Motion Observability Analysis of the Simplified Color Correlogram for Visual Tracking

Qi Zhao and Hai Tao

Department of Computer Engineering
University of California at Santa Cruz, Santa Cruz, CA 95064
{zhaoqi,tao}@soe.ucsc.edu

Abstract. Compared with the color histogram, where the position information of each pixel is ignored, a simplified color correlogram (SCC) representation encodes the spatial information explicitly and enables an estimation algorithm to recover the object orientation. This paper analyzes the capability of the SCC (in a kernel based framework) in detecting and estimating object motion and presents a principled way to obtain motion observable SCCs as object representations to achieve more reliable tracking. Extensive experimental results demonstrate the reliability of the tracking procedure using the proposed algorithm.

1 Introduction

The computer vision community has witnessed the development of several excellent tracking algorithms using color statistics, e.g., color histogram, as representations. There statistics features can be convolved with an isotropic kernel to allow gradient estimation of the representation [1]. One inherent limitation of such kernel based methods is the singularity problem, where the representation is blind to certain motion.

Most existing kernel based tracking algorithms are concerned only with the tracking of object locations [1], or object locations and scales [2]. Since isotropic kernels are often used [1,3,4], rotational motion can not be estimated using these methods. Recently, Zhao and Tao [5] proposed the simplified color correlogram (SCC) representation to efficiently track location and orientation simultaneously. Although the kernel is also rotationally-symmetric, the underlying SCC representation is sensitive to orientation changes. This property makes the representation capable of tracking rotational motion as well as translational motion. As in most kernel based algorithms, the assumption is that the statistics of the SCC feature be sufficient to determine the motion of the object [4]. However, this assumption needs to be validated. This paper shows that under certain degenerated cases, ill-conditioned cases may occur, i.e., translational and/or rotational motion may not cause changes to the SCC, therefore the motion is not observable. In this study, we derive a criterion to evaluate the numerical stability of the tracking solution, according to which, schemes for the SCC selection are designed.

The paper is organized as follows. Section 2 reviews the simplified color correlogram (SCC). Section 3 analyzes the properties of the SCC in a kernel based

Y. Yagi et al. (Eds.): ACCV 2007, Part I, LNCS 4843, pp. 345–354, 2007.

framework, and proposes the solution to obtain the motion observable SCCs. Section 4 discusses implementation details. Section 5 shows experimental results and demonstrates the improvement over the standard mean shift algorithm, and section 6 concludes the paper.

2 Introduction to Simplified Color Correlogram (SCC)

Color correlogram expresses the correlation between color pairs in an image, which has been commonly used in the literature of image retrieval/indexing [6]. Zhao and Tao [5] has recently proposed a simplified version of color correlogram (SCC) for tracking purpose. Instead of including pixel pairs along all directions and with a set of distances, the SCC only counts pairs along one or several axes, i.e., pre-selected directions, with predefined distances. Formally, the SCC with L axes are defined as

$$S_{u,v} = Pr(I(p_1) = u \wedge I(p_2) = v \mid f(p_1 - p_2) = (\theta, d)). \tag{1}$$

Here, f is a function to obtain the direction and the distance of a pixel pair, representing the spatial relationship of two pixels p_1 and p_2. $\theta \in \{\theta^l, l = 1, ..., L\}$ and $d \in \{d^l, l = 1, ..., L\}$, where L is the number of axes, θ^l is the direction of axis l and d^l is the pair distance along axis l. We use the L_2 norm to measure the distance between pixels.

Though the conclusions about the singularity problem in the kernel based representations made in this study is not restricted to the SCC representation, we focus our discussions on the SCC due to the following reasons [5]:

1. The SCC achieves a natural integration of color and spatial information.
2. Since certain directions are emphasized, the SCC is effective in manifesting rotational variations.
3. The SCC is computationally inexpensive.
4. Being a middle ground of the template based methods and the pure statistics based representations, the SCC can have desired properties from both sides.

