# Layered image model using binary PCA transparency masks

Zoran Zivkovic
ISLA Lab, University of Amsterdam, The Netherlands
zivkovic@science.uva.nl

### Abstract

The "layered image model" [13] represents an image sequence as a composition of 2D layers where each layer corresponds to a different object. A layer is described by its appearance and its transparency mask. The transparency masks are used to combine the layers. In this paper we present a probabilistic layered model that uses the "logistic principal component analysis (PCA)" to describe the masks. The Gaussian based factor analysis was used previously but it does not consider the constraints imposed on the transparency values. The "logistic PCA" models the transparency values that are between 0 and 1 more naturally using Bernoulli distributions. The presented model can be used to automatically extract low dimensional representation of the transparency maps of the moving objects from a video sequences more efficiently.

## 1 Introduction

In the layered representation [13] a video sequence of a 3D scene is decomposed into a set of 2D layers where each layer corresponds to a different moving object. This is a potentially very effective representation for automatically analyzing video sequences since the representation greatly simplifies the geometry but still accounts for the occlusions between the layers [8].

A generative probabilistic layered image model is presented by Jojic and Frey [8] and further extended by a number of authors. Each layer in the layered model is described by its appearance and its transparency mask. The sprite appearances are combined using the transparency masks. Various appearance models were proposed: Gaussian per pixel [8], factor analysis [6], index maps [7], Gaussian with local image deformations [9] etc. Various models were also proposed for the transparency maps: Gaussian [8], factor analysis [6], binary mask with local image deformations [9] etc.

Principal component analysis (PCA) and factor analysis (FA) are often used for modelling image data [12, 1]. Both techniques try to find a low dimensional representation of the data by linear projection. Layered model presented by Frey et. al. [6] is using factor analysis for layer appearance and the transparency map. The model can be used to automatically extract low dimensional representation of the moving objects from a video sequence. For example, images of a person walking can be mapped to a 1-dimensional manifold that measures the phase of the persons gait. The FA can be applied to find the low dimensional representation of both the layer transparency masks and the layer appearance. However, the Gaussian based factor analysis does not consider the constraints

imposed on the transparency values. Furthermore, a number of authors noted that more efficient and robust inference can be achieved by using Bernoulli distribution instead of Gaussian for the transparency masks [14, 9].

The relation between the low dimensional representations of the mask and the appearance is complex in general. For example a single-colored object might change its transparency mask, 2D shape, while the appearance remains the same. Therefore in this paper we leave the choice of the appearance model free and focus on the low dimensional representation of the transparency masks. Natural model for the masks is to use the Bernoulli distribution [14, 9]. The "logistic PCA" using Bernoulli distributions was proposed in the machine learning community [11, 10]. The recent study [16] shows that the "binary PCA" is much more accurate than the standard PCA in representing binary image data and probability maps. In this paper we will present a layered model where the "logistic PCA" is used for the transparency masks. The presented model can be used to automatically extract low dimensional representation of the moving objects transparency mask (shape) from a video sequence more efficiently and robustly.

This paper is organized as follows. In Section 2 we describe the layered image model. In Section 3 the model is extended by including the "logistic PCA" transparency masks. In Sections 4 we explain a generalized expectation maximization (EM) inference scheme for the model. In Section 5 we present experimental results, and in Section 6 we list our conclusions and some topics for further research.

## 2  Layered image model

In the layered model an image $\mathbf{x}$ is decomposed into a set of $L$ layers corresponding to objects that occlude each other. Each layer is described by its appearance parameters $\Lambda_l$ and the transparency mask $\mathbf{m}_l$.

The transparency mask describes which part of the image is covered by the object. The mask value for the $l$-th layer and $d$-th pixel will be denoted by $\mathbf{m}_{dl} \in \{0,1\}$. A natural way to model the mask data is using Bernoulli distributions:

$$p(\mathbf{m}_{dl}|\alpha_l) = \alpha_{dl}^{\mathbf{m}_{dl}} \bar{\alpha}_{dl}^{\bar{\mathbf{m}}_{dl}} \tag{1}$$

where $\alpha_{dl}$ is the probability that $\mathbf{m}_{dl} = 1$, $\bar{\mathbf{m}}_{dl} = 1 - \mathbf{m}_{dl}$ and $\bar{\alpha}_{dl} = 1 - \alpha_{dl}$.

