colour from torso. If there is a large difference between torso and legs colour, back-projection will not correctly highlight the legs' pixels.

### 3.6.2.2 Scanned Area Back-Projection

In the second proposed method, the above method is modified to find the best colour separation between the torso and the legs. If the colour of torso and legs are similar, then this modification will not harm the results. But if there is a separation of colours, then this method will provide the best colour boundary between torso and the legs. A scan of the area in the vicinity of torso is made to find the largest colour mismatch. This technique empirically selects a rectangular area in the centre of the bounding box that has width and height equal to 25% of the width of the bounding box and 20% of the height ( $H_t$ ) of the torso region. An area identical in width and height is selected under it. The scan starts at  $0.4 * H_t$  above the torso line and stops at  $0.8 * H_l$ , where  $H_l$  is the height of legs region. Colour histogram of both top and bottom rectangular areas are created and Bhattacharyya coefficient is computed. Then the two rectangular areas are moved down and the Bhattacharyya coefficient is calculated in similar fashion. This process is repeated till the end is reached. This process is shown in the figures below:

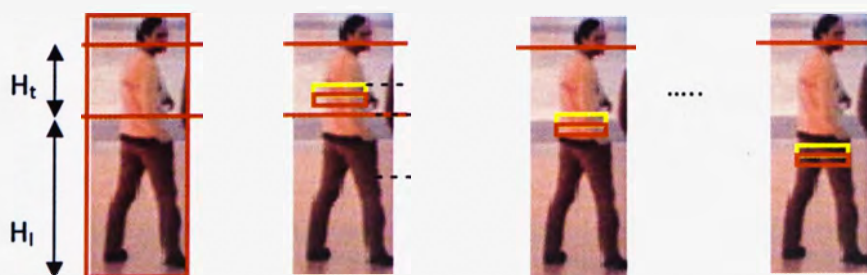


Figure 3-21: Background Rejection using Scan Search

The top and bottom regions generating the lowest Bhattacharyya coefficient gives us the least similarity in colour which indicates the best separation between torso and legs. So the line between these two regions is considered to be the new torso line. Any histogram matching technique could be used but we opted for Bhattacharyya coefficient.

### 3.6.2.3 Segmentation Selection

Once both the techniques are applied for segmenting the legs' pixels, we need to pick the best segmented image. A Naïve Bayes Classifier (NBC) is trained on different segmented images of legs for selecting the best segmented image. Once this has been done, then the colour pixels corresponding to the segmented pixels are used to extract the colour information of the legs. The classifier is trained on a single class only. The class here represents a segmented legs image. Both segmented images are passed through the classifier and the one with the highest probability value is selected to be the best segmented image.

NBC uses Bayes' theorem with an assumption that there are no conditional probabilities. This means that a feature of a class is not dependent on any other feature of the class. It uses Maximum Likelihood for parameter estimation. NBC has the advantage of ease of training and requiring a small training data to estimate the parameters of features used. Using Gaussian distribution, these parameters are the mean and variance of features. NBC uses *Maximum A Posteriori* (MAP) for classification as shown below:



$$p(C \mid f_1, \dots, f_n) = \arg \max p(C = c) \prod_{i=1}^n p(F_i = f_i \mid C = c)$$

Equation 3-9

The task of classifier is to identify a class given a list of features. NBC uses probability to accomplish this. On the left hand side of the above equation, we are asking a question about the class: Given a list of features, what is the most likely class? We find the answer by finding the product of probabilities of a particular feature in a certain class and the probability of the class.

To train the classifier, the spatial distribution of pixels is taken into consideration. Each image is divided into  $n$  different sections horizontally. In the current research,  $n$  is empirically selected to be 8. Horizontal division of the image was selected to make the foreground invariant to the location of the legs in the image as shown below:


Segmented Legs			Number of Pixels
			120
			126
			124
			118
			108
			111
			112
			78

Table 3-1: Location Invariance of Legs

The above segmented images have different locations for legs. The width of ROI for the first image is larger than the middle one. Similarly the ROI for the third image does not centre the legs properly. All of these are common scenarios one might expect in a practical situation. We can never assume properly centred and scaled ROI. So the



horizontal division of segmented legs provides invariant to the location of the legs in the image.

The image is made scale invariant in y-direction by dividing the number of pixels per division by the height  $h$  of the division. We do not want to use width  $w$  because it will scale the image per division area making it dependent on the width of ROI. The following tables show that we get the same number of pixels per division height  $h$  for scale variations:

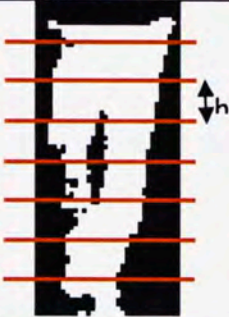
Segmented Legs	Number of Pixels / h
	24
	25.2
	24.8
	23.6
	21.6
	22.2
	22.4
	15.6

Table 3-2: Scale Invariance of Legs

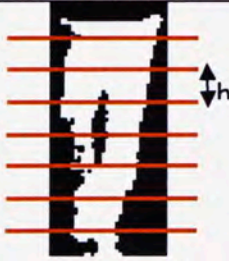
Segmented Legs	Number of Pixels / h
	24
	25.2
	24.8
	23.6
	21.6
	22.2
	22.4
	15.6

Table 3-3: Scale Invariance of Legs

The number of pixels per  $h$  serves as an input to the classifier. Therefore the NBC uses these 8 features to learn their probabilities using Gaussian distribution. For a Gaussian distribution, we need to find the mean and variance of these features. When two

images are passed through the classifier, the image the highest MAP is selected to be the best segmented image. Using the example above, the following table shows the calculation of MAP:

Input	Z	$p(F_i = f_i   C = c)$
1	24	0.9
2	25.2	0.95
3	24.8	0.5
4	23.6	0.6
5	21.6	0.7
6	22.2	0.9
7	22.4	0.95
8	15.6	0.9

**Table 3-4: Finding Maximum A Posteriori**

In the above table the conditional probability is found by using Gaussian distribution:

$$p = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[\frac{-(z-\mu)^2}{2\sigma^2}\right]$$

Equation 3-10

For each input, the mean  $\mu$  and variance  $\sigma^2$  are found during training. Then the conditional probability for input z is found using the Gaussian distribution formula given above. Using , we get:

$$p(C | 24, \dots, 15.6) = 0.9 * 0.95 * 0.5 * 0.6 * 0.7 * 0.9 * 0.95 * 0.9 = 0.138$$

It should be noted that  $p(C = c)$  has no significance in our case we have a single class classifier. Therefore this probability is left out of the calculations. For both segmented images from the two methods, the inputs are run through NBC and the image maximizing the posterior probability is chosen as the best segmented image.

Once the legs' pixels have been identified, a colour histogram of those pixels is created. Because of the different pose of people, the position of legs is different. Hence we need to ignore the location or position of the legs. Since a histogram does not capture the spatial distribution, it is well suited for this situation. To make the histogram scale invariant, it is normalised.

#### 3.6.2.4 Simple Histogram

A derivative of *Colour Context People Descriptor* was also used in the research. A simple bottom histogram was used instead of segmented one described in sections 3.6.2.1 and 3.6.2.2. This helped compare the performance of segmented and non-segmented histograms across the datasets. This version of *Colour Context People Descriptor* is referred to as *Hybrid Colour Context People Descriptor*.

### 3.7 Conclusion

There are many types of colour spaces available. Instead of using just one colour model, one should consider many different colour models. By using more than one colour model, we can understand the effects of colour on re-identification. In this

research, four colour spaces, RGB, HSV, YIQ, XYZ, have been used. Once the colour space has been selected, the next step is to model the colour such that it can be used for matching purposes. Histogram provides a lightweight way of encapsulating the colour information. The following are the main points pertaining to colour histograms:

- Well established model
- Inherently invariant to rotation
- Invariant to scale if normalised
- Fast to compute
- Many matching techniques for comparison

Rotation invariance is a major attribute of colour histogram but it comes with a cost.

Rotation invariance discards spatial distribution of colours. Hence a colour histogram is unable to preserve colour distribution in the underlying image. Colour distribution is an important factor to consider in re-identification process. It helps differentiate similarly coloured objects and provides better accuracy.

Using colour as the only appearance feature, creating a robust appearance descriptor is of prime importance. Histogram is an efficient way of encoding the colour information but lacks the ability to preserve spatial scattering of colour. A simple colour histogram of body has already been used in the relevant research. As a step towards adding spatial colour information to the simple histogram, two other methods have been proposed.

The first method is to separate the torso colour histogram from bottom histogram. This method is referred to as Top-Bottom histogram. The resulting histogram is concatenation of top and bottom histograms.

*Colour Context People Descriptor* has been described as the second method of including spatial colour distribution. It is created by concatenating *Colour Context Histogram* of the torso and the histogram of the legs using background rejection. Because of the wide variety of leg poses, a need to reject the background arises. A methodology has been proposed for legs' that rejects background by using back-projection to reveal the pixels of legs. Once the background has been identified, only the legs' pixels are used to create the legs' colour histogram. Furthermore, pose and scale invariance is provided by using normalised colour histograms. Finally, the location of people in the bounding box is ignored when classifying the legs' pixels. This makes the legs location independent with respect to the bounding box. *Colour Context People Descriptor* effectively separates the colour of the torso from the legs. Furthermore, the descriptor takes into account the spatial distribution of the colours in the torso area with scale invariance.

# 4 Experiments

---

An incremental approach was taken during the experimentations. First round of experiments used a histogram of people in the bounding box. This is the approach described in the literature review and does not include any spatial information of colour distribution. The second stage was to separate the torso colour histogram and lower region colour histogram to include spatial distribution of the colours. The third phase was to replace the lower region colour histogram with legs' colour histogram obtained by using the back-projection described earlier. Finally, the torso colour histogram was replaced by *Colour Context Histogram* which resulted in the *Colour Context People Descriptor*. Bhattacharyya coefficient was used to match the histograms. Different parameters were used during the experimentation. Four colour models, namely, RGB, HSV, XYZ and YIQ, were used. For histogram bins, 3 different sizes of 3x3x3, 8x8x8 and 12x12x12 were used. For colour context histogram, radial distances of 1 and 2 were used and angles of 4, 6, 10 and 12 were used.

The results could not be compared with other related research as none of them have used a public data set.

## 4.1 Parameters

There were a total of 132 experiments performed on all three datasets. The following four colour spaces were used:

- RGB
- HSV
- YIQ
- XYZ

The above colour spaces were used in the following four methods during the experimentation:

- Histogram (H)
- Top-Bottom Histogram (TBH)
- Top-Bottom Back-Projection Histogram (TBBPH)
- Colour Context People Descriptor (CCPD)
  - Proportional Area Back-Projection / Scanned Area Back-Projection
  - Simple Histogram - Hybrid Colour Context People Descriptor (HCCPD)

The Histogram method creates a normalised colour histogram of person in the bounding box. The histogram bin sizes of 3x3x3, 8x8x8 and 12x12x12 were chosen for the experiments. The combination of colour spaces and the sizes of bins created a total of 12 different histograms. The normalised histograms of two persons are matched using Bhattacharyya Coefficient.

The Top-Bottom Histogram creates two normalised colour histograms. The first one is for the top of the person and the second one is for the bottom of the person. The process of creating the histogram is described in section [3.5]. During the matching process, the top histograms of two people are compared using Bhattacharyya Coefficient. Similarly, the bottom histograms are matched. The total matching score is obtained by adding the Bhattacharyya Coefficient for top and bottom histograms. This also results in 12 different Top-Bottom histograms.

The Top-Bottom Back-Projection Histogram is similar to Top-Bottom Histogram except that it uses background rejection as described in section [3.6.2]. The method and score calculation is also the same. There are 12 different Top-Bottom Back-Projection histograms.

*Colour Context People Descriptor* is described in section [3.6]. Bhattacharyya Coefficient is used for matching the normalised colour histograms. The total matching score for two different *Colour Context People Descriptors* is obtained by matching the histograms of each region in the torso and then legs. Apart from the 4 colour spaces and 3 histogram bin sizes mentioned earlier, the descriptor used 2 radial distances of 1 and 2 and 6 angles of 4, 6, 10 and 12. This resulted in 96 different *Colour Context People Descriptors*.

Hybrid Colour Context People Descriptor is same as Colour Context People Descriptor except that the lower body histogram is a simple histogram as used in Top-Bottom Histogram. The need for this descriptor arose after first round of experimentations. It



was noticed that Colour Context People Descriptor performed best in all datasets except NICTA. Further analysis revealed that there was a big performance difference between Top-Bottom Histogram and Top-Bottom Back-Projection Histogram methods. Since the only difference between these two methods is the foreground segmentation, the hypothesis was that the segmentation did not work properly in NICTA dataset. To evaluate this hypothesis, Hybrid Colour Context People Descriptor was used. It proved the hypothesis to be correct as using this hybrid descriptor improved the results in case of NICTA dataset.

For each dataset, experiments were carried out to find average, best and worst ROC curves per method. For each of the four methods, the average ROC curve was calculated for all colours. This gives us classification method performance indication without taking into account the affects of a different colour spaces. For best and worst ROC curves, these experiments points out the best set of parameters such as colour space, histogram bin size etc.

In the analysis sections, the average, best and worst ROC curves were compared to examine the performance of different methods. The average ROC curves ignore colour space affects whereas the best and worst ROC curves include parameter information. It then compared the best techniques by charting the number of cases, i.e. individual people from dataset, in which a specific method performed the best. Similarly, it also compares the worst techniques where a certain method was the worst for a case. Furthermore, a similar comparison of best and worst colour spaces was done to study the effects of colour spaces.

In conclusion, all four methods are compared for each dataset using ROC curves. The best and worst techniques and colour spaces for each dataset are also compared.

## **4.2 Datasets**

There were two primary datasets used for experimentation. The first dataset is the (Context Aware Vision using Image-based Active Recognition) CAVIAR dataset [11]. This dataset is available for public download. It consists of indoor shopping mall sequences. The people in the sequences are hand-labelled. The ground truth for the CAVIAR dataset is available in XML format. The ground truth consists of centroid location and length and width of the bounding box. During the experiments, the ground truth data was used to get the bounding box around a person. The second dataset was generated by recording indoor video at one of NICTA's research labs. To get the ground truth for NICTA dataset, Meanshift tracking algorithm [16] was used. The ground truth extracted contained centroid location and length and width of the bounding box. The algorithm was manually initialised. A third dataset was generated that combined all the images from both CAVIAR and NICTA datasets.

The reason for extracting the ground truth using tracking was to mimic a practical scenario. In real-life situations, a tracking algorithm is often used for tracking people. The colour information of people being tracked can then be used for matching. Meanshift, being a popular tracking algorithm, has been picked for creating the NICTA data set.

### **4.2.1 CAVIAR Dataset**

The dataset was manually pruned to identify 21 different people over 3987 images. A unique id is assigned to each person for identification purposes. The following table shows the people used in the experimentation:









								
Id	1	3	4	5	33	8	9	34

Table 4-1: People in CAVIAR dataset

								
Id	13	14	15	19	20	21	22	23

Table 4-2: People in CAVIAR dataset




					
Id	25	27	29	32	36

**Table 4-3: People in CAVIAR dataset**

### 4.2.2 NICTA Dataset

Another dataset was created at NICTA by recording indoor video. These videos were not hand-labelled. The ground truth was obtained by implementing the Meanshift tracking algorithm [16]. The meanshift algorithm was manually initialised by selecting a bounding box around the person of interest. Any of the background subtraction techniques could have been used to initialise the model for meanshift tracking. This ensured that the algorithm is tested on data that will be obtained by conventional computer vision algorithms instead of hand labelling, which is impractical.

For the NICTA dataset, 12 different people were identified over 3022 images and given a unique Id for recognition. The following table shows the people used in the experiments:

						
Id	1	2	3	4	5	6

**Table 4-4: People in NICTA dataset**

						
Id	7	8	9	10	11	13

**Table 4-5: People in NICTA dataset**



### 4.2.3 NICTA-CAVIAR Merged Dataset

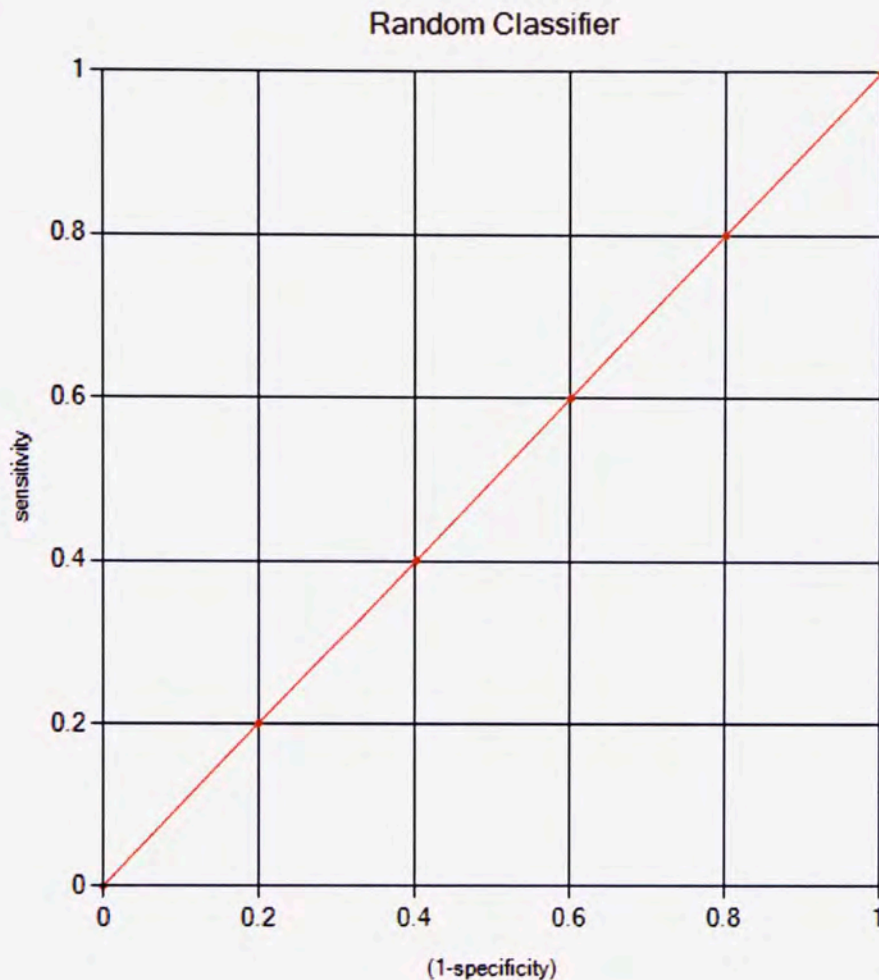
For a third data set, the NICTA and CAVIAR data sets were combined. This data set included 33 different people over 7009 images.

## 4.3 ROC Curves

Receiver Operating Characteristic or ROC curves are commonly used for evaluating the performance of classifiers. To plot an ROC curve, we need True Positive Rate (TPR) and False Positive Rate (FPR). TPR on Y-axis is plotted against FPR on X-axis. An alternatively terminology for TPR and FPR is Sensivity and  $(1 - \text{Specificity})$  respectively. In this thesis, the later terminology is used for plotting ROC curves.

