# Fast, Robust Colour Vision for the Monash Humanoid

**Andrew Price Geoff Taylor Lindsay Kleeman**

Intelligent Robotics Research Centre
Monash University
Clayton 3168 , Australia
andrew.price@eng.monash.edu.au geoffrey.taylor@eng.monash.edu.au
lindsay.kleeman@eng.monash.edu.au

## Abstract

*The Monash Humanoid Robot Project has been in progress for one year. A lightweight plastic humanoid robot has been designed and various parts are in production. This paper details the first of the sensor systems to be examined for the robot*

*A simple, fast, modified HSV colour model is presented to aid in the reliable separation and filtering of colours within images. Early experimental work has focused on tracking of moving targets, using a camera mounted on a pan-tilt head. The system uses image subtraction and colour separation to identify and track the dominant colour of a moving target.*

## 1. Introduction

The Monash Humanoid Robot Project commenced at the Intelligent Robot Research Centre in 1999. The project brief was to design and construct a humanoid robot capable of moving and operating within the normal human environment. The robot was to support a number of research efforts including humanoid motion, sensory perception (vision, hearing, tactile sensing and smell), control systems, and human-robot interaction. One of the greatest challenges in any robot project is sensor integration. The simple act of attaching a sensor to a robot and have it perform in a desired fashion is a significant issue in the design. In addition, when a robot enters life as a research system, it is bound to suffer damage or modification as ideas change and accidents happen. Metal robots are particularly inconvenient on both counts. The Monash humanoid [1] is based almost entirely on plastic, which not only reduces the time for replacement of damaged components but also offers unique possibilities for sensor embedding. In addition, components with very similar characteristics can be generated quickly and easily. The Monash Humanoid was designed to emulate the sensor capabilities of a human. The head component is a mono-piece construction capable of accepting microphones, odor sensors (in the roof of the mouth) and cameras as well as various voice reproduction systems. The design of the head incorporates sufficient lightweight components such that it is possible to mount a localized processor for image analysis inside as well. Vision is provided by two lightweight low cost CCD board cameras mounted in the eye sockets.

Figure 1. The Monash Humanoid.

This paper presents early research efforts into equipping the Monash humanoid with colour vision sensing. The current emphasis of our research is on speed (25 Frames/sec) and robustness for low level processing tasks, so as to maximize the available time for high level decision making processes. A modified Hue Saturation Value (HSV) model to aid in determining appropriate characteristics for colour filters for use in fast video capture and tracking of multiple objects of multiple colours is described in Section 2. Section 3 presents results showing the ability of the model, and the filters it produces, to separate difficult combinations of colours such as red, orange, yellow and white under incandescent lighting. In addition colour filtering, a number of fast, low level image processing tools have been developed using MMX instuctions as described in Section 4. Finally, Section 5 presents an early application for the image processing tools developed here. This involves a tracking system, using cameras mounted on a pan-tilt head, with the ability to sense a moving object and 'lock' on to its dominant colour. Results presented in this paper lead towards a simple, efficient means for reliable, adaptive colour filtering and establish a mechanism by which the humanoid can track multiple objects of different colours, with and without *a priori* knowledge of the colours involved.

## 2. Colour Filtering in a modified HSV space.

Lighting conditions often cause significant problems in robotics vision systems. It was for this reason that a modified Hue, Saturation, Value (HSV) space was devised to aid in the process of designing accurate colour filter models and to evaluate the effects of specular reflection when lighting conditions were variable and non-determinable in advance. Lighting problems are compounded when cameras with auto iris features are used. Tuning vision systems to suit dark regions may cause more brightly illuminated regions to wash out to white. Conversely tuning to bright regions cause darker regions to wash out to black [2][3].

The majority of mobile robots have limited processing capabilities on board. This means that every consumer of those resources must minimize their requirements. While the native format of many image capture systems is some form of RGB space, filtering in such an environment is not always as accurate as would be desired. And yet transforming the image to some other space consumes valuable processor resources. To compound this problem, in a highly dynamic environment, a high frame rate is desirable, requiring all systems to consume minimum processing time.