Similar to the conventional color histogram, the SCC is also integrated into a kernel framework to allow efficient localization, thereby the singularity problem exists. This paper presents two methods to approach this issue in the SCC context, where both translation and rotation are concerned. One is to select more than one axis to form a multi-axis SCC, therefore motion ignored by pairs along one axis can be recovered along other axes, as will be justified later. This strategy suffices in most cases. However, efficiency consideration suggests an alternative of obtaining one optimal axis and its corresponding pair distance, so that the resulting SCC is the most sensitive to all different motion. Details of the two approaches are provided in the next section.

3 Motion Observability Analysis

The SCC based tracking method enables the detection of both translational and rotational motion, therefore reliable tracking in this context requires that both types of motion be distinctly observed and reliably recovered.

3.1 Representing Objects Using SCC

In this subsection, we first address necessary notations for the later analysis. Specifically, each pair of pixels counted in the SCC representation can be parameterized using a 3-dimensional vector $\Phi = [cx, cy, \theta]^T$, where (cx, cy) are the image coordinates of the midpoint of each pair and θ is the angle between the axis of the object and the object coordinate system. Consider for a moment a target model/candidate for the SCC with one axis l and the index l is omitted to keep the notation simple.

Target Model. Similar to [4], we define the matrix form of the target model as

$$M = \alpha U_M^T K(0). \tag{2}$$

In Eqn.2, $U_M = [\mathbf{u_{11}}, \mathbf{u_{12}} \cdots \mathbf{u_{1m}}, \mathbf{u_{21}} \cdots \mathbf{u_{2m}} \cdots \mathbf{u_{m1}}, \mathbf{u_{m2}} \cdots \mathbf{u_{mm}}]$, where $\mathbf{u_{rs}} = [\delta(I(\Phi_{111}) - I_{rs}), \delta(I(\Phi_{112}) - I_{rs}) \cdots \delta(I(\Phi_{WHO}) - I_{rs})]^T, r, s = 1 \cdots m$. $I(\Phi_{ijk})$ represents the colors of the pixel pair Φ_{ijk} in the image I. W, H and O are the numbers of cxs, cys and θs considered in the SCC. If the colors of pixels in the pair Φ_{ijk} are r and s, then the corresponding element in the vector is assigned 1, otherwise it is 0. The subscript M represents "model" and α normalizes the representation. $K(0) = [K(\frac{\Phi_{111}}{h}), K(\frac{\Phi_{112}}{h}) \cdots K(\frac{\Phi_{WHO}}{h})]^T$, where K is a kernel function that assigns a smaller weight to the locations and orientations that are farther from the center of the object, h is the kernel radius, and the kernel is centered at 0.

By definition, L different axes yield L different target models $M^1, ..., M^L$.

Target Candidate. Similarly, the target candidate is defined as

$$C(\Phi_0) = \beta U_C^T K(\Phi_0), \tag{3}$$

where Φ_0 is the initialized location and orientation in the current frame. β and U_C are defined the same way as α and U_M for the target model, and the subscript C denotes "candidate". $K(\Phi_0) = [K(\frac{\Phi_{111}-\Phi_0}{h}), K(\frac{\Phi_{112}-\Phi_0}{h}) \cdots K(\frac{\Phi_{WHO}-\Phi_0}{h})]^T$.

L target candidates $C^1, ..., C^L$ can be defined for L axes.

3.2 The Objective Function and Solution for Reliable Tracking

We focus on the single-axis case in this subsection and the multi-axis cases would be analyzed in the next subsection. For the mean shift based tracking algorithms [1,5], the objective is to seek the maximum of the Bhattacharyya coefficient [7]. Its well known connection with the Matusita metric opens the possibility that we analyze the Matusita metric other than the Bhattacharyya coefficient to better illustrate the inherent problem of kernel based tracking [4].