The point  $(0,1)$  on ROC curve represents a perfect classifier. Such a classifier will identify all positive and negative samples correctly. This can be further understood by realizing that at this point, FPR is 0 while TPR is 1. On the other side of the spectrum, the point  $(1,0)$  on ROC curve corresponds to a classifier that incorrectly identifies all samples. At this point FPR is 1 and TPR is 0. A classifier at point  $(0,0)$  identifies all samples as negatives and point  $(1,1)$  will classify all samples as positive.

As a rule-of-thumb, north-west corner of the ROC curve represents good classifiers. A random classifier will identify half samples correctly and other half incorrectly. The figure below shows an ROC curve for a random classifier:



**Figure 4-1: ROC for Random Classifier**

In order to find the TPR and FPR, for each candidate, the whole data set is searched and matching is performed. All the results are ranked according to their scores. Then the results are checked by divided them as percent of entire data set. The process of creating ROC curves is described in the paragraph below.

Let us consider we have a data set of 100 images. Before conducting the matching process, we have to decide the scope of ROC curves. The scope is used to find the number of matches as a function of section of the data set. When all the results are obtained, top X% of the results are checked to find the actual match. For matching,

each image is matched to the rest of the images. The score of the results are ranked from most to least. Then for each scope, we count the number of correct i.e. True Positives (TP) and incorrect i.e. False Positives (FP) matches. For instance, for scope of 10%, we search the top 10% of the results and count the number of TPs and FPs. Then TPR is found by dividing TP by Total Number of Actual Positives in the dataset. Similarly, FPR is found by dividing FP by Total Number of Actual Negatives in the dataset. A point corresponding to intersection of TPR on y-axis and FPR on x-axis is plotted on the ROC curve. Similarly, all other points on ROC curve are found by increasing the scope.

In this research, scope of 5%, 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90% and 100% were used.

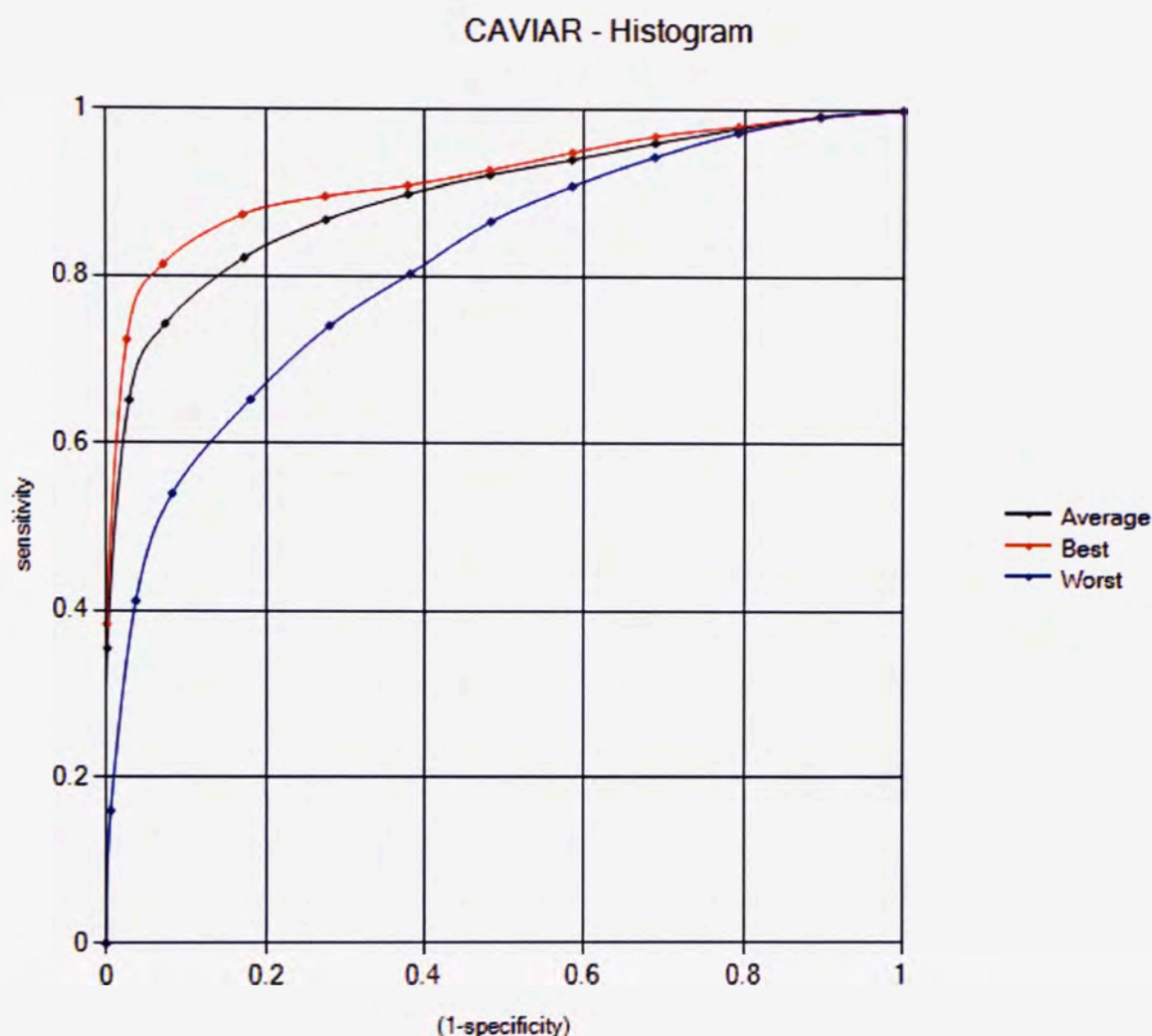
## **4.4 CAVIAR Dataset Benchmark**

The first set of experiments was conducted on the CAVIAR data set. The four methods described earlier were evaluated and their results were analysed. It was clear from the results that adding spatial colour information increases the performance on re-identification.

### **4.4.1 Body Histogram**

Using the bounding box around a person, a normalised colour histogram of the area is created. The following figures show the ROC curve for body histogram. The best ROC curve was for an HSV histogram with 8x8x8 bins and the worst ROC curve was for YIQ histogram with 3x3x3 bins.



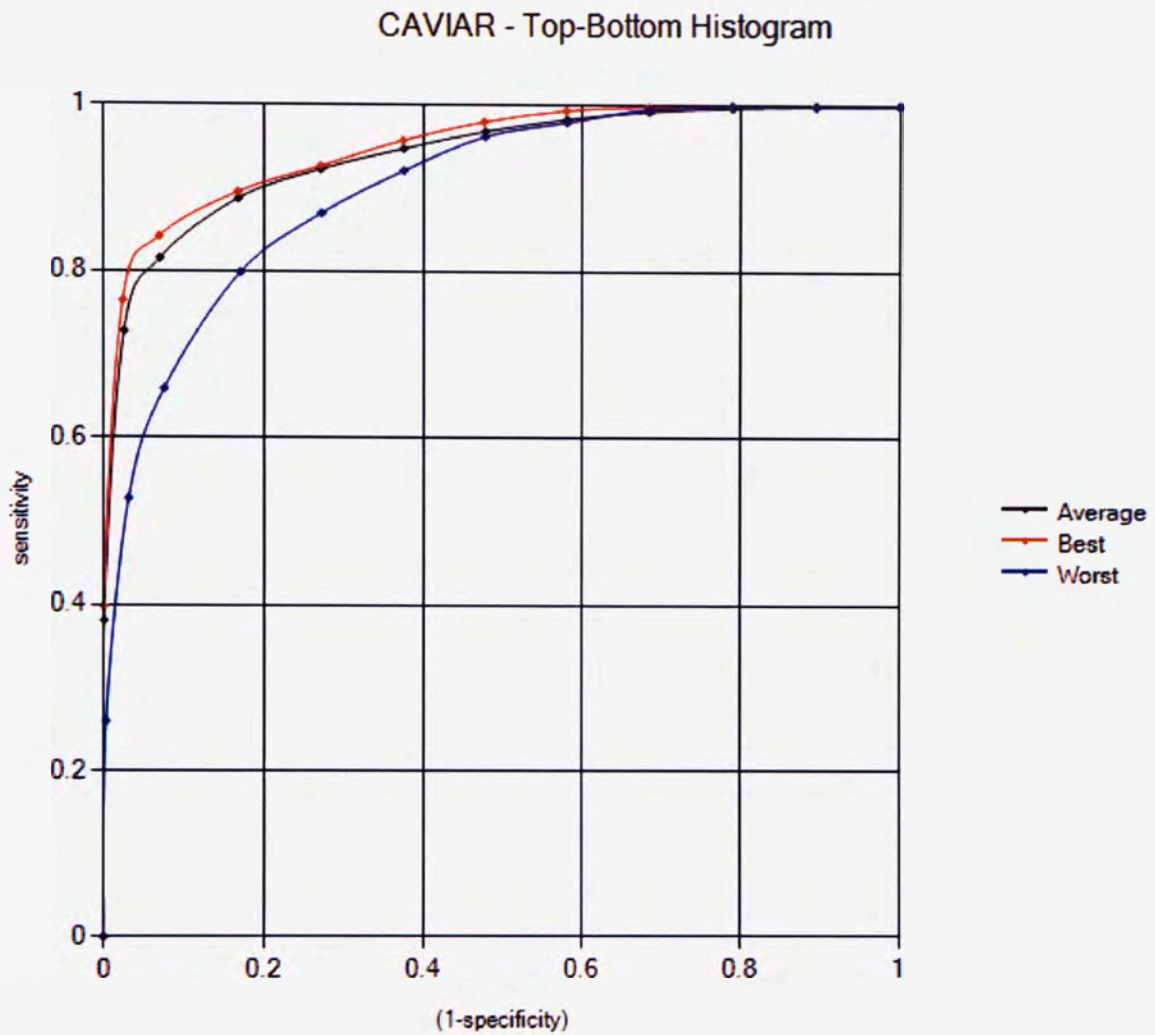


**Figure 4-2: CAVIAR - Body Histogram ROC**

One can see the big difference between the worst and best results. This indicates that using body histogram will give us unpredictable results as it does not seem to be stable way of re-identification.

#### 4.4.2 Top-Bottom Histogram

Using the bounding box around a person, a top-bottom histogram is created. This is first step towards incorporating the spatial information of colour. The ROC curve is shown below:



**Figure 4-3: CAVIAR - Top-Bottom Histogram ROC**

One can immediately note the closing gap between the best and worst results. This indicates that separating the torso and legs histograms makes the top-bottom histogram more stable for re-identification purposes. The best and worst results are achieved by using XYZ colour space with 12x12x12 histogram and YIQ colour space with 3x3x3 histogram respectively.

### 4.4.3 Top-Bottom Back-Projection Histogram

This method is an extension of the previous method. The best results were obtained by using HSV colour space with histogram bins 8x8x8. Whereas, YIQ colour space with a histogram of 3x3x3 bins performed the worst as shown below:

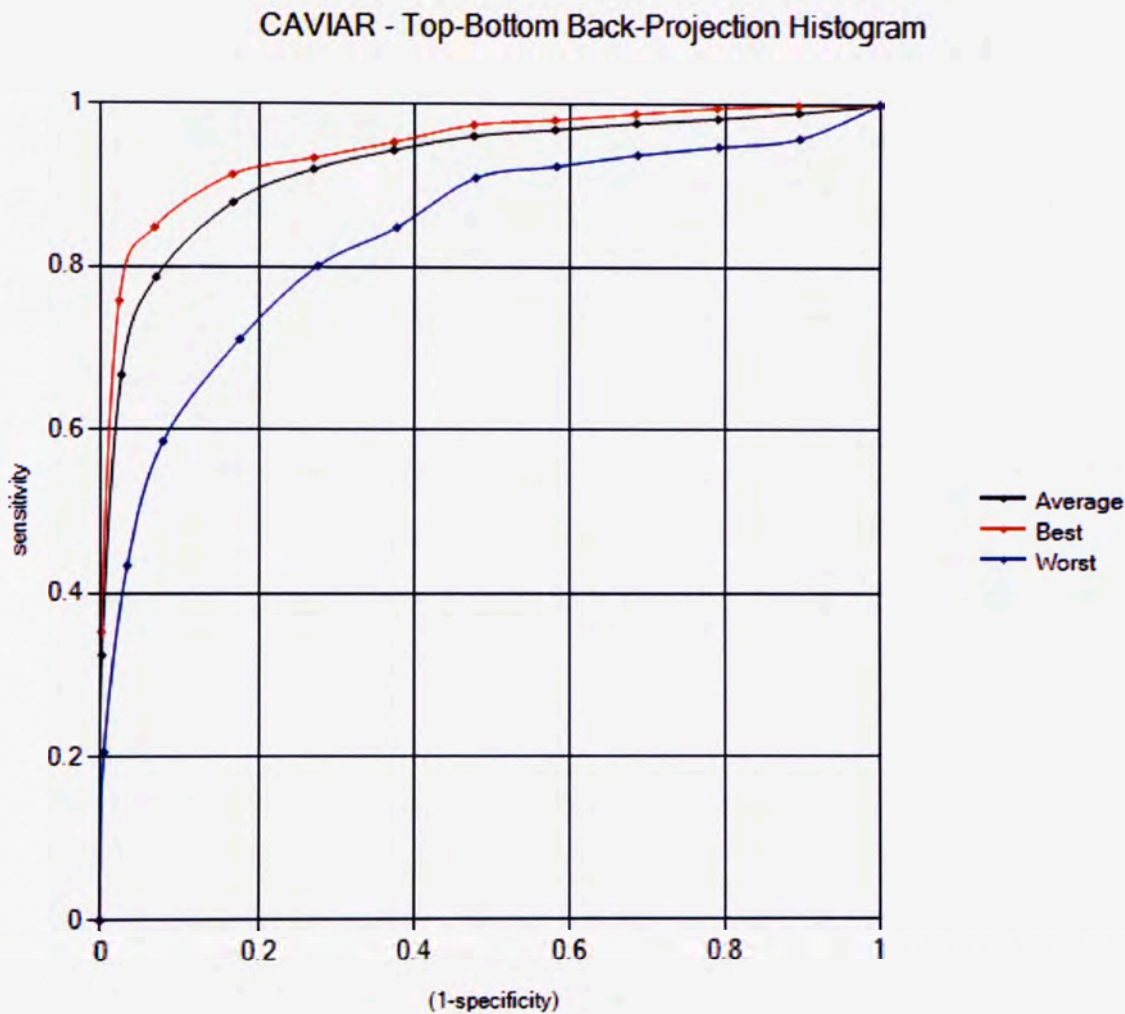


Figure 4-4: CAVIAR - Top-Bottom Back-Projection Histogram ROC

Again the difference between the best and worst case scenario using top-bottom back-projection histogram is better than the simple body histogram method. However, using the background rejection methodology described in section [3.6.2] has affected the



results. Top-bottom histogram method seems to perform better as compared to this. This indicates that the classification of foreground pixels is not good enough to work in different images.

#### 4.4.4 Colour Context People Descriptor

This method described in section 3.6 adds more spatial colour information to create a descriptor. The best descriptor consisted of a YIQ histogram with 12x12x12 bins, 6 angles and a single radial distance. The worst performing descriptor was also in YIQ colour space with histogram of 3x3x3 bins, 4 angles and a single radial distance. The results are shown in the figure below:

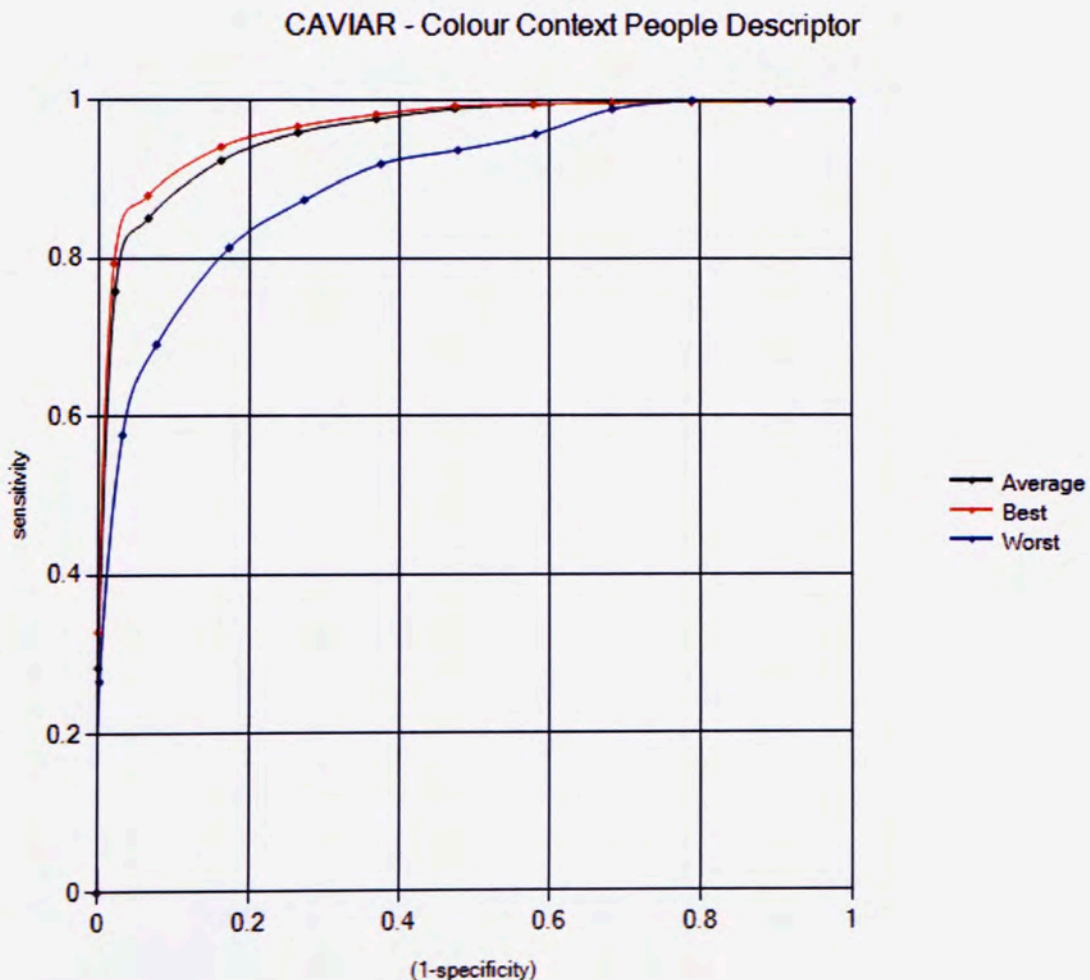


Figure 4-5: CAVIAR - Colour Context People Descriptor ROC

Once again this descriptor is more stable than body histogram as it contains spatial distribution of colour.