The appearance model describes the pixel values. The probability of the $d$-th image pixel value $\mathbf{x}_d$ for the $l$-th layer is given by $p(\mathbf{x}_d; \Lambda_l)$. The pixel value $\mathbf{x}_d$ is for example the 3 dimensional RGB value. In this paper we will use a simple appearance model, similar to [8], consisting of a Gaussian per pixel

$$p(\mathbf{x}_d; \Lambda_l) = \mathcal{N}(\mathbf{x}_d; \mu_{dl}, \Psi_{dl}I) \tag{2}$$

where the covariance matrix is isotropic $\Psi_{dl}I$ and $I$ is a $3 \times 3$ identity matrix.

Assuming the pixel values to be independent an image is described by:

$$\prod_d p(\mathbf{x}_d, \mathbf{m}_{d1}, ..., \mathbf{m}_{dL}|\Omega) \tag{3}$$

where $\Omega = \{\Lambda_1, ..., \Lambda_L, \alpha_1, ..., \alpha_L\}$ are the parameters of the model. The unobserved mask variables $\mathbf{m}_{d1}, ..\mathbf{m}_{dL}$ determine which pixels belong to which objects/layers as described further.

To alow the objects to switch between layers an additional discrete labelling variable $c$ needs to be included which assigns objects to different layers, see [8]. For simplicity here we will assume that each object stays always in the same layer.

The layered model per pixel $p(\mathbf{x}_d, \mathbf{m}_{d1}, ..., \mathbf{m}_{dL}|\Omega)$ can be written using recursive equation:

$$p(\mathbf{x}_d, \mathbf{m}_{dl}, ..., \mathbf{m}_{dL}, \mathbf{o}_{dl}) = p(\mathbf{x}_d; \Lambda_l)^{\mathbf{m}_{dl}\mathbf{o}_{dl}} p(\mathbf{x}_d, \mathbf{m}_{dl+1}, ..., \mathbf{m}_{dL}, \mathbf{o}_{dl+1}) p(\mathbf{m}_{dl}|\alpha_l)^{\mathbf{o}_{dl}} \qquad (4)$$

where $\mathbf{o}_{dl} = \prod_1^{l-1} \bar{\mathbf{m}}_{dl}$ is the occlusion of the $d$-th pixel by the previous layers closer to the camera. The equation describes stacking the layers on top of each other with background layer $l = L$ at the bottom. In other words a pixel value $\mathbf{x}_d$ can be explained by the $l$-th layer appearance model $p(\mathbf{x}_d; \Lambda_l)$ if it is not occluded and already modelled by the previous layers $1, ..., l-1$, i.e. $\mathbf{o}_{dl} = 0$, and if the current layer mask $\mathbf{m}_{dl} = 1$. If the pixel value is not described by the current and the previous layers, i.e. $\mathbf{o}_{dl} = 0$ and $\mathbf{m}_{dl} = 0$, than it is described by the layers that lie below $p(\mathbf{x}_d, \mathbf{m}_{dl+1}, ..., \mathbf{m}_{dL}, \mathbf{o}_{dl})$. For simplicity the background layer $l = L$ will be without the mask $p(\mathbf{x}_d, \mathbf{m}_{dL}, \mathbf{o}_{dl}) = p(\mathbf{x}_d; \Lambda_L)^{\mathbf{o}_{dL}}$. For the top layer there are no occlusions $p(\mathbf{x}_d, \mathbf{m}_{dl}, ..., \mathbf{m}_{dL}|\Omega) = p(\mathbf{x}_d, \mathbf{m}_{d1}, ..., \mathbf{m}_{dL}, \mathbf{o}_{dl-1} \equiv 0)$.