The modified HSV space begins with the conventional HSV model [4] and is extended beyond the usual 360 degrees of Hue (Figure 2a). By extending the model in such a way some colours appear more than once however the purpose of the frame work is to generate a mechanism where all colours can be viewed in a continuous distribution. The most obvious benefit of which is that colours centred around 0 degrees Hue are also represented as a continuous distribution around 360 degrees Hue.

By mapping the modified HSV space onto three linear axes as shown in figure 2b it was possible to form a useful 3D representation of the distribution of colour within an image. By eliminating the effects of the circular

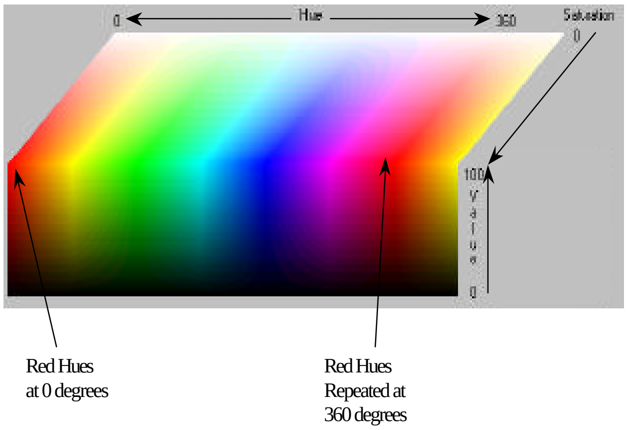



Figure 2. The Modified Hue Saturation Value Model.

discontinuity all of the colours associated with any object are now visible in localized regions. The objective of the model is not to represent each colour once, but to improve the visualization of the colour distribution and so make filter design more homogenous.

The object of filter design within the modified HSV model was to differentiate between the distributions of colour attributable to each object within the image with particular emphasis on difficult separation tasks including red-orange-yellow and dark shades of green and blue with black. In addition the model had to contend with the influence of incandescent floodlights ranging in intensity from 600 to 1100 lux at floor level, and up to 3000 lux on the surface of objects. By observing objects of different colours both in isolation and combination it was evident that a filter within the modified HSV space would have the shape of a spheroid or part thereof, truncated by the limits of the saturation and value axes. The volume of the spheroid, when mapped onto the modified HSV model would optimally contain all of the colours which are evident in a particular object whilst excluding the colours of different objects and unassociated colours which are noise.

A filter, in a three dimensional space consists of some form of volume representing a subset of all the points in the space. The volume may be used as a means of including or excluding certain data. From the distributions of Figure 3 it is apparent that in the modified HSV model, the size and shape of colour distributions vary considerably. This is particularly evident in darker shades such as dark blue and light green (almost yellow). Colours whose saturation and value are both very high make simple test examples for any vision system. Red, green, blue, cyan, yellow and magenta are all examples whose distributions are easily separated. However it is difficult to obtain pure colours in the real world.

By projecting the distribution of pixels in an image onto planes, it was noted that the distribution in either the hue-saturation or the hue-value planes could be contained with suitable optimality by either an ellipse or part of an ellipse.

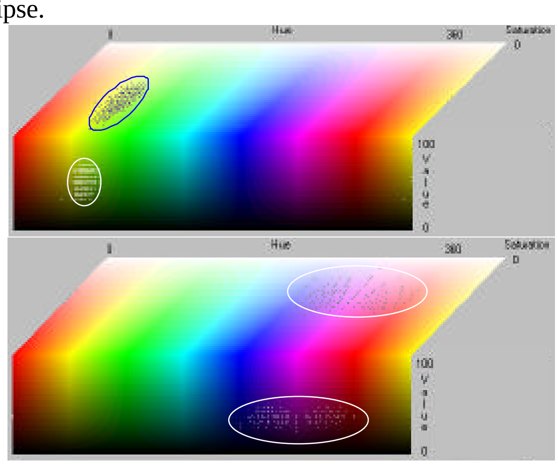


Figure 3 Elliptical shaped distributions

Extrapolating this realization to three dimensions yields the probability that an ellipsoid can be used to contain distributions in the modified HSV model. An ellipsoid is essentially a ‘squashed sphere’ as shown in Figure 4.