#### 4.4.5 Hybrid Colour Context People Descriptor

The best descriptor consisted of a YIQ histogram with 12x12x12 bins, 6 angles and a single radial distance. The worst performing descriptor was also in YIQ colour space with histogram of 3x3x3 bins, 4 angles and a single radial distance. The results are shown in the figure below:

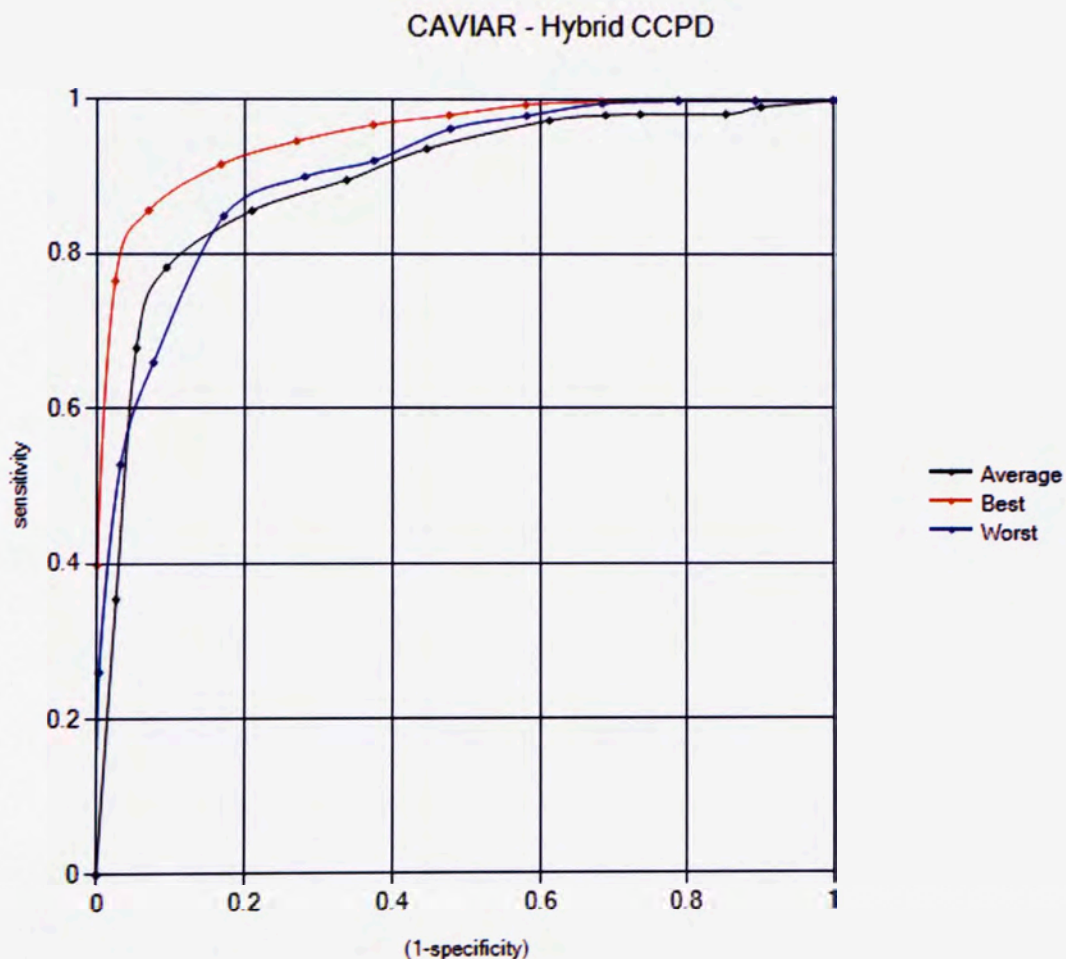
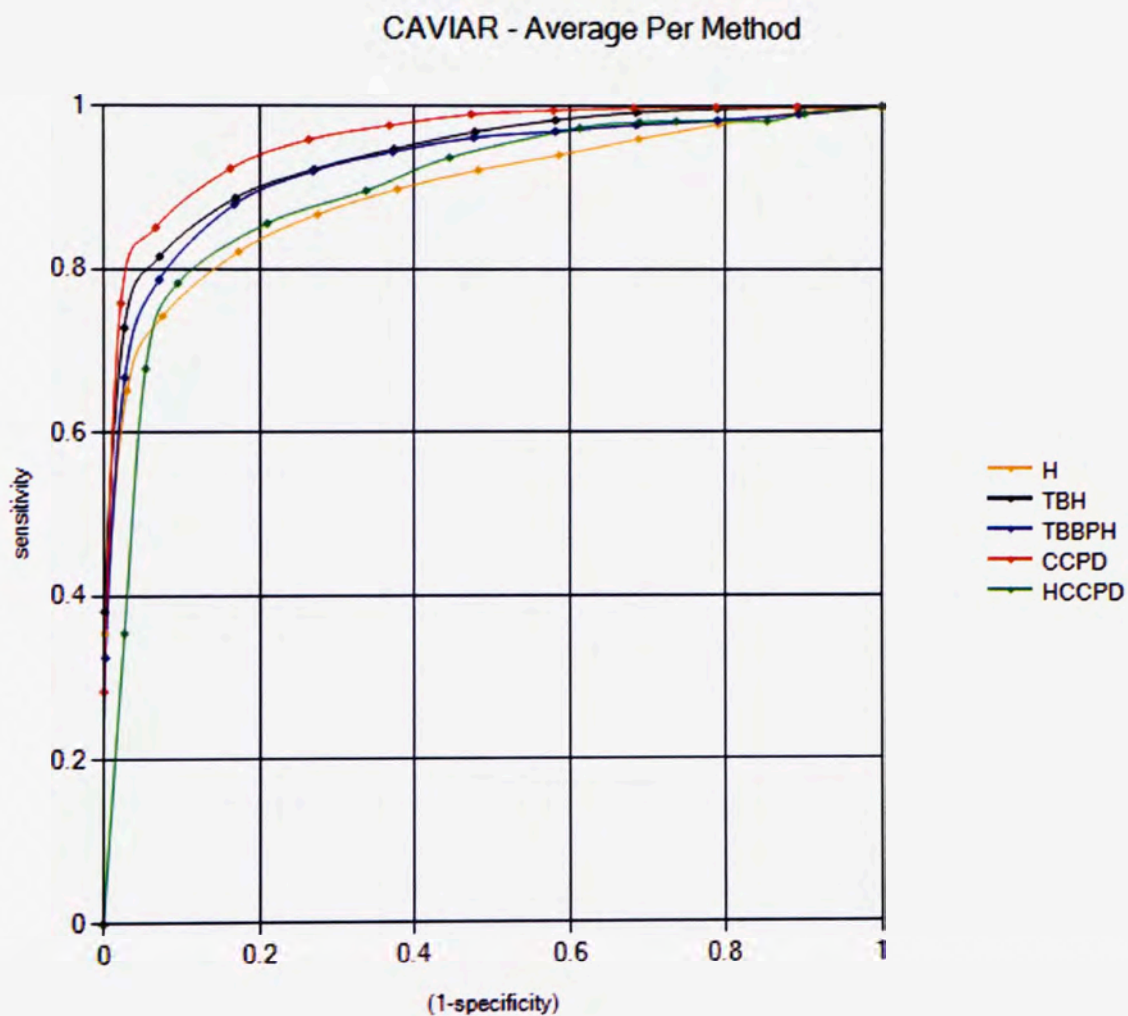


Figure 4-6: CAVIAR – Hybrid Colour Context People Descriptor

#### 4.4.6 Analysis

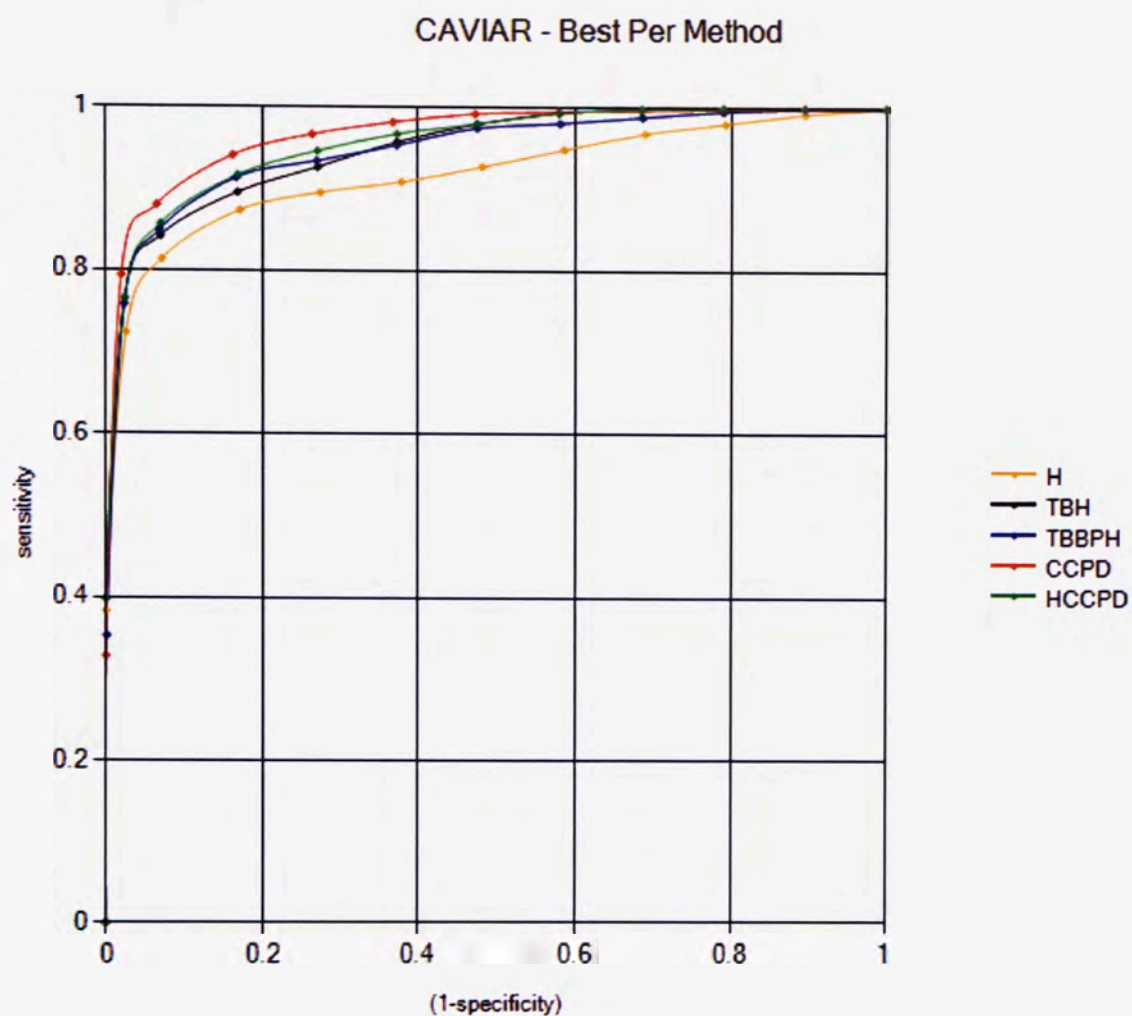
We have already seen that adding spatial colour information improves the classification performance. In this section, the average, best, worst results are shown along with colour space results. By looking at the average ROC curve below, one can observe that overall the methods that include spatial distribution of colours perform better than body histogram method. Therefore, we can conclude that the re-identification performance can be increased by adding spatial colour information.



**Figure 4-7: CAVIAR - Average ROC per Method**

Similar results can be seen for the best performing cases as shown in the images below:

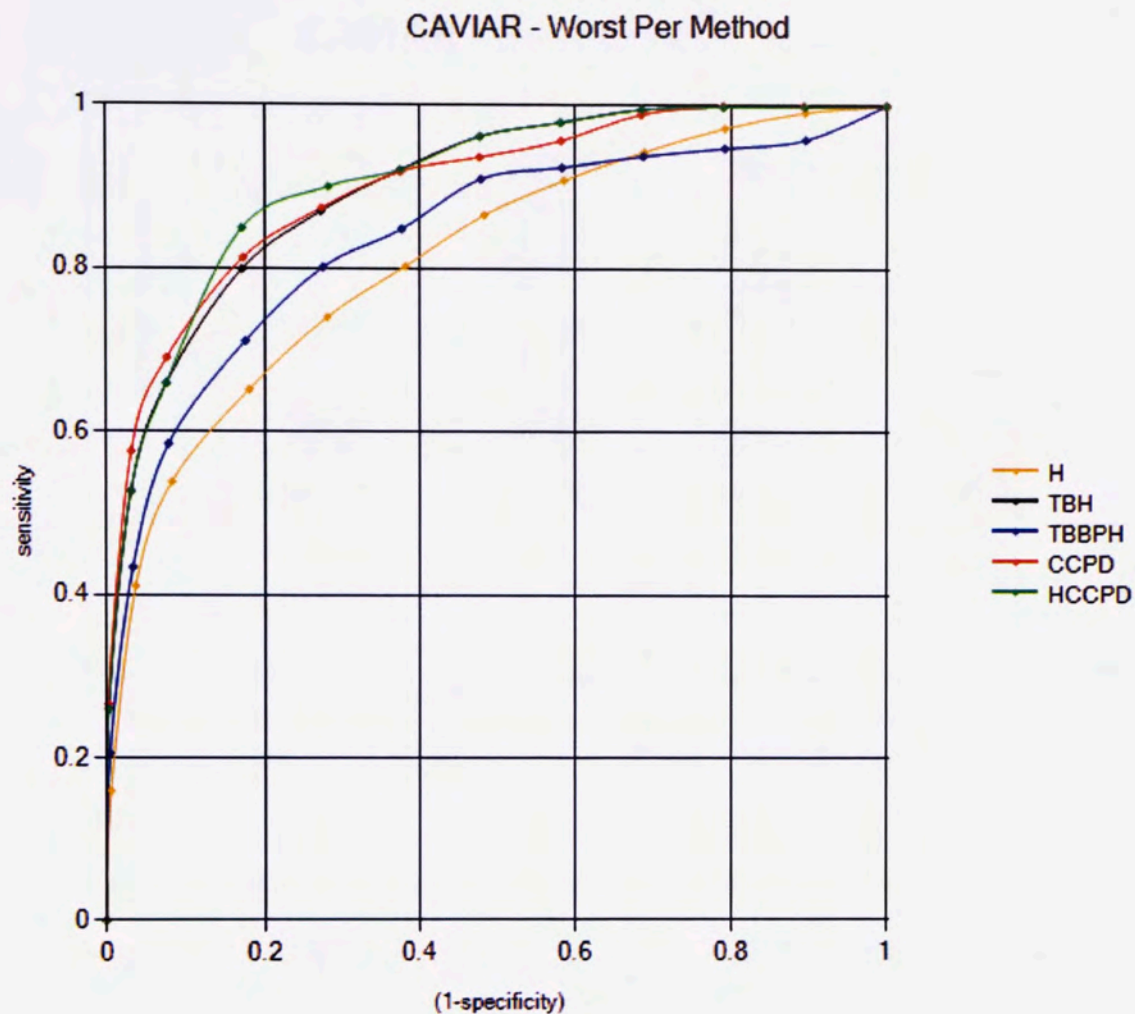




**Figure 4-8: CAVIAR - Best ROC per Method**

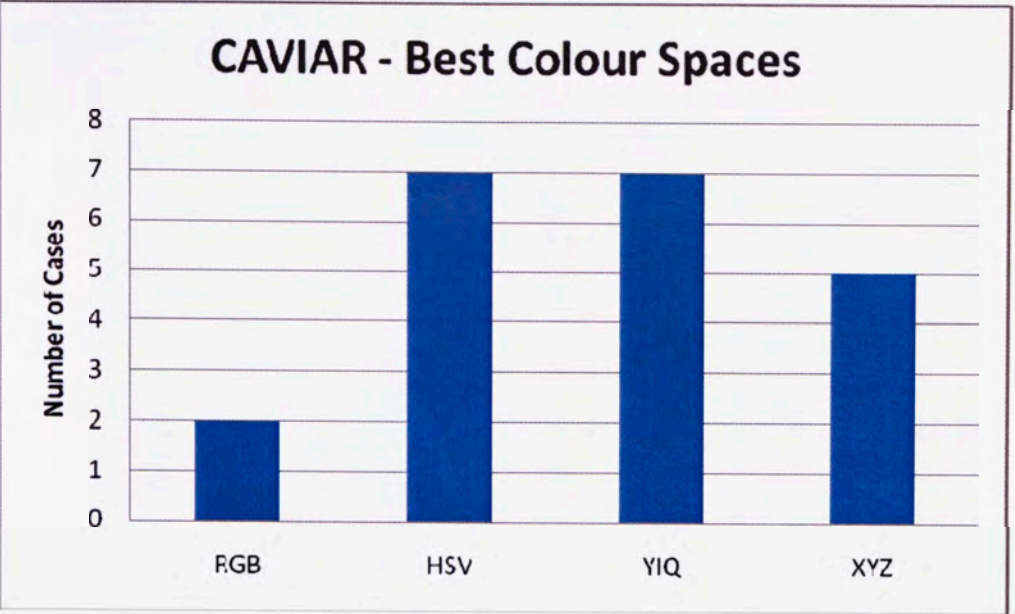
The difference of performance is more obvious when looking at the worst case scenarios. As compared to body histogram, *Colour Context People Histogram* improves the true positive rate by about 75%. This is shown in the images below:





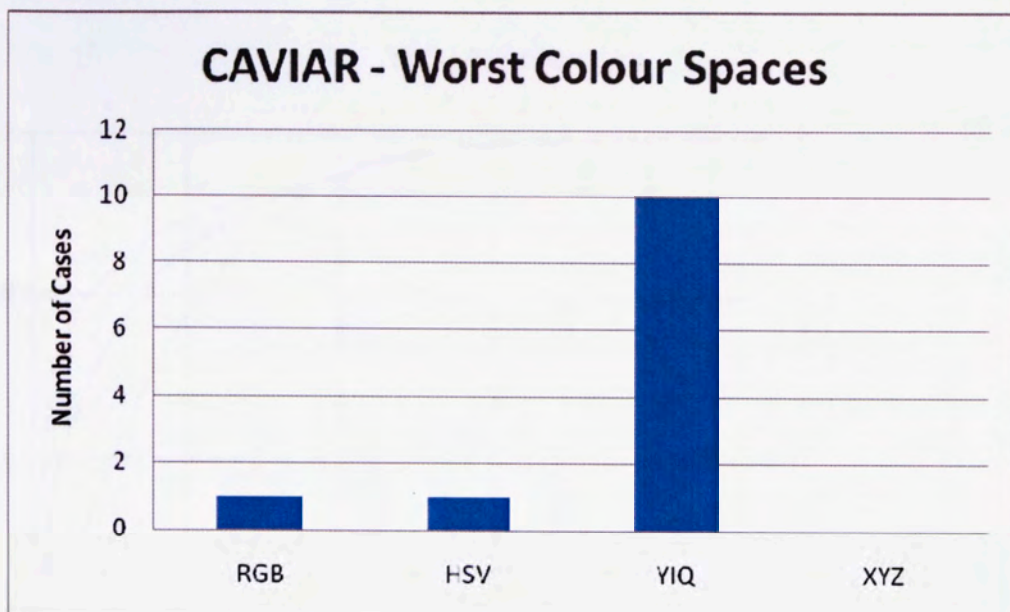
**Figure 4-9: CAVIAR - Worst ROC per Method**

In the colour space, HSV and YIQ were the best colour spaces to use. They were followed by XYZ and RGB. In the worst colour spaces, RGB and HSV performed the least in 1 case each. YIQ was the worst in 19 cases and XYZ in none. The figure below shows the best colour spaces for this dataset:



**Figure 4-10: CAVIAR - Best Colour Spaces**

It was interesting to note throughout the experimentation process that YIQ appeared to be both the worst and best colour space to use for re-identification. When more histogram bins were used YIQ performed better than others. But when a smaller histogram was used, YIQ did not provide much discriminative power compared to other colour spaces and performed the worst. The result of colour space analysis is shown below:



**Figure 4-11: CAVIAR - Worst Colour Spaces**

It should be pointed out that YIQ colour space with histogram of size 3x3x3 was the worst performing colour space in most cases.