A common extension is to include unknown layer transformation function $\mathscr{T}_{tl}$, for example translation, rotation, scaling etc. The transformation $\mathscr{T}_{tl}$ transforms the layer $l$ before it is combined with other images. We will denote the transformed appearance parameters by $\mathscr{T}_{tl}\Lambda_l$ and the transformed mask parameters as $\mathscr{T}_{tl}\alpha_l$. The recursive equation per pixel and per layer becomes:

$$p(\mathbf{x}_d, \mathbf{m}_{dl}, ..., \mathbf{m}_{dL}, \mathbf{o}_{dl}) =$$
$$p(\mathbf{x}_d; \mathscr{T}_{tl}\Lambda_l)^{\mathbf{m}_{dl}\mathbf{o}_{dl}} p(\mathbf{x}_d, \mathbf{m}_{dl+1}, ..., \mathbf{m}_{dL}, \mathbf{o}_{dl+1}) p(\mathbf{m}_{dl}|\mathscr{T}_{tl}\alpha_l)^{\mathbf{o}_{dl}} p(\mathscr{T}_{tl}) \qquad (5)$$

where $\mathscr{T}_{t1}, ..., \mathscr{T}_{tL}$ are additional unobserved variables. As in [8] we consider a discrete set of transformations $\mathscr{T}_{tl} \in \{\mathscr{T}_1, ..., \mathscr{T}_T\}$ and the prior distribution over the transformations is denoted as $p(\mathscr{T}_{tl}) = p_{tl}$.

## 3 Logistic PCA masks

We would like to design a layered image model to automatically extract low dimensional representation of the moving objects transparency masks. For example, images of a person walking can be mapped to a 1-dimensional manifold that measures the phase of the persons gait, see Figure 1. Principal component analysis (PCA) commonly used to find a low dimensional representation of the data by linear projection. We describe here a similar model but for Bernoulli distributions, the so called "logistic PCA", to describe the transparency masks of the layered model from the previous section. As in [10] instead of the $\alpha_{dl}$ in (1) for the Bernoulli mask model we will use the log-odds parameter $\Theta_{dl} = \log(\alpha_{dl}/(1 - \alpha_{dl}))$ and the logistic function $\sigma(\Theta_{dl}) = (1 + e^{-\Theta_{dl}})^{-1}$. The mask model can be written equivalently as:

$$p(\mathbf{m}_{dl}|\Theta_{dl}) = \sigma(\Theta_{dl})^{\mathbf{m}_{dl}} \sigma(-\Theta_{dl})^{\bar{\mathbf{m}}_{dl}} \qquad (6)$$

Logistic PCA assumes that the log-odds mask parameter $\Theta_l$ is given by the so called "mean" log-odds mask parameter $\Delta_l$ plus a linear combination of $S \ll D$ basis vectors
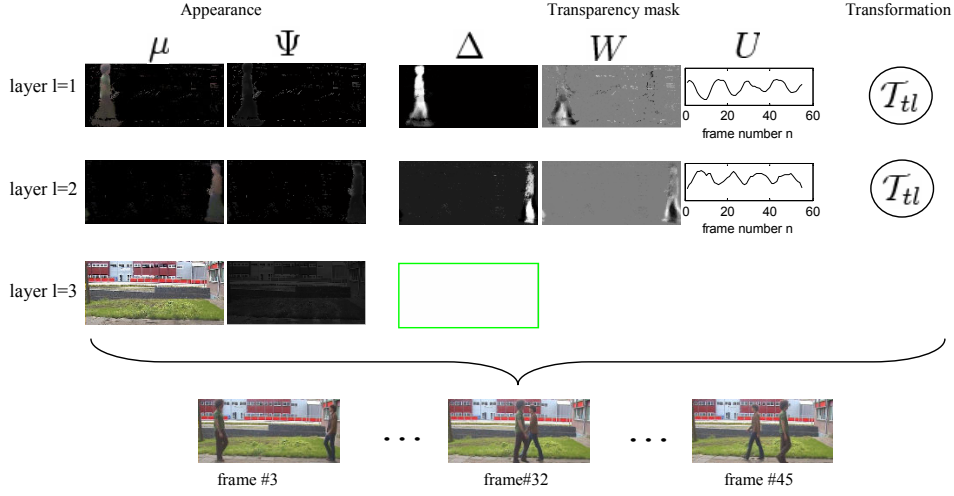
Figure 1: The parameters of the layered model learned automatically from the image sequence of two people walking in opposite directions and occluding each other.