Using the notations given in section 3.1, the objective of tracking using the Matusita metric is defined as

$$min_\Phi(D_M(\Phi)) = min_\Phi||\sqrt{M} - \sqrt{C(\Phi)}||^2 = min_\Phi(\sum\nolimits_{u,v}(\sqrt{M_{u,v}} - \sqrt{C_{u,v}(\Phi)})^2), \tag{4}$$

where Φ is the object location and orientation in the current frame.

A Newton-style iterative procedure is applied to convert this optimization problem to a more explicit form (Derivations provided in the Appendix)

$$d(C(\Phi_0))^{-\frac{1}{2}} U_C^T J_K(\Phi_0) \Delta\Phi = 2(\sqrt{M} - \sqrt{C(\Phi_0)}), \tag{5}$$

where Φ_0 is the initialized object location and orientation for the current frame and $d(C(\Phi_0))$ denotes the matrix with $C(\Phi_0)$ on its diagonal. In Eqn.5, $J_K(\Phi_0) = [\nabla_\Phi K(\frac{\Phi_{111}-\Phi_0}{h}), \nabla_\Phi K(\frac{\Phi_{112}-\Phi_0}{h}) \cdots \nabla_\Phi K(\frac{\Phi_{WHO}-\Phi_0}{h})]^T$, where $\nabla_\Phi K(\frac{\Phi_{ijk}-\Phi_0}{h}) = \frac{2}{h^2}(\Phi_{ijk} - \Phi_0) g(||\frac{\Phi_{ijk}-\Phi_0}{h}||^2)$, and $g(x) = -k^{'}(x), k(||x||^2) = K(x)$.

Denoting $A = d(C(\Phi_0))^{-\frac{1}{2}} (U)_C^T J_K(\Phi_0)$ and converting the matrix before $\Delta\Phi$ to a square matrix for further analysis, we obtain

$$A^T A \Delta\Phi = 2A^T(\sqrt{M} - \sqrt{C(\Phi_0)}). \tag{6}$$

Therefore the solution to the optimization problem in this 3-dimensional case is unique if and only if the 3×3 matrix $A^T A$ is of full rank. Additionally, the stability of the solution depends on the magnitude of its condition number. In the single-axis cases, the SCC with the parameters (θ, d) corresponding to the smallest condition number of $A^T A$ is the optimal SCC.

3.3 SCC with Multiple Axes

If multi-axis correlograms are used, denote

$$A_L = \begin{bmatrix} d(C^1(\Phi_0))^{-\frac{1}{2}} (U_C^1)^T J_K^1(\Phi_0) \\ \vdots \\ d(C^L(\Phi_0))^{-\frac{1}{2}} (U_C^L)^T J_K^L(\Phi_0) \end{bmatrix},$$

and

$$A^l = d(C^l(\Phi_0))^{-\frac{1}{2}} (U_C^l)^T J_K^l(\Phi_0), \quad l = 1, .., L,$$

then we have

$$A_L^T A_L = \sum\nolimits_{l=1}^{L} (A^l)^T A^l.$$

In this paper, we explore further into this multi-axis problem considering the simple yet effective two-axis cases. A useful property of the semi-positive definite matrices $(A^1)^T A^1$ and $(A^2)^T A^2$ states as

$$\begin{aligned} &min(cond((A^1)^T A^1), cond((A^2)^T A^2)) \\ &\leq cond(A_2^T A_2) \leq max(cond((A^1)^T A^1), cond((A^2)^T A^2)), \end{aligned} \tag{7}$$

where $A_2^T A_2 = (A^1)^T A^1 + (A^2)^T A^2$.

These inequalities indicate that the condition number of a two-axis SCC is between the two condition numbers of the corresponding single-axis SCCs. A consequence is that for a SCC defined with two axes, unfavorable condition numbers are less possible to be generated, since it requires both corresponding single-axis SCCs to have sufficiently large condition numbers. To make full use of this point, the two axes should be as independent as possible and in our work, two orthogonal axes are used.