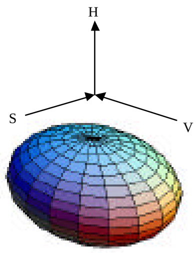



Figure 4 Ellipsoid: A more optimal filter shape in modified HSV space

For the purpose of developing a filter the Cartesian form of the surface equation for a spheroid is used. The generalized form allows the positioning of the ellipsoid within the modified HSV model

$$\frac{(x+k)^2}{a^2}+\frac{(y+l)^2}{b^2}+\frac{(z+m)^2}{c^2}\le 1 \qquad \begin{array}{l}(0\le k\le 360)\\(0\le l\le 100)\\(0\le m\le 100)\end{array}$$

Equation 4

where $(k,l,m)$ is a point in the modified HSV model, $k$ is the nominal hue of the filter, $l$ is the nominal saturation and $m$ is the nominal value.

Thus any ellipsoid filter within the modified HSV space can be specified by a minimum of six parameters, three for the centroid of the ellipsoid, three defining the length of the semi-axis in each plane.

Figure 5 shows the results of a filter whose centroid is located at Hue = 0, Saturation = 100 and Value = 100. This filter design corresponds to shades of red.

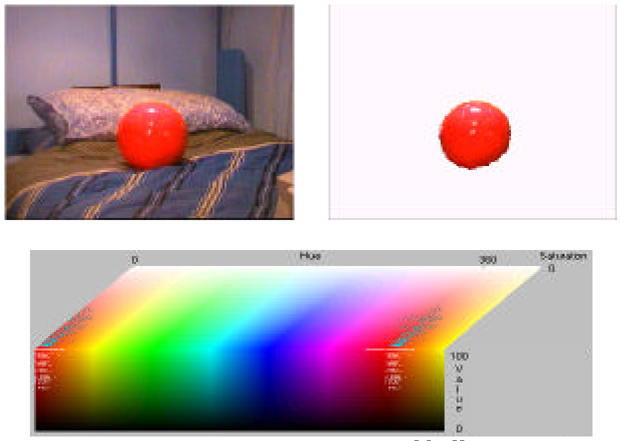


Figure 5. Extracting a red ball.
(Note the repeat of half the red pixels at 0 Hue. The filter itself is constructed around 360 Hue)

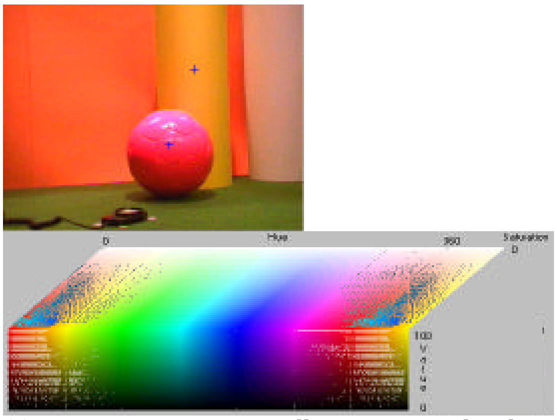
## 3. Modified HSV Filtering Results

Figures 6 and 7 demonstrate two examples of filtering that can be achieved reliably and repeatably using filters devised in a modified HSV space.

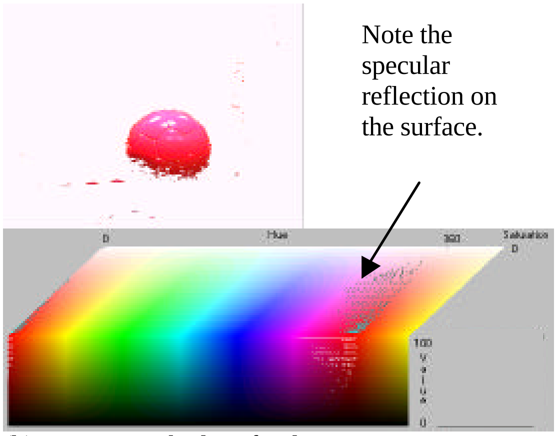
Figure 6 demonstrates separation of red, yellow, orange and white. This test was performed under lighting conditions varying from 600 to 1200 lux at floor level. (The intensity above floor level was considerably higher. Intensity at floor level was used as a base for a wide range of tests). Figure 7 Demonstrates the ability of the modified HSV model to successfully separate very dark shades, namely dark blue and black.