(images) contained in the rows of the $S \times D$ matrix $W_l$. The linear combination is obtained through the coefficients contained in $U_l$:

$$\Theta_{dl} = \Delta_{dl} + \sum_s U_{sl} W_{sdl}. \tag{7}$$

The $\Delta_l, W_l$ and $U_l$ are the parameters of the logistic PCA.

# 4   Learning model parameters

The layered image model with the binary PCA for a set of $N$ independent images can be written as:

$$\prod_{nd} p(\mathbf{x}_{nd}, \mathbf{m}_{nd1}, ..., \mathbf{m}_{ndL}, \mathscr{T}_{nt1}, ..., \mathscr{T}_{ntL} | \Omega) \tag{8}$$

where the masks $\mathbf{m}_{nd1}, ..\mathbf{m}_{ndL}$ and the transformations $\mathscr{T}_{nt1}, ..., \mathscr{T}_{ntL}$ are the unobserved variables and $\Omega = \{\Lambda_1, ..., \Lambda_L, \Delta_l, ..., \Delta_L, W_1, ..., W_L, U_1, ..., U_L, p_{tl}\}$ are the parameters of the model. The index $n$ indicates that there is each layer can have a different appearance mask $\mathbf{m}_{ndl}$ and a different transformation $\mathscr{T}_{ntl}$ for each image. The log-likelihood of a given set of $N$ images is given by:

$$\mathscr{L}(\Omega) = \sum_{nd} \ln p(\mathbf{x}_{nd} | \Omega) \tag{9}$$

where the unknown masks and transformations are integrated out:

$$p(\mathbf{x}_{nd} | \Omega) = \sum_{\text{all masks and transf.}} p(\mathbf{x}_{nd}, \mathbf{m}_{nd1}, ..., \mathbf{m}_{ndL}, \mathscr{T}_{nt1}, ..., \mathscr{T}_{ntL} | \Omega) \tag{10}$$

The goal is to find the parameters $\Omega$ that maximize the log-likelihood (9).

## 4.1  Approximate inference

The log-likelihood is a complex function. The EM algorithm [3] presents an iterative solution but computing it would be intractable for such a model [8]. Therefore as in [8] we use a variational approximate method. We will denote the hidden variables by $h$, layer masks and hidden transformations in our case. Variational techniques replace the intractable computation of the posterior distribution $p(h|\mathbf{x})$ with a search for a simplified distribution $q(h)$, that is made close to $p(h|\mathbf{x})$ by minimizing the "free energy" function:

$$F = \int_h q(h) \frac{p(h)}{\ln p(\mathbf{x},h|\Omega)} \geq -\mathscr{L}(\Omega) \tag{11}$$

Minimizing $F$ w.r.t. $q(h)$ minimizes the relative entropy between $q(h)$ and $p(h|\mathbf{x})$. Minimizing $F$ w.r.t. $q(h)$ and the model parameters $\Omega$ minimizes an upper bound on the negative log-likelihood of the data $\mathscr{L}(\Omega)$ [5].

Similar to [8] we will use the following simplified factorized distribution:

$$q(h) = \prod_{ndl} \mathbf{r}_{ndl}{}^{\mathbf{m}_{ndl}} \bar{\mathbf{r}}_{ndl}^{\bar{\mathbf{m}}_{ndl}} \mathbf{q}_{ntl} \tag{12}$$

The parameter estimation is then performed iteratively using a generalized EM algorithm steps:

E step:    Minimizing $F$ w.r.t. the variational    (13)

parameters $\mathbf{r}_{ndl}$ and $\mathbf{q}_{ntl}$.    (14)

M step:    Minimizing $F$ w.r.t. $\Omega$.    (15)

These two steps are repeated iteratively until convergence. See [5] for a tutorial.

## 4.2  Updating mask parameters

The update equations for the E and M steps above are already given in the various extensions [6, 9] of the initial work by Jojic and Frey [8]. An extensive tutorial can be found also in [5]. Therefore, and because of the limited space, we will not repeat all the update equations. Instead we will focus on the extension proposed in this paper: the logistic PCA model applied to the transparency masks and the update equations for the logistic PCA parameters $\Delta$, $W$ and $U$.