3.4 Visual Interpretations and Verifications on Example Patterns

Using the SCC as an object representation, an image patch can have different SCCs, with some being more favorable than others in terms of motion observability. To provide a visual interpretation on the SCC representation, we make further analysis into the matrix A, which is

$$d(C(\Phi_0))^{-\frac{1}{2}}U_C^T J_K(\Phi_0) = \begin{bmatrix} \frac{2}{h^2}C_{1,1}^{-\frac{1}{2}} \sum\limits_{I(\Phi_{ijk})=I_{11}} (\Phi_{ijk}-\Phi_0)g(||\frac{\Phi_{ijk}-\Phi_0}{h}||^2) \\ \vdots \\ \frac{2}{h^2}C_{1,m}^{-\frac{1}{2}} \sum\limits_{I(\Phi_{ijk})=I_{1m}} (\Phi_{ijk}-\Phi_0)g(||\frac{\Phi_{ijk}-\Phi_0}{h}||^2) \\ \vdots \\ \frac{2}{h^2}C_{m,m}^{-\frac{1}{2}} \sum\limits_{I(\Phi_{ijk})=I_{mm}} (\Phi_{ijk}-\Phi_0)g(||\frac{\Phi_{ijk}-\Phi_0}{h}||^2) \end{bmatrix}. \tag{8}$$

To obtain a unique solution for $\Delta\Phi$ in Eqn.5, at least three of the row vectors of the above matrix need to be linearly independent. In this paper, we analyze two typical image patterns that are inherently unobservable to certain motion.

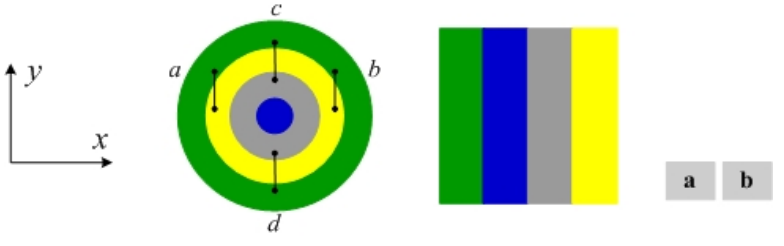

Fig. 1. Illustration for visual interpretation of SCC choice

Concentric Circles (Fig.1(a)): In color histogram based kernel methods, *concentric circles* are regarded as a degenerated case [4], where translation can not be detected. In this SCC based kernel methods, due to the spatial information encoded in the pixel pairs, translation along the SCC axis can now be observed.

Without loss of generality, we set the SCC axis to be along the y direction, as shown in Fig.1(a), and validate the translation observability through Eqn.8 by examining the weighted distance vectors $(\Phi_{ijk}-\Phi_0)g(||\frac{\Phi_{ijk}-\Phi_0}{h}||^2)$ for pixel pairs of two certain distinct colors. The cx component (recall that $\Phi = [cx, cy, \theta]^T$) of this term is cancelled out by every two corresponding pixel pairs, i.e., two symmetric pairs w.r.t. the SCC axis, like pairs a and b in Fig.1(a). However, the cy component of the term can not be always cancelled out by corresponding pairs, i.e., symmetric ones w.r.t. the x direction. For example, pair c in Fig.1(a) is of color (i, j), while its corresponding pair (pair d) is of color (j, i), therefore they

can neither cancel out each other for C_{ij}, nor for C_{ji}. Since the representation is rotationally-symmetric, the θ components are all cancelled out. As a result, the row vectors in Eqn.8 is $\kappa[0, 1, 0]^T$. The intuition behind this is that among the three degrees of motion, only the translation along the SCC axis can cause sufficient changes to the SCC. Although one may use two axes to detect translation in both dimensions, the blindness to rotation is the inherent limitation of *Concentric Circles*.