Filters can be reliably constructed in the modified HSV space to take into account a broad range of lighting conditions. This has the effect of minimizing problems caused when auto iris cameras adjust to changes in light, making the appearance of colours change and potentially destabilizing less robust filter techniques. By mapping the filter back into RGB space, the native form of data from many video capture cards, and using a simple lookup table a high speed (25 frames/second) filter process is achieved in optimal time (ie one comparison per pixel).

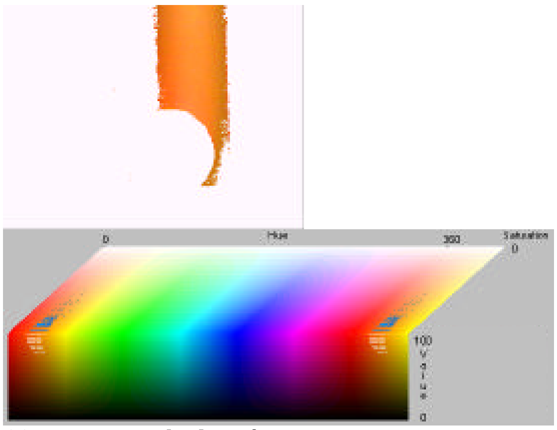
Recent research has been directed towards adaptive selection of filter parameters to provide a vision system with flexible, autonomous and robust colour filtering. The vision system is described later in Section 5.



(a) An image containing yellow, orange red and various reflections.




(b) Extracting shades of red.



(c) Extracting shades of orange

Figure 6. Separating yellow, orange and red from white with incandescent lighting.

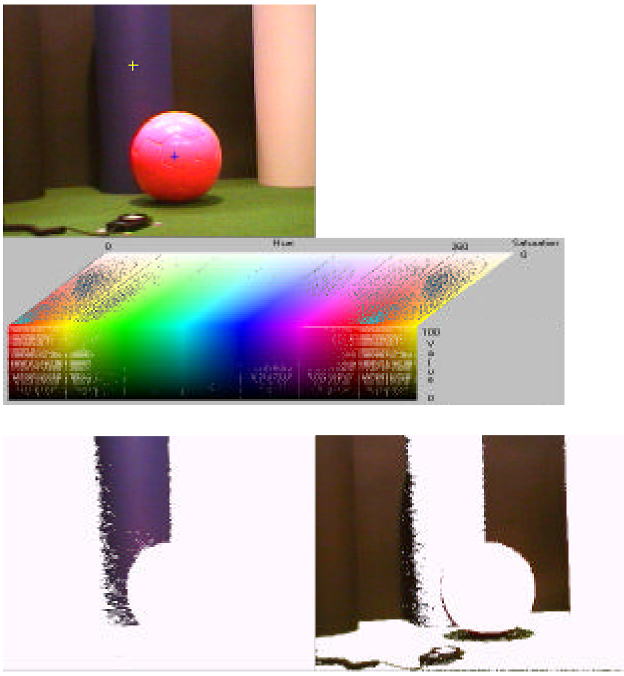


Figure 7. Separating Dark Blue and black

## 4. Fast Image Processing with MMX

In addition to the colour filter design described above, much effort has gone into building a set of fast and robust image processing tools for low level operations such as image subtraction, thresholding, edge detection, binary morphological filtering and binary image analysis. The ultimate goal is to perform these operations on 400×300 pixel stereo images at video frame rates (25 frames/sec), thus providing a solid foundation for building higher level computer vision systems. In some instances, the image processing can be accelerated by exploiting *a priori* knowledge: analyzing a small 'region of interest' rather than the entire image, or performing the analysis at a reduced resolution where appropriate. However, a particular interest in our development effort has been the use of accelerated hardware techniques, specifically, the use of MMX instructions on Intel Pentium based machines. MMX includes a SIMD (Single Instruction, Multiple Data) instruction set that allows a limited number of operations to be performed on up to eight data elements at a time. While this addition to the Intel architecture was initially intended to enhance 'multi-media' applications, we have found that MMX is capable of performing a function similar to specialized digital signal processing hardware. This makes MMX ideal for low level image processing.