## 4.5 NICTA Dataset Benchmark

The second set of experiments was conducted on the NICTA data set. Again, the four methods described earlier were evaluated. Analysing the results of these experiments, it was concluded that methods containing spatial colour information performed better than body histogram.

### 4.5.1 Body Histogram

The results for NICTA data set were similar to that of CAVIAR. For body histogram the best performing histogram had 12x12x12 bins in YIQ colour space with the worst also in the same colour space with 3x3x3 histogram. The results are shown in the figures below:

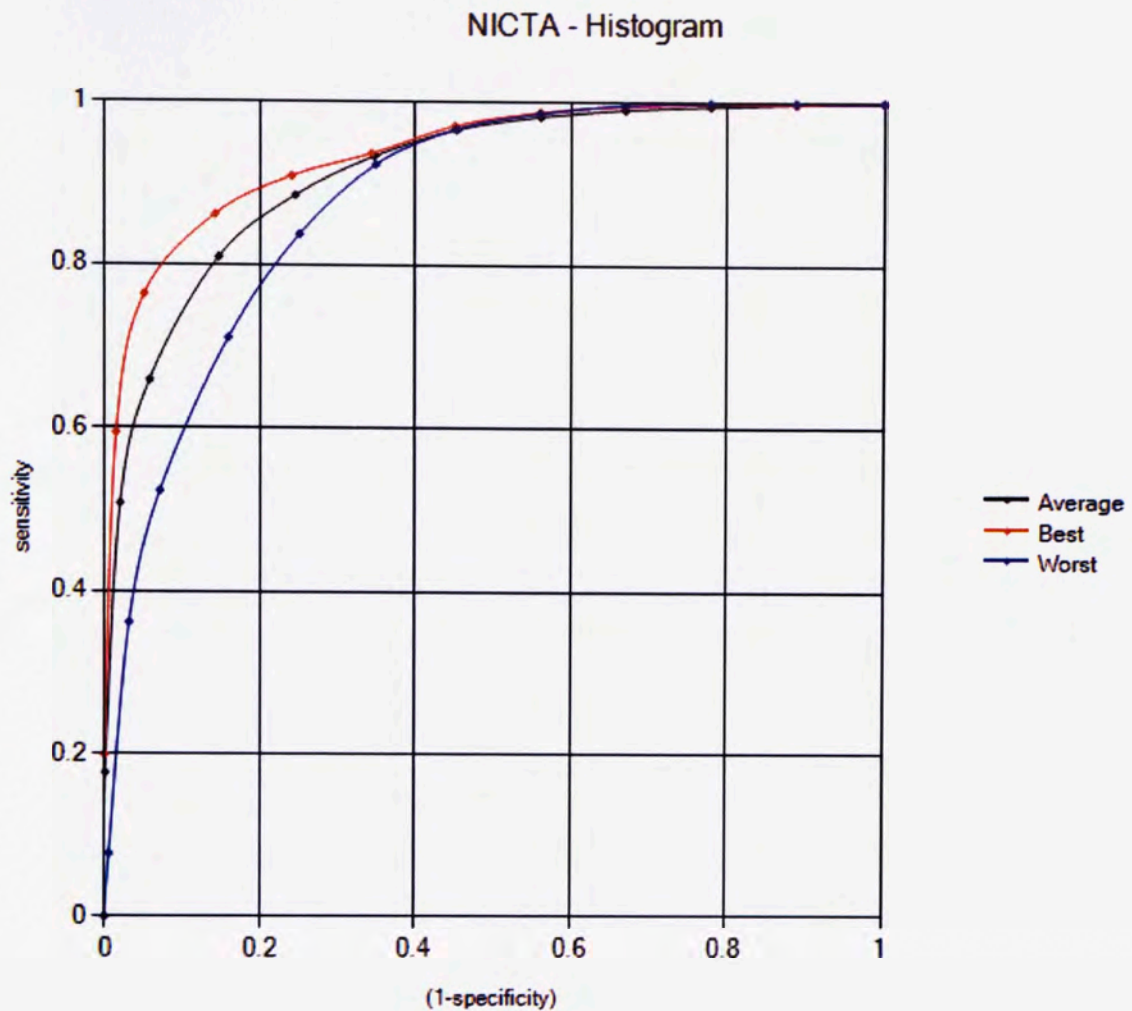


Figure 4-12: NICTA – Histogram ROC

#### 4.5.2 Top-Bottom Histogram

Similarly, top-bottom histogram improved the results as compared to the body histogram. The best and worst top-bottom histograms were in YIQ colour space with 12x12x12 and 3x3x3 bins respectively. These results are shown below:



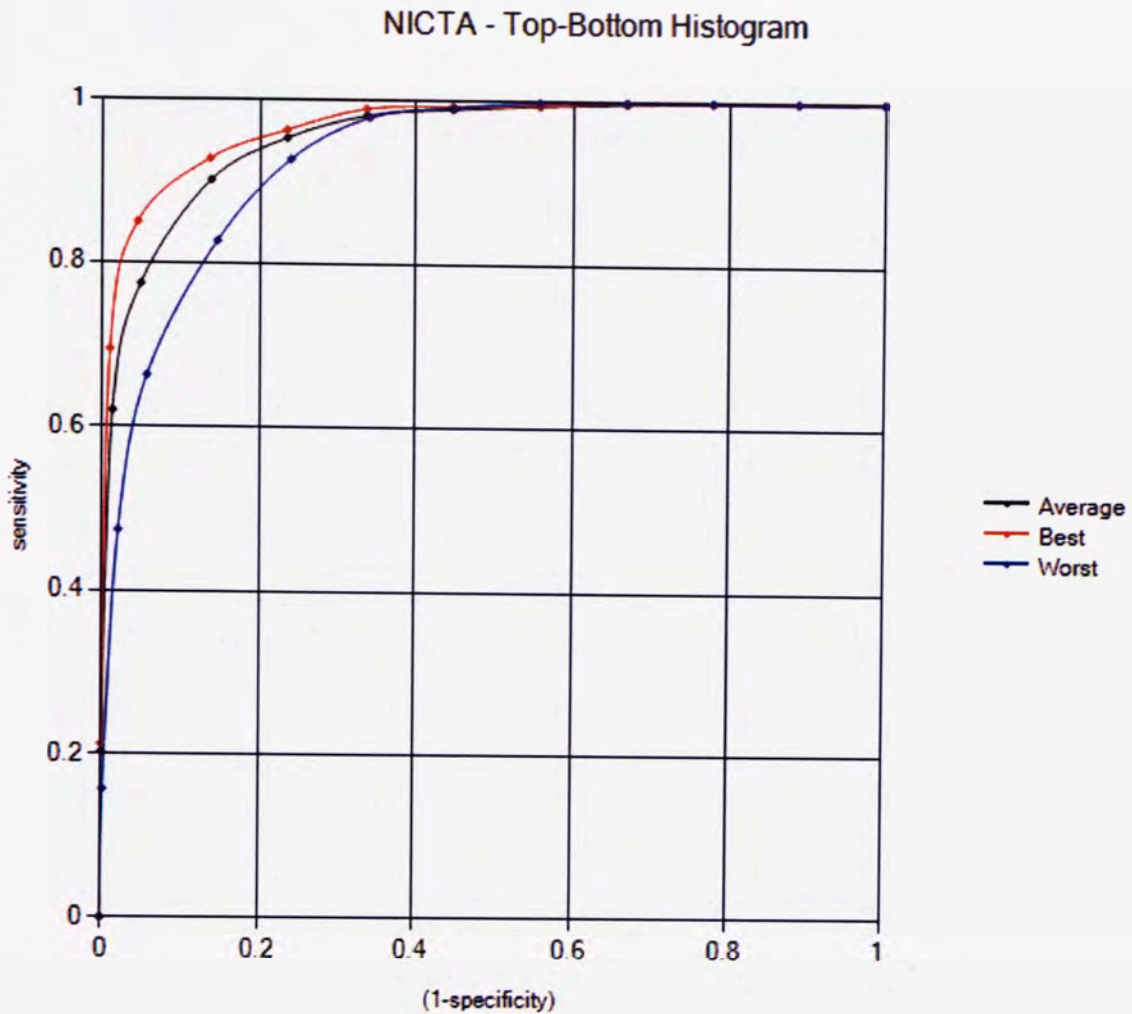
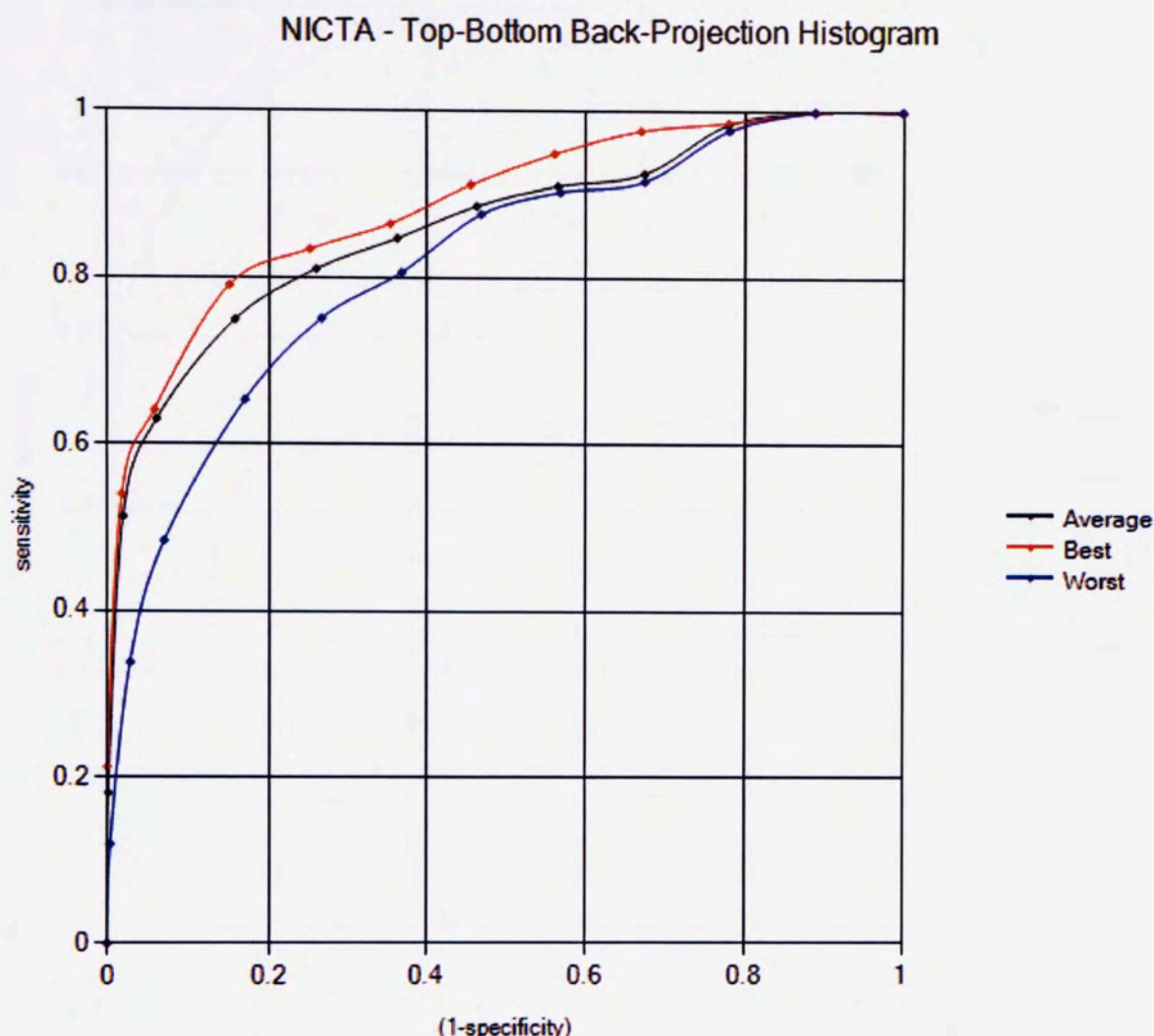


Figure 4-13: NICTA – Top-Bottom Histogram ROC

### 4.5.3 Top-Bottom Back-Projection Histogram

The use of background rejection scheme did not improve upon the top-bottom histogram method. However, it achieved better results than body histogram. The best results were obtained using 8x8x8 histogram in HSV colour space. The worst results were achieved when using YIQ colour space with 3x3x3 histogram as shown below:

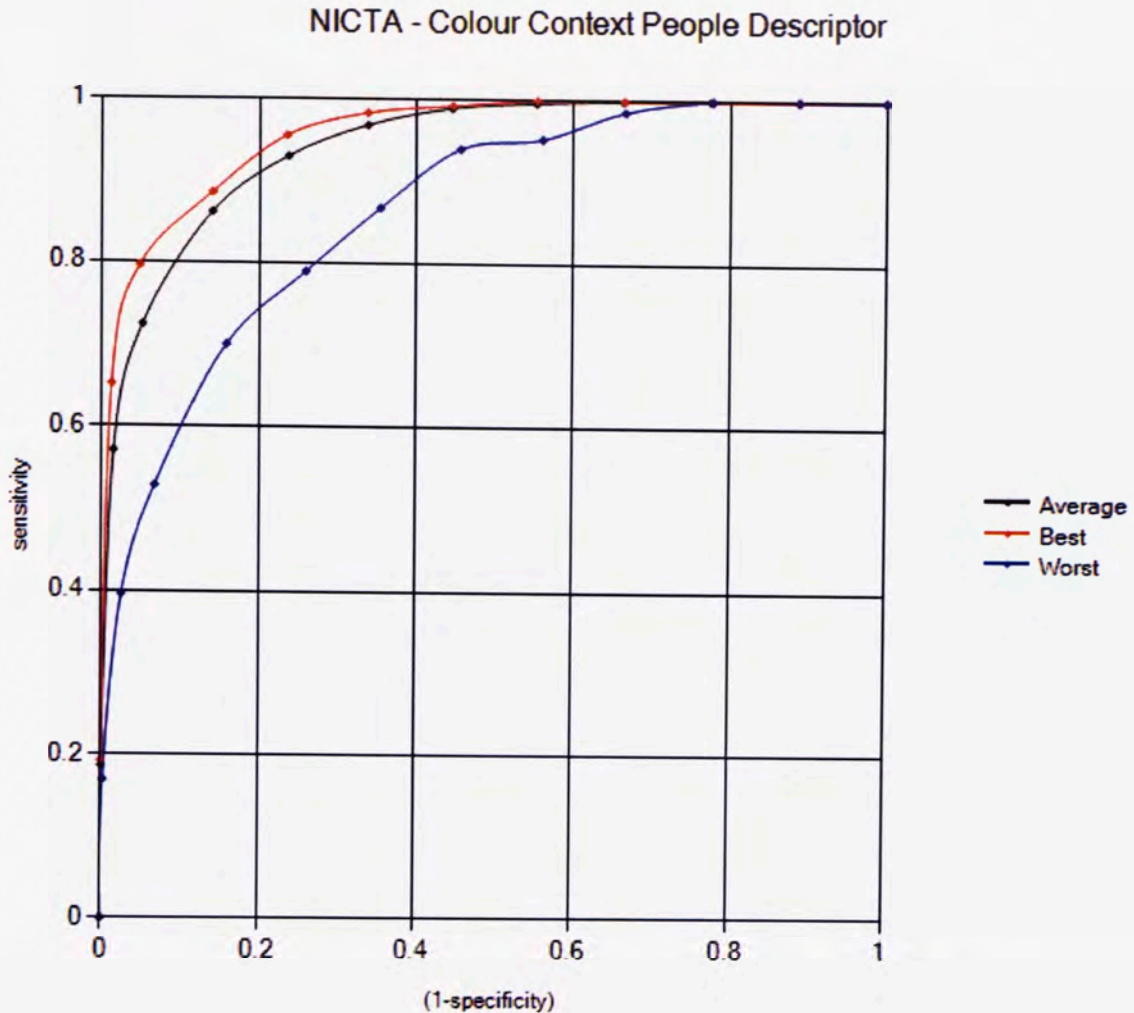


**Figure 4-14: NICTA – Top-Bottom Back-Projection Histogram ROC**

#### **4.5.4 Colour Context People Descriptor**

The best and worst ROC curves were achieved using YIQ colour space. The best ROC curve had 12x12x12 histogram with 6 angles and 1 radial distance. The worst ROC curve was with 3x3x3 histogram with 4 angles and 1 radial distance. These results are shown in the figures below:

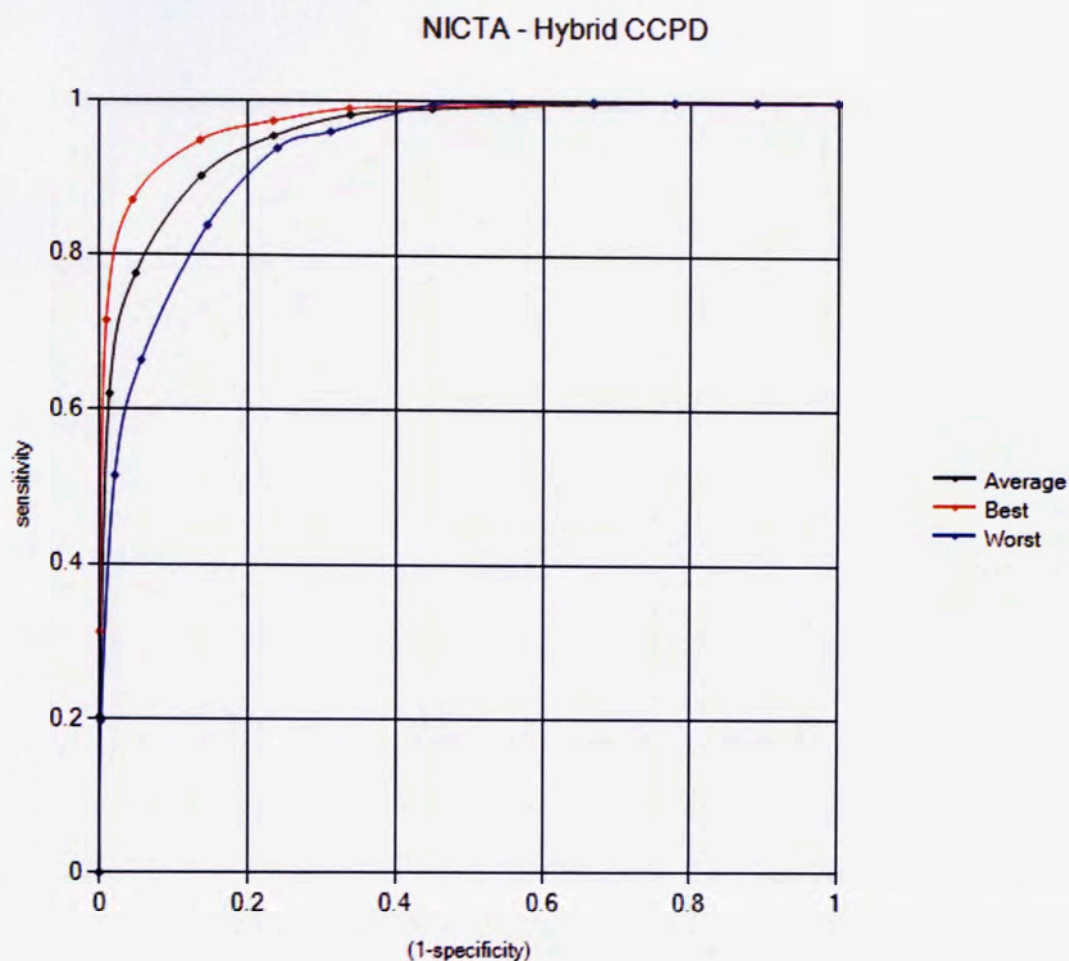




**Figure 4-15: NICTA – Colour Context People Descriptor ROC**

#### **4.5.5 Hybrid Colour Context People Descriptor**

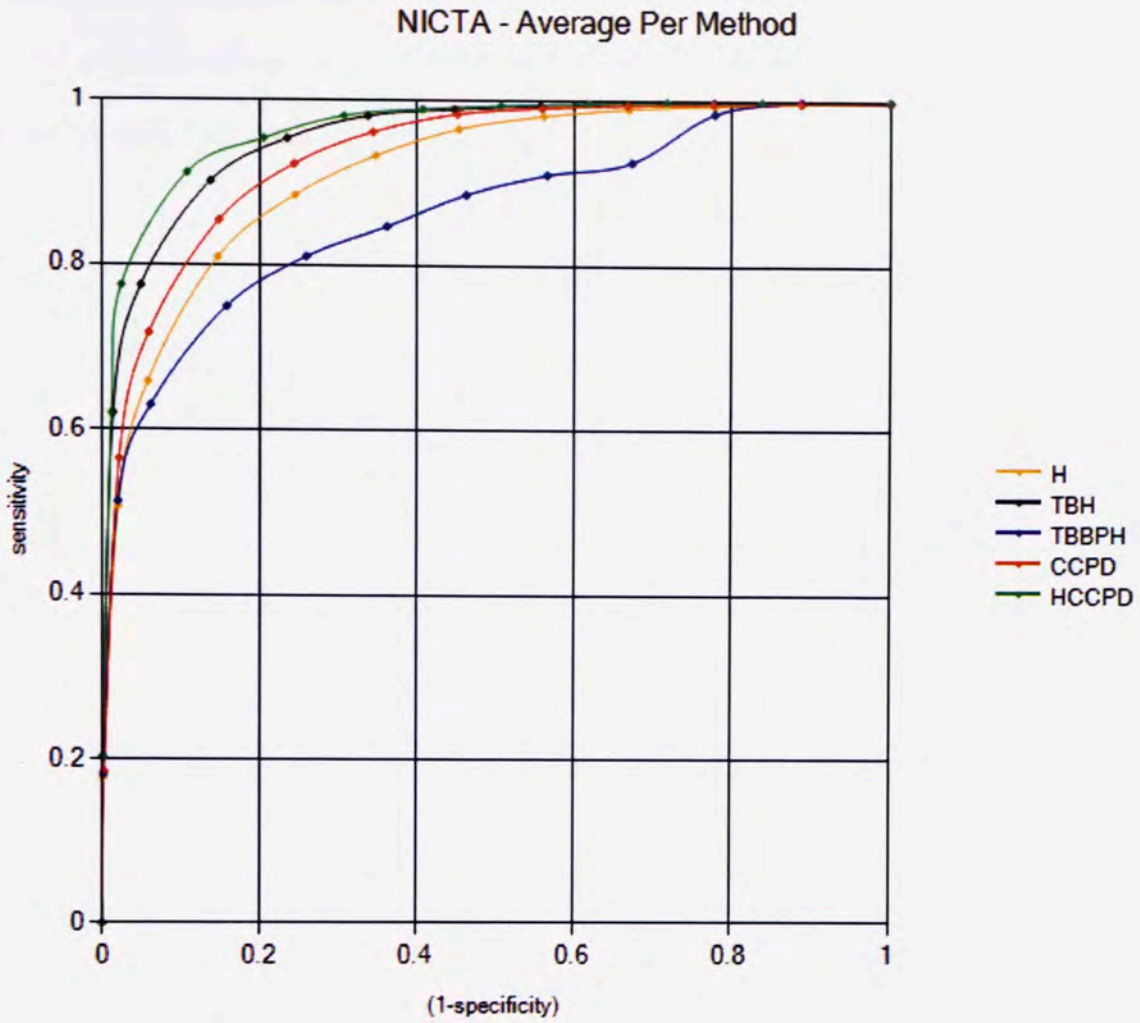
For this descriptor also, the best and worst ROC curves were achieved using YIQ colour space. Similarly, the best and worst ROC curves had 12x12x12 histogram with 6 angles & 1 radial distance and a 3x3x3 histogram with 4 angles & 1 radial distance respectively. These results are shown in the figures below:



**Figure 4-16: NICTA – Hybrid Colour Context People Descriptor ROC**

#### 4.5.6 Analysis

Looking at the results of experimentation from NICTA data set, it can be observed that including spatial colour distribution improves the ROC curve. It has been noticed, however, that the use of background rejection in NICTA data set has not been very useful. It can be seen in the figures below that Top-Bottom Histogram improves the true positive rate over Histogram, but then using background rejection i.e. Top-Bottom Back-Projection Histogram decreases the true positive rate:



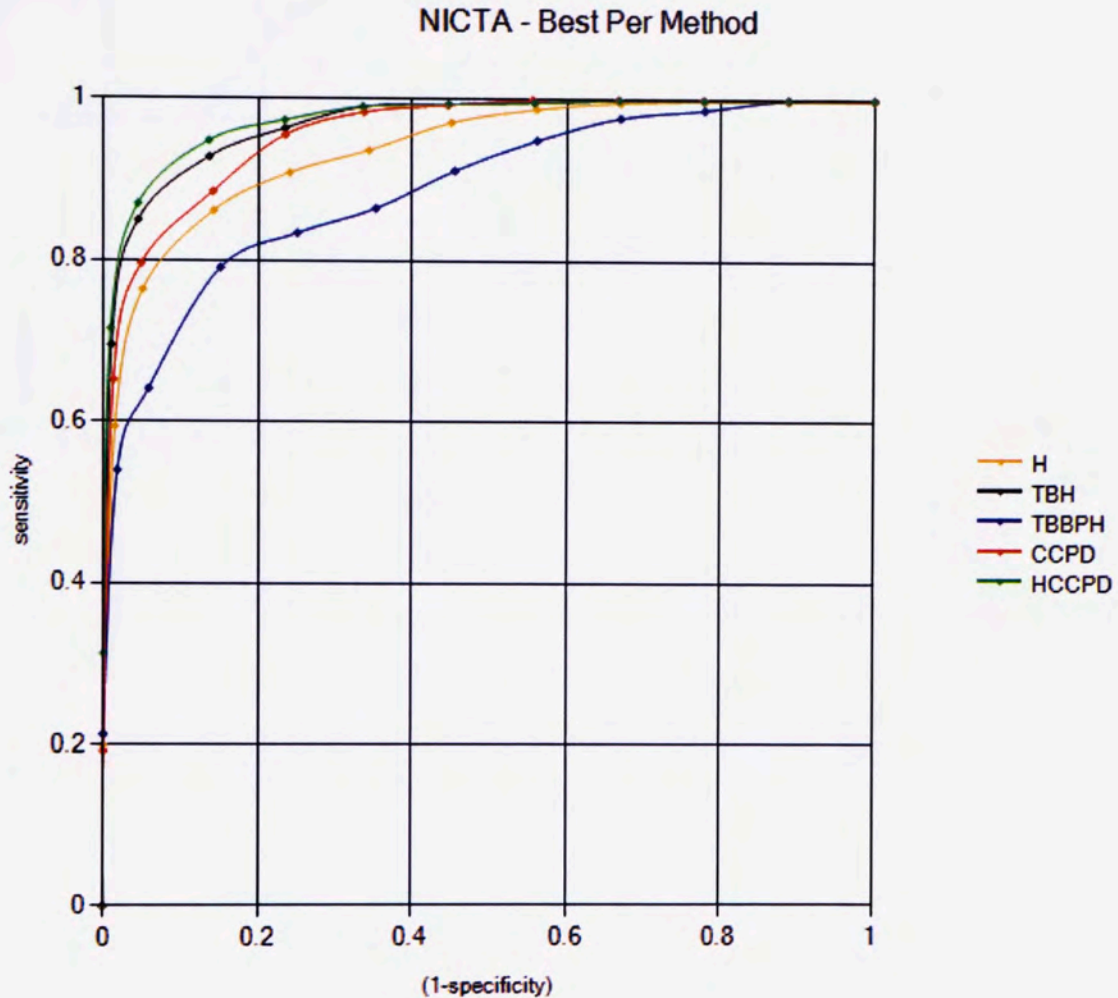
**Figure 4-17: NICTA – Average ROC per Method**

The results shown above were the primary reason for using the Hybrid Colour Context People Descriptor. As seen from the graph, it out performed all techniques. This indicates that the foreground segmentation method adopted for *Colour Context People Descriptor* was not successful for this dataset.

The figures below show the best ROC curve for each method. For colour histogram, it has a YIQ histogram of size 12x12x12. For Top-Bottom Histogram, a YIQ histogram with 12x12x12 bins was the best. For Top-Bottom Back-Projection Histogram, an HSV

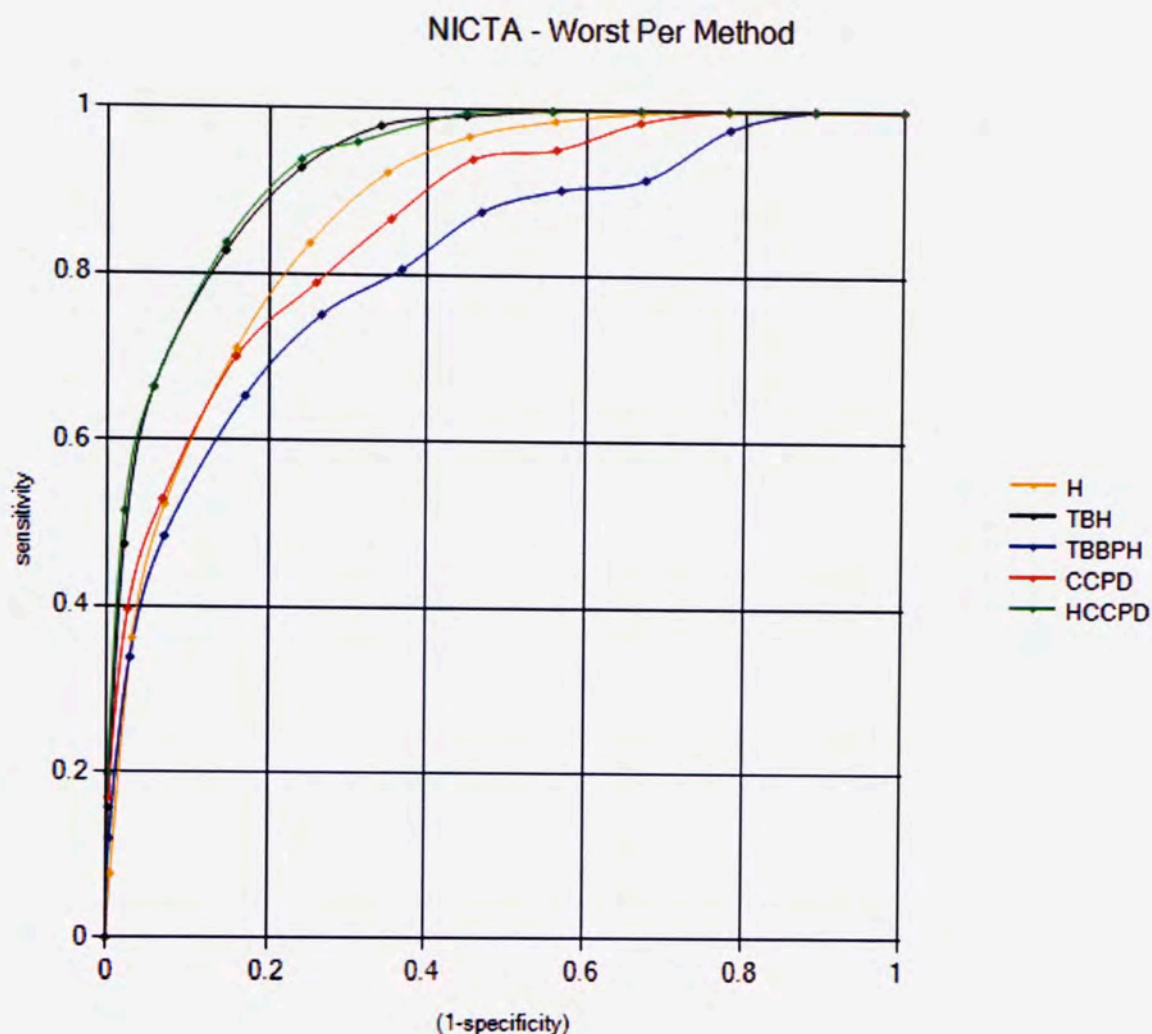


histogram of 8x8x8 performed the best. Lastly, for *Context Colour People Descriptor*, and the hybrid version, a YIQ histogram with 12x12x12 bins, 6 angles and 1 radial distance was the best.



**Figure 4-18: NICTA – Best ROC per Method**

For the worst cases, a YIQ histogram with 3x3x3 bins was the worst performing configuration for Histogram, Top-Bottom Histogram and Top-Bottom Back-Projection Histogram. For Hybrid Colour Context People Descriptor and *Colour Context People Descriptor*, a YIQ histogram of size 3x3x3 with 4 angles and a single radial distance was the worst. These results are shown in the figures below:



**Figure 4-19: NICTA – Worst ROC per Method**

Compiling the overall results of best techniques, it can be easily seen that Histogram performed the worst as compared to methods incorporating spatial colour distribution.

The best and worst colour space for NICTA data set was YIQ as in CAVIAR data set. Again, when more histogram bins were used YIQ performed better than others and vice versa. Other colour spaces almost equally as the figures below show:

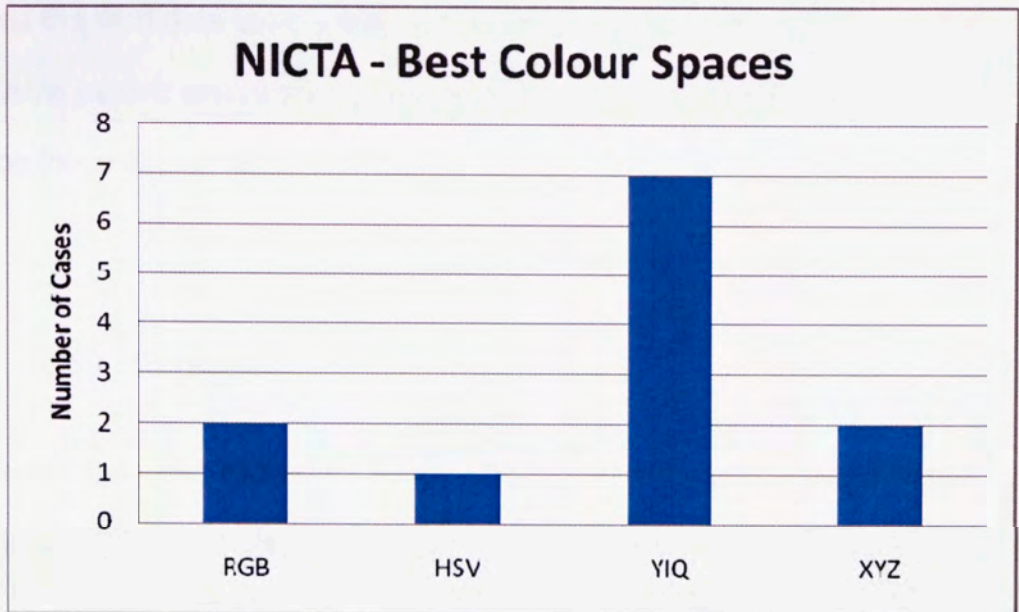


Figure 4-20: NICTA – Best Colour Spaces

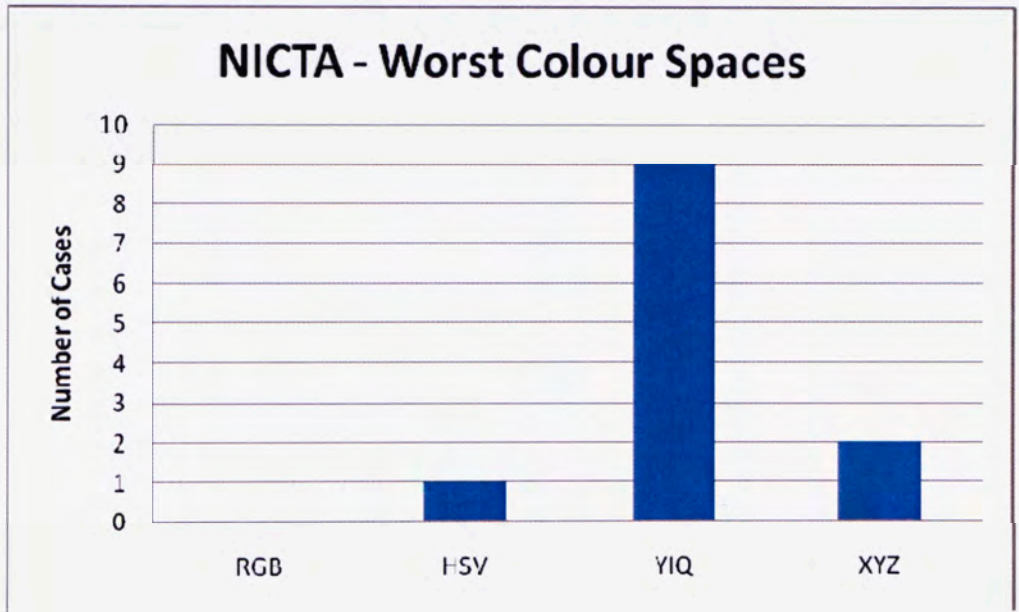


Figure 4-21: NICTA – Worst Colour Spaces

For this dataset also, YIQ colour space with histogram of size 3x3x3 was the worst in most of the cases.

4.6 NICTA-CAVIAR Benchmark



This was the third data set created by combining the data from CAVIAR and NICTA data sets. Same experiments were performed on this data set. The following sections describe these experiments and results.

### 4.6.1 Body Histogram

As mentioned earlier, YIQ colour space featured in both best and worst colour spaces. When a histogram with more bins was used, it performed well. On the other hand for a smaller histogram it performed the worst. The best and worst Histogram results were achieved by using YIQ colour space with histogram bins of 12x12x12 and 3x3x3.