Parallel Stripes (Fig.1(b)): Independent of the axis choice in the SCC, the *parallel stripes* pattern is sensitive to motion along the x direction, while blind to motion along the y direction. However, its observability to rotation depends on the direction of the SCC axis. If the axis is defined to the be along the x direction, then the elements for the orientation dimension in $(\Phi_{ijk} - \Phi_0)g(||\frac{\Phi_{ijk}-\Phi_0}{h}||^2)$ cancel out. Intuitively, this means that slight rotation does not cause enough change to the SCC. On the other hand, if the axis is some degrees away from the x direction, then rotation makes a difference in the SCC by causing some pixels in the boundary of two stripes to be the other color.

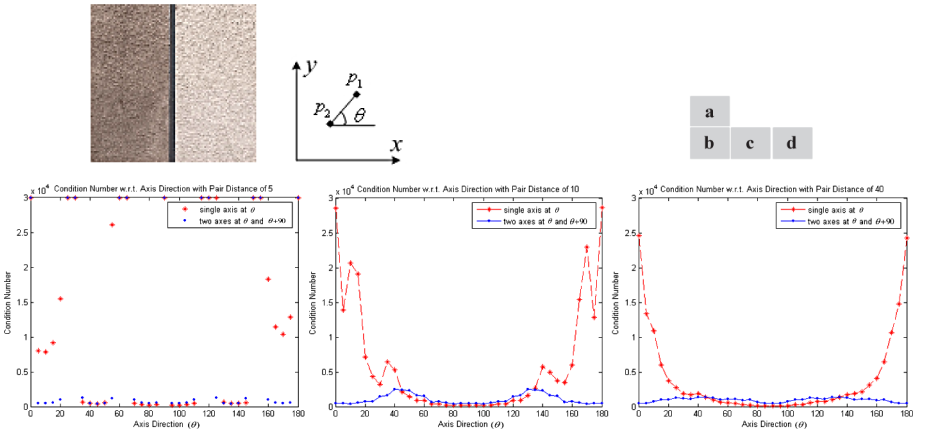

Fig. 2. Quantitative relationship between condition numbers and parameters (θ, d): (a) Image patch (96×96), (b)-(d) Condition numbers w.r.t. axis directions with different pair distances: (b) d = 5 (Symbols at the top line represent overflow values), (c) d = 10, (d) d = 40

We evaluated the quantitative relationship between condition numbers and parameters (θ, d) for the image patch of the *Parallel Stripes* pattern, shown in Fig.2(a). The axis direction θ is defined in the object coordinate system, i.e., with along the x direction being 0 degree, and increases counterclockwise. We observe the condition numbers of the single-axis SCCs with different axis directions and those of two-orthogonal-axis SCCs, where the directions of the axes are θ and $\theta + 90$. Fig.2(b-d) show the relationship of condition numbers w.r.t.

different axis directions with pair distances of 5, 10, and 40, respectively. From the illustrated outputs, following conclusions can be made:

1. The pair distance can not be too small. This is due to the discrete nature of the image and the fact that the colors for the SCC are assigned in a nearest-neighbor manner. For short pixel pairs, small image rotation may not cause any changes to the SCC. As shown in Fig.2, the results with pair distance of 5 pixels (Fig.2(b)) are much poorer than those with longer distances(Fig.2(c-d)).

2. Some axis directions are more favorable than others in terms of stability. For the patch shown in Fig.2(a), the most favorable axis direction for single-axis SCCs is along the y direction ($\theta = 90$ in Fig.2(c-d)).

3. When two axes are used, due to the inequalities given in Eqn.7, no matter what the directions of the axes are, the condition number is between the two condition numbers generated independently by the two single-axis SCCs. Fig.2(c-d) show that the average condition numbers provided by the two-orthogonal-axis SCCs are significantly smaller than those generated by the single-axis ones.