The variational parameters are updated in the E-step. The layer appearance parameters are updated in the M-step. In the M step we need also to minimize the free energy function $F$ w.r.t. the logistic PCA parameters $\Delta$, $W$ and $U$. It can be shown, see (5) and (12), that the only part of the function $F$ that depends on the layer $l$ mask parameters is given by:

$$\sum_{n,t,d} \mathbf{q}_{ntl}(\mathbf{w}_{dl}\mathbf{r}_{ndl}\log(\mathscr{T}_{ntl}\alpha_{dl}) +$$

$$\mathbf{w}_{dl}\bar{\mathbf{r}}_{ndl}\log(\mathscr{T}_{ntl}\bar{\alpha}_{dl})) \tag{16}$$

where $\mathbf{w}_{dl} = \prod_1^{l-1} \bar{\mathbf{r}}_{ndl}$. So in order to minimize the free energy function $F$ w.r.t. the logistic PCA parameters $\Delta$, $W$ and $U$ for each layer we need to consider only these terms.

For simplicity as previously in the similar models [4] we assume the transformation $\mathcal{T}_{ntl}$ to be a permutation matrix that rearranges the pixels. For example, to account for all translations in a $J \times J$ image, $\mathcal{T}_{ntl}$ can take on $J^2$ values the permutation matrices that account for all discrete translations. The discrete 2D image translation is a common transformation to align images. Furthermore, an efficient solution for the E step is available [4]. Furthermore, note that by transforming an image to log-polar coordinates, shifts correspond to rotations and scalings [15]. Let $\mathcal{T}_{ntl}^{-1}$ denote the inverse transformation. The terms above (16) can be rewritten in the following form:

$$\sum_{n,t,d} \mathbf{q}_{ntl} (\mathcal{T}_{ntl}^{-1} \mathbf{w}_{dl} \mathbf{r}_{ndl} \log(\alpha_{dl}) +$$
$$\mathcal{T}_{ntl}^{-1} \mathbf{w}_{dl} \bar{\mathbf{r}}_{ndl} \log(\bar{\alpha}_{dl})) \tag{17}$$

After integrating over all possible transformations we get:

$$\sum_{n,d} \widehat{\mathbf{wr}_{ndl}} \log(\alpha_{dl}) + \widehat{\bar{\mathbf{wr}}_{dl}} \log(\bar{\alpha}_{dl}) \tag{18}$$

where:

$$\widehat{\mathbf{wr}_{ndl}} = \sum_t \mathbf{q}_{ntl} \mathcal{T}_{ntl}^{-1} \mathbf{w}_{dl} \mathbf{r}_{ndl} \tag{19}$$

$$\widehat{\bar{\mathbf{wr}}_{dl}} = \sum_t \mathbf{q}_{ntl} \mathcal{T}_{ntl}^{-1} \mathbf{w}_{dl} \bar{\mathbf{r}}_{ndl}. \tag{20}$$

Finally this can be written using the log-odds parameters $\Theta_{dl} = \log(\alpha_{dl}/(1 - \alpha_{dl}))$ as:

$$\omega_{ndl} (\widehat{M_{ndl}} \Theta_{ndl} + \log \sigma(-\Theta_{ndl})) \tag{21}$$

where

$$\omega_{ndl} = (\widehat{\mathbf{wr}_{ndl}} + \widehat{\bar{\mathbf{wr}}_{dl}}) \text{ and} \tag{22}$$
$$\widehat{M_{ndl}} = \widehat{\mathbf{wr}_{ndl}} / \omega_{ndl}. \tag{23}$$

Note that $\widehat{M_{ndl}} \in [0,1]$. If we consider $\widehat{M_{ndl}}$ as data then (21) presents log-likelihood under Bernoulli model with the log-odds parameters $\Theta_{ndl}$. Additionally each data point is weighted by its weight $\omega_{ndl}$. The goal in the M-step is to find the logistic PCA parameters $\Delta_l$, $W_l$ and $U_l$ that maximize the weighted log-likelihood (21). The maximum can not be found in closed form. There exist an efficient iterative procedure for the logistic PCA [10]. The procedure can be extended for the weighted case (21). For completeness of the text the iterative update equations for the weighted logistic PCA are given in Appendix A.