Currently, MMX has been used to implement thresholding, subtraction, convolution, edge detection and morphological filtering. In all cases the code is hand-written in assembly language, and processing time is reduced by a factor of 2.5 or more when compared to code written in a high-level language (C++). For instance, subtraction (a central component of stereo correspondence analysis) of two 8-bit 400×300 pixel images on a 450MHz Pentium III takes approximately 5.2 ms using code written in C++ and generated by *gcc* (with appropriate optimizations enabled), and only 1.3 ms for code written with MMX instructions as inline assembly in C++ and generated using the same compiler. This represents a reduction in processing time by a factor of four, with significant implications for stereo image processing.

Unfortunately, MMX was found to have limited applicability to image processing techniques such as connectivity analysis (binary segmentation) and filtering via look-up tables. This is mainly due to the limited instruction set: MMX registers cannot be used to index memory locations, or as the source for conditional jumps. Implementing connectivity analysis using MMX instructions may be possible using a non-conventional implementation of the algorithm, and further work is proceeding in this area.

# 5. Pan-Tilt Object Tracking Using Dynamic Colour Filtering

A colour-based object tracking system was implemented to explore the capabilities of the image processing platform developed by the authors. It is also anticipated that this work may find use in future areas of investigation such as visual servoing. The system uses a pair of cameras mounted on a Biclops pan-tilt head, although only a single camera is used for this application. The tracking system is able to fixate on any moving object within the image, and the pan-tilt head orients the camera so that the target is always near the centre of the image.

The steps followed to acquire a visual target are illustrated in figures 8 to 11. Figure 8 shows the view from the pan-tilt head of the author waving a multi-coloured (blue, green and orange) mouse pad in an attempt to gain the attention of the tracking system. By subtracting the intensity components of successive frames in the video stream, and with the camera in a fixed position, the system is able to identify some of the pixels associated with the moving object. Figure 9 shows a typical difference image after thresholding and morphological filtering to remove noise. Parts of the mouse pad and forearm are clearly visible. Note that some of the 'difference pixels' belong to the moving object while others belong to the portion of background uncovered as the object moved.

If the camera were to remain fixed, the difference pixels themselves could be used to track the movement of the object. However, our aim is to have the pan-tilt head orient the camera so that the moving object is always near the centre of the image. Unfortunately, any movement of the camera causes the entire background to shift between successive frames, so that the movement of individual objects is obscured. It is therefore necessary to extract some feature of the object which is invariant to camera motion and can be used for tracking. Colour was chosen as a prominent and easily identified feature suitable for this purpose. To facilitate the implementation of colour filtering, this system uses a simplified version of the filters described earlier. The target colour is selected by compiling a histogram of the hue/saturation components for the difference pixels. In figure 10, colours associated with difference pixels are plotted as white dots. By accumulating the histogram over five successive frames, difference pixels belonging to the uncovered background (which change colour from frame to frame, assuming the background is cluttered) accumulate less than the difference pixels associated with the target (which remain a fixed colour between frames). The hue/saturation at the global maximum in the histogram is chosen as a candidate for the dominant colour of the moving target. The white square in figure 10 indicates that the global maximum for the mouse pad occurs at a blue hue.

Figure 11 was generated by passing figure 8 through a band-pass colour filter centred at the candidate hue/saturation, and shows that the system correctly identified the dominant colour of the moving target. Before tracking commences, the candidate colour is validated by comparing the centroid of the target colour (white pixels in figure 11) with the centroid of the difference pixels (white pixels in figure 9), to ensure that the colour belongs to the moving object. The white cross in figure 8 locates the centroid of the target colour, after validation. Tracking is implemented by driving the pan-tilt joints at velocities proportional to the distance of the target colour centroid from the centre of the image. The system continues to track the same colour until the target becomes stationary, and then reverts back to searching for a new moving target and corresponding candidate colour.