These results are shown in the following two figures:

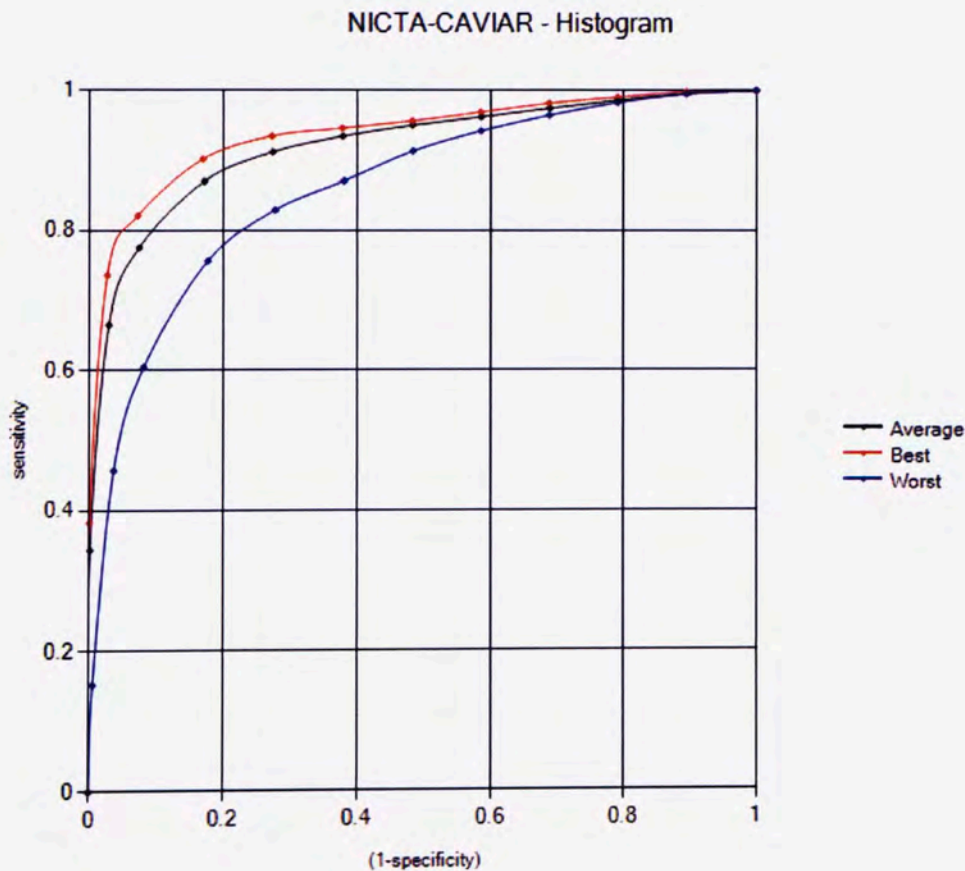


Figure 4-22: NICTA-CAVIAR - Histogram ROC

### 4.6.2 Top-Bottom Histogram

This method had the same best and worst performing parameters as the Histogram method in the section before. The best and worst results were with YIQ colour space with 12x12x12 and 3x3x3 bins. But by including the spatial scattering of colour it has much better performance than the Histogram method in the previous section. The results are shown as:

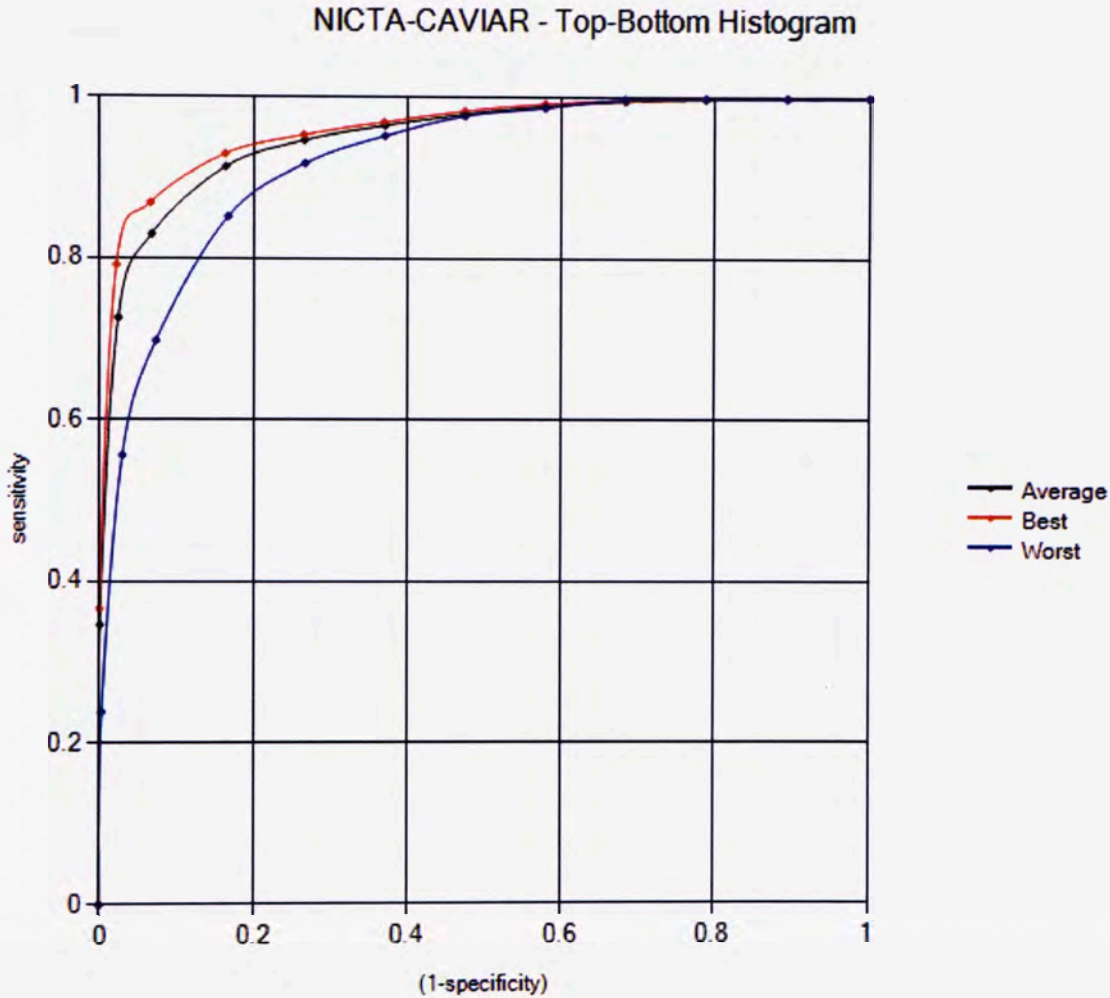


Figure 4-23: NICTA-CAVIAR – Top-Bottom Histogram ROC

### 4.6.3 Top-Bottom Back-Projection Histogram

This method adds background rejection to the method in the previous section. The best and worst ROC curves were achieved by using an HSV 8x8x8 histogram and YIQ 3x3x3 histogram. These results are given below:

#### NICTA-CAVIAR - Top-Bottom Back-Projection Histogram

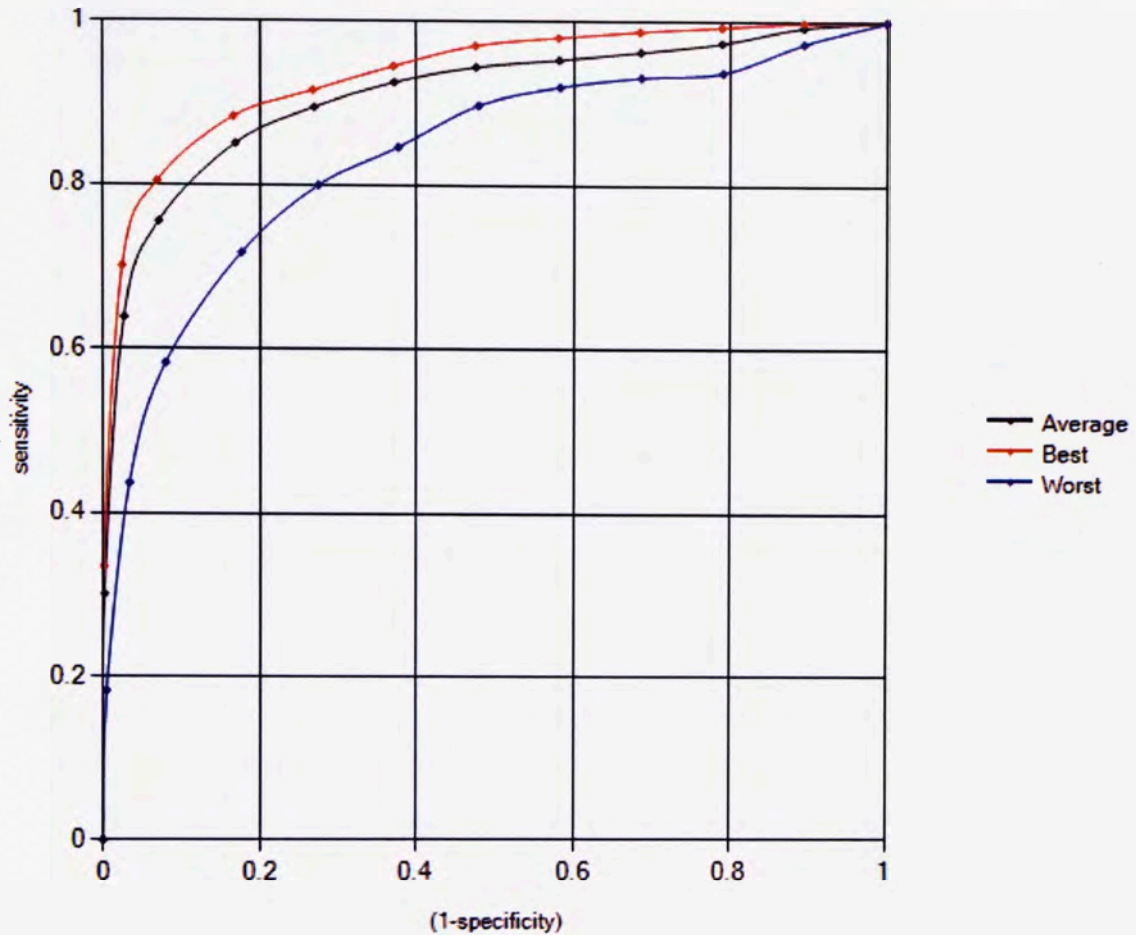
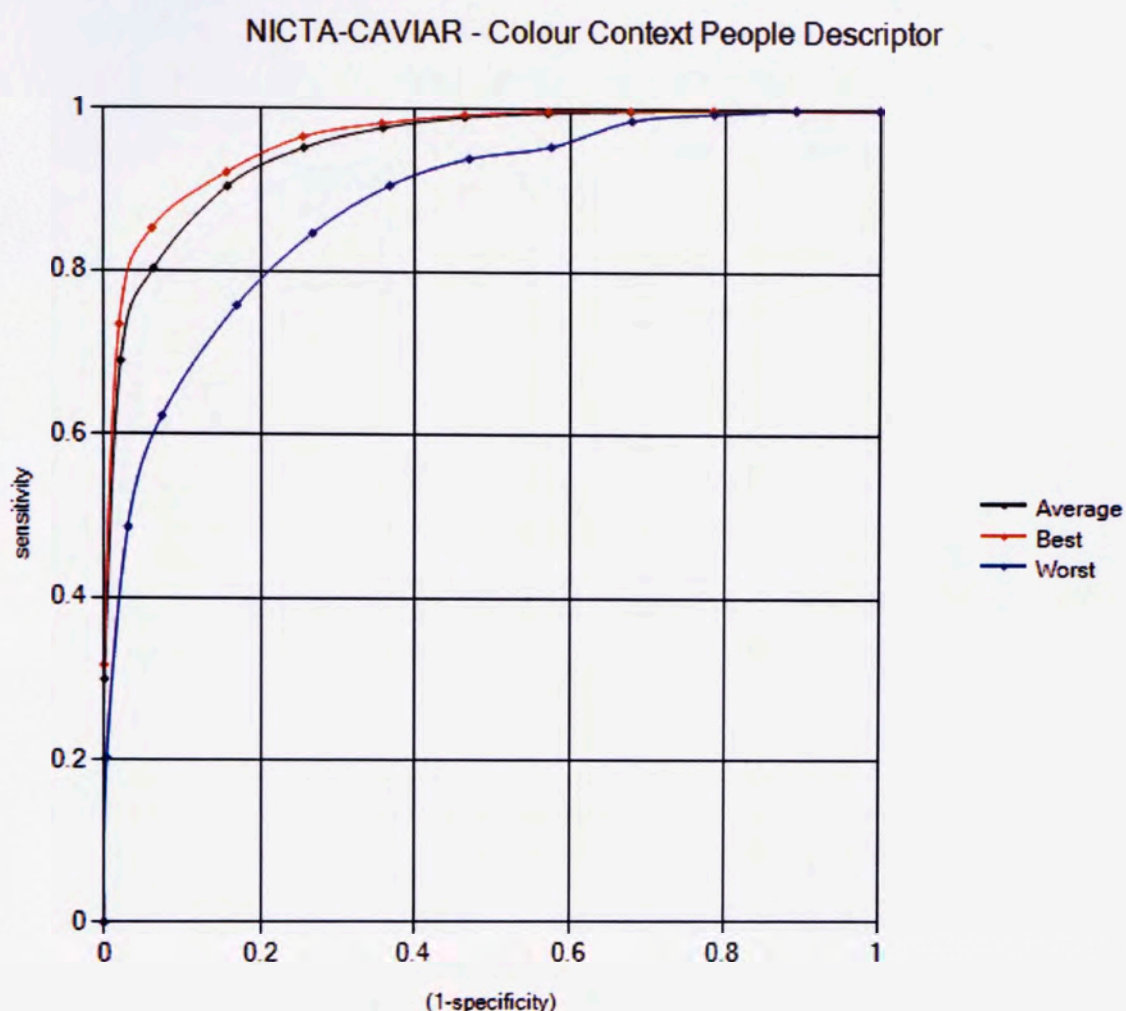


Figure 4-24: NICTA-CAVIAR – Top-Bottom Back-Projection Histogram ROC

#### 4.6.4 Colour Context People Descriptor

Again the best and worst performing ROC curves were achieved using YIQ colour space. The best configuration consisted of 12x12x12 histogram with 6 angles and a single radial distance where as, the worst was 3x3x3 histogram with 4 angles and radial distance of 1. These results are given below:

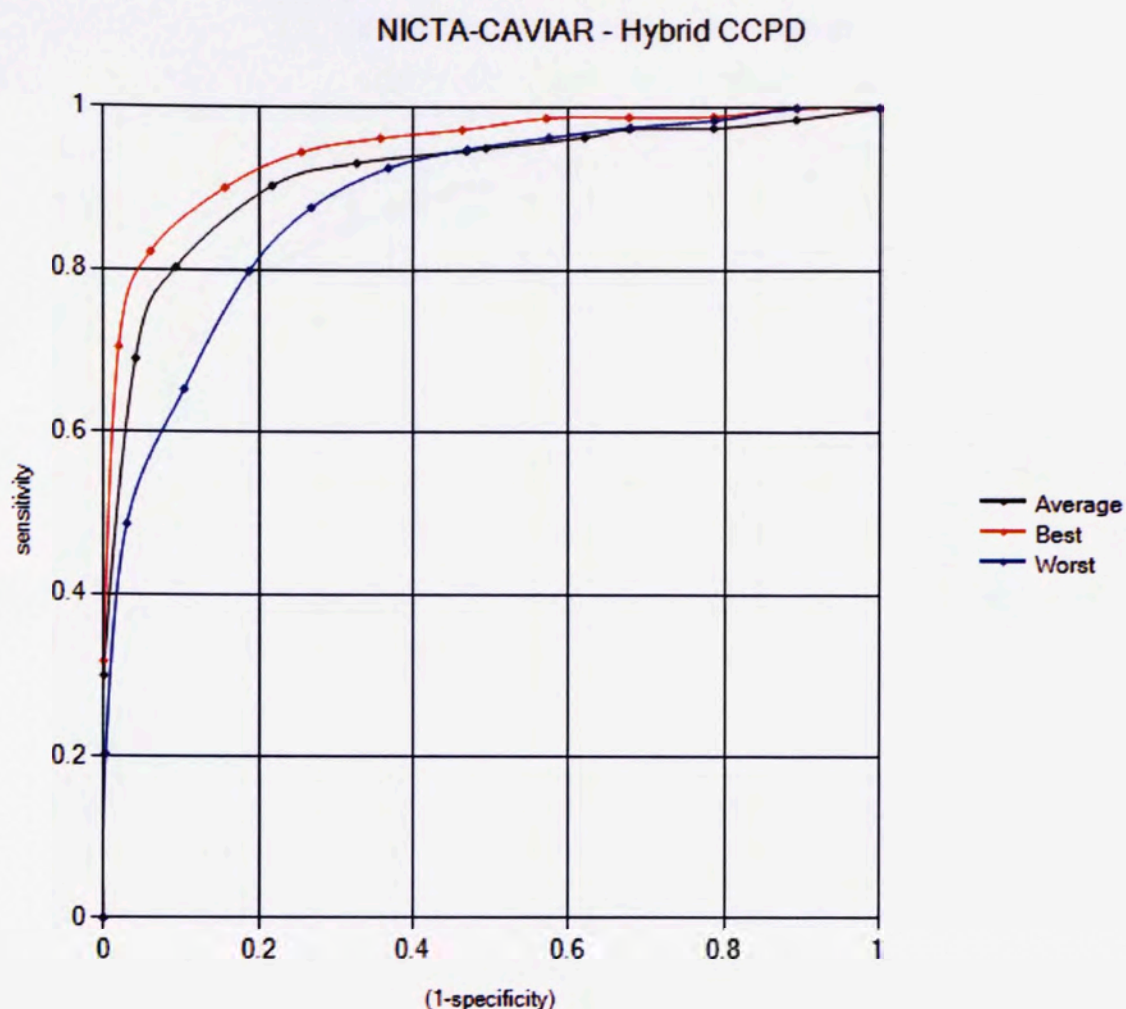




**Figure 4-25: NICTA-CAVIAR – Colour Context People Descriptor ROC**

#### **4.6.5 Hybrid Colour Context People Descriptor**

The hybrid version of Colour Context People Descriptor had same parameters for best and worst ROC curves. YIQ colour space was dominant in both curves. The best descriptor was a 12x12x12 histogram with 6 angles and a single radial distance. The worst curve was produced by a 3x3x3 histogram with 4 angles and single radial distance. These results are given below:

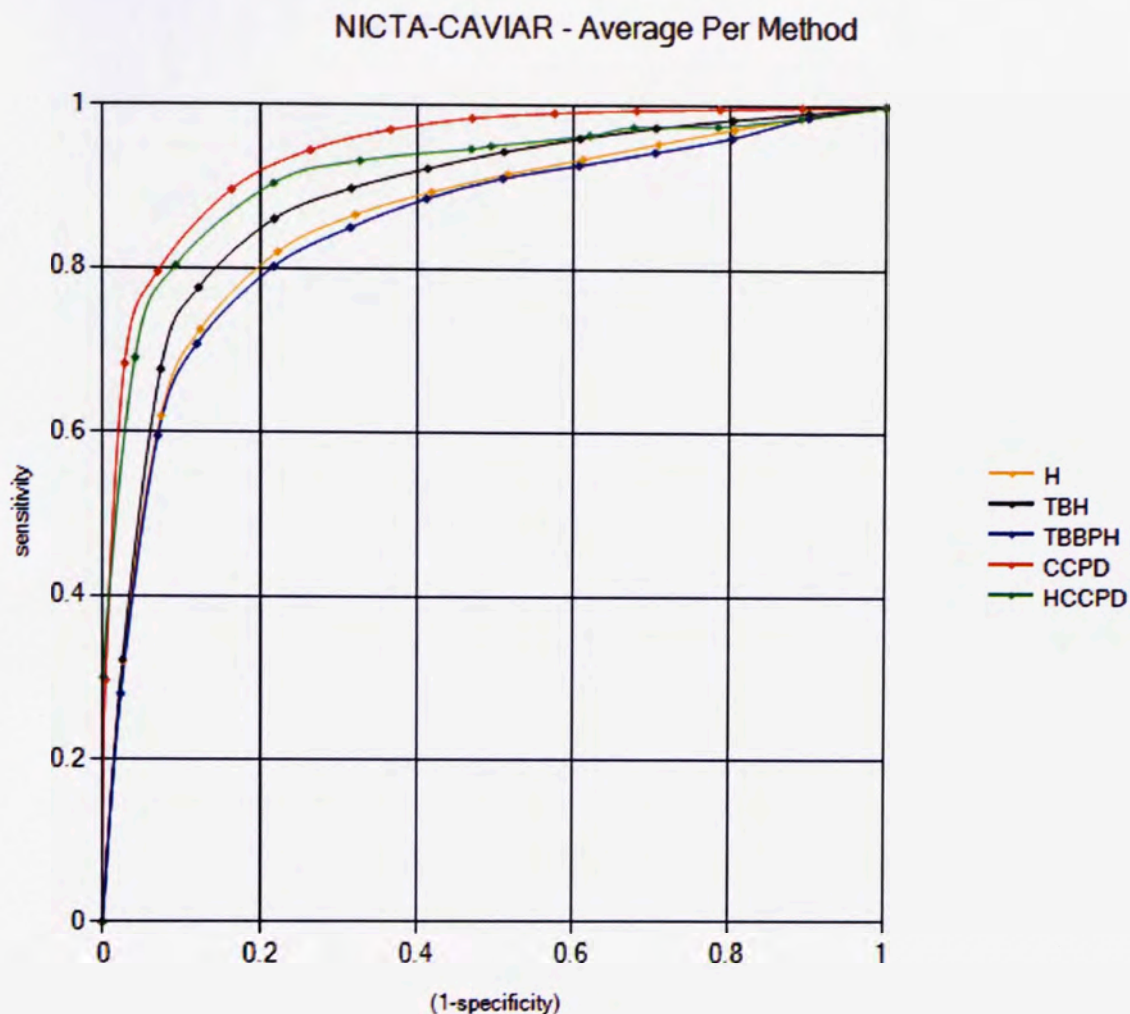


**Figure 4-26: NICTA-CAVIAR – Hybrid Colour Context People Descriptor ROC**

#### 4.6.6 Analysis

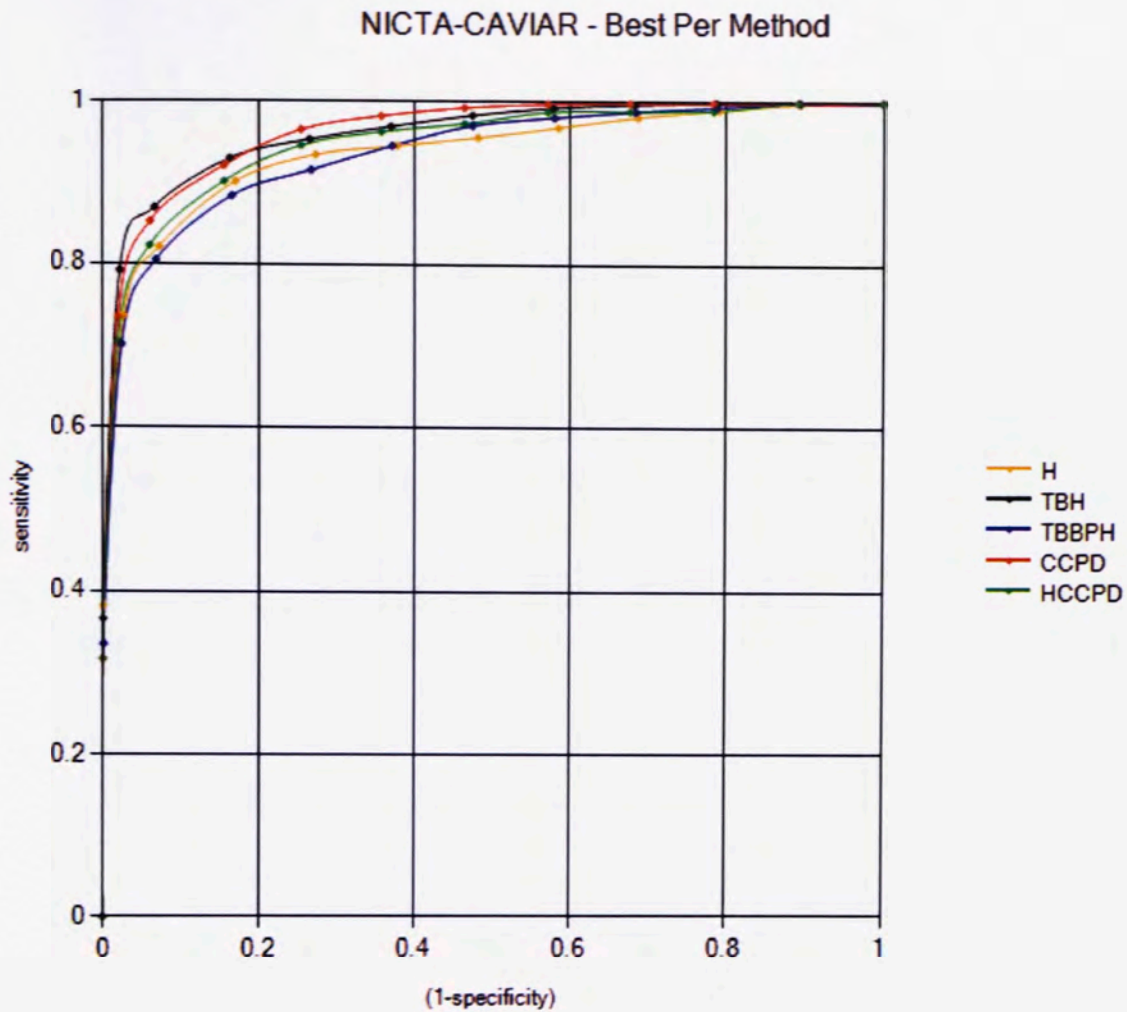
Following the trend of previous data sets, the spatial colour related methods performed well as compared to the histogram method. The results of average ROC curves are shown in the figures below:





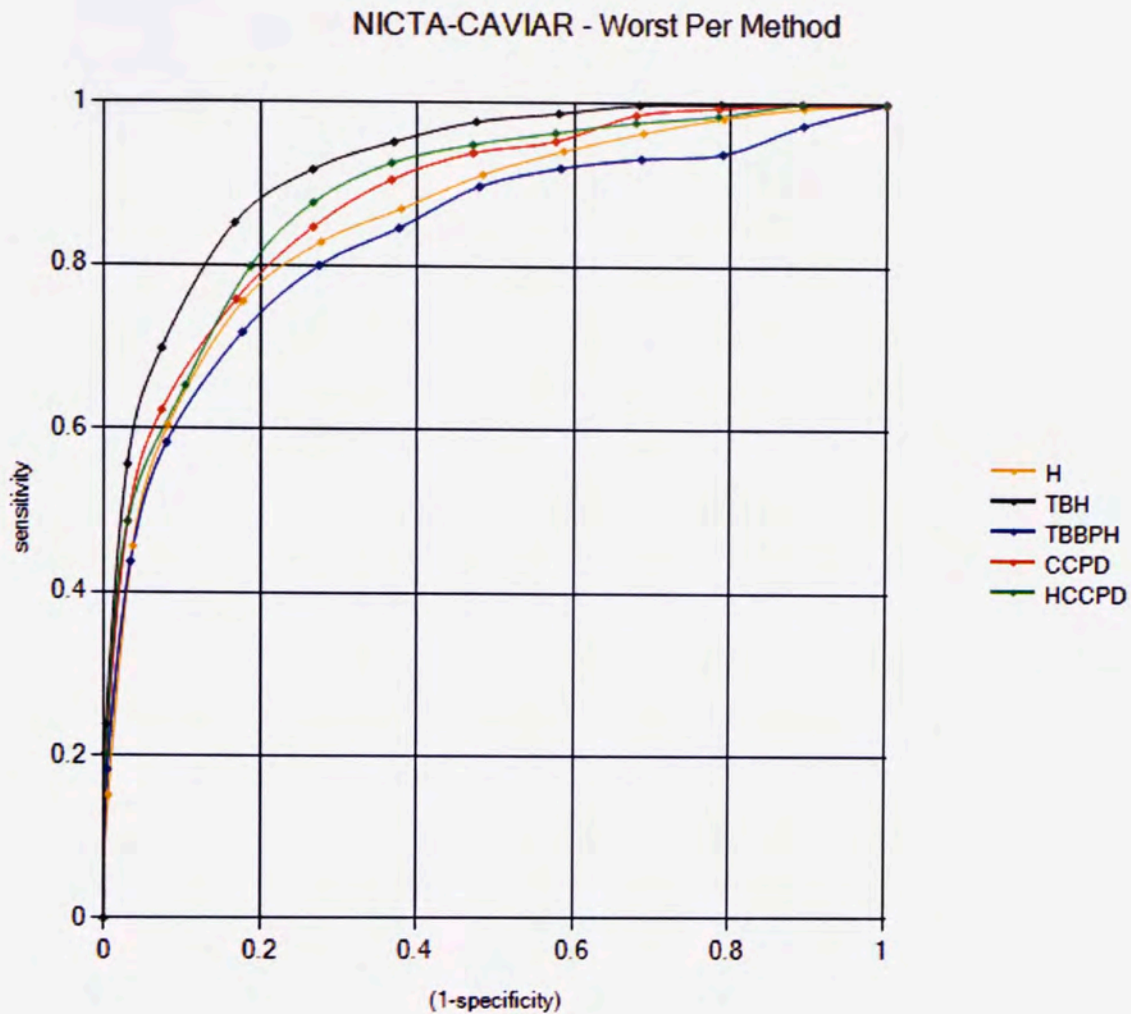
**Figure 4-27: NICTA-CAVIAR – Average ROC per Method**

There was not much difference in the methods for best ROC curve. However, the methods containing spatial colour distribution information fared slightly better than the Histogram method. Also *Colour Context People Descriptor* has performed better overall as compared to other methods as can be seen below:



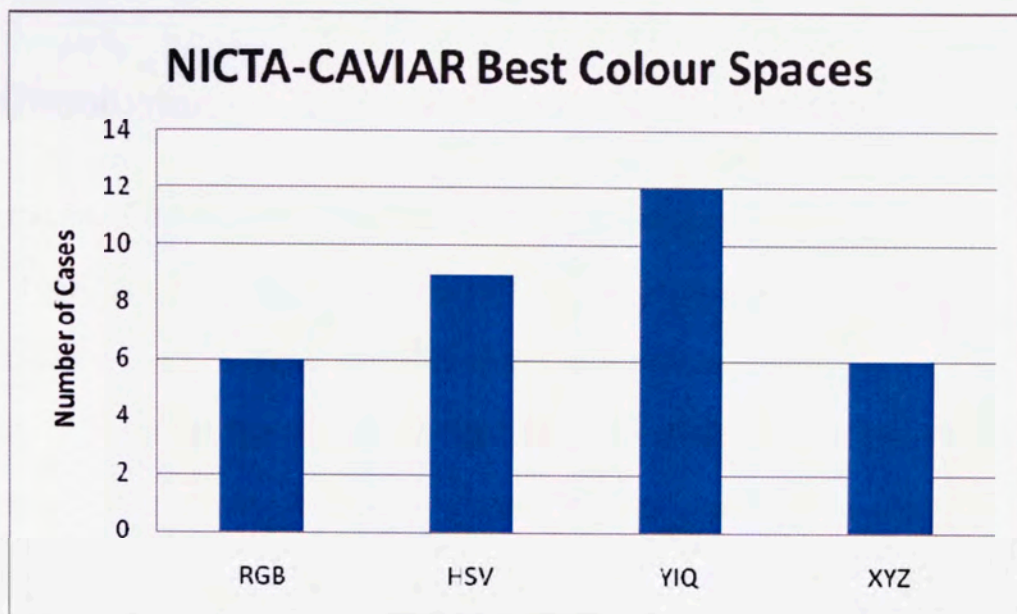
**Figure 4-28: NICTA-CAVIAR – Best ROC per Method**

However, when comparing the worst ROC curve for each method, the advantage of using methods with spatial colour distribution becomes evident. The figure below shows that two of the spatial methods performed much better than the Histogram method.



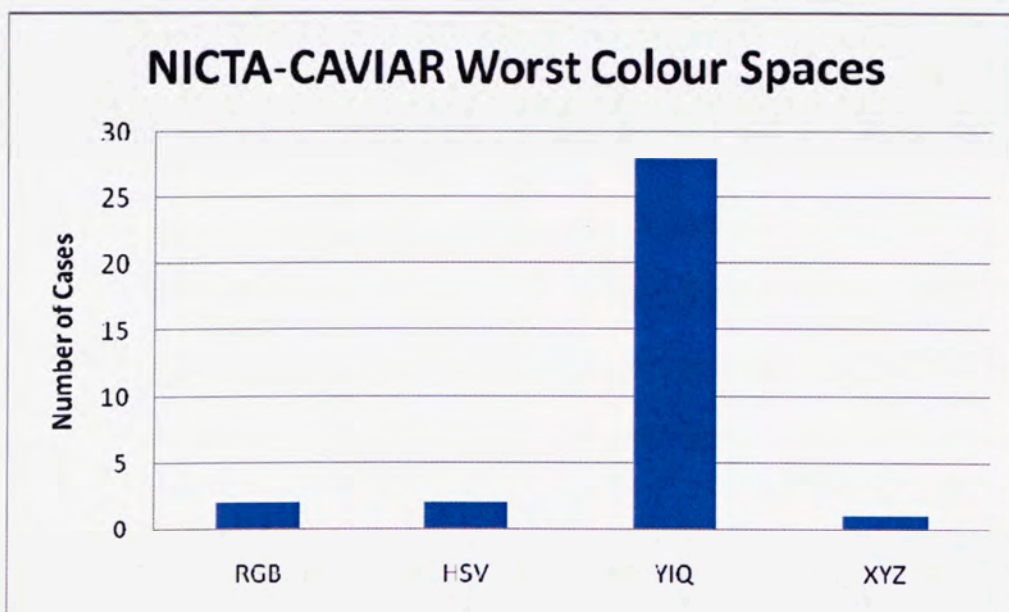
**Figure 4-29: NICTA-CAVIAR – Worst ROC per Method**

Again YIQ with more histogram bins provided the best results as compared to when using with smaller histograms. HSV colour space was the second best. Interestingly, RGB, HSV and XYZ performed better than YIQ when using smaller histograms. These results are shown in the following two figures:



**Figure 4-30: NICTA-CAVIAR – Best Colour Spaces**

Keep in line with the previous observation, YIQ colour space histogram of size 3x3x3 was the worst performing colour in most of the case as shown below:



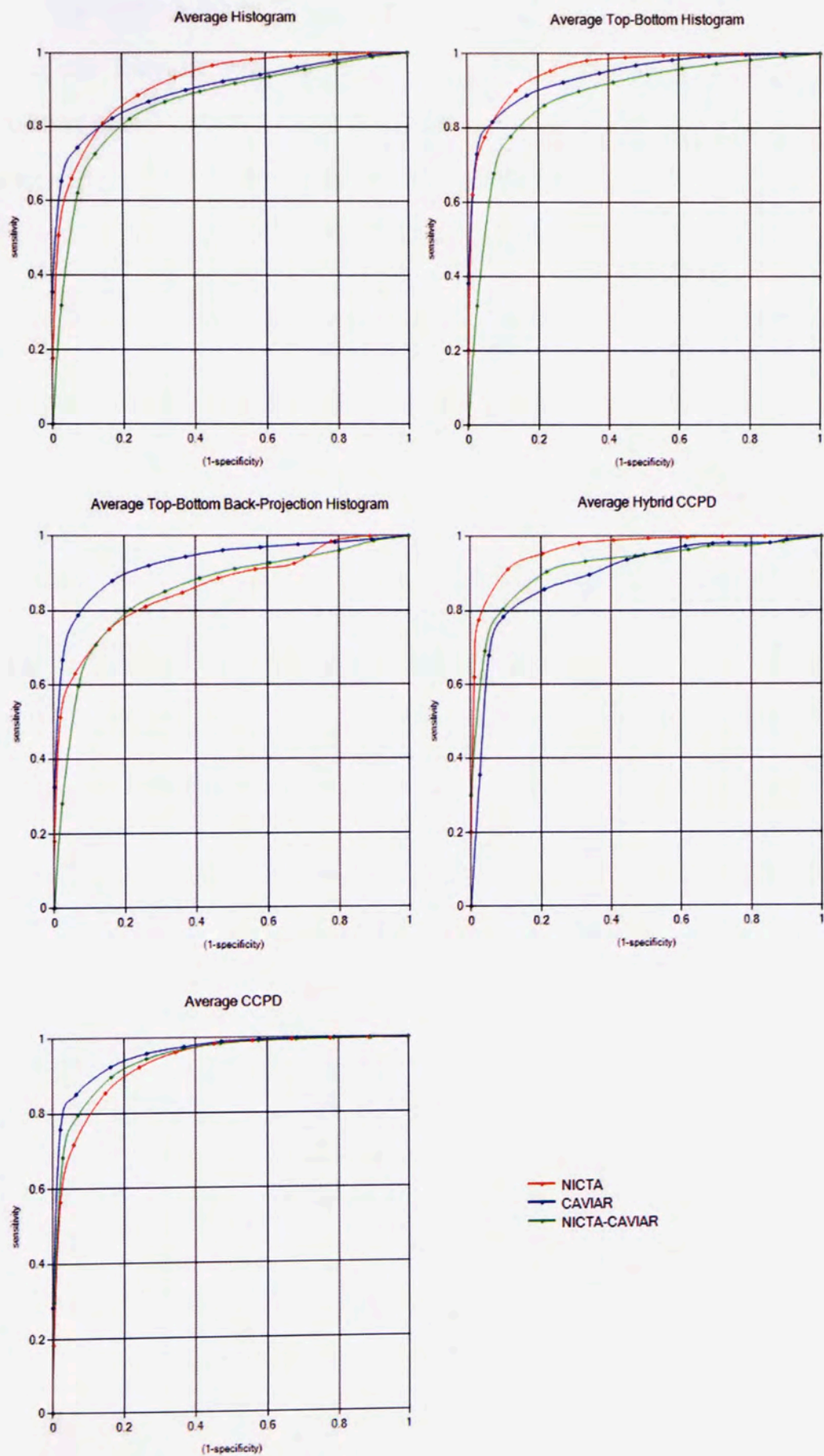
**Figure 4-31: NICTA-CAVIAR – Worst Colour Spaces**

## 4.7 Conclusion

Given the four methods described in this thesis, extensive experimentation was carried out on three data sets. It has been shown that Histogram method is the worst when it comes to using it for re-identification of people in images or videos. On the other hand, methods discussed in this thesis that include the spatial dispersal of colour performed the best. It has been seen that *Colour Context People Descriptor* is a robust way of capturing the colour of people. Overall, *Colour Context People Descriptor* provided the best true positive rate performance.