4 Implementation Details

For most real objects, textures are irregular enough to avoid the extreme cases, as shown in section 3.4, thereby a two-orthogonal-axis SCC suffices in most cases. However, in applications where speed is an important factor, single-axis SCC is greatly favored for efficiency considerations. In this case, we search for the optimal axis direction in the orientation space to obtain the SCC with the smallest condition number.

The pair distance is an important parameter in that it influences not only the SCC's sensitivity to rotation, but also the stability of the solution. On one hand, the larger the pair distance, the more observable the orientation changes, as discuessed in section 3.4; on the other hand, a SCC with a large pair distance tends to have too few pairs counted (both pixels should be in the tracking window), which decreases the stability of the tracking. By trial and error, we set the default distance to be $max(1/8(l + w), 10)$, where l and w are the length and width of the kernel size. The lower bound of 10 pixels ensures the stability of the tracker when the object is small.

In this paper, similar to [5], mean shift algorithm is extended to a translation-rotation joint domain to locate the object position and orientation simultaneously, in a gradient descent manner. However, the proposed idea can be easily incorporated into any tracking framework other than the mean shift one.

5 Experiments

The usefulness of the proposed schemes to ensure reliable tracking have been demonstrated on vehicle and pedestrian sequences under various environmental conditions. In the following, the first two real-time tracking tasks compute and use the optimal single-axis SCCs as representations, and the two-orthogonal-axis SCCs are used for the other sequences.

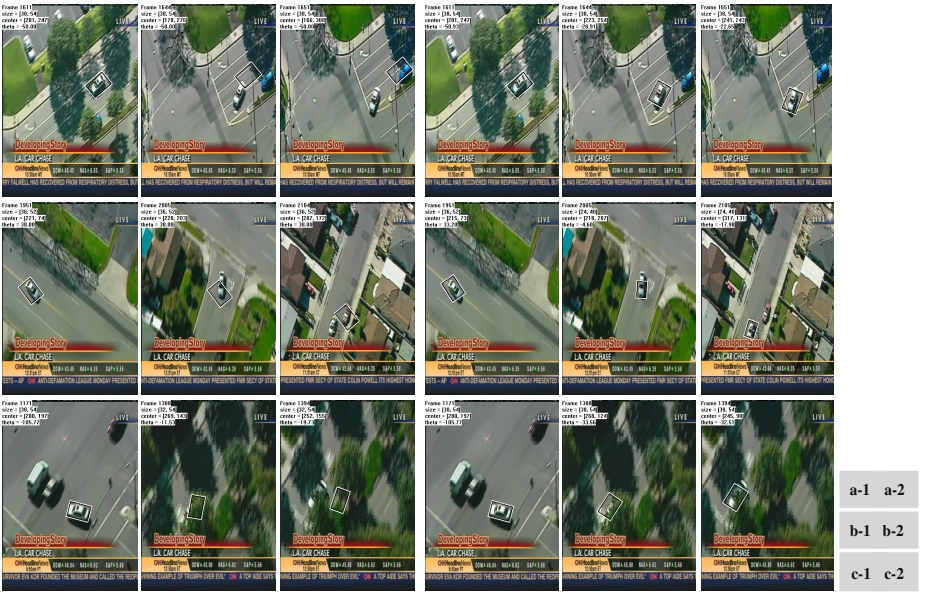

Fig. 3. *Car-Chasing* Sequence

Car-Chasing: The *Car-Chasing* is a live video sequence of 2250 frames. It has been tested using the standard mean shift (MS) based tracking algorithm, the single-axis SCC based tracking algorithms with and without the optimal SCC selection.