## 4.3  Practical algorithm

For the sake of clarity we summarize the practical algorithm:

**Initialization**: In case of a static background the background layer appearance parameters can be initialized by the mean value and the variance of each pixel for the whole sequence. The other layer appearances can be initialized by arbitrary mean and some

large variance. For the masks we initialize the logistic parameter $\Delta_l$ by some small random values around zero. Within first few iterations it is useful to update the parameters for each layer separately starting from the background layer and going upwards. This is similar to the greedy layer estimation presented in [14]. Furthermore, for the first few iterations we keep the basis vectors of the logistic PCA $W$ and the coefficients $U$ to zero and then initialize them by some small random values, for example sampled from a zero mean Gaussian distribution with the standard deviation 0.001.

**1:** For each image calculate the approximation of the posterior distribution by maximizing $F$ w.r.t. the variational parameters $\mathbf{r}_{ndl}$ and $\mathbf{q}_{ntl}$, see [5].

**2:** For each image calculate $\omega_{ndl}$ and $\widetilde{M}_{ndl}$.

**3:** Update the appearance parameters $\Lambda_l$ and if required $p_t$.

**4:** Update the logistic PCA mask parameter estimates $\Delta_l$, $W_l$ and $U_l$ using the update equations from Appendix A. The updated $\Delta_l$, $W_l$ and $U_l$ will not maximize the free energy function $F$ but they will increase its value. This can be seen as a "generalized M step" [5].

**5:** Stop if increase of the free energy function $F$ is below some threshold, otherwise go to **1**.

There is often not enough data to estimate $p_t$ reliably and we will use in this paper a uniform prior distribution over the transformations: $p_t = 1/D$ [8].

## 5 Experiments

### 5.1 Extracting low dimensional representations

To demonstrate how the layered model can automatically extract low dimensional representation of the transparency maps of the moving objects from a video sequences we recorded a 55 frame sequence of two people walking into opposite directions and occluding each other during the sequence. In Figure 1 we present the model parameters automatically learned from the sequence. Note that $U$ nicely captures the cyclical walking motion while $W$ models the corresponding deformations.

### 5.2 Reconstructing transparency maps

In order to compare the quality of the Gaussian and Bernoulli based models we conducted the following experiments. We used a sequence captured by a surveillance camera. The camera was observing people walking in front of a static background. The sequence contains 400 frames. There were 5 people present in the sequence. Only single person was present per frame. Therefore we constructed a model consisting of 2 layers. The appearance of the front layer was modelled by a Gaussian mixture with 5 components to accommodate for the 5 different people. The transparency mask is modelled by 5 component logistic PCA. The parameters learned from the sequence are presented in Figure 2. The learned Gaussian mixture components nicely correspond to the 5 different people present in the video. The components of the logistic PCA presented at the bottom row in Figure 2 seem to capture the walking deformations and also the different walking directions. The first 2 people were observed walking in both directions and this is clearly visible in their appearance parameters.
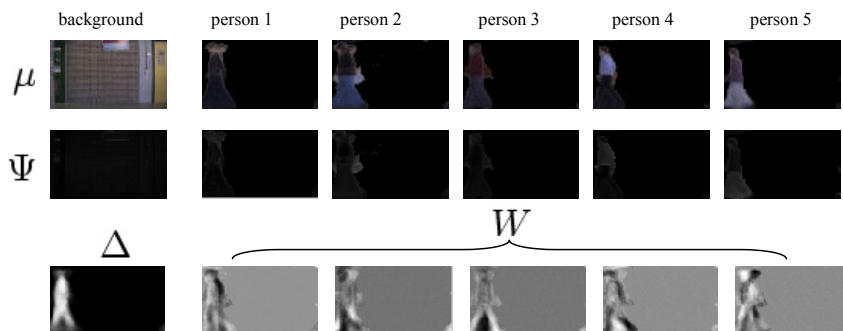
Figure 2: The parameters of the model learned from a 400 frames sequence containing 5 walking people. Only single person was present per frame. The appearances of the 5 people and the static background are presented at the top rows. The first 2 people were observed walking in both directions. At the bottom we present the parameters of the logistic PCA ($S = 5$ components) that is used to model the transparency maps segmenting the people from the background. The mean and the 5 basis images are shown.