In order to compare the performance of each of these methods, the results are compared across each data set. It can be seen that the average ROC curve for Histogram is lower than the other three methods. There is a need to apply background rejection technique to reduce noise in the legs region. The method using back-projection described in this thesis has worked reasonably well with CAVIAR data set but not so much with NICTA data set. Because of this, Top-Bottom Back-Projection method has appeared in some worst cases for NICTA data set. This in turn as effected the results of NICTA-CAVIAR data set. Nevertheless, it has been shown through the experimentation that using spatial colour scattering one can improve the re-identification of people in video. The following series of figures show the comparison of average ROC curves across each data set:





**Figure 4-32: Average ROC Per Method**

Like the ROC curves, the best and worst colour spaces are also compared across the data sets. It can be seen the YIQ is both the best and the worst colour space for people re-identification. When a larger histogram was used, YIQ colour space out performed RGB, HSV and XYZ. On the other hand, when a histogram with smaller bin size was used, YIQ colour space performed the worst. These comparisons are shown in the following two figures:

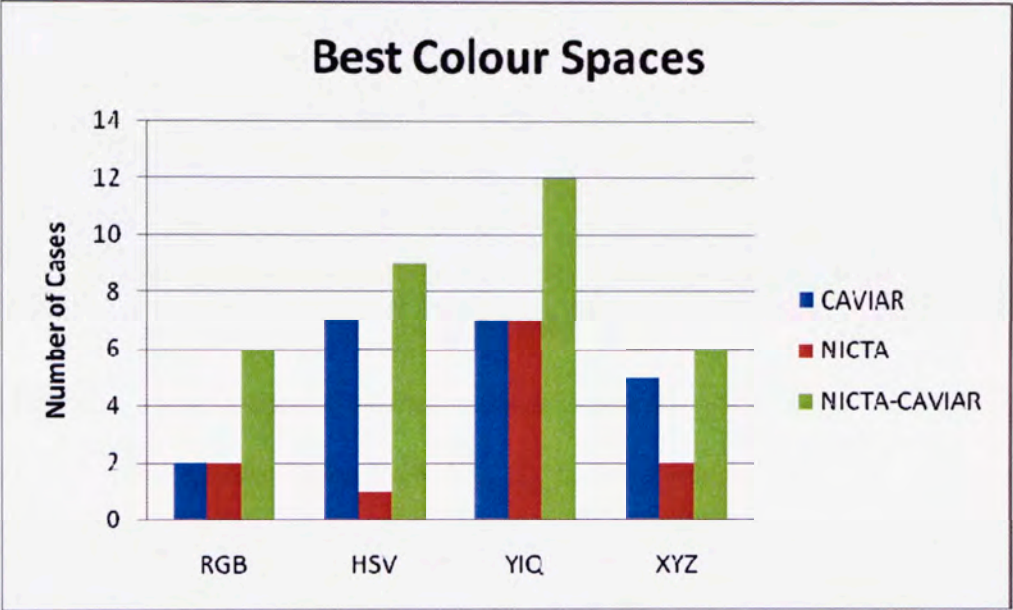


Figure 4-33: Best Colour Spaces

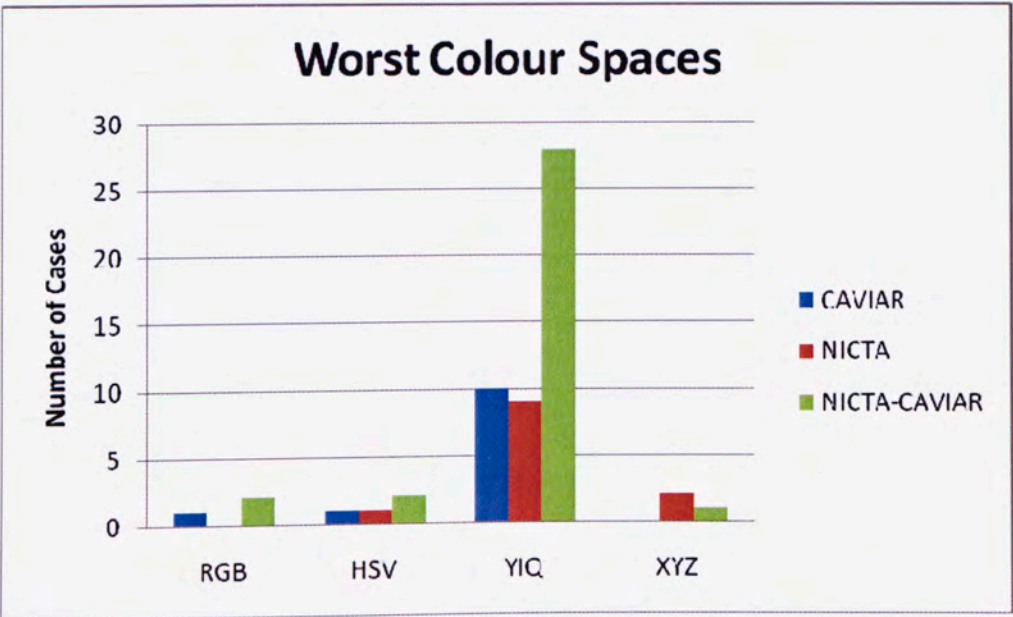


Figure 4-34: Worst Colour Spaces

Finally, all the methods used in the experimentations were compared across all three data sets. On the whole, spatial colour related techniques provided the best performance. Histogram was the worst performing method. The *Colour Context People Descriptor* proposed in this thesis performed the best. It should be noted here that the use of background rejection method did not perform well in NICTA data set. The proposed descriptor uses the same method for creating bottom histogram. Considering the weakness of background rejection method used in this research, the *Colour Context People Descriptor* still outperformed all the other methods. This highlights to the robustness of the *Context Colour Histogram* used in the descriptor.

It can be seen from the results that including the spatial colour information improves the re-identification of people. Separating the torso and lower region histograms gives us great improvement. Moreover, including the spatial colour distribution of the torso region further improves the classification performance.

# 5 Conclusion

---

Colour information is an important feature in videos. It has to be used judiciously to improve colour-based re-identification. This research has taken focused on including spatial scattering of colour while using efficient colour modelling structure of histogram. Just splitting the top and bottom histogram as in Top-Bottom Histogram greatly improves the re-identification. Because same person can appear in different places where the background is different, there is a need to ignore the background in the legs area. The pose of legs during a video capture, as a person walks, varies widely. In most cases, we end up with unwanted background.

For background rejection in this research, a colour histogram of pixels, which are likely to be legs, is created and then by using back-projection a legs' mask is recovered. Using the mask, only the legs pixels are used to create the legs histogram thus ignoring the background colour. Two approaches were discussed for finding the probable legs' pixels. One was a naïve approach and the other one was based on search an area. The first approach assumed that the partition between torso and legs is correct and selected a rectangular area under the torso line to create the histogram to be back-projected. While this worked in most situations, it caused poor segmentation in other. Therefore, a search-based approach was proposed to look for the biggest colour mismatch. To make a decision between which segmented mask to use, a Naïve Bayes Classifier was trained for classification.

But the background rejection method used in this thesis gave mixed results. While it worked fine in CAVIAR data set, it reduced the accuracy in NICTA data set. More work is required to improve the background rejection process. Further investigation is required for implementing a more robust foreground mask recovery.

The *Colour Context People Descriptor* proposed in this thesis outperformed all the methods despite the weakness of background rejection technique. The main source of robustness of *Colour Context People Descriptor* came from *Colour Context Histogram* proposed in this thesis. The descriptor uses the overall spatial distribution of colours by separating the torso colour information from the legs. Furthermore, it considers the spatial colour distribution of the torso region.

We have seen from the literature review that colour histograms contribute to the best re-identification results. However, when the colour information is ignored, the re-identification rate is low. Histograms are an efficient way to model colours. It is inherently rotation variant and can be made scale variant through normalisation. But the downside of rotation invariance is that a histogram does not retain spatial colour distribution. It has been shown through the experimentation that taking into account spatial distribution of colours improves the re-identification performance.

Furthermore, using *Colour Context People Descriptor* along with other features is definitely going to further improve the re-identification rate. This is the future direction of this research where features like height or gait could be used along with *Colour Context People Descriptor* to increase the accuracy of re-identification of people in video.



# References

- [1] M. J. Swain, D. H. Ballard. "Indexing via color histograms", DARPA Image Understanding Workshop, 1990.
- [2] D. Wojtaszek, R. Laganiere, "Using Color Histograms to Recognize People in Real Time Visual Surveillance", In Int. Conf. Multimedia, Internet and Video Technologies, volume 3, Greece, pages 261--264, September 2002.
- [3] L. Goldmann, M. Karaman, J. T. S. Minquez, T. Sikora, "Appearance-Based Person Recognition for Surveillance Applications", 2006.
- [4] M. H., D. Klunder, K. Kraiss, "Color and Texture Features for Person Recognition", 2004 IEEE International Joint Conference on Neural Networks, 2004.
- [5] C. Nakajima, M. Pontil, B. Heisele, T. Poggio, "Full-body person recognition sytem", Pattern Recognition, vol. 36, pp. 1997-2006, 2003.
- [6] C. BenAdelkader et al. "Person identification using automatic height and stride estimation", ICPR Vol. 4 pp.377-380 2002.
- [7] C. Madden, M. Piccardi. "Height Measurement as a Session-based Biometric for People Matching Across Disjoint Camera Views", IVCNZ 2005.
- [8] O. Javed, K. Shafique, M. Shah, "Appearance Modeling for Tracking in Multiple Non-overlapping Cameras", CVPR vol. 2, pp 26 – 33, 20 – 25 June 2005.
- [9] S. Belongie, J. Malik and J. Puzicha, "Shape matching and object recognition using shape contexts." IEEE Trans. Pattern Analysis and Machine Intell., 24(4):509-522, April 2002.
- [10] S. Park and J. K. Aggarwal, "Simultaneous tracking of multiple body parts of interacting persons", Computer Vision and Image Understanding, 2005.
- [11] CAVIAR Project, <http://homepages.inf.ed.ac.uk/rbf/CAVIAR/>





- [12] G. Jaffre and P. Joly, "Costume: A new feature for automatic video content indexing", In Proceedings of RIAO, pp 314 – 325, 2004.
- [13] N. Gheissari Et Al., "Person Reidentification Using Spatiotemporal Appearance", CVPR, 2006.
- [14] M. S. Nixon Et Al., "Automatic gait recognition", IEE Colloquium on Motion Analysis and Tracking 3, pp 1 – 6, 1999.
- [15] N. Dalal and B. Triggs, "Histogram of oriented gradients for human detection", IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pp 1063 – 6919, 2005.
- [16] D. Comaniciu, V. Ramesh and P. Meer, "Kernel-based object tracking", IEEE Transactions on Pattern Analysis and Machine Intelligence, pp 564 – 575, 2003.
- [17] G. Bradski, "Computer Vision Face Tracking For Use in a Perceptual User Interface", Proc. IEEE Workshop Applications of Computer Vision, pp. 214 – 219, October 1998.
- [18] A. Bhattacharyya, "On a measure of divergence between two statistical populations defined by their probability distribution", Bulletin of the Calcutta Mathematical Society 35, pp. 99 – 110, 1943.
- [19] S. Kullback and R. A. Leibler, "On information and sufficiency", Annals of Mathematical Statistics 22, pp. 79 – 86, 1951.
- [20] RGB Colour Space, <http://gimp-savvy.com/BOOK/index.html?node50.html>
- [21] HSL and HSV Colour Space, [http://en.wikipedia.org/wiki/HSL\\_color\\_space](http://en.wikipedia.org/wiki/HSL_color_space)
- [22] R. W. Hunt. "Measuring Colour 3<sup>rd</sup> Edition", Fountain Press, England, pp. 39 – 46, 1998.
- [23] YIQ Colour Space, <http://en.wikipedia.org/wiki/YIQ>
- [24] K. Plataniotis & A Venetsanopoulos, "Color Image Processing and Applications 1<sup>st</sup> Edition", Springer, pp. 4, 2000.

# Appendix



## Best Individual Results – CAVIAR

Object	Id	Method	Colour Space	Histogram	Angles	Radii
	36	CCPD	RGB	8x8x8	6	2
	21	CCPD	YIQ	12x12x12	12	1
	14	CCPD	XYZ	12x12x12	10	2
	34	CCPD	YIQ	12x12x12	10	2
	8	H	RGB	3x3x3		







	9	CCPD	XYZ	8x8x8	4	1
	22	CCPD	HSV	3x3x3	6	1
	32	CCPD	HSV	8x8x8	6	1
	29	CCPD	XYZ	12x12x12	6	1
	1	CCPD	HSV	3x3x3	4	1
	15	TBH	HSV	12x12x12		
	25	CCPD	YIQ	3x3x3	10	1








	5	CCPD	YIQ	12x12x12	10	1
	7	TBH	XYZ	12x12x12		
	20	CCPD	YIQ	12x12x12	12	1
	3	TBBPH	HSV	12x12x12		
	4	TBH	HSV	3x3x3		
	13	CCPD	XYZ	3x3x3	6	1
	23	CCPD	YIQ	3x3x3	6	1



	33	TBBPH	YIQ	8x8x8		
	19	TBBPH	HSV	8x8x8		



Worst Individual Results - CAVIAR

Object	Id	Method	Colour Space	Histogram	Angles	Radii
	36	H	YIQ	3x3x3		
	21	TBBPH	YIQ	3x3x3		
	14	H	YIQ	3x3x3		
	34	H	YIQ	3x3x3		
	8	H	YIQ	3x3x3		
	9	TBBPH	YIQ	3x3x3		

	22	H	YIQ	3x3x3		
	32	TBBPH	YIQ	3x3x3		
	29	H	YIQ	3x3x3		
	1	H	YIQ	3x3x3		
	15	TBBPH	YIQ	3x3x3		
	25	H	HSV	12x12x12		
	5	H	YIQ	3x3x3		








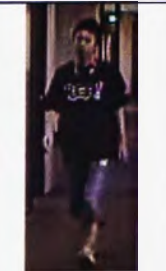
	7	TBBPH	YIQ	3x3x3		
	20	H	YIQ	3x3x3		
	3	TBBPH	YIQ	3x3x3		
	4	TBBPH	RGB	3x3x3		
	13	H	YIQ	3x3x3		
	23	H	YIQ	3x3x3		




	33	H	YIQ	3x3x3		
	19	H	YIQ	3x3x3		





Best Individual Results - NICTA

Object	Id	Method	Colour Space	Histogram	Angles	Radii
	10	CCPD	YIQ	12x12x12	12	2
	8	H	XYZ	8x8x8		
	9	CCPD	XYZ	12x12x12	4	1
	13	H	YIQ	12x12x12		





	1	TBH	RGB	8x8x8		
	2	CCPD	YIQ	12x12x12	12	1
	3	CCPD	YIQ	3x3x3	12	2
	4	TBBPH	YIQ	12x12x12		
	5	TBH	RGB	3x3x3		







	6	TBH	YIQ	8x8x8		
	7	CCPD	YIQ	12x12x12	10	2
	11	TBH	HSV	8x8x8		

**Worst Individual Results - NICTA**

Object	Id	Method	Colour Space	Histogram	Angles	Radii
	10	TBBPH	XYZ	3x3x3		
	8	TBBPH	YIQ	3x3x3		




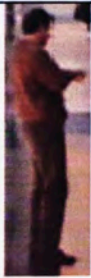









	9	TBBPH	YIQ	3x3x3		
	13	CCPD	YIQ	3x3x3	4	2
	1	TBBPH	XYZ	3x3x3		
	2	TBBPH	YIQ	3x3x3		







	3	H	YIQ	3x3x3		
	4	H	YIQ	3x3x3		
	5	TBBPH	HSV	3x3x3		
	6	H	YIQ	3x3x3		
	7	H	YIQ	3x3x3		
	11	H	YIQ	3x3x3		










**Best Individual Results – NICTA-CAVIAR**





Object	Id	Method	Colour Space	Histogram	Angles	Radii
	n11	TBH	HSV	8x8x8		
	c22	CCPD	HSV	3x3x3	6	1
	c1	CCPD	HSV	3x3x3	4	1
	c14	CCPD	XYZ	12x12x12	10	2
	n4	TBBPH	YIQ	12x12x12		
	c23	CCPD	YIQ	3x3x3	6	1

	c34	CCPD	YIQ	12x12x12	10	2
	c13	CCPD	XYZ	3x3x3	6	1
	c33	TBBPH	YIQ	8x8x8		
	c32	CCPD	HSV	8x8x8	6	1
	n7	CCPD	YIQ	12x12x12	10	2
	n13	TBH	YIQ	12x12x12		

	c8	H	RGB	3x3x3		
	c15	TBH	HSV	12x12x12		
	c19	TBBPH	HSV	8x8x8		
	n5	TBH	RGB	3x3x3		
	n8	TBH	RGB	8x8x8		
	n10	CCPD	HSV	3x3x3	10	1

	c20	CCPD	YIQ	12x12x12	12	1
	c9	CCPD	XYZ	8x8x8	4	1
	c36	CCPD	RGB	8x8x8	6	2
	c21	CCPD	YIQ	12x12x12	12	1
	c5	CCPD	YIQ	12x12x12	10	1
	n3	CCPD	YIQ	3x3x3	12	1
	c7	TBH	XYZ	12x12x12		









	c4	TBH	HSV	3x3x3		
	c25	CCPD	YIQ	3x3x3	10	1
	n2	CCPD	YIQ	12x12x12	10	2
	c29	CCPD	XYZ	12x12x12	6	1
	n6	TBH	RGB	8x8x8		



	n9	CCPD	XYZ	12x12x12	12	2
	c3	TBBPH	HSV	12x12x12		
	n1	CCPD	RGB	12x12x12	12	2








## Worst Individual Results – NICTA-CAVIAR

Object	Id	Method	Colour Space	Histogram	Angles	Radii
	n11	TBBPH	YIQ	3x3x3		
	c22	H	YIQ	3x3x3		
	c1	H	YIQ	3x3x3		
	c14	H	YIQ	3x3x3		
	n4	H	YIQ	3x3x3		
	c23	H	YIQ	3x3x3		






	c34	H	YIQ	3x3x3		
	c13	H	YIQ	3x3x3		
	c33	H	YIQ	3x3x3		
	c32	TBBPH	YIQ	3x3x3		
	n7	TBBPH	YIQ	3x3x3		
	n13	TBBPH	RGB	3x3x3		

	c8	H	YIQ	3x3x3		
	c15	TBBPH	YIQ	3x3x3		
	c19	H	YIQ	3x3x3		
	n5	TBBPH	HSV	3x3x3		
	n8	TBBPH	YIQ	3x3x3		
	n10	TBBPH	YIQ	3x3x3		



	c20	H	YIQ	3x3x3		
	c9	TBBPH	YIQ	3x3x3		
	c36	H	YIQ	3x3x3		
	c21	TBBPH	YIQ	3x3x3		
	c5	H	YIQ	3x3x3		
	n3	H	YIQ	3x3x3		
	c7	TBBPH	YIQ	3x3x3		



	c4	TBBPH	RGB	3x3x3		
	c25	H	HSV	12x12x12		
	n2	TBBPH	YIQ	3x3x3		
	c29	H	YIQ	3x3x3		
	n6	H	YIQ	3x3x3		

	n9	TBBPH	XYZ	3x3x3		
	c3	TBBPH	YIQ	3x3x3		
	n1	TBBPH	YIQ	3x3x3		