The possible problems in vehicle tracking using the MS tracker are revealed in Fig.3: (1) Loss of tracking tends to occur when the car makes turns (Fig.3(a-1)); (2)Fixed orientation of the tracking window causes scale adaptation difficult to realize and the mismatch of the window to the object makes the tracker sensitive to background clutter (Fig.3(b-1)). In Fig.3(c-1), although the window direction adapts to the object, using a non-optimal single-axis SCC, the tracker is less stable. In Fig.3(a-2),(b-2),(c-2), we show results of the optimal single-axis SCC based tracker, which tracks the car throughout the entire sequence. Results demonstrate that challenging issues like object rotations, heavy occlusions, background clutters, scale changes and motion blur are handled elegantly.

PETS 2001 Data: The proposed schemes are further evaluated on the *PETS 2001 Dataset* [8]. Compared with the color histogram based methods, the SCC based tracker successfully removes the restrictions brought up by certain object shapes and/or camera viewpoints. The proposed optimal single-axis SCC selection scheme further ensures both reliability and efficiency of the tracker. Sample frames of the experimental outputs are shown in Fig.4.

Multiple Human Parts: Other experiments are shown in Fig.5. The two-orthogonal-axis SCCs are used to track multiple human parts. Although only

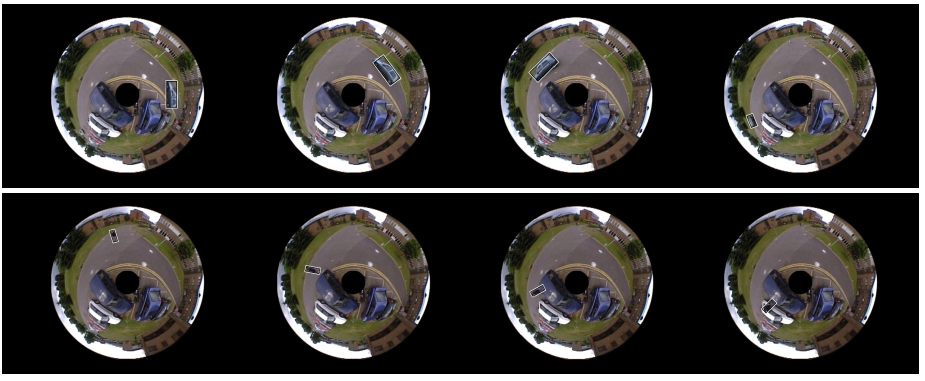

Fig. 4. *PETS 2001* Sequences

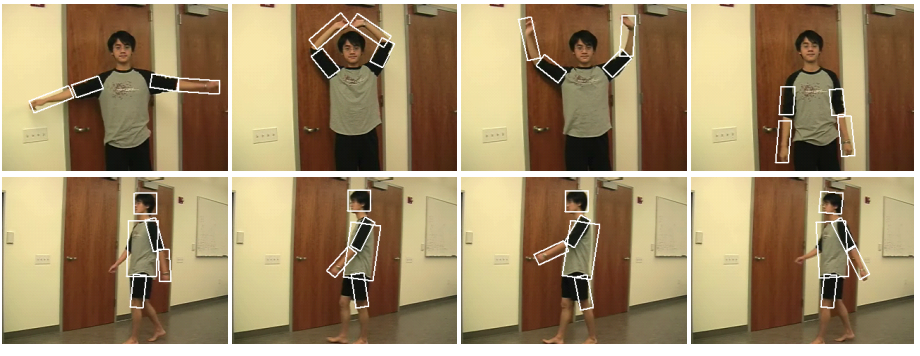

Fig. 5. *Stretching* and *Walking* Sequences

color information is extracted, the reliable tracking results indicate the algorithm's potential in being a useful module in any human tracking or behavior analysis tasks.

6 Conclusion

This paper analyzes the capability of the SCC (in a kernel based framework) in recovering both translation and rotation. A criterion to evaluate the SCC in terms of motion estimation is provided to guide the SCC selection. Two-orthogonal-axis SCCs are proved to be practical sufficient, while in tasks where speed requirements are high, optimal single-axis SCCs are desired. The discussion is focused on, but not limited to, the SCC representation. The SCC in an extended mean shift tracking framework is not computationally expensive. The tracker runs comfortably at 30 fps on a PIV 3.20GHz PC.