Once the model is constructed we used additional 50 images to test the quality of the model. The ground truth segmentation of the 50 additional images is obtained by manually segmenting the persons from the background. Using the model, we compress the images to the PCA scores for the transparency masks. We use then the model to project the PCA scores back to mask images. Finally, we use most likely transformation $argmax(\mathbf{q}_{ntl})$ to shift the reconstructed mask to the proper position. The mask reconstructed by the layered model will be denoted $\hat{X}_n$. See some examples in Table 5.2. We then measure the difference between the manually segmented image and the segmentation using the layered model. We measured the error in three ways. (i) Quadratic loss: the sum of the squared differences per pixel value, $e_2 = (1/D)\sum_d (X_{nd} - \hat{X}_{nd})^2$. (ii)logistic loss: the sum of the log-likelihood of the ground truth masks given the reconstructions, $e_{log} = 1/D\sum_d X_{nd}\ln\hat{X}_{nd} + (1 - X_{nd})\ln(1 - \hat{X}_{nd})$. As the reconstruction from the Gaussian model can be outside $(0,1)$, we first map values outside this interval to $\varepsilon = 10^{-6}$ and $1 - epsilon$ respectively.(iii) Zero-one loss: first we threshold the segmentation by the model at $\hat{X}_{nd} > 1/2$ to get a binary reconstruction $\hat{X}_n^{01}$, then we measure the number of pixels that differ from the ground truth, $e_{01} = (1/D)\sum_d |X_{nd} - \hat{X}_{nd}^{01}|$.

The results for $S = 5$ components are reported in Table 5.2. We also constructed a layered model similar to [6] where we used the Gaussian probabilistic PCA to model the transparency masks. Clearly, the layered model using logistic PCA leads to big improvements. This is also visible in Figure 5.2.

Since the camera was static, in Table 5.2 we also show the results obtained using standard background subtraction scheme [2] which builds a model only for the background layer. The layered model which considers all layers leads to much better results, see Table 5.2. Another common technique to improve segmentation results after background subtraction is to apply some morphological operators on the segmentation results. We used image closing operator with a $3 \times 3$ element. The results improve slightly but the layered model is still superior. When a larger template is used for image closing of when image opening is performed, the results get only worse.

| | | |
|---|---|---|
| original | |  |
| manual | |  |
| background subtraction [2] | |  |
| background subtraction + image closing | |  |
| layered model + normal PCA | |  |
| layered model + logistic PCA | |  |

Table 1: Segmentation examples for various approaches.

| | $e_2$ | $e_{01}$ | $e_{log}$ |
|---|---|---|---|
| layered model + logistic PCA ($S = 5$ components) | 0.023 (0.006) | 0.03 (0.01) | 0.086 (0.023) |
| layered model + norm. PCA ($S = 5$ components) | 0.029 (0.006) | 0.04 (0.01) | 0.098 (0.023) |
| background subtraction + image closing | 0.052 (0.017) | 0.052 (0.017) | 0.64 (0.23) |
| background subtraction | 0.059 (0.024) | 0.059 (0.024) | 0.66 (0.28) |

Table 2: Segmentation error w.r.t. manually segmented images of the walking people sequence. The mean error per pixel over 50 hand-segmented images is reported. The standard deviation over images is reported within the brackets.

# 6 Conclusions

The generative probabilistic layered image model presented by Jojic and Frey [8] was further extended by a number of authors. We focus here on modelling the transparency masks. The natural way to model the masks is to use Bernoulli distributions. We presented probabilistic layered image model that models the masks using Bernoulli distributions and extracts the low dimensional representation of the transparency masks using the "logistic PCA" [10]. Gaussian component and factor analysis as in [6] does not take into account that the transparency mask has values limited between [0,1]. The logistic PCA describes the mask more naturally and leads to crisper masks and better segmentation results as we demonstrated.

A disadvantage of the logistic PCA is that it requires solving two $S \times S$ linear systems for each data point (see Appendix) which might be prohibitive if the number of components $S$ is large. Furthermore, projecting data to the low-dimensional PCA space requires iterations in the case of logistic PCA, while for normal PCA the projection is linear. Finally, the logistic PCA used here is not a full generative model as there is no prior distribution on the low-dimensional coefficient matrix $U$. A computationally slightly more expensive model which incorporates Gaussian priors on $U$ is described in [11].