Using the fast MMX image processing operations described in the previous section, the system is able to acquire and track targets at 25 frames/sec. Most objects exhibiting some prominently coloured region are capable of being tracked, provided the same colour does not dominate the background. Skin tones, such as a waving hand, are difficult to track in the current implementation since the image of a person may contain a number of skin

Figure 8. View from the pan-tilt head of the author waving a multi-coloured mouse pad in a cluttered background.

Figure 9. Difference pixels from subtraction of successive frames highlights movement (here, the mouse pad and the author's arm) .

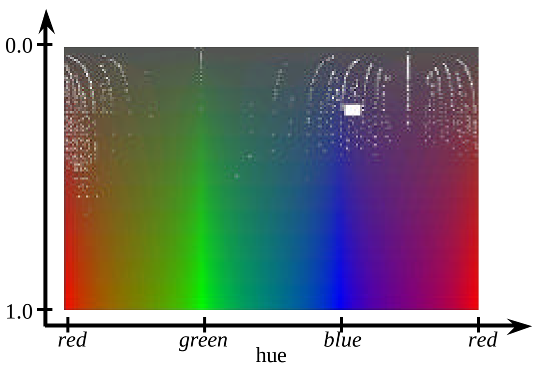



Figure 10. Histogram of difference pixel colours accumulated over five successive frames. The white square locates the maximum (in this case, the blue of the mouse pad).

coloured regions. The system could be improved by incorporating binary segmentation and further analysis, but a fast binary segmentation algorithm is still in development. Nevertheless, the current implementation provides a solid basis for further work.

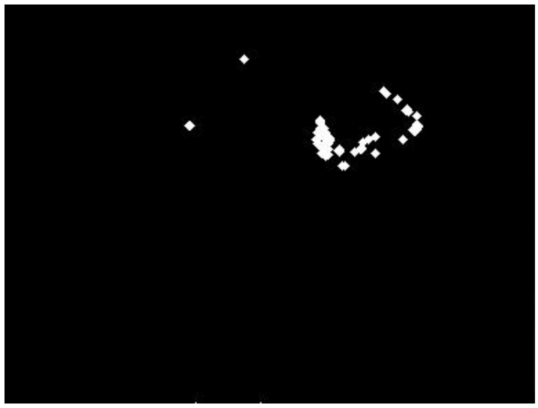


Figure 11. Filtering of the dominant object colour. White pixels identify the blue region of the moving mouse pad. The centroid of these pixels give the location of the target (shown by the white cross in figure 8).

## 6. Conclusions

This paper presents a robust, simple approach to determining appropriate filter constraints for the various colours within an image based on a modified Hue, Saturation, Value model. The model, and the filters it produces have been repeatedly tested over a wide range of lighting conditions and found to reliably separate difficult combinations such as orange, red and yellow as well as black and dark blue without complex and time consuming mathematics or transformations.

Using the fast, robust colour filtering and MMX image processing routines, a pan-tilt object tracking system was developed to dynamically acquire and track the dominant colour of moving targets. The high frame rates achieved by this system (25 frames/sec) demonstrate the value of exploiting MMX for data-intensive image processing operations. The tracking system provides a solid basis for further research into vision sensing.

## References

[1] Price A.R. , Jarvis R. , Russell R. A. , Kleeman L. "A Lightweight Plastic Robotic Humanoid", Submitted to IROS2000, 2000.

[2] Kneen, J., Price A.R.,``RoboCup 97 : An Omnidirectional Perspective," Official Book RoboCup 1: Springer Lecture notes in Artificial Intelligence 1395, pp. 320-331, 1998.

[3] Price A.R. , "Mobility and Vision for Mobile Robots in Non-Deterministic, Competitive Environments", Ph.D. Thesis, Deakin University, 1999.

[4] Foley, J.D., A. v.Dam, S.K. Feiner, J. F. Hughes and R. L. Phillips, Introduction to Computer Graphics. Reading, Massachusetts, Addison-Wesley, 1993