References

1. Comaniciu, D., Ramesh, V., Meer, P.: Kernel-based object tracking. PAMI 25(5), 564–577 (2003)
2. Collins, R.: Mean-shift blob tracking through scale space. In: CVPR, pp. 234–240 (2003)
3. Fan, Z., Wu, Y.: Multiple collaborative kernel tracking. In: CVPR II, pp. 502–509 (2005)
4. Hager, G., Dewan, M., Stewart, C.: Multiple kernel tracking with SSD. In: CVPR I, pp. 790–797 (2004)
5. Zhao, Q., Tao, H.: Object tracking using color correlogram. In: VS-PETS, pp. 263–270 (2005)
6. Huang, J.: Color-spatial image indexing and applications. PhD thesis, Cornell University (1998)
7. Kailath, T.: The divergence and Bhattacharyya distance measures in signal selection. IEEE Trans. on Comm. Tech. 15(1), 52–60 (1967)
8. `http://visualsurveillance.org/PETS2001/`

Appendix: Derivation for Eqn.5

Optimization on the objective function of $min||\sqrt{M} - \sqrt{C(\Phi)}||^2$ *results the following equation:* $d(C(\Phi_0))^{-\frac{1}{2}} U_C^T J_K(\Phi_0) \Delta\Phi = 2(\sqrt{M} - \sqrt{C(\Phi_0)})$, *where* $\Delta\Phi$ *is the motion vector in terms of both translation and rotation, and* Φ_0 *is the initialized object center in the current frame.*

Applying the Taylor Expansion on $\sqrt{C(\Phi)}$ and dropping higher order terms yields

$$\sqrt{C(\Phi)} = \sqrt{C(\Phi_0)} + \frac{d\sqrt{C(\Phi)}}{d\Phi}|_{\Phi=\Phi_0} \Delta\Phi. \tag{9}$$

Since $C(\Phi_0) = U_C^T K(\Phi_0)$, therefore $\frac{dC(\Phi)}{d\Phi}|_{\Phi=\Phi_0} = U_C^T \nabla K(\Phi_0)$. Introducing this into Eqn.9, we have

$$\sqrt{C(\Phi)} = \sqrt{C(\Phi_0)} + \frac{1}{2} d(C(\Phi_0))^{-\frac{1}{2}} U_C^T \nabla K(\Phi_0) \Delta\Phi, \tag{10}$$

where $d(C(\Phi_0))$ is the matrix with $C(\Phi_0)$ on its diagonal.

Rewrite the objective function in terms of the motion vector $\Delta\Phi$, we obtain

$$argmin_{\Delta\Phi} ||\sqrt{M} - \sqrt{C(\Phi_0 + \Delta\Phi)}|| \tag{11}$$

Substituting Eqn.10 into Eqn.11, the resulting objective function is

$$argmin_{\Delta\Phi} ||\sqrt{M} - \sqrt{C(\Phi_0)} - \frac{1}{2} d(C(\Phi_0))^{-\frac{1}{2}} U_C^T \nabla K(\Phi_0) \Delta\Phi||, \tag{12}$$

the solution of which equates to the solution of the linear system

$$\frac{1}{2} d(C(\Phi_0))^{-\frac{1}{2}} U_C^T \nabla K(\Phi_0) \Delta\Phi = \sqrt{M} - \sqrt{C(\Phi_0)}. \tag{13}$$

Denoting $\nabla K(\Phi_0)$ as $J_K(\Phi_0)$ and scaling both sides up by a factor of 2 results

$$d(C(\Phi_0))^{-\frac{1}{2}} U_C^T J_K(\Phi_0) \Delta\Phi = 2(\sqrt{M} - \sqrt{C(\Phi_0)}). \tag{14}$$