# Appendix I: Weighted Binary PCA update equations

**U-update:** First intermediate quantities are computed:

$$H_{nd} = \omega_{nd}\Theta_{nd}^{-1}\tanh(\Theta_{nd}/2) \; , \; A_{nss'} = \Sigma_d H_{nd}W_{sd}W_{s'd} \text{ and} \qquad (24)$$

$$B_{ns} = \Sigma_d(2\omega_{nd}\widehat{M}_{nd} - 1 - H_{nd}\mu_d)W_{sd} \qquad (25)$$

Row $n$ of $U$ is computed by solving linear system: $\Sigma_{s'}A_{nss'}U_{ns'} = B_{ns}$.

**W-update:** First intermediate quantities are computed:

$$A_{dss'} = \Sigma_n H_{nd}U_{ns}U_{ns'} \text{ and } B_{ds} = \Sigma_n(2\omega_{nd}\widehat{M}_{nd} - 1 - H_{nd}\mu_d)U_{ns} \qquad (26)$$

Column $d$ of $W$ is computed by solving the linear system $\Sigma_{s'}A_{dss'}W_{s'd} = B_{dl}$.

$$\Delta-\textbf{update}: \Delta_{nd} = (\Sigma_n H_{nd})^{-1}\Sigma_n(2\omega_{nd}\widehat{M}_{nd} - 1 - H_{nd}(UV)_{nd}) \qquad (27)$$

# References

[1] M. Black and A. Jepson. EigenTracking: Robust matching and tracking of articulated objects using a view-based representation. *Int. Journal of Computer Vision*, 26(1):63–84, 1998.

[2] C.Stauffer and W.Grimson. Adaptive background mixture models for real-time tracking. *In Proc. of the Conf. on Computer Vision and Pattern Recognition*, 1999.

[3] A. Dempster, N. Laird, and D. Rubin. Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Stat. Society, Series B (Methodological)*, 1(39):1–38, 1977.

[4] B. Frey and N. Jojic. Transformation-invariant clustering using the EM algorithm. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 25:1–17, 2003.

[5] B. Frey and N. Jojic. A comparison of algorithms for inference and learning in probabilistic graphical models. *IEEE Trans. on Pattern Analysis and Machine Intel.*, 27(9), 2005.

[6] B. Frey, N. Jojic, and A. Kannan. Layered density models and unsupervised video analysis. *In Proc. IEEE Conf. on Computer Vision and Pattern Recognition*, 2003.

[7] N. Jojic and Y. Caspi. Capturing image structure with probabilistic index maps. *In Proc. IEEE Conf. on Computer Vision and Pattern Recognition*, 2004.

[8] N. Jojic and B. Frey. Learning flexible sprites in video layers. *In Proc. IEEE Conf. on Computer Vision and Pattern Recognition*, 2001.

[9] A. Kannan, N. Jojic, and B. Frey. Layers of appearance and deformation. *10th Int. Workshop on Artificial Intelligence and Statistics (AISTATS)*, 2005.

[10] A. Schein, L. Saul, and L. Ungar. A generalized linear model for principal component analysis of binary data. *In Proc. Int. Workshop on Art. Intel. and Statistics*, pages 14–21, 2003.

[11] M. Tipping. Probabilistic visualisation of high-dimensional binary data. *In Advances in Neural Information Processing Systems (NIPS)*, 1999.

[12] M. Turk and A. Pentland. Eigenfaces for recognition. *Journal of Cognitive Neuroscience*, 3:71–86, 1991.

[13] J. Wang and E. Adelson. Layered representation for motion analysis. *In Proc. IEEE Conf. on Computer Vision and Pattern Recognition*, pages 361–366, 1993.

[14] C. Williams and M. Titsias. Learning about multiple objects in images: Factorial learning without factorial search. *In Advances in Neural Information Processing Systems (NIPS)*, 2003.

[15] G. Wolberg and S. Zokai. Robust image registration using log-polar transform. *In Proc. IEEE Int. Conf. image processing, vol. 1*, pages 493–496, 2000.

[16] Z.Zivkovic and J. Verbeek. Transformation invariant component analysis for binary images. *In Proc.IEEE Conference on Computer Vision and Pattern Recognition*, 2